

# Thèse de Doctorat

Mohammad ABUOUEDEH

*Mémoire présenté en vue de l'obtention du  
grade de Docteur de l'Université de Nantes  
sous le sceau de l'Université Bretagne Loire*

**École doctorale :** *Education, Langages, Interaction, Cognition, Clinique (ELICC)*

**Discipline :** *Sciences du Langage : linguistique et phonétique générale*

**Spécialité :** *Phonétique*

**Unité de recherche :** *Laboratoire de Linguistique de Nantes (LLING)*

**Soutenu le** 26 Juin 2018

## De l'impact des variations temporelles sur les transitions formantiques

### JURY

Présidente :	<b>Sabrina BENDJABALLAH</b> , Directrice de Recherche CNRS (HDR), Université de Nantes, France
Rapporteurs :	<b>Rachid RIDOUANE</b> , Directeur de Recherche CNRS (HDR), Université Sorbonne Nouvelle, Paris 3, France <b>Nathalie VALLEE</b> , Chargée de Recherche CNRS (HDR), Université Grenoble Alpes, France
Examineurs :	<b>Eleonora CAVALCANTE ALBANO</b> , Professeure des universités, Universidade Estadual de Campinas, Brésil <b>Sabrina BENDJABALLAH</b> , Directrice de Recherche CNRS (HDR), Université de Nantes, France
Invité :	<b>Jalal-Eddin AL-TAMIMI</b> , University Lecturer, University of Newcastle, Royaume-Uni
Directeur de Thèse :	<b>Orin PERCUS</b> , Professeur des universités (HDR), Université de Nantes, France
Co-encadrant de Thèse :	<b>Olivier CROUZET</b> , Maître de conférence, Université de Nantes, France



## THÈSE DE DOCTORAT

**Mohammad ABUOUEH**

Mémoire présenté en vue de l'obtention du  
**grade de Docteur de l'Université de Nantes**  
sous le seau de l'Université Bretagne Loire

**École doctorale** : Éducation, langages, Interaction, Cognition, Clinique (ELICC)

**Discipline** : Sciences du Langage : Linguistique et phonétique générale

**Spécialité** : Phonétique

**Unité de Recherche** : Laboratoire de linguistique de Nantes (LLING)

**Soutenu le** : 26 Juin 2018

# De l'impact des variations temporelles sur les transitions formantiques

Jury :

- Présidente** : Sabrina BENDJABALLAH, Directrice de Recherche CNRS (HDR), Université de Nantes, France
- Rapporteurs** : Rachid RIDOUANE, Directeur de Recherche CNRS (HDR), Université Sorbonne Nouvelle, Paris 3, France  
Nathalie VALLEE, Chargée de Recherche CNRS (HDR), Université Grenoble Alpes, France
- Examineurs** : Eleonora C. ALBANO, Professeure des universités, Univeridade Estadual de Campinas, Brésil  
Sabrina BENDJABALLAH, Directrice de Recherche CNRS (HDR), Université de Nantes, France
- Invité** : Jalal-Eddin AL-TAMIMI, University Lecturer, University of Newcastle, Royaume -Uni
- Directeur de Thèse** : Orin PERCUS, Professeur des universités (HDR), Université de Nantes, France
- Co-encadrant** : Olivier CROUZET, Maître de conférence, Université de Nantes, France



**À mon père**



# Résumé

Cette étude permet d'explorer les variations spectrales provoquées par deux types de changements temporels qui résultent respectivement de l'opposition de longueur vocalique (variations intrinsèques) et des variations du débit de parole (variations extrinsèques). Trois protocoles expérimentaux ont été conçus, deux en arabe jordanien et un en français, pour examiner ce phénomène. Un intérêt particulier a été porté aux occlusives produites dans des séquences CVC dans le but d'évaluer la coarticulation de la consonne en position initiale et finale. La durée des voyelles et la fréquence des trois premiers formants au début ou à la fin et au milieu de chaque séquence ont été mesurées dans chaque condition de longueur et de débit. Les équations de locus ont été utilisées afin de décrire la relation CV/VC quand elle subit ces deux types de variations. Selon les résultats, la qualité de la voyelle et de la consonne est influencée par l'opposition de durée et par le débit de parole de manière distincte. Le changement généré par les variations temporelles est évalué à l'aide des équations de locus et à travers des analyses de dispersion. Ces outils révèlent qu'il existe un chevauchement coarticulatoire plus important quand la durée de la voyelle est réduite, peu importe l'origine des variations temporelles dans les deux langues (intrinsèque et extrinsèque). Par contre, les variations temporelles impactent la classification des occlusives de manière distincte selon leur origine.

**Mots clés :** variations temporelles, débit de parole, longueur vocalique, coarticulation, équations de locus, analyses de dispersion, arabe jordanien, français, analyses acoustiques, phonétique, classification des occlusives.



# Abstract

The aim of this study is to examine the different spectral variations that are impacted by two types of temporal modifications : phonological vowel length contrast (intrinsic variations) and varying speaking rates (extrinsic variations). Three experiments, two on Jordanian Arabic and one on French, were conducted to investigate this phenomenon. We were interested in stop consonants produced in the initial and the final positions in CVC sequences. Vowel duration and the frequency of the first three formants were measured for each vowel length and speaking rate. In addition, locus equations were computed to measure the impact of time variations on CV / VC coarticulation. According to the results, the vowel and consonant qualities are influenced both by vowel duration contrasts and by speaking rate modifications in different ways. Such modifications are evidenced by locus equations and by dispersion analyses as well. These two quantification tools revealed that there is greater coarticulatory overlap when the duration of the vowel is reduced, regardless of the origin of temporal variations in the two languages (intrinsic and extrinsic). On the other hand, temporal variations affect the classification of the stop consonant differently according to their origin.

**Key Words :** temporal variations, speaking rate, vowel length, coarticulation, locus equations, dispersion analyses, Jordanian Arabic, French, acoustical analyses, phonetics, stop consonant classification.



# Remerciements

Tout d'abord, je tiens à remercier toutes les personnes qui m'ont aidé durant l'élaboration de cette thèse.

Je remercie notamment Orin Percus d'avoir accepté d'être mon directeur et de m'avoir suivi tout au long de ces années. Sans sa gentillesse, sans sa générosité, je n'aurais pas pu commencer ce voyage scientifique.

Je remercie également chaleureusement la personne qui m'a inspiré, conseillé et guidé pendant toutes les phases de la thèse, mon co-encadrant Olivier Crouzet. Ce manuscrit n'aurait pas pu voir le jour sans ses idées, son intérêt et son soutien. Un grand merci pour ses relectures, ses avis constructifs et son aide lors la rédaction.

Mes remerciements vont également aux membres du jury, Nathalie Vallée, Rachid Ridouane, Eleonora Albano, Sabrina Bendjaballah et Jalal-Eddin Al-Tamimi, pour avoir accepté d'évaluer ce travail et pour le temps qu'ils ont consacré à la lecture de ce manuscrit.

J'adresse mes remerciements aux membres permanents du LLING pour leur bonne humeur et pour toutes les connaissances qu'ils m'ont transmises : Hamida Demirdache pour m'avoir facilité la vie à plusieurs reprises, Jean-Pierre Angoujard pour son suivi dans la progression de la thèse, Ali Tifrit pour les pauses café et son humour, Anne Croll pour son sourire et ses conversations, Anamaria Falas pour sa bienveillance et sa gentillesse et Jean-Marc Beltzung pour ses discussions. Je suis reconnaissant à Myriam Le Coz, gestionnaire administrative et financière du LLING, toujours disponible et prête à m'aider concernant les questions administratives.

Je ne sais pas comment remercier Sarah Grenouilleau pour tout ce qu'elle a fait pour moi. Elle était là pour illuminer mon chemin quand il faisait nuit, elle m'a soutenu et cru en moi pendant les moments les plus difficiles, je la remercie de tout mon cœur. Son aide à la mise au net des figures et ses nombreuses relectures de la thèse m'ont également été précieuses. Sans son aide, ses corrections, jour et nuit, la thèse n'aurait pas abouti à cet état. Les moments heu-

reux partagés ensemble et ses beaux sourires m'ont toujours donné le courage et la force de continuer.

Je remercie mes amis et mes collègues du LLING, les doctorants et les ex-doctorants, qui ont su créer une ambiance idéale pour la rédaction de la thèse, particulièrement ceux qui n'ont pas seulement partagé le bureau des doctorants, mais également les moments durs et moins durs, joyeux et moins joyeux avec moi. Merci à Jieun Bark d'avoir été là à l'écoute, pour son sourire et son soutien, rendant les années de la thèse moins dures, pour les pauses café et les balades à l'intérieur et à l'extérieur du bâtiment Censive. Je remercie aussi Typhanie Prince pour son soutien, son sourire et ses conversations encourageantes et positives ; Laurence Voeltzel pour son humour, ses rires, ses blagues et son soutien qui me mettaient toujours de bonne humeur ; Juliette Angot et Lucas Tual pour leur intérêt, leurs sourires et leur présence et enfin, Ashraf Allawama, Oana Lungu, Hongyuan Sun, Jiyoung Choi, Natasa Knezevic, Jovana Gajić, Julien Millasseau, Kryzzya Gomez, Emmanuella Annan, Samantha Becerra et Elizabeth Heredia.

Mes pensées sont aussi tournées vers mes collègues des bâtiments Censive et Tertre avec les pauses café, les déjeuners au TU, leur sourire et leur gentillesse : Caroline Biron, Lolita Rousseau, Stan Bossard, Matteo Deroma, Fanny Prou et Pauline Peter.

Je n'oublie pas non plus les locuteurs de l'arabe jordanien et du français pour leur aide dans les enregistrements des trois expériences.

Enfin, je remercie celles et ceux qui me sont chers que j'ai quelque peu délaissés ces derniers mois pour achever cette thèse, plus particulièrement ma famille, mes amis en Jordanie et mes amis à Nantes. Leurs attentions et encouragements m'ont accompagné au cours de ces années. Je suis reconnaissant à ma mère, mes sœurs Iqbal, Anfal et Ibtihal, mes frères Hakim, Yousef et Aid, mes neveux et nièces Farah, Nour, Issa, Liliane et Malak et mes cousins Sha-her Abuoudeh, Khamis Al-Khattab et Adel Al-Khattab pour leur soutien dans les moments difficiles. Les encouragements et le soutien de mes amis en France et en Jordanie m'ont aussi beaucoup aidé, j'ai une pensée toute particulière à mon ami et colocataire Ahmad Qabbaa, merci pour les moments qu'on a passé ensemble pendant ces années, Madhat et Sylvie Al-Naser, Ibrahim Abkal, Ahmad Ajlouni, Imran Abushamala, Salim Maddah, Ali Aid, Saeed Ben Zekri, Mohammad AbuTaha, Ahmad Al-Dahdouh, Mohammad Abelwahed, Aurore Clément, Thibaud et Maylis Lehugeur, Sophie, Gaël, Alban, Aurore, Lucette et Ludovic Grenouilleau.

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>I Aspects théoriques</b>	<b>5</b>
<b>1 Caractérisation phonétique</b>	<b>7</b>
1.1 Notions de bases . . . . .	7
1.1.1 Le spectrogramme . . . . .	8
1.1.2 Le modèle source-filtre . . . . .	9
1.1.3 Les formants . . . . .	11
1.1.4 Les transitions formantiques . . . . .	13
1.2 Description générale des sons de la parole . . . . .	16
1.2.1 Les voyelles . . . . .	16
1.2.1.1 Le plan F1-F2 . . . . .	17
1.2.1.2 F3 . . . . .	19
1.2.2 La durée vocalique . . . . .	19
1.2.3 Caractéristiques dynamiques de la voyelle . . . . .	33
1.2.4 Les consonnes occlusives . . . . .	40
1.2.4.1 Le bruit de relâchement . . . . .	43
1.2.4.2 Les transitions formantiques des occlusives . . . . .	54
1.2.4.3 Le VOT . . . . .	58
1.2.4.4 La force d'articulation . . . . .	61
1.2.4.5 La position syllabique . . . . .	63
1.3 Invariance et variabilité . . . . .	64
1.4 Conclusion . . . . .	72
<b>2 La coarticulation</b>	<b>73</b>
2.1 Qu'est-ce que la coarticulation ? . . . . .	73
2.2 Approches théoriques de la coarticulation . . . . .	79
2.2.1 La propagation des traits . . . . .	80
2.2.2 La coproduction . . . . .	84

2.2.3	La variabilité adaptative . . . . .	90
2.3	Faits et modèles de la coarticulation . . . . .	93
2.3.1	Le locus . . . . .	93
2.3.2	L' <i>undershoot</i> . . . . .	99
2.3.3	Le modèle d'Öhman . . . . .	103
2.3.4	La résistance coarticulatoire . . . . .	108
2.3.5	Le degré de contrainte articuloire (DAC) . . . . .	113
2.3.6	La théorie H&H . . . . .	119
2.3.7	Les équations de locus . . . . .	122
2.3.8	Le <i>Task-dynamics</i> . . . . .	124
2.4	Conclusion . . . . .	131
<b>3</b>	<b>Les équations de locus</b>	<b>133</b>
3.1	Qu'est-ce que les équations de locus ? . . . . .	133
3.2	Les premières études . . . . .	134
3.3	Outil de mesure de coarticulation . . . . .	137
3.3.1	En style de parole . . . . .	139
3.4	Un descripteur du lieu d'articulation . . . . .	143
3.4.1	Vers une validation des équations de locus . . . . .	150
3.4.1.1	Chez les enfants . . . . .	150
3.4.1.2	Dans d'autres langues . . . . .	154
3.4.1.3	Critiques de l'invariant relationnel . . . . .	157
3.4.1.4	Dans différents modes d'articulation . . . . .	160
3.4.1.5	Dans une articulation compensatoire . . . . .	165
3.4.1.6	Dans un modèle articuloire . . . . .	167
3.5	La position syllabique . . . . .	170
3.6	La relation perceptuelle . . . . .	174
3.6.1	<i>The orderly output constraint</i> . . . . .	179
3.6.2	L'influence du degré de coarticulation sur la catégorisation . . . . .	184
3.7	L'origine articuloire . . . . .	193
3.8	Facteurs influençant le calcul des équations de locus . . . . .	207
3.8.1	Les mesures $F2_{onset}$ . . . . .	208
3.8.2	Les mesures de $F2_{mid}$ . . . . .	210
3.8.3	Les erreurs de mesures . . . . .	212
3.9	Conclusion . . . . .	214
<b>II</b>	<b>Problématiques et aspects expérimentaux</b>	<b>217</b>
<b>4</b>	<b>Problématiques</b>	<b>219</b>

4.1	Un espace consonantique riche . . . . .	220
4.1.1	Caractéristiques acoustiques de l'arabe jordanien . . . . .	221
4.1.1.1	Le système vocalique de l'arabe jordanien . . . . .	221
4.1.1.2	Les occlusives de l'arabe jordanien . . . . .	223
4.2	Variations temporelles intrinsèques et extrinsèques . . . . .	230
4.2.1	Équations de locus et temporalité de la parole . . . . .	232
4.3	Conclusion . . . . .	237
<b>5</b>	<b>La longueur vocalique et la coarticulation</b>	<b>241</b>
5.1	Expérience I . . . . .	241
5.1.1	Participants . . . . .	241
5.1.2	Stimuli . . . . .	242
5.1.3	Procédure . . . . .	243
5.1.3.1	Les anomalies et les valeurs extrêmes . . . . .	244
5.1.3.2	L'estimation de $F2_{@burst}$ . . . . .	246
5.1.4	Résultats . . . . .	249
5.1.4.1	La durée vocalique . . . . .	249
5.1.4.2	Propriétés spectrales des formants . . . . .	252
5.1.4.3	La coarticulation en position initiale (CV) . . . . .	257
5.1.4.4	Les équations de locus (CV) . . . . .	264
5.1.4.5	La coarticulation en position finale (VC) . . . . .	270
5.1.4.6	Les équations de locus (VC) . . . . .	276
5.1.5	Discussion . . . . .	279
5.2	Expérience II . . . . .	283
5.2.1	Participants . . . . .	283
5.2.2	Stimuli . . . . .	284
5.2.3	Procédure . . . . .	284
5.2.4	Résultats . . . . .	284
5.2.4.1	La durée vocalique . . . . .	284
5.2.4.2	Propriétés spectrales des formants . . . . .	286
5.2.4.3	Coarticulation en position initiale (CV) . . . . .	288
5.2.4.4	Les équations de locus (CV) . . . . .	291
5.2.4.5	Coarticulation en position finale (VC) . . . . .	295
5.2.4.6	Les équations de locus (VC) . . . . .	298
5.2.5	Discussion . . . . .	300
5.3	Conclusion . . . . .	302
<b>6</b>	<b>Le débit de parole et la coarticulation</b>	<b>305</b>
6.1	Méthodologie . . . . .	306
6.1.1	Participants . . . . .	306

6.1.2	Stimuli . . . . .	306
6.1.3	Procédure . . . . .	306
6.1.4	Résultats . . . . .	307
6.1.4.1	Analyses de durée . . . . .	307
6.1.4.2	Plan F1-F2 . . . . .	309
6.1.4.3	Coarticulation en position initiale (CV) . . . . .	312
6.1.4.4	Équations de locus (CV) . . . . .	316
6.1.4.5	Coarticulation en position finale (VC) . . . . .	318
6.1.4.6	Équations de locus (VC) . . . . .	322
6.2	Discussion . . . . .	324
6.2.1	Chevauchement temporel . . . . .	324
6.2.2	Position syllabique de l'occlusive . . . . .	324
6.2.3	Opposition voisée vs. non voisée . . . . .	325
6.3	Conclusion . . . . .	326
<b>7</b>	<b>Variations temporelles et classification</b>	<b>329</b>
7.1	Qu'est-ce que les LDA ? . . . . .	329
7.2	La classification et la temporalité . . . . .	330
7.2.1	L'impact de la durée physique sur la classification . . . . .	332
7.2.1.1	Première expérience . . . . .	332
7.2.1.2	Deuxième expérience . . . . .	336
7.2.1.3	Troisième expérience . . . . .	339
7.2.2	L'impact des catégories temporelles sur la classification . . . . .	342
7.2.2.1	Catégories intrinsèques . . . . .	342
7.2.2.2	Catégories extrinsèques . . . . .	344
7.3	Conclusion . . . . .	345
	<b>Conclusion générale</b>	<b>347</b>
	<b>A Les Tables</b>	<b>357</b>
	<b>B Les Corpus</b>	<b>359</b>
	<b>C Les Figures</b>	<b>363</b>
	<b>Bibliographie</b>	<b>365</b>

# Introduction

La production de la parole comprend la réalisation des consonnes et des voyelles qui sont dépendantes les unes des autres dans les domaines articuloire et acoustique. Dans ce dernier, les transitions formantiques représentent le lien reliant les deux segments. Elles fournissent alors des informations importantes sur l'identité spectrale de ces sons. Or, ces informations ainsi que l'identité de ces sons peuvent être influencées par plusieurs phénomènes dont les variations temporelles vocaliques. Ces dernières peuvent être associées à l'opposition de durée phonologique dans certaines langues (voyelle longue vs. courte), qui est intrinsèquement liée à la voyelle. Autrement dit, les voyelles longues sont physiquement plus longues que les voyelles courtes. Cette particularité de longueur est ainsi inhérente à chacun de ces deux types de voyelles. Les variations temporelles d'une voyelle peuvent aussi dépendre des variations des débits de parole. Ce phénomène extrinsèque à la voyelle change également la durée, où les voyelles produites en débit lent sont physiquement plus longues que les mêmes voyelles prononcées en débit rapide. Ces changements de durée vocalique sont alors étrangers aux particularités acoustiques des voyelles.

L'impact des variations temporelles sur les informations spectrales a été traité à de nombreuses reprises par différents chercheurs dans différentes langues. Ce sujet a alimenté plusieurs débats dans la littérature où la question de l'effet temporel s'est posée sur sa présence ou son absence. C'est pourquoi cette question méritait d'être examinée sous un angle nouveau. En effet, cette nouvelle recherche sur le sujet propose d'étudier l'impact de ces différents types de variations temporelles sur les qualités spectrales de la voyelle et de la consonne ainsi que sur la relation coarticuloire entre la consonne et la voyelle. Pour ce faire, plusieurs analyses ont été mises en place sur l'arabe jordanien et le français, apportant de nouvelles données qui serviront à éclaircir ces interrogations et à vérifier les postulats mentionnés dans la littérature sur les deux types de variations temporelles extrinsèque et intrinsèque. Ces analyses ont également pour but d'enrichir le débat et d'examiner l'influence de la temporalité sur la classification des consonnes.

Cette recherche est alors composée de deux parties, théorique et expérimentale, réparties en sept chapitres, faits de manière à répondre aux questions citées ci-dessus. Ces chapitres cherchent à explorer la participation de la temporalité dans la variabilité du signal acoustique. La partie théorique est constituée de trois chapitres introductifs, visant à décrire certaines notions, qui aideront à appréhender les problématiques examinées et le traitement de celles-ci par

les chercheurs dans la littérature. Le premier chapitre est consacré à la définition des propriétés spectrales des sons de la parole les plus importantes. En effet, les phonéticiens ont toujours cherché des indices permettant de comprendre les mécanismes de la production et de la perception de la parole. Les mouvements des articulateurs, qui modifient le flux d'air sortant des poumons dans le conduit vocal, forment les caractéristiques principales distinguant les différents sons de la parole. Ces caractéristiques reflètent des informations spectrales qui donnent également des indications sur l'identité de ces segments. Plusieurs indices spectraux ont alors été identifiés qui contribuent à différencier les catégories phonétiques de la parole. Caractérisant un segment, ils sont sujets à des variations issues de plusieurs sources dont les sons adjacents et les variations temporelles.

Une partie de la variation des caractéristiques acoustiques peut alors être justifiée par les effets de la coarticulation, c'est-à-dire quand un segment est co-articulé avec un autre segment. La notion de la coarticulation sera décrite dans le deuxième chapitre afin de la définir, de l'expliquer et de mettre en avant les différentes théories et modèles montrant comment et pourquoi la coarticulation a lieu. L'étude de ces derniers permettra de mieux réfléchir sur les moyens qui pourront être utilisés dans l'analyse de cette recherche.

Parmi les modèles proposés, les équations de locus ont été choisies ici, car elles représentent un outil paramétrique fiable qui aide à mesurer le degré de coarticulation. Ces équations ont également été l'intérêt de plusieurs études au travers d'une autre problématique parallèle. De fait, les équations de locus sont considérées comme des indices relationnels invariants qui caractérisent le lieu d'articulation des consonnes. Le troisième chapitre sera donc destiné à expliquer le fonctionnement de ces équations et à exposer les travaux qui les ont employées.

Concernant la seconde partie, elle est composée d'un chapitre développant les problématiques de ce travail et de trois chapitres présentant les aspects expérimentaux étudiés. Le chapitre des problématiques détaille les objectifs de cette recherche dans leurs contextes. Les systèmes vocalique et consonantique de l'arabe jordanien seront alors précisés ainsi que la différence entre les variations temporelles intrinsèques. Par ailleurs, des études ayant utilisé les équations de locus pour mesurer l'impact temporel seront ensuite exposées.

Le cinquième chapitre, quant à lui, consiste à décrire la méthodologie et énonce les résultats des deux expériences menées sur la production de la parole en arabe jordanien. Ces deux expériences ont pour but d'analyser l'impact des variations temporelles intrinsèques aux voyelles dans plusieurs contextes consonantiques dans ce dialecte. La première expérience se focalisera sur l'examen de l'influence temporelle issue de l'opposition phonologique sur les propriétés spectrales des cinq occlusives non voisées, des huit voyelles de l'arabe jordanien et sur leurs relations coarticulatoires. À propos de la deuxième expérience, elle sera conçue de manière à comparer les effets des variations temporelles dans les contextes des occlusives voisées et non voisées. Cette expérience permettra de vérifier si les effets potentiels des variations temporelles intrinsèques sont présents ou absents dans les deux catégories consonantiques (voisée et non voisée). Elle servira également comme une expérience contrôle pour évaluer les résultats obser-

vés dans la première expérience.

L'impact des variations temporelles extrinsèques à la voyelle sur les informations spectrales et sur la coarticulation sera abordé dans le sixième chapitre. Celui-ci exposera un protocole expérimental qui a été mis en place de façon à examiner si les variations temporelles extrinsèques liées aux changements du débit de parole amènent à des conséquences comparables à celles observées avec les variations temporelles intrinsèques énoncées dans le chapitre précédent. Pour ce faire, les occlusives voisées et non voisées du français sont testées lorsqu'elles sont produites avec trois voyelles dans trois débits de parole, c'est-à-dire lent, moyen et rapide.

L'effet des deux types de variations temporelles sur la classification des catégories consonantiques sera ensuite examiné dans le septième chapitre par le biais des analyses discriminantes linéaires. Ces analyses permettront d'avoir des indications sur l'importance des informations temporelles dans la perception de la parole.



**Première partie**  
**Aspects théoriques**



# Chapitre 1

## Caractérisation phonétique des catégories phonémiques

Un principe fondamental dans l'étude de la parole est d'admettre que les énoncés peuvent être présentés comme une séquence de segments sous-jacents (Stevens, 1983). Ces derniers peuvent être décrits en termes de traits ou de classes naturelles (Jakobson et al., 1963). La production d'un segment phonétique est organisée de manière à achever les cibles acoustiques souhaitées ou les propriétés qui identifient les traits du segment. Par conséquent, les structures articulatoires, actionnées par les mécanismes physiologiques, sont programmées de façon à réaliser une série de configurations cibles. Le rôle de ces mécanismes physiologiques de la parole est de produire des signaux acoustiques qui ont une signification communicative (MacNeilage et Ladefoged, 1976). Ces signaux sont perçus par l'être humain grâce à des propriétés physiques existantes dans chaque signal. Quelles sont ces propriétés acoustiques ? Quelles sont les informations nécessaires pour différencier un son d'un autre ? Pour répondre à ces questions, des catégorisations ont été créées pour classer les différents sons de la parole en fonction du lieu et du mode d'articulation. Au sein de chaque catégorie, un son possède une ou plusieurs caractéristiques acoustiques qui le distinguent des autres sons. Dans ce chapitre, un aperçu général des caractéristiques phonétiques des catégories phonologiques sera présenté .

### 1.1 Notions de bases

Lors de la production de la parole, un signal acoustique est formé par les mouvements des organes du conduit vocal. Ces mouvements mènent à un ensemble de perturbations des molécules de l'air dans le flux d'air pulmonique qui est ensuite propagé vers l'extérieur dans toutes les directions. Lorsque les variations de pression atteignent le tympan, il se met à vibrer. Le système auditif traduit ces vibrations en pulsations neurales reconnues en tant que sons. Un son est donc perçu quand les variations de pression générées par le conduit vocal affectent suffisamment le tympan (Johnson, 2003). Le signal de la parole représente le point final de la

production de la parole et le point de départ de la perception.

Un des problèmes majeurs de l'étude de la voix humaine (et du son en général) est son invisibilité. L'examen des aspects acoustiques de la parole s'est longtemps occupé de la conversion des sons complexes en une représentation visible. Parmi les diverses représentations utilisées, la plus efficace est fournie par le spectrographe qui a été considéré comme un outil fiable dans l'étude des caractéristiques acoustiques des sons de la parole. L'objectif de cette section est de définir les notions de bases qui faciliteront la compréhension de l'étude des caractéristiques phonétiques des segments telles que le spectrogramme, le modèle source-filtre, le formant et les transitions formantiques.

### 1.1.1 Le spectrogramme

L'apparition du *spectrographe* durant les années 1940 par les chercheurs des *Laboratoires Bell* a contribué de manière considérable à l'avancement de la recherche dans le domaine de la production et de la perception de la parole (Koenig et al., 1946). Le spectrographe produit une image visuelle interprétable (*spectrogramme*) des sons. Dans un spectrogramme, une représentation tridimensionnelle du signal acoustique est obtenue, avec le temps en abscisse, la fréquence en ordonnée et l'intensité identifiée par le dégradé de la couleur noire (+ de noir, + d'intensité, Figure 1.1). En examinant plusieurs spectrogrammes des mêmes sons produits par

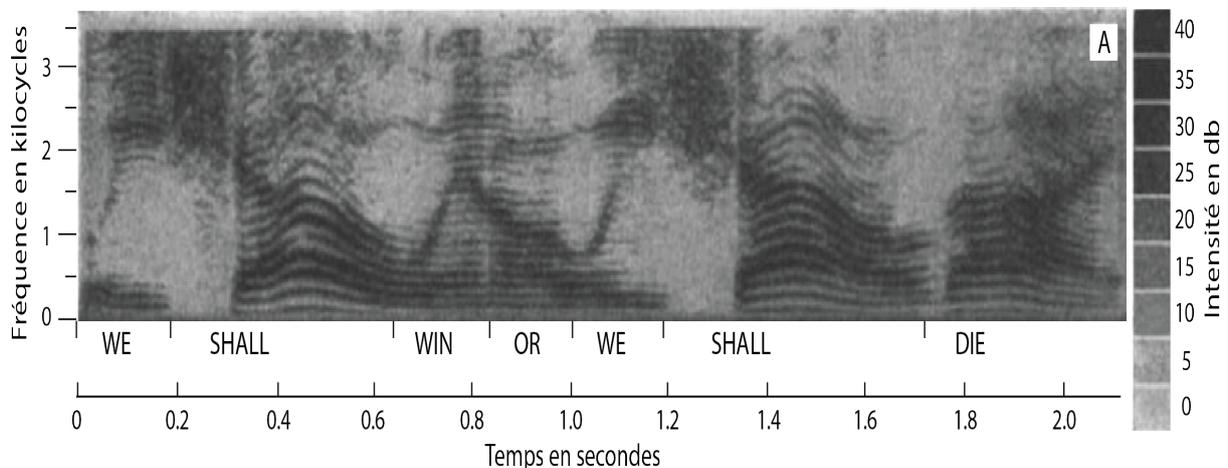


FIGURE 1.1 – Spectrogramme par Koenig et al. (1946) montrant des propriétés spectrales (fréquence, temps et intensité) de la phrase « *we shall win or we shall die* ».

différents locuteurs, un chercheur peut constituer une description des traits acoustiques communs de ces échantillons. Cette opération aide à l'identification des informations les plus importantes qui caractérisent chaque son. La représentation spectrographique est donc une manière efficace d'étudier les indices et les critères des sons de la parole.

Le spectrogramme a inspiré les chercheurs des *Laboratoires Haskins* pour l'invention de *The Pattern Playback Machine* à la fin des années 1940. Cette machine a servi aux chercheurs à identifier les informations spectrales nécessaires à la perception des sons de la parole. Elle per-

met la reproduction de la parole intelligible à partir d'un spectrogramme dessiné à la main. Il faut tout simplement recopier le spectrogramme original d'une séquence quelconque sur un papier, c'est-à-dire dessiner une image identique du spectrogramme original (Cooper et al., 1951). L'image est ensuite passée dans la machine qui, à son tour, produit la séquence d'origine de manière claire et intelligible. Les spectrogrammes peuvent donc être redessinés et modifiés de façon à ce qu'ils contiennent de moins en moins d'informations spectrales. Ces spectrogrammes sont écoutés par des auditeurs dans le but d'identifier les indices les plus importants pour la perception de la parole. Cette technique a permis aux chercheurs de combiner l'analyse et la synthèse de la parole et d'accéder à une manipulation libre de la parole synthétique. Ce processus de combinaison a conduit à la reconnaissance des indices acoustiques les plus pertinents, "*the acoustic cues*", pour la perception de la parole. Les années 1950 ont connu le développement des recherches sur les indices acoustiques de la parole grâce au *Pattern Playback* qui a mené à une description acoustique approfondie des consonnes et des voyelles (Cooper et al. (1952); Delattre et al. (1955); Liberman et al. (1958), entre autres). De plus, la relation entre les mouvements articulatoires et les informations acoustiques fournies par le spectrogramme est devenue plus claire et plus compréhensible.

Les caractéristiques acoustiques des signaux de la parole, citées plus haut, peuvent être prédites par le modèle *source-filtre*. C'est un moyen de modéliser les différentes constriction dans le conduit vocal par un tube fermé à une extrémité et ouvert de l'autre.

### 1.1.2 Le modèle source-filtre

La théorie acoustique de la production de la parole propose un modèle idéalisé du conduit vocal afin de prédire comment les différentes formes et actions du conduit vocal contribuent au signal acoustique (Stevens et Arthur, 1956; Fant, 1960). Selon cette théorie, l'excitation de la source du signal peut être modulée indépendamment des caractéristiques du filtre du conduit vocal. Le processus de production de la parole peut être décrit par le modèle *source-filtre* développé par Fant (1960). La réalisation des sons implique la présence d'une **source** d'énergie acoustique. Il en existe deux types primaires : le voisement et le bruit. Le premier correspond à une onde complexe périodique, le second à une onde complexe aperiodique. Quand le flux d'air venant des poumons passe à travers le larynx, il peut provoquer la vibration des plis vocaux (les cordes vocales). Dans ce cas, une quantité d'énergie laryngale est émise ; le signal de la parole résultant est dit voisé. Une onde périodique<sup>1</sup> est alors produite (Figure 1.2(a)). Le spectre d'amplitude<sup>2</sup> de cette onde expose l'énergie de la fréquence fondamentale ( $f_0$ ) de la vibration laryngale et de ses harmoniques (les multiples de la fréquence fondamentale) (Figure 1.2(d)). Le débit de cette vibration dépend de la taille des plis vocaux et de la tension musculaire qu'ils

1. Il s'agit en réalité d'une onde complexe quasi-périodique, or dans les analyses acoustiques, elle est considérée en tant que périodique pour simplifier le calcul.

2. Le spectre d'amplitude (ou le spectre) est une représentation bidimensionnelle du signal acoustique affichant la fréquence (en abscisse) en fonction de l'intensité (en ordonnée).

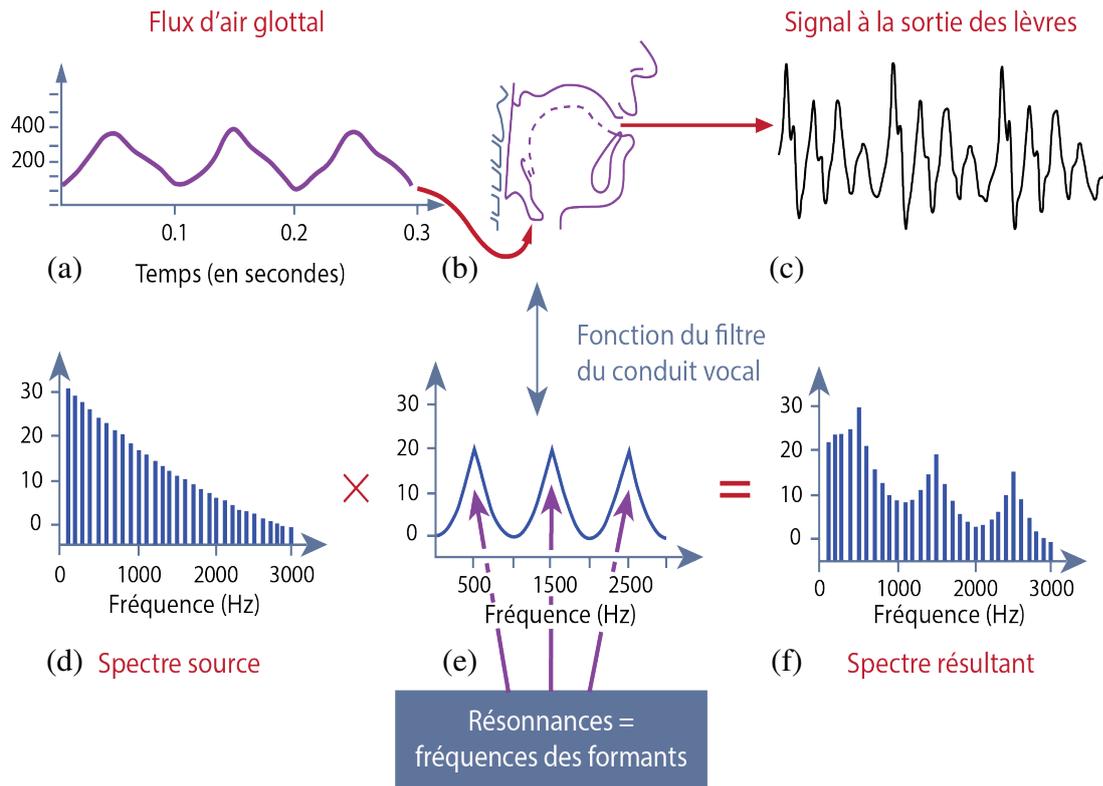


FIGURE 1.2 – Représentation de la théorie source-filtre (d'après Flemming (2005)).

subissent.

Il est également possible que l'air passe par les plis vocaux sans qu'ils ne vibrent. Dans ce cas, l'énergie (dite supra-glottale) est générée grâce à la turbulence de l'air située à un point dans le conduit vocal selon le son produit. Cette énergie implique des variations de pression apériodiques sur une gamme large de fréquences. Le signal de la parole résultant est appelé non voisé.

Un son peut combiner plus d'une source comme le voisement et le bruit de friction pour les fricatives voisées. Après son passage à travers le larynx, le flux d'air entre dans le conduit vocal. Ce dernier réagit comme un **filtre** modifiant l'onde de la source. Un filtre acoustique est un dispositif qui laisse passer certaines fréquences et en atténue d'autres. Les sons de la parole sont les résultats du filtrage de la source dans le conduit vocal (Figure 1.2). Le locuteur déplace les articulateurs (la langue, les lèvres et la mâchoire) d'une configuration à une autre pour produire les sons de la parole. Une manière très simplifiée pour décrire le conduit vocal, souvent adaptée dans la littérature, est une série de tubes. À tout moment de la production des sons, les positions de la langue, de la mâchoire et des lèvres peuvent être représentées par des tubes de longueurs différentes et de zones transversales uniformes (Figure 1.2(b)) (Sawusch, 2005). Chaque tube a sa propre fréquence de résonance selon sa longueur. Les caractéristiques des résonances du conduit vocal d'un son sont affichées dans la Figure 1.2(e) où les pics représentent les fréquences de résonance. En effet, le conduit vocal donne forme au flux d'air et le son résultant correspond aux effets combinés du larynx (**source**) et du conduit vocal (**filtre**)

(Figure 1.2(c) et (f)).

Tout changement dans la configuration du conduit vocal modifie les fréquences auxquelles les cavités résonnent. Le conduit vocal est donc un résonateur variable, autrement dit quand sa forme varie, les résonances ou les formants changent de fréquence de façon continue produisant une variété de sons. Le modèle source-filtre illustre d'une manière efficace les conséquences des modifications qui ont lieu dans les différentes parties du conduit vocal. La conversion des segments audibles en représentation visuelle interprétable a constitué les bases de la recherche de la perception de la parole. Le spectrogramme a permis de visualiser les modifications des sons sources (périodique et apériodique) dans le conduit vocal. Grâce au spectrogramme, des informations phonétiques importantes ont pu être identifiées à l'œil nu, telles que le bruit de relâchement, les trois premiers formants, les transitions formantiques et le bruit de friction (Figure 1.3). Ces dernières sont considérées comme des indices qui permettent de déterminer certaines caractéristiques des segments de la parole comme le lieu et le mode d'articulation.

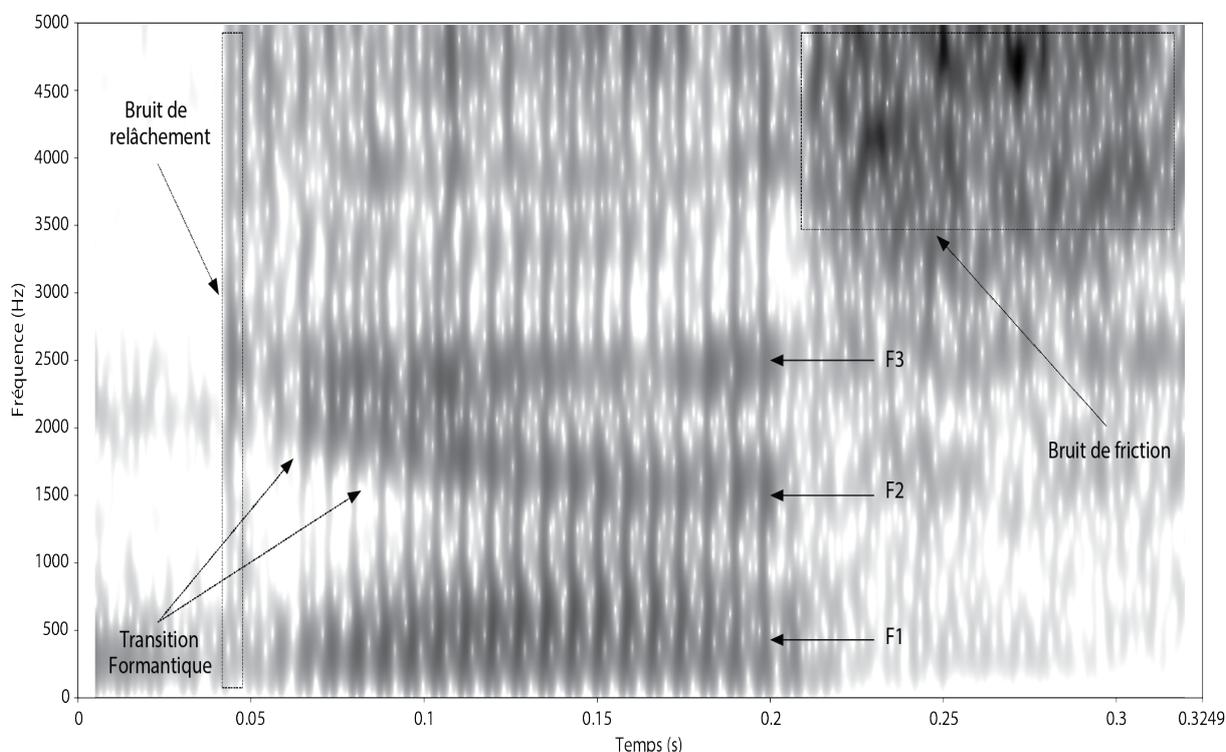


FIGURE 1.3 – Représentation spectrographique de la séquence /ga:s/ affichant des informations phonétiques importantes.

### 1.1.3 Les formants

Selon le modèle *source-filtre*, les vibrations des plis vocaux sont considérées comme la source des sons voisés. Cette source est ensuite filtrée par le biais du conduit vocal menant à la formation des formants qui sont les résonances naturelles du conduit vocal. Les formants sont alors issus de la modification de l'énergie produite par la source dans le conduit vocal. Il existe

un nombre indéfini de formants, mais les trois premiers - qui ont les fréquences les plus basses - sont les plus importants pour la perception. Ces premiers formants portent la plus grande partie de l'information sonore. Le formant avec la fréquence la plus basse est appelé F1 et ceux qui le suivent F2, F3, F4, etc.

Les formants peuvent être reconnus dans deux représentations. La première est le spectre d'amplitude où les formants sont les pics de l'enveloppe spectrale (Figure 1.2(e)). La seconde est le spectrogramme où les formants correspondent aux régions de plus grande intensité ; autrement dit, aux zones horizontales les plus foncées sur le spectrogramme (Figure 1.1). Les fréquences des formants fournissent des informations sur la position des organes articulatoires et donc sur le type de son produit (ouvert, fermé, antérieur, postérieur, etc.). Les trois premiers formants, souvent même les deux premiers, contiennent tous les principaux traits pertinents aidant à la perception (par exemple, le lieu d'articulation ou l'ouverture du conduit vocal). Les formants supérieurs contribuent surtout à caractériser la voix individuelle et ils ont un rôle linguistique très limité dans l'intelligibilité totale du son (Delattre, 1961).

Chaque formant peut être décrit par deux caractéristiques : la fréquence au milieu du formant appelée souvent fréquence formantique et la bande passante qui est la mesure de l'étendue de l'énergie dans le domaine fréquentiel (Kent et Read, 1996). Les fréquences des formants sont corrélées à la forme et aux positions des articulateurs pendant la réalisation des segments. Toute partie du conduit vocal a des influences sur les formants et chaque formant est dépendant de la forme globale du système (Harrington, 2010). Selon Fant (1959), les variables articulatoires principales sont : 1) le point d'articulation, 2) le degré de constriction du rétrécissement principal entre la langue et la partie opposée dans le conduit vocal et 3) le degré de constriction et l'allongement du passage des lèvres. Une valeur maximale du premier formant (F1) exige que la constriction principale soit située au-dessus du larynx et que la cavité buccale soit ouverte. Une augmentation de la constriction dans la cavité buccale conduit à une baisse de F1 (Lindblom et Sunberg, 1971; Harrington, 2010). Autrement dit, les sons de la parole qui exigent une petite ouverture de la bouche ont des fréquences de F1 moins élevées. Au contraire, lorsqu'ils sont réalisés avec une bouche ouverte, le F1 est caractérisé par des fréquences hautes (Borden et al., 1994). Le deuxième formant (F2) est associé aux changements dans la cavité orale provoqués par les mouvements de la langue, des lèvres et de la mâchoire. Une valeur maximale du F2 est liée à une constriction linguale dans la zone palatale. Des constriction plus postérieures à cette zone mènent à une élévation du F2 ainsi que du troisième et du quatrième formants (F3, F4). F2 a une valeur maximale basse lorsque la constriction linguale est située dans la partie supérieure du pharynx (Harrington, 2010; Borden et al., 1994). Une diminution de l'ouverture des lèvres ou une augmentation de la longueur du passage des lèvres produit un abaissement des formants. La protrusion des lèvres a un effet marqué sur F3 des voyelles antérieures et sur F2 des voyelles postérieures (Lindblom et Sunberg, 1971; Harrington, 2010). L'importance relative des formants des sons voisés diminue à partir de F2. Les deux premiers formants, F1 et F2, sont les déterminants principaux de la qualité de la voyelle. F3 et F4 contribuent signifi-

tivement à la qualité phonétique des voyelles antérieures, mais ils ont une importance mineure pour celles postérieures (Fant, 1960). F3, F4 et f0 fournissent des informations considérables sur les caractéristiques personnelles de la voix comme le sexe, l'âge et la taille du conduit vocal.

Les analyses acoustiques de la parole ainsi que les études de perception indiquent que les corrélats acoustiques dominants des voyelles sont les trois premiers formants. Ces derniers ne sont pas complètement stables, notamment aux frontières (le début et la fin des formants), car ils sont influencés par les segments adjacents. Les changements rapides de fréquence aux deux extrémités des formants s'appellent *transitions formantiques*.

#### 1.1.4 Les transitions formantiques

Lors de la production de la parole, les locuteurs modifient leur conduit vocal de façon à prononcer des consonnes et des voyelles (phases ouverte et fermée) de manière continue. Les transitions formantiques correspondent alors aux changements rapides de fréquence dans les formants vocaliques qui reflètent les mouvements articulatoires du passage d'une consonne à une voyelle ou l'inverse (Figure 1.3). Delattre (1961) les définit ainsi : « *Entre la tenue d'une consonne et la tenue de la voyelle suivante [...], il se produit un mouvement articulatoire ouvrant combiné à un déplacement complexe des organes. Ce mouvement physiologique est reflété sur un spectrogramme acoustique par des changements de fréquence généralement rapides et continus dans les formants, [...]. Tout comme les notes de résonance du système des cavités changent continuellement pendant le déplacement des organes, de même les formants acoustiques changent continuellement de fréquence. C'est à ces changements de fréquence des formants que la terminologie acoustique a donné le nom de transitions* »<sup>3</sup>. Par conséquent, la durée des transitions formantiques d'une séquence est la même que celle de son changement articulatoire. Par exemple, si la transition articulatoire d'une consonne à une voyelle prend 50 ms, les transitions acoustiques sont de 50 ms environ (Kent et Read, 1996).

Les transitions formantiques sont considérées comme un indice acoustique important pour la perception de la combinaison consonne-voyelle. Les voyelles peuvent être identifiées par leurs formants sans les transitions (en isolation par exemple) puisqu'elles sont généralement fortes et longues, alors que les consonnes sont souvent faibles et courtes. La portion des transitions de la voyelle est transférée comme une contrepartie acoustique pour la consonne (Cooper et al., 1952). Plusieurs études de perception ont confirmé l'importance des transitions formantiques pour la perception des consonnes. Pourtant, l'importance de ces transitions varie en fonction du mode et du lieu d'articulation (Delattre, 1961). Par exemple, les transitions avec les formants vocaliques suffisent à distinguer les occlusives /b, d/ sans besoin d'autres indices comme le bruit de relâchement alors que pour l'occlusive /g/, il est indispensable d'avoir les deux indices pour la reconnaître (bruit de relâchement et transitions formantiques). De même, le bruit de friction fournit l'information nécessaire qui aide à la perception des fricatives /s,

---

3. Page 407.

ʃ, z, ʒ/ et domine les autres indices existants tandis que pour les fricatives /f, θ, v, ð/, il faut que les transitions formantiques et le bruit de friction soient réunis pour les différencier (Harris, 1958; Wagner et al., 2006)<sup>4</sup>. Les transitions peuvent être positives (au-dessus du formant) ou négatives (en-dessous du formant) selon leurs directions. Elles peuvent être horizontales lors de l'absence de transition (transition zéro). Un régime général existe qui fonctionne pour les trois formants : plus les transitions sont lentes, plus la consonne est vocalisée. Par conséquent, les transitions des consonnes non voisées sont plus rapides que leurs contreparties voisées (Delattre, 1961). À travers les transitions formantiques (aspect acoustique), il est donc possible d'avoir des informations sur le mode et le lieu d'articulation (aspect articulatoire).

La transition du premier formant est globalement montante lorsque la consonne est en position initiale et descendante en position finale, peu importe la voyelle suivante/précédente. Autrement dit, la valeur fréquentielle au début ou à la fin de F1 est inférieure à celle de son milieu<sup>5</sup>. En effet, selon la théorie acoustique de la production de la parole, la fréquence du F1 est corrélée avec le degré de l'ouverture de la cavité buccale (+ d'ouverture, + de fréquence). Il est donc attendu à ce que la fréquence du début de la transition du F1 (proche de la consonne = + de constriction) soit inférieure à son milieu (voyelle = + d'ouverture). La transition du F1 peut alors être utilisée comme un indice du mode d'articulation des consonnes selon le degré de constriction (Delattre, 1961; Kent et Read, 1996). Elle est également associée au voisement des consonnes (Cooper et al., 1952; Liberman et al., 1958; Delattre, 1961; Liberman et al., 1967; Benki, 2001). Les consonnes voisées sont marquées par une transition de F1 plus longue et plus lente en comparaison de leurs contreparties non voisées. Cela signifie que les valeurs de fréquence du F1 au début du formant sont proches ou similaires à celles de son milieu pour les non voisées alors que la fréquence au début de la transition du F1 pour les voisées est plus basse que son milieu (Delattre et al., 1955; Liberman et al., 1958). Dans le cas de la parole synthétique, la diminution de l'intensité ou la suppression totale de la transition du F1 transforme l'occlusive voisée en une non voisée (Liberman et al., 1958).

La transition du deuxième formant est un indice important pour préciser le lieu d'articulation des consonnes, notamment des occlusives. Elle peut changer de direction en fonction de la voyelle suivante et/ou précédente pour une consonne donnée (Delattre et al., 1955; Öhman, 1966). Par exemple, pour la consonne /d/, la transition du deuxième formant est montante quand elle est produite avec la voyelle /i/ tandis qu'elle est descendante devant le /u/ (Figure 1.4). Malgré cette différence de direction, les deux transitions se dirigent vers un même point virtuel de fréquence. Ce dernier est appelé *locus* (Delattre et al., 1955) (cf. section 2.3.1).

La transition du troisième formant est également associée à l'identification du lieu d'articulation. Elle offre un indice indispensable du lieu d'articulation pour compléter l'information

4. Ces constatations ne sont pas universelles et elles peuvent varier d'une langue à une autre (Wagner et al., 2006).

5. La transition du F1 peut être également montante en initiale et descendante en finale lorsque la constriction est dans la cavité pharyngale (Delattre, 1961). Elle sert ainsi à distinguer les consonnes buccales des consonnes pharyngales.

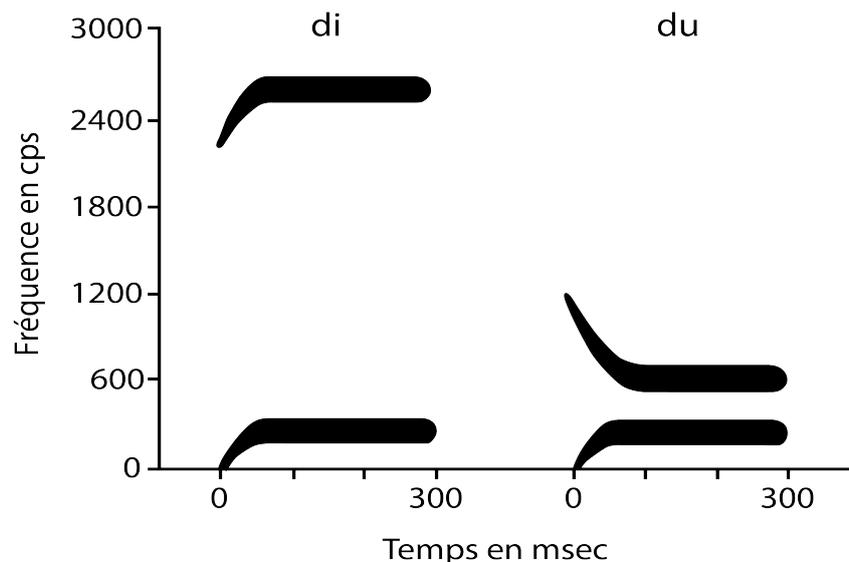


FIGURE 1.4 – Le premier et le deuxième formant avec les transitions formantiques (emprunté à Liberman et al. (1967))

de la transition du deuxième formant en cas d'ambiguïté. Par exemple, les occlusives /g/ et /d/ ont les mêmes transitions du premier et du deuxième formants lorsqu'elles sont produites avec la voyelle /ε/, c'est donc la transition du troisième formant qui permet de les distinguer (Figure 1.5).

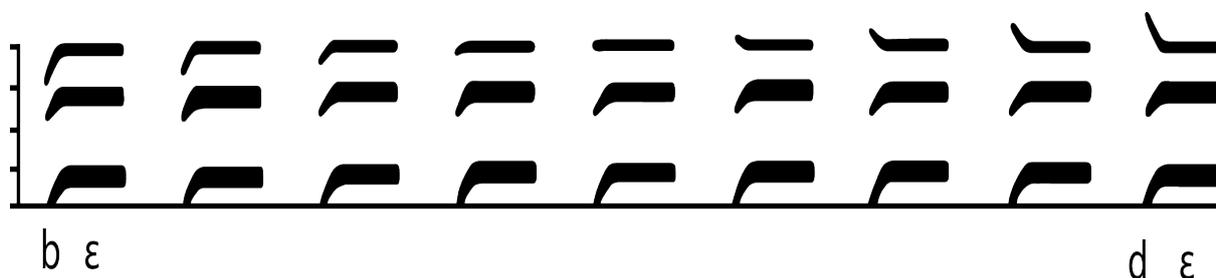


FIGURE 1.5 – Illustration par des formants synthétiques de l'importance de la transition du troisième formant (d'après Delattre (1961)).

Dans cette section, des notions de base nécessaires à la compréhension et à l'étude des caractéristiques phonétiques ont été exposées. La section suivante a pour objectif de définir les catégories des sons de la parole et d'expliquer les indices acoustiques les plus importants aidant à leur caractérisation. En effet, la production de la parole contient deux phases principales : une phase ouverte (les voyelles) et une phase fermée (les consonnes). Les voyelles sont réalisées avec une grande ouverture du conduit vocal et elles sont classées en fonction de la position de la langue et de l'ouverture des lèvres. Les consonnes représentent la phase fermée dans la parole à cause de la constriction étroite ou de la fermeture totale du conduit vocal. Les caractéristiques acoustiques des consonnes sont plus complexes que celles des voyelles. Les consonnes sont divisées en deux groupes, les voisées ou les non voisées selon la source d'énergie (vibration des plis vocaux vs. constriction dans le conduit vocal). Elles sont classées selon le lieu et le

mode d'articulation. Un intérêt particulier sera porté aux occlusives<sup>6</sup> dans la section suivante ; les autres modes d'articulation seront définis de manière générale.

## 1.2 Description générale des sons de la parole

La compréhension de la parole entraîne principalement un processus qui permet d'obtenir des informations provenant d'un stimulus auditif. Ce processus implique la discrimination des différents sons et l'évaluation de ceux qui sont similaires. Différentes manières peuvent être adaptées pour distinguer les sons de la parole comme diviser les sons selon le type de sources d'énergie ou selon le degré de l'ouverture du conduit vocal. Les sons produits avec la vibration des plis vocaux et avec un grand degré d'ouverture du conduit vocal sont appelés voyelles, ceux réalisés avec ou sans la vibration des plis vocaux et avec une ouverture plus petite (ou fermeture complète) sont nommés consonnes. Dans cette section, une description générale des caractéristiques phonétiques sera donnée pour ces deux catégories.

### 1.2.1 Les voyelles

La réalisation des voyelles représente la phase ouverte dans le processus de la production la parole dans laquelle le flux d'air passe librement sans obstacles. Les voyelles sont les sons les plus simples à décrire et à analyser acoustiquement (Kent et Read, 1996). En effet, elles sont souvent caractérisées par un ensemble de descripteurs acoustiques, notamment les fréquences des trois premiers formants. Les changements de position des articulateurs modifient la forme du conduit vocal et par conséquent les fréquences des formants puisque la forme du conduit vocal détermine les valeurs des fréquences des formants, comme signalé précédemment. Les voyelles partagent certaines caractéristiques communes telles que : 1) le voisement, les voyelles sont toutes voisées (même si elles sont dévoisées dans certains contextes), 2) le degré d'ouverture, les constriction dans le conduit vocal restent larges comparées à celles des consonnes (Ladefoged, 1975).

Selon la théorie acoustique de la production de la parole, les voyelles peuvent être reproduites à l'aide d'un tube fermé à une extrémité (le larynx) et ouvert à l'autre (les lèvres). Chaque voyelle a un point de rétrécissement propre, appelé lieu de constriction, qui divise le tube en deux cavités : antérieure et postérieure (Stevens et Arthur, 1956). La variation du lieu de constriction de la cavité antérieure à celle postérieure (ou l'inverse) provoque le changement des deux premières fréquences de résonance (Fant, 1960). En effet, ces changements ne sont pas linéaires<sup>7</sup> ; c'est-à-dire qu'il existe des zones où une grande modification du lieu de

6. Le terme *consonnes occlusives* (en anglais *stop consonants*) est plus large que celui de *consonnes plosives* (en anglais *plosive consonants*). Ce dernier est utilisé pour désigner les occlusives orales uniquement (Durand, 2005). Par simplicité, nous utilisons le terme *consonnes occlusives* ou *occlusives* dans notre étude pour désigner les occlusives orales.

7. Une manière claire de comprendre ce que *linéaire* vs. *non linéaire* signifie est de présenter l'analogie du

constriction conduit à un effet négligeable sur les fréquences des formants et d'autres zones où un petit changement articulaire peut mener à des conséquences acoustiques importantes. La structure acoustique d'une voyelle est alors déterminée par des résonances qui, lorsqu'elles sont combinées avec la source du signal, engendrent les formants.

### 1.2.1.1 Le plan F1-F2

La langue joue un rôle indispensable dans la production des voyelles. La théorie acoustique de la production de la parole indique la présence de relations entre l'élévation de la langue et la fréquence du F1 et entre son reculement et la fréquence de F2. De ce fait, lorsque les voyelles sont affichées en fonction des deux premiers formants dans un domaine bi-dimensionnel où les fréquences du F1 décroissent sur l'axe des abscisses et celles du F2 baissent sur l'axe des ordonnées, une forme ressemblant à un quadrilatère articulatoire vocalique émerge (Figure 1.6). Cette forme, qui a d'abord été proposée par Essner (1947) et par Joos (1948), est devenue ensuite

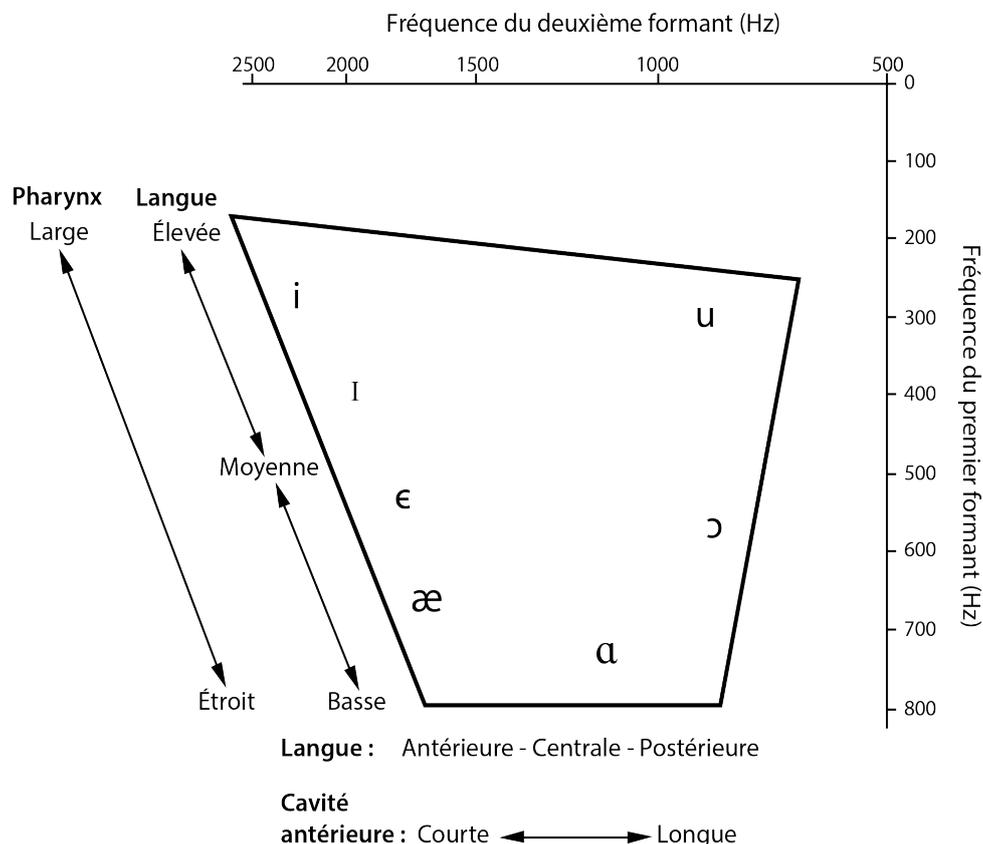


FIGURE 1.6 – Représentation de la relation entre la position de la langue, la taille de la cavité et les fréquences des formants de quelques voyelles (d'après Borden et al. (1994)).

un moyen standard pour comparer la qualité des voyelles dans les études phonétiques. Sur le bouton de la radio que Johnson (2003) propose. Il explique que le contrôle linéaire est comme le bouton du volume d'une radio. En tournant le bouton, le son devient plus fort (plus le bouton est tourné, plus le volume augmente). Au contraire, le contrôle non linéaire est comme le bouton de réglage de la radio. En tournant ce bouton, la radio reçoit différentes chaînes. Un petit tournage de ce bouton peut provoquer un grand changement du son émis par la radio et un grand tournage peut conduire à très peu de changement.

plan vertical (F1), le volume de la cavité pharyngale est réduit quand la langue se déplace d'une position haute à une autre plus basse. La raison principale de cette réduction est le reculement de la racine de la langue dans la cavité pharyngale. Les voyelles avec des positions hautes de la langue ont alors des cavités pharyngales larges qui résonnent à des fréquences basses de la source. Celles avec des positions basses ont des cavités pharyngales petites qui résonnent à des fréquences hautes de la source. La corrélation entre l'ouverture des lèvres et la fréquence du F1 peut être prédite par la position de la mâchoire. Les voyelles hautes sont caractérisées par des positions élevées de la mâchoire et des apertures réduites des lèvres et les voyelles basses sont liées à des positions basses de la mâchoire et des apertures larges des lèvres.

Quand la langue est éloignée du palais, c'est-à-dire dans une position basse, les voyelles sont dites ouvertes et quand elle prend une position proche du palais, les voyelles sont dites fermées. Sur le plan horizontal, les fréquences du F2 sont corrélées avec la longueur de la cavité de résonance orale. L'avancement de la langue conduit à une cavité antérieure plus petite, et son reculement mène à une cavité plus grande. Les fréquences de F2 diminuent quand la cavité antérieure est petite et augmentent lorsqu'elle est longue. L'arrondissement des lèvres allonge la cavité orale provoquant ainsi un abaissement de fréquence du deuxième formant. Quand la langue est placée vers le palais dur, les voyelles sont nommées antérieures ou palatales. Lorsqu'elle est positionnée vers le centre, les voyelles sont dites centrales et quand elle se trouve dans la partie postérieure, les voyelles produites sont dites postérieures.

Les formants d'une voyelle changent continuellement en fonction du temps, c'est-à-dire qu'il n'existe pas *d'état d'équilibre* à proprement parler (*cf.* section 1.2.3). Les voyelles sont prononcées généralement en présence d'autres sons (précédents ou suivants) qui influenceront leurs soi-disant *cibles vocaliques* (Lindblom, 1963a). Le conduit vocal - contrairement à un tube - se modifie alors de façon permanente. Harrington (2010) indique que la cible vocalique peut être vue comme un point temporel qui a lieu au plus proche du milieu temporel d'une monophthongue<sup>8</sup> ou bien comme une portion au milieu temporel d'une voyelle. Cette portion présente les plus petites modifications spectrales et subit le moins les influences contextuelles. En effet, il n'existe pas de méthode standard dans la recherche de la parole pour identifier où la cible vocalique se produit car, soit les voyelles n'ont pas d'état d'équilibre clair soit, l'intervalle qui change spectralement le moins peut être différent pour les trois formants. Par exemple, Van Son et Pols (1990) utilisent cinq méthodes pour mesurer la cible vocalique :

1. *Méthode Centre* : les valeurs des formants sont déterminées au milieu temporel de la voyelle.
2. *Méthode Moyenne* : il s'agit de calculer la moyenne des fréquences de la trajectoire formantique pour chaque formant.

---

8. La voyelle est également appelée monophthongue, qui signifie son (*-phthongue*) seul/unique (*mono-*). Une autre classe relative aux voyelles existe, nommée diphtongue (deux sons). Cette dernière ne peut pas être décrite par une seule forme du conduit vocal ou un seul ensemble de formants. Par conséquent, la qualité d'une diphtongue au début est différente de celle à la fin (Kent et Read, 1996).

3. *Méthode Énergie* : la cible est fixée au point maximal de l'énergie.
4. *Méthode Formant* : la cible est prélevée à la valeur maximale ou minimale selon la voyelle et/ou le formant.
5. *Méthode Stationnaire* : la cible est déterminée automatiquement grâce à un logarithme qui précise la partie la plus stable (– de variation de fréquence) dans chacun des trois formants.

Les auteurs soulignent que les différences entre les cinq méthodes sont marginales et qu'elles conduisent à des résultats similaires.

En résumé, il apparaît que les voyelles peuvent être décrites par les deux premiers formants. Ces derniers portent les informations les plus importantes qui permettent la distinction entre les différentes catégories vocaliques. Le troisième formant, quant à lui, est également un indice majeur qui pourrait contribuer à la perception de la voyelle.

### 1.2.1.2 F3

Le troisième formant est moins responsable que les deux premiers formants dans la distinction de la qualité vocalique. L'utilisation de F3 dans la parole synthétique ajoute de l'intelligibilité sans modifier la qualité de la voyelle (Delattre, 1951). Dans des études sur la perception des voyelles dans la parole synthétique avec deux formants, Delattre et al. (1952) ont montré que les auditeurs privilégient des fréquences plus élevées du F2 au détriment des valeurs naturelles correspondantes. Ils expliquent que ce phénomène est provoqué par l'absence du F3. De même, ce dernier contribue à la distinction entre arrondie *vs.* non arrondie pour les voyelles antérieures. L'arrondissement des lèvres conduit à un abaissement de la fréquence du F3. Par exemple, F3 de la voyelle antérieure arrondie [y] est plus basse (proche de F2) que sa contrepartie non arrondie [i] dû à l'arrondissement des lèvres (Harrington, 2010). De plus, une diminution de la fréquence du F3 en-dessous de son niveau normal avec les voyelles qui ont la couleur-r de l'anglais américain comme dans "*bird*" est attendue. Cette baisse est provoquée par l'élévation de l'apex de la langue vers la position de rétroflexe (Delattre, 1951).

## 1.2.2 La durée vocalique

La durée d'un son est le temps réel nécessaire pour le produire. La quantité de temps peut être mesurée et spécifiée mais elle n'est pas absolue mais plutôt relative (Al-Ani, 1970). Elle change en fonction du contexte linguistique, du locuteur, de la situation conversationnelle, etc. La durée de la voyelle est ainsi considérée comme un paramètre principal dans la perception. En effet, la durée relative de chaque voyelle est différente et elle peut varier de manière considérable allant de très courte à très longue. Plusieurs langues distinguent les voyelles qui sont relativement similaires au niveau spectral mais qui diffèrent en durée. Il s'agit donc de l'opposition phonologique des voyelles qui consiste à les diviser en voyelles longues *vs.* courtes. Par

exemple, la voyelle longue /a:/ et celle courte /a/ en arabe sont spectralement similaires mais elles sont distinguées principalement par l'indice temporel puisque la durée de la première est approximativement le double de la seconde (Al-Ani, 1970; Tsukada, 2009). D'un autre côté, il existe des différences spectrales et temporelles qui discriminent des voyelles dans certaines langues. Cette opposition est appelée tendue-relâchée (comme la voyelle tendue /i/ et celle relâchée /ɪ/ en anglais). Les voyelles tendues (+ de tension musculaire) sont plus longues que celles relâchées (– de tension musculaire). Elles occupent généralement des positions périphériques dans le plan F1-F2 ; autrement dit, elles sont plus loin du centre que les voyelles relâchées (Harrington, 2010).

La durée de la voyelle peut être également liée à d'autres raisons physiologiques. Des études ont montré que la durée acoustique d'une voyelle dépend significativement du degré de l'ouverture de la voyelle. Les voyelles ouvertes sont ainsi plus longues que celles fermées (Lisker, 1974). Cette différence de durée est expliquée principalement par l'effet mécanique de la masse relativement importante de la mâchoire.

Les différences temporelles citées plus haut sont dites intrinsèques aux voyelles. Des facteurs contextuels, dits extrinsèques, influencent également la durée vocalique tels que :

- l'accent lexical : les voyelles accentuées sont plus longues que celles non-accentuées.
- le débit de parole : les voyelles sont plus longues en débit lent qu'en débit rapide.
- le voisement des consonnes adjacentes : les voyelles deviennent plus courtes quand elles sont suivies de consonnes non voisées et plus longues lorsqu'elles sont produites avec des voisées (Klatt, 1976).

L'impact des variations temporelles extrinsèques sur la qualité de la voyelle a longtemps été débattu. D'après certains chercheurs, notamment Lindblom (1963a), les voyelles ont tendance à se centraliser lorsqu'elles subissent des variations temporelles provoquées par des phénomènes comme le débit de parole et l'accentuation. Au contraire, d'autres chercheurs indiquent que l'impact de ces variations sur la qualité vocalique est négligeable (entre autres, Gay (1978a); O'Shaughnessy (1986)) (cf. section 2.3.2).

La Figure 1.7 contient les moyennes des durées des voyelles provenant de quatre études différentes analysées par Hillenbrand et al. (2000). Il est à remarquer qu'il existe une variabilité globale des durées des voyelles reflétant la variation des conditions méthodologiques et celle de l'échantillon des locuteurs. Cependant, les voyelles présentent les mêmes tendances. D'un côté, les voyelles tendues sont toujours plus longues que leurs contreparties relâchées. D'un autre côté, la durée varie en fonction du degré de l'ouverture de la bouche. Les deux études sur la parole spontanée dans la Figure 1.7 montrent des moyennes de durées plus courtes que celles des deux autres études. Les durées sont plus longues dans l'expérience de Hillenbrand et al. (1995) que celles dans l'expérience de Black (1949) car dans la première, elles sont associées principalement à l'utilisation de l'occlusive voisée en position finale (/hVd/) tandis que dans la seconde, l'occlusive finale est non voisée (/tVp/).

Le rôle de la durée dans l'identification des voyelles a été étudié par le biais des expériences

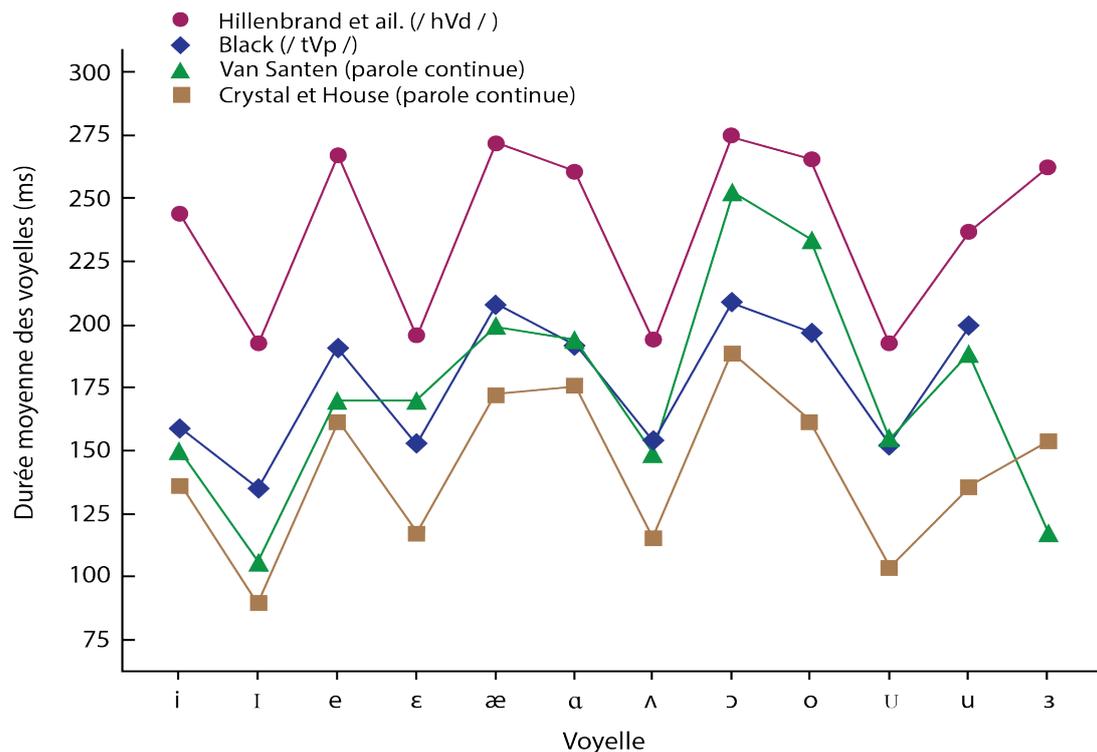


FIGURE 1.7 – Moyennes des durées vocaliques issues de quatre études : Hillenbrand et al. (1995), Black (1949), van Santen (1992) et Crystal et House (1988) (d'après Hillenbrand et al. (2000)).

utilisant des classifieurs statistiques pour déterminer la séparabilité des voyelles basés sur la combinaison de différentes mesures acoustiques. Plusieurs études ont porté sur l'importance de l'information temporelle (durée vocalique) dans la classification des voyelles, néanmoins elles ne présentent pas toujours de résultats compatibles. Par exemple, Zahorian et Jagharghi (1993) se servent des classifieurs discriminants pour identifier les signaux dans une base de données de 2922 séquences CVC (C pour consonne et V pour voyelle) avec 11 voyelles de l'anglais américain produites dans plusieurs contextes consonantiques. Les auteurs ont testé la capacité de la durée vocalique - entre autres facteurs - à améliorer la catégorisation des voyelles. Zahorian et Jagharghi (1993) ont souligné que la durée n'améliore pas significativement la classification lorsqu'elle est ajoutée aux mesures spectrales (amélioration inférieure à 1 %). D'après les auteurs, « [...] *the duration cues appear to play a small role in vowel discrimination. Duration cues might, however, be more significant for vowels in naturally spoken continuous speech* »<sup>9</sup>. Des conclusions différentes ont été obtenues par Hillenbrand et al. (1995) où ils ont utilisé des classifieurs discriminants pour identifier des voyelles dans une base de données de 1668 séquences /hVd/ (12 voyelles × 139 locuteurs). Les auteurs mettent en place trois groupes de classification par rapport au nombre d'échantillons de fréquences utilisés. Le premier groupe est constitué d'un point de mesure prélevé à l'état d'équilibre des formants. Le deuxième est com-

9. « [...] il apparaît que l'indice de la durée joue un rôle minime dans la discrimination des voyelles. Toutefois, cet indice peut être significatif pour les voyelles produites en parole naturelle », page 1977 [notre traduction].

posé de deux échantillons déterminés à 20 % et à 80 % de la durée de la voyelle. Le troisième correspond à trois points de mesures extraits à 20 %, à 50 % et à 80 % de la durée de la voyelle. Dans chaque groupe, quatre paramètres de classification ont été testés avec et sans la durée vocalique : 1)  $F1 + F2$  ; 2)  $F1 + F2 + F3$  ; 3)  $f_0 + F1 + F2$  et 4)  $f_0 + F1 + F2 + F3$ . Les résultats soulignent que l'ajout de la durée vocalique entraîne une amélioration globale conséquente pour catégoriser les voyelles (+ de 11 %), notamment avec les conditions les plus simples comme celui d'un échantillon de  $F1$  et de  $F2$  (à l'état d'équilibre). Des résultats similaires ont été également observés dans l'étude de Hillenbrand et al. (2001) où les analyses discriminantes ont été effectuées sur huit voyelles produites par 12 locuteurs en contextes consonantiques (CVC) en anglais. Une amélioration globale de la classification des voyelles de 10 % est notée quand la durée est incluse avec les informations spectrales. Hillenbrand et al. (2001) remarquent que l'effet de la durée sur la classification de certaines voyelles est plus grand que sur d'autres. Une grande amélioration de la séparation entre les catégories est observée pour  $/æ/$  (+ de 25 %) et  $/ε/$  (+ de 22 %). Elle est également non négligeable pour  $/ɪ/$  (+ de 11 %) et pour  $/ɑ/$  (+ de 13 %). La classification n'a néanmoins augmenté que de 5 % pour  $/ʌ/$  et de -1 % pour  $/i, u, ʊ/$ . En anglais australien, Watson et Harrington (1999) ont aussi remarqué une hausse de 18 % du taux total de classification lorsque la durée des voyelles est ajoutée aux cibles des trois premiers formants. En adéquation avec les résultats de Hillenbrand et al. (2001), l'effet de la durée sur la classification varie en fonction de la voyelle.

Une autre série d'études a fourni des preuves directes du rôle de la durée dans l'identification de la voyelle. Il s'agit des expériences de perception utilisant la parole synthétique ou la parole naturelle modifiée. Tiffany (1953) a enregistré 12 voyelles de l'anglais américain réalisées par quatre locuteurs entraînés sous une variété de conditions comme la variation de la hauteur ou du contexte phonétique (voyelles en isolation *vs.* en contexte CVC). Les voyelles produites en isolation ont été maintenues pendant des durées différentes (80, 200 et 500 ms). Les voyelles ont été écoutées par des auditeurs qui ont de l'expérience en phonétique. Les résultats montrent que la durée a un effet direct sur la perception de certaines voyelles. Les voyelles qui ont une durée intrinsèquement longue comme  $/e/$  et  $/ɑ/$  ont été mieux identifiées dans les durées longues (200 et 500 ms) alors que celles avec une durée intrinsèquement courte comme  $/ɪ/$  et  $/ʊ/$  ont été mieux reconnues dans la durée courte (80 ms).

De nombreuses recherches ont utilisées des voyelles synthétiques afin d'examiner l'influence de la durée vocalique (entre autres, Stevens (1959); Bennett (1968); Ainsworth (1972); Huang (1986)). Stevens (1959) a synthétisé des séquences  $/dVs/$  avec des durées qui varient entre 25 et 400 ms où  $V = /i, ɪ, ε, æ, u, ʊ, ʌ, ɑ/$ . Il a été demandé aux auditeurs d'identifier la voyelle au milieu des séquences. Les résultats signalent que les auditeurs se servent des informations spectrales et temporelles pour reconnaître les voyelles. Par exemple, les voyelles avec des propriétés caractéristiques à la voyelle  $/æ/$  ont été perçues comme  $/ε/$  lorsque les durées étaient inférieures à 100 ms. Celles avec des fréquences qui correspondent à la voyelle  $/ɑ/$  ont été entendues comme  $/ʌ/$  quand le stimulus a une durée inférieure à 100 ms. Ces tendances

sont moins marquées pour le /i/ vers le /ɪ/ et pour le /u/ vers le /ʊ/.

Bennett (1968) effectue également une expérience pour tester quel rôle peut jouer la durée des voyelles dans leur perception en anglais britannique et en allemand. L'auteur émet l'hypothèse selon laquelle l'importance de la durée vocalique comme un indice de perception diminue quand les voyelles sont spectralement loin (dans le domaine F1-F2) mais elle augmente lorsque les voyelles sont spectralement proches. Des locuteurs anglais et allemands (20 pour chaque langue) ont écouté des mots synthétiques de type CVC. Deux paires de voyelles ont été sélectionnées par langue. Les paires avec des voyelles qui sont relativement proches en qualité étaient /ʌ/ vs. /ɜ:/ pour l'anglais et /ʊ/ vs. /o:/ pour l'allemand tandis que celles loin en qualité étaient /æ/ vs. /ɑ:/ pour l'anglais et /ɔ/ vs. /o:/ pour l'allemand. Le chercheur a fait varier les fréquences des trois premiers formants et la durée des voyelles en quatre étapes. Les résultats mettent en avant que l'impact de la durée est inversement proportionnel à la distance entre les qualités vocaliques de chaque paire. Ils montrent également que les formes spectrales sont en général plus importantes que la durée dans la reconnaissance de la parole en anglais et en allemand. La durée vocalique devient plus influente que la qualité dans la discrimination uniquement lorsque les deux voyelles sont très proches en qualité.

Sur le même sujet d'étude, Ainsworth (1972) conçoit un protocole expérimental dans le but d'examiner l'effet de la durée dans la perception des voyelles. Il synthétise les deux premiers formants de 11 voyelles en isolation (V) et en contexte consonantique (/hVd/) où V = /i, ɪ, æ, a, ʌ, ɜ, ε, u, ʊ, o, ɔ/. La durée de la voyelle a été variée de 120 à 600 ms avec un pas de 120 ms. Huit locuteurs de l'anglais britannique ont participé à cette expérience. Les résultats de cette étude sont en accord avec ceux de Bennett (1968); autrement dit, les voyelles qui ont les formes spectrales les plus proches sont plus fortement dépendantes de la durée que les autres. Les voyelles fermées (/u, ʊ/ et /i, ɪ/) sont moins affectées que les autres quand elles sont en isolation comme l'a observé Stevens (1959). Lorsque les voyelles sont en contexte consonantique, l'influence de la durée est majeure même pour les voyelles fermées. L'auteur résume que « [...] *the duration of a vowel sound does affect its perceived identity, especially if the vowel is located near the center of the F1-F2 space where it may easily be confused with one of its neighbors if judgment is based on spectral information alone* »<sup>10</sup>.

Une expérience de perception sur la parole synthétique effectuée par Huang (1986) a rapporté des résultats équivoques. L'auteur examine l'importance de la durée dans la perception de l'opposition tendue vs. relâchée en anglais américain. Il synthétise des séquences /dVs/ dans le but de comparer les paires tendue-relâchée /i, ɪ/, /u, ʊ/, /æ, ε/ et /ɑ, ʌ/. Deux paires relâchée-relâchée (/ɪ, ε/ et /ʊ, ʌ/) ont également été testées. Cinq auditeurs ont écouté un continuum de neuf pas de fréquence entre les voyelles de chaque paire à quatre durées distinctes soit 40, 90, 140 et 235 ms. Pour chaque paire, les sujets ont tendance à percevoir des voyelles relâchées

10. « [...] la durée d'une voyelle affecte son identité perçue notamment si elle est localisée à proximité du centre de l'espace F1-F2 où elle peut être facilement confondue avec une de ses voisines si le jugement [des auditeurs] est basé sur l'information spectrale uniquement », page 651 [notre traduction].

avec les durées courtes, par exemple, le /i/ à 40 et 90 ms sera identifié comme un /ɪ/. Dans le cas du continuum des deux paires de voyelles basses (/æ, ε/ et /ɑ, ʌ/), il est aussi remarqué que les auditeurs ont identifié majoritairement les stimuli avec les durées les plus courtes comme des voyelles relâchées peu importe les fréquences des formants. De façon inattendue, les auditeurs ont confondu les paires relâchée-relâchée qui ont des durées semblables, particulièrement pour les voyelles basses. En effet, les différences de durées qui séparent une voyelle tendue d'une autre relâchée doivent être plus grandes que celles existant dans la parole naturelle pour que le changement d'une voyelle tendue vers une relâchée, ou l'inverse, se produise. Les différences de durées similaires à celles dans la parole naturelle présentent un changement faible ou inexistant entre les voyelles.

Les études sur la parole synthétique citées plus haut ont utilisé des valeurs de fréquences fixes pour les formants (les cibles vocaliques). Les variations spectrales des voyelles en fonction du temps (*cf.* la section suivante) et d'autres informations acoustiques ont ainsi été ignorées. Des chercheurs précisent que ces études basées sur des voyelles synthétiques statiques ont sur-estimées l'importance de l'information temporelle (Sawusch, 1996; Hillenbrand et al., 2000). Ces chercheurs ont établi des protocoles expérimentaux pour examiner l'importance de la durée dans la parole naturelle et dans une méthode la parole synthétique améliorée.

Sawusch (1996) s'interroge sur la participation de la durée vocalique dans la distinction entre les voyelles /ε/ et /æ/ en parole naturelle. Le point de départ de son étude réside dans l'analyse des deux mots "head" (/hɛd/) et "had" (/hæd/) produits par un locuteur de l'anglais américain. Les mesures spectrographiques ont montré que les fréquences des formants dans ces mots sont quasiment identiques au milieu de la syllabe<sup>11</sup>. L'auteur explique que des modèles de classification, qui utilisent exclusivement les fréquences des formants, auraient ainsi classifiés les deux voyelles comme étant la même. Pourtant, les auditeurs ont discriminé les deux voyelles en identifiant les deux mots comme le locuteur l'a envisagé. Malgré la similarité de leurs fréquences formantiques, les voyelles dans les deux mots sont différentes. Tout d'abord, la durée de la voyelle dans le mot /hæd/ est de 235 ms alors que celle dans le mot /hɛd/ est de 157 ms. De plus, les transitions formantiques et le changement dans le spectre des deux voyelles sont distincts. Par conséquent, Sawusch (1996) suggère que cette paire de voyelles a pu être distinguée perceptivement par un ou tous ces paramètres acoustiques.

L'auteur effectue deux protocoles expérimentaux afin de vérifier si la durée vocalique contribue à la discrimination des deux voyelles. Dans le premier, la durée des voyelles dans les mots /hæd/ et /hɛd/ réalisés naturellement a été modifiée par la suppression ou la réplication des pulsations de la voyelle. Le /æ/ a été raccourci en sept étapes allant de sa durée originale (235 ms) à celle du /ε/ (157 ms). Quant à la durée de celui-ci, elle a été allongée en sept étapes partant de sa durée originale à celle du /æ/. Les 14 stimuli ont été écoutés 18 fois par 16 locuteurs de l'anglais américain. Les résultats affichent que tous les stimuli dérivés de /hɛd/ ont été iden-

11. f0 = 152 vs. 132 Hz ; F1 = 607 vs. 608 Hz ; F2 = 1740 vs. 1745 Hz ; F3 = 2340 vs. 2271 Hz, respectivement pour /ε/ vs. /æ/

tifiés comme "head" et la grande majorité de ceux issus de /hæd/ ont été perçus comme "had". D'après ces résultats, le chercheur souligne que les auditeurs ne traitent pas la durée vocalique comme un indice perceptuel pour l'identité de la voyelle dans ces stimuli.

Dans la seconde expérience, Sawusch (1996) a créé des séries de parole synthétique qui sont basées sur les mots naturels de la première expérience. Dans chaque série, les fréquences des trois premiers formants durant la portion vocalique (voyelle et transitions vers la seconde consonne : VC) ont été manipulées pour changer celles qui correspondent à /hæd/ à celles appropriées à /hɛd/. Quatre variants de chaque série ont été générés. Deux ont une durée longue correspondant à celle de la voyelle dans /hæd/ alors que les deux autres ont une durée courte correspondant à celle de la voyelle dans /hɛd/. Le /h/ naturel de /hæd/ a été ajouté à une série longue et à une autre courte dans la portion VC synthétique tandis que le /h/ naturel de /hɛd/ a été intégré aux portions VC des deux autres séries. Simultanément, la fréquence de f0 et la forme du spectre ont été neutralisées. Les résultats montrent que la manipulation des fréquences influence légèrement la décision des auditeurs. Tous les stimuli longs ont été identifiés comme "had" et ceux courts ont été perçus comme "head". D'après l'auteur, ces résultats laissent supposer que les auditeurs n'utilisent pas les fréquences des formants ou leurs mouvements comme des indices perceptuels dominants afin de distinguer les voyelles dans ces deux mots. Au contraire, les auditeurs ont basé leur jugement principalement sur la durée vocalique. Sawusch (1996) résume que « [...] *vowel duration was not a strong perceptual cue to vowel identity for highly natural tokens but was used by listeners when other sources of information were made ambiguous* »<sup>12</sup>.

Des tendances similaires ont été observées par Hillenbrand et al. (2000) lorsqu'ils ont employé deux synthétiseurs de qualités différentes. Les chercheurs ont mis en place une expérience pour examiner le rôle de la durée dans la perception des voyelles à l'aide d'une version synthétisée de 300 séquences /hVd/ naturelles sélectionnées dans la base de données de Hillenbrand et al. (1995) où V = /i, ɪ, e, ɛ, æ, a, ɔ, o, ʊ, u, ʌ, ɜ/. Quatre versions de chaque séquence ont été synthétisées :

1. durée originale : les voyelles ont été synthétisées avec leurs durées originales,
2. durée neutre : les voyelles ont été synthétisées avec une durée fixée à la moyenne totale des durées de toutes les voyelles (272 ms),
3. durée courte : la durée a été déterminée à deux écart-types en-dessous de la moyenne (l'écart-type = 65 ms donc  $274 - 2 \times 65 = 144$  ms),
4. durée longue : la durée a été fixée à deux écart-types au-dessus de la moyenne (400 ms).

Parmi les deux synthétiseurs utilisés, l'un est un synthétiseur de formants (Klatt et Klatt, 1990) alors que l'autre est un synthétiseur sinusoïdal qui représente le spectre original de la

12. « [...] la durée de la voyelle n'était pas un indice perceptuel fort pour l'identité de la voyelle en parole naturelle mais elle est utilisée par les auditeurs quand d'autres sources sont rendues ambiguës », page 4 [notre traduction].

voyelle de manière plus précise que celle du synthétiseur de formants (McAulay et Quatieri, 1986). L'objectif de la seconde méthode est de comparer l'importance du niveau de la perte d'information spectrale sur le rôle de la durée dans la perception des voyelles. Les séquences des deux méthodes ont été écoutées durant deux sessions différentes par 15 locuteurs de l'anglais américain.

Condition de durée	Synthétiseur 1	Synthétiseur 2	Classification statistique
Durée originale	91.7 %	96.0 %	98.0 %
Durée neutre	90.4 %	94.1 %	97.3 %
Durée courte	82.4 %	91.4 %	88.3 %
Durée longue	89.5 %	90.9 %	87.0 %

TABLE 1.1 – Moyennes globales des taux de classification des auditeurs avec le synthétiseur de formants (synthétiseur 1) et avec le synthétiseur sinusoïdal (synthétiseur 2) dans les quatre conditions de durée. La troisième colonne représente les moyennes des taux de classification issues des analyses discriminantes (classification statistique) (Hillenbrand et al., 2000).

Les résultats de la première session (avec un synthétiseur de formants) et de la deuxième session (avec un synthétiseur sinusoïdal) mettent en évidence que l'impact de la variation de la durée sur la perception est globalement faible (Table 1.1). La grande majorité des signaux a été identifiée avec une précision élevée quand ils sont synthétisés avec leurs durées originales, neutres, raccourcies et allongées. Par contre, il existe des effets significatifs des voyelles avec une durée typiquement longue, notamment /æ/ et /ɑ/, qui ont tendance à être perçues comme leurs voisines courtes lorsque elles sont raccourcies. Quant aux voyelles typiquement courtes, notamment /ʌ/ et /ɛ/, elles ont tendance à être reconnues comme leurs adjacentes longues lorsqu'elles sont allongées. Les paires de voyelles /i/-/ɪ/ et /u/-/ʊ/ ainsi que l'ensemble /ɪ/-/e/-/ɛ/ ont rarement été influencées par les changements de durée. Il est à remarquer que le raccourcissement des voyelles a un impact plus important sur l'identité de la voyelle que son allongement dans la première session. Cependant, cet impact est statistiquement similaire pour les deux types de changement temporel dans la seconde session.

La comparaison entre les deux méthodes suggèrent que lorsque l'information spectrale est plus riche (cas du synthétiseur 2), le degré de l'impact de l'information temporelle diminue. De la même façon, l'indice de durée est plus conséquent quand une partie de l'indice spectral est absente. Ce phénomène est observé lors des trois manipulations de durée, spécifiquement dans le cas de la condition de durée courte.

Hillenbrand et al. (2000) soulignent que les résultats de cette étude laissent suggérer que l'importance de la durée augmente dans la distinction des voyelles avec un grand chevauchement spectral comme /ɑ/-/ɔ/-/ʌ/ et /æ/-/ɛ/, conduisant alors à une relation plus forte de la durée avec leur discrimination. D'un autre côté, les voyelles non sensibles aux variations temporelles telles /i/-/ɪ/ et /u/-/ʊ/ sont assez distinctes les unes des autres par leurs propriétés spectrales et, par conséquent, la durée joue un rôle minime dans leur séparation. Les auteurs ont

ainsi conçu un modèle de classification pour vérifier si un modèle peut montrer que /ɑ/-/ɔ/-/Λ/ et /æ/-/ɛ/ ont une sensibilité plus grande que /i/-/ɪ/, /u/-/ʊ/ et /ɪ/-/e/-/ɛ/ à la variation temporelle. Les paramètres utilisés pour les analyses discriminantes sont la durée et les informations spectrales des quatre conditions temporelles (fréquences de f0 et des trois premiers formants à 20 % et à 80 % de la voyelle).

Par comparaison avec les tests de perception, les résultats de ce modèle révèlent un taux de classification global légèrement plus élevé pour les conditions de durées originale et neutre alors qu'il est légèrement plus bas pour les deux autres conditions (Table 1.1). Similairement aux classifications des auditeurs, le modèle a catégorisé le /ɔ/ comme un /ɑ/ ou un /Λ/ et le /æ/ comme un /ɛ/ lorsque la voyelle est raccourcie (Table 1.2). Quand celle-ci est allongée, le

	Auditeurs	Modèle
/ɔ/ → /ɑ/ ou /Λ/	43 %	54.2 %
/æ/ → /ɛ/	20.7 %	25 %
/ɑ/ → /Λ/	9.4 %	8 %
/e/ → /ɪ/ ou /ɛ/	2.2 %	4 %
/i/ → /ɪ/	0.6 %	0 %
/u/ → /ʊ/	0 %	0 %

TABLE 1.2 – Pourcentages des changements de l'identité des voyelles entraînés par leurs raccourcissements identifiés par les auditeurs et par le modèle de l'analyse discriminante (Hillenbrand et al., 2000).

/Λ/ est classé comme un /ɑ/ ou un /ɔ/ et le /ɛ/ comme un /æ/. De plus, les voyelles /i/-/ɪ/ ainsi que /u/-/ʊ/ n'ont jamais été confondues dans le modèle lorsque la durée est modifiée. En résumé, d'après Hillenbrand et al. (2000), les résultats de ce modèle et ceux des tests de perception soutiennent l'idée que le rôle de la durée dans la reconnaissance de la voyelle ne dépend pas seulement de la magnitude et de la cohérence des différences temporelles observées entre les voyelles mais également du degré auquel ces voyelles sont séparées par les indices spectraux.

Les études citées plus haut ont révélé que l'importance de la durée dans la distinction des différentes catégories vocaliques est influencée principalement par le degré de la séparation spectrale entre les voyelles. Cette tendance est également constatée dans l'opposition phonologique de durée (voyelles longue vs. courte). Hadding-Koch et Abramson (1964) analysent l'impact de la durée vocalique dans la discrimination des voyelles longues de leurs contreparties courtes en suédois. Les auteurs choisissent trois paires de voyelles (longue vs. courte : /ɛː, ɛ/, /ɥː, ɥ/, /øː, ø/) parmi les neuf existantes en suédois pour effectuer des tests de perception. Ces paires représentent trois degrés de différence spectrale qui sépare la voyelle longue de sa contrepartie courte. Des vrais mots suédois (paires minimales) ont d'abord été produits dans une phrase porteuse par un locuteur natif. Les voyelles cibles ont ensuite été manipulées par les expérimentateurs. Les durées des voyelles longues ont été raccourcies progressivement en

couplant des portions de leurs milieux en plusieurs étapes jusqu'à obtenir les durées de leurs contreparties courtes. Les transitions formantiques n'ont pas été touchées dans le découpage. Les séquences coupées ont été écoutées dans trois expériences.

Dans la première, la durée de la voyelle /ɛ:/ du mot /vɛ:g/ (durée = 240 ms) a été raccourcie jusqu'à la durée de la voyelle /ɛ/ dans le mot /vɛg/ (durée = 135 ms). Ces voyelles ont été sélectionnées pour mettre en évidence les paires où très peu de différence qualitative est perçue. Le raccourcissement de la voyelle s'est produit en sept étapes. Dans chacune d'entre elles, la durée de la voyelle diminue de 15 ms sauf entre les étapes 1 et 2 et entre 3 et 4 où elle diminue de 20 ms et 10 ms, respectivement. Les stimuli ont été écoutés par 13 locuteurs suédois qui ont donné 130 réponses pour chaque stimuli. La Figure 1.8.A représente le pourcentage des réponses des locuteurs qui ont entendu le stimulus comme un /vɛ:g/. Entre les étapes 3 et 4, où la différence temporelle est de 10 ms, la courbe des réponses de /vɛ:g/ baisse de 79 % à 25 %. La durée vocalique semble alors avoir le rôle principal pour distinguer les deux voyelles car la différence entre les fréquences des trois premiers formants de /ɛ:/ et de /ɛ/ est négligeable.

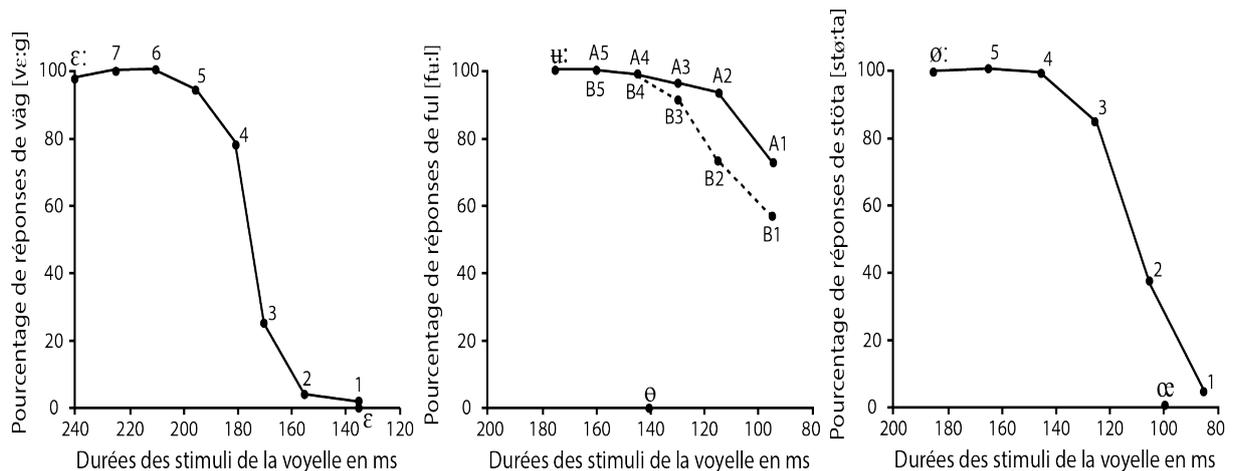


FIGURE 1.8 – Pourcentages des réponses des locuteurs suédois sur les stimuli entendus dans les trois expériences (d'après Hadding-Koch et Abramson (1964)).

La deuxième expérience consiste à examiner les voyelles /ɯ:/ (durée = 175 ms) et /ø/ (durée = 140 ms) dans les mots /fɯ:l/ et /føl/. Elles ont été choisies pour illustrer les voyelles ayant une différence conséquente de qualité. Un changement spectral important est observé vers le milieu de la voyelle dans le mot /fɯ:l/, suggérant ainsi la présence d'une diphtongaison de la voyelle. Ce phénomène est courant dans les voyelles longues et fermées en suédois. Les auteurs ont examiné quelle partie de la voyelle (initiale ou finale) fournit l'indice le plus fort pour la perception de /ɯ:/. Deux séries de découpage de la voyelle ont ainsi été effectuées. La première commence au début et s'achève au milieu de la voyelle tandis que la seconde débute de la fin et se termine au milieu. Dans les deux séries, les transitions formantiques restent intactes. La voyelle est coupée progressivement en cinq étapes dans chacune de ces séries. La durée est réduite de 15 ms à chaque fois en partant de la durée originale de la voyelle longue (175 ms) jusqu'à 100 ms qui est une durée inférieure à celle de la voyelle courte (140 ms). La

Figure 1.8.B affiche les 130 réponses des auditeurs sur le mot /fʊ:l/ dans les deux séries (série avec la portion stable non coupée (A) et série avec la portion diphtonguée non coupée (B)). D'après ces résultats, les auteurs indiquent que la durée a un rôle minimal dans la distinction de ces voyelles. Le stimulus est toujours perçu comme un /ʊ:/ quand la durée de la voyelle est égale ou un peu inférieure à celle de /ə/. De plus, les stimuli les plus courts à 90 ms ont été entendus dans la majorité des cas comme une voyelle longue. Les résultats montrent également que la partie stable de cette voyelle porte plus d'informations que celle diphtonguée puisque la perception du stimulus comme une voyelle longue est meilleure lorsque la portion finale est supprimée. La mesure des fréquences des trois premiers formants révèle que les deux voyelles sont séparées par une différence spectrale importante.

La troisième expérience teste l'importance temporelle dans l'opposition des voyelles /ø:/ (durée = 185 ms) et /œ/ (durée = 100 ms) dans les mots /stø:ta/ et /stœta/. Ces voyelles ont été sélectionnées afin de représenter les paires où une différence de la qualité entendue est moyenne (intermédiaire par rapport aux deux paires précédentes). La voyelle cible dans /stø:ta/ a été raccourcie de 20 ms graduellement en cinq étapes en partant de sa durée originale (185 ms) à 85 ms (soit 15 ms inférieur à la durée de /œ/). Les résultats dévoilent une grande influence de la durée sur la décision des auditeurs (Figure 1.8.C). Entre les étapes 3 et 2, la courbe chute de 84 % à 38 % démontrant que la longueur de la voyelle est un indice primordial dans la séparation des deux voyelles. L'analyse des trois premiers formants souligne que les différences de fréquences sont intermédiaires en comparaison des deux cas précédents.

Les trois expériences ont permis d'exposer le degré de participation de la quantité (durée vocalique) par rapport à celui de la qualité (information spectrale) dans trois cas différents. Une comparaison des schémas de la Figure 1.8 laisse suggérer que quand la distance spectrale est importante entre les deux voyelles (comme pour /ʊ:/-/ə/), l'effet de la durée dans leur distinction est réduit ou absent. De même, lorsque la différence spectrale est minimale (/ɛ:/-/ɛ/) ou moyen (/ø:/-/œ/), l'information temporelle est indispensable pour distinguer les deux voyelles.

Le même phénomène a également été observé en arabe syrien par Allatif (2008) (voir aussi Allatif (2004, 2007)). L'auteur examine l'influence des variations temporelles sur la perception de trois paires de voyelles : /i:/-/i/, /a:/-/a/ et /u:/-/u/. Dans une première expérience, Allatif (2008) étudie l'impact des variations temporelles provoquées par le changement du débit de parole et des schémas mélodiques sur la quantité (durée) et sur la qualité (informations spectrales) des voyelles. Le but de cette expérience est de tester si les voyelles courtes et longues vont se chevaucher dans leurs durées et/ou dans leurs informations spectrales lorsqu'elles subissent différentes conditions d'allocution. Trois locuteurs de l'arabe syrien ont produit des mots réels de type /zVr/ dans une phrase porteuse où V = /i/, i, a:, a, u:, u/. Les mots ont été produits dans quatre conditions expérimentales soit deux débits de parole (normal et rapide) et deux schémas mélodiques (déclaratif et interrogatif). Chaque mot a été répété 10 fois dans les quatre conditions. L'auteur explique que le débit rapide est supposé réduire la durée et centraliser les

fréquences dans l'espace F1-F2 tandis que l'interrogation a pour effet d'augmenter la durée et de rendre les fréquences plus périphériques dans le domaine F1-F2. Les résultats montrent que les voyelles courtes en débit normal avec le schéma déclaratif (ND) ont une durée équivalente à 60 % de celle des voyelles longues. Le débit rapide avec le schéma déclaratif (RD) réduit la durée des voyelles longues et courtes de 19 % comparée à celle du débit normal, ce qui correspond aux durées les plus courtes dans les quatre conditions. Quant à l'interrogation en débit normal (NI), elle augmente la durée des voyelles (longues et courtes) de 11 %. La condition des voyelles en débit rapide avec le schéma interrogatif réduit leurs durées de 13 %.

Il apparaît que les voyelles longues sont significativement plus longues que celles courtes en ND et en NI. Toutefois, l'écart temporel entre les voyelles longues et courtes est réduit dans les conditions de débit rapide même si la différence demeure toujours significative. Allatif (2008) a examiné l'impact de ces variations temporelles sur les fréquences formantiques dans le domaine F1-F2 (Figure 1.9). Les fréquences des formants des voyelles courtes sont plus centrées compa-

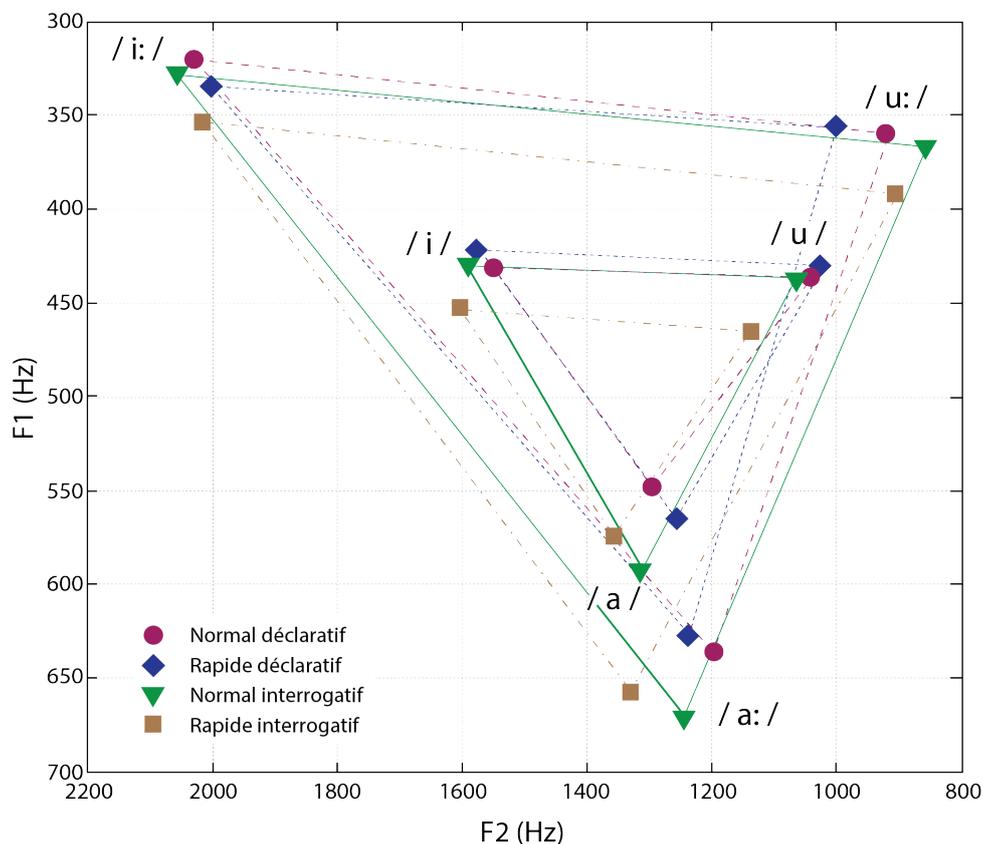


FIGURE 1.9 – Moyennes des 3 voyelles courtes /i, a, u/ et leurs correspondantes longues /i:, a:, u:/ de l'arabe syrien dans 4 conditions expérimentales : 2 schémas mélodiques (déclaratif et interrogatif) et 2 débits de parole (normal et rapide), réalisées par 3 locuteurs masculins avec 10 répétitions par locuteur et par voyelle (d'après Allatif (2008)).

rées à celles des voyelles longues. Pour toutes les voyelles confondues, le débit rapide réduit la distance entre les F1 des voyelles longues et courtes. Ce débit provoque principalement la centralisation des voyelles longues. L'interrogation, quant à elle, réduit les distances entre les F1

des voyelles longues et courtes. Il a été constaté que, dans toutes les conditions, les fréquences des F1 des voyelles longues restent significativement distinctes de celles des courtes. De plus, le débit rapide diminue également la distance entre les F2 des voyelles longues et courtes alors que l'interrogation l'augmente pour certaines voyelles et la réduit pour d'autres. Contrairement à F1, l'écart entre les valeurs du F2 des deux catégories n'est pas significatif pour les trois paires dans les quatre conditions. En schéma déclaratif, l'écart de F2 entre les voyelles /a:/-/a/ est significatif en débit normal néanmoins, il devient non significatif en débit rapide. En schéma interrogatif, la différence n'est pas significative dans les deux débits. L'écart de F2 de /i:/-/i/ est le plus important parmi les trois paires puisqu'il est significatif dans les quatre conditions. Quant à la différence de F2 entre /u:/-/u/, il n'est significatif que dans le schéma interrogatif. En résumé, Allatif (2008) déclare que, face aux différentes conditions expérimentales, l'écart de « [...] la qualité vocalique n'est pas robuste pour A/AA [/a:/-/a/], robuste pour I/II [/i:/-/i/] et à moitié robuste pour U/UU [/u:/-/u/] (elle résiste à deux perturbations sur 3) »<sup>13</sup>.

Dans une deuxième expérience, l'auteur analyse l'impact de la variation temporelle sur la perception de l'opposition des voyelles longues vs. courtes. Il manipule la durée des deux catégories vocaliques de façon à neutraliser la différence temporelle entre la voyelle longue et sa contrepartie courte. Les séquences produites par un des trois locuteurs dans le schéma déclaratif en débit normal ont été choisies pour la manipulation (6 voyelles × 10 répétitions). Chaque voyelle courte a été allongée pour que sa durée soit égale à sa contrepartie longue. L'allongement est réalisé en copiant le cycle délimité par deux passages à zéro au début de la partie stable de la voyelle. Cette portion copiée est ensuite collée après le cycle original. L'opération est répétée jusqu'à la durée souhaitée. Au contraire, les voyelles longues ont été réduites par la suppression des portions de l'état stable jusqu'à obtenir la durée de leurs contreparties courtes. Les mots non manipulés et ceux manipulés ont été présentés dans un ordre aléatoire et écoutés par dix auditeurs. Ces derniers devaient écrire le mot qu'ils entendaient lorsque celui-ci était clair auditivement. Dans le cas contraire, ils cochaient la case "signal pas clair".

Voyelles non manipulées			Voyelles manipulées				
	Identification correcte	Identification erronée	Signal pas clair		Réussite de la modification	Échec de la modification	Signal pas clair
/i/	94	4	2	/i/ → /i:/	21	40	39
/i:/	100	0	0	/i:/ → /i/	15	68	17
/a/	87	3	10	/a/ → /a:/	81	4	15
/a:/	98	0	2	/a:/ → /a/	79	11	10
/u/	77	19	4	/u/ → /u:/	52	20	28
/u:/	95	2	3	/u:/ → /u/	43	43	15

TABLE 1.3 – Pourcentages des réponses des dix auditeurs concernant les signaux non modifiés (à gauche) et modifiés (à droite) écoutés. L'auteur considère que la manipulation est réussie lorsque le sujet identifie sans ambiguïté le signal modifié dans le sens de la manipulation (Allatif, 2008).

Les résultats des données non modifiées montrent que certaines séquences ont été mal identifiées ou jugées comme "pas claires" (Table 1.3). Les taux d'identification correcte des voyelles longues sont tous élevés (98 %). Quant aux voyelles courtes, les taux d'identification correcte sont moins importants, notamment pour le /u/ qui a été perçu comme /u:/ 19 % des cas. Le /i/ et le /a/ sont rarement confondus avec leurs contreparties longues.

En ce qui concerne les données modifiées, les résultats indiquent que la perception des voyelles est influencée à des niveaux différents lors de la neutralisation de la durée. La paire /a:/-/a/ est la plus sensible à la modification de la durée, laissant suggérer alors que l'opposition au sein de cette catégorie dépend essentiellement de la durée. La paire /i:/-/i/ est la moins sensible aux changements temporels, étant donné que les deux voyelles sont distinguées principalement par leurs informations spectrales. Quant à la paire /u:/-/u/, sa sensibilité à la variation de la durée se situe entre les deux autres catégories. D'après ces observations, Allatif (2008) déclare que les auditeurs n'utilisent pas le paramètre de la durée au même degré pour affirmer l'identité vocalique. L'auteur résume que « Pour l'opposition I/II, l'information sur la catégorie de la voyelle est essentiellement fournie par la distance spectrale entre le /i/ bref et le /i:/ long, puis par la durée de la voyelle [...]. Pour l'opposition A/AA, l'information est supportée principalement par la durée de la voyelle, secondairement par la différence spectrale [...]. Pour l'opposition U/UU, l'information sur la catégorie de la voyelle n'est pas fournie massivement par un seul indice. Les indices sont : durée de la voyelle, distance spectrale entre le /u/ bref et le /u:/ long ». <sup>14</sup>

Dans cette section, il a été montré que la durée vocalique est influencée par plusieurs facteurs. Les changements de la durée d'une voyelle peuvent avoir ou non des conséquences sur la fréquence des formants et sur l'identité de la voyelle. Ainsi, la durée vocalique joue dans certains cas un rôle secondaire et dans d'autres un rôle principal dans la perception de la voyelle. En effet, l'importance de l'information temporelle augmente lorsque les deux voyelles sont spectralement proches ou lors de la perte d'informations spectrales. Les mêmes tendances sont observées dans les langues ayant une opposition phonologique de durée où la différence temporelle peut être plus importante que celle spectrale (comme pour la paire /a:/-/a/ en arabe syrien). Contrairement à l'état d'équilibre - ou cible vocalique -, la durée est considérée comme un trait dynamique de la voyelle. Des recherches sur la perception de la parole ont révélé que d'autres traits dynamiques, tels que les transitions formantiques et le changement spectral intrinsèque, portent des informations aussi importantes que les indices statiques (la cible vocalique). La section suivante présente les caractéristiques dynamiques de la voyelle et leur importance dans la perception.

### 1.2.3 Caractéristiques dynamiques de la voyelle

L'idée selon laquelle les voyelles sont perçues grâce à leurs cibles (statiques) a été remise en question par plusieurs recherches. Une des premières études s'intéressant à ce sujet est celle de Peterson et Barney (1952). Les auteurs enregistrent des séquences /hVd/ produites par 76 locuteurs américains (33 hommes, 28 femmes et 15 enfants) où V = /i, ɪ, ε, ɜ, æ, α, ɔ, ʊ, u, ʌ/. La mesure de l'état d'équilibre de la voyelle dans chaque séquence a été déterminée pour les trois premiers formants "loin des influences des deux consonnes adjacentes". Une représentation du domaine F1-F2 de ces données est affichée dans la Figure 1.10. Les syllabes /hVd/ ont été écoutées par 70 auditeurs américains pour une tâche de perception des dix voyelles. Les auditeurs ont identifié les voyelles comme elles ont été envisagées par les locuteurs dans

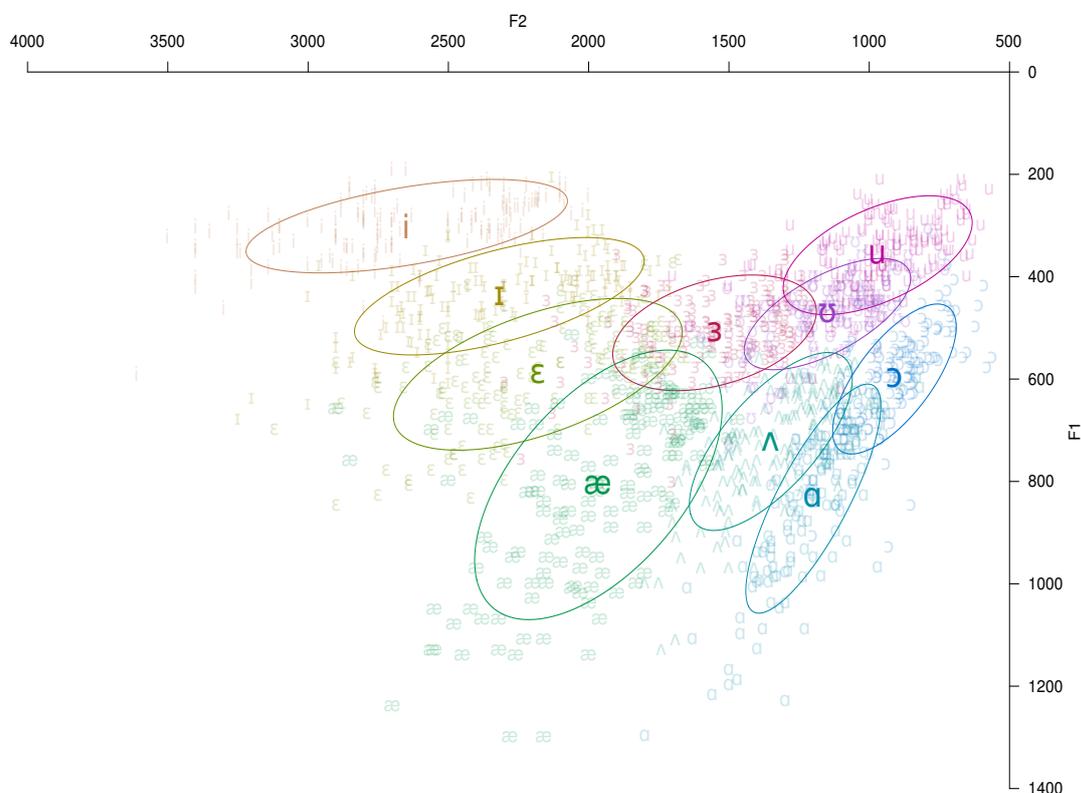


FIGURE 1.10 – Représentation des mesures de F1-F2 (en Hz) des 10 voyelles de l'anglais américain produites par 33 hommes, 28 femmes et 15 enfants (d'après Peterson et Barney (1952))<sup>15</sup>.

plus de 94 % des cas. D'après ce test de perception, les auditeurs n'ont pas eu de difficulté à reconnaître les voyelles malgré le grand chevauchement acoustique noté dans le domaine de F1-F2 (Figure 1.10). Il est donc clair que les auditeurs utilisent d'autres indices (supplémentaires) que celles des cibles vocaliques de F1 et de F2 (Hillenbrand, 2013). Peterson et Barney (1952) expliquent alors que l'ensemble des informations acoustiques d'une voyelle ne peut pas être décrit par une simple "section mesurée" (ou un point de mesure) mais il exige une représentation plus complexe. Ils ajoutent que les parties initiales et finales de la voyelle sont des

éléments essentiels et leur examen peut fournir des informations fondamentales.

En effet, les voyelles peuvent être produites de manière quasiment stable et en isolation. Cependant, cette réalisation n'est qu'un artefact de laboratoire (Strange et al., 1976). Les voyelles sont normalement coarticulées avec des consonnes et dans un contexte syllabique. Elles varient alors en durée et en fréquence de façon continue. Strange et ses collègues étaient parmi les premiers à analyser la caractéristique dynamique de la voyelle (Strange et al., 1976, 1983; Jenkins et al., 1983). Dans la première étude, les chercheurs ont conçu deux expériences de perception avec plusieurs conditions expérimentales afin d'étudier l'importance de la stabilité de la voyelle en parole naturelle (Strange et al., 1976). La première expérience consiste à examiner la perception des voyelles en contexte consonantique symétrique (/pVp/) comparée à celle de la voyelle en isolation. La seconde analyse l'impact de la variation du contexte consonantique sur la perception de la voyelle (CVC où  $C \neq C$ ) et les résultats sont comparés à ceux de la première expérience.

Les résultats des deux expériences soulignent que les auditeurs identifient mieux les voyelles en contexte consonantique identique ou varié que celles en isolation réalisées par les mêmes locuteurs. Les voyelles en contexte consonantique prononcées par les différents locuteurs (âges et sexes confondus) ont été correctement catégorisées à 80 % alors que les voyelles en isolation ont été reconnues dans moins de 60 % des cas dans la même condition. De plus, cette différence entre les deux conditions contextuelles a également été observée quand les auditeurs écoutaient les séquences articulées par un même locuteur. Les auteurs ont alors suggéré que des informations importantes de l'identité de la voyelle sont contenues dans les transitions formantiques (parties dynamiques de la voyelle). Ils argumentent que, d'un point de vue perceptuel, « *the definition of a vowel ought to include a specification of how the relevant acoustic parameters change over time* »<sup>15</sup>.

À partir des résultats obtenus dans l'article de Strange et al. (1976), les chercheurs ont émis l'hypothèse que le début et la fin des portions transitionnelles des syllabes fournissent des informations importantes sur l'identité de la voyelle. Ils mettent en place une série d'expériences de perception dans l'objectif de tester leur hypothèse (Jenkins et al., 1983; Strange et al., 1983). Une méthodologie spécifique a été utilisée dans ces expériences. Trois sources d'informations ont été prises en compte : 1) l'information du milieu de la voyelle (l'état d'équilibre), 2) la durée vocalique et 3) les transitions formantiques (Strange, 1987). Adaptée pour explorer la pertinence de ces sources d'information, la méthodologie implique la modification systématique des syllabes CVC produites par un ou plusieurs locuteurs. Les séquences enregistrées ont été modifiées en supprimant des portions dans le signal et en changeant la relation temporelle entre les portions non supprimées. Dans un premier temps, la portion qui correspond à l'état d'équilibre de la voyelle a été enlevée (remplacée par du silence), tandis que les transitions au

15. Les données de Peterson et Barney (1952) sont disponibles en ligne *via* le lien : <https://www.cs.cmu.edu/Groups/AI/areas/speech/database/pb/0.html>

16. « la définition d'une voyelle doit contenir une mention de comment les paramètres acoustiques pertinents changent en fonction du temps », page 221.

début et à la fin de la séquence ainsi que les relations temporelles entre les différentes portions n'ont pas été changées. Les séquences subissant cette modification sont appelées syllabes avec un centre silencieux (SC)<sup>17</sup> (Figure 1.11).

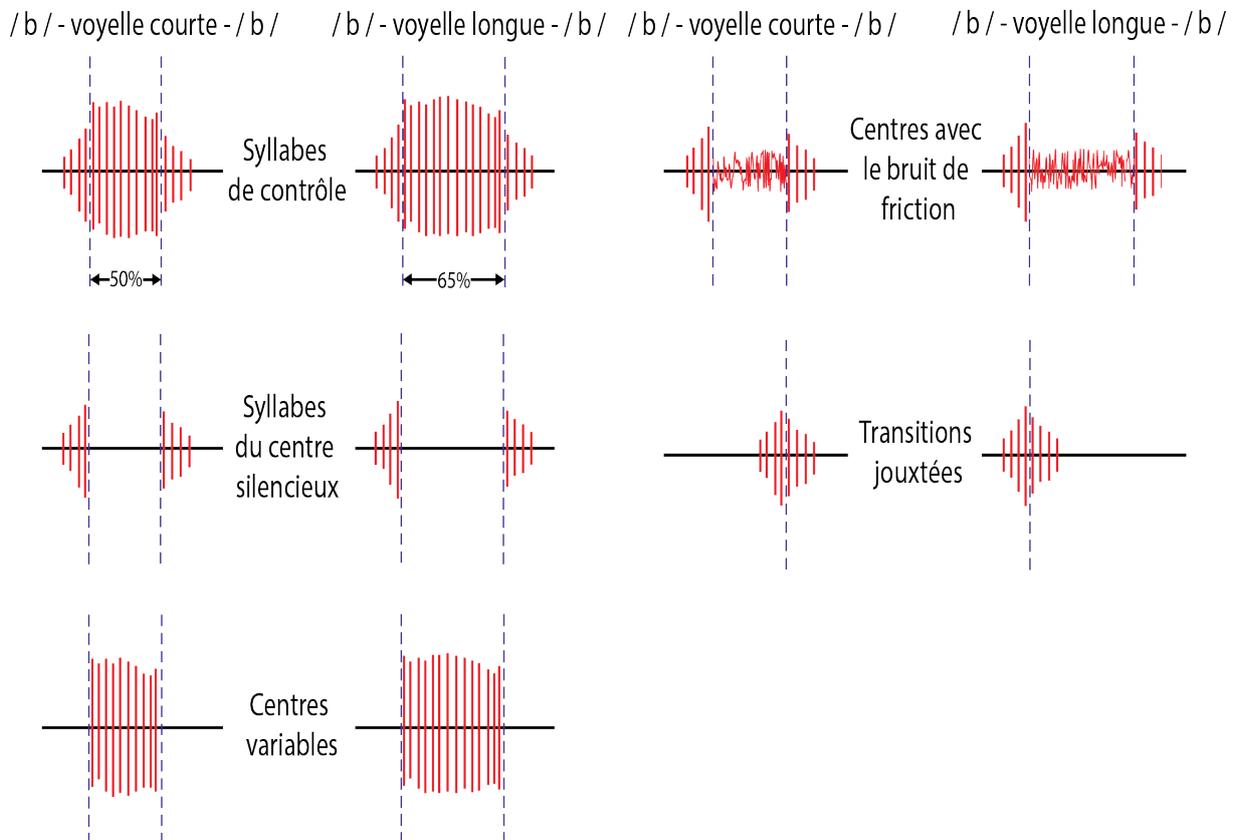


FIGURE 1.11 – Représentation schématique des formes d'onde des stimuli de contrôle, des syllabes du centre silencieux, celles des centres variables, celles du centre avec du bruit de friction et celles avec des transitions jouxtées pour une voyelle courte et pour une voyelle longue (d'après Strange et Jenkins (2013) adaptée de Jenkins et al. (1983)).

Les syllabes SC contiennent alors deux sources, la durée intrinsèque de la voyelle et les transitions formantiques. Dans un deuxième temps, la durée de l'intervalle du silence a été manipulée dans le but d'examiner la contribution relative de la durée à la perception de la voyelle. Deux intervalles silencieux constants ont été créés. Il s'agit de la durée de la voyelle la plus longue (LSC) et celle de la voyelle la plus courte (SSC). Dans un troisième temps, les deux portions de transition (initiale et finale) ont été supprimées mais l'état d'équilibre a été conservé afin d'évaluer l'efficacité perceptuelle de l'information contenue dans le milieu de la syllabe. Ces séquences sont appelées centres variables (VC)<sup>18</sup> (Figure 1.11). Dans cette condition (VC), deux sources d'information ont été préservées : la cible et la durée de la voyelle. Une condition du centre fixe (FC) a été ajoutée où la durée de l'état d'équilibre a été fixée à la durée de la voyelle la plus courte.

17. De l'anglais *Silent-Center syllables*.

18. De l'anglais *Variable Centers*.

Dans ces expériences, les voyelles de l'anglais américain ont été produites dans des séquences /b/+ voyelle +/b/ par un ou quatre locuteurs. Dans une première expérience, Jenkins et al. (1983) ont présenté les syllabes entières /bVb/ comme des syllabes contrôle (C). Ces syllabes ont été modifiées de manière à avoir des séquences SC, VC, HS (*Hiss-Center*) et AS (*Abutted Syllables*). Dans les séquences HS, les centres silencieux ont été remplacés par un bruit de friction de /f/. Dans celles AS, l'intervalle temporel du silence séparant les transitions initiale et finale est réduit à zéro. Les portions de transition initiale et finale ont été présentées séparément (Figure 1.11). Des groupes séparés d'auditeurs ont écouté les stimuli dans des conditions variées. Il en ressort que les syllabes de contrôle et celles du centre silencieux ont été identifiées avec un taux de 93 % environ. Les centres variables ont été reconnus avec un taux légèrement plus bas de 87 % (statistiquement similaires). Quant aux syllabes des centres avec un bruit de friction, elles ont été identifiées avec un taux statistiquement plus faible que celui des syllabes de contrôle (80 %). Un écart plus grand a été observé pour les syllabes jouxtées ainsi que pour celles contenant les portions initiale et finale séparées (respectivement, 70 %, 54 % et 53 %). Ces résultats montrent que les transitions initiales et finales ensemble et leur relation temporelle originale fournissent une information importante sur l'identité de la voyelle.

Dans une deuxième expérience, Jenkins et al. (1983) ont répété la première expérience en ajoutant la condition des centres fixes où les portions du centre ont été fixées à 60 ms. Toutes les séquences ont donc la même durée. Avec ces syllabes, les auditeurs sont privés de l'information de la durée relative et du changement spectral pour les voyelles longues. Les résultats de cette expérience sont quasiment identiques à la première. Les centres fixes a été correctement identifiée à 70 % alors que les centres variables ont été reconnus à 80 %. Selon les auteurs, ces résultats indiquent que la contribution de l'information de la durée relative est faible pour l'identification des voyelles de l'anglais américain (voir aussi Strange et Jenkins (2013)). Ils concluent que les voyelles ont une nature dynamique intrinsèque. D'après ce postulat, l'acte de coarticulation des phonèmes dans des syllabes donne naissance à des informations acoustiques dans lesquelles les consonnes et les voyelles sont co-spécifiées dans une configuration spectrale qui varie dans le temps.

Strange et al. (1983) ont également examiné les paramètres dynamiques des voyelles. Dans une première expérience, la syllabe contrôle correspond aux séquences /bVb/ produites par un locuteur américain dans différents débits de parole. Ces séquences ont été modifiées selon les conditions des centres silencieux, des centres variables et des centres fixes. Dans cette dernière condition, les durées du centre silencieux de toutes les syllabes ont été réduites à la durée de la voyelle la plus courte (57 ms) et ensuite, allongées à la durée de la voyelle la plus longue (163 ms). Les voyelles dans presque toutes les conditions (syllabes contrôle, centres silencieux, centres silencieux raccourcis et centres variables) ont été identifiées correctement avec un taux qui dépasse les 90 %. Néanmoins, les centres silencieux allongés n'ont été reconnus que dans 87 % des cas. La plupart des erreurs résident dans la mauvaise identification des voyelles courtes en leurs contreparties longues. Ces résultats confirment également que les portions de transition

et la relation temporelle sont importantes pour spécifier les voyelles.

Dans une deuxième expérience, Strange et al. (1983) ont répété la première expérience avec des stimuli issus de quatre locuteurs (deux femmes et deux hommes). Les résultats de cette expérience sont similaires à ceux mis en évidence dans la première. L'étude de Strange et al. (1983) a mis en avant que les informations spectrales dynamiques étaient influentes même quand les relations temporelles spécifiant les différences intrinsèques de la durée vocalique sont éliminées. Strange et Jenkins (2013) commentent la dynamicité spectrale et temporelle des voyelles en expliquant « *Acoustic analysis found consistent differences between tense and lax vowels (including lax, but long /æ/) not only in vocalic duration but also in the length and manner of transition into and out of the "vowel target". This reflected the findings of Lehiste and Peterson (1961) that lax vowel production involved not only a short "target" but also a slower release into the final consonant while tense vowels involved both a longer hold in the target position as well as a more rapid closing gesture, as indicated by more rapid offset transitions* »<sup>19</sup>.

Les études de Strange et ses collègues (entre autres) ont montré que les formants des voyelles changent de façon continue lorsqu'elles sont produites en contexte consonantique. Neary et Assman (1986) avancent que cet aspect dynamique des voyelles existe aussi lorsqu'elles sont produites en isolation. Ils précisent que cette dynamicité spectrale intrinsèque contient des informations importantes pour la perception des voyelles en isolation. Le changement spectral inhérent à la voyelle correspond aux variations relativement lentes<sup>20</sup> des fréquences formantiques qui sont associées aux voyelles elles-mêmes en l'absence de contexte consonantique. Neary et Assman (1986) ont conçu un protocole expérimental pour examiner l'importance de l'information dynamique dans la perception de la voyelle en isolation. Une série de 40 voyelles monophthongues de l'anglais canadien prononcées en isolation (10 voyelles × 4 locuteurs dont 2 femmes et 2 hommes) a été utilisée dans cette expérience<sup>21</sup>. Ces voyelles ayant une durée qui varie entre 300 et 550 ms ont servi de stimuli de référence. Deux portions de 30 ms chacune ont été extraites de chaque voyelle. Le noyau du segment (A), correspondant au milieu de la première portion de 30 ms, est situé à 24 % de la durée totale du signal. Le noyau de la portion finale (B), quant à lui, a été déterminé à 64 % de la durée totale de la voyelle (Figure 1.12).

Quatre conditions d'identification ont été établies : 1) le segment de référence, correspondant à la réalisation originale non modifiée, 2) l'ordre naturel (A-B) : le noyau (A) est suivi de 10 ms de silence et de la portion finale (B), 3) le noyau répété (A-A) : le noyau (A) est suivi de

19. « Les analyses acoustiques ont trouvé des différences cohérentes entre les voyelles tendues et relâchées (incluant le /æ/ relâché mais long) pas seulement dans la durée vocalique mais également dans la longueur et le mode de transition vers l'intérieur et l'extérieur de la "cible vocalique". Ceci reflète les résultats de Lehiste et Peterson (1961) où la production des voyelles relâchées n'implique pas seulement une "cible" courte mais également un relâchement plus lent vers la consonne finale. Toutefois, les voyelles tendues entraînent un maintien plus long dans la position de la cible et un geste de fermeture plus rapide, ce qui est indiqué par des transitions finales plus rapides », page 96 [notre traduction].

20. Contrairement aux transitions rapides qui sont associées aux consonnes.

21. Ces données proviennent d'un corpus plus grand de 100 voyelles de l'anglais canadien produites en isolation par 5 femmes et 5 hommes traité dans l'article de Assman et al. (1982).

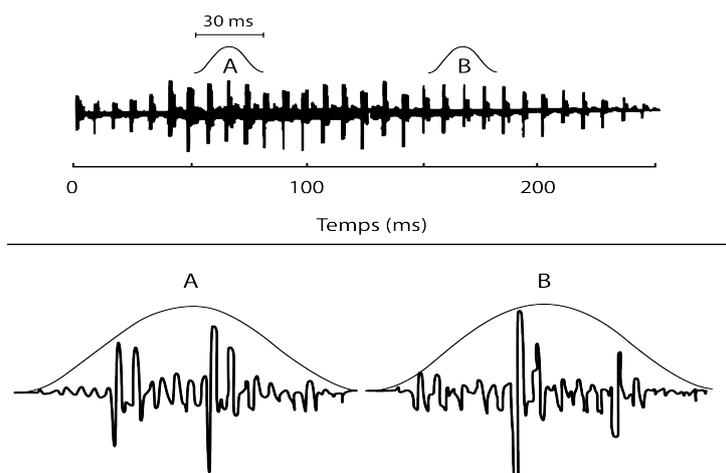


FIGURE 1.12 – Illustration du processus de découpage : en haut, les positions des portions d'un segment typique ; en bas, les détails des deux portions découpées (d'après Neary et Assman (1986)).

10 ms de silence et du même noyau (A), 4) l'ordre inversé (B-A) : la portion finale (B) est suivie de 10 ms de silence et du noyau (A). Les quatre conditions ont été écoutées par 64 auditeurs canadiens. Les résultats montrent que les voyelles ont été mieux perçues dans la première et la quatrième conditions avec un taux d'erreur de 13 % pour la voyelle sans modification et 14 % pour l'ordre naturel (A-B). Le taux de reconnaissance des deux conditions sont statistiquement similaires. Quant aux deux dernières conditions, le taux d'erreur dans l'identification des voyelles est statistiquement plus élevé, avec 31 % pour la condition du noyau répété (A-A) et 38 % pour celle de l'ordre inversé (B-A). Les auteurs déclarent que ces résultats fournissent une preuve claire de l'existence d'informations fiables contenues dans des portions de 30 ms proche du milieu et de la fin du signal pour l'identification des voyelles. Ils ajoutent que l'augmentation marquée de l'erreur dans la condition du noyau répété (A-A) indique que la composante de la portion finale (B) offre une contribution significative et indépendante dans la perception. Le taux d'erreur élevé en condition d'ordre inversé (B-A) révèle que l'ordre temporel naturel des deux portions est essentiel.

Cette expérience a donc confirmé l'importance des changements spectraux intrinsèques de la voyelle. Neary et Assman (1986) ont ensuite analysé l'étendue de ces changements au sein de chacune des dix voyelles. Pour ce faire, les voyelles produites en isolation par les dix locuteurs canadiens (5 femmes et 5 hommes) de l'étude de Assman et al. (1982) ont été utilisées (10 voyelles  $\times$  10 locuteurs). Deux mesures de F1 et de F2 ont été prélevées au début et à la fin du signal<sup>22</sup>.

Les résultats attestent un mouvement statistiquement significatif des voyelles /I, ε, æ, e, o/ dans F1 et/ou F2 (Figure 1.13, à gauche). Cette figure illustre clairement les changements spectraux inhérents aux voyelles. Il est à remarquer que les voyelles antérieures relâchées /I, ε,

22. Les mesures au début du segment ont été prélevées le plus tôt possible lorsque les formants sont mesurables et lorsque le pic d'amplitude globale est de 15 dB. Celles à la fin du segment ont été fixées le plus tard possible quand les formants sont encore mesurables et quand le pic d'amplitude globale est de 15 dB.

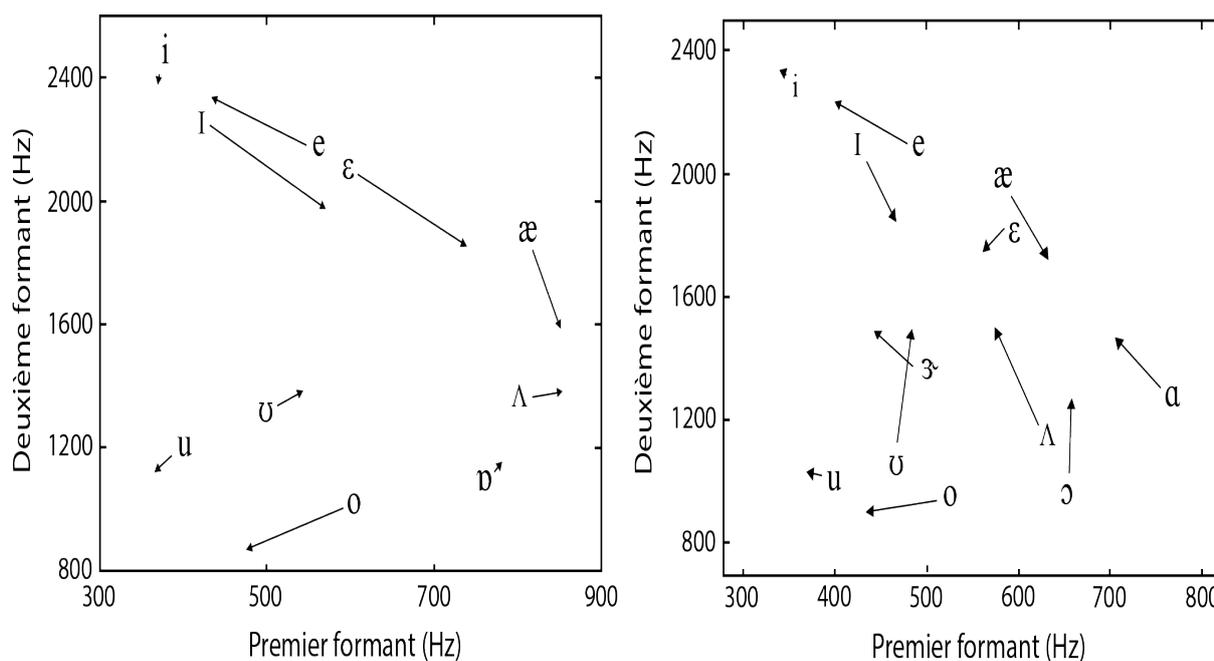


FIGURE 1.13 – À gauche, les moyennes des fréquences des F1 et F2 mesurées au début et à la fin des voyelles de l’anglais canadien réalisées en isolation par 10 locuteurs (Neary et Assman, 1986). À droite, les moyennes des fréquences des F1 et F2 des voyelles de l’anglais américain prononcées dans des séquences /hVd/ par 45 locuteurs masculins (Hillenbrand et al., 1995). La tête des flèches indique la partie finale des voyelles (d’après Hillenbrand (2013)).

/æ/ ont le même comportement. Autrement dit, le mouvement de ces voyelles se dirigent vers une zone centrale où les valeurs de F1 s’élèvent et celles de F2 baissent. Les auteurs expliquent que cette tendance est également observée pour les voyelles en contexte consonantique.

Hillenbrand (2013) offre une comparaison entre les données de l’anglais canadien et celles des voyelles produites en contexte consonantique de l’anglais américain issues de l’étude de Hillenbrand et al. (1995) (Figure 1.13). La figure de droite affiche les valeurs formantiques mesurées à 20 % et à 80 % de la durée vocalique. Dans ces données, les voyelles ont été prononcées par 45 locuteurs masculins dans des séquences /hVd/. Il existe des différences entre les deux dialectes, cependant, quelques caractéristiques communes ont pu être dégagées. Dans les deux cas, les voyelles /i/ et /u/ sont relativement stationnaires et les /e, o, ɪ, æ/ offrent les mêmes tendances (à des degrés différents). Les fréquences du F1 diminuent pour les voyelles /e/ et /o/ et celles du F2 augmentent pour la première et baissent pour la seconde vers la fin du signal. Quant aux voyelles /ɪ/ et /æ/, les fréquences du F1 mais également celles du F2 décroissent. Hillenbrand (2013) signale que la paire de voyelles /æ/ et /ɛ/ de l’étude de Hillenbrand et al. (1995) dévoile une situation intéressante. Le /æ/ est élevé et antérieur dans ce dialecte, ce qui le place dans une zone très proche du /ɛ/, provoquant un grand degré de chevauchement spectral. Cependant, les auditeurs de cette étude ont rarement confondu le /æ/ et le /ɛ/ avec un taux d’identification correct qui dépasse les 94 % pour les deux voyelles. L’auteur explique que la grande intelligibilité de ces voyelles est maintenue malgré le chevauchement car elle est liée en

partie aux différences de durée entre les deux voyelles.

En résumé de cette sous-section, les voyelles se distinguent des autres catégories de son par la simplicité de leur description acoustique et cela est dû aux fréquences des trois premiers formants. Le domaine F1-F2 est la représentation la plus connue de la description acoustique des voyelles permettant la différenciation de la qualité des voyelles. Plusieurs recherches (citées plus haut) ont montré que cette représentation des informations statiques de la voyelle est trop simplifiée. En effet, la plupart des voyelles monophthongues exposent un mouvement spectral pendant leur réalisation, même pour celles produites en isolation. Ces changements spectraux aident à la séparation statistique des voyelles et à leur identification par les auditeurs. Il semble alors que les informations statiques ne sont pas nécessaires à la perception des voyelles (Hillenbrand, 2013). Néanmoins, la représentation classique des voyelles selon leurs caractéristiques statiques et en fonction de leur conceptualisation comme des points stables reste toujours largement utilisée dans les domaines acoustique et perceptuel. Cette vision statique est une simplification adéquate qui peut être utile dans certaines situations mais dans d'autres cas, elle pourrait être inappropriée.

La section précédente a ainsi présenté une description des propriétés acoustiques des segments caractérisant la partie ouverte de la parole soit les voyelles. La section suivante sera consacrée à la description des caractéristiques acoustiques d'une des classes de la partie fermée soit les consonnes occlusives.

#### 1.2.4 Les consonnes occlusives

Les consonnes occlusives ont des propriétés articulatoires et acoustiques plus compliquées que celles des voyelles ou celles d'autres classes de consonnes comme les fricatives. Elles se distinguent par ailleurs des autres catégories consonantiques par la fermeture totale momentanée à un point donné dans le conduit vocal pendant la production et par la brièveté du bruit de friction engendré par le relâchement. Les lieux d'articulation des occlusives sont décrits en fonction du point où a lieu la fermeture dans le conduit vocal. Par exemple, pour les occlusives bilabiales, il s'agit du rapprochement des deux lèvres jusqu'à la clôture totale, pour les occlusives alvéolaires, de l'élévation de l'apex de la langue contre les alvéoles et pour les occlusives vélaires et uvulaires, de l'élévation du corps de la langue jusqu'au contact du dos de la langue avec le palais. La production des occlusives comprend trois phases principales :

- la clôture totale du conduit vocal,
- le relâchement de la clôture qui génère un bruit de frottement des articulateurs
- et la transition articulatoire rapide vers le son suivant (Figure 1.14).

Avant la fermeture, les articulateurs concernés se déplacent de leur position initiale vers la formation de la consonne en question. Pendant la fermeture, les articulateurs entrent en contact pour bloquer le flux d'air oral, augmentant la pression dans la bouche. Cette fermeture est

associée à un silence acoustique qui peut être accompagné ou non de la barre horizontale du voisement selon la nature de l'occlusive : voisée ou non voisée respectivement (Figure 1.15). Ce silence est un indice important pour la perception de l'occlusive. En effet, si ce silence acoustique est rempli par n'importe quel autre type de son - à part le voisement, l'occlusive n'est pas perçue (Halle et al., 1957).

Il existe deux contraintes physiologiques pour que le voisement ou la vibration des plis vocaux ait lieu pendant la fermeture. Premièrement, les plis vocaux doivent être dans une configuration appropriée, c'est-à-dire assez rapprochés (en *adduction*). Deuxièmement, le flux d'air qui passe à travers les plis vocaux doit être suffisant. Si l'une ou les deux conditions ne sont pas réunies, le voisement ne se produira pas (Ohala, 1983). En outre, plus ces deux contraintes sont regroupées pendant la production d'un segment non voisé, plus la probabilité que ce segment devienne voisé est élevée. La durée de la fermeture joue, de même, un rôle important sur le voisement puisque plus la durée est longue, moins le voisement est marqué. Par conséquent, la tendance à ce que les occlusives voisées longues (géménées) deviennent non voisées est forte (Ohala, 1983).

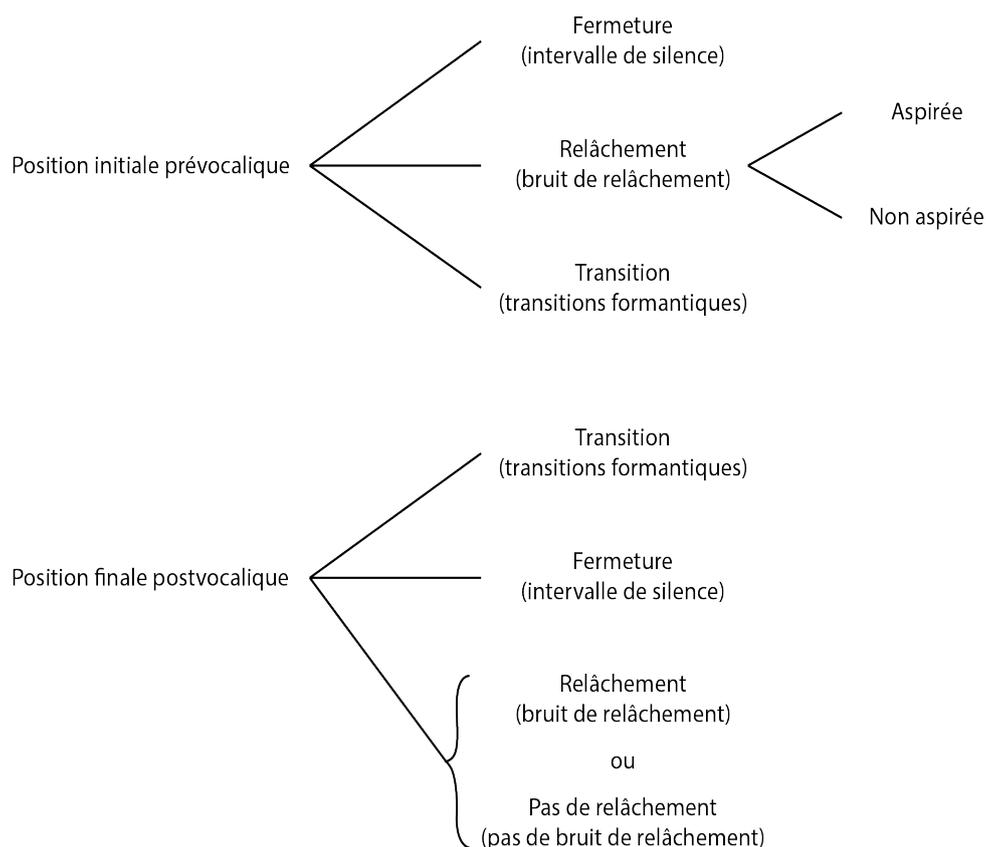


FIGURE 1.14 – Phases principales et conséquences acoustiques des occlusives en position syllabique initiale et finale (d'après Kent et Read (1996)).

Au moment du relâchement des articulateurs, la pression est brusquement relâchée. Un bruit sera généré suite à cet événement, appelé bruit de relâchement (*burst noise* en anglais). Ce dernier peut comprendre trois phases soit *le transient*, *la friction* et quelques fois *l'aspiration*

(Repp et Lin, 1989; Fant, 1973; Harrington, 2010). Le transient correspond au moment du relâchement traduit acoustiquement par une barre verticale foncée visible sur un spectrogramme (Figure 1.15). Il marque alors le moment de l'ouverture de l'occlusive. Celui-ci est souvent suivi d'un court bruit de friction. Les propriétés acoustiques de ce bruit sont similaires à celles des consonnes fricatives produites aux même lieux d'articulation. Lorsque l'occlusive est dite aspirée, le bruit de friction sera suivi d'un bruit caractérisé par une énergie inférieure à 1 KHz (Harrington, 2010). Le bruit de l'aspiration est généré quand le flux d'air passe par les plis vocaux partiellement fermés et par le pharynx. La source de ce bruit, la glotte, n'est donc pas la même que celle des deux autres bruits (le transient et la friction). Ce bruit correspond essentiellement à celui de la fricative glottale /h/ (Kent et Read, 1996).

Le relâchement d'une occlusive en position initiale entraîne un changement dans le conduit vocal dès l'occlusion jusqu'à la configuration de la voyelle. La transition articulatoire d'une occlusive à une voyelle est associée à une transition acoustique sous la forme de changements rapides dans les fréquences des formants appelés transitions formantiques. Ces dernières représentent un indice acoustique important pour la perception de la voyelle et de la consonne (*cf.* section 1.1.4).

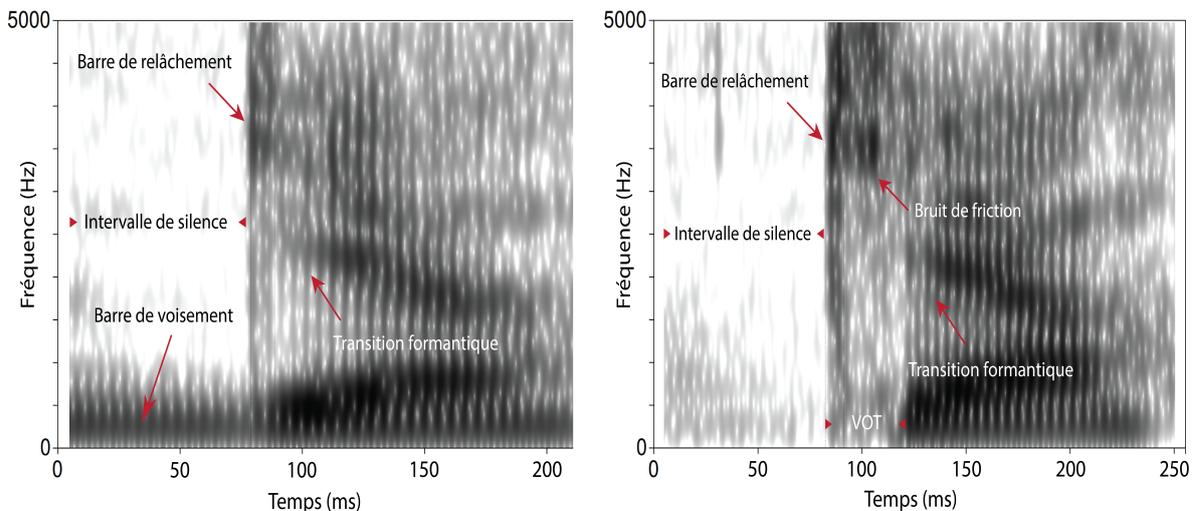


FIGURE 1.15 – Spectrogrammes des séquences /ga/ à gauche et /ka/ à droite produites par une locutrice du français affichant l'intervalle de silence, la barre de relâchement, la transition du F2, la barre de voisement pour l'occlusive voisée, le bruit de friction et le VOT pour l'occlusive non voisée.

Lorsqu'une occlusive est prononcée en position finale du mot (postvocalique), sa réalisation commence par une transition de la configuration de la voyelle précédente (phase ouverte) à la fermeture totale (Figure 1.14). Similairement à la position initiale, la phase de fermeture est accompagnée d'un intervalle de silence avec ou non une barre de voisement. L'occlusive sera ensuite relâchée ou non. Dans le premier cas, une barre verticale du bruit de relâchement sera visible sur le spectrogramme tandis que dans le second, la fermeture est maintenue jusqu'à la

fin de la séquence et de la dissipation de la pression orale. Le bruit de relâchement n'apparaît donc pas.

Les caractéristiques acoustiques des phases de la réalisation des occlusives citées plus haut constituent des indices indispensables pour leur identité en tant que consonnes occlusives. Ces dernières se différencient entre elles par le voisement et les lieux d'articulation. Les sections suivantes décrivent les indices qui contribuent à la perception des différentes occlusives. Parmi eux, le bruit de relâchement et les transitions formantiques aident à la reconnaissance du lieu d'articulation alors que le VOT les sépare entre consonnes voisée et non voisée (Figure 1.15).

#### 1.2.4.1 Le bruit de relâchement

La phase du relâchement débute à la fin de la fermeture de la consonne et se termine à la première pulsation de la voyelle suivante (dans des séquences CV). Comme mentionné précédemment, elle pourrait être divisée en trois entités sur un spectrogramme : la barre d'occlusion (le transient), le bruit de friction et, en fonction des cas, un silence ou un bruit d'aspiration avec ou sans barre de voisement. Les fréquences du bruit de relâchement ou *burst*<sup>23</sup> changent en fonction du lieu d'articulation. Par conséquent, il est considéré comme un indice important dans la perception des différentes occlusives. Les labiales ont tendance à dominer dans les fréquences basses, les alvéolaires sont associées à de l'énergie dans les fréquences hautes et les vélares sont caractérisées par des fréquences moyennes du bruit de relâchement (Kent et Read, 1996).

Liberman et al. (1952) utilisent la technique du *pattern playback* pour générer des stimuli de la parole synthétique. L'objectif de cette étude est d'identifier quelles sont les fréquences possibles du bruit de relâchement qui permettent la perception des consonnes occlusives /p, t, k/. Douze positions de fréquence des bruits de relâchement schématiques variant entre 360-4320 Hz ont été combinées à sept voyelles avec deux formants statiques (Figure 1.16.A et B respectivement). La dimension verticale de chaque bruit de relâchement est de 600 Hz et celle horizontale est de 15 ms. Un exemple du stimulus après l'association du bruit de relâchement avec une voyelle est donnée dans la Figure 1.16. Au total, 84 séquences ont été créées par le *pattern playback* et ensuite écoutées deux fois par 30 auditeurs. Il leur a été demandé d'identifier les consonnes sur une échelle de certitude de quatre points.

Les réponses des participants ne dépendent pas seulement de la fréquence du bruit de relâchement mais également de la voyelle suivante pour les consonnes /k/ et /p/. L'occlusive vélaire /k/ a été identifiée lorsque les fréquences du bruit de relâchement sont égales ou légèrement supérieures à la fréquence du F2. La distribution des réponses montre un mouvement des réponses de /k/ sur l'échelle des fréquences qui monte des voyelles postérieures à celles antérieures, semblant suivre les fréquences du deuxième formant des voyelles (Figure 1.16). Les voyelles /ε/ et /e/ connaissent une seconde distribution de la reconnaissance du /k/ lorsque

---

23. Les deux termes seront utilisés alternativement dans cette étude : *bruit de relâchement* en français et *burst* en anglais.

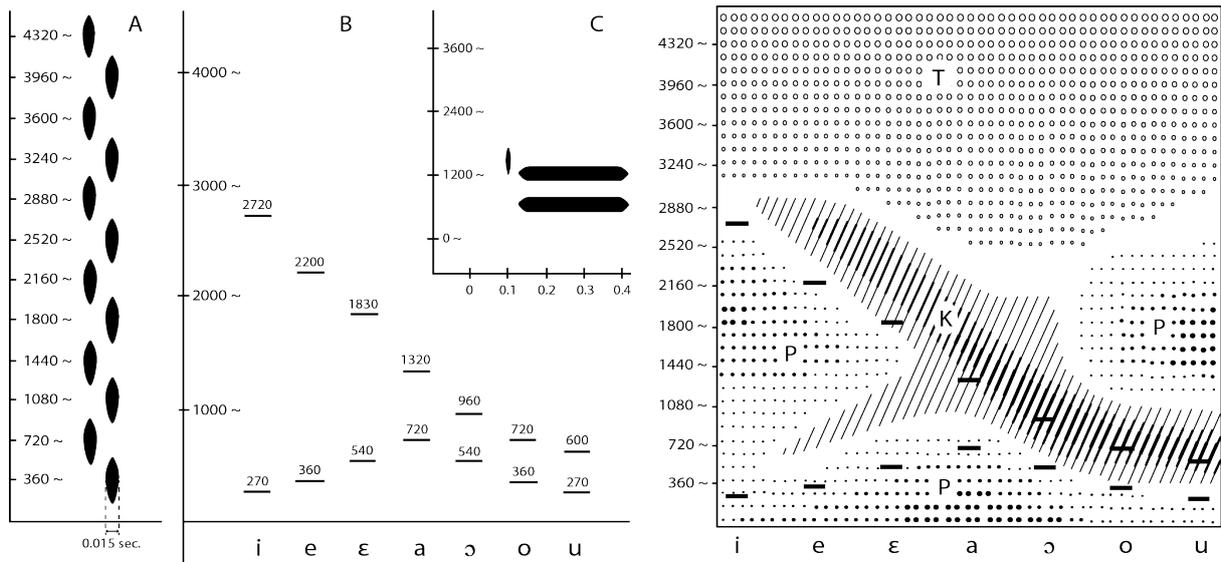


FIGURE 1.16 – À gauche, représentation des stimuli synthétiques utilisés dans l'expérience de perception des différents bruits de relâchement où A correspond aux 12 fréquences du *burst* employées, B désigne les fréquences des deux premiers formants des sept voyelles utilisées et C affiche un exemple de stimulus après l'association d'un bruit de relâchement avec une voyelle. À droite, résultats de l'identification des auditeurs avec le niveau de certitude (d'après Liberman et al. (1952)).

les fréquences du bruit de relâchement sont légèrement au-dessus de celles du premier formant. L'identification du /k/ avec la voyelle /i/ est faible. Par conséquent, Liberman et al. (1952) énoncent que d'autres indices sont nécessaires à la perception du /k/ devant cette voyelle. Quant à l'occlusive bilabiale /p/, elle est reconnue quand les fréquences du *burst* sont relativement basses et toujours inférieures à la fréquence du F2 (à l'exception du F2 des voyelles /o, u/). Contrairement aux /k/ et /p/, l'occlusive alvéolaire /t/ est très peu impactée par le contexte vocalique. Elle a été identifiée quand les fréquences du bruit de relâchement sont élevées (+ de 3000 Hz) et supérieures à la fréquence du second formant.

Halle et al. (1957) s'interrogent, quant à eux, sur l'importance du bruit de relâchement dans la discrimination des différentes occlusives. Ils posent la question suivante : « *Is a listener able to identify stop bursts isolated from their context as /p/, /t/ or /k/?* »<sup>24</sup>. Pour y répondre, les auteurs ont établi un protocole expérimental dans lequel trois locuteurs ont produit les occlusives /b, d, g, p, t, k/ avec les voyelles /i, ɪ, ʌ, ɑ, u/ dans des mots monosyllabiques de l'anglais. Le *burst* des occlusives non voisées en position finale a été sélectionné pour une expérience de perception. Les premières 20 ms du bruit de relâchement (sans trace de transition) ont été retenues et entendues par les participants de l'expérience. Les auditeurs ont reconnu correctement les différentes occlusives à partir de leur *burst* en isolation avec un taux relativement élevé de bonnes réponses (77 %). Les auteurs émettent ainsi l'hypothèse que l'indice qui rend possible l'identification des différentes occlusives seulement à partir des *bursts* réside dans le spectre.

24. « L'auditeur est-il capable d'identifier les occlusives uniquement à partir de leurs bruits de relâchement ? »

Ils effectuent alors des analyses sur le spectre du bruit de relâchement des occlusives afin de tester cette hypothèse. En examinant le spectre de leur corpus, trois classes associées à trois lieux d'articulation ont été déterminées :

- l'énergie du bruit de relâchement des labiales /p, b/ se concentre dans les fréquences basses (500-1500 Hz).
- celle des alvéolaires /t, d/ se situe dans les fréquences hautes (supérieures à 4000 Hz) mais également dans la zone de fréquence autour de 500 Hz.
- celle des vélares /k, g/ est localisée dans une zone de fréquence intermédiaire (1500-4000 Hz).

Halle et al. (1957) ajoutent qu'il existe une différence spectrale marquée de l'opposition relâchée vs. tendue. Le spectre des occlusives relâchées (voisées) contient des composantes d'énergie forte dans les fréquences basses tandis qu'une chute d'énergie est attestée dans les fréquences hautes. Cette perte d'énergie est la conséquence de la diminution de la pression de l'air associée à la production des occlusives voisées, contrairement aux occlusives tendues qui ont moins de perte d'énergie. Ils ont observé, de plus, que le *burst* des vélares /k, g/ varie en fonction de la voyelle adjacente antérieure vs. postérieure. Devant les voyelles antérieures, la concentration de l'énergie est localisée entre 2000 et 4000 Hz alors que, dans le cas où ces occlusives sont produites avec des voyelles postérieures, elle se situe dans des fréquences plus basses. Les auteurs expliquent que cette variation spectrale est due partiellement au changement de volume du conduit vocal qui bloque le flux d'air selon les voyelles (voyelles antérieures = volume plus important).

Malécot (1958) a également examiné la quantité d'informations que le bruit de relâchement des occlusives /b, d, g, p, t, k/ fournit sur la perception de leurs lieux d'articulation (bilabial, alvéolaire ou vélaire) et sur le voisement (voisée ou non voisée). Il explique que les occlusives dans la position finale en anglais alterne librement entre relâchées et non relâchées et que les voisées varient entre dévoisées et complètement voisées. Dans cette étude, il s'est intéressé aux occlusives non voisées relâchées /p, t, k/ et aux occlusives complètement voisées relâchées /b, d, g/ produites en position finale. L'auteur précise que le bruit de relâchement des occlusives en position finale et celui en position initiale sont quasiment identiques. Il explique que cette similarité du *burst* est attendue puisque la séparation des articulateurs en position finale et le mouvement de l'ouverture en position initiale sont très similaires.

Une série d'expériences a été conçue lors de cette étude de manière à explorer l'importance du bruit de relâchement dans la perception des occlusives réalisées en séquences VC. Dans un premier temps, l'auteur a prononcé des séquences CV et VC et le V en isolation où V = /ε/ ou /ɔ/. Les occlusives en position initiale représentent un point de référence pour celles en position finale. Ces dernières ont été réalisées relâchées et non relâchées. Dans un second temps, ces réalisations ont été manipulées et ensuite écoutées par 50 locuteurs de l'anglais américain. Une première expérience a été conçue pour examiner le bruit de relâchement des /b, d, g/ en position finale. Sept types de stimuli ont été utilisés dans cette expérience. Trois n'ont pas été

manipulés : 1) les syllabes CV, 2) les syllabes VC relâchées et 3) les syllabes VC non relâchées. Les autres conditions ont été manipulées : 4) les syllabes VC avec le bruit de relâchement supprimé, 5) la voyelle en isolation avec le bruit de relâchement ajouté, 6) les séquences avec les bruits de relâchement inter-changés et 7) le bruit de relâchement en isolation. Les résultats montrent qu'avec le bruit de relâchement, les occlusives /b, d, g/ en position finale sont aussi intelligibles que quand elles sont en position initiale (100 % de réponses correctes dans les deux positions). Le taux de bonnes réponses diminue légèrement lorsque les occlusives ne sont pas originellement relâchées ou quand le bruit de relâchement est supprimé (90 % et 95 % respectivement). En échangeant le *burst* des occlusives par un autre, tout en gardant la voyelle et la transition originale, les auditeurs ont tendance à identifier le lieu d'articulation du *burst* et non pas celui des transitions formantiques de la séquence originale. Selon ces constatations, Malécot (1958) déclare que le bruit de relâchement contient plus d'informations sur le lieu d'articulation que les transitions formantiques. Ce raisonnement est soutenu par les résultats des stimuli du *burst* en isolation et ceux ajoutés à la voyelle en isolation où le taux d'identification correcte du lieu d'articulation était généralement élevé (85 % et 79 % respectivement).

Parallèlement à la première expérience, une deuxième a été mise en place dans l'objectif d'étudier l'importance du bruit de relâchement dans la perception des occlusives non voisées /p, t, k/. Les conditions de la première expérience ont été appliquées sur la deuxième. Les occlusives avec un bruit de relâchement en position finale ont été identifiées aussi bien que quand elles sont en position initiale (98 % et 100 % respectivement). Les résultats indiquent également que l'identification du lieu d'articulation des occlusives non voisées dépend plus de leur bruit de relâchement que leurs contreparties voisées. En effet, l'absence naturelle ou artificielle du bruit de relâchement conduit à une baisse plus importante de l'identification des occlusives non voisées par rapport à celle des occlusives voisées (76 % ou 79 %, 90 % ou 95 % respectivement pour les non voisées et pour les voisées). L'auteur explique que ces résultats montrent aussi que les transitions formantiques des occlusives non voisées contiennent moins d'informations sur le lieu d'articulation que les voisées. En effet, dans le cas des non voisées, le taux de l'identification correcte dans la condition des voyelles stables (sans les transitions) avec le bruit de relâchement ajouté est élevé (97 %) tandis que celui du bruit de relâchement en isolation reste relativement élevée (80 %). Des résultats similaires à ceux de la première expérience ont été observés lorsque les bruits de relâchement sont inter-changés.

La troisième expérience a été élaborée pour tester le rôle du bruit de relâchement dans la distinction entre occlusive voisée vs. non voisée. Les mêmes stimuli des deux premières expériences ont été utilisés dans cette dernière. Toutefois, les sujets ont six choix de réponses (/b, d, g, p, t, k/) au lieu de trois comme les expériences précédentes. Les occlusives voisées et non voisées ont un taux d'identification similaire (95 % et 94 % respectivement). La suppression des bruits de relâchement n'a pas influencé la perception des occlusives voisées (98 %) alors qu'une baisse importante de réponses correctes a été remarquée pour les occlusives non voisées (57 %). De plus, les auditeurs perçoivent des occlusives voisées lorsque les séquences des

occlusives non voisées sans le bruit de relâchement sont combinées avec le *burst* de leurs contreparties voisées (prédominance à 92 % du bruit de relâchement). Dans la manipulation inverse, autrement dit, quand les transitions des occlusives voisées sont associées aux *bursts* des non voisées, les auditeurs reconnaissent le lieu d'articulation des bruits de relâchement que dans 26 % des cas (prédominance des transitions formantiques). Ces résultats confirment que la quantité d'informations existantes dans les transitions formantiques des occlusives non voisées est moins importante en comparaison de leurs correspondantes voisées. En outre, ces observations indiquent que le bruit de relâchement des non voisées et des voisées contient des informations sur le voisement puisque leur absence entraîne la confusion de leur identification à différents niveaux. Malécot (1958) résume que « *Voiceless p t k releases and voiced b d g releases contain sufficient cues for conveying both place and manner of articulation of American English plosives in final position* »<sup>25</sup>.

Dans une série d'articles, Stevens et Blumstein ont tenté de décrire et de vérifier expérimentalement l'existence d'un indice acoustique du lieu d'articulation des occlusives dans le bruit de relâchement et dans le début des transitions formantiques (Stevens et Blumstein, 1978; Blumstein et Stevens, 1979, 1980; Stevens et Blumstein, 1981). Ils argumentent que des propriétés acoustiques invariantes du lieu d'articulation peuvent être trouvées dans la forme globale du spectre au début du relâchement de la consonne. Ils soutiennent l'idée que chaque lieu d'articulation a une forme spectrale unique qui est indépendante du contexte vocalique. Stevens et Blumstein (1978) ont alors développé des modèles spécifiques de formes spectrales globales pour chaque lieu d'articulation (bilabial, alvéolaire et vélaire) à partir des analyses acoustiques précédentes. Ces modèles ont été dérivés de simples sections spectrales de 26 ms prélevées au début du relâchement de la consonne dans des syllabes CV (Figure 1.17).

La forme globale du spectre au début du relâchement des labiales est diffuse descendante. Autrement dit, l'intensité baisse lorsque la fréquence augmente. Au contraire, la forme globale du spectre des alvéolaires est diffuse montante. En effet, quand la fréquence croît, l'intensité s'élève aussi. Quant aux vélaires, elles ont une forme globale compacte où le pic d'intensité se trouve dans une zone de fréquence centrale (Figure 1.17). Stevens et Blumstein (1978) ont conçu un protocole expérimental dans le but d'examiner leur hypothèse. Différentes séries de stimuli de la parole synthétique ont été générées incluant le continuum de fréquences consonantiques<sup>26</sup> allant de /b/ à /d/ puis à /g/ ou de /b/ à /g/ avec les voyelles /i, a, u/.

Pour tous les stimuli, la durée de la voyelle a été fixée à 250 ms et le bruit de relâchement de 5 ms a été déterminé entre 5 et 15 ms avant le début du voisement. Trois conditions pour chaque série de stimuli ont été créées. La première correspond aux séquences CV avec le bruit de relâchement et les transitions formantiques ; la deuxième contient les transitions formantiques

25. « Les bruits de relâchement des non voisées /p, t, k/ et ceux des voisées /b, d, g/ contiennent suffisamment d'indices pour fournir le lieu et le mode d'articulation des occlusives en position finale de l'anglais américain », page 380.

26. Les fréquences consonantiques correspondent à celles au début des transitions formantiques de chaque consonne.

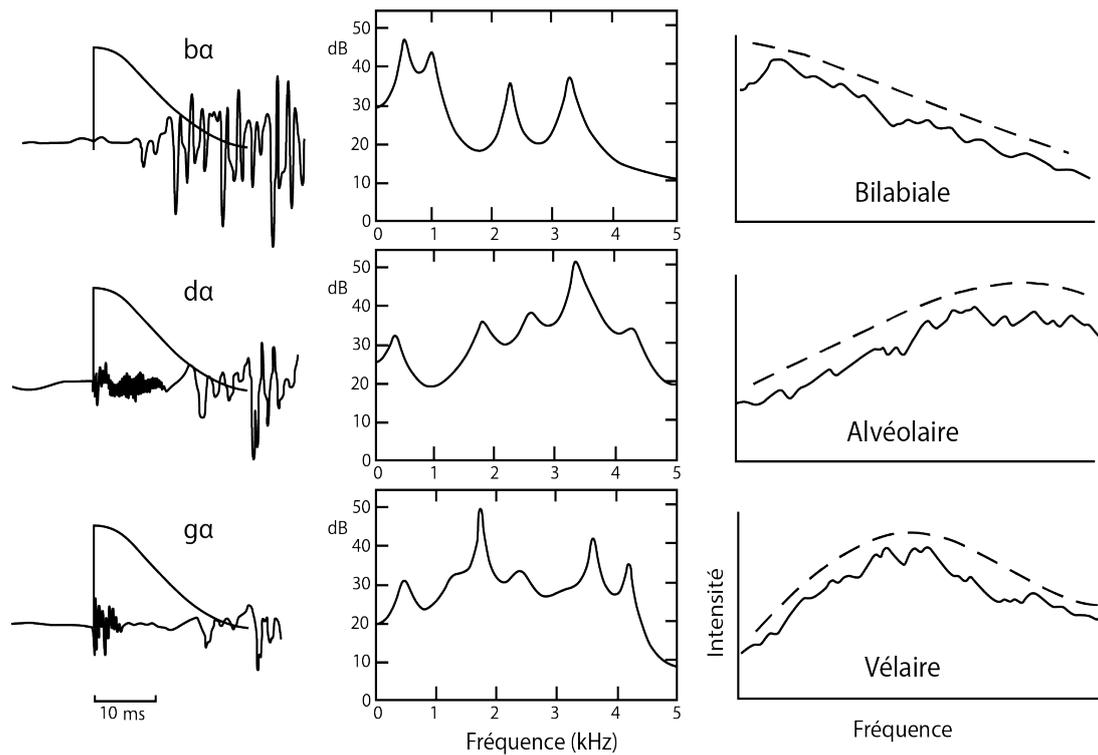


FIGURE 1.17 – Exemples des formes d’onde (à gauche) et leurs spectres (au centre) au moment du relâchement des trois occlusives voisées. Sur chaque forme d’onde, un tracé montre la fenêtre temporelle (de 26 ms de largeur) qui a été utilisée pour l’échantillonnage du spectre (d’après Stevens et Blumstein (1978)). À droite, représentation schématique indiquant la tendance de la forme globale de chaque spectre du lieu d’articulation : diffuse descendante pour les bilabiales, diffuse montante pour les alvéolaires et compacte pour les vélaire (d’après Kent et Read (1996)).

sans le bruit de relâchement alors que la troisième comprend le bruit de relâchement sans les transitions formantiques. Dans la première condition, les pics spectraux du bruit de relâchement se situent toujours aux mêmes fréquences au début des transitions. Les séquences synthétiques réalisées ont été écoutées par 20 locuteurs de l’anglais américain.

Les résultats mettent en avant que les sujets ont perçu les séquences de la première condition mieux que celles de la deuxième (90 % vs. 81 %), tandis que les séquences de la troisième condition n’ont été identifiées qu’à 18 % par les auditeurs. Les auteurs révèlent que ce faible taux est expliqué par le fait que le spectre du bruit de relâchement ne subit pas de changement lisse et continu puisque le stimulus contient deux débuts, le *burst* pour la consonne et l’état d’équilibre pour la voyelle, sans le passage transitionnel. La discontinuité entre le spectre du bruit de relâchement et celui de l’état d’équilibre de la voyelle rend la consonne moins intelligible. La Figure 1.18 illustre les pourcentages de l’identification des séries de stimuli de /b, d, g/ dans les trois contextes vocaliques /i, a, u/ (avec et sans le bruit de relâchement).

Les réponses à chaque stimulus sont généralement plus catégorielles pour les stimuli avec le bruit de relâchement que celles sans. Autrement dit, les pentes aux frontières phonétiques sont plus élevées dans la première condition que dans la deuxième. Le début du spectre, cor-

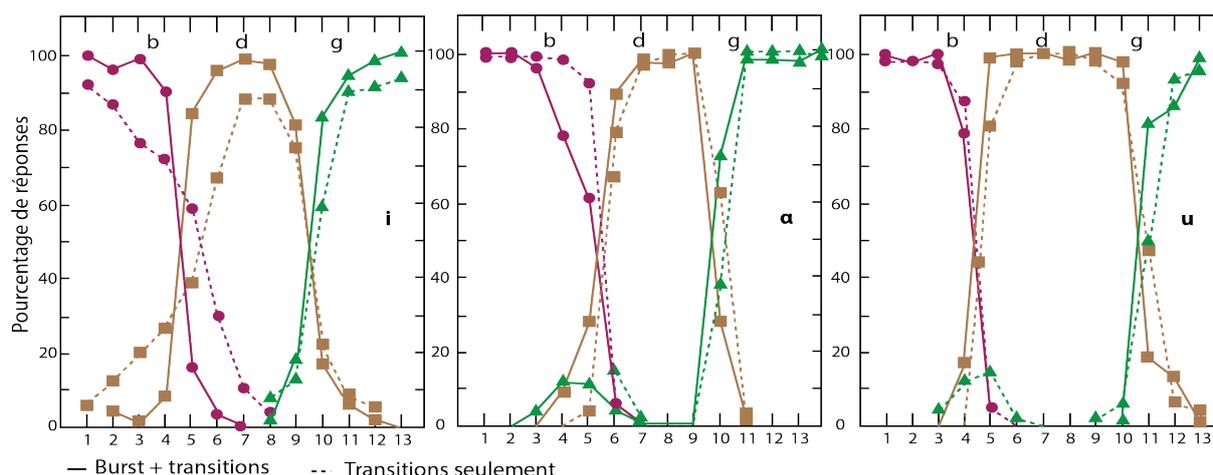


FIGURE 1.18 – Moyennes des pourcentages de l'identification des continuum des stimuli de /b, d, g/ avec les trois voyelles. Les lignes continues représentent les réponses des stimuli avec le bruit de relâchement et les transitions formantiques et les lignes discontinues indiquent les réponses des stimuli avec les transitions sans le *burst* (d'après Stevens et Blumstein (1978)).

respondant aux premières 26 ms après le relâchement, des séquences les mieux perçues par les auditeurs <sup>27</sup> a été analysé afin d'examiner si la forme globale du spectre de chaque lieu d'articulation est distincte. Le début du spectre pour les stimuli identifiés comme un /d/ montre une augmentation générale de l'amplitude des pics spectraux en fonction de la fréquence, contrairement aux stimuli du /b/ qui dévoilent une décroissance. Les spectres du /b/ et du /d/ sont caractérisés par la diffusion de l'énergie, d'où le terme "diffuse", sur une gamme de fréquences centrales (entre 1 et 3 KHz). Quant au spectre du /g/, il expose un pic spectral proéminent dans une gamme de fréquences centrales (entre 1 et 3 KHz), d'où le terme "compact". Selon les auteurs, les résultats de cette analyse confirment l'hypothèse de départ selon laquelle la forme globale du spectre de relâchement est une propriété acoustique invariante du lieu d'articulation indépendamment du contexte vocalique. Stevens et Blumstein (1978) résument que « *Results of this study support this interpretation and further suggest that invariant cues for place of articulation for stop consonants can be characterized independent of the following vowel by an event, i.e., the spectrum at stimulus onset occurring at a point in time, rather than by a sequence of events over time. In this view, then, the formant transitions are not the primary cues signaling place of articulation. Instead, their primary function, [...] seems to be to join the onset spectrum to the vowel smoothly without introducing any additional discontinuities* »<sup>28</sup>.

Blumstein et Stevens (1979) fournissent une mesure quantitative plus fiable de l'importance des informations contenues dans le bruit de relâchement et dans le début des transitions forman-

27. Celles dont l'identification est proche de 100 %

28. « Les résultats de cette étude soutiennent cette interprétation et suggèrent que les indices invariants du lieu d'articulation des occlusives peuvent être caractérisés par un événement indépendamment de la voyelle suivante. Autrement dit, le spectre au début du stimulus se produit à un point donné dans le temps et non lors d'une série d'événements dans le temps. D'après ce point de vue, les transitions formantiques ne sont donc pas des indices principaux permettant de désigner le lieu d'articulation. Leur fonction principale [...] est de lier le début du spectre (de la consonne) à celui de la voyelle de façon lisse et sans produire de discontinuités », page 1367.

tiques. Ils utilisent des séquences CV et VC produites par six locuteurs de manière à vérifier si la forme spectrale globale au début du relâchement peut désigner le lieu d'articulation en parole naturelle indépendamment du voisement (occlusives voisées et non voisées), de la position syllabique (initiale et finale) et de la voyelle adjacente (où C = /b, d, g, p, t, k/ et /i, a, u, e, o/). Les auteurs ont mis en place des modèles caractérisant les propriétés au début du relâchement observées dans l'étude précédente : diffuse descendante ou plate (bilabiales), diffuse montante (alvéolaires) et compacte (vélares). Les séquences ont été classifiées selon la forme globale du spectre des 26 premières ms après le relâchement de la consonne par rapport aux trois modèles<sup>29</sup>. Les résultats montrent que les modèles ont catégorisé le lieu d'articulation des consonnes (voisées et non voisées) en position initiale dans 85 % des cas. Toutefois, le taux de classification des occlusives en position finale est moins important (76 %) au moment du bruit de relâchement. Les analyses statistiques révèlent que le contexte vocalique a un impact significatif sur la classification pour les bilabiales et les vélares mais il est non significatif pour les alvéolaires. Blumstein et Stevens (1979) expliquent que cet impact du contexte vocalique est provoqué par la coarticulation du corps de la langue dans l'anticipation de la voyelle suivante. Les auteurs déclarent que cette influence est limitée puisque tous les modèles ont catégorisés correctement toutes les consonnes avec un taux supérieur à celui du hasard.

Dans une troisième étude, Blumstein et Stevens (1980) effectuent des tests de perception sur des syllabes CV synthétiques afin de déterminer si la partie initiale d'une syllabe peut fournir des indices sur le lieu d'articulation des consonnes occlusives indépendamment du reste de la syllabe. Les auteurs ont généré des séquences CV, où C = /b, d, g/ et V = /i, a, u/, contenant le bruit de relâchement (d'une durée de 5 à 10 ms) et le début des transitions formantiques (transitions en mouvement). Ces transitions ont été produites dans quatre durées différentes (10, 20, 29 et 49 ms). Le bruit de relâchement et le début du voisement sont séparés par un intervalle compris entre 5 et 15 ms en fonction de l'occlusive. Un autre ensemble de stimuli a été créé sans les transitions (transitions plates) de façon à examiner l'importance de celles-ci et du bruit de relâchement (Figure 1.19).

La forme globale du spectre du début des syllabes correspond à celle des modèles explorés dans les études précédentes (Stevens et Blumstein, 1978; Blumstein et Stevens, 1979). Les 36 stimuli (3 voyelles × 3 consonnes × 4 durées) ont ensuite été écoutés par 34 locuteurs de l'anglais américain. Les résultats soulignent que les portions courtes du stimuli fournissent des informations suffisantes pour percevoir le lieu d'articulation dans les différents contextes vocaliques. En effet, le taux d'identification de l'occlusive est supérieur à 80 % en moyenne avec 10 ms de transitions en mouvement. Quand les stimuli contiennent des transitions formantiques en mouvement, le taux de classification est généralement amélioré pour toutes les durées. La reconnaissance de la bilabiale et de l'alvéolaire ne semble pas être influencée par la durée des

29. Il est à remarquer que, pour les voisées /b, d, g/, les 26 ms suivant le relâchement de l'occlusive peuvent contenir le bruit de relâchement et le début des transitions formantiques alors que, pour les non voisées, seul le bruit de relâchement est inclus puisque le début du voisement commence généralement après les premières 26 ms.

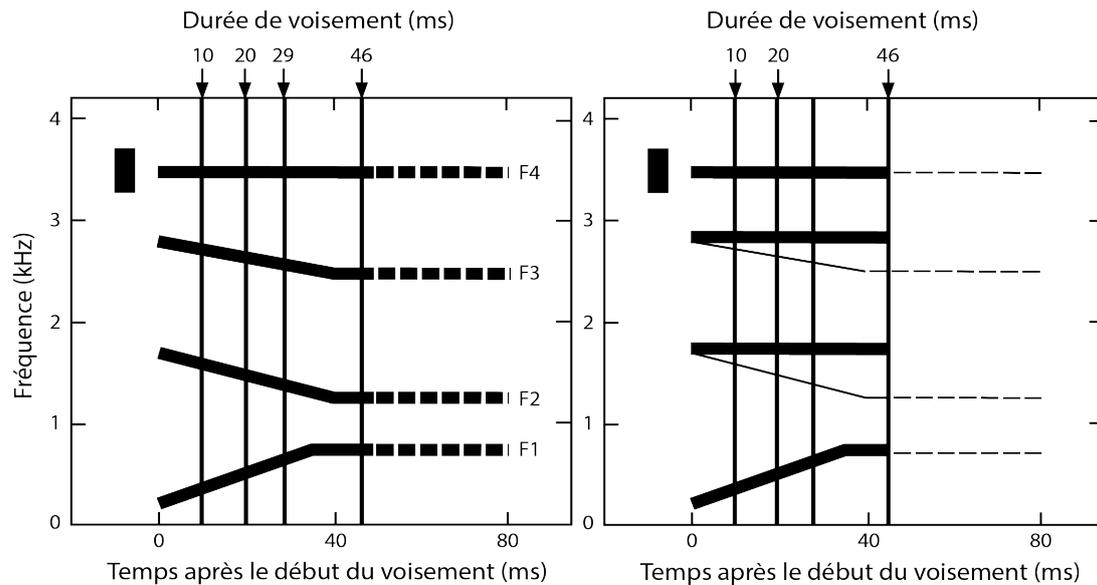


FIGURE 1.19 – Représentation schématique d'un stimulus synthétique montrant le bruit de relâchement et les quatre premiers formants avec (à gauche) et sans (à droite) les transitions formantiques. Les transitions formantiques du premier formant restent toujours stables pour indiquer le mode d'articulation (occlusive). Les lignes verticales correspondent aux durées des quatre stimuli (10, 20, 29 et 46 ms). Les lignes horizontales discontinues désignent la suite de la syllabe si elle a été réalisée en entier (d'après Blumstein et Stevens (1980)).

stimuli tandis que celle de la vélaire, notamment avec la voyelle /i/, est impactée par ce facteur. Le taux de reconnaissance passe de 48 % avec la durée de 10 ms à 90 % avec celle de 46 ms. Les auteurs expliquent qu'une durée relativement longue (supérieure à 20 ms) est nécessaire aux auditeurs pour construire une représentation du début du spectre "compact" dans leur système auditif. Par conséquent, des stimuli de /g/ avec une durée courte sont souvent interprétés comme étant diffus (occlusive bilabiale ou alvéolaire).

Dans une autre expérience, les mêmes stimuli ont été générés mais cette fois-ci sans le bruit de relâchement dans l'objectif d'examiner l'importance de ce dernier et celle des transitions formantiques dans la perception des occlusives. Les résultats dévoilent que les portions courtes au début des stimuli CV sans le bruit de relâchement fournissent également des informations suffisantes pour reconnaître le lieu d'articulation dans la plupart des stimuli. Cependant, les taux de classification des séquences /bi/ avec des transitions plates et /gi/ avec des transitions en mouvement sont inférieurs au hasard (33 %). Similairement à la première expérience, les transitions en mouvement améliorent en grande majorité l'identification du lieu d'articulation. Toutefois, tous les stimuli sans les transitions ont été catégorisés avec un taux supérieur à celui du hasard (à l'exception de /bi/). Selon Blumstein et Stevens (1980), ce résultat laisse suggérer que les sujets ne se basent pas sur les directions des mouvements des transitions (indices secondaires selon les auteurs) mais, lorsque ces indices sont disponibles, ils peuvent être utilisés pour prendre la décision phonétique adéquate. Les effets de la durée du stimulus (sans le bruit de relâchement) sur les réponses des auditeurs dans cette expérience sont similaires à ceux ob-

servés dans celle précédente (avec le bruit de relâchement). Une comparaison entre les résultats de la première et de la deuxième expérience mettent en exergue que la présence du bruit de relâchement n'influence pas la reconnaissance de la bilabiale. Quant à l'alvéolaire et à la vélaire, la performance de l'identification est améliorée grâce à la présence du bruit de relâchement. D'après les auteurs, le taux convenable de reconnaissance de la consonne qui persiste quand le bruit de relâchement est omis soutient l'hypothèse que celui-ci et le début du voisement contribuent tous les deux à la constitution d'une propriété acoustique plus large. Cette propriété est l'indice du lieu d'articulation des occlusives voisées qui correspond à la forme globale du début de relâchement. Blumstein et Stevens (1980) indiquent que les résultats globaux appuient leur hypothèse de départ et résumant que « *These observations are consistent with the view that the gross properties of the spectrum sampled over the initial 10-20 ms of a stop consonant provide the principal cues to place of articulation for stop consonants. In this view, the transitions of the formants from the release of the consonant to the vowel provide the acoustic material that links the transient events at the onset to the slowly varying spectral characteristics of the vowel. These transitions ensure that no further abrupt discontinuities in the spectrum occur following the initial transient at the release* »<sup>30</sup>.

Les études de Stevens et Blumstein citées plus haut ont dévoilé l'importance du bruit de relâchement et de la forme globale de son spectre dans la perception du lieu d'articulation. Ohde et Stevens (1983) ont analysé les détails contribuant à la perception du lieu d'articulation présents dans les informations cohérentes que le *burst* peut contenir. L'objectif de leur recherche est d'expliquer la nature des propriétés du bruit de relâchement que les auditeurs utilisent pour identifier le lieu d'articulation des bilabiales et des alvéolaires. Des formes variées du spectre ont été générées au moment du relâchement des consonnes. En manipulant l'amplitude du *burst*, la manière dont celle-ci change dans les différentes gammes de fréquence est modifiée. Ohde et Stevens (1983) conçoivent deux expériences de perception afin d'étudier les réponses des auditeurs à ces stimuli pour déchiffrer quelles sont les propriétés acoustiques au moment du relâchement de la consonne qui sont les plus pertinentes pour la perception des deux catégories consonantiques examinées. La première expérience analyse les occlusives non voisées /p/ et /t/ produisant une série de neuf étapes dans un continuum allant des caractéristiques acoustiques adéquates pour une bilabiale à celles correspondantes à une alvéolaire dans le contexte du /a/ (Figure 1.20). Pour chaque élément de cette série, cinq stimuli avec des amplitudes différentes pour le bruit de relâchement (-10, -5, 0, +5, +10 dB relatives à une amplitude standard prédéfinie) ont été générés pour un total de 45 stimuli (9 étapes × 5 amplitudes). Les stimuli ont été écoutés par 11 locuteurs de l'anglais américain qui ont répondu sur une échelle de certitude de 1 à 6 où 1 = /p/ et 6 = /t/.

---

30. « Ces observations sont en accord avec l'hypothèse que les grandes propriétés du spectre analysées dans les premières 10 à 20 ms d'une occlusive fournissent les indices principaux du lieu d'articulation. Selon ce point de vue, les transitions formantiques du relâchement de la consonne à la voyelle présentent un lien acoustique entre les deux événements. Ces transitions assurent qu'il n'existerait plus de discontinuité brusque dans le spectre après le bruit de relâchement », page 660.

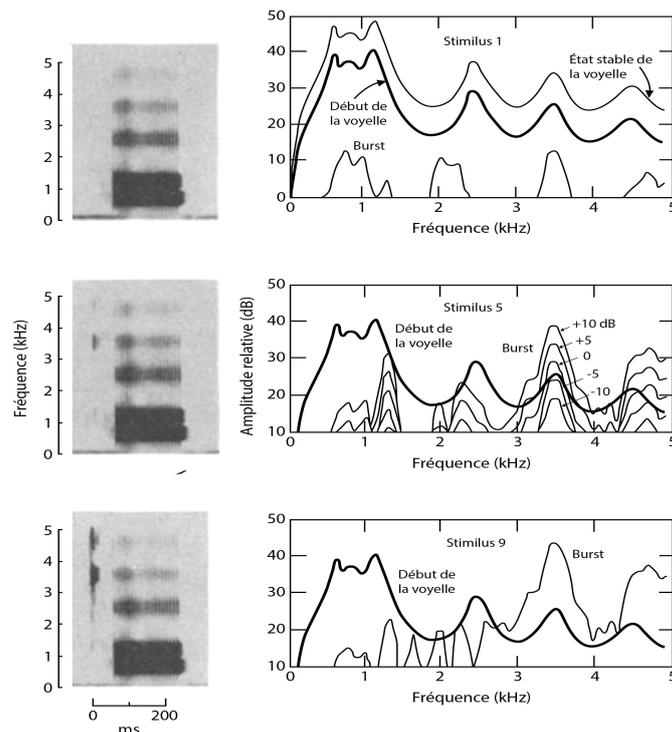


FIGURE 1.20 – À gauche, les spectrogrammes représentant les stimuli des étapes au début et à la fin de la série avec une étape intermédiaire dans le continuum /pa/-/ta/. À droite, les spectres de la voyelle analysés à 7 ms après le début du voisement et les spectres du *burst* analysés à 4 ms après le début du relâchement. Pour les étapes 1 et 9, seul le spectre avec une amplitude de 0 dB est visible alors que, pour l'étape 5, toutes les amplitudes du *burst* sont représentées (d'après Ohde et Stevens (1983)).

Les résultats indiquent que les stimuli avec des amplitudes relatives basses du bruit de relâchement ont tendance à être reconnus comme un /p/ plus que comme un /t/. Le changement de l'amplitude relative du *burst* affecte significativement les réponses des auditeurs dans les 9 étapes. L'effet de cette variation est plus marqué dans les stimuli qui sont spectralement ambigus, notamment les étapes 4, 5 et 6. Les auteurs ont également observé que, lorsque l'amplitude du spectre du bruit de relâchement située dans les zones de F4-F5 (hautes fréquences) est inférieure à celle de la voyelle, les réponses ont tendance à être considérées comme un /p/ alors que, quand l'amplitude du spectre est plus grande dans ces zones, les réponses sont plutôt interprétées comme un /t/. Ils ajoutent que l'identification de la consonne n'est pas déterminée uniquement par l'amplitude relative du spectre dans les fréquences élevées mais aussi par la forme globale du spectre dans les fréquences basses (dans la zone de F2-F3).

Dans la seconde expérience, l'effet de l'amplitude a été examiné pour les occlusives voisées /b, d/. Les procédures employées dans le cadre de la première expérience ont été suivies pour générer les stimuli de /ba, da/. Les résultats de cette expérience sont en accord avec ceux constatés dans celle précédente où la variation de l'amplitude du spectre influence la perception des deux occlusives. Toutefois, cet impact est moins important que celui pour les non voisées.

Selon les auteurs, les observations des deux expériences laissent suggérer qu'il existe deux

propriétés acoustiques qui sont engagées dans la distinction entre les alvéolaires et les labiales en anglais. La première est déterminée par les changements de l'amplitude dans les fréquences hautes (+ d'amplitude = alvéolaire, - d'amplitude = bilabiale). La seconde propriété est désignée par les fréquences des pics du spectre associés à F2 et à F3 suivant le relâchement de la consonne. Ces dernières ont des amplitudes basses pour les non voisées et élevées pour les voisées.

Les études dans cette section ont mises en avant l'importance du bruit de relâchement dans la perception des occlusives. Il contient les informations nécessaires, et quelquefois suffisantes, pour distinguer les différents lieux d'articulation des occlusives. Le bruit de relâchement est associé aux transitions formantiques qui forment le passage entre la consonne et la voyelle. Ces dernières sont également un indice indispensable dans la perception des occlusives. Elles seront présentées dans la section suivante.

#### 1.2.4.2 Les transitions formantiques des occlusives

Comme il a été défini dans la section 1.1.4, les transitions formantiques sont des changements rapides de fréquences qui reflètent le passage articulaire d'une consonne à la voyelle suivante (ou d'une voyelle à la consonne suivante). Elles portent ainsi des informations sur la voyelle (*cf.* section 1.2.1) et sur la consonne (*cf.* section 1.1.4 et 1.2.4.1). L'objectif de cette section est de mettre en exergue quelques recherches qui ont exploré l'importance de cet indice dans la perception des occlusives.

L'observation des liens entre les transitions formantiques et la perception du lieu d'articulation des occlusives a été établie très tôt après l'invention du spectrographe. Dans leur recherche innovatrice "*Visible speech*", Potter et al. (1947) ont noté que les transitions formantiques entre une occlusive et une voyelle, particulièrement celles du deuxième formant, semblent caractériser le lieu d'articulation de ces consonnes. Selon ces auteurs, lorsque l'occlusive est suivie d'une voyelle, le mouvement du deuxième formant est toujours montant pour /b/ ou /p/. Le début de la voyelle est ainsi inférieure à son état stable. Ils ont également remarqué que les transitions peuvent être montantes ou descendantes pour /d/ et /t/ en fonction de l'état stable de la voyelle suivante puisque le point de départ de ces occlusives est situé dans une zone centrale de fréquences. Quant aux /g/ et /k/, le deuxième formant commence souvent à partir d'une position légèrement au-dessus de son état stable ; par conséquent, les transitions sont toujours descendantes. La constatation des changements systématiques des transitions formantiques par rapport au contexte a mené à des expériences de perception de façon à examiner leur pertinence dans la distinction du lieu d'articulation.

Une des premières études expérimentales traitant de l'importance perceptuelle des transitions formantiques est celle de Cooper et al. (1952) et Liberman et al. (1954)<sup>31</sup> avec le *pattern playback machine*. En observant le spectrogramme des différentes occlusives, les auteurs ont

---

31. Les tests ont été présentés et discutés dans les deux articles.

noté que les transitions du premier formant sont toujours montantes pour ce mode d'articulation et la longueur des transitions signale l'opposition du voisement (+ de transition = voisée, – de transition = non voisée). Les transitions du deuxième formant, quant à elles, semblent changer en fonction du lieu d'articulation. Toutefois, elles sont également influencées par le contexte vocalique. Cooper et al. (1951) mettent en place un ensemble de tests de perception pour examiner le sens perceptuel des différentes transitions formantiques du deuxième formant. Les deux premiers formants d'une voyelle (en état d'équilibre) ont d'abord été dessinés et ont ensuite été associés à une variété de transitions (Figure 1.21). Dans un premier test, une série

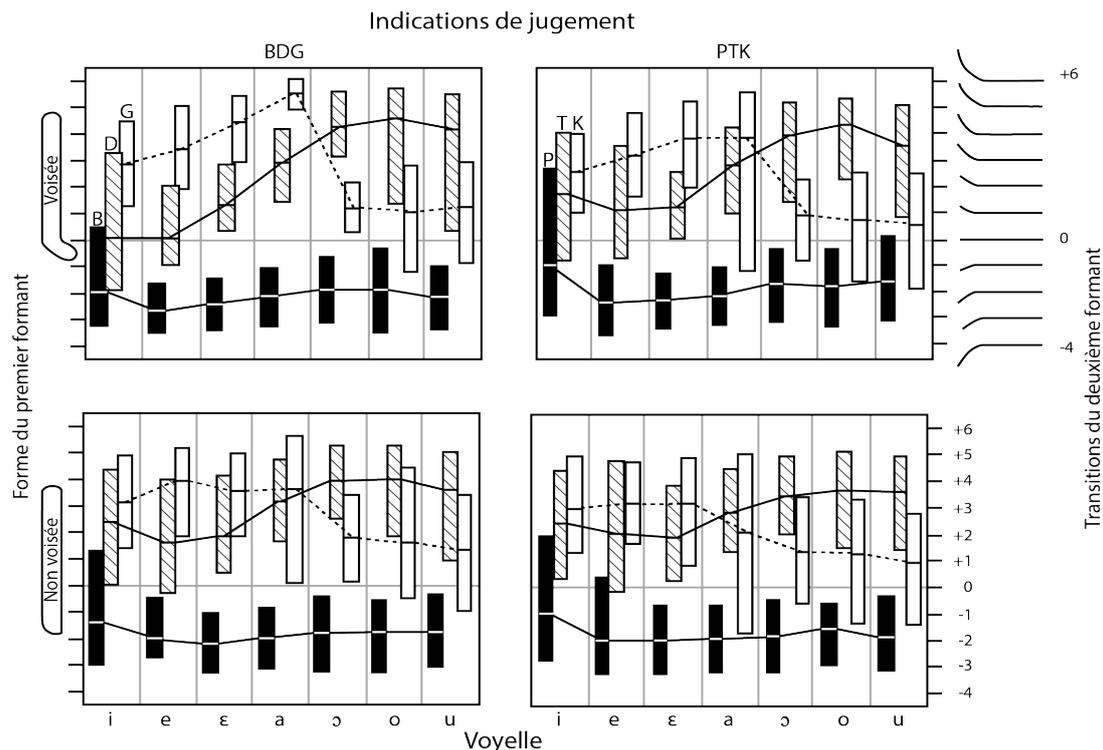


FIGURE 1.21 – Représentation des réponses des auditeurs montrant la contribution des transitions formantiques dans la perception des occlusives. Le schéma en haut à gauche affiche les réponses des 33 sujets sur les trois choix possibles /b, d, g/. Les stimuli ont un premier formant montant (à gauche) combiné à 11 transitions différentes du second formant (à droite) avec 7 contextes vocaliques (en abscisse). Les lignes connectées désignent la médiane des réponses pour chaque voyelle. Le schéma en haut à droite montre les réponses d'un autre groupe de 33 participants où les choix étaient limités à /p, t, k/. Les schémas en bas indiquent les réponses pour les mêmes stimuli mais avec une transition très courte et montante de F1 (d'après Cooper et al. (1952)).

de 11 transitions formantiques des F2 avec une transition montante aiguë du F1 a été créée dans sept contextes vocaliques. Le deuxième test contient les mêmes stimuli mais avec une transition montante très courte du premier formant. Les stimuli ont été écoutés par deux groupes de participants de 33 personnes. Le premier groupe avait le choix entre /b, d, g/ et le second entre /p, t, k/.

Les résultats affichés dans la Figure 1.21 (désignés par des barres verticales) dévoilent que

la perception des occlusives ne dépend pas seulement de la direction ou de l'étendue des transitions du second formant mais également du contexte vocalique. La longueur de la barre donne une indication sur l'étendue des différentes transitions incluses dans une catégorie consonantique. Elle signale également le degré du chevauchement ou de confusion entre ces sons. Selon les auteurs, l'interprétation globale de ces réponses est que le son perçu est une bilabiale (/b/ ou /p/) lorsque les transitions du second formant sont montantes et une alvéolaire ou une vélaire (/d/, /t/, /g/ ou /k/) quand elles sont descendantes. Cooper et al. (1952) expliquent que le chevauchement important entre les sons révèle que la direction des transitions ne suffisent pas comme indice pour faire la distinction entre les trois catégories.

Liberman et al. (1954) ont ajouté les transitions du troisième formant sur les stimuli des tests précédents dans le but d'améliorer la perception des trois catégories. Trois directions de transition ont été utilisées, montantes, descendantes ou plates. Lorsque le F3 avec une transition montante est ajouté aux stimuli où les transitions de F2 sont également montantes, une bilabiale est perçue par les auditeurs dans tous les contextes vocaliques. Quand cette transition montante du F3 est combinée aux transitions descendantes du F2, une vélaire est entendue. L'ajout du F3 avec une transition plate (transition 0) ne change pas le son produit et aucune amélioration n'est remarquée comparée aux stimuli constitués de deux formants. Enfin, une alvéolaire est perçue lorsque le F3 avec une transition descendante est ajoutée aux stimuli ambigus (dans la zone de chevauchement dans la Figure 1.21). Ces résultats ont été confirmés par une étude de perception similaire dirigée par Sfford et al. (1958). D'après ces derniers, les transitions formantiques du troisième formant peuvent jouer un rôle dans la perception du lieu d'articulation des occlusives. Liberman et al. (1954) précisent que, quand les transitions du F2 échouent à fournir l'information nécessaire à l'identification du lieu d'articulation, l'ajout du bruit de relâchement ou de la transition de F3 résout cette ambiguïté.

Les études menées par les chercheurs des laboratoires Haskins à l'aide du *playback machine* ont démontré que les variations dans les transitions du F2 entre une consonne et une voyelle peuvent être suffisantes pour produire des stimuli identifiés comme des occlusives distinctes. D'après les expériences présentées plus haut, il est clair qu'une occlusive donnée est associée à une variété de transitions en fonction de la voyelle adjacente. Il reste à expliquer comment les transitions du F2 sont liées à l'identification du lieu d'articulation des occlusives. Comme montré précédemment, les directions des transitions ne sont pas des indices fiables pour déterminer l'identité de la consonne. Toutefois, l'examen de ces directions a amené ces chercheurs à formuler la théorie du *locus* (Delattre et al., 1955). D'après cette théorie, chaque lieu d'articulation a un point virtuel de fréquence vers lequel les transitions du F2 se dirigent (*cf.* section 2.3.1). Par exemple, le locus de /d/ est de 1800 Hz impliquant des transitions du F2 montantes lorsqu'il est produit avec les voyelles antérieures (où F2 est supérieur à 1800 Hz) et descendantes quand il est réalisé avec les voyelles postérieures (où F2 est inférieur à 1800 Hz) (Figure 1.22). Selon Delattre et al. (1955), le locus représente le point de départ des différentes transitions du F2 d'une seule consonne qui est invisible mais prédictible grâce leur directions.

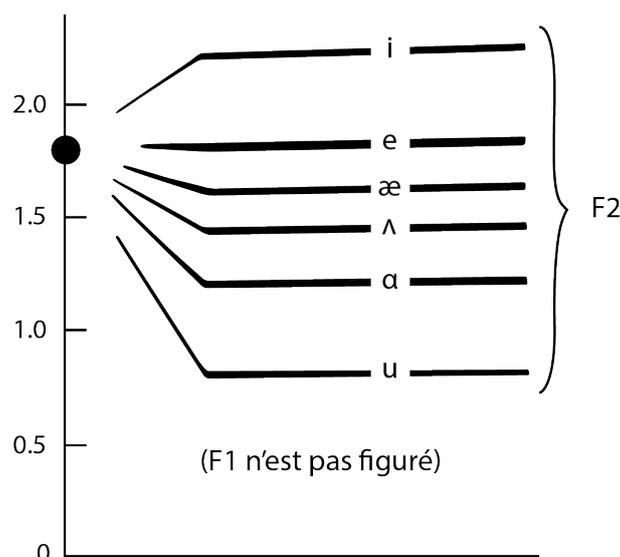


FIGURE 1.22 – Représentation schématique des transitions de F2 de l'occlusive /d/ réalisée avec plusieurs voyelles qui se dirigent vers son locus virtuel (cercle noir) situé à 1800 Hz (d'après Kent et Read (1996)).

Ils déclarent alors que le locus est une propriété invariante indiquant le lieu d'articulation des occlusives. D'après cette idée, les transitions formantiques (du F2 et quelquefois du F3) sont suffisantes pour marquer le lieu d'articulation des occlusives.

Dans une étude spectrographique sur la parole naturelle, Halle et al. (1957) soulignent que les régularités observées dans les transitions formantiques de la parole naturelle sont plus complexes que ce qui a été proposé dans la théorie de locus. Les auteurs ont remarqué que les transitions formantiques ne dépendent pas seulement de l'état d'équilibre de la voyelle suivante mais également de la position syllabique (initiale ou finale) et de la propriété tendue-relâchée de la consonne et de la voyelle (*cf.* section 2.3.3 et section 1.3 pour d'autres critiques). Halle et al. (1957) ont alors conçu un protocole expérimental afin d'examiner l'importance des transitions formantiques dans la perception des occlusives en parole naturelle. Dans un premier temps, trois locuteurs de l'anglais américain ont réalisé des syllabes fermées /VC/ et ouvertes /V/ où C = /b, d, g, p, t, k/ et V = /i, a, u, ɪ, ʌ/. Dans un second temps, le bruit de relâchement des consonnes finales a été supprimé. Les syllabes fermées (sans le *burst*) et ouvertes ont été écoutées par huit auditeurs. Ces derniers devaient identifier la consonne à la fin de chaque séquence et choisir une des six occlusives ou la mention "rien"<sup>32</sup> lorsqu'ils n'entendaient pas de consonne finale.

Les résultats montrent que les sujets n'ont jamais confondu une syllabe ouverte se terminant par une occlusive non voisée avec une syllabe fermée tandis qu'elle a parfois été reconnue comme une syllabe fermée lorsque l'occlusive finale est voisée. D'un autre côté, les syllabes fermées, notamment celles avec des occlusives voisées, ont souvent été identifiées comme ou-

32. Le choix "rien" a été éliminé dans le cas des syllabes fermées produites avec des voyelles relâchées puisque cette association est impossible en anglais.

vertes. Les auteurs révèlent que les "erreurs" d'identification ont été corrélées avec la position du premier formant de la voyelle précédente. Les transitions montantes du F1 jouent un rôle principal dans la reconnaissance de la classe de l'occlusive. Autrement dit, ces dernières sont souvent jugées comme fermées lorsqu'elles sont présentes dans une syllabe. La présence des transitions du F1 est renforcée lorsqu'il est en position haute. Cette position permet des transitions plus prononcées que celles basses. Par conséquent, quand le F1 est élevé, les syllabes fermées ont rarement été entendues comme ouvertes. Les résultats désignent également que les occlusives réalisées avec les voyelles relâchées /ɪ, ʌ/ ont été mieux identifiées que dans les autres contextes vocaliques. D'après Halle et al. (1957), cette constatation est due aux transitions marquées des occlusives dans les contextes de ces voyelles<sup>33</sup>. Les syllabes fermées avec la voyelle /u/ ont souvent été perçues comme ouvertes car les transitions du F2 dans ce contexte sont plates (transitions 0). Le taux global de l'identification correcte avec les voyelles /i, a/ est relativement bas.

Les études dans les deux sections précédentes ont souligné que le bruit de relâchement et les transitions formantiques contiennent des informations importantes dans la perception du lieu d'articulation des occlusives. L'absence d'un de ces indices dans une séquence peut provoquer la confusion entre les différents lieux d'articulation des occlusives. Ces dernières peuvent également être séparées en fonction du voisement. Comme il a été avancé plus haut, l'opposition voisée vs. non voisée est liée aux transitions du F1. Liberman et al. (1958) ont observé que lorsque la transition du F1 est allongée pour commencer plus tôt que celles du F2 et du F3, l'occlusive est toujours perçue comme une voisée (transitions du F1 + longues = voisée). D'autres indices tel que le VOT et la force d'articulation participent également à la distinction entre occlusive voisée vs. non voisée.

### 1.2.4.3 Le VOT

La vibration des plis vocaux fournit un signal quasi-périodique appelé "voisement". Ce dernier conduit à une excitation harmonique d'une bande de fréquences basses durant l'intervalle de la clôture orale et continuant après le relâchement de l'occlusive jusqu'au son suivant. Ce phénomène est visible sur le spectrogramme par une barre horizontale dans les fréquences basses. Lors de la séparation des articulateurs responsables de la fermeture de la consonne, le flux d'air sous pression bloqué dans le conduit vocal est relâché brusquement donnant lieu au bruit de relâchement (*cf.* section 1.2.4.1). La durée séparant la barre du bruit de relâchement du début des vibrations des plis vocaux est appelé le VOT, de l'anglais "*Voice Onset Time*" (Figure 1.15). Celui-ci est considéré comme l'indice principal qui distingue les deux catégories phonémiques : voisée et non voisée (Lisker et Abramson, 1964, 1967).

Le mécanisme de production du voisement permet une meilleure compréhension de l'opposition voisée vs. non voisée. La prononciation d'un son exige le développement de la pression

33. Les auteurs exposent que les transitions formantiques des occlusives sur un spectrogramme sont plus faciles à distinguer dans les contextes des /ɪ, ʌ/ que dans les autres contextes.

sous-glottale et un mouvement de l'air subséquent à travers la glotte vers les cavités supra-glottales. Lorsque la pression de l'air sous une glotte fermée est suffisamment élevée, les plis vocaux sont séparés et un souffle d'air est relâché dans les cavités supra-laryngales. Cette opération provoque une chute de la pression de l'air sous-glottale et une tension musculaire du larynx entraînant la refermeture rapide des plis vocaux. Ce processus est répété jusqu'à ce que la pression de l'air sous-glottale soit dissipée. Le débit de la répétition laryngale ( $f_0$ ) dépend de la longueur, de la masse et de la tension des plis vocaux en fonction de la quantité de la pression de l'air pulmonaire. Les souffles, relâchés de façon répétitive par la glotte, excitent les fréquences de résonances du conduit vocal produisant un son voisé. Si la glotte est ouverte comme elle l'est au repos, il existera alors peu d'obstacles devant le flux d'air pulmonaire et un son non voisé sera réalisé.

Lisker et Abramson (1964) exposent que le concept du VOT ne présente aucune difficulté physiologique. Il décrit tout simplement à quel moment le voisement commence par rapport au moment du bruit de relâchement. Pendant le VOT, il est possible d'entendre une aspiration si le conduit vocal résonne aux turbulences de l'air qui passent par la glotte. En effet, lors de la mesure du VOT, le moment du relâchement est considéré comme le point de référence et a une valeur de zéro. Trois valeurs de VOT sont possibles en fonction de la barre de l'occlusion et du début du voisement :

1. Une valeur négative est donnée au VOT lorsque la vibration des plis vocaux débute avant la barre de l'occlusion.
2. Une valeur positive du VOT est notée quand la vibration démarre après la barre.
3. La valeur du VOT est de zéro si la vibration et le moment du relâchement commencent simultanément (Lisker et Abramson, 1964).

Lisker et Abramson (1964) calculent le VOT des occlusives en position initiale produites dans des mots en isolation ou en parole continue dans 11 langues. Les auteurs observent que l'opposition voisée *vs.* non voisée (mais également aspirée *vs.* non aspirée) est différenciée principalement par le VOT dans toutes les langues. En effet, les VOT des occlusives non voisées ont des valeurs généralement positives tandis que ceux des occlusives voisées ont des valeurs globalement négatives ou égales à zéro (Figure 1.23). Quant aux occlusives aspirées, elles ont des VOT plus grands que leurs contreparties non aspirées.

Il est à remarquer que les VOT varient d'une langue à une autre et varient selon le débit de parole. Lisker et Abramson (1964, 1967) ont constaté que le VOT en parole continue est plus réduit que celui des mots en isolation. De plus, le VOT diffère en fonction du lieu d'articulation (Table 1.4). Cho et Ladefoged (1999) expliquent que le VOT est influencé par trois facteurs principaux liés au lieu d'articulation :

1. La position du point de fermeture : lorsqu'elle est plus postérieure, la durée du VOT est plus longue.
2. La taille de la zone du contact lingual : quand elle est grande, la durée du VOT est longue.

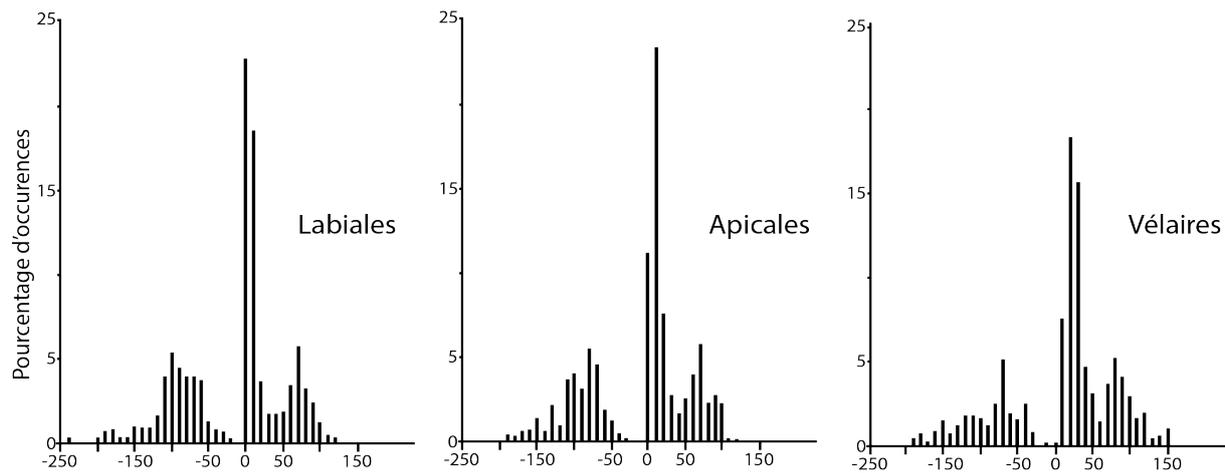


FIGURE 1.23 – Représentation des distributions de la fréquence globale des valeurs du VOT dans toutes les langues pour les trois catégories d’occlusives (d’après à Lisker et Abramson (1964)).

	Néerlandais	Espagnol portoricain	Hongrois	Cantonais	Arménien oriental	Coréen	Tamil
/p/	10	4	2	9	3	18	12
/t/	15	9	16	14	15	25	8
/k/	25	29	29	34	30	47	24

TABLE 1.4 – Moyennes des VOT pour 7 langues (parmi 11) issus des données de Lisker et Abramson (1964) montrant la variation entre les langues et entre les lieux d’articulation (dans Cho et Ladefoged (1999)).

### 3. La vitesse du mouvement articuloire : une vitesse élevée conduit à une durée de VOT courte.

Le premier facteur expose que la variation de VOT peut être expliquée par le changement de la taille de la cavité supra-glottale derrière le point de fermeture. Dans le cas d’une occlusive vélaire, le volume de la cavité derrière le point d’articulation est plus petit que celui d’une alvéolaire ou d’une bilabiale. Simultanément, la cavité devant le point d’articulation d’une occlusive vélaire a un volume plus grand que celle d’une alvéolaire ou d’une bilabiale. Pour produire une occlusive (entre autres), il est nécessaire de créer une variation de la pression du flux d’air dans le conduit vocal. Dans le cas des occlusives vélares, le flux d’air est compressé dans un volume plus petit<sup>34</sup> générant une pression plus importante. Après le relâchement, cette pression nécessite plus de temps pour créer une pression trans-glottale adéquate afin que les plis vocaux se remettent à vibrer. Concernant les alvéolaires et les bilabiales, le volume est plus grand et, par conséquent, la pression est moins importante. De plus, le point d’articulation des bilabiales est le plus antérieur provoquant un VOT relativement inférieur aux deux autres catégories. Les alvéolaires, quant à elles, ont des VOT intermédiaires entre les bilabiales et les vélares car leur

34. Cela comprend l’air des poumons jusqu’au point de fermeture.

point de fermeture est entre les deux autres (Cho et Ladefoged, 1999).

Selon le deuxième facteur, le VOT peut varier en fonction de l'étendue de la zone du contact de la langue. Les occlusives vélares sont produites avec le dos de la langue contre le palais mou. Cette zone est relativement plus large que celle des alvéolaires ce qui conduit à un relâchement plus lent. Par conséquent, les articulateurs prennent plus de temps pour se séparer avant qu'une pression trans-glottale appropriée se produise.

La vitesse du mouvement des articulateurs joue également un rôle important dans la variation de la durée du VOT. En effet, la lèvre inférieure se meut plus rapidement que le corps de la langue tandis que le dos de la langue est plus lent que l'apex. Ces variations de vitesse peuvent être expliquées par les différences de masse entre les articulateurs engagés<sup>35</sup>. Une vitesse élevée du mouvement articuloire entraîne une diminution rapide de la pression derrière le point de fermeture. Un temps plus court est ainsi nécessaire avant qu'une pression trans-glottale convenable soit accumulée.

En résumé, il a été montré dans cette section que les variations du moment du début du voisement mesurées par le VOT fournissent des informations qui aident à la distinction entre les deux catégories voisée et non voisée. De plus, ces variations contribuent à la séparation entre les lieux d'articulation des occlusives. Outre le voisement, *la force d'articulation* est une deuxième dimension phonétique jouant un rôle non négligeable dans la différenciation de l'occlusive voisée et celle non voisée. Cette propriété est présentée dans la section suivante.

#### 1.2.4.4 La force d'articulation

La force d'articulation est toujours citée dans la littérature en tant qu'indice contribuant à la distinction des deux catégories d'occlusives voisées et non voisées dans plusieurs langues (Lisker et Abramson, 1964). Les occlusives non voisées, appelées *fortis* ou tendues, sont généralement réalisées avec plus de force articuloire que les occlusives voisées dites *lenis* ou relâchées. Cette différence de force d'articulation engendre une variation de pression conduisant à des changements d'intensité du bruit de relâchement. Delattre (1941) définit la force d'articulation par les termes suivants : « Nous croyons qu'il faut entendre par la force d'articulation consonantique la somme d'énergie nécessaire pour fournir la totalité des efforts musculaires qui prennent part à l'émission d'une consonne. Ces efforts musculaires sont connus sous les noms de : expiration, tension, tenue, détente, contraction, pression, rapprochement, éloignement, resserrement, occlusion, etc. et sont les modifications apportées à l'état de repos des organes phonateurs : poumons, larynx, pharynx, voile du palais, langue, joues, lèvres, etc. »<sup>36</sup>.

Jakobson et Halle (1962) soulignent que dans la production des sons relâchés, le conduit vocal expose le même comportement que leurs contreparties tendues mais avec une "atténuation significative". Cette atténuation se manifeste par une diminution de la pression de l'air dans la cavité orale, par une déformation faible du conduit vocal par rapport à sa situation neutre

35. Le mouvement de la mâchoire accélère également la vitesse de la lèvre inférieure.

36. Page 220.

et par un relâchement plus rapide de la constriction. L'origine de cette opposition a suscité de nombreux commentaires dans la littérature. Par exemple, Harris et al. (1965) ont examiné si la quantité d'énergie musculaire joue un rôle dans la distinction entre les occlusives bilabiales tendue et relâchée. Les occlusives /p, b/ ont été prononcées dans des syllabes CVC en positions initiale et finale par cinq locuteurs. Des électrodes ont été placées sur les lèvres et la mâchoire pour détecter leurs mouvements pendant la production des mots. Ces électrodes transforment les activités musculaires en signaux électriques interprétables. Les résultats indiquent en général une légère différence de l'effort musculaire entre tendue vs. relâchée mais cette différence est faible et quelquefois inexistante chez certains locuteurs. Les auteurs expliquent que l'effet observé de l'effort musculaire ne permet pas d'offrir une distinction phonémique entre /p/ et /b/. Ils ajoutent que la différence essentielle entre ces deux occlusives se trouve ailleurs.

Malécot (1970) critique également les travaux qui considèrent que l'opposition tendue vs. relâchée est uniquement le résultat de la variation de l'énergie musculaire. L'auteur explique les recherches qu'il a menées pendant 15 ans sur ce sujet et qui soutiennent l'idée que la force d'articulation est un attribut significatif des consonnes et qu'elle contribue à l'opposition tendue/relâchée. Toutefois, il considère que le fait de penser que l'énergie articulatoire est le facteur principal dans l'opposition tendue/relâchée n'est qu'un cas de synesthésie<sup>37</sup>. C'est une impression proprioceptive erronée basée principalement sur les pressions intrabuccales générées par l'action du blocage du flux d'air par la glotte, sur le relâchement ou la constriction du passage buccal, sur le sphincter vélo-pharyngé et sur la durée de la fermeture de l'occlusive. En effet, l'auteur expose que l'impulsion et les pics de pulsation de la pression développés au cours de la fermeture de l'occlusive sont fortement associés aux classes consonantiques. Dans tous les contextes phonétiques, les valeurs sont plus élevées pour les non voisées que pour leurs correspondantes voisées. Il ajoute que la durée de la clôture des occlusives impacte significativement la distinction des deux catégories puisqu'elle est plus longue pour les occlusives non voisées que pour leurs contreparties voisées.

Malécot (1970) souligne que la discrimination entre les classes voisée et non voisée est basée sur certains traits redondants. Les catégories se différencient principalement par la présence/absence de barre de voisement, par les durées de la voyelle suivante et de la clôture, par la pression intra-buccale, par le débit des transitions formantiques, etc. Lorsqu'un trait est absent, il est compensé par un autre existant.

Les sections précédentes ont présentées quelques caractéristiques qui permettent de distinguer les différentes catégories des occlusives. Un facteur important pouvant influencer ces caractéristiques est la position syllabique. En effet, l'occlusive en position initiale n'a pas exactement les mêmes caractéristiques phonétiques que quand elle est en position finale. La section suivante présente les changements engendrés par les différentes positions syllabiques de l'occlusive.

---

37. La synesthésie est définie comme une sensation subjective d'un sens autre que celui qui est en train d'être simulé (Malécot, 1970).

### 1.2.4.5 La position syllabique

Parmi les différentes positions de l'occlusive, il est possible d'en distinguer deux selon sa position avec la voyelle : l'occlusive pré-vocalique ou bien occlusive en position initiale (CV) et celle post-vocalique nommée également occlusive en position finale (VC). Comme il a été signalé précédemment (Figure 1.14), la position syllabique de l'occlusive inverse les étapes de sa production. Quand elle est en position initiale (attaque), la réalisation de l'occlusive commence par la fermeture suivie du relâchement et de la transition vers le son suivant tandis qu'en position finale (coda), la prononciation débute par la transition du son précédent qui est succédé par la fermeture et le relâchement (ou non). Plusieurs études ont montré qu'il existe des asymétries articulatoires entre les consonnes en position syllabique initiale *vs.* finale (Keating et al., 2001). Un résumé des résultats observés dans la littérature concernant la position syllabique a été présenté par Fougeron (1998) dans lequel elle énonce que « En position initiale, les consonnes ont un geste d'ouverture de la glotte plus long et plus ample ; les voyelles sont précédées d'un coup de glotte ou sont glottalisées. Les consonnes et les voyelles initiales ont une activité musculaire labiale plus grande. Au niveau du velum, les consonnes orales et nasales initiales ont une position du velum plus élevée. Au niveau de la langue, les consonnes ont une position de la langue plus élevée et une pression de la langue contre le palais plus importante »<sup>38</sup>. Ces asymétries génèrent des changements acoustiques importants qui peuvent conduire à des modifications perceptuelles. Par exemple, Sharf et Hemeyer (1972) trouvent que les auditeurs identifient mieux les consonnes en position finale qu'en position initiale après la suppression du bruit de relâchement (seulement avec les transitions formantiques). Selon les auteurs, les transitions formantiques des consonnes en position initiale portent moins d'informations que celles en position finale.

Keating et al. (2001) ont étudié l'étendue et la durée du contact lingo-palatal des consonnes /tʃ, dʒ, ʃ, ʒ, s, z, t, d, k, l, n/ en position initiale *vs.* finale dans des séquences CVC. Les résultats révèlent que le contact lingo-palatal est significativement plus grand en position initiale qu'en position finale pour les consonnes /t, d, k, l, n/. Au contraire, aucune différence statistique n'est observée pour le contact des fricatives et de l'affriquée /dʒ/ dans les deux positions. Toutefois, l'affriquée /tʃ/ présente un effet opposé puisque le contact en position finale est statistiquement plus important que celui en position initiale. Quant à la durée du contact, elle est globalement plus longue en position initiale qu'en position finale, peu importe la consonne. Les auteurs concluent que « [...] *the word-level asymmetry in consonant articulation, by which word-initial consonants are stronger in articulation than word-final consonants, is robust for certain consonants, the non-sibilant non-continuants /t, d, k, n/ and /l/, but is reflected in greater acoustic duration for other word-initial consonants* »<sup>39</sup>.

38. Page 29.

39. « [...] l'asymétrie de l'articulation des consonnes au niveau du mot, selon laquelle l'articulation des consonnes initiales est plus forte que celle finale, est valide pour certaines consonnes (les non sibilantes non continues /t, d, k, n, l/). Elle est pourtant reflétée dans une durée acoustique plus importante pour d'autres consonnes initiales », page 28 [notre traduction].

En conclusion, les études précédentes mettent en avant que la position de la consonne peut avoir un impact sur sa qualité et sur son interaction avec les sons adjacents. Lors de l'examen d'une consonne, il faut donc prendre en compte sa position syllabique. Dans cette section et dans toutes celles précédentes décrivant les caractéristiques acoustiques des voyelles et des occlusives, l'attribut dominant est le changement continu des propriétés acoustiques des segments en fonction des différents facteurs tels que le contexte adjacent, les variations temporelles, la position syllabique, etc. Des chercheurs ont essayé d'identifier des caractéristiques stables (appelées invariants) qui permettent aux locuteurs de percevoir les sons malgré la présence de ces facteurs. La théorie du locus et la forme globale constituent deux exemples d'invariant supposé. D'autres chercheurs ont argumenté que ces invariants sont toutefois influencés par les facteurs cités plus tôt. La section suivante propose quelques recherches discutant de ce débat.

### 1.3 Invariance et variabilité

Au moment du relâchement de l'occlusive, un ensemble d'événements acoustiques existe qui sont constitués principalement du bruit de relâchement suivi des changements spectraux rapides. Ces derniers reflètent les structures articulatoires lorsqu'elles passent de la configuration fermée du conduit vocal à celle ouverte. Ces événements fournissent des informations qui permettent aux auditeurs d'identifier les lieux d'articulation des occlusives (points de fermeture) (Ohde et Stevens, 1983). Les données issues des mesures acoustiques et des expériences de perception ont mené à plusieurs hypothèses concernant la manière avec laquelle ces informations acoustiques sont traitées par les auditeurs pour la reconnaissance du lieu d'articulation des consonnes dans les syllabes CV. Une des hypothèses avance que plusieurs indices acoustiques sont utilisés par l'auditeur mais que ceux-ci sont dépendants du contexte vocalique adjacent (entre autres, Dorman et al. (1977)). Par conséquent, l'auditeur doit considérer l'environnement phonétique lors de l'identification du lieu d'articulation. Une autre hypothèse suggère la présence de propriétés acoustiques invariantes (qui ne changent pas) pour chaque lieu d'articulation qui sont indépendantes du contexte vocalique. Selon cette hypothèse, l'identification du lieu d'articulation résulte d'un processus plus direct car il existe une relation plus simple entre la propriété acoustique détectée dans le système auditif et le trait phonétique sous-jacent (Ohde et Stevens, 1983). Plusieurs recherches ont été conduites afin d'examiner la présence de ces propriétés dans les occlusives (par exemple, Delattre et al. (1955) et Stevens et Blumstein (1978)).

Dorman et al. (1977) exposent que les indices qui apportent les informations importantes dans la perception du lieu d'articulation des occlusives (le bruit de relâchement et les transitions formantiques) changent en fonction de plusieurs facteurs dont le contexte vocalique, le voisement de l'occlusive (voisée vs. non voisée) et les variations inter- et intra-locuteurs (dimensions du conduit vocal, habitudes articulatoires, état émotionnel, etc.). Ils ajoutent que les prédictions de l'importance perceptuelle attachée à ces indices peuvent être significatives dans

le meilleur des cas malgré ces variations. Cependant, la probabilité qu'un seul indice soit le déterminant unique du lieu d'articulation dans tous les contextes est extrêmement faible. Les auteurs conçoivent deux expériences de façon à analyser l'importance perceptuelle du bruit de relâchement et des transitions formantiques.

La première expérience étudie le rôle des bruits de relâchement et des transitions formantiques dans la perception de la parole naturelle en supprimant systématiquement ces indices des occlusives /b, d, g/ produites avec neuf voyelles (/i, ɪ, ε, æ, ʌ, a, ɔ, u, ʊ/) par un locuteur américain dans des séquences /CVd/. Un total de 23 auditeurs a écouté trois types de stimuli correspondant 1) aux séquences /CVd/ originales, 2) aux séquences /CVd/ après la suppression du bruit de relâchement (transitions + voyelle), 3) aux séquences /Vd/, prononcées précédemment par le même locuteur, avec le bruit de relâchement des séquences CVd correspondantes (bruit de relâchement + voyelle). La deuxième expérience est une réplique de la première mais réalisée avec un autre locuteur. Les résultats mettent en avant que l'importance perceptuelle du bruit de relâchement et des transitions formantiques en tant qu'indices désignant le lieu d'articulation varie en fonction de la consonne, de la voyelle et du locuteur.

Selon les réponses des auditeurs, le bruit de relâchement de la bilabiale est un indice faible pour cette occlusive. Cependant, son rôle est plus important dans la deuxième expérience car la durée du bruit de relâchement du second locuteur est plus long que celui du premier<sup>40</sup>. Dans les deux expériences, le bruit de relâchement est plus efficace devant les voyelles arrondies que celles non arrondies. Quant aux séquences avec les transitions formantiques, elles étaient aussi performantes que celles originales devant toutes les voyelles à l'exception de /u/ et de /ʊ/ dans les deux expériences.

Concernant l'alvéolaire dans les séquences des deux locuteurs, le bruit de relâchement est un bon indice devant les voyelles antérieures mais pas dans les autres contextes alors que les transitions formantiques sont des indices forts devant les voyelles centrales et postérieures mais faibles avant celles antérieures.

Les différences inter-locuteurs sont plus marquées pour les stimuli de la vélaire. Pour le premier locuteur, le bruit de relâchement est un indice faible devant toutes les voyelles (sauf /ʊ/) tandis que pour le second, il est plus efficace devant les voyelles postérieures et arrondies que devant celles antérieures et non arrondies. Par contre, les transitions formantiques de la bilabiale et de l'alvéolaire sont plus importantes que celles de la vélaire pour les deux locuteurs.

Dorman et al. (1977) exposent que les résultats de ces expériences laissent suggérer qu'il existe une performance réciproque entre les bruits de relâchement et les transitions formantiques. Quand l'importance perceptuelle est élevée pour l'un, elle diminue pour l'autre. Ces relations réciproques sont influencées par les contextes vocaliques adjacents. Ils concluent que les invariants acoustiques peuvent être spécifiés pour un segment phonétique, cependant cette

---

40. Le second locuteur a produit des bruits de relâchement plus longs que ceux réalisés par le premier pour toutes les occlusives avec 10 vs. 4 ms pour /b/, 12 vs. 6 ms pour /d/ et 12 vs. 20 ms pour /g/, respectivement. Cette durée plus grande a amélioré l'identification des consonnes par les auditeurs à plusieurs niveaux.

spécification doit être en lien avec le contexte de ce segment.

L'existence des caractéristiques invariantes dans les transitions formantiques des occlusives a été examinée par Kewley-Port (1982) en anglais américain. L'objectif de cette recherche est de vérifier si les indices invariants proposés dans la littérature peuvent être observés et mesurés dans la parole naturelle. Un locuteur a prononcé des syllabes CV où C = /b, d, g/ et V = /i, ɪ, e, ε, æ, a, o, u/ dans une phrase porteuse (3 consonnes × 8 voyelles × 5 répétitions). Dans un premier temps, l'auteure analyse les débuts des transitions et les états stables des trois premiers formants ainsi que la durée des transitions. Les résultats indiquent que les états stables des voyelles sont statistiquement similaires dans les différents contextes consonantiques. Par conséquent, ils ne peuvent pas à eux seuls jouer de rôle dans la catégorisation des occlusives. Quant aux débuts des transitions des trois occlusives, ils sont statistiquement différents pour les trois formants dans tous les contextes vocaliques confondus. Les débuts des transitions des trois occlusives ont été comparés par contexte vocalique (quand le contexte était connu). Il a été constaté que les débuts du F1 seuls ne sont pas un corrélât fiable du lieu d'articulation puisqu'ils ne montrent pas de différence significative dans cinq des huit contextes vocaliques. Les débuts de F2 quant à eux sont statistiquement distincts dans tous les contextes. Toutefois, les analyses statistiques par paires dévoilent que la différence est significative dans six contextes car le /b/ et le /d/ dans le contexte de /i/ et le /d/ et le /g/ dans le contexte de /o/ sont similaires. Pour le F3, les débuts des transitions sont statistiquement différents dans cinq contextes mais seulement la voyelle /æ/ présente des catégories distinctes pour /b/, /d/ et /g/. Selon ces résultats, Kewley-Port (1982) avance que les transitions formantiques seules ne peuvent pas séparer les trois catégories consonantiques dans tous les contextes.

Des analyses discriminantes ont été effectuées dans l'objectif de tester si la combinaison des débuts des transitions de F2 et de F3 peut catégoriser les trois lieux d'articulation. Les résultats révèlent que quand le contexte vocalique est connu (/bi/, /di/, /gi/, etc.), le taux de classification est de 97 % alors que, quand il est inconnu (/b/, /d/, /g/), le taux de classification est de 68 %. Selon l'auteure, la combinaison des deux informations ne peut pas être considérée comme un indice invariant fiable du lieu d'articulation dans tous les contextes vocaliques.

Kewley-Port (1982) examine également la validité de la théorie du locus par l'estimation du locus virtuel du F2 et du F3 pour les trois occlusives. Des droites de régression linéaire prolongeant les transitions du F2 et du F3 ont été calculées. Deux mesures ont ensuite été fixées sur la pente de régression à deux moments dans le temps. Le premier, appelé le locus des transitions, est au moment du relâchement de l'occlusive et le second, nommé transition prolongée, est à 10 ms avant celui-ci<sup>41</sup> (Figure 1.24). Les résultats indiquent que les bilabiales ont au moins deux loci de F2 à 1645 Hz avec les voyelles antérieures et à 1090 Hz avec celles postérieures alors que le F3 a un locus vers 2337 Hz. Quant aux alvéolaires, un locus invariant émerge pour le F2 et le F3, respectivement à 1797 et 2581 Hz. Les vélaires ne semblent pas avoir

41. Selon Delattre et al. (1955), la fréquence du locus virtuel peut être extrapolée à partir des transitions formantiques.

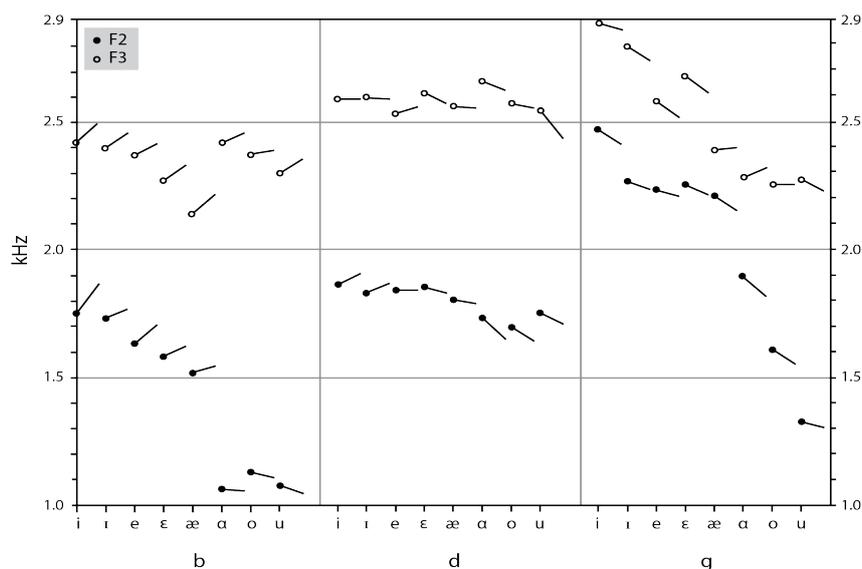


FIGURE 1.24 – Représentation des loci de transitions de F2 (cercles noirs) et de F3 (cercles blancs) des trois occlusives dans les huit contextes vocaliques. Les lignes représentent les premières 10 ms qui projettent les transitions formantiques (d’après Kewley-Port (1982)).

de loci fixes pour F2 et F3. Au contraire, les fréquences des loci de transitions augmentent de /u/ à /i/. Selon l’auteure, malgré l’existence d’un locus invariant pour le /d/, il se chevauche avec les loci des autres consonnes. De ce fait, il ne peut pas être considéré comme un indice fiable du lieu d’articulation. Kewley-Port (1982) conclut que les résultats révèlent l’absence d’un indice invariant du lieu d’articulation dans les transitions formantiques. De plus, les indices observés dans les transitions ne sont pas toujours fiables dans tous contextes vocaliques.

Kewley-Port (1983) critique également la conclusion de Blumstein et Stevens (1980) sur l’existence d’un invariant acoustique dans un intervalle temporel fixe situé dans les premières 20 ms après le relâchement. D’après l’auteure, cette conclusion ne convient pas à la nature dynamique et variable de la parole. La production des segments phonétiques inclut des changements rapides et continus des articulateurs reflétés par des variations de fréquences et d’énergies dans le temps sur le plan acoustique. Par conséquent, l’hypothèse de Stevens et Blumstein selon laquelle la forme globale du spectre située dans une fenêtre temporelle fixe de 20 ms après le relâchement ne peut pas être adéquate car l’intervalle fixe ne capture pas les variations du spectre qui se produisent dans des intervalles temporels plus courts. En outre, la durée du bruit de relâchement varie en fonction du lieu d’articulation et elle est estimée entre 5 et 10 ms pour les labiales et les alvéolaires et entre 20 et 30 ms pour les vélaires. Dès lors, des informations spectrales du bruit de relâchement ne peuvent pas être renseignées pour les vélaires dans l’étude de Stevens et Blumstein car le relâchement relativement lent nécessite un intervalle temporel supérieur à celui de 20 ms.

Kewley-Port (1983) propose ainsi une version modifiée de l’hypothèse de Stevens et Blumstein où la dimension temporelle est intégrée dans la représentation du spectre et la durée de l’intervalle est augmentée (40 au lieu de 20 ms). Dans cette représentation, le spectre des premières

40 ms est affichée toutes les 5 ms (Figure 1.25). De ce fait, les variations qui ont lieu dans des durées courtes peuvent être captées. L'auteure définit trois critères qui permettent l'identification visuelle du lieu d'articulation des occlusives à partir de la représentation tridimensionnelle du spectre :

1. L'inclinaison du spectre au début du *burst* (dans la première fenêtre temporelle de 5 ms) dans les fréquences entre 0 et 3500 Hz. Deux catégories d'inclinaison sont possibles : montante et descendante/plate.
2. L'existence d'un début tardif de l'énergie dans les fréquences basses. Il s'agit de l'apparition du pic de F1 à partir de la quatrième fenêtre temporelle (à partir de 20 ms). Deux catégories sont possibles : début tardif et début non tardif.
3. L'existence de pics de fréquences centrales qui s'étendent dans le temps. Ce critère indique la présence d'un pic proéminent entre 1000 et 3500 Hz qui se produit dans trois ou quatre fenêtres temporelles. Deux catégories sont possibles : oui (ces pics existent) et non (ces pics n'existent pas).

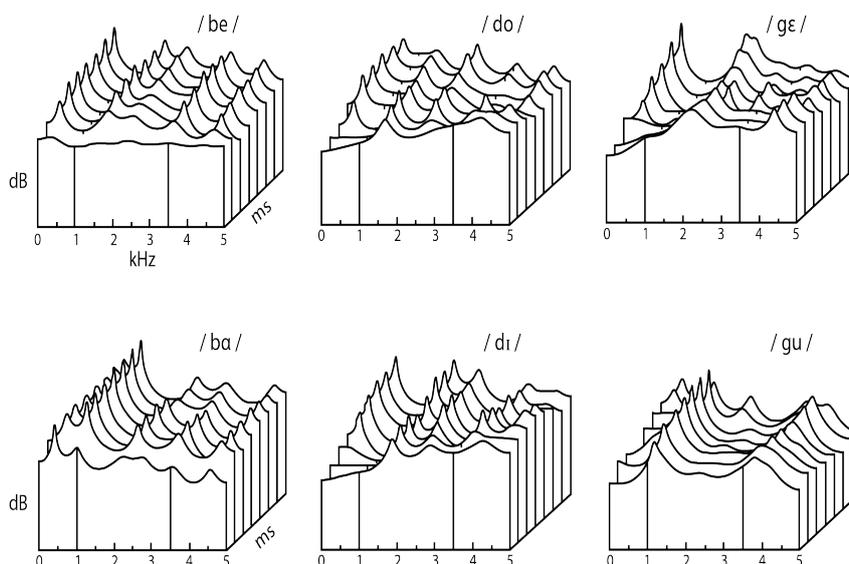


FIGURE 1.25 – Représentations du spectre courant des occlusives /b, d, g/ articulées avec des contextes vocaliques différents où l'énergie est en ordonnée, la fréquence en abscisse et le temps est en troisième dimension en ms (d'après Kewley-Port (1983)).

L'auteure souligne que lors de l'application de ces critères sur la représentation tridimensionnelle du spectre, comme celle affichée dans la Figure 1.25, un invariant relationnel peut être obtenu pour désigner le lieu d'articulation. Autrement dit, la combinaison de ces critères peut conduire à catégoriser correctement les occlusives dans les différents contextes. Une matrice a été créée contenant les critères qui aident à l'identification des trois occlusives /b, d, g/ (Table 1.5). Lorsque les deux catégories d'un critère peuvent se produire pour une occlusive, le symbole "?" est attribué.

Kewley-Port (1983) met en place une expérience dans le but d'évaluer l'efficacité de cette hypothèse de l'invariant relationnel. Des séquences CV ont été prononcées par deux locuteurs

Inclinaison du <i>burst</i>	Début tardif du F1	Pic de fréquences centrales	Consonne assignée
Descendante ou plate	Non	Non	/b/
Montante	?	Non	/d/
?	Oui	Oui	/g/

TABLE 1.5 – Matrice utilisée pour assigner le lieu d’articulation des occlusives à partir des catégories des critères définies plus haut. Le " ? " signifie que les deux valeurs de la variable sont possibles pour cette occlusive (Kewley-Port, 1983).

(une femme et un homme) de l’anglais américain où  $C = /b, d, g/$  et  $V = /i, e, a, o, u/$ . De plus, les données de l’article de Kewley-Port (1982) ont été ajoutées à celles des deux locuteurs<sup>42</sup>. Le spectre de chaque réalisation a été analysé après le relâchement toutes les 5 ms pendant les premières 40 ms de manière à obtenir la représentation tridimensionnelle, appelé spectre du temps variant (Figure 1.25). Il a été demandé à trois participants ("juges") d’identifier visuellement les occlusives qui correspondent à chacune de ces représentations en se servant de la matrice dans la Table 1.5. Les résultats révèlent que les juges ont réussi à classifier correctement les occlusives à partir du spectre du temps variant dans 88 % des cas. Selon l’auteure, la représentation du spectre du temps variant fournit les informations qui ne sont pas données par le spectre avec un temps fixe telles que les changements rapides de fréquence et d’énergie au cours du temps et l’apparition tardive du pic de F1 des vélaires.

Dans leur étude, Kewley-Port et al. (1983) ont testé dans trois expériences si l’analyse du spectre avec un temps variable offre un plus grand nombre d’informations que celui avec un temps fixe (hypothèse de Kewley-Port (1983) vs. celle de Stevens et Blumstein (1978)). La première expérience a été conçue pour vérifier si les premières 20 ms après le relâchement sont suffisantes à la reconnaissance des lieux d’articulation. Un sous-ensemble des séquences CV des deux locuteurs masculins de l’étude précédente a été utilisé dans cette expérience ( $3 C \times 5 V \times 2$  locuteurs). Dans un premier temps, les 30 séquences sans modification ont été écoutées dix fois dans un ordre aléatoire par six auditeurs. Les participants ont identifié correctement les occlusives dans 98.8 % des cas. Ce taux de classification sert ensuite de référence pour les séquences modifiées. Dans un second temps, les 30 séquences CV ont été modifiées de manière à obtenir uniquement les portions initiales du signal. Cinq découpages différents ont été réalisés pour chaque séquence CV. Le premier découpage pour le /d/ et le /g/ contient seulement le bruit de relâchement donnant des portions avec une durée moyenne de 14 et 21 ms respectivement. Le deuxième découpage inclut le bruit de relâchement et la première pulsation de la voyelle. Quand au /b/, il a été difficile d’avoir le bruit de relâchement séparé du contexte de la phrase porteuse qui le précède<sup>43</sup>. En conséquence, le premier découpage de cette occlusive

42. Soit un total de 162 séquences dont 72 du locuteur masculin de l’expérience de Kewley-Port (1982) ( $3 C \times 8 V \times 3$  répétitions) et 90 des deux nouveaux locuteurs (féminin et masculin) ( $3 C \times 5 V \times 3$  répétitions  $\times 2$  locuteurs).

43. Dans la phrase porteuse "Teddy said" CV, le voisement de l’occlusive finale de *said* continue jusqu’au bruit

contient le bruit de relâchement et la première pulsation constituant des portions avec une durée moyenne de 13 ms. Le deuxième découpage comprend le bruit de relâchement et deux pulsations de la voyelle. Les troisième, quatrième et cinquième découpages des trois occlusives incluent le bruit de relâchement et la troisième, la cinquième et la septième pulsations respectivement. Les 150 portions coupées (occlusives tronquées) ont été entendues par neuf auditeurs dans six sessions.

Les résultats montrent que les taux de l'identification correcte de /b/ et /d/ sont autour de 90 % pour les stimuli les plus courts (premiers découpages) et montent à 100 % à partir des troisièmes découpages. La performance de la reconnaissance du /g/ est différente des deux autres occlusives. Les stimuli avec seulement le bruit de relâchement (durées plus courtes) ont été identifiés correctement dans 70 % des cas. Les autres stimuli plus longs n'ont jamais atteint un taux de classification de 100 %. Les réponses des stimuli qui correspondent aux deux durées proposées par Blumstein et Stevens (1980) (20 ms) et par Kewley-Port (1983) (40 ms) ont été sélectionnées afin de vérifier si elles sont suffisantes à la reconnaissance des occlusives. Les taux d'identification des stimuli avec 20 ms sont de 96 %, 94 % et 73 % alors que ceux avec 40 ms sont de 99 %, 98 % et 90 % pour /b/, /d/ et /g/ respectivement. Selon les auteurs, ces résultats mettent en avant que les premières 20 ms du signal ne permettent pas d'obtenir une bonne intelligibilité de toutes les occlusives.

Dans la deuxième expérience, les auteurs ont synthétisé les trois occlusives /b, d, g/ dans les trois contextes vocaliques /i, a, u/ en employant deux procédures différentes à l'aide du synthétiseur de Klatt (1979). La première respecte les propriétés du spectre avec un temps fixe proposées par Stevens et Blumstein alors que la seconde applique les caractéristiques du spectre avec un temps variant suggérées par Kewley-Port. Chaque séquence synthétisée a été modelée à partir de la syllabe naturelle correspondante par des techniques d'appariement visuel du spectre. Ces séquences CV ont été générées avec trois durées distinctes : 20, 30 et 40 ms. Les séquences CV naturelles et synthétiques ont été écoutées par dix participants qui ont donné un degré de certitude à leur réponse (pas sûr, sûr, très sûr). Les stimuli des séquences naturelles ont été reconnus à 94 %. Quant aux séquences synthétiques, le taux global de l'identification révèle que les stimuli produits avec la seconde procédure (spectre avec un temps variable) ont été mieux reconnus que ceux de la première (78 % vs. 68 %). De plus, le niveau de certitude est généralement plus élevé pour les réponses des stimuli de la seconde procédure. Il faut toutefois préciser que l'influence de la durée des stimuli n'a pas été observée pour les réponses des deux procédures malgré l'effet remarqué dans les réponses en parole naturelle.

Les stimuli de la deuxième expérience ont été synthétisés par le biais d'une autre méthode dans la troisième expérience. En effet, selon les auteurs, les procédures de l'appariement visuel du spectre peuvent être critiquées puisqu'elles dépendent de la subjectivité de l'expérimentateur concernant la précision visuelle du spectre. Deux ensembles de stimuli ont alors été synthétisés au moyen de coefficients de prédiction linéaire dérivés à partir des stimuli naturels de relâchement de l'occlusive cible.

employés dans la deuxième expérience. Le premier ensemble a été réalisé par des coefficients qui spécifient le début du spectre comme l'ont défini Stevens et Blumstein. Autrement dit, l'information spectrale était constante durant le stimulus. L'autre ensemble a été synthétisé par des coefficients mis à jour tous les 5 ms. Les stimuli de cet ensemble ont des spectres variants dans le temps comme l'a proposé Kewley-Port. Des algorithmes ont traité automatiquement les deux ensembles de stimuli sur un ordinateur indépendamment de l'expérimentateur. Chaque séquence a deux durées, 20 et 40 ms. Les deux ensembles de stimuli synthétisés ainsi que ceux naturels ont été écoutés par dix auditeurs<sup>44</sup>. D'après leurs réponses, le taux de classification des stimuli naturels est également de 94 % pour ce groupe de participants. Les stimuli synthétisés avec un spectre variant dans le temps ont aussi été mieux identifiés que ceux avec un spectre fixe (87 % vs. 59 %). Les analyses statistiques montrent qu'un effet significatif de la durée des stimuli naturels et synthétiques dynamiques est noté pour le /g/ alors qu'il ne l'est pas pour les stimuli synthétiques fixes. L'effet temporel sur les réponses des /b/ et des /d/ pour les trois types de stimuli est non significatif. Les résultats de cette troisième expérience dévoilent qu'ils sont en accord avec les propositions de Kewley-Port (1983) et ne soutiennent pas celles de Blumstein et Stevens (1980). Dans les deux cas, un temps supérieur à 20 ms est nécessaire pour percevoir le /g/. L'hypothèse prédite est donc qu'une augmentation de la durée des stimuli entraîne une amélioration de la reconnaissance de l'occlusive vélaire, mais que cette hypothèse n'a pas pu être révélée dans l'expérience avec les procédures proposées par Blumstein et Stevens (avec un temps fixe).

D'après les résultats des trois expériences, les informations nécessaires à l'identification des occlusives se trouvent dans les premières 20 à 40 ms du signal. Il a également été mis en évidence que les propriétés spectrales dynamiques situées au début des séquences CV spécifient le lieu d'articulation des occlusives.

Dans cette section, la notion de l'invariance et de la variabilité ont été présentées. Plusieurs phonéticiens ont recherché des propriétés spectrales invariantes qui permettent la perception du lieu d'articulation des occlusives. Il a été montré que les indices invariants proposés dans la littérature (comme le locus et la forme globale du spectre) sont dépendants des contextes vocaliques. Autrement dit, l'indice d'une occlusive produite avec une voyelle donnée est modifié lorsqu'elle est articulée avec une autre. D'après les études de Dorman et al. et Kewley-Port, la nature dynamique de la parole implique des propriétés acoustiques variables. En effet, la combinaison de celles-ci est nécessaire pour une meilleure reconnaissance. Par conséquent, l'impact du contexte adjacent ou de la coarticulation rend la question de l'invariant plus difficile. La coarticulation est ainsi considérée comme un facteur principal de la dynamique de la parole. Elle offre une meilleure compréhension de la variabilité observée dans les différents indices de la parole<sup>45</sup>.

---

44. Les trois groupes de participants dans les trois expériences sont différents.

45. La notion de coarticulation sera présentée dans le chapitre suivant.

## 1.4 Conclusion

Ce chapitre a été consacré à la description des principales caractéristiques phonétiques des voyelles et des consonnes occlusives. La théorie source-filtre fournit une explication de la relation acoustique-articulatoire des sons de la parole. Selon cette théorie, la production de la parole peut être divisée en deux parties où la source du son est le larynx et le filtre qui la modifie est le conduit vocal. Le résultat de l'opération source-filtre peut être séparé en deux grandes catégories : les voyelles et les consonnes. Ces dernières se distinguent par plusieurs éléments. Des mouvements rapides des articulateurs sont engagés lors de l'articulation des consonnes. Elles sont réalisées par la mise en place d'une constriction ou d'une fermeture d'une zone dans le conduit vocal où la combinaison de plusieurs sources est possible (friction, aspiration et voisement). Les voyelles, quant à elles, sont prononcées par des articulateurs plus lents. Elles sont produites avec un conduit vocal relativement ouvert et le voisement est la seule source utilisée.

Le spectrogramme a permis l'identification des traits acoustiques qui reflètent la source filtrée et qui aident à la perception des sons de la parole. Ces traits ont été testés expérimentalement par le biais de la synthèse vocale. En effet, les catégories de son se caractérisent par des indices acoustiques en commun et diffèrent entre elles par d'autres. Par exemple, les voyelles se distinguent des consonnes par la présence de formants relativement stables avec une intensité élevée et par une durée longue. Elles se différencient principalement entre elles par les fréquences des deux premiers formants. Quant aux occlusives, elles se distinguent des autres consonnes par la brièveté de leur réalisation et par la présence du bruit de relâchement. Elles se différencient entre elles par la fréquence et l'intensité du bruit de relâchement, par les transitions formantiques et par le VOT.

Chaque segment est alors caractérisé par un ensemble d'indices qui lui donne son identité. Toutefois, ces indices sont impactés par les contextes adjacents. Autrement dit, les indices des voyelles et ceux des consonnes s'influencent mutuellement lors de leurs articulations ou *coarticulation*. Cette dernière est considérée comme un facteur principal dans la modification des différents indices des sons de la parole. Les chercheurs se sont intéressés à la coarticulation afin de mieux comprendre les variations spectrales des différents sons de la parole. Cette notion sera présentée dans le chapitre suivant.

# Chapitre 2

## La coarticulation

L'une des propriétés de la parole est d'être continue. Les auditeurs sont capables de découper cette continuité en segments séparés et indépendants (Morais et al., 1979)<sup>1</sup>. Ces derniers apparaissent au niveau perceptif, pourtant il n'existe pas de démonstration définitive réelle au niveau articulatoire ou acoustique. Autrement dit, les segments peuvent avoir une représentation mentale relativement fixe qui pourrait correspondre partiellement à la notion de *phonème* ou de *phone*. Sur le plan articulatoire, les segments ne sont pas articulés de façon indépendante ou séparée mais étant coarticulés, les segments voisins se chevauchent et s'influencent. Sur le plan acoustique, le chevauchement articulatoire des segments est reflété par des mouvements transitionnels continus sur un spectrogramme. Il peut être difficile de dire avec précision le début ou la fin de ces segments. Le terme *coarticulation* est composé du préfixe *co-* qui indique l'adjonction et la simultanéité et le mot *articulation* qui désigne l'action de produire les sons de la parole. La notion de coarticulation révèle ainsi la prononciation conjointe et simultanée entre deux ou plusieurs sons. Ce chapitre a pour objectif de définir et d'expliquer la coarticulation et d'exposer les différentes théories et modèles expliquant comment et pourquoi la coarticulation a lieu.

### 2.1 Qu'est-ce que la coarticulation ?

La coarticulation se réfère globalement au fait qu'un segment phonologique n'est pas réalisé de la même manière lorsqu'il est produit avec d'autres segments. Cependant, il varie systématiquement pour devenir plus ou moins comme le segment adjacent (Kühnert et Nolan, 1999). Les mouvements des différents articulateurs effectués pour la production d'une succession de segments phonétiques se chevauchent dans le temps et interagissent ; par conséquent, la configuration du conduit vocal est influencée à tout moment par plus d'un segment. Deux types de

---

1. Les chercheurs Morais et al. (1979) ont montré que les illettrés n'arrivent pas à ajouter ou à supprimer des phonèmes à des non-mots en portugais en comparaison avec des ex-illettrés qui ont réussi à effectuer cette tâche. Ils concluent : « La prise de conscience de la parole comme une séquence de phonèmes n'est donc pas acquise spontanément au cours du développement cognitif mais exige un entraînement spécifique, lequel, pour la plupart des personnes, est fourni probablement par l'apprentissage de la lecture dans le système alphabétique », page 331.

coarticulation sont distingués : la coarticulation de droite à gauche ou dite anticipatoire et la coarticulation de gauche à droite ou dite persévérante. D'un côté, la coarticulation anticipatoire est reliée essentiellement aux effets temporels, c'est-à-dire que pour un segment donné les mouvements vers la configuration d'un caractère cible commencent avant les autres. D'un autre côté, la coarticulation persévérante est souvent considérée comme un effet mécano-inertiel (Kühnert et Nolan, 1999). L'articulateur répond au décalage entre les commandes neuronales successives et persiste en mouvement ou dans sa position après que la commande soit terminée. Ce type de coarticulation est visible dans la variabilité de la position des cibles<sup>2</sup> en fonction des contextes précédents (Gay, 1978b). La production de la consonne /g/ dans /gid/ est influencée par l'avancement de la langue causée par la production du /i/ (*coarticulation anticipatoire*). Dans le mot /t<sup>h</sup>a:l/ la voyelle /a:/ est influencée par le recul de la langue provoqué par la production de la consonne emphatique /t<sup>h</sup>/ (*coarticulation persévérante*). Les segments phonétiques sont sensibles au contexte, d'où le problème de la notion de l'invariant qui a occupé les phonéticiens pendant de nombreuses années (Chapitre 1). Cette notion peut être défini par la difficulté à donner une description physique ou acoustique d'une catégorie linguistique quelconque qui soit constante et toujours indépendante du contexte (Lindblom, 1990). Autrement dit, « [...] il s'agit de la présence de caractéristiques physiques dont l'auditeur pourrait systématiquement se servir pour identifier les unités phonétiques produites »<sup>3</sup>.

Kühnert et Nolan (1999) évoquent l'analogie entre la machine à écrire et le conduit vocal dans le but d'expliquer pourquoi le phénomène de la coarticulation a lieu dans la parole. Les lettres sont les unités de base des mots écrits, comme peuvent l'être les phonèmes (ou les phones) dans les mots parlés. Une machine à écrire est composée d'une série de marteaux porte-caractères. Pour écrire un mot, les marteaux qui portent les lettres correspondantes frappent sur le papier. Les lettres tapées sont indépendantes les unes des autres et elles sont à chaque fois réalisées de façon identique. La raison pour laquelle les phonèmes ne sont pas produits de la même manière que les mots sur une machine à écrire est qu'il n'existe pas de conduit vocal différent pour chaque phonème, au contraire de la machine à écrire où chaque lettre a un marteau séparé. Un conduit vocal doit alterner entre une configuration et une autre et s'adapter afin de satisfaire les exigences des sons dans la séquence. Étant contraint par les lois de la physique et les limites physiologiques, il produit les séquences en temps réel. Il ne peut pas changer instantanément d'une configuration cible à une configuration suivante, mais il se transforme progressivement d'une configuration cible à la suivante. Le conduit vocal étant limité en temps et en espace, les configurations cibles des phonèmes ne sont pas toujours atteintes toutefois elles s'influencent les unes les autres.

Une autre analogie de la coarticulation - abondamment citée dans la littérature - est celle de Hockett (1955) : « *Imagine a row of Easter eggs carried along a moving belt the eggs are of*

2. Le terme *cible* (*target* en anglais) a été utilisé principalement dans la littérature pour décrire la tendance d'un articulateur à s'approcher d'une position quelconque d'un phonème dans des conditions différentes (MacNeilage, 1970).

3. Meunier (2005), page 352.

*various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source ; the mess that emerges from the wringer represents the output of the speech transmitter* ». <sup>4</sup> Cette comparaison porte un ton négatif selon Kent et Minifie (1977). Ils mettent en avant ici que le processus de la parole est considéré comme un comportement "pathologique" qui génère ou aboutit à dégrader l'information et non comme un processus de transformation qui ordonne des unités linguistiques en séquences de mouvements articulatoires complexes. Néanmoins, cette analogie montre les difficultés et les défis que les phonéticiens rencontrent en abordant les questions de la production de la parole.

Le terme *coarticulation* est mentionné pour la première fois par Menzerath et de Lacerda dans "*Koartikulation, Steuerung und Lautabgrenzung : eine experimentelle Untersuchung*" publié en 1933. Toutefois, le postulat que les sons de la parole s'influencent mutuellement et varient selon le contexte était connu antérieurement. L'idée à l'époque était que chaque son a un état stable (position statique) et que les différents sons sont connectés par des "*glides*" courts (une articulation et un son transitionnels). Rousselot (1901-1908) explique qu'une voyelle peut être plus influente qu'une autre dans une séquence : « Entre les deux voyelles, la plus influente est la seconde, comme cela se montre très bien pour *a tu* dans "il *a tou*-rné". Dès le milieu de l'*a*, la ligne des lèvres s'élève à la sollicitation de l'*u* » <sup>5</sup>. Laclotte (1899) discute également de l'influence des segments adjacents les uns sur les autres : « Le problème se pose ainsi : 1° Quelle est l'influence d'une articulation subséquente sur l'articulation qui la précède ; 2° Dans le cas où le groupe renferme deux voyelles, l'influence de la seconde peut-elle s'étendre jusqu'à la première ? » <sup>6</sup>. Il a étudié le mouvement des lèvres et de la langue afin de répondre à la première question traitant de la coarticulation persévérante. Il explique qu'il suffit tout simplement de regarder dans un miroir pour constater l'influence de la voyelle sur la consonne dans les séquences /ba/ /bu/. Dans la première séquence, le /b/ est produit avec des lèvres écartées tandis que, dans la seconde, il est réalisé avec des lèvres arrondies. Dans les deux cas, les lèvres prennent la position qui convient à la voyelle pendant la prononciation de la consonne. Il a pu ensuite démontrer par le biais d'un palais artificiel que la langue prend la position de la voyelle /i/ dans la zone antérieure du palais quand la consonne /k/ est produite dans la séquence /ki/. Dans une étude aérodynamique sur des paires de type CV : /ba/ – /bi/, /da/ – /di/, /za/ – /zi/, /ʒa/ – /ʒi/, Laclotte trouve que le flux d'air et la pression des vibrations linguales de la consonne sont modifiés en fonction de la voyelle (Figure 2.1). En effet, la langue n'occupe pas la même position lors de la production de /i/ et de /a/. Le premier est produit avec la langue

4. « Imaginons une ligne d'œufs de Pâques portés sur un tapis roulant, les œufs sont de tailles variées et de couleurs différentes. À un moment donné, le tapis roulant emmène les œufs jusqu'à une essoreuse à rouleaux qui, à son tour, les écrase et les mélange à un certain degré. Les œufs avant l'écrasement représentent les phonèmes et la masse produite par l'essoreuse à rouleaux représente la sortie de la parole » , page 210 [notre traduction].

5. Page 947

6. Page 178

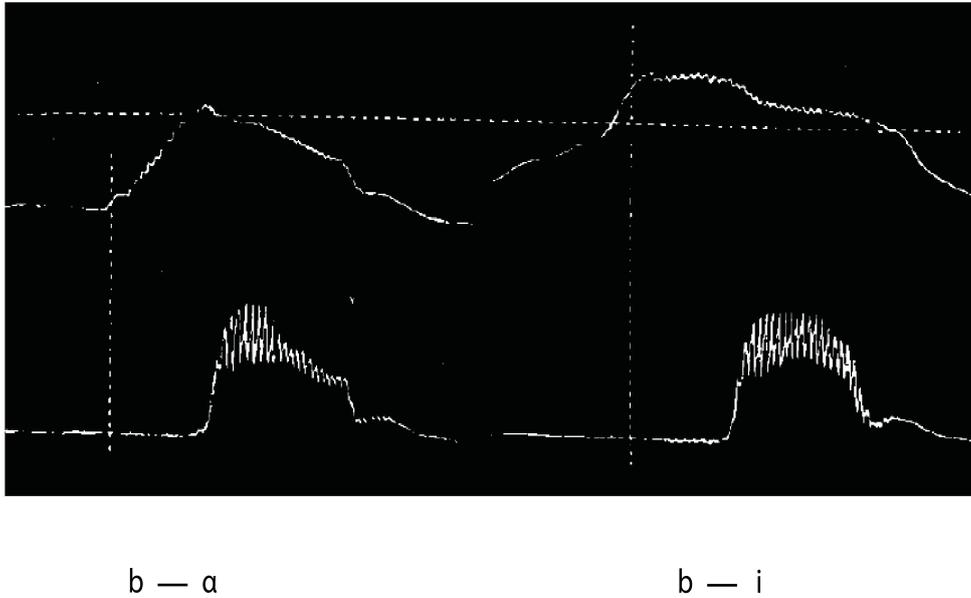


FIGURE 2.1 – Deux tracés de la production des deux séquences /ba/ – /bi/. Le tracé supérieur représente la pression et les vibrations linguales. Le tracé inférieur correspond à la masse et à la vitesse du courant d'air mesurées à la sortie de la bouche. Les lignes verticales indiquent le moment de relâchement de l'occlusion et la ligne horizontale permet d'apprécier l'élévation de la langue (emprunté à Laclotte (1899)).

rapprochée (élevée) vers le palais alors que le second est prononcé avec une langue éloignée (abaissée) du palais. Il a trouvé que pour toutes les consonnes dans les séquences étudiées, le tracé de la langue (tracé supérieur de la Figure 2.1) est plus élevé lorsqu'elle est produite avec le /i/ qu'avec le /a/. Il a aussi analysé des paires de séquences de type VCV (/ela/ – /eli/ et /eba/ – /ebi/) afin de vérifier si l'influence de la première voyelle ne s'étendait pas jusqu'à la seconde voyelle qui précède la consonne. Les tracés indiquent que le /e/ est toujours produit avec une langue plus élevée quand il est prononcé avec un /i/. L'auteur a examiné également si l'influence de la voyelle peut dépasser la syllabe (dans les séquences /vāta/–/vāti/ et /vata/–/vati/). Le /v/ est réalisé avec une langue plus élevée dans les séquences contenant un /i/ final. Il conclut que « l'influence d'une voyelle peut s'étendre jusque sur la syllabe précédente [...] »<sup>7</sup>.

Menzerath et de Lacerda (1933) ont utilisé des kymogrammes<sup>8</sup> et des mesures de flux d'air dans le but d'analyser la prononciation des consonnes bilabiales avec des voyelles en allemand

7. Page 188

8. Kymogramme : tracé obtenu à l'aide du kymographe. « Cet appareil qui était fondé sur un principe mécanique, était composé de deux parties : (1) une partie 'enregistrement' qui comprend un système pour capter le flux d'air phonatoire issu de la bouche et du nez, par l'intermédiaire d'une embouchure pour la pression buccale et d'une ou deux olives de verre pour la pression nasale. Le flux d'air est conduit grâce à des tuyaux de caoutchouc vers des membranes qui vont entrer en vibration sous l'effet de la pression d'air. Simultanément, on procède à l'enregistrement des vibrations laryngées par application d'un diaphragme sur le larynx au niveau des cordes vocales. (2) une partie 'inscription' dont l'élément principal est constitué d'un cylindre (tambour de Marey) lui-même entraîné par un système de rotation mécanique. Des stylets s'appuient sur le cylindre recouvert d'un papier enduit de noir de fumée, et inscrivent des tracés dont les variations d'amplitude sont proportionnelles à la quantité d'air expiré [...] » (Chafcouloff 2004, page 27).

dans des séquences CV et VC. N'ayant pas réussi à trouver des positions articulatoires stables, ils ont proposé que l'articulation est gouvernée par *la coarticulation et le contrôle*. La coarticulation (*Koartikulation*) indique que les articulateurs sont déjà en préparation pour le son suivant lors de la production du segment précédent. L'activité préparatoire de l'articulation commence le plus tôt possible. Le contrôle ou la gouvernance (*steuerung*) désigne que la déviation d'un articulateur de sa cible articulatoire est provoquée par la présence d'une autre cible provenant d'un autre son. Par exemple, dans la séquence /am/, la consonne impose la direction des mouvements articulatoires durant la production de la voyelle. Selon Menzerath et de Lacerda (1933), la structure de toute séquence implique des mouvements complexes entrecroisés et chaque articulation est une coarticulation.

Joos (1948) étudie le comportement du deuxième formant dans les séquences /gæg/ et /dæd/ dans son étude spectrographique innovatrice. Il explique que l'effet de chaque consonne est étendu jusqu'au milieu de la voyelle et, par conséquent, les deux effets se chevauchent au milieu. Il ajoute que le deuxième formant de /æ/ produit avec le /g/ est toujours plus élevé que celui prononcé avec le /d/ (Figure 2.2). L'auteur montre que l'influence de chaque consonne

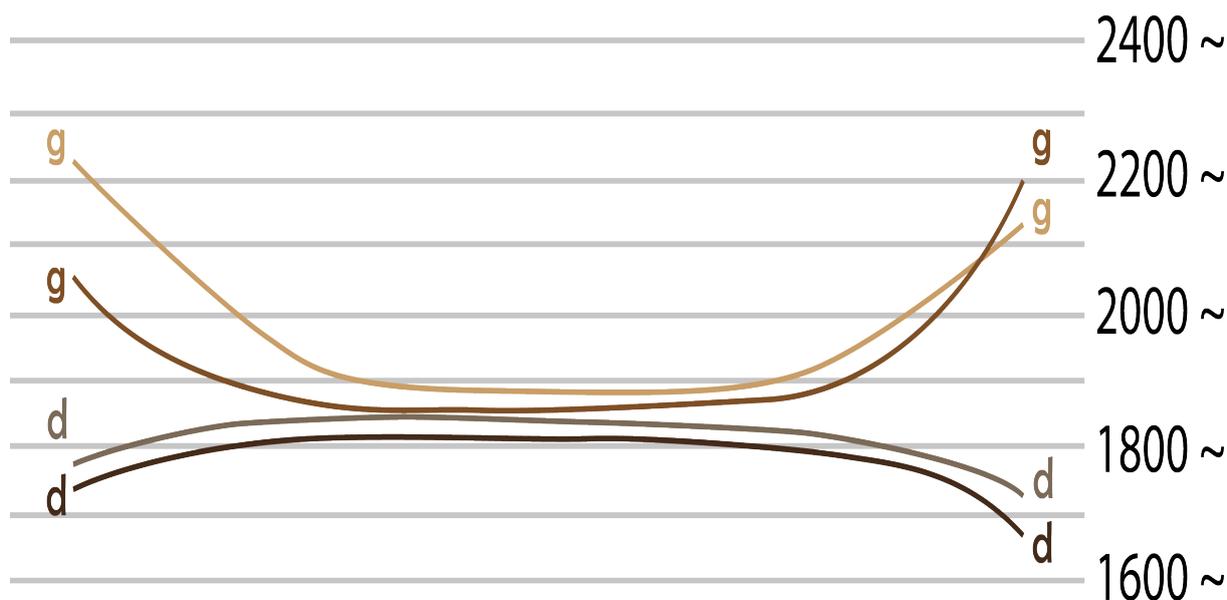


FIGURE 2.2 – Représentation schématique du deuxième formant de /æ/ dans quatre réalisations différentes (d'après Joos (1948)).

diminue progressivement au cours du temps. L'impact du /g/ initial sur le /æ/ disparaît graduellement et ne s'arrête pas brusquement. Quant au /g/ final, l'effet sur la voyelle ne commence pas soudainement, il augmente doucement à partir du milieu de la voyelle jusqu'à la fin. La même constatation est remarquée dans la séquence /dæd/. L'explication simple et incontestable est qu'il n'existe pas de transformation instantanée d'une position articulatoire à une autre. Auparavant, l'interprétation souvent proposée de ce phénomène était l'existence d'un "glide" qui intervenait entre le début et le deuxième son et un autre entre le deuxième et le troisième son dans une séquence comme /gæd/. Le chercheur montre que cette interprétation n'est

pas valable puisque les influences du premier son et du troisième s'étendent tout au long de la voyelle. Un découpage de la parole en petites portions reliées entre elles par des "glides" semble être impossible sur le plan acoustique ou articulatoire. Joos (1948) critique ainsi la théorie des "glides" et met en exergue que la considération de ceux-ci comme des articulations intercalées entre les sons n'est qu'un "mythe". Il précise que chaque moment durant l'articulation appartient à deux ou plusieurs sons. La contribution d'un son dans l'articulation totale se fait d'une manière lente et croissante. Il propose le terme "slur"<sup>9</sup> comme une notion technique descriptive du phénomène de coarticulation. Selon lui, la coarticulation est réalisée (ou comme il l'exprime « l'articulation est "slurred" ») parce que le cerveau l'a dirigée pour l'être. Au milieu du mot /dæd/, les muscles ont été excités par le cerveau de manière à articuler non pas un /æ/ qui serait la moyenne des différents /æ/ qui peuvent être prononcés, mais le /æ/ approprié au contexte /d--d/. Il suggère donc que chaque locuteur possède tout un système de sons vocaliques pour le contexte /d--d/, un autre système pour le contexte /s--p/, etc. Dans chaque cas, il produit l'allophone adéquat en envoyant les commandes d'innervation à l'organe de parole.

Daniloff et Hammarberg (1973) déclarent que la notion de coarticulation présuppose l'existence de cibles canoniques de l'articulation. La coarticulation est issue de l'interaction entre ces segments canoniques par l'intermédiaire d'un mécanisme qui est inconnu. Ce dernier pourrait impliquer un déploiement des traits, c'est-à-dire la diffusion d'un trait inhérent d'un segment canonique à un autre pour lequel ce trait n'est pas inhérent. La coarticulation est mesurée à partir de la forme canonique idéale. La meilleure estimation de cette forme est donnée quand un son de parole est réalisé en isolation ou lorsqu'il est jugé peu coarticulé. Une fois la forme ou la cible canonique mesurée, la coarticulation pourrait être évaluée par l'étendue de la déviation systématique de la cible quand le segment est produit dans un contexte phonétique quelconque.

La variabilité et la non-discrétisation des sons de la parole attestent d'un accord général dans la littérature. La coarticulation est un phénomène universel qui expose une grande partie de cette variabilité, autrement dit les influences systématiques et réciproques entre les segments adjacents et quelques fois non-adjacents. Cela explique pourquoi la coarticulation a occupé une place centrale dans les études phonétiques récentes. Plusieurs théories et modèles de coarticulation ont alors été proposées afin de comprendre ce phénomène. La différence entre une théorie et un modèle est souvent ambiguë et complexe. D'un point de vue épistémologique, une théorie est « une manière de concevoir et de percevoir les faits et d'organiser leur représentation. Elle sert à conceptualiser et à expliquer un ensemble d'observations systématiques relatives à des phénomènes et à des comportements complexes »<sup>10</sup>. Cependant, un modèle est « développé dans le prolongement d'une théorie, dont il est en fait une projection. [...] Le modèle apparaît

9. Comme le mot "coarticulation" n'était pas encore très usité à l'époque, l'auteur a proposé le terme "slur" de la terminologie musicale (liaison en anglais) de façon à désigner le phénomène de la coarticulation. Il explique qu'il ne faut pas le confondre avec le sens habituel de ce mot mais il faut le considérer selon cette définition : « *It means, for example, that the effect of each consonant upon the vowel in a CVC sequence usually extends past the middle of the vowel, so that the two consonantal effects overlap, and it also means every other detail of particular articulations that cannot be fully accounted for by one phone alone* » (Joos 1948, page 108).

10. Willett (1996), page 6.

donc comme une partie concrète de la théorie qui est directement en rapport avec un ensemble de comportements »<sup>11</sup>. Par conséquent, le but des théories de la coarticulation est de déterminer son origine, sa nature et sa fonction alors que l'objectif des modèles de la coarticulation est d'en prédire les détails et de combler le vide entre la représentation des unités discrètes et la réalisation articulatoire et acoustique (Farnetani et Recasens, 1999). Dans la deuxième section de ce chapitre, des approches théoriques de la coarticulation (la propagation des traits, la coproduction et la variabilité adaptative) sont présentées dans l'intention de comprendre le contexte général des faits et modèles de la coarticulation expliqués dans la troisième section. Parmi les différents modèles, certains ont été choisis en fonction de leur pertinence par rapport au sujet développé dans cette recherche. La troisième section décrit les modèles qui ont livré une réflexion sur le traitement des problèmes de la coarticulation et ceux ayant fourni des outils employés pour l'analyse des données acoustiques. Les modèles énoncés sont au nombre de huit : le locus, l'*undershoot*, le modèle d'Öhman, la résistance coarticulatoire, le degré de contrainte articulatoire, la théorie H et H, le modèle *task-dynamics* et les équations de locus.

## 2.2 Approches théoriques de la coarticulation

La parole a souvent été considérée comme une séquence discrète d'unités ou de segments. Ce postulat a été confronté avec le chevauchement complexe des mouvements articulatoires pendant la production de la parole. À ce propos, Lindblom (1982) explique que deux problèmes majeurs se posent à cause de cette complexité de la parole. Il s'agit du problème de segmentation, « [...] *the speech signal cannot be unambiguously segmented into temporally non-overlapping chunks corresponding to e.g. phonemes, syllables and words* »<sup>12</sup> et du problème de l'invariant, « [...] *invariant acoustic correlates on linguistic units cannot be satisfactorily demonstrated* ». <sup>13</sup> Le développement des technologies dans la recherche de la production de la parole (sur les plans acoustique et articulatoire) a remis en question l'interprétation de la notion du segment (discret) dans le flux de la parole (Kent et Minifie, 1977). Pourtant, la perception de la parole exige la définition d'une notion où, à un niveau donné, la production de la parole a besoin d'une chaîne de segments discrets en entrée et en sortie réalisée par des gestes chevauchés et complexes (Bell-Berti et Krakow, 1991). Deux classes d'explications générales ont été fournies dans la littérature visant à une interprétation de la coarticulation : la théorie de **propagation des traits** appelée théorie avec une temporalité extrinsèque de la production de la parole (Joos, 1948; Kozhevnikov et Chistovich, 1965; Henke, 1966; Moll et Daniloff, 1971) et celle de **coproduction** dite théorie avec une temporalité intrinsèque (Fowler, 1980; Bell-

11. Willett (1996), page 9.

12. « [...] le signal de la parole ne peut pas être segmenté sans ambiguïté en portions non chevauchées temporellement, ce qui correspond par exemple à des phonèmes, des syllabes et des mots » (Lindblom 1982, page 9 [notre traduction]).

13. « [...] les corrélats linguistiques invariants dans les unités linguistiques ne peuvent pas être démontrées de façon satisfaisante » (Lindblom, page 9 [notre traduction]).

Berti et Harris, 1981; Browman et Goldstein, 1986; Saltzman et Munhall, 1989). La première approche considère que la dimension temporelle est exclue de la spécification d'un segment phonologique sur le plan articulatoire d'une séquence tandis que la deuxième l'incorpore (Fowler, 1980). D'autres classes d'explications ont également été proposées comme les modèles de **la variabilité adaptative** qui traitent de la coarticulation d'un point de vue communicatif et sociolinguistique (Lindblom, 1983, 1989, 1990).

### 2.2.1 La propagation des traits

La théorie de propagation des traits (en anglais "*feature spreading*") a été avancée dans les articles de Daniloff et Hammarberg (1973) et Hammarberg (1976). La coarticulation est considérée comme un processus physiologique pur provoqué par les contraintes mécano-inertielles de l'appareil phonatoire qui comprend une dichotomie entre l'intention et l'exécution. Les articulateurs sont ainsi incapables d'effectuer les commandes comme ils devraient le faire. L'approche de propagation des traits suggère que la coarticulation est une partie de la composante phonologique qui sert à résoudre la dichotomie existante entre intention et exécution. Ce postulat est soutenu par trois arguments (Farnetani et Recasens, 1999, 2010) : (1) la phonologie est antérieure à la phonétique, autrement dit, la composante phonologique implique une réalisation phonétique des sons de la parole ; (2) les segments phonologiques sont des entités abstraites et ne peuvent pas être changés par le mécanisme physique de la parole ; (3) celui-ci peut uniquement exécuter les commandes d'un niveau plus haut. Les variations associées à la coarticulation font ainsi partie de l'entrée du mécanisme de la parole. Les segments ont des propriétés inhérentes et dérivées. Ces dernières sont dues à la coarticulation, ce qui influence les propriétés d'un segment. Les règles phonologiques précisent quels traits à modifier, la représentation phonétique, qui est l'entrée du mécanisme de la parole, spécifie les détails de l'articulation et de la coarticulation (Farnetani et Recasens, 2010). Les consonnes et les voyelles se distinguent par des traits phonétiques. Par exemple, les consonnes peuvent être décrites par le trait du voisement (voisées ou non voisées), le lieu d'articulation (bilabiales, vélares, etc.) et/ou le mode d'articulation (occlusives, fricatives, etc.). Quant aux voyelles, elles peuvent être définies en fonction de la position horizontale de la langue (antérieures, centrales ou postérieures), par la position verticale (ouvertes, moyennes ou fermées) ou par l'arrondissement des lèvres (arrondies ou non-arrondies). Un changement de trait implique une modification de l'identité de la consonne ou de la voyelle et, par conséquent, l'identité du mot qui les contient. Par exemple, les consonnes /p/ et /b/ se différencient par le trait du voisement : une absence de voisement de /b/ changera le mot "beau" en "peau". Toutefois, certains traits ne sont pas distinctifs. Ajouter un arrondissement des lèvres lors de la production des consonnes ne change pas leur identité, de même qu'ajouter un trait de nasalité sur les voyelles en anglais (Fowler, 2007). Selon les modèles de propagation des traits, les traits non-distinctifs sont propagés dans une direction anticipatoire à n'importe quel son pour lequel ce trait n'est pas spécifié (c'est-à-dire, ce trait n'est

pas distinctif pour ce son). De ce fait, l'arrondissement des lèvres doit être diffusé à n'importe quelle consonne précédant une voyelle arrondie. La nasalisation, elle, doit se propager sur les voyelles précédant une consonne nasale. La coarticulation persévérante est considérée comme une inertie. Les articulateurs ne peuvent pas s'arrêter directement après la réalisation d'une cible, ils effectuent un mouvement passif - conséquence de la réalisation d'une cible - vers la position de repos ou vers la cible suivante. Dès lors, l'arrondissement des lèvres peut continuer pendant un segment qui suit une voyelle arrondie.

Cette approche a été appuyée par plusieurs études expérimentales qui s'opposent à l'idée que la coarticulation est le produit de l'inertie. Daniloff et Moll analysent la coarticulation labiale et vélaire (Daniloff et Moll, 1968; Moll et Daniloff, 1971). Les résultats révèlent que l'abaissement du vélum lors de l'anticipation d'une consonne nasale ou l'arrondissement des lèvres lors de l'anticipation d'une voyelle arrondie peut commencer un, deux, trois ou quatre segments avant celui en question. Ces constatations indiquent clairement que la coarticulation anticipatoire ne peut pas être un effet d'inertie mais une propagation des traits délibérée. De plus, les ajustements spatiaux produits en anticipation des segments subséquents sont considérés comme des processus délibérés de propagation des traits (Daniloff et Hammarberg, 1973). Quand deux segments adjacents sont produits avec le même articulateur, leurs lieux d'articulation ont tendance à être identiques. Par exemple, le lieu d'articulation de l'occlusive alvéolaire voisée /d/, suivie par la fricative dentale non voisée /θ/, devient dental à l'instar du mot "width". Une telle adaptation (ou "accommodation") d'une séquence a pour but de lisser et de minimiser les sons transitionnels qui se produisent entre les segments. Les auteurs ajoutent que ces ajustements spéciaux ne sont pas à considérer comme étant un phénomène inéluctable ou universel. Certains processus articulatoires impliquent une propagation des traits telle que la nasalisation des voyelles devant les nasales, le dévoisement des consonnes, etc. Toutefois, ils ne servent pas à éliminer les transitions entre les segments. Des phénomènes comme ceux-ci sont expliqués en terme d'économie d'effort. Cette dernière désigne la réduction de l'effort des mouvements d'ajustement articulatoire que les organes doivent effectuer. De plus, certaines langues semblent avoir un degré minimal d'*accommodation* entre les sons à l'exemple du russe. Des différences coarticulatoires entre le français et l'anglais ont été trouvées lors de la production des occlusives vélares avec des voyelles antérieures et postérieures (Ladefoged, 1967). Daniloff et Hammarberg proposent que tous les processus de la coarticulation anticipatoire (de droite à gauche) sont représentés dans la grammaire d'une langue par les règles phonologiques de propagation des traits (Daniloff et Hammarberg, 1973; Hammarberg, 1976). D'un autre côté, les mécanismes de la coarticulation persévérante (de gauche à droite) sont décrits par les auteurs comme un effet passif de l'inertie du système articulatoire.

D'après les modèles de propagation des traits, le planning articulatoire, dit aussi "*look-ahead*", détermine quels sont les mouvements exigés pour le segment suivant. Dans ces modèles, l'ampleur temporelle d'une telle adaptation anticipatoire est limitée uniquement par les caractéristiques des autres segments, syllabes ou structures syntaxiques (Bell-Berti et Krakow,

1991). Avec cette limitation temporelle, la coarticulation peut être illimitée au sein d'une syllabe et minimale entre les syllabes. Cette hypothèse a été proposée par Kozhevnikov et Chistovich (1965) dans leur modèle de la syllabe articulatoire. Selon ce modèle, les commandes d'une voyelle sont actionnées simultanément avec celles des consonnes qui précèdent la voyelle. Un degré de coarticulation maximal est ainsi attendu pour les syllabes de type  $C_nV$  et une coarticulation minimale dans les autres types de syllabes. De ce fait, le modèle de Kozhevnikov et Chistovich ne peut pas prédire la coarticulation anticipatoire dans les syllabes de type VC. Contrairement au modèle de la syllabe articulatoire, les données de Moll et Daniloff (1971) montrent que la coarticulation vélaire peut être étendue d'une consonne nasale jusqu'à deux voyelles précédentes. Le modèle de Henke (1966), à l'encontre de celui de Kozhevnikov et Chistovich (1965), n'impose pas de frontière pour la coarticulation anticipatoire. Les segments sont spécifiés pour des cibles articulatoires en termes de traits binaires (+ ou -). Les traits non-spécifiés prennent la valeur de 0. Les règles coarticulatoires assignent un trait d'un son à tous les segments précédents qui sont non-spécifiés par le biais d'un mécanisme "look-ahead". Les traits propagés sont bloqués par un trait non-spécifié. Par exemple, il est attendu que l'abaissement du vélum pour une consonne nasale se propage jusqu'à la première voyelle dans une séquence  $CV_nN$ .<sup>14</sup> La nasalité ne dépasse pas la première voyelle car le vélum sera en conflit lors du début de la consonne orale. Une représentation de ce cas est affichée dans la Table 2.1 (a) où le début de l'abaissement du vélum commence bien plus tôt que la consonne nasale. Le trait de nasalité est diffusé sur toutes les voyelles précédentes. La propagation de ce trait est arrêtée par la présence d'une consonne orale qui est en conflit avec l'abaissement du vélum (il faut un vélum élevé pour produire une orale).

L'approche de propagation des traits a reçu plusieurs critiques notamment par Bell-Berti et Harris (1979, 1981, 1982) et Fowler (1980). Elles se résument par le fait qu'un modèle basé sur la propagation des traits binaires échoue à expliquer la présence vs. absence de la coarticulation, la nature graduelle de celle-ci, les différents degrés de coarticulation exposés par les mêmes segments dans les langues et les changements temporels pendant la réalisation des segments. Le modèle de la résistance coarticulatoire de Bladon et Al-Bamerni (1976) et le modèle de fenêtre ("*the window model*") de Keating (1990) ont permis de surmonter quelques limitations de la théorie de propagation des traits. Le premier modèle suggère que la spécification en termes de traits binaires peut être accompagnée par une spécification graduelle de la résistance coarticulatoire. Dans leur étude acoustique sur la coarticulation des allophones de /l/, Bladon et Al-Bamerni (1976) mettent en évidence que les variations articulatoires décroissent de /l/ clair, /l/ sombre à /l/ sombre et syllabique. Ces différences graduelles ne peuvent pas être représentées par des analyses binaires sur des traits qui bloqueraient la coarticulation pour l'allophone sombre spécifié comme [+ sombre]. Les chercheurs proposent un index numérique de la résistance coarticulatoire assigné à la spécification des traits de chaque allophone. Le mo-

14. Où C est une consonne orale,  $V_n$  correspond à n'importe quel nombre de voyelle et N est une consonne nasale.

**Modèles de prédictions coarticulatoires**

(a) propagation des traits	(b) coproduction
C V <sub>1</sub> N	C V <sub>1</sub> N
↑ ←	→ : ↑ ←
C V <sub>1</sub> V <sub>2</sub> N	C V <sub>1</sub> V <sub>2</sub> N
↑ ←	↑ ←
C V <sub>1</sub> V <sub>2</sub> V <sub>3</sub> N	C V <sub>1</sub> V <sub>2</sub> V <sub>3</sub> N
↑ ←	↑ ←
C V <sub>1</sub> V <sub>2</sub> V <sub>3</sub> V <sub>4</sub> N	C V <sub>1</sub> V <sub>2</sub> V <sub>3</sub> V <sub>4</sub> N
↑ ←	↑ ←

TABLE 2.1 – Prédications des modèles de propagation des traits et de coproduction. (a) Les modèles de propagation des traits prédisent que l’abaissement du vélum dans l’anticipation d’une consonne nasale s’étend jusqu’au début de la séquence vocalique précédant la nasale, indépendamment de la durée de la séquence vocalique ou le nombre de segments. (b) Les modèles de coproduction prédisent que l’abaissement du vélum pendant une séquence vocalique précédant une consonne nasale commence à un temps stable avant la nasale, indépendamment de la durée de la séquence vocalique (d’après Bell-Berti et Krakow (1991)).

dèle de Keating prend en compte les variations continues en espace et en temps observées dans la production de la parole ainsi que les différences inter-segmentales et interlinguistiques de la coarticulation. Keating met en avance que les traits binaires et les règles phonologiques ne peuvent pas expliquer la nature graduelle de la coarticulation et les différences coarticulatoires entre les langues. Ce modèle propose ainsi deux hypothèses qui ont été inspirées de la théorie de propagation des traits : (1) la sous-spécification<sup>15</sup> phonologique peut persister dans la représentation phonétique ; (2) la sous-spécification phonétique est une notion continue et n’est pas catégorielle. La représentation phonologique est exprimée sous la forme de traits binaires. Les traits non-spécifiés peuvent devenir spécifiés par une règle phonologique ou ils peuvent rester non-spécifiés. Les règles phonologiques par lesquelles les segments obtiennent un trait varient en fonction des langues (Farnetani et Recasens, 1999). Le modèle de Keating souligne que chaque unité phonologique délimite le maximum et le minimum de la position permise pour chaque articulateur. Ce dernier est libre de prendre n’importe quelle trajectoire pour un

15. « Underspecification (n.) In feature theories of phonology, a term characterizing various approaches which see it as desirable that information should be omitted from underlying phonological representations. The representations should be minimally specified, or underspecified. There is a departure from the concept of "full" specification present in early generative phonology : the view that the output of the phonological component must contain fully specified binary feature matrices. Underspecification theory is concerned with the extent to which feature distinctions should appear in a phonological representation, not as a binary choice of [+ feature] v. [- feature], but as a choice between [+ feature] and no marking at all. It therefore looks in particular at which feature values are predictable and may thus be left unspecified in a representation without harming the surface form. ». Définition par le dictionnaire *Linguistics and Phonetics* de Crystal (2008).

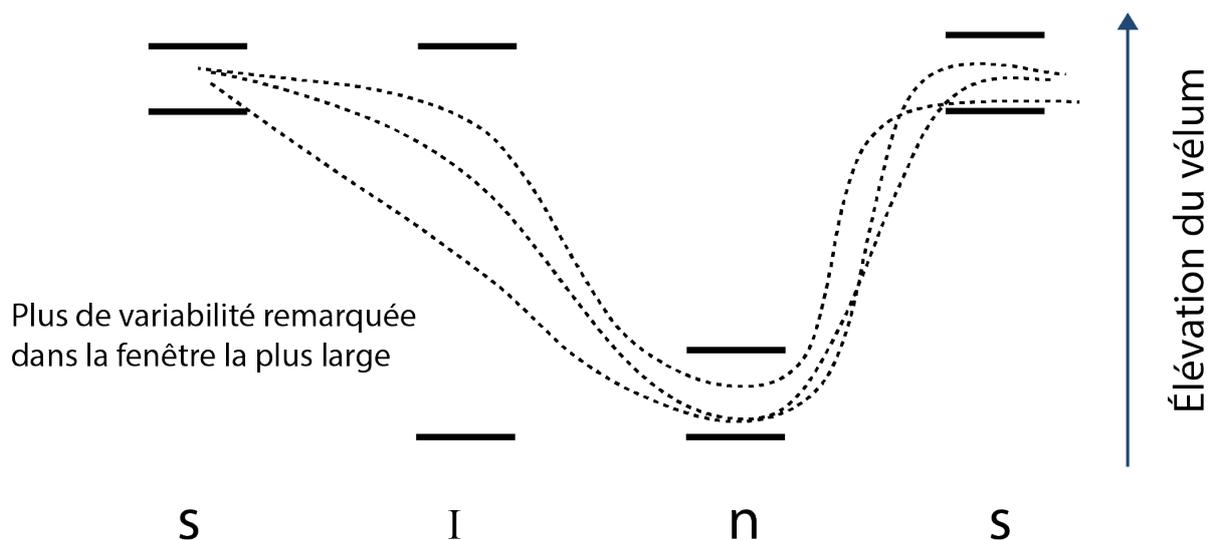


FIGURE 2.3 – Démonstration du modèle de Keating (1990) pour le mot anglais /sɪns/ indiquant la variation du degré d'élévation du vélum pour chaque segment. Les deux lignes horizontales au-dessus de chaque phonème représentent l'étendue du mouvement possible du vélum ou la fenêtre pour le segment. Le /s/ est produit avec un vélum élevé et une fenêtre étroite désignant une variabilité réduite. Le /ɪ/ est réalisé avec un vélum relativement élevé et une fenêtre large montrant une variabilité élevée. Le /n/ est prononcé avec un vélum baissé et une petite fenêtre révélant une variabilité restreinte (d'après Huckvale (2016), inspiré de Keating (1990)).

segment en passant par une fenêtre. Le modèle de fenêtre suppose que la taille de la fenêtre indique l'étendue de la variation de coarticulation possible d'un articulateur pour un segment (fenêtre + grande = + de variabilité) (Figure 2.3).

### 2.2.2 La coproduction

Le deuxième type d'explication générale est la coproduction qui peut être définie par un chevauchement temporel dans la production de deux ou plusieurs sons. Meynadier (2013) explique que « la notion de coproduction n'est pas strictement synonyme de coarticulation, mais est plus générale. Elle renvoie plus précisément au processus biomécanique intrinsèque à l'acte moteur impliquant la production simultanée ou chevauchée de mouvements articulaires coordonnés réalisant les sons enchaînés dans la parole. La coarticulation n'en est en fait qu'une conséquence phonétique tant spatiale (attraction d'une position articulaire sur une autre) que temporelle (synchronisation des gestes) ». D'après les modèles de coproduction, la coarticulation est une fonction du chevauchement des gestes pour lesquels le début de la réalisation garde une relation stable avec les autres aspects articulaires d'un segment (Table 2.1 (b)). Bell-Berti (1980) montre que le départ de l'abaissement du vélum dans le cas d'une consonne nasale maintient une relation temporelle stable avec la constriction dans le conduit oral qui est caractérisée en unité de temps et non pas en nombre de segments. L'arrondissement des lèvres ne doit pas s'engager au début d'une série de consonnes précédant une voyelle arrondie. Une diminution de l'arrondissement est attendue pendant une série de consonnes entre deux voyelles arrondies,

puisque l'arrondissement de la deuxième voyelle commence après la fin de la première (Fowler, 2007). Dans les modèles de coproduction, il n'existe pas de composante "*look-ahead*" qui avertit le plan d'un segment en fonction du contexte. La coarticulation n'affecte pas la nature des gestes, mais elle se manifeste sous la forme de la cooccurrence et de la coordination entre les gestes successifs. Les gestes sont modifiés uniquement pour retarder leurs débuts en raison des exigences des segments précédents. La ligne verticale discontinue dans la Table 2.1 (b) de la première séquence indique le moment où l'abaissement du vélum de la consonne nasale pourrait commencer si son début n'avait pas été retardé par les exigences de la consonne orale initiale. Ce phénomène représente la coarticulation persévérante (Bell-Berti et Krakow, 1991). Bell-Berti et Harris (1981) proposent une explication spécifiée de la coproduction connue comme "*frame theory*". Selon cette théorie, la coarticulation anticipatoire commence à un intervalle fixe avant le début défini acoustiquement d'une voyelle arrondie ou d'une consonne nasale.

Des évidences de la coproduction ont déjà été remarquées dans les observations de la coarticulation voyelle-à-voyelle. Les ajustements articulatoires peuvent se coproduire avec l'articulation des consonnes précédentes et les gestes consonantiques sont surimposés sur un geste de voyelle-à-voyelle dans les séquences VCV (Öhman, 1966). Fowler (1980) interprète la coarticulation comme étant un résultat de la coproduction des consonnes et des voyelles. La parole est ainsi considérée comme une configuration continue des liens de coproduction. Elle critique les modèles qui introduisent la coarticulation comme une propagation des traits. Selon Fowler, la dichotomie entre le niveau de représentation et celui de la production se situe dans la nature différente des entités descriptives appartenant aux deux niveaux. Elle explique que les unités de représentation sont abstraites, statiques et atemporelles dans tous les modèles de propagation des traits. Cette dualité nécessite un processus de traduction qui les transforme en mouvements articulatoires. Dans ce processus de transformation, la planification de la parole fournit les cibles spatiales et un mécanisme central. Celui-ci précise à quel moment l'articulateur doit se déplacer. La théorie de coproduction, développée dans les recherches de Fowler (1980), Bell-Berti et Harris (1981) et Browman et Goldstein (1986, 1989), résout le problème de la dichotomie du plan temporel des segments en modifiant les unités dans le plan phonologique. Ce dernier doit spécifier l'acte qui doit être exécuté étant donné que les unités phonologiques doivent être des actions planifiées. Autrement dit, des gestes phonétiques sont spécifiés de manière dynamique avec une dimension temporelle intrinsèque (Farnetani et Recasens, 1999). La spécification des gestes articulatoires, en termes de paramètres dynamiques, permet la détermination automatique des représentations cinématiques des mouvements de la parole. Les gestes ont leurs propres structures temporelles intrinsèques qui leur tolèrent un chevauchement temporel lors de l'exécution. Le degré de chevauchement des gestes est alors contrôlé au niveau de la planification de la parole. Les modèles de propagation des traits postulent l'existence de règles d'assimilation dans le but d'expliquer la coarticulation. À partir de ces règles, les traits sont échangés ou modifiés avant d'être traduits en cibles articulatoires. Quant aux modèles de coproduction, il est proposé que les gestes ne subissent pas de changement lorsqu'ils sont actualisés pendant la production

de la parole. Leur structure temporelle intrinsèque permet aux gestes de se chevaucher en temps. Les gestes ne sont pas modifiés par le contexte mais coproduits avec le contexte.

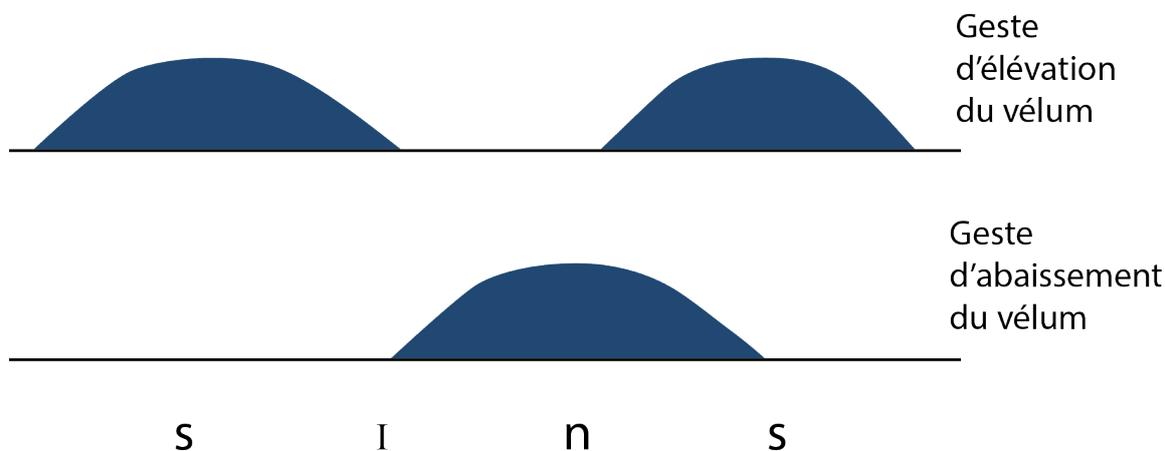


FIGURE 2.4 – Illustration du degré d'activation du vélum selon la théorie de coproduction du mot /sms/ en anglais (d'après Huckvale (2016), inspirée de Fowler et Saltzman (1993)).

Browman et Goldstein (1993) signalent que les nouvelles versions des modèles de propagation des traits (par exemple Keating (1990)) échouent à résoudre le problème de la dualité entre les aspects cognitif et physique. Cet échec est dû à l'introduction des règles d'implémentation phonétique avec les règles phonologiques de propagation. D'après la théorie de propagation des traits développée ci-dessus, la description de la parole est divisée en deux domaines séparés impliquant deux types de représentations distincts : la structure phonologique ou cognitive et la structure phonétique ou physique. Or selon Browman et Goldstein (1993), la relation entre les aspects physique et cognitif de la parole (ou bien phonétique et phonologique) sont intrinsèquement contraints pour être tout simplement deux niveaux de description - microscopique et macroscopique - du même système. Browman et Goldstein (1990b) désignent que la relation entre les propriétés microscopiques et macroscopiques de la parole est une contrainte mutuelle ou réciproque. Ils ajoutent que l'existence d'une telle réciprocity est soutenue par deux lignes différentes de recherche. La première essaye de montrer comment les propriétés macroscopiques du contraste et de la combinaison des unités phonologiques découlent des propriétés microscopiques. Celles-ci peuvent être définies comme étant les propriétés détaillées de l'articulation de la parole et les relations entre l'articulation, l'aérodynamique, l'acoustique et l'audition de la parole. Au contraire, la seconde ligne expose que des contraintes existent qui révèlent que les propriétés acoustiques ou articulatoires détaillées des unités phonologiques sont déterminées, en partie, par le système macroscopique du contraste et de combinaison trouvées dans une langue donnée.

Fowler et Saltzman (1993) énoncent que chaque unité phonologique spécifie un geste articulaire indépendamment du contexte. Le temps d'exécution de ces gestes peut être plus important que la durée du segment, ce qui provoque un chevauchement temporel entre les gestes. Ce chevauchement amène à une réalisation qui est dépendante du contexte (Figure 2.4). L'activa-

tion d'un geste croît ou décroît progressivement dans le temps ; elle effectue ainsi son influence sur la forme du conduit vocal et sur le signal acoustique. La Figure 2.4 met en évidence le chevauchement qu'il peut y avoir entre les segments du mouvement du vélum lors de la production des 4 phones en anglais. Les sommets de chaque geste représentent une activation maximale du vélum. Le chevauchement entre les gestes explique la coarticulation anticipatoire (l'influence de droite à gauche) et persévérante (l'impact de gauche à droite). Fowler et Saltzman (1993) indiquent que les variations provoquées par la coproduction dépendent de l'étendue partagée par les gestes et les articulateurs (le degré de chevauchement spatial). Ils soulignent que les gestes sont mis en œuvre dans la parole par des structures coordinatives. Celles-ci sont les dépendances fonctionnelles transitoires entre les articulateurs qui contribuent à un geste. Par exemple, la production d'une occlusive bilabiale implique le lien fonctionnel entre la lèvre supérieure, la lèvre inférieure et la mâchoire. Une diminution de la contribution d'un articulateur au geste de la fermeture labiale est compensée alors par une augmentation de la contribution d'un autre articulateur. Par conséquent, une cible gestuelle est accomplie grâce à la coordination entre les articulateurs issues des manœuvres compensatoires (Saltzman et Munhall, 1989). Fowler et Saltzman (1993) mettent en relief que le degré de coarticulation ou la variabilité coarticulatoire dépend de l'étendue où les gestes temporellement chevauchés partagent les articulateurs. Lorsque des gestes subséquents ont un articulateur en commun, comme la mâchoire dans les séquences /VbV/, le chevauchement temporel entre les gestes vocalique et consonantique se produit avec des perturbations spatiales minimales. Une perturbation spatiale élevée est attendue lorsque deux gestes chevauchés partagent l'articulateur engagé directement à la production de la cible et imposent des demandes opposées. Browman et Goldstein (1989); Saltzman et Munhall (1989) mettent en exergue que le chevauchement spatial se produit lorsque deux ou plusieurs gestes sont coproduits partageant quelques unes ou toutes leurs composantes coarticulatoires. Dans ce cas, les influences sur l'articulateur en commun des gestes spatialement et temporellement chevauchés sont dites mélangées ("*blended*"). Ce mélange gestuel conduit à des résultats variés qui dépendent de la force du mélange ("*blending strength*") associée aux gestes chevauchés. Les gestes "forts" ont tendance à annuler l'influence des gestes "faibles". Une similarité des forces entraîne une influence réciproque des deux gestes. Cette interprétation inclut que les résultats du mélange varient en fonction des gestes impliqués et dépendent en grande partie du degré de résistance coarticulatoire associé au geste en question.

Recasens (1984a,b, 1985, 1989) suggère que le facteur majeur qui affecte la magnitude et l'étendue temporelle des effets coarticulatoires d'un segment phonétique sur un autre est la compatibilité mutuelle des gestes qui constituent les segments. Chaque segment, soit consonne soit voyelle, a une étendue de résistance au chevauchement coarticulatoire différente. La magnitude de la résistance est déterminée par l'incompatibilité mutuelle des segments adjacents avec leurs gestes. Les études de Recasens révèlent que la source principale de la résistance coarticulatoire est le niveau de contrainte que les gestes d'une consonne ou d'une voyelle effectuent sur le dos de la langue. Recasens et al. (1997) proposent *le modèle de degré de contrainte articulaire* ou

le modèle de DAC afin de traiter la complexité de la coarticulation linguale et de ses effets sur la production de la parole. Selon ce modèle, le degré, l'étendue temporelle et la direction de la coarticulation linguale sont conditionnés par la sévérité des exigences imposées sur la langue pour la production des voyelles et des consonnes. Ces deux dernières sont assignées des valeurs de DAC spécifiques. Les voyelles antérieures sont plus contraintes<sup>16</sup> que celles postérieures. Ce phénomène est dû aux propriétés biomécaniques engagées dans le déplacement du dos de la langue horizontalement et verticalement. La voyelle la moins contrainte est le schwa parce qu'il n'a pas de cible articulatoire claire et que le déplacement du dos de la langue est minimal. Les consonnes exigeant un engagement important du dos de la langue et une précision articulatoire élevée ont des valeurs de DAC supérieures. Les labiales sont les consonnes qui ont les valeurs de DAC les plus basses puisque le corps de la langue n'est pas impliqué lors de leurs réalisations. Le modèle de DAC a confirmé la notion de la force des gestes de Fowler et Saltzman (1993) selon laquelle les gestes sont spécifiés par des demandes concurrentes qui prédominent ou subissent l'influence des autres gestes.

### Les modèles hybrides

Il est clair que les deux types de modèles explorés plus haut, la propagation des traits et la coproduction, proposent deux interprétations fondamentalement différentes sur la nature et l'organisation du contrôle moteur de la parole. Selon la théorie de coproduction, les gestes articulatoires ont leurs propres durées intrinsèques. De ce fait, l'étendue temporelle de la coarticulation anticipatoire d'un geste doit être constante ou "*time-locked*" (Bell-Berti, 1980; Bell-Berti et Harris, 1981; Bell-Berti et Krakow, 1991). Toutefois, plusieurs études ont soutenu le modèle de *look-ahead* (Benguerel et Cowan, 1974; Sussman et Westbury, 1981; Lubker, 1981) et montrent que le début de l'arrondissement des lèvres ou l'abaissement anticipatoire du vélum n'est pas fixe mais s'étend en fonction du nombre des segments neutres qui précèdent le segment influant. Dans une étude trans-linguistique sur la coarticulation anticipatoire des lèvres en suédois et en anglais, Lubker et Gay (1982) observent que les locuteurs suédois ont un arrondissement plus intense, plus long et moins variable que les locuteurs américains. De plus, l'arrondissement chez les locuteurs suédois commence plus tôt que chez les locuteurs américains. Ils expliquent que ces différences sont provoquées par des exigences de perception en suédois du fait que la protrusion soit contrastive. Les résultats montrent également que trois (sur cinq) locuteurs suédois utilisent la stratégie "*look-ahead*" et les deux autres emploient celle de la coproduction. Ils concluent que la coarticulation anticipatoire dépend de la langue et de la stratégie que les locuteurs adaptent pendant la production de la parole. Une étude sur la coarticulation du vélum en français de Benguerel et al. (1977) dévoile que les voyelles ouvertes /a, ɑ/ produites dans

16. La notion de *contrainte* a été introduite dans la littérature pour expliquer les degrés de résistance coarticulatoire qui sont liés aux demandes situées dans le conduit vocal. Selon Sussman et Westbury (1981), le terme "contrainte" reflète la résistance coarticulatoire provoquée par deux facteurs possibles : (a) les exigences aérodynamiques du segment ; (b) l'antagonisme des deux gestes, comme l'opposition : arrondissement vs. non-arrondissement des lèvres (Farnetani et Recasens, 1999).

des contextes non-nasaux attestent un abaissement clair du vélum. Les chercheurs constatent que les voyelles orales peuvent tolérer un certain degré d'ouverture vélopharyngée (avec un maximum  $S_{max}$ ). Ils exposent que si l'abaissement du vélum commence "trop tôt",  $S_{max}$  sera dépassé avant le début de la consonne nasale. Une voyelle nasalisée (phonémiquement) est ainsi produite au lieu d'une orale. Deux types de mouvement vélaire sont alors à distinguer dans les séquences des consonnes nasales prononcées avec des voyelles orales : (a) l'abaissement lent du vélum associé à la voyelle commence pendant la réalisation de celle-ci, (b) il est ensuite suivi d'un abaissement plus rapide du vélum pour la nasale qui conduit à une ouverture de la porte vélaire. Ces résultats ont également été remarqués par Al-Bamerni et Bladon (1982) sur des données en anglais. Les auteurs observent que les locuteurs utilisent deux stratégies de production : un geste d'ouverture vélaire à une seule phase et un geste à deux phases. Ce dernier commence avec le début de la première voyelle non-nasale (abaissement lent du vélum) et s'en suit une phase avec une vélocité importante coordonnée au geste de la fermeture orale de la consonne nasale. Sur la même lignée, Perkell et Cohen trouvent que deux phases existent lors de l'arrondissement des lèvres dans des séquences  $iC_nu$  (Perkell et Cohen, 1986; Perkell, 1990). Ils constatent que la formation de la protrusion des lèvres se fait en deux phases. La première, relativement lente, commence aussi tôt que possible. Une formation graduelle d'un arrondissement est remarquée dès la fin de la voyelle /i/. La seconde, plus rapide, débute à un instant fixe avant la cible (Figure 2.5). Les auteurs interprètent ces mouvements à deux phases comme

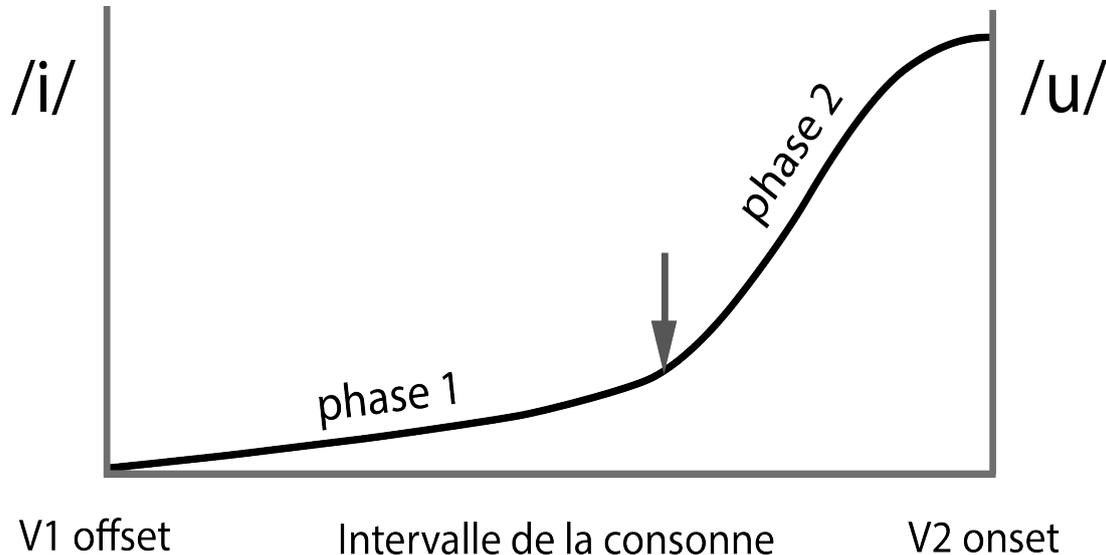


FIGURE 2.5 – Représentation du modèle hybride proposé par Perkell et Cohen (1986); Perkell (1990) montrant les deux phases de l'arrondissement des lèvres dans des séquences  $iC_nu$  (d'après Farnetani et Recasens (2010)).

une stratégie mixte de coarticulation et proposent le modèle hybride comme un modèle de coarticulation anticipatoire. Celui-ci représente un compromis entre les modèles "look-ahead" et "time-locked". Selon le modèle hybride, le début du mouvement de l'arrondissement des lèvres décrit une stratégie "look-ahead" (1<sup>re</sup> phase). L'augmentation rapide de la protrusion, avec un

intervalle fixe avant la voyelle arrondie, reflète alors une stratégie "*locked-time*" (2<sup>e</sup> phase).

Le modèle hybride a été critiqué par de nombreuses études, notamment Boyce et al. (1990); Perkell et Matthies (1992). Ces chercheurs remarquent que le début de la coarticulation anticipatoire est due au geste de la protrusion des lèvres spécifique à la consonne. Ce phénomène a été repéré dans des séquences de type /iC<sub>n</sub>i/. Selon ces résultats, l'argument de l'existence des deux phases d'arrondissement dans la coarticulation anticipatoire du modèle hybride n'est donc pas valable. Toutefois, Bell-Berti et Krakow (1991) mettent en avant que l'abaissement du vélum peut avoir une ou deux phases comme l'ont indiqué Al-Bamerni et Bladon (1982). Bell-Berti et Krakow ont pu prédire dans quel contexte l'abaissement est réalisé avec une ou deux phases. Une seule phase d'abaissement se produit quand la séquence contenant la voyelle a une durée réduite. *A contrario*, il existe deux phases lorsque la séquence de la voyelle est longue. Au sein d'une séquence, un nombre restreint de voyelles (une ou deux) ou un débit rapide de la parole du locuteur entraînent la réduction de l'intervalle de la séquence.

Les deux approches théoriques de la coarticulation présentées ci-dessus ont pour objectif d'expliquer et de traiter la coarticulation au niveau articulatoire/acoustique. Une autre approche, abondamment citée dans la littérature, considère la coarticulation et ses variations d'un point de vue communicatif et sociolinguistique. Elle est appelée la variabilité adaptative et sera exposée dans la section suivante.

### 2.2.3 La variabilité adaptative

La théorie de la variabilité adaptative a été proposée par Lindblom (1983, 1989, 1990) afin d'expliquer les principes utilisés dans la production de la parole. Cette théorie prend en compte les aspects biologique, sociologique et communicatif de la parole. Dès lors, les contraintes de la production, de l'audition et de l'apprentissage de la parole interagissent d'une manière complexe pour définir les sons de la langue. La variabilité adaptative indique que la variation phonétique n'est pas une simple conséquence biomécanique dans le mécanisme de la production de la parole, mais elle est une adaptation continue de la production de la parole aux exigences des situations communicatives. La variation a lieu en raison du changement des stratégies de production de la parole. Il existe des situations où la production de parole doit être réalisée avec un grand degré de contraste perceptuel ; d'autres en nécessitent moins et permettent plus de variabilité. Par conséquent, les caractéristiques acoustiques d'un même segment montrent un large éventail de variations reflété tout au long d'un *continuum* de la sur- à la sous-articulation ou hyper- hypo-articulation (Farnetani et Recasens, 2010). Lindblom et Lindgren (1985) constatent que quand les locuteurs changent de style de parole (claire, neutre et conversationnelle informelle), des variations importantes sont remarquées dans le VOT, dans la durée de la voyelle et dans les fréquences des formants consonantiques et vocaliques. Certains principes de cette théorie sont inspirés de l'étude de la réduction vocalique de Lindblom (1963a). Dans cette étude, la variation temporelle est le facteur principal qui modifie les informations spectrales ou

bien qui *fait perdre des informations spectrales* à un segment. Un phone produit avec un temps réduit (en débit rapide) contient moins d'informations que quand il est réalisé avec un temps "normal" (débit moyen). Ces résultats n'ont pas pu être confirmés par d'autres recherches (Gay (1978a); O'Shaughnessy (1986), entre autres)<sup>17</sup>. Ces chercheurs indiquent que les changements spectraux (notamment la vélocité des transitions formantiques) sont provoqués par l'effet temporel mais d'autres facteurs jouent aussi un rôle très important comme l'accent tonique ou la tension musculaire. Il semble que les locuteurs réorganisent leurs gestes articulatoires afin de s'adapter au manque temporel lors de la production de la parole en débit rapide. *La perte d'information spectrale* d'un segment est alors diminuée ou minimale. Des études approfondies ont été menées pour examiner les facteurs pertinents occasionnant une réduction vocalique ou des modifications de l'information spectrale d'un phone (Lindblom et Lindgren, 1985; Krull, 1989; Lindblom et al., 1992). Ces études montrent que le changement du style de parole (parole formelle, informelle, lue, langage enfantin, etc.) peut faire subir de fortes transformations physiques aux mouvements articulatoires. L'hypothèse de Lindblom est que le changement du style de parole vers la parole spontanée ("*hypo-speech*") conduit à un degré de coarticulation maximal. Elle a été vérifiée par plusieurs études (Krull, 1987, 1989; Duez, 1992). Krull (1987, 1989) utilise *l'équation de locus* comme un outil de mesure du degré de coarticulation qui est engendré par le changement des différents styles de parole. Il s'agit de la relation linéaire entre le début et le milieu du deuxième formant (cf. chapitre 3 Les équations de locus). Les paramètres (la pente et l'ordonnée à l'origine) de cette équation permettent de quantifier la coarticulation. L'équation de locus a été proposée à l'origine par Lindblom (1963b) afin de mesurer le degré de la réduction vocalique. Elle est considérée comme une manière efficace servant à calculer le degré du chevauchement d'une consonne avec un ensemble de voyelles (Krull, 1987, 1989; Sussman et al., 1991; Fowler, 1994). Krull (1987, 1989) mais aussi Duez (1992) trouvent que le style de parole spontané contient les segments subissant le degré de coarticulation le plus important en comparaison avec de la parole lue ou des mots isolés. Ces résultats sont dans la continuité de la théorie de la variabilité adaptative ; puisqu'en parole spontanée, l'économie des gestes articulatoire est maximale.

Lindblom (1983, 1989, 1990) explique que le mécanisme de la parole, comme tout autre mécanisme biologique, a tendance à économiser l'effort. De ce fait, l'information acoustique d'un segment, au lieu d'être stable, est le résultat de l'interaction entre les nécessités perceptuelles de l'auditeur et les exigences de l'économie de la parole du locuteur. Une tendance générale d'un comportement moteur "*low-cost*" domine lors du déplacement des mouvements articulatoires pour produire des segments phonétiques. Cette tendance représente le côté économique de la parole. Lindblom et al. (1975) constatent que - par le biais d'un modèle paramétrique (Lindblom et Sunberg, 1971) - la coarticulation de la voyelle vers la consonne résulte d'une stratégie de production "*low-cost*". Ils analysent des séquences VCV symétriques avec des consonnes occlusives dentales ou rétroflexes et les voyelles /i/ ou /u/. Cette étude explore l'interaction

---

17. Voir section 2.3.2

réci-proque entre les paramètres de l'apex et du corps de la langue quand le contexte vocalique change. Des valeurs formantiques ont été simulées à partir des paramètres articulatoires. Dans chaque combinaison VCV, les valeurs correspondant à l'ouverture de la mâchoire et au lieu d'articulation ont été maintenues constantes. L'élévation et la rétraction du corps de la langue ont été variées avec des seuils maximaux fixes. Les résultats de ce modèle ont été comparés aux caractéristiques acoustiques des séquences VCV naturelles. Les valeurs formantiques artificielles ont été ajustées jusqu'à atteindre la meilleure combinaison possible. Cette dernière est atteinte lorsque la configuration de la langue est compatible avec une fermeture apicale, caractérisée par un déplacement minimal des voyelles adjacentes. Autrement dit, une cible apicale est réalisée en fonction d'un nombre de configurations données du corps de la langue et celui-ci a tendance à toujours effectuer les mouvements les plus petits (Lindblom, 1983; Farnetani et Recasens, 1999).

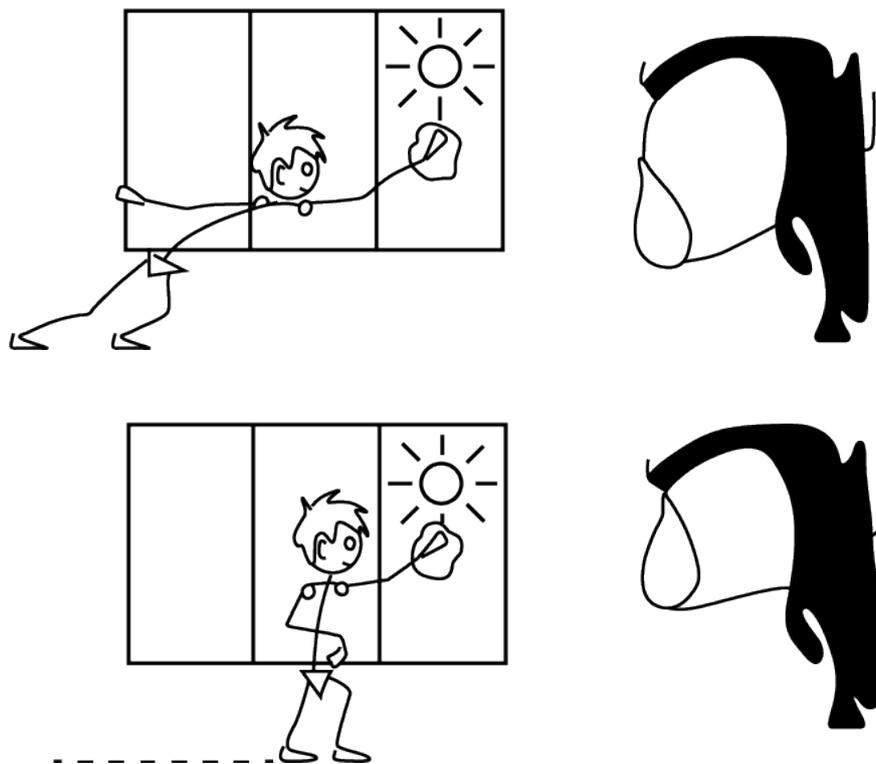


FIGURE 2.6 – Analogie entre des gestes articulatoire et non-articulatoire. En haut de la Figure, une représentation du comportement compensatoire ; en bas, démonstration du comportement normal (d'après à Lindblom (1983)).

Lindblom (1983) illustre l'économie des gestes articulatoires en la comparant à une situation de la vie quotidienne. Un laveur de vitre peut effectuer son travail d'une manière naturelle ou non-naturelle (Figure 2.6). Les deux approches du laveur de vitres peuvent être mises en parallèle avec deux manières analogues pour produire la voyelle /i/ : l'abaissement de la mâchoire synchrones au geste compensatoire d'élévation de la langue (en haut de la Figure 2.6) ou l'élévation de la langue et de la mâchoire dans une interaction synergique (en bas de la Figure 2.6). Cet exemple désigne que l'explication de la coarticulation ne peut pas être spécifique

à la parole puisque la coordination des articulateurs est parallèle à l'organisation des gestes non-articulatoires. L'économie de la production de la parole est permise tant que la situation communicative et la perception de l'auditeur l'autorisent. Il ajoute que plusieurs similitudes ont été trouvées en comparant les mouvements de la parole avec les autres types de mouvements physiologiques (par exemple, l'écriture à la main, les gestes du langage des signes et jouer du piano). Les mouvements physiologiques et ceux de la parole suivent une règle générale où les positions extrêmes sont évitées autant que possible. Ce phénomène reflète le fait que la dépense de l'énergie a tendance à être minimisée.

Les trois approches théoriques expliquées plus haut construisent les bases générales à partir desquelles découlent plusieurs modèles de coarticulation. Ces derniers ont pour objectif de prédire les processus du phénomène de la coarticulation et d'étudier la relation articulaire et acoustique entre les différents sons de la parole. Ils présentent alors une application directe des théories de la coarticulation.

## 2.3 Faits et modèles de la coarticulation

Les modèles de coarticulation représentent un ensemble de propositions, relatives aux observations et aux mesures réalisées sur certains aspects coarticulatoires, qui cherchent à prédire les mécanismes de la coarticulation. Ces modèles sont développés dans le prolongement des approches théoriques abordées précédemment, jouant ainsi un rôle important dans leur formulation. En effet, les modèles de coarticulation ont pour but d'offrir une analyse plus approfondie de ce phénomène par l'intermédiaire d'expériences sur la production et la perception de la parole. Par conséquent, ils permettent d'obtenir une interprétation cohérente et rationnelle des caractéristiques fondamentales de la coarticulation. Les sections suivantes présentent les modèles de coarticulation qui nous ont inspiré dans notre recherche et qui montrent le développement de la réflexion sur la notion de coarticulation.

### 2.3.1 Le locus

La notion du locus formantique a été proposée par Delattre et al. (1955) comme une solution au *problème de l'invariant*. Grâce aux analyses spectrographiques, il a été démontré que le lieu d'articulation d'une occlusive est perçu non seulement à l'aide du bruit de relâchement mais aussi, et parfois essentiellement, grâce aux changements rapides de la fréquence des formants vocaliques. Ces changements sont appelés *transitions formantiques* et ils reflètent le mouvement de l'ouverture vers la position de la voyelle suivante (dans une séquence CV). Le bruit de relâchement, quant à lui, représente l'explosion causée par le relâchement de la pression de l'air à un point de restriction donné. Selon Delattre et ses collègues, puisque le point d'articulation pour chaque consonne est en général fixe, il serait envisageable de trouver une position de fréquence fixe - ou bien *un locus* - qui décrit cette stabilité de lieu d'articulation. Peterfalvi (1966)

définit le locus comme étant « le point d'origine d'un formant dans l'échelle des fréquences. Ce point peut être le point de départ réel ou un point virtuel résultant de la convergence des différentes transitions conduisant à l'identification d'un même phonème. On constate alors que le locus propre à une consonne donnée reste le même quelle que soit la voyelle suivante »<sup>18</sup>. Menées aux *Laboratoires Haskins* et présentées dans Cooper et al. (1952); Liberman et al. (1954), des expériences révèlent que les différentes occlusives ayant le même lieu d'articulation ont des transitions similaires au deuxième formant indépendamment du contexte vocalique. Delattre et ses collègues ont remarqué qu'il est possible de synthétiser les occlusives /p, b, m/, qui partagent le même lieu d'articulation (les lèvres), en gardant les mêmes transitions formantiques du deuxième formant. Les transitions des premiers formants indiquent le mode d'articulation et celles des troisièmes ont été utilisées pour améliorer la qualité des occlusives. Les formants sont dessinés à la main et transformés en sons au moyen du *Pattern Playback* dans le but d'effectuer des tests de perception. Ces chercheurs ont fait produire les occlusives /t, d, n/, qui ont en commun le même lieu d'articulation (la zone dentale), à partir de la même transition du deuxième formant. De manière similaire, la transition du deuxième formant est la même pour les occlusives /k, g, ŋ/. Ces dernières partagent le même lieu d'articulation (le vélum). Par conséquent, les différentes transitions formantiques du deuxième formant peuvent être décrites comme un mouvement allant du locus acoustique jusqu'à l'état stable de la voyelle.

Delattre et al. (1955) ont étudié les trois occlusives voisées /b, d, g/ et ont trouvé que la fréquence du locus du /b/ est située entre 600 et 800 Hz et le /d/ a un locus de 1800 Hz. Quant au /g/, le locus varie en fonction de l'antériorité vs. postériorité de la voyelle. Le /g/ produit avec les voyelles antérieures a un locus de 3000 Hz, alors que le /g/ prononcé avec les voyelles postérieures a un locus de 1300 Hz (Figure 2.7). Les valeurs de locus sont basées sur des expériences avec deux formants simplifiés (F1 et F2). Le troisième formant (F3) est ajouté pour améliorer la qualité du son synthétique. Le locus de F2 d'une consonne reste le même quelle que soit la voyelle. Les auteurs ont fixé la position du locus virtuel à 50 ms avant le relâchement de la consonne dans leurs signaux de parole synthétique. Selon cette expérience, deux conditions sont donc nécessaires pour entendre une de ces occlusives en parole synthétique : les transitions formantiques doivent se diriger vers le locus de l'occlusive et un silence de 50 ms doit exister entre le début des transitions (ou le relâchement) et le locus virtuel. Le locus est ainsi calculé par l'extrapolation à 50 ms avant l'apparition du formant. Les chercheurs indiquent que lorsque les transitions du deuxième formant commencent à 1800 Hz et que le locus est égal au début des transitions, les auditeurs entendent des /bi, da, gu/ au lieu de /di, da, du/. L'occlusive est identifiée comme un /d/ à chaque fois uniquement quand le locus est virtuel (interpolation à 50 ms avant le début des transitions formantiques) (Figure 2.8).

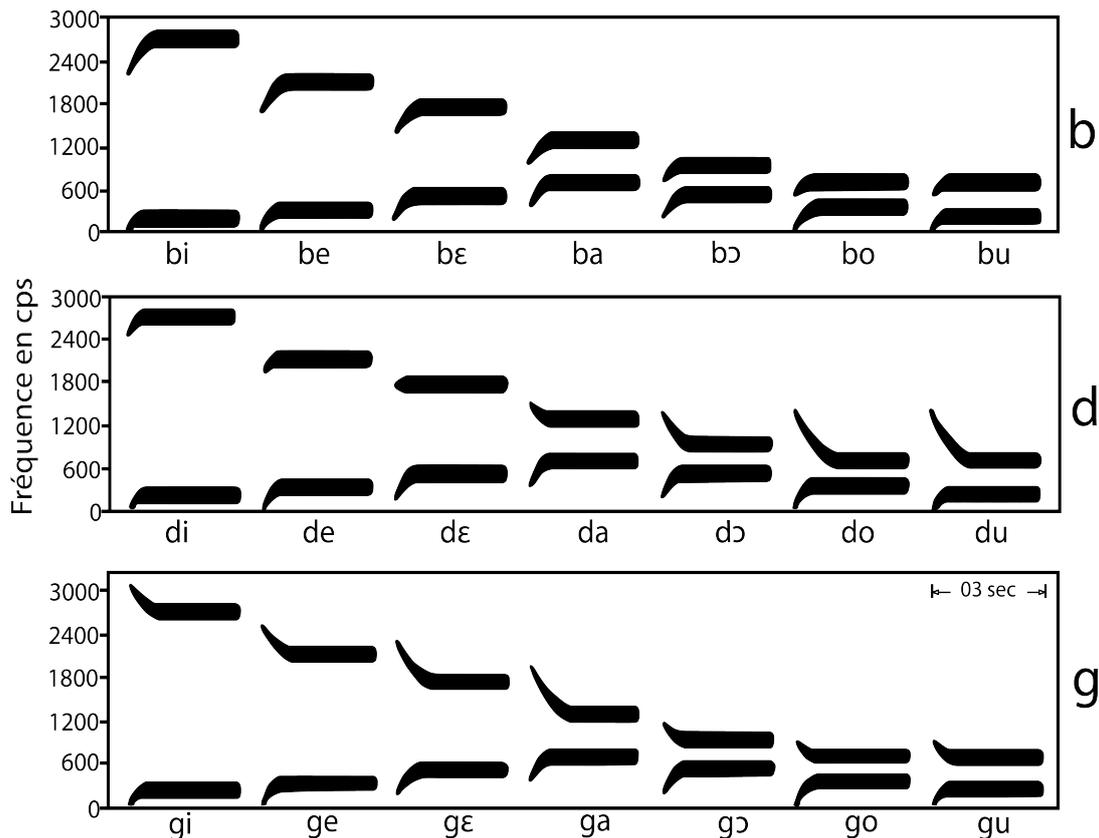


FIGURE 2.7 – Spectrogramme synthétique des deux premiers formants et les transits des consonnes /bdg/ en plusieurs contextes vocaliques (d'après Delattre et al. (1955)).

Stevens et House (1956) ont travaillé sur la validité de "l'hypothèse du locus" à l'aide d'un analogue électrique du conduit vocal<sup>19</sup>. Ils ont obtenu des résultats similaires à ceux rapportés dans les articles de Cooper et al. (1952); Delattre et al. (1955); Liberman et al. (1954) avec quelques petites nuances. Ils ont remarqué que les transitions du premier formant révèlent le mode d'articulation. Le locus du premier formant des occlusives est égal à zéro. Celui du deuxième formant des bilabiales n'est jamais supérieur à la fréquence de F2 de la voyelle adjacente et elle varie entre 700 et 1500 Hz. Les alvéolaires ont un locus qui est autour de 2000 Hz. Les vélaires possèdent un locus généralement supérieur à la fréquence du deuxième formant qui fluctue entre 600 et 2500 Hz.

La théorie du locus a été critiquée notamment par Öhman (1966) et Kewley-Port (1982) qui montrent que le contexte vocalique influence la direction des transitions formantiques et par conséquent le locus, plus précisément dans des séquences VCV et dans la parole naturelle (cf. ci-dessous sous-section 2.3.3). Delattre (1969) construit deux expériences sur des locuteurs du français afin de répondre aux critiques soulignées dans l'étude d'Öhman (1966). Il propose de redéfinir le terme du locus comme étant « *the frequency toward which formant transitions must*

19. Un analogue électrique du conduit vocal est un synthétiseur de parole utilisé pour obtenir des relations expérimentales entre certains paramètres articulatoires idéalisés et les fréquences des formants (Stevens et al., 1953).

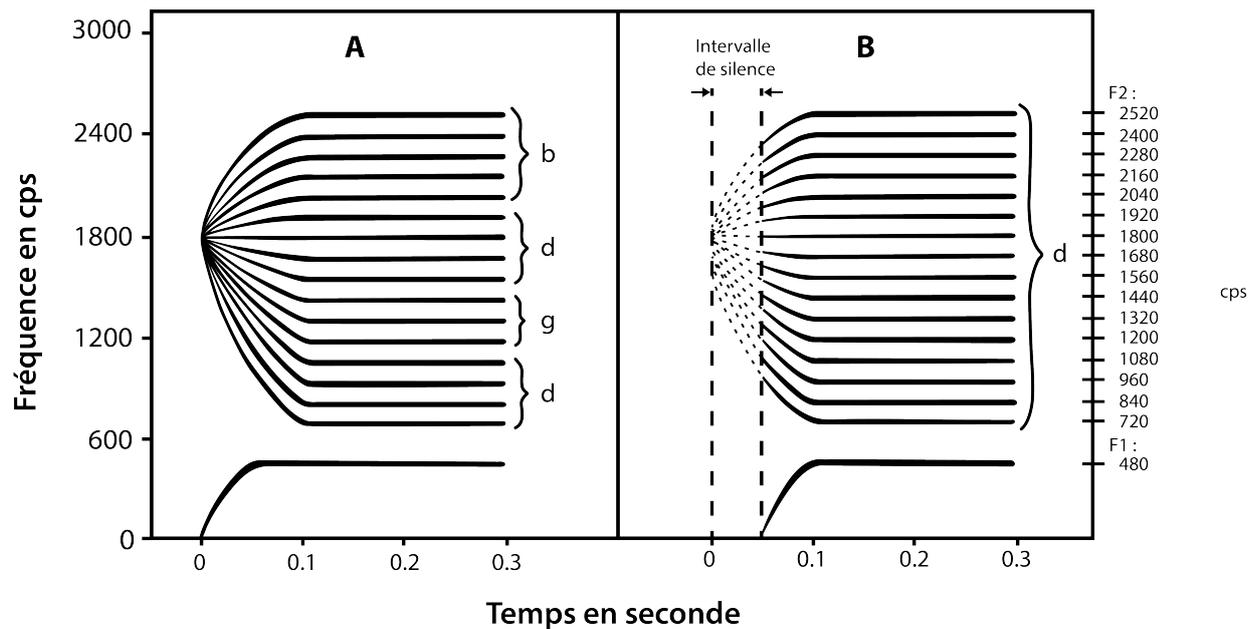


FIGURE 2.8 – Illustration des deux situations confrontées dans la théorie du locus. Dans A, les transitions formantiques commencent à 1800 Hz correspondant au locus de la consonne /d/. Dans ce cas, les locuteurs entendent /b/, /d/ ou /g/ en fonction de la voyelle. Dans B, les transitions se dirigent vers le locus virtuel situé à 50 ms avant leurs débuts. Les auditeurs entendent un /d/ dans tous les contextes vocaliques (d'après à Delattre et al. (1955)).

*point in order to contribute maximally to the perception of a given place of articulation* »<sup>20</sup>. Selon lui, cette définition implique que les transitions formantiques qui ne se dirigent pas (ou se dirigent légèrement) vers un locus donné ne contribuent pas (ou contribuent peu) à la perception du lieu d'articulation cible dans la parole naturelle. Autrement dit, les transitions formantiques du deuxième formant doivent se diriger vers le locus afin que celui-ci participe efficacement à la perception du lieu d'articulation. Si ce n'était pas le cas, le locus ne peut pas (tout seul) donner l'information nécessaire pour le lieu d'articulation cible. C'est donc grâce à d'autres traits (en plus du locus) que le lieu d'articulation d'un son est perçu. La première expérience de perception est basée sur de la parole synthétique. Les formants, les transitions formantiques et le bruit de relâchement des séquences synthétiques simulent des réalisations naturelles des séquences gV (V = /i, e, ε, a, ʌ, ɤ, ʊ, y, ø, œ, ɔ, o, u/). Il s'agit de manipuler ces séquences afin d'examiner les effets mutuels du bruit de relâchement et des transitions formantiques quand ils sont associés ou séparés. Cette expérience permet de montrer les changements auditifs provoqués par l'abstraction du bruit de relâchement des séquences dans lesquelles les occlusives vélaires sont prononcées avec des voyelles arrondies et non-arrondies<sup>21</sup>. Il remarque que les auditeurs réussissent à identifier la vélaire /g/ produite avec les voyelles non-arrondies dans

20. « la fréquence vers laquelle les transitions formantiques **devraient** se diriger pour contribuer *au maximum* à la perception d'un lieu d'articulation. » Page 13, [notre traduction]. Il est à souligner l'importance du terme "au maximum" dans cette phrase.

21. Delattre (1969) sépare deux groupes de voyelles en fonction de l'arrondissement (arrondie/non-arrondie) car il a trouvé que le comportement des transitions formantiques des occlusives vélaires en contexte de voyelles antérieures arrondies est similaire à celles produites avec des voyelles postérieures arrondies.

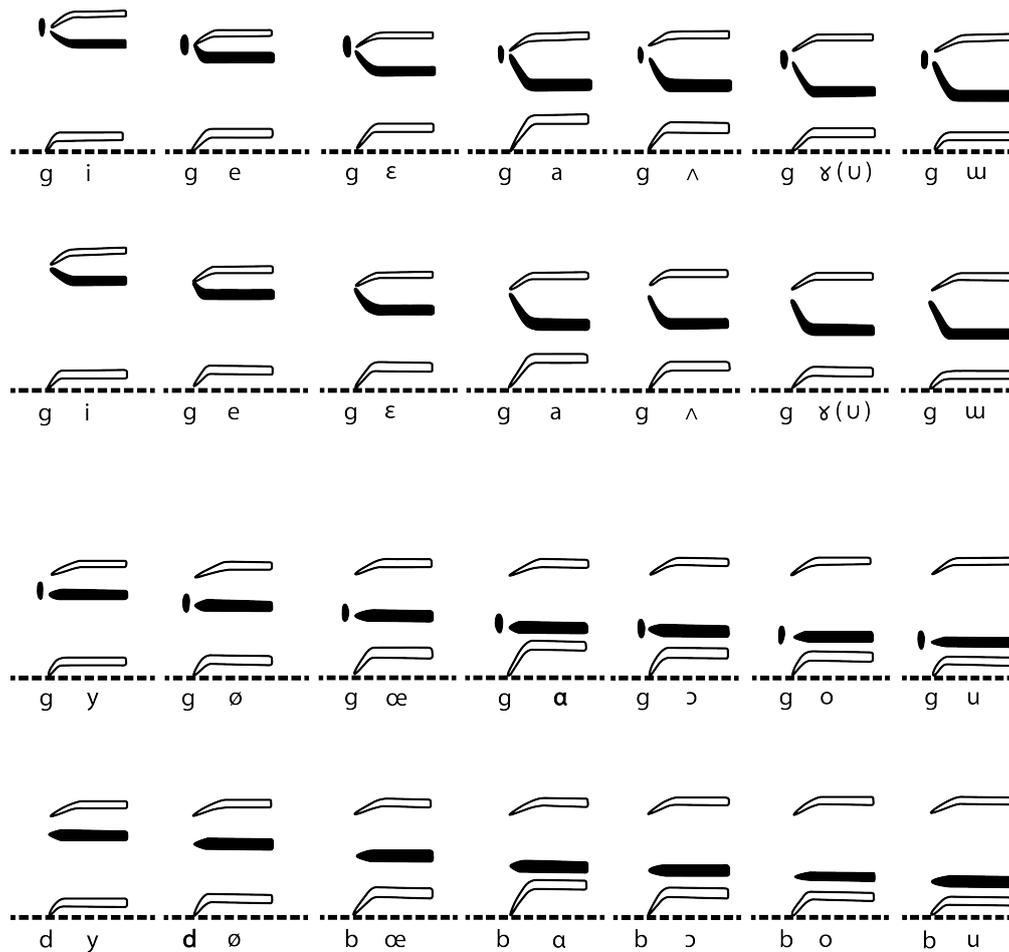


FIGURE 2.9 – Illustration des spectrogrammes synthétiques montrant les transitions formantiques avec le bruit de relâchement (1<sup>re</sup> et 3<sup>e</sup> lignes) et sans le bruit de relâchement (2<sup>e</sup> et 4<sup>e</sup> lignes) de la vélaire /g/ produite avec des voyelles non-arrondies (1<sup>re</sup> et 2<sup>e</sup> lignes) et des voyelles arrondies (3<sup>e</sup> et 4<sup>e</sup> lignes). **Les réponses des auditeurs sont transcrites en bas de chaque séquence** (d'après à Delattre (1969)).

des séquences CV lorsque les transitions du deuxième formant s'orientent vers le locus de la vélaire avec ou sans le bruit de relâchement (première et deuxième lignes dans la Figure 2.9). Par contre, les auditeurs perçoivent le /g/ réalisé avec les voyelles arrondies uniquement en présence du bruit de relâchement (troisième ligne dans la Figure 2.9). En l'absence de ce dernier, les auditeurs entendent une alvéolaire /d/ ou une bilabiale /b/ en fonction du locus le plus proche vers lequel les transitions s'orientent (1800 Hz pour les alvéolaires et 700 Hz pour les bilabiales) (quatrième ligne dans la Figure 2.9). Cette expérience révèle l'importance du bruit de relâchement et les limites du locus dans la perception du lieu d'articulation.

La seconde expérience a pour but de vérifier les résultats de la première (sur de la parole synthétique) avec de la parole naturelle. Le processus du dévoilement progressif ("*gating*") a été utilisé sur des séquences VCV. La vélaire /g/ a été produite et enregistrée dans plusieurs contextes vocaliques et la séquence a été ensuite découpée en quatre portions : voyelle + transition, voyelle + transition + bruit de relâchement, bruit de relâchement + transition + voyelle,

transition + voyelle. Les auditeurs devaient identifier chaque extrait écouté séparément. Comme précédemment, les portions contenant des voyelles non-arrondies (avec ou sans bruit de relâchement) ont été reconnues comme vélaires. Celles avec des voyelles arrondies et sans le bruit de relâchement n'ont jamais été perçues comme vélaires alors que les portions avec le bruit de relâchement ont été distinguées comme vélaires : /g/ ou /k/. Delattre déduit ainsi que les éléments qui influencent la perception du lieu d'articulation d'une occlusive sont valides en parole synthétique et en parole naturelle. Il ajoute aussi que les transitions se comportent toujours selon le concept du locus. Lorsqu'une transition sans bruit de relâchement ne se dirige pas vers le locus théorique de la cible, la consonne discernée est celle dont le locus correspond à la transition. Par exemple, dans la séquence /ygo/, les transitions (sans le bruit de relâchement) s'orientent vers le locus des bilabiales et elles sont perçues comme /yp-, -bo/. Dans la séquence /aga/, les transitions (sans le bruit de relâchement) du deuxième formant se dirigent vers le locus des alvéolaires et ont été entendues comme /at-, -da/.

Il conclut qu'il ne faut pas confondre *le point de convergence* des transitions formantiques avec le locus. Le premier concept est déterminé par des mesures spectrographiques de la parole naturelle et, par conséquent, il est sujet à l'influence articulatoire des segments adjacents. Il est donc variable et conditionné par la coarticulation. Le second concept est défini par des jugements auditifs des variables acoustiques et synthétiques contrôlées. Il montre ce que les transitions formantiques devraient être afin de contribuer au maximum à la perception d'un lieu d'articulation. Il est ainsi fixe pour un dialecte et une fréquence fondamentale donnés. Selon Delattre, la coarticulation pourrait réduire considérablement le rôle perceptuel des transitions formantiques à l'égard de la perception d'une consonne cible, mais elle ne peut pas impacter le locus. Ce dernier est un moyen utile par lequel on aurait l'habileté de comprendre la coarticulation et d'évaluer la contribution des transitions formantiques à la perception de la parole. Il explique que l'interprétation physiologique des spectrogrammes de la parole naturelle révèle uniquement comment le locuteur articule et non pas comment il la perçoit. La manipulation de la parole synthétique et l'utilisation des tests psycho-acoustiques ont permis aux chercheurs de dévoiler les paramètres pertinents de la perception d'un segment. Le locus théorique d'un segment s'avère être une notion mentale fixe. Le locuteur reconnaît le segment qui correspond à un locus donné lorsqu'il entend une séquence où les transitions formantiques se dirigent vers une fréquence donnée. Si les transitions formantiques de ce segment changent et ne s'orientent plus vers son locus théorique à cause de la coarticulation, le locuteur ne percevra pas le segment en question en l'absence d'autres indices. Dans ce cas, les transitions formantiques se dirigent vers le point de convergence qui est un concept physique et changeable. Le segment est entendu grâce à d'autres informations spectrales comme le bruit de relâchement. En d'autres termes, la perception d'un son dépend d'un ensemble d'indices dont certains sont redondants. Un indice peut alors varier quand d'autres se maintiennent. De ce fait, les transitions formantiques, se dirigeant vers un locus théorique, sont un indice parmi d'autres. Elles sont compensées par d'autres indices lorsqu'elles sont modifiées à cause de la coarticulation.

Le locus statique a ensuite été développé de manière à considérer l'aspect dynamique de la parole en proposant les équations de locus (Lindblom, 1963b; Sussman et al., 1991). Ces dernières indiquent que les changements des transitions formantiques d'une consonne sont le résultat d'un processus ordonné et prédictible (Sussman et al., 1998b) (*cf.* chapitre 3). De ce fait, les études sur la coarticulation ont cherché à comprendre les raisons des différentes variations acoustiques/articulatoires engendrées par la coarticulation au lieu de s'intéresser aux invariants. Le modèle de l'*undershoot* représente un bon exemple de cette tendance.

### 2.3.2 L'*undershoot*

Lindblom (1963a) s'intéresse aux aspects dynamiques dans la production de la voyelle et le phénomène de la réduction vocalique en suédois. Il crée ainsi le modèle de l'*undershoot* et propose des équations afin de mesurer le degré de la réduction vocalique. Des séquences CVC de trois contextes consonantiques (/bVb, dVd, gVg/) et huit voyelles ont été produites par un locuteur natif du suédois. Les séquences ont été prononcées dans quatre débits différents. Il met en valeur la relation entre l'accent tonique et la durée vocalique, c'est-à-dire que plus la syllabe est accentuée, plus la durée vocalique est longue. Selon lui, le débit de parole, l'accent tonique et les différents contextes consonantiques engendrent une réduction vocalique. Les voyelles qui subissent un ou plusieurs de ces facteurs ont tendance à changer de qualité vers un schwa. Il trouve que les fréquences des formants sont modifiées quand la durée de la voyelle diminue (Figure 2.10).

Un son est spécifié par un ensemble de signaux envoyé par le système nerveux. Lindblom explique que les articulateurs répondent aux signaux de contrôle d'une manière graduelle et lente. Étant donné que la vitesse du mouvement articulaire est limitée, l'étendue à laquelle les articulateurs atteignent leurs positions cibles dépend du temps relatif à l'excitation de ces signaux. Si ces derniers sont temporellement éloignés, la réponse pourrait devenir stable pour une cible donnée. En revanche, si les commandes ont lieu dans un intervalle de temps restreint, le système devrait répondre à plusieurs signaux simultanément, ce qui entraîne la coarticulation. Autrement dit, lorsqu'un locuteur produit une série de segments phonétiques, elle est réalisée par un ensemble de commandes dans des intervalles temporels très courts. La vitesse des articulateurs étant restreinte (pour aller d'une cible articulaire à une autre), il est possible qu'une nouvelle commande soit configurée avant même que les articulateurs terminent d'effectuer la commande précédente.

Son étude démontre que lorsque la voyelle devient de plus en plus courte, les articulateurs auraient de moins en moins le temps pour compléter les mouvements articulaires transitionnels de la consonne vers la voyelle et de la voyelle vers la consonne dans les séquences CVC. En effet, les commandes neuronales correspondant aux phonèmes resteraient invariantes, mais les organes articulaires échoueraient à atteindre leurs positions articulaires lorsque les phonèmes sont produits dans un état idéal et stable. Ceci est provoqué par des limitations physiolo-

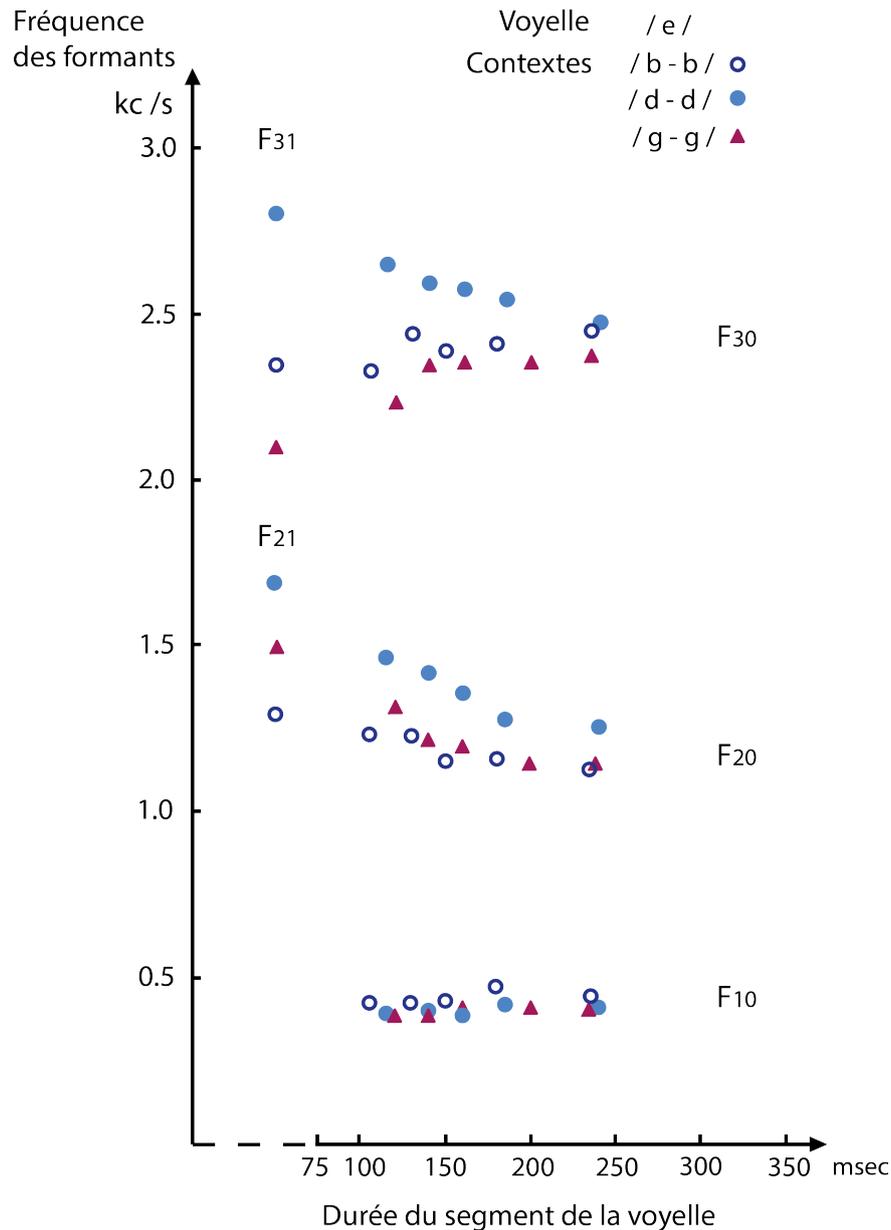


FIGURE 2.10 – Moyennes des fréquences des premier (F10), deuxième (F20) et troisième (F30) formants mesurées à l'état stable de la voyelle /e/, affichées en fonction de la durée vocalique. La voyelle est produite dans trois contextes consonantiques : /b-b/, /d-d/ et /g-g/. Lorsque la voyelle est plus courte, les fréquences du deuxième et du troisième formants changent vers les valeurs du début des transitions de /bV/, /dV/, /gV/, F2i et F3i (d'après Lindblom (1963a)).

giques. Au niveau acoustique, ce phénomène est marqué par un échec (en anglais "*undershoot*") des fréquences des formants à atteindre leurs cibles acoustiques. L'accent tonique est associé à la force de l'énoncé comparée à la quantité de l'effort musculaire. Les résultats de sa recherche laissent suggérer l'hypothèse suivante : lorsque la voyelle est produite dans des syllabes peu accentuées, elles sont articulées avec moins d'effort et moins d'ajustement des organes ; ce qui conduit à leurs modifications et par conséquent à la réduction vocalique. Selon lui, dans les deux phénomènes étudiés (l'accent tonique et le débit de parole), la durée vocalique est le responsable

principal de la réduction vocalique ou *undershoot*.

Dans une recherche sur l'influence du débit de parole et de l'accent tonique sur les propriétés spectrales des formants, Gay (1978a) remet en question le modèle d'*undershoot*. L'objectif central de l'étude est de vérifier l'existence d'un changement systématique des cibles acoustiques (vocalique et consonantique) en fonction du débit de parole. Il trouve que les milieux des formants (cibles vocaliques) ne sont pas influencés par le débit de parole. En revanche, des changements de fréquences ont été remarqués aux débuts des transitions formantiques (cibles consonantiques) lorsque le débit de parole change. Les résultats indiquent que les cibles vocaliques sont toujours atteintes dans les deux débits étudiés. Lorsque la durée de la voyelle décroît, la fréquence des cibles consonantiques se rapproche de celle des cibles vocaliques. Il explique que ces résultats constants impliquent qu'il existe une modification des gestes articulatoires. Il propose deux processus possibles qui pourraient être à l'origine de cette modification :

1. Le mouvement de l'articulateur vers la cible vocalique commencerait plus tôt. Il serait plus proche du moment de l'occlusion de la consonne précédente. Ce constat est mis en évidence par le changement de la position des débuts des transitions formantiques (*les onsets*). Lorsque le débit de parole change, les *onsets* de la transition se rapprochent de la cible de la voyelle.
2. Un même mouvement articulatoire pourrait être produit plus rapidement en débit rapide qu'en débit lent. Ce postulat est démontrable par le calcul de la vitesse de la transition formantique. Elle est mesurée par la formule suivante :

$$V_t = \left| \frac{F2_{onset} - F2_{mid}}{d} \right| \quad (2.1)$$

Où  $V_t$  = vitesse de la transition,  $F2_{onset}$  = la fréquence du début des transitions du deuxième formant,  $F2_{mid}$  = la fréquence du milieu du deuxième formant,  $d$  = la durée entre le début des transitions formantiques et le milieu des formants.

D'un autre côté, Gay constate que l'accent tonique lexical influence les propriétés spectrales dans les deux débits, il modifie aussi la durée de la voyelle. Il ajoute que ces données indiquent que le débit de parole et l'accent tonique lexical affectent la durée de la voyelle. Toutefois, ils n'ont pas les mêmes effets sur les différents paramètres spectraux. Ces derniers semblent être contrôlés indépendamment par des mécanismes physiologiques distincts. Les résultats obtenus par Gay (1978a) ne sont pas en accord avec Lindblom (1963a). Selon le premier, les formants d'une voyelle se centralisent vers un schwa uniquement pour une voyelle non-accentuée, même si elle est de la même durée que sa contrepartie accentuée. Ses résultats suggèrent ainsi que le degré de réduction est lié principalement à l'accent tonique, indépendamment de la durée du segment. Il conclut qu'une réduction globale de l'effort articulatoire pourrait engendrer une réduction des paramètres suivants : la fréquence fondamentale, l'amplitude, la durée et la qualité de la voyelle. Il précise qu'il existe deux mécanismes distincts pour contrôler les changements lors des variations de débits de parole et lors des changements liés à l'accent tonique. Le premier

implique un changement de durée délibéré et une compression horizontale de la séquence ; le second entraîne une modulation de l'effort articulatoire global dont le changement de durée est une conséquence. Il explique également que la coordination des mouvements articulatoires est réglée de manière à ce que les informations spectrales des unités segmentales soient préservées durant les changements des différents débits de parole. Les résultats de cette étude suggèrent que ce réglage implique principalement une compression horizontale sur la dimension temporelle. La dimension verticale (fréquentielle) reste, quant à elle, relativement stable. D'un autre côté, l'accent tonique joue un rôle dans les modifications spectrales et temporelles puisque les segments non-accentués subissent une réduction sur les deux dimensions (verticale et horizontale), l'amplitude globale et la fréquence fondamentale.

Dans une étude acoustique et électromyographique (EMG)<sup>22</sup> sur l'influence du débit de parole et de l'accent tonique sur la qualité de la voyelle, Harris (1978) met en avant également que l'*undershoot* des fréquences des formants est indépendant de la durée de la syllabe. Elle explique que les effets causés par les variations du débit de parole sur la cible vocalique sont distincts de ceux provoqués par les changements de l'accent tonique. Les locuteurs accomplissent les différents débits de parole avec leurs propres stratégies. Celles-ci diffèrent d'une personne à une autre. Les segments accentués sont accompagnés par un nombre de signaux plus important que les segments non-accentués lors de l'excitation des muscles. Les syllabes accentuées sont ainsi plus longues et les formants sont plus périphériques (non-centrés) que celles non-accentuées. Cette hypothèse a été déjà proposée dans la littérature, notamment par Öhman (1967).

O'Shaughnessy (1986) approfondit le travail de Gay (1978a) et étudie l'influence du débit de parole sur les formants en français canadien. Il examine les voyelles produites dans plusieurs contextes consonantiques dans trois débits de parole. Les résultats de son étude sont en accord avec ceux de Gay, il constate que la variation temporelle provoquée par les différents débits de parole n'influence pas les cibles vocaliques. De plus, il soumet l'idée qu'il existe un impact minimal sur les cibles consonantiques. Les transitions formantiques sont situées entre le début des premières pulsations des plis vocaux et l'état stable de la voyelle. Il expose que dans la plupart des cas lorsque le débit augmente, la vitesse des transitions formantiques augmente aussi et par conséquent il n'existe pas d'*undershoot*. Les cibles restent donc relativement stables lorsque le débit de parole change. Toutefois, la vitesse des transitions formantiques est modifiée par le changement du débit de parole (+ le débit est lent, – la vitesse de transition est rapide). Il remarque également que les effets du débit de parole ne sont pas linéaires et varient en fonction du contexte phonétique (consonantique ou syllabique).

Plus récemment, Weismer et Berry (2003) analysent l'influence de l'alternance du débit de parole sur les changements spectraux au début, à la fin et au milieu du deuxième formant

---

22. Électromyographie (EMG) : technique d'électrophysiologie consistant à fournir des informations graphiques sur les activités électriques des fibres musculaires accompagnant la contraction des muscles et les effets subséquents ainsi que le développement de la tension musculaire (Hirose, 2010).

(paramètres statiques) ainsi que sur les variations des trajectoires du F2 en fonction du temps (paramètre dynamique). Six locuteurs de l'anglais américain ont prononcé les séquences /bɪd/, /bæŋ/ et /boɪ/ dans une phrase porteuse dans plusieurs débits de parole. Les résultats ont montré que l'impact du débit de parole sur les paramètres statiques et dynamiques dépend de façon importante du contexte vocalique et du locuteur. Ces résultats laissent suggérer que certaines voyelles sont plus sujettes à la réduction vocalique que d'autres et que certains locuteurs réorganisent leurs articulateurs de manière à éviter l'*undershoot*, lorsque le débit de parole est modifié. Ces dépendances (du contexte et du locuteur) peuvent ainsi justifier les résultats variés – et parfois contradictoires – sur la présence de l'*undershoot* dans les études explorées ci-dessus.

Les deux modèles précédents se sont concentrés sur la coarticulation d'une consonne (vers une voyelle) en négligeant l'influence de tout autre contexte vocalique adjacent. Le modèle d'Öhman offre des arguments indiquant que la relation coarticulatoire CV ou VC est directement impactée par un segment adjacent.

### 2.3.3 Le modèle d'Öhman

Dans son travail, Öhman (1966) s'intéresse à décrire comment les occlusives voisées sont coarticulées dans le contexte voyelle-consonne-voyelle (VCV). Il souligne que lors de la production d'une consonne intervocalique, les transitions du deuxième formant sont dépendantes des voyelles et la consonne ne joue pas de rôle dans cette action. La Figure 2.11 montre les spectrogrammes des séquences /øgy/ et /øga/ produites par un locuteur suédois. Les débuts des deux séquences sont identiques (/øg/) mais la voyelle de fin de séquence change. Pourtant, les transitions du deuxième formant ne sont pas les mêmes pour la portion initiale VC des deux séquences. La transition est montante pour le début de la première séquence. Au contraire, la transition est descendante pour le début de la seconde séquence. Dans les deux séquences, les transitions du deuxième formant de la portion VC initiale s'orientent vers les transitions formantiques de la portion finale. Cette constatation prouve que la voyelle finale influence la voyelle initiale malgré la présence de la consonne. La Figure 2.12 présente les spectrogrammes des deux séquences /ydø/ et /odø/ qui ont la même partie finale /dø/ et deux voyelles initiales différentes. Les transitions du deuxième formant dans la partie VC semblent aussi être modifiées en fonction de la voyelle adjacente. Toutefois cette modification est moins marquée que dans l'exemple précédent (Figure 2.11). Ces observations démontrent que la voyelle finale exerce une certaine influence sur les segments précédents, allant jusqu'à modifier la structure formantique de la voyelle initiale et/ou ses transitions formantiques avec la consonne. Un phénomène similaire est donc noté lorsque cette fois la voyelle initiale est changée.

Afin de mesurer l'influence vocalique sur les transitions formantiques, Öhman (1966) mesure l'écart entre la fréquence à la fin et au début de la transition du deuxième formant dans les

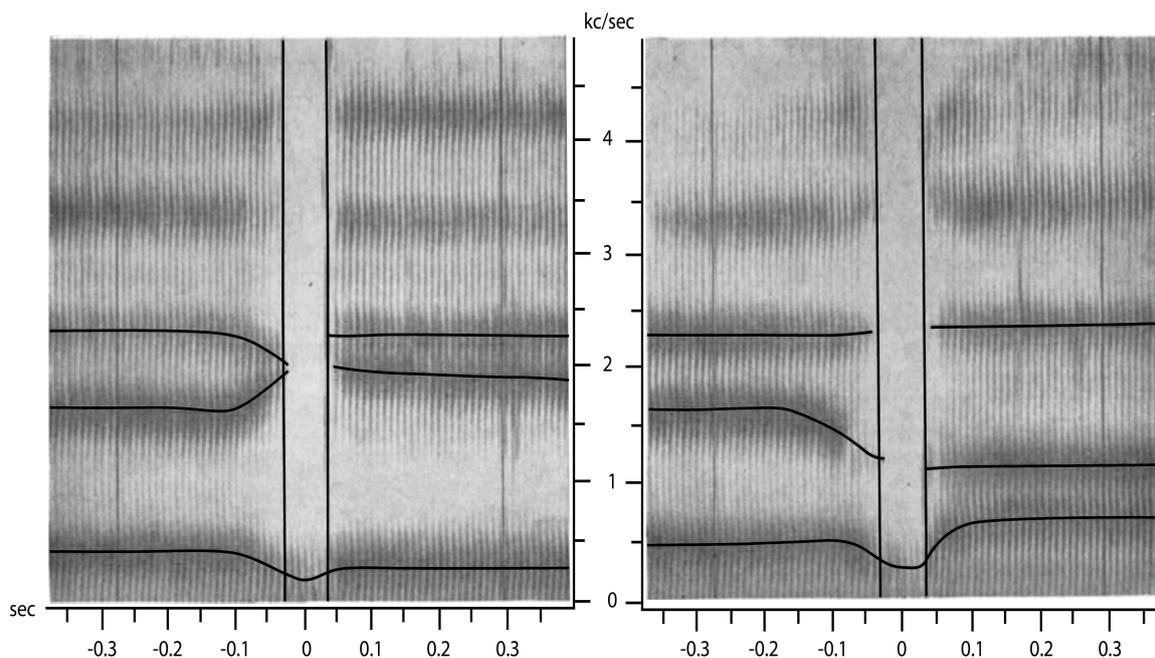


FIGURE 2.11 – Représentation spectrographique des transitions formantiques des séquences /øgy/ à gauche et /øga/ à droite prononcées par un locuteur suédois. Il est à remarquer la différence des transitions VC dans les séquences en fonction de la voyelle finale (empruntée à Öhman (1966)).

portions VC/CV de chaque séquence :

$$Et = F2_t - F2_v \quad (2.2)$$

Où  $Et$  = l'étendue de la transition,  $F2_t$  = la fréquence du point terminal de la transition<sup>23</sup> du deuxième formant,  $F2_v$  = la fréquence au milieu du deuxième formant.

Grâce à ce calcul, le changement de transition d'une consonne vers une voyelle fixe, lorsque la voyelle précédente varie, peut être comparé dans les deux séquences. Il est également possible de comparer le changement de transition de la voyelle vers la consonne lorsque la voyelle suivante varie. Les résultats montrent que, dans les portions VC et CV, les transitions sont dépendantes de la voyelle finale/initiale respectivement. En mesurant les fréquences au milieu des trois premiers formants des voyelles initiales et finales, l'auteur précise que les consonnes intervocaliques étudiées /b, d, g/ ont un impact minimal sur les voyelles. La vélaire /g/ semble influencer la fréquence du deuxième formant des voyelles /o, u/. Pourtant, la dispersion globale des fréquences des formants (mesurées au milieu) est négligeable, indiquant une stabilité générale de la production des voyelles dans les différents contextes consonantiques. Les mesures de la fréquence au point terminal de la transition révèlent un chevauchement important entre les consonnes. Par exemple, les trois occlusives se retrouvent dans une même zone fréquentielle (1450 – 1500 Hz) lorsqu'elles sont produites dans les contextes suivants : /øbø/, /øgo/, /øda/. Öhman précise que ces résultats attestent de l'inexistence d'un point fixe pour chaque consonne

23. Ceci correspond au début de la transition dans les portions CV et à fin de la transition dans les portions VC.

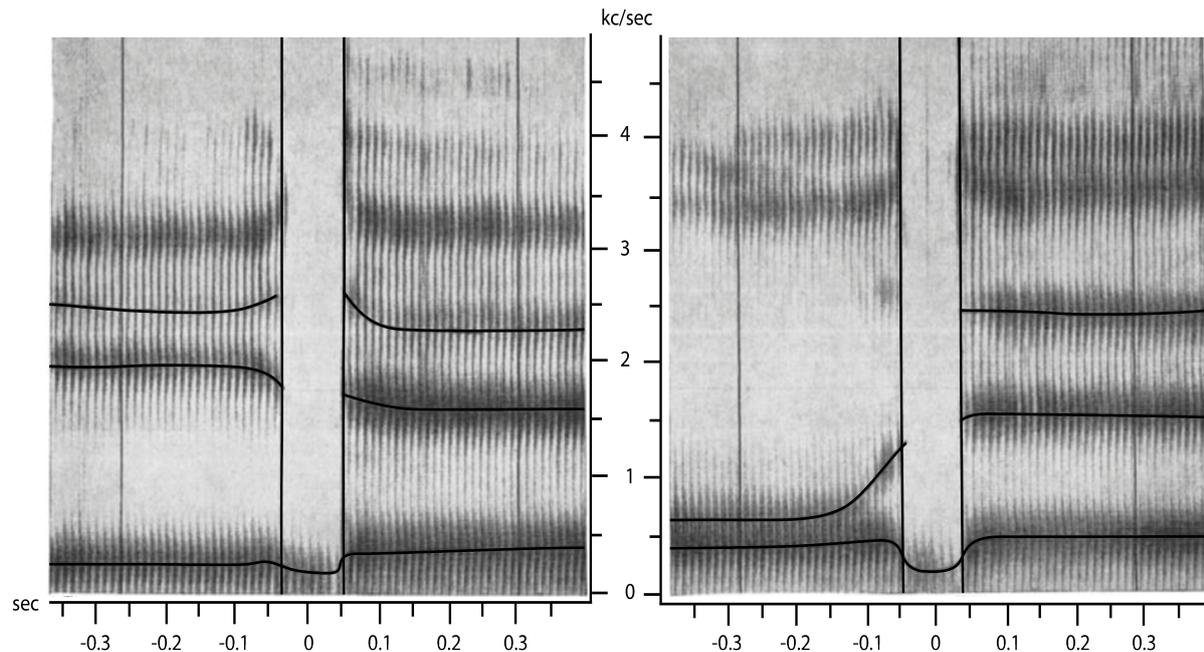


FIGURE 2.12 – Représentation spectrographique des transitions formantiques des séquences /ydø/ à gauche et /odø/ à droite prononcées par un locuteur suédois. Il est à remarquer la différence des transitions CV dans la première et la seconde séquences en fonction de la voyelle initiale (empruntée à Öhman (1966)).

puisque la fréquence de chaque consonne n'est pas indépendante du contexte vocalique. L'effet de coarticulation VCV a été trouvé en suédois et en anglais américain, contrairement au russe. Öhman explique qu'en russe la liberté de coarticulation est faible. Lorsque les mouvements des deux voyelles s'opposent, le mouvement consonantique domine.

Les résultats d'Öhman (1966) contredisent la théorie du locus de Delattre et al. (1955) dans laquelle un point fixe de fréquence (vers lequel les transitions formantiques se dirigent) existerait pour chaque lieu d'articulation<sup>24</sup>. Öhman montre que le travail de Delattre et al. (1955) sur le locus formantique est basé sur des séquences CV synthétiques et que, dans la parole naturelle, l'effet de la coarticulation peut faire varier les valeurs des fréquences des transitions formantiques et la direction (positive-négative) de ces transitions. Il explique que dans les séquences d'occlusives intervocaliques, le geste articulatoire est modifié en fonction de la voyelle qui suit ou précède la consonne ; ceci est reflété sur le spectrogramme comme des changements dans les transitions formantiques. Dans des séquences VCV, les deux voyelles interagissent malgré l'existence d'une consonne intervocalique. Des images cinéradiographiques aux rayons X<sup>25</sup> ont été utilisées afin de trouver une interprétation physiologique à ces phénomènes. Les voyelles

24. Néanmoins, Delattre (1969) a répondu aux critiques d'Öhman, ses arguments sont présentés dans la sous section 2.3.1

25. « Les rayons X sont une forme de rayonnement électromagnétique à haute fréquence qui peuvent traverser le corps humain. Lorsqu'un faisceau de rayons X est projeté à travers le corps, il sera absorbé différemment selon la structure et la densité du tissu » (Amelot 2009, page 66). C'est donc un moyen très efficace pour obtenir des informations dynamiques des mouvements des articulateurs. Pourtant, l'utilisation des rayons X représente un risque d'endommager les tissus sains et il est maintenant exclu pour des raisons d'éthique (Vaxelaire et al., 2009).

sont produites par des mouvements diphtongaux (relativement lents) du corps de la langue sur lesquels les gestes articulatoires des consonnes sont superposés. Les voyelles sont donc coarticulées à travers une consonne médiane. Autrement dit, les mouvements vers la seconde voyelle commencent indépendamment de la consonne intervocalique. Les gestes de la voyelle sont déjà activés au début de ceux de la consonne précédente et la formation de la configuration de la voyelle se fait pendant la fermeture de la consonne.

Une étude radiographique a été effectuée par Gay (1977) dans le but d'analyser les gestes articulatoires dans les séquences VCV (C=/p, t, k/, V=/i, a, u/). Les mouvements des lèvres, de la mâchoire, de l'apex et du corps de la langue ont été examinés. Les données indiquent que les mouvements anticipatoires vers la seconde voyelle commencent pendant la fermeture de la consonne intervocalique. Ces mouvements sont réalisés après que la fermeture des consonnes soit complète (entre 5-60 ms après la fermeture). Par conséquent, la coarticulation anticipatoire entre les deux voyelles ( $\overleftarrow{VCV}$ ) est restreinte, cela inclut les mouvements du corps de la langue et de la mâchoire pour préparer la prononciation de la seconde voyelle. La coarticulation anticipatoire est donc limitée aux deux segments adjacents ( $\overleftarrow{VC}$  et  $\overleftarrow{CV}$ ). La coarticulation persévérante de la première voyelle vers la seconde ( $\overrightarrow{VCV}$ ) semble également être absente. Quant à la coarticulation de la première voyelle vers la consonne ( $\overrightarrow{VC}$ ), elle est variable. Elle dépend du degré de l'engagement du corps de la langue pendant la production de la consonne. Les effets de la coarticulation persévérante de la consonne intervocalique vers la seconde voyelle ( $\overrightarrow{CV}$ ) apparaissent uniquement pour la voyelle /a/. Les résultats de Gay (1977) mettent en évidence que la temporalité des mouvements articulatoires dans les séquences VCV est affectée par la consonne intervocalique. Les mouvements articulatoires vers la seconde voyelle commencent toujours après la fermeture de la consonne, même pour le /p/ qui n'entraîne pas d'engagement direct du dos de la langue. Ces résultats ne concordent pas avec ceux du modèle d'Öhman selon lequel, le mouvement articulatoire dans la séquence VCV est essentiellement diphtongal. Le modèle d'Öhman est basé sur l'argument que le mouvement du corps de la langue vers la seconde voyelle commence au moment du début de la clôture de la consonne. Gay propose que les gestes articulatoires entre les deux voyelles ne sont pas de simples mouvements diphtongaux. Les gestes consonantiques sont formés et terminés pendant la première et la seconde voyelle. Ils pourraient ainsi apporter des modifications sur les mouvements vocaliques. Ces données sous-entendent que les mouvements anticipatoires commencent et les effets principaux de la coarticulation persévérante se terminent en même temps pendant la période de clôture de la consonne. L'auteur signale que ce phénomène suggère que le relâchement de la consonne et le mouvement vers la voyelle suivante sont organisés et produits comme un événement articulatoire intégral.

Le principe de superposition, qui gouverne les transitions voyelle-à-voyelle et la fermeture consonantique, a été examiné par Carré et Chennoukh (1995). Le modèle *DRM* (*Distinctive Region Model*)<sup>26</sup> a été utilisé dans cette étude. Ce modèle, capable de prédire les lieux de

26. Le Distinctive Region Model (DRM) est un modèle qui simule sur ordinateur la production de la parole

constriction tout au long du conduit vocal, permettrait ainsi la compréhension des variations acoustiques provoquées par les mouvements articulatoires. Carré et Chennoukh ont modélisé la coarticulation VCV des séquences énoncées dans l'article d'Öhman (1966) par le biais du DRM. Le principe de superposition d'Öhman (1966) a été appliqué afin de simuler la produc-

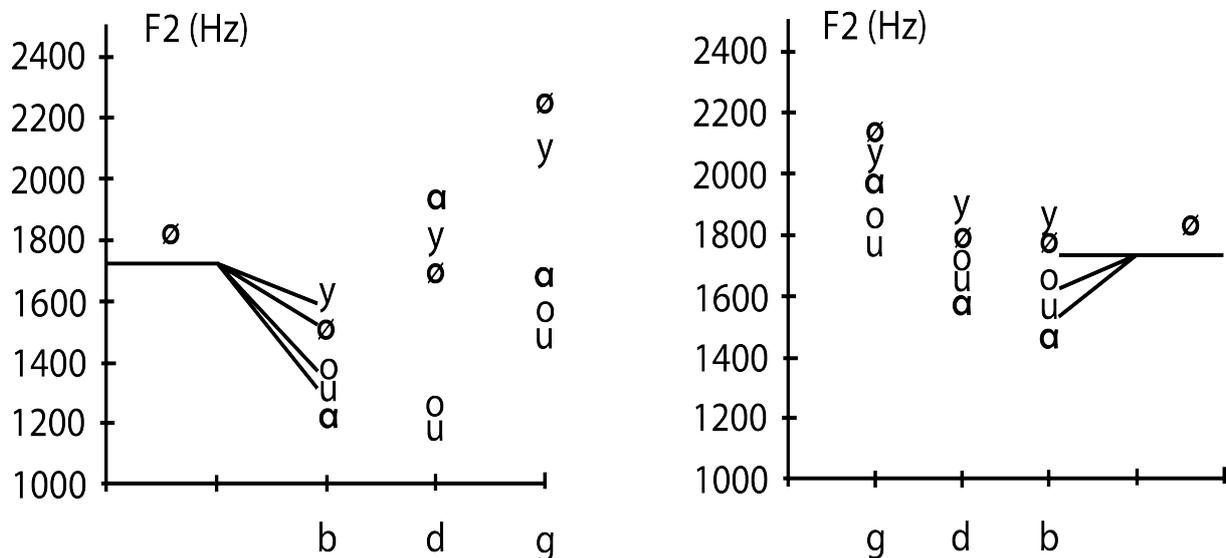


FIGURE 2.13 – Les valeurs du début et de la fin du deuxième formant obtenues par le modèle *DRM* pour les séquences /øCV/ et /VCø/ (d'après Carré et Chennoukh (1995)).

tion de VCV. Les mouvements entre la première et la seconde voyelle ont été synchronisés avec la clôture de la consonne. Les transitions de voyelle-à-voyelle ont été simulées, puis le geste de la consonne a été superposé sur les transitions. Par exemple, dans la production de la séquence /abi/, les transitions entre le /a/ et le /i/ sont d'abord calculées et le geste de /b/ est superposé plus tard. Les résultats de cette étude soutiennent le modèle VCV et la dominance des voyelles sur la consonne intervocalique. Ils montrent également que les directions des transitions formantiques et les fréquences des trois premiers formants sont similaires à ce qui a été rapporté dans l'étude d'Öhman (1966) (Figure 2.13). Les fréquences au début ou à la fin des transitions du F2 changent en fonction de la voyelle trans-consonantique précédente ou suivante. Carré et Chennoukh (1995) expliquent de même que la mise en place de la consonne avec la seconde voyelle conduit à une simulation correcte de la production des CVC. Ce postulat contredit l'hypothèse de Gay (1977) où la seconde voyelle commence après la fin de la formation de la consonne et appuie l'explication d'Öhman (1966).

L'idée de la dominance des voyelles sur la consonne intervocalique – observable dans les Figures 2.11, 2.12 et 2.13 – laisse penser que des différences existent au niveau des puissances coarticulatoires entre les segments. Autrement dit, certains segments sont plus influents que d'autres. Les deux modèles suivants – la résistance coarticulatoire et le DAC – suggèrent des

dans un tube acoustique (Mrayati et al., 1988; Carré et Chennoukh, 1995). Ce tube est divisé en plusieurs régions spécifiques représentant les différents lieux d'articulation. Dans cette étude, le tube a été divisé en 8 régions. Chaque changement dans le tube est relié à des variations dans les fréquences des formants.

techniques d'analyse et de mesure pour les degrés de dominance/influence coarticulatoire des différents sons de la parole.

### 2.3.4 La résistance coarticulatoire

Bladon et Al-Bamerni (1976) proposent un principe général qui a pour objectif d'explicitier la variation coarticulatoire. Ce modèle est exposé à travers une étude acoustique sur la variation des propriétés spectrales des allophones du /l/, clair ([l]), sombre ([ɫ]) et syllabique ([ɫ̩]) en anglais britannique standard nommé "*Received Pronunciation*". Les deux allophones principaux (clair et sombre) sont sujet à plusieurs variations de qualité selon le contexte phonétique. En anglais britannique standard, le lieu d'articulation principal de ces deux allophones est alvéolaire, il devient dental lorsqu'ils sont réalisés avec /θ, ð/ et post-alvéolaire quand ils sont prononcés avec un /r/. Le /l/ clair est réalisé quand il est suivi d'une voyelle ou d'un /j/ (ex. #[l]V) et le /l/ sombre est produit dans toutes les autres situations (ex. C[ɫ]#). Bladon et Al-Bamerni signalent que d'autres variations de qualité existent en fonction du contexte. Dans leur travail, ils s'interrogent sur les questions suivantes :

- Quel est le degré de la variation phonétique dans le /l/ ?
- Comment identifier ces variations qui pourraient être considérées comme un résultat de la coarticulation ?
- La direction de la coarticulation (de droite à gauche ou de gauche à droite) fonctionne-t-elle de la même manière que le décrivent Kozhevnikov et Chistovich (1965)<sup>27</sup> ?
- Quels principes linguistiques pourraient régulariser la coarticulation ?

Le protocole expérimental mené dans cette étude était basé sur la production des variations que le phonème /l/ peut subir dans divers contextes : la variation de qualité causée par les voyelles adjacentes, celle provoquée par un deuxième /l/ adjacent et le dévoisement occasionné par une consonne non voisée adjacente.

Ils considèrent l'existence d'une relation articulatoire-acoustique entre le degré de rétraction de la langue et la valeur de la fréquence de F2. Plus la langue est rétractée, plus la fréquence de F2 est basse. Par conséquent, il est possible de distinguer la qualité claire vs. sombre de /l/ grâce à la fréquence de F2 : plus la fréquence est basse, plus le son est sombre. Selon eux, il serait possible de mesurer une valeur canonique de F2 pour chaque son afin de quantifier le degré de coarticulation d'un son (le /l/ dans cette étude) dans un contexte syllabique donné. Ils proposent la mesure de la moyenne des valeurs de F2 de /l/ dans des environnements différents. Cette mesure est considérée comme une valeur canonique de référence pour un contexte quel-

27. Kozhevnikov et Chistovich (1965) ont défini la syllabe comme étant une séquence CV, où C pourrait être n'importe quel nombre de consonnes. La coarticulation anticipatoire est maximale à l'intérieur de la syllabe et minimale entre les frontières syllabiques. Les commandes neuronales de la syllabe entière sont spécifiées simultanément dès le début du premier son dans la mesure où les commandes ne sont pas contradictoires. Les commandes articulatoires en conflit qui rivalisent entre elles sont différées jusqu'à ce que les commandes précédentes soient complètes. La coarticulation n'est donc pas inhibée d'une façon marquée par les frontières syllabiques environnantes, à l'exception des frontières de haut-niveau comme un groupe d'intonations ou une pause.

conque. Par exemple, dans le contexte #[l]V, la moyenne de F2 de [l] coarticulé avec plusieurs voyelles est de 1218 mels<sup>28</sup> ; cette valeur est appelée la valeur canonique de F2 de ce contexte. Pour le contexte C[ɫ]#, la valeur canonique est la moyenne de F2 de [ɫ] produit avec plusieurs consonnes (840 mels). Le degré de coarticulation (nommé également moyenne de la distance articulatoire) est donc la déviation de cette valeur canonique.

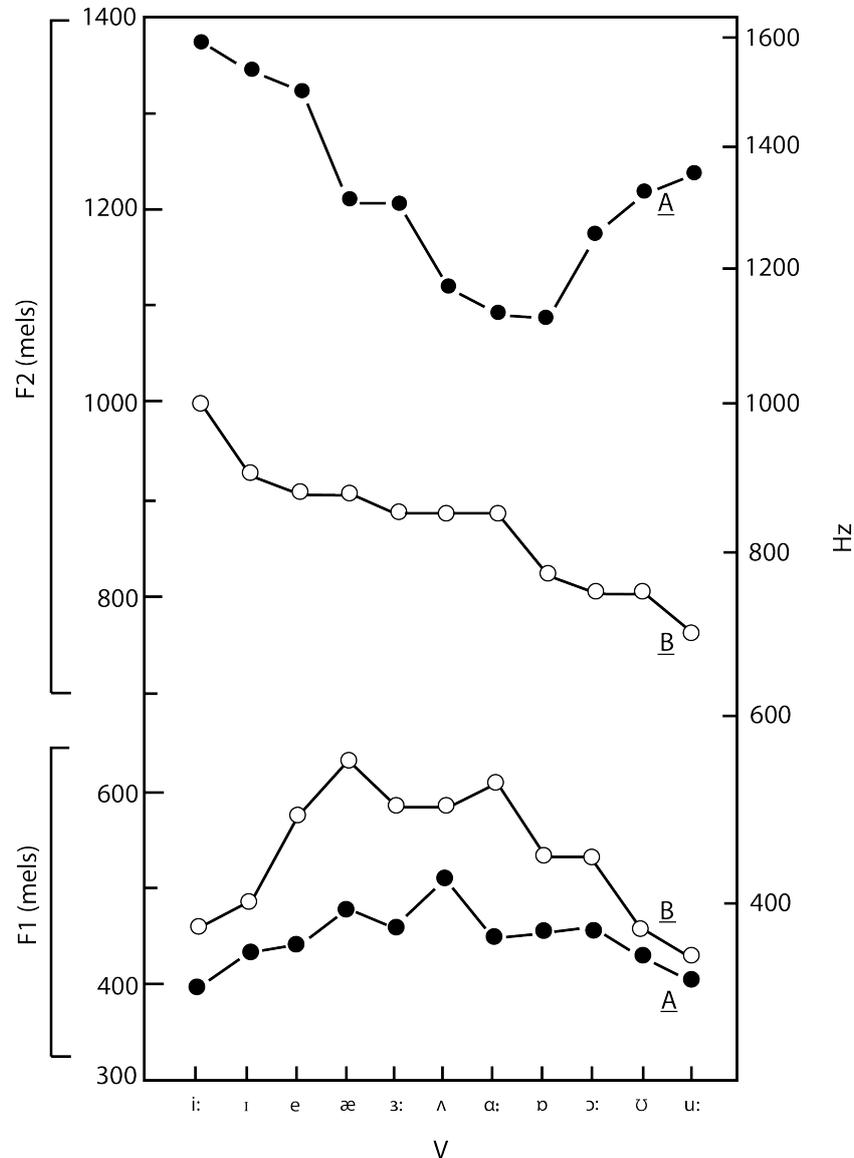


FIGURE 2.14 – Les fréquences (en mels à gauche et en Hz à droite) de F1 et F2 du /l/ en deux contextes différents : A #[l]V et B V[ɫ]#, en fonction de la voyelle adjacente. Les valeurs sont les moyennes des quatre locuteurs (d'après Bladon et Al-Bamerni (1976)).

28. Les auteurs ont transformé la fréquence en mel étant donné que la relation entre la qualité perçue et la fréquence en Hz des formants n'est pas linéaire. La hauteur du son est directement liée à la fréquence. La hauteur étant une sensation, lorsque la fréquence de vibration croît, une augmentation de la hauteur est entendue. Quand la fréquence décroît, une diminution de la hauteur est perçue. Toutefois, cette relation n'est pas linéaire. Grâce à des tests psychoacoustique (Stevens et al., 1937), cette relation peut être représentée sur une échelle logarithmique, appelée "mel" (Borden et al., 1994). D'autres échelles non-linéaires (*bark* et *ERB*) sont également utilisées pour représenter la relation entre les changements de la hauteur des fréquences sonores et la variation perçue.

Les résultats montrent que le degré de coarticulation varie d'une manière considérable en fonction du contexte vocalique pour les deux allophones, mais il est plus important pour le [l] clair (Figure 2.14). La coarticulation de droite à gauche est aussi marquée que celle de gauche à droite dans ce contexte. L'impact de coarticulation est minimal pour le [ɫ] (sombre et syllabique) parce qu'il est phonologiquement loin de la voyelle suivante. Lorsque les deux latérales [ɫ, l] sont coarticulées, la latérale claire devient plus sombre. La même influence est notée lorsque le [l] clair est coarticulé avec le [ɫ] sombre non-syllabique. La coarticulation anticipatoire est plus faible que celle persévérante dans ce contexte. Un intérêt a aussi été porté sur le degré de dévoisement de /l/ et, par conséquent, sur l'effet coarticulatoire des consonnes non voisées sur la latérale. Le degré de dévoisement est déterminé à l'aide du pourcentage de la durée totale non voisée du /l/. Un léger effet de dévoisement est constaté pour les trois allophones dans la coarticulation de droite à gauche. Au contraire, le dévoisement est plus important dans la coarticulation de gauche à droite, notamment derrière les occlusives non voisées. En effet, il est moins important avec les fricatives non voisées. La latérale sombre et syllabique est la moins influencée.

Ils concluent que la variation coarticulatoire des trois allophones de /l/ étudiés suggère un principe de contrôle sur lequel dépend le mécanisme de l'encodage de la parole. Ils définissent la *résistance coarticulatoire (RC)* comme une propriété associée aux caractéristiques phonétiques des segments de la parole. Cette propriété est sous la forme d'une valeur numérique qui varie selon les segments et elle est indépendante de la direction coarticulatoire (anticipatoire ou persévérante) et du contexte (qualité vocalique, latérale ou dévoisement). Dans une succession de sons, la coarticulation est uniquement inhibée par la résistance coarticulatoire. Chaque allophone est associé à une valeur de RC qui pourrait être dépendante de la langue. Un segment à 3 RC indique un degré élevé de résistance et un segment à 1 RC désigne un degré bas de résistance. La coarticulation est donc permise jusqu'à ce qu'elle soit inhibée par un degré de RC élevé. Dans le cas de l'anglais britannique, la résistance coarticulatoire augmente d'un ordre croissant pour les allophones de la latérale : [ɫ] < [ɫ] < [l]. Autrement dit, le [ɫ] est, lui-même, plus résistant que [ɫ] qui est plus résistant que [l]. Selon eux, la RC dépend de plusieurs facteurs : le segment, la langue, le contexte, le dialecte et le locuteur. Kent et Minifie (1977) expliquent que la solution de la coarticulation est aussi complexe que la multiplicité de ces facteurs. Par conséquent, les valeurs de RC ne se limitent pas à un ensemble de nombres mais représentent les contributions de nombreux effets inconnus ou peu connus. Ils considèrent ainsi que les valeurs de RC ne génèrent pas de prédictions, mais elles font partie d'une description quantitative de l'articulation de la parole. Kühnert et Nolan (1999) sont en désaccord avec cette proposition qui implique de résumer les différents facteurs de la RC en une valeur numérique. Pour eux, un nombre n'est pas suffisant pour décrire le comportement global d'un segment. La notion de résistance coarticulatoire est utilisée actuellement dans un sens plus général désignant que la coarticulation est graduelle et varie selon les segments. Elle est souvent définie comme l'étendue du blocage de l'influence coarticulatoire d'un segment phonétique sur un autre segment

adjacent (Recasens et Espinosa, 2009).

Recasens (1985) propose une autre manière (différente de celle discutée précédemment) de mesurer le degré de la résistance coarticulatoire d'un segment. Celle-ci est déterminée par le calcul des valeurs de l'écart-type de la fréquence de chaque formant dans les différentes conditions coarticulatoires. Le degré de dispersion verticale (l'échelle de fréquence) est ainsi évalué au moyen de l'écart-type. La précision de cette échelle permet d'estimer la capacité d'un segment à résister à l'influence d'un autre. Quand une consonne est produite dans plusieurs contextes vocaliques et consonantiques, une variation des valeurs formantiques de cette consonne est attendue. Une grande variation reflète une résistance coarticulatoire faible ; autrement dit, les gestes articulatoires de la consonne sont facilement influencés par ceux des segments voisins. Une petite variation traduit une résistance coarticulatoire élevée, les gestes articulatoires de la consonne sont difficilement influencés par les segments adjacents. La valeur de l'écart-type permet le calcul de cette variation. Une valeur élevée de l'écart-type révèle une variation importante et, de ce fait, un degré de résistance coarticulatoire faible. Une valeur basse de l'écart-type signale une variation minimale et, ainsi, une résistance coarticulatoire importante.

Afin d'étudier la résistance coarticulatoire, Recasens (1985) examine l'effet de la coarticulation de la voyelle sur la consonne et de la consonne sur la voyelle dans des séquences CV (prononcées dans des mots  $CV_1CV_2$ )<sup>29</sup> en catalan. Les fréquences des trois premiers formants au début des transitions formantiques sont évaluées et considérées comme des valeurs consonantiques ( $F1_{onset}$ ,  $F2_{onset}$ ,  $F3_{onset}$ ). Les fréquences au milieu des formants sont prélevées et jugées comme des valeurs vocaliques ( $F1_{mid}$ ,  $F2_{mid}$ ,  $F3_{mid}$ ). En examinant l'influence de la voyelle sur la consonne, Recasens remarque que les F2 et les F3 des consonnes varient en fonction de la voyelle suivante. Les  $F3_{onset}$  subissent moins de modifications que les  $F2_{onset}$ . Les fréquences des  $F2_{onset}$  et  $F3_{onset}$  diminuent lorsque le degré de constriction des voyelles antérieures décroît (– de constriction, – de fréquence) et quand le degré d'arrondissement des voyelles postérieures augmente (+ d'arrondissement, – de fréquence). Les valeurs formantiques du schwa catalan sont proches de celles de /a/. D'un autre côté, la fréquence du  $F1_{onset}$  décroît quand le degré de la hauteur de la voyelle et l'arrondissement des lèvres croissent. Étant donné que le /a/ est la voyelle la plus basse, celle-ci a la fréquence du  $F1_{onset}$  la plus élevée. Le /i/ étant la voyelle la plus élevée, la fréquence du  $F1_{onset}$  est la plus basse. Ces résultats indiquent une influence importante de la voyelle sur la consonne. Toutefois, chaque consonne a un degré de résistance à la coarticulation différent en fonction de la voyelle utilisée.

Les fréquences du  $F2_{onset}$  laissent suggérer que les vélaires et les labiales détiennent une variabilité élevée puisque la valeur de l'écart-type est importante pour ces deux catégories (plus importante pour les vélaires). L'existence de deux allophones qui sont liés directement à l'antériorité-postériorité de la voyelle adjacente explique la grande variabilité des fréquences

29.  $V_1 = /i, e, \varepsilon, a, \text{ɔ}, o, u/$ ,  $V_2 = /a, \text{ə}/$ . Les consonnes sont symétriques dans toutes les séquences et correspondent à une des consonnes suivantes /p, b, t, d, k, g, c, ʃ, f, s, z, ʒ, j, ʎ, r, w, l/. La première séquence CV est étudiée dans cet article.

du  $F2_{onset}$  des vélaires. Le lieu d'articulation des vélaires change en fonction de la voyelle voisine. Lorsqu'une vélaire est coarticulée avec une voyelle antérieure, le lieu d'articulation est palato-vélaire et quand elle est produite avec une voyelle postérieure, le lieu d'articulation devient post-vélaire. Ainsi, la valeur fréquentielle de la consonne dépend fortement de la voyelle voisine. Il expose que la grande variabilité des  $F2_{onset}$  des labiales est occasionnée par l'emplacement libre du dos de la langue pendant la fermeture de la consonne lors de la prononciation des voyelles. Aucune restriction qui pourrait être faite par la langue n'est demandée dans la production des occlusives labiales. Les dentales et les alvéolaires acceptent une variabilité réduite, les /r, l/ sont les plus résistants des alvéolaires. L'auteur précise que la résistance coarticulatoire élevée de ces consonnes est provoquée par le couplage physique entre la pointe et le dos de la langue. Celle-ci est engagée dans la production de la voyelle et de la consonne, créant une dépendance entre les deux segments (et par conséquent un conflit entre les gestes articulatoires). Ce postulat sous-entend que la coarticulation minimale (la résistance coarticulatoire maximale) est engendrée par un degré de contrainte articulatoire élevé sur le dos de la langue. Par exemple, le degré de contrainte du dos de la langue est élevé pour le /l/, parce que la production de la latérale engage la formation simultanée de deux lieux d'articulation : alvéolaire (par le contact apical) et pharyngo-vélaire (par le contact dorsal). La langue a donc une forme concave plus importante lors de la production du /l/ que dans celle des autres alvéolaires, ce qui conduit à une diminution importante de la fréquence du F2 (plus importante que pour les autres alvéolaires). Les fréquences du  $F3_{onset}$  montrent une grande variabilité pour les vélaires et les palatales en comparaison des autres lieux d'articulation (les palatales sont plus résistantes que les vélaires). Lorsque les palatales et les palato-vélaires sont produites avec des voyelles antérieures, la cavité frontale devient petite et le passage dorso-palatal est très réduit, induisant une élévation de la fréquence du  $F3_{onset}$ . Une diminution de la fréquence de  $F3_{onset}$  est visible quand les vélaires et les palatales sont prononcées avec des voyelles postérieures en fonction de l'arrondissement des lèvres. La variabilité restreinte des autres catégories consonantiques est générée par la stabilité du  $F3_{onset}$  dans les différents contextes vocaliques. L'écart-type de la fréquence des  $F1_{onset}$  indique que la variabilité augmente en fonction du degré de l'ouverture consonantiques, notamment des consonnes avec une fermeture linguale. L'effet de la voyelle sur  $F1_{onset}$  est ainsi bloqué par les consonnes qui impliquent une fermeture linguale complète.

En examinant l'influence de la consonne sur la voyelle, la fréquence du  $F2_{mid}$  de /i/ montre un degré de résistance coarticulatoire élevé. Les articulations nécessitant une élévation du dos de la langue vers le palais (le /i/ mais aussi les palatales) conduisent à un blocage important de la coarticulation avec le segment adjacent. La fréquence du  $F2_{mid}$  varie selon le degré de constriction dorso-palatale et pharyngale de la consonne. La fréquence du  $F2_{mid}$  est élevée quand les voyelles sont produites avec une palatale ou une palato-vélaire. Ces dernières engagent une fermeture dorso-palatale et un passage pharyngal large. Il est à noter également que l'influence consonantique sur le /ə/ est similaire à celle de /a/ (le /ə/ est moins résistant que le /a/). Les palatales créent une augmentation de la fréquence du  $F2_{mid}$  alors que les consonnes

/w, l/ provoquent une diminution de celle-ci. La voyelle /u/ est la moins résistante. Ce phénomène est dû à l'impact important des palatales lors de la production de cette voyelle. Ces consonnes entraînent un avancement considérable de la constriction dorsale de la voyelle /u/. La fréquence du  $F3_{mid}$  dépend de la constriction palatale et de la taille de la cavité antérieure du conduit vocal.  $F3_{mid}$  semble être plus sensible que  $F2_{mid}$  aux changements de la constriction palatale et au degré de l'arrondissement des consonnes adjacentes. La fréquence du  $F1_{mid}$  augmente en fonction du degré de l'ouverture orale (+ d'ouverture, + de fréquence). L'écart-type met en évidence que la variabilité croît quand le degré de l'ouverture est plus important. Tous ces résultats exposent que les consonnes peuvent exiger un grand degré de contrôle articulaire, telles que les palatales. Les consonnes et les voyelles peuvent être produites par des gestes similaires, comme le /i/ et les consonnes palatales.

L'auteur révèle que le degré de résistance coarticulaire est directement lié à une notion supplémentaire celle du *Degré de Contrainte Articulaire*. Ce dernier permet de vérifier si les gestes d'un segment sont compatibles avec les effets coarticulaires des gestes adjacents ou non. Les gestes fortement contraints réduisent énormément l'effet coarticulaire des gestes opposés, comme l'articulation des palatales et des consonnes qui exigent deux lieux de constriction. Les gestes qui sont sujets à un degré bas de contrainte (telle que la production des labiales ou du schwa) deviennent compatibles avec les autres gestes articulaires. Cette étude a ouvert la voie à la proposition du modèle de DAC.

### 2.3.5 Le degré de contrainte articulaire (DAC)

Recasens et al. (1997) ont cherché un modèle qui prédirait à quel point la coarticulation serait permise par un segment. Ils proposent ainsi le modèle du *degré de contrainte articulaire* (DAC). La notion de contrainte articulaire et l'importance du dos de la langue dans la coarticulation ont déjà été discutées dans plusieurs recherches (Recasens, 1984a,b, 1985, 1990). Il peut être cité, entre autres, le travail sur la résistance coarticulaire de la voyelle vers la consonne et de la consonne vers la voyelle en catalan. Recasens trouve que le degré de la RC est différent pour tous les phonèmes et suggère que ces différences sont dues aux contraintes articulaires de chaque phonème. La coarticulation est permise tant que l'articulateur engagé n'est pas en conflit avec le besoin dynamique d'un autre articulateur pendant la production d'un phonème. La réalisation articulaire des segments est conditionnée par la force des contraintes imposée aux articulateurs lors de la formation de l'occlusion ou de la constriction des segments adjacents (Recasens, 2011). Recasens et al. (1997) suggèrent ainsi que les gestes articulaires d'un phonème varient en fonction du degré de compatibilité avec les gestes voisins. Par conséquent, le degré de contraintes détermine l'étendue de la résistance d'un geste articulaire sur un geste adjacent. Ils prédisent le comportement coarticulaire d'un ensemble de segments (/ʃ, ɲ, i, k, l, s, n, a, p, ə/) en leur distribuant des valeurs de DAC, en se basant sur les informations articulaires et acoustiques rapportées dans la littérature. Cette prédiction est ensuite vérifiée

par le biais d'un protocole expérimental.

Tous les segments phonétiques sont associés aux gestes articulatoires. Ces derniers sont des mouvements contrôlés dans le but de former une configuration cible. Les segments phonétiques peuvent être produits par un ou plusieurs gestes linguaux (comme l'apex de la langue dans /t/ et l'apex et le dos de la langue pour /l/). L'engagement d'un ou plusieurs articulateurs linguaux dans la production d'un segment entraîne un déplacement des autres zones de la langue. L'effet de cette association forte entre les différents articulateurs linguaux (l'apex, la lame et le dos de la langue) augmente avec la rétraction de la langue ou avec une constriction linguale. Le concept du DAC est dérivé de l'hypothèse sur le degré de l'engagement des articulateurs de la parole dans la formation de la clôture ou la constriction, établie sur des informations obtenues dans des expériences de production de la parole. Les auteurs révèlent l'existence d'un lien important entre le dos de la langue et le DAC. Une échelle fondée sur le degré de l'engagement du dos de la langue a été suggérée, elle est appelée échelle de DAC. Sur celle-ci, la valeur de 1 indique un DAC minimal, celle de 3 un DAC maximal et 2 est un DAC intermédiaire.

Les alvéo-palatales /ʃ/ et /ɲ/, la palatale /i/, la vélaire /k/ et le /l/ sombre sont fortement contraintes et, par conséquent, elles ont un DAC maximal (DAC = 3). Cette valeur élevée du DAC est générée par l'engagement maximal du dos de la langue à la formation de la constriction ou de l'occlusion. Les labiales et le schwa sont légèrement contraints et de ce fait, ils ont un DAC minimal (DAC = 1). Cette valeur réduite du DAC est expliquée par l'implication faible ou absente du dos de la langue dans la production de ces sons. Des études, notamment Recasens (1985, 1984b), ont montré que les segments avec une position neutre de la langue sont facilement influencés par les segments adjacents (résistance coarticulatoire minimale).

Les alvéolaires ont un DAC intermédiaire (DAC = 2). Le dos de la langue n'est pas directement sollicité pendant la production d'un segment mais il est influencé par son association importante avec l'articulateur utilisé dans la formation de la constriction. Par exemple, l'articulation des consonnes alvéolaires /n, s/ engage l'élévation de la lame de la langue, induisant une élévation du dos de la langue. Les auteurs mettent en place trois scénarios qui prédisent l'étendue de l'influence consonantique sur la voyelle :

1. La consonne et la voyelle sont produites avec des positions articulatoires comparables. Les deux segments impliquent une élévation (comme dans /ɲi/ et /ni/) ou un abaissement (comme dans /la/) du dos de la langue. Peu d'influence de la consonne sur la voyelle est attendue pour /ɲi/ et /ni/ soit parce que les deux sons de la première séquence ont le même DAC élevé (DAC = 3) ou soit parce que la voyelle est plus contrainte que la consonne pour la seconde (DAC de /i/ = 3 et celui du /n/ = 2). Un impact de la consonne sur la voyelle est prévu lorsque la consonne est plus contrainte que la voyelle, comme c'est le cas dans la séquence /la/ (le DAC du /l/ = 3 et celui du /a/ = 2).
2. La consonne et la voyelle sont produites avec des positions articulatoires opposées, tel que dans /na/, /ɲa/ et /li/. Les sons /ɲ, n, i/ amènent à une élévation du dos de la langue alors que /l, a/ sont réalisés avec le dos de la langue abaissé. La séquence /ɲa/ doit donc

montrer un effet coarticulaire de la consonne sur la voyelle. La consonne a un degré de contrainte élevé (DAC = 3) tandis que la voyelle a un degré de contrainte moyen (DAC = 2). L'impact de la coarticulation de la voyelle sur la consonne est réduit lorsque les deux segments ont le même DAC, comme dans /na/ (DAC = 2) et dans /li/ (DAC = 3).

3. Un des deux segments adjacents a un DAC égal à 1. Dans ce scénario, un des segments n'expose pas d'activation du dos de la langue, tel que le /p/ dans la séquence /pi/. Cette combinaison doit présenter un effet minimal de la consonne sur la voyelle (ou de la voyelle sur la consonne, par exemple /də/).

L'influence de la voyelle sur la consonne et de la voyelle sur la voyelle (dans des séquences VCV) peut aussi être prédite par ce modèle. Les auteurs expliquent que le dos de la langue (et, dès lors, le DAC) est lié à la sensibilité coarticulaire<sup>30</sup> de la voyelle dans la littérature. Cette sensibilité diminue selon l'engagement du dos de la langue /ə/ > /a/ > /i/, pour les voyelles étudiées. Les voyelles avec un DAC élevé (et donc un degré de sensibilité bas) exercent une grande influence sur les consonnes avec un DAC bas. Un impact moyen est attendu sur les consonnes avec un DAC intermédiaire, tandis que peu d'effet est attendu sur les consonnes avec un DAC élevé. Les auteurs soulignent que la direction de la coarticulation (anticipatoire et persévérante) des consonnes varie en fonction des exigences des gestes consonantiques et détermine l'orientation des effets vocaliques. Certaines consonnes avec un DAC élevé exercent un effet plus important en coarticulation persévérante qu'en coarticulation anticipatoire, alors que l'inverse est vrai pour d'autres consonnes. Une consonne exigeant une composante anticipatoire forte permet plus d'effet anticipatoire que persévérant sur la voyelle. Autrement dit, lorsque l'articulation de la consonne nécessite des gestes anticipatoires importants, elle tolère peu la coarticulation persévérante de la voyelle. Le protocole expérimental, conçu afin de vérifier toutes les prédictions précédentes, consiste en la production de séquences VCV (où V = /a, i/ et C = /p, n, l, s, ʃ, ʒ, k/). Le contact linguo-palatal et la fréquence de F2 ont été mesurés pour ces séquences. Les valeurs des DAC ont été contrôlées à l'aide de l'*électropalatographie* (EPG)<sup>31</sup>, qui offre une observation précise de la surface de contact du dos de la langue avec le palais. Parallèlement, les analyses spectrales permettent de comparer les données articulatoires avec les données acoustiques. Les résultats obtenus de ce protocole confirment la validité du modèle du DAC. Les consonnes et les voyelles sont assignées à des valeurs différentes du DAC selon le degré de l'engagement du dos de la langue pendant leur articulation. Le DAC des alvéo-palatales, des palatales, des vélaires et du /l/ sombre est égal à 3, ce qui sous-entend que le dos de la langue est fortement sollicité. Les segments subissant une activité non-contrôlée du

30. La sensibilité coarticulaire indique la capacité d'un segment à être modifié par un segment adjacent (Recasens, 1985, 1991)

31. « L'électropalatographie repose sur l'utilisation d'un palais artificiel recouvert d'électrodes de contact. Un courant de très faible amplitude est injecté dans le corps du locuteur. La langue touchant une ou deux séries d'électrodes sert d'interrupteur du circuit électrique. On dispose ainsi en temps réel d'une information sur les zones du palais touchées par la langue. L'acquisition des données est contrôlée par un PC. L'affichage des données peut prendre plusieurs formes de trames EPG isolées ou successives du nombre de contact par zones, surimpression de trames ... » (Marchal et Legou 2009, page 92). Regarder aussi Stone (2010).

dos de la langue ont un DAC de 2 comme dans /n, a/. Lorsque l'activité du dos de la langue est minimale ou absente (pour les bilabiales ou le schwa), la valeur du DAC est de 1. Les résultats laissent suggérer que le mode d'articulation peut faire augmenter la valeur du DAC. Les prédictions de la coarticulation expliquées précédemment sont donc valables. Recasens et al. (1997) proposent que la formation de la clôture ou de la constriction nécessaire pour prononcer la consonne doit être réalisée pendant l'articulation de la voyelle adjacente. Ils émettent l'idée qu'un antagonisme maximal est produit dans une coarticulation de C-à-V afin de garantir que le geste consonantique soit réalisé et que, pour un DAC donné, les consonnes soient plus contraintes que les voyelles. Une corrélation entre l'effet de la coarticulation et la durée du segment a été obtenue. En effet, un son prononcé avec une durée plus longue exerce une influence plus grande qu'un son réalisé avec une durée réduite. Cette constatation est en accord avec le modèle de l'*undershoot*. La diminution du temps nécessaire à l'articulation d'un segment provoque l'échec de celui-ci à atteindre sa cible articulatoire. De fait, le segment perd de l'information et permet une coarticulation plus importante (Lindblom, 1963a). Les consonnes prononcées avec une élévation du dos de la langue, comme les alvéo-palatales et les vélares, favorisent la coarticulation persévérante sur celle anticipatoire. Ce phénomène est suscité par l'abaissement lent du dos de la langue au moment du relâchement consonantique, comparé à l'élévation plus rapide de celui-ci à l'*onset* de la consonne. Ceci est en lien avec les exigences mécano-inertielles de la langue. La direction coarticulatoire des autres consonnes dépend de l'action de la voyelle. Cette dernière peut conduire ou non à une élévation du dos de la langue durant la production de la consonne. Une coarticulation anticipatoire est notée lorsque la voyelle entraîne une élévation réduite du dos de la langue.

Recasens et Espinosa (2009) reprennent les mêmes consonnes et les mêmes voyelles (en y ajoutant, toutefois, la voyelle /u/) utilisées dans des séquences VCV examinées par Recasens et al. (1997) quelques années plus tôt. L'objectif est de comprendre la relation entre la résistance coarticulatoire et l'agression coarticulatoire de la consonne vers la voyelle et de la voyelle vers la consonne. L'agression coarticulatoire détermine l'influence qu'un segment fait subir à un autre. Les auteurs se sont posés deux questions : les sons très résistants doivent-ils être aussi agressifs ? Ceux moins résistants sont-ils moins agressifs ? Les données articulatoires ont été recueillies à l'aide d'un articulographe électromagnétique (EMA)<sup>32</sup> où les deux dimensions (verticale et horizontale) du mouvement lingual sont analysées pour chaque segment. Le déplacement bidimensionnel (x, y) des trois positions linguales est mesuré en centimètres : l'apex, la lame et le dos de la langue. Les valeurs acoustiques et articulatoires sont évaluées au milieu de chaque segment pour chaque locuteur. Pour obtenir la résistance coarticulatoire de la consonne C, la moyenne globale des trois contextes vocaliques est calculée, puis la moyenne des dis-

32. L'articulographie électromagnétique (EMA ou EMMA de l'anglais *Electromagnetic Midsagittal Articulatory*) « est une méthode qui permet d'observer l'évolution temporelle de la position de plusieurs articulateurs (les lèvres, la langue, la mâchoire et le voile du palais) à partir d'un suivi de points, avec une résolution spatiale inférieure au millimètre. Les données EMA permettent d'évaluer les mouvements haut-bas et également les mouvements avant-arrière des articulateurs » (Amelot 2009, page 71). Regarder aussi Stone (2010).

tances euclidiennes<sup>33</sup> entre les valeurs des positions (x, y) pour chaque contexte vocalique et la moyenne globale est mesurée. Quant à la résistance coarticulatoire des voyelles, elle est acquise tout d'abord par l'estimation de la moyenne globale des sept contextes consonantiques, puis par le calcul de la moyenne des distances euclidiennes entre les valeurs des positions (x, y) pour chaque contexte consonantique et la moyenne globale. Une valeur de distance euclidienne élevée (dispersion maximale) signifie une coarticulation minimale, tandis qu'une valeur réduite (dispersion minimale) indique une coarticulation maximale. La valeur de l'agression coarticulatoire est calculée pour chaque consonne sur le milieu de V1 et celle pour chaque voyelle est mesurée sur le milieu de C de chaque locuteur en utilisant la même méthode que celle de la distance euclidienne. L'agression coarticulatoire des consonnes est mesurée en calculant la moyenne des distances euclidiennes entre les valeurs des positions (x, y) de chaque consonne dans les trois contextes vocaliques et la moyenne globale. L'agression coarticulatoire des voyelles, quant à elle, est déterminée par le calcul de la moyenne des distances euclidiennes entre les valeurs des positions (x, y) de chaque voyelle dans les sept contextes consonantiques et la moyenne globale. Une dispersion importante (+ de distance euclidienne) signifie une agression coarticulatoire faible ; à l'inverse, une dispersion réduite (– de distance euclidienne) révèle une agression coarticulatoire élevée. Les résultats mettent en avant que les différences de variabilité linguale (de la voyelle sur la consonne ou de la consonne sur la voyelle) pour /p/ > /n/ > /l, k/ > /s/ > /ʃ, ɲ/ > et /u/ > /a/ > /i/ sont conformes à l'échelle de degré de contrainte proposée dans le modèle du DAC. La direction, verticale et/ou horizontale, du déplacement articulaire diminue dans l'ordre suivant : alvéo-palatales > alvéolaires > labiales. Il est aussi à remarquer que le mode d'articulation influence le degré de contrainte d'un segment. Par exemple, si deux segments ont le même lieu d'articulation /l/ vs. /s/, la consonne /s/ est plus contrainte que le /l/ parce que la friction est plus contrainte que la latéralité. L'élévation et l'antériorité de la langue pendant la production de la voyelle /i/ semblent contraindre largement toutes les régions de la langue. Ce phénomène conduit à une résistance coarticulatoire plus importante que les autres voyelles. Les résultats de l'agression coarticulatoire sont en accord avec le modèle du DAC. Les segments qui sont les plus résistants sont aussi les plus agressifs.

Recasens (2013) explique que la direction de la coarticulation est une composante importante du modèle du DAC. Les effets coarticulatoires d'un segment peuvent favoriser la direction anticipatoire (les effets de droite à gauche) ou la direction persévérante (les effets de gauche à droite). Par conséquent, un segment peut avoir un impact plus important sur le segment

33. La distance euclidienne est la distance géométrique dans un espace multidimensionnel, elle est calculée par la formule suivante :

$$distance(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (2.3)$$

(cf. Hill et Lewicki (2005)). C'est un autre moyen qui permet de mesurer la dispersion (variation) d'un segment dans les différents contextes. Par exemple, Recasens (1985) a employé l'écart-type dans ce but. Recasens et Espinosa (2009) ont utilisé l'écart-type et le coefficient de variation dans cette étude comme un indice de dispersion ; ceux-ci ne sont pas expliqués dans cet article parce qu'ils donnent les mêmes résultats que ceux obtenus par la distance euclidienne.

précèdent que sur celui qui suit ou inversement. La dominance d'une ou des deux directions coarticulatoires dépend des caractéristiques articulatoires de la voyelle ou de la consonne en question. Par exemple, le [ɸ] en position finale en anglais est réalisé avec l'abaissement du dos de la langue et avec le mouvement de rétraction, qui commencent souvent avant l'élévation de l'apex. Ce [ɸ] est plus prédominant en direction anticipatoire qu'en direction persévérante (les impacts de C-à-V). De plus, l'influence de la voyelle sur la consonne et sur la voyelle transconsonantique dans des séquences VCV semble être conditionnée par la direction des effets coarticulatoires exercés par la consonne sur les voyelles adjacentes. L'impact anticipatoire de la consonne sur la première voyelle interfère avec l'effet persévérant de celle-ci alors que l'influence persévérante de la consonne sur la seconde voyelle interagit avec l'effet anticipatoire de cette dernière. Le chercheur souligne que les consonnes qui favorisent une coarticulation anticipatoire C-à-V (comme le [ɸ]) doivent bloquer les effets de la coarticulation persévérante provoqués par la voyelle ; elles permettent donc une coarticulation anticipatoire de la voyelle plus importante qu'une coarticulation persévérante. D'un autre côté, les consonnes exerçant une dominance coarticulatoire persévérante vers la voyelle voisine (comme le [ɲ]) empêcheraient les effets anticipatoires induits par la voyelle suivante ; elles permettent ainsi une coarticulation persévérante de la voyelle plus importante que la coarticulation anticipatoire de celles-ci.

Une étude électropalatographique a été réalisée sur l'influence d'une consonne sur une autre consonne adjacente dans des séquences C1C2 en trois dialectes du catalan : majorquin, valencien et catalan oriental (Recasens, 2011). L'auteur teste si les effets de C2 sur C1 refléteraient une planification articulatoire pour préparer le phonème cible et si les effets de C1 et C2 seraient plus conditionnés par l'état instantané. Le lieu d'occlusion pour les consonnes relativement non-contraintes /t, n/ et /l/ clair devrait se rétracter systématiquement vers le lieu de constriction des consonnes subséquentes grandement contraintes /s, r, ʃ/ depuis le début de l'articulation de C1 dans C1C2. Dans les séquences où C1 est plus contrainte que C2, c'est-à-dire /s, r, ʃ/ suivi par /t, n/, C1 ne devrait pas être affectée par C2. Dans le même temps, C2 devrait s'adapter partiellement à C1 de telle façon que sa réalisation phonétique serait comparable à celle observable dans les autres positions. Le mode d'articulation de C1 pourrait contribuer d'une façon importante à la production finale de C2. Ainsi, le lieu d'articulation de la dentale C2 = /t/ devrait être plus aisément réalisé après C1 = /r/ qu'après /s, ʃ/, car la durée brève de /r/ laisse assez de temps à la pointe de la langue pour atteindre les dents. Les résultats indiquent que l'adaptation du lieu d'articulation dans les consonnes linguales antérieures dans les trois dialectes étudiés est conforme aux prédictions citées plus haut et ainsi au modèle du DAC.

L'adaptation des articulateurs dans les différents contextes (consonantique ou vocalique) souligne la flexibilité coarticulatoire de ces derniers pour assurer la réalisation du continuum du signal de parole. D'autres variations adaptatives peuvent également être produites afin de répondre aux exigences communicatives ou sociolinguistiques. La théorie H & H permet l'exploration de l'impact des variations adaptatives d'un point de vue communicatif sur la relation coarticulatoire entre les segments adjacents.

### 2.3.6 La théorie H&H

La théorie hyper et hypo-articulation (théorie H & H) de Lindblom (1990) est basée sur la relation entre les mécanismes de production et de perception de la parole. Lindblom explique que la perception de la parole implique la discrimination entre les items stockés dans le lexique de l'auditeur. L'accès lexical dépend ainsi de la discriminabilité du stimulus acoustique et non de l'invariance de celui-ci. Selon lui, la clarté articulatoire des locuteurs varie en fonction des exigences d'information des auditeurs. Autrement dit, selon le contexte, les locuteurs modifient leur manière d'articuler pour que les auditeurs puissent comprendre correctement leur message. Les locuteurs hyper-articulent lorsque les auditeurs ont des besoins importants en information acoustique et ils réduisent leur effort articulatoire (hypo-articulent) quand les auditeurs complètent l'information de l'entrée acoustique par d'autres sources, par exemple le contexte lexical (Figure 2.15). Ils hypo-articulent uniquement quand les auditeurs sont capables de distinguer la cible par les différents items lexicaux. Cette théorie représente un prolongement de son modèle de l'*undershoot* selon lequel l'invariant n'existe pas. Il explique que l'inexistence de l'invariant du signal de parole est une conséquence de l'organisation adaptative entre le locuteur et l'auditeur. Lindblom (1990) s'interroge sur la relation entre les locuteurs et les auditeurs, plus particulièrement sur les processus utilisés par ces deux groupes de personnes pour réussir à converser de manière intelligible. Il met en avant deux critères articulatoires que les locuteurs mais également les auditeurs prennent en compte : la plasticité et l'économie. Le locuteur change sa façon de parler (en allant de l'hypo- à l'hyper-articulation) en s'adaptant à l'auditeur et *vice versa* ; le critère de plasticité est donc utilisé. Lorsqu'un locuteur produit de la parole, il a tendance à faire le moins d'effort possible (si le contexte le permet) pour transmettre le message, le critère de l'économie est alors employé. Les locuteurs planifient de prononcer des énoncés qui sont suffisamment intelligibles pour l'auditeur ; autrement dit, le locuteur économise de l'effort articulatoire sans oublier l'intelligibilité. De plus, les locuteurs font une estimation instantanée du besoin de l'information acoustique de l'auditeur et adaptent leur énoncé selon ces besoins (Harrington, 2010). Il existe ainsi un système d'équilibre entre locuteur et auditeur. La conséquence directe des arguments apportés par Lindblom, concernant les interactions entre les locuteurs et les auditeurs, est l'abandon de la recherche d'une simple unité primitive de perception de la parole. Le locuteur et l'auditeur sont en relation de coopération continue, ils s'harmonisent réciproquement aux différentes situations conversationnelles. La continuité existant entre la forme hypo-articulée et la forme hyper-articulée explique en partie les deux mécanismes de production et de perception de la parole. Lindblom (1996) signale que la perception de la parole (mais aussi la perception en général) réussit à traiter des informations partielles grâce à l'interaction entre le signal et la connaissance ("*signal-plus-knowledge*"). Il ajoute que la perception de la parole ne présuppose pas l'existence d'un invariant dans le signal acoustique. Le seul invariant possible dans la production de la parole est l'intention du locuteur à faire passer un message à l'auditeur. Le locuteur "idéal" s'engage à ce que ce message soit clair et compréhensible pour l'auditeur.

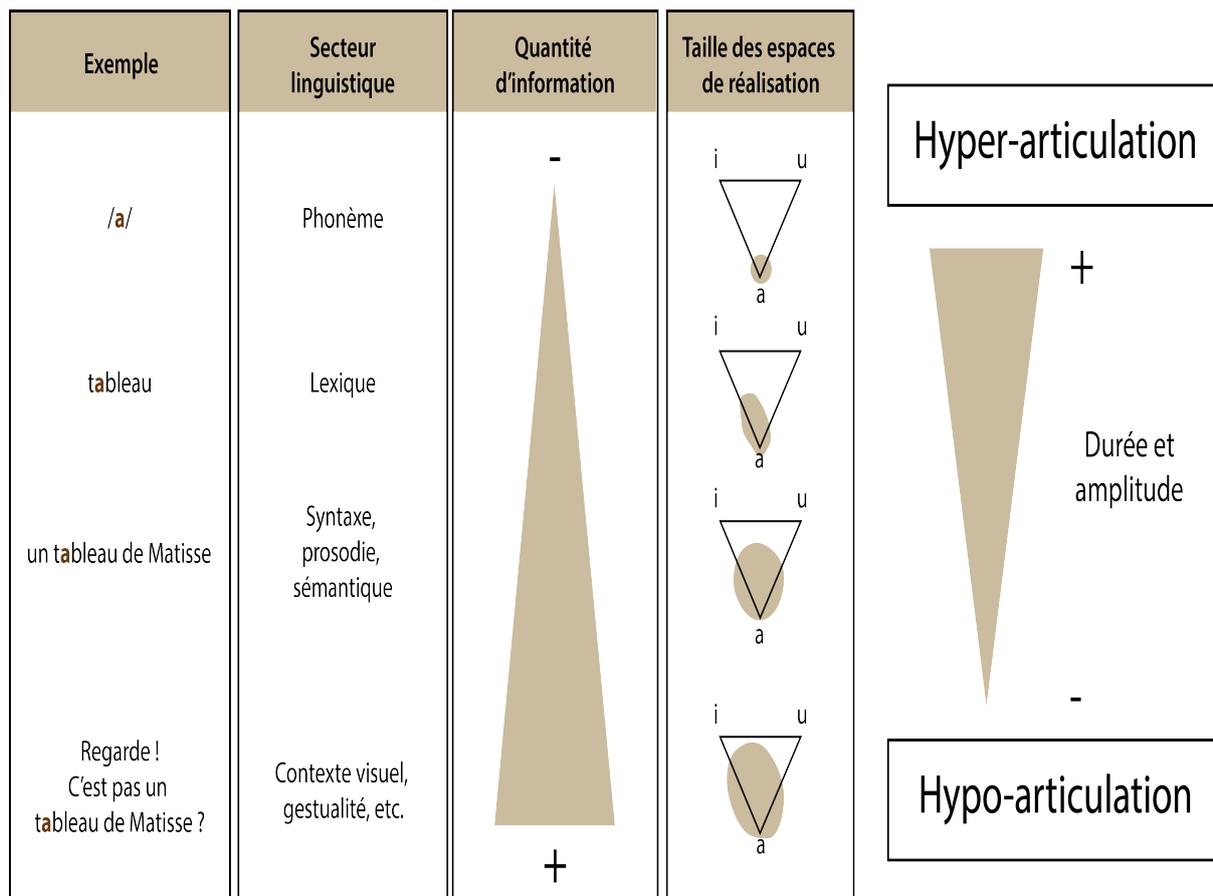


FIGURE 2.15 – Représentation de la variation phonétique en fonction de la quantité d'information contextuelle et de la relation entre l'hypo- hyper-articulation, la durée et la force de la production des segments (+ de durée/d'amplitude = + d'hyper-articulation) (d'après Al-Tamimi (2007), inspirée de Meunier (2005), adaptée de Lindblom (1990).

Lindblom (1996) explique que le fait "parler clairement" n'est pas qu'une simple augmentation de l'effort musculaire du conduit vocal, c'est un processus d'organisation adaptative que le locuteur ajuste pour faciliter la compréhension de son message. En effet, « [...] *clear speech owes both its existence and its phonetic characteristics to the way the production and perception of speech work. Clear speech is a manifestation of the adaptive organization of speech production which is evident in the interplay between the goals of satisfying the listener's informational needs, on the one hand, and of minimizing articulatory complexity, on the other. Accordingly, clear speech appears perfectly compatible with the present claims, namely, that the listener and the speaking situation define the speaker's task, that that task varies, and that the speaker is capable of adapting to it also at the level of individual phonetic segments* »<sup>34</sup>.

34. « [...] la parole claire doit son existence et ses caractéristiques phonétiques à la façon dont la production et la perception de la parole fonctionne. La parole claire est une manifestation de l'organisation adaptative de la production de la parole qui cherche à satisfaire les besoins informationnels de l'auditeur d'un côté et à minimiser la complexité articulaire d'un autre côté. En conséquence, la parole claire semble être parfaitement compatible avec les propositions présentes, notamment celle où l'auditeur et la situation de la parole définissent la tâche du locuteur. Cette dernière varie et le locuteur est capable de s'adapter à elle ainsi qu'à chaque segment phonétique » page 1688, [notre traduction].

Le test de *bite-block* est utilisé afin d'examiner la validité de cette théorie. Dans ce test, le locuteur a pour tâche de parler en mettant un objet dans la bouche, un stylo par exemple, l'empêchant d'articuler normalement. Le test de *bite-block* apporte la preuve que le locuteur s'adapte à la situation et arrive à faire comprendre son message à l'auditeur malgré la difficulté. La même situation est vraie du côté de l'auditeur qui réalise un effort à son tour pour comprendre le message qu'il entend. Le locuteur a donc le choix d'adapter et de réorganiser ses articulateurs en fonction du besoin dans l'intention de transmettre son message. Quant à l'auditeur, il se contente d'un minimum d'informations acoustiques qui lui permet d'avoir accès au message voulu par le locuteur ; autrement dit, l'auditeur cherche le contraste suffisant qui ait un pouvoir discriminatif permettant un accès lexical correct. Pour cette raison, Lindblom met en exergue qu'aucun niveau (articulatoire, acoustique ou perceptif) n'est privilégié. Par contre, il soumet l'idée selon laquelle le niveau perceptif peut avoir la priorité car il porte toute l'information sensorielle cruciale à l'identification de la parole.

L'auteur expose que les locuteurs favorisent davantage l'utilisation de la forme hypo-articulatoire dans leurs conversations, plus particulièrement avec les mots fréquents qui sont mieux prédictibles par les auditeurs. Ces derniers exigent ainsi peu d'informations acoustiques afin d'accéder à ces mots. Cette forme d'hyper-articulation correspond à une réduction articulatoire où une cible n'est pas atteinte (un *undershoot*), comme c'est le cas dans une articulation non-accentuée ou dans une prononciation en débit rapide. Cette réduction articulatoire (vocalique ou consonantique) n'est généralement pas perceptible. Sur le plan acoustique, l'interaction entre les gestes et la durée très courte des segments conduit à l'absence ou à la perte de l'information acoustique. Le manque de temps entraîne l'absence d'élaboration des gestes (Lindblom, 1963a). La parole hyper-articulée est souvent associée à moins de centralisation et moins de co-articulation ; autrement dit, des fréquences formantiques périphériques sont attendues pour les voyelles ayant peu de chevauchement coarticulatoire avec les consonnes. Les mots fréquents sont ainsi moins souvent hyper-articulés. Aylett et Turk (2006) ont examiné l'influence de la prédictibilité des mots sur la fréquence des F1 et F2 des voyelles dans un grand corpus, c'est-à-dire l'impact du contexte lexical sur le contenu spectral des voyelles (voir aussi Richard (2004)). Les résultats indiquent une centralisation importante des fréquences pour les voyelles qui sont prononcées dans des mots ayant une prédictibilité élevée. Ces résultats soutiennent la théorie H&H car le locuteur effectue une économie en produisant les mots fréquents. Cette économie est justifiée par la diminution du besoin en informations acoustiques de l'auditeur due à l'accès lexical facile de ces mots. Une économie de l'effort articulatoire génère une "non-atteinte" des cibles articulatoires et par conséquent, les valeurs fréquentielles des formants se dirigeront vers une zone centrale. En même temps, les locuteurs prédisent que les auditeurs auront besoin de plus d'informations acoustiques pour les mots moins fréquents, incitant les locuteurs à hyper-articuler les mots ou à contrôler les articulateurs de manière à ce qu'ils atteignent leurs cibles articulatoires. Ce phénomène conduit à des valeurs formantiques périphériques.

Les variations des degrés du chevauchement coarticulatoire, provoquées par les modifica-

tions adaptatives dues aux contextes (vocalique et consonantique) et / ou aux exigences communicatives, peuvent être captées par un outil dynamique appelé *équations de locus*.

### 2.3.7 Les équations de locus

Les équations de locus sont des modèles linéaires<sup>35</sup> exprimant la relation de dépendance entre les configurations articulatoires réalisées dans la production d'une séquence consonne – voyelle coarticulée dans des séquences CV ou VC (Lindblom, 1963b). Ces équations sont de la forme :

$$F2_{onset} = k \times F2_{mid} + c \quad (2.4)$$

Où les coefficients  $k$  et  $c$  sont respectivement la pente et l'ordonnée à l'origine. Elles sont issues d'une régression linéaire appliquée à la relation entre la fréquence du deuxième formant mesurée au milieu de la voyelle ( $F2_{mid}$ ) et celle de ce même formant prélevée au moment du relâchement de la consonne ( $F2_{onset}$ ). La fréquence du  $F2_{mid}$  est conventionnellement placée en abscisse et celle du  $F2_{onset}$  est placée en ordonnée. Afin d'obtenir une équation de locus (LE pour Locus Equation), des paires de fréquences  $F2_{mid} - F2_{onset}$  ont été extraites sur de nombreuses réalisations individuelles impliquant une catégorie consonantique unique produite en contexte avec différentes voyelles (Figure 2.16).

La plupart des travaux ont porté sur les occlusives (*cf.* par exemple Sussman et al. (1991)) même si certains se sont intéressés à des propriétés qui seraient généralisées à différents modes d'articulation (Sussman et al., 1996), voire aux propriétés qu'elles pourraient présenter en lien avec les voyelles (Al-Tamimi, 2007). Deux approches parallèles cherchent à rendre compte de l'interprétation théorique des équations de locus : la "force" de la coarticulation et l'indication du lieu d'articulation. Dans la lignée du travail séminal de Lindblom (1963a), les équations de locus permettent d'exprimer la « force » de la coarticulation (le degré de coarticulation) entre la consonne et la voyelle (Krull, 1989; Duez, 1992; Iskarous et al., 2010). En théorie, si aucune coarticulation n'existe entre la consonne et la voyelle, la réalisation de la consonne est totalement indépendante de celle de la voyelle. Changer la réalisation de la voyelle qui est coarticulée avec la consonne ne devrait donc avoir aucun impact sur le locus consonantique (Delattre et al., 1955), lequel devrait être stable quelle que soit la fréquence du formant vocalique. La pente exprimant la relation entre les deux valeurs devrait être égale à 0. À l'inverse, si les configurations du conduit vocal associées à la réalisation de la consonne et de la voyelle sont dans une interdépendance massive, une relation "parfaite" serait observée entre ces deux valeurs. Par exemple, un changement de 100 Hz dans le formant de la voyelle correspondrait à un changement de 100 Hz dans le locus de la consonne. La pente de l'équation serait alors égale à 1. Les valeurs des pentes des équations de locus mesurées se distribuent donc, en théorie, entre ces deux va-

35. Le terme "modèle", ici employé, doit être compris comme un modèle pratique et non théorique. Autrement dit, elles peuvent être considérées comme un modèle qui met en pratique les principes exploités dans la théorie du locus sur l'importance de la transition du F2. C'est pourquoi le terme "modèle" sera par la suite utilisé.

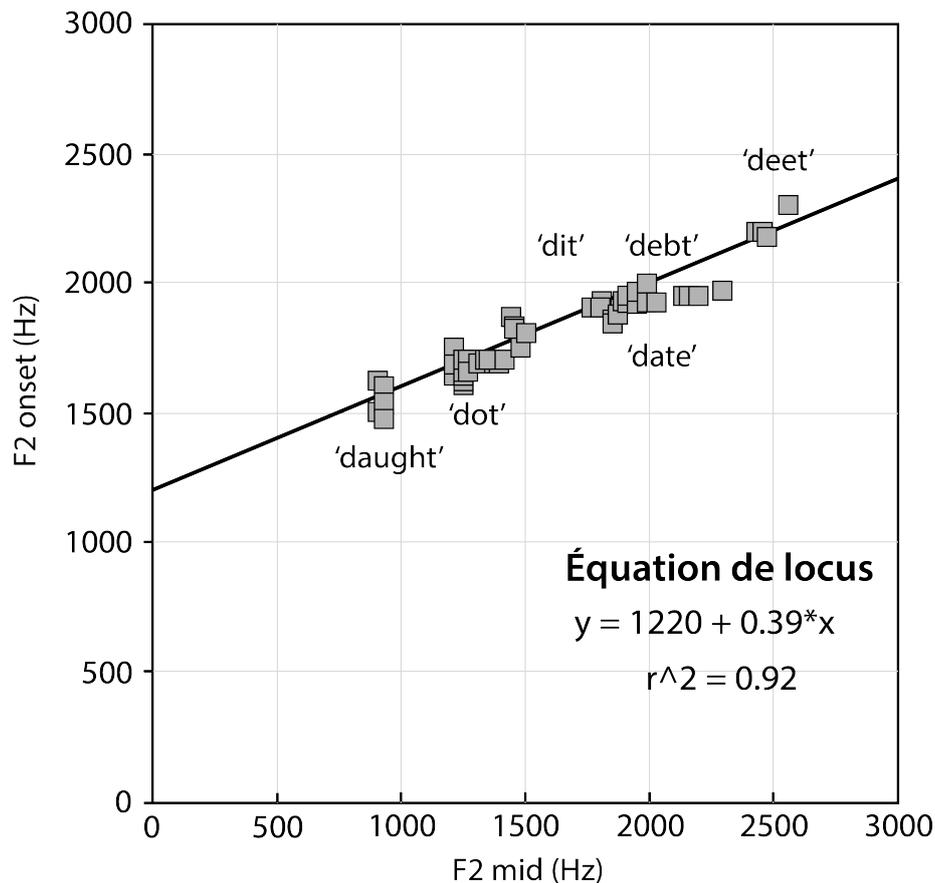


FIGURE 2.16 – Représentation de l'équation de locus de la consonne /d/ dans les séquences dVt (d'après Lindblom et Sussman (2012)).

leurs (0 et 1) : plus la pente est élevée, plus le degré de coarticulation est important. Un autre ensemble de recherches s'est attaché à évaluer la relation entre les propriétés des équations de locus (pente mais aussi ordonnée à l'origine) et les catégories phonologiques (lieu d'articulation notamment) (entre autres, Nearey et Shammass (1987); Sussman et al. (1991, 1996)).

Sussman et al. (1991) montrent que les coefficients de l'équation de locus ne sont pas seulement un indicateur de degré de coarticulation mais aussi un bon descripteur du lieu d'articulation des occlusives. Ainsi, sur la base de la valeur de pente de la droite de régression obtenue sur les trois lieux d'articulation les plus étudiés, les vélares révèlent un degré de coarticulation relativement fort (pente élevée autour de 0.9), les alvéolaires ont quant à elles un degré de coarticulation relativement faible (autour de 0.4) et celui des bilabiales est intermédiaire (environ 0.7). Ils formulent alors l'hypothèse que la relation linéaire entre  $F2_{onset}$  et  $F2_{mid}$  telle qu'elle est exprimée par les équations de locus représente, sinon un invariant, un indice pour décrire le lieu d'articulation des occlusives. Si les travaux associés à cette problématique se réfèrent également à la question du taux de coarticulation CV, ils orientent plus fortement le débat sur la « fonction » linguistique de ces équations. Alors que la relation entre classes sonores et équations de locus soulève des débats (notamment Sussman et al. (1998b); Iskarous et al. (2010); Lindblom et Sussman (2012)), cette possibilité d'une catégorisation des données acoustiques associée aux

lieux d'articulation suscite également des interrogations sur leur possible fonction perceptive (Brancazio et Fowler (1998), par exemple). En effet, les séquences individuelles se distribuent dans des espaces acoustiques relativement distincts et des analyses discriminantes mettent en évidence l'impact des mesures associées sur la classification statistique des séquences (Sussman et al., 1991). Si ces paramètres ne sont ni « invariants » ni suffisants (Brancazio et Fowler, 1998), ils contribueraient de manière non négligeable à la classification (Sussman et al., 1998b). Selon Fowler (1994), les équations de locus ne fournissent pas d'informations permettant de spécifier le lieu d'articulation (voir aussi Iskarous et al. (2010)). Par contre, elles caractériseraient les propriétés de chevauchement coarticulatoire, qui peut varier non seulement avec la résistance coarticulatoire due au lieu ou mode d'articulation, mais aussi avec l'accent tonique. Selon Fowler (1994), les équations de locus ne peuvent donc pas être utilisées comme des descripteurs du lieu d'articulation. À l'inverse, Sussman et al. (1996) ont estimé des équations de locus dérivées de séquences CV/t/ dans lesquelles le mode d'articulation et le voisement de la consonne initiale variaient chez 22 locuteurs américains ( $C \in /d, t, n, z, s/$ ; cette consonne étant associée à 10 voyelles différentes). Les pentes des équations de locus mesurées pour des séquences contenant des consonnes ayant le même lieu d'articulation (alvéolaire uniquement) et différents modes d'articulation (voisement, nasalité et friction) seraient similaires. Les équations de locus seront développées dans **le chapitre 3**.

En résumé, les équations de locus offrent un moyen dynamique pour décrire le chevauchement coarticulatoire, tandis que leur fiabilité à désigner le lieu d'articulation reste une source de débat. Un autre modèle, décrivant la relation coarticulatoire entre les segments d'un point de vue dynamique, est le modèle *task-dynamics*.

### 2.3.8 Le *Task-dynamics*

Le modèle *task-dynamics* offre une explication systématique et générale du contrôle du mouvement qui a été créé à l'origine pour décrire les tâches effectuées par les membres du corps humain (Saltzman et Kelso, 1983). Il a ensuite été étendu à l'assimilation de la production de la parole (Saltzman, 1985; Saltzman et Munhall, 1989). Ce modèle est développé à partir de la théorie de la coproduction et est élaboré grâce au travail collaboratif entre psychologues et linguistes (Saltzman et Kelso, 1983; Saltzman, 1985; Saltzman et Munhall, 1989; Browman et Goldstein, 1989; Fowler et Saltzman, 1993). Il est basé sur les principes biologiques et physiques généraux des mouvements coordonnés mais il est exprimé plutôt en termes dynamiques que physiologiques. Ce modèle présente une analyse détaillée de la production de la parole et de la relation coarticulatoire entre les articulateurs où les mouvements complexes sont divisés en ensemble de tâches fonctionnelles indépendantes.

Saltzman et Kelso (1983) définissent que « *The term task dynamics follows directly from the view (1) that the degree of freedom comprising action units are constrained by the particular tasks that animals perform, and (2) that action units are specified in the language of dynamics,*

*not, as is more frequently assumed, in terms of kinematic or muscular variables. Thus we propose, and seek to elaborate here, an invariant control structure that is specified dynamically according to task requirements and that gives rise to diverse kinematic consequences* »<sup>36</sup>.

Les activités répétitives ordinaires (par exemple, marcher, parler, écrire) ou extraordinaires (par exemple, danser, chanter, courir) qu'un être humain effectue semblent être invariantes malgré les variations observées de la performance. Lorsqu'une action est effectuée, elle n'est jamais reproduite à l'identique, cependant, l'organisation générale de cette action est perçue comme étant invariante. Le modèle *task-dynamics* propose une compréhension des aspects variables et stables des actions. Le terme "task" désigne les performances des tâches à accomplir dans la vie réelle tandis que le terme "dynamics" est utilisé afin de donner une explication dynamique des forces qui provoquent des mouvements cinématiques observables (Saltzman, 1991). La dynamique permet la réalisation et l'achèvement de ces tâches. Une description dynamique est préférable parce qu'elle fournit une considération de la forme du mouvement, de la stabilité et de la déformation légitime de cette forme. Selon ce modèle, un seul mouvement peut impliquer plusieurs tâches discrètes et abstraites. La production de la parole exige une succession de mouvements dont chacun est modulé comme un ensemble de tâches. Par exemple, pour prononcer la consonne /f/, la langue est élevée, les lèvres sont légèrement arrondies et les plis vocaux sont écartés pour assurer le non-voisement et le passage du flux d'air. Il existe au moins six tâches distinctes engagées dans la production d'un /f/ en isolation : garder la porte vélo-pharyngée fermée, contrôler le degré et le lieu de constriction de la langue, le degré de protrusion des lèvres, la taille de l'aperture des lèvres et la taille de l'aperture glottale (Hawkins, 1992). Un même articulateur peut être sollicité pour plusieurs tâches. Dans l'exemple précédent, la mâchoire contribue à la réalisation de la bonne aperture des lèvres et d'une constriction linguale convenable.

Deux notions importantes sont à prendre en compte lors du traitement de ce modèle. Il s'agit des **gestes phonétiques** (dits aussi gestes articulatoires ou tout simplement gestes) et de la **coordination**. Les *gestes phonétiques* sont des actions linguistiquement signifiantes des structures du conduit vocal. Selon la phonologie articulatoire, les gestes sont considérés comme les éléments primitifs du système phonologique sur le plan linguistique (Browman et Goldstein, 1986, 1989, 1990a). Physiquement parlant, ils sont des mouvements coordonnés du conduit vocal accomplissant un but phonétique. Fougeron (2005) définit les gestes articulatoires de cette façon : « Un Geste articulatoire est une unité d'action : l'action de formation et de relâchement d'une constriction à un endroit spécifique dans le conduit vocal. Le mot pas, par exemple, commence par un geste d'occlusion labiale, alors que le mot cas commence par un Geste d'oc-

---

36. « Le terme *task dynamics* provient directement des idées suivantes : (1) le degré de liberté constituant les unités d'actions sont contraintes par des tâches particulières que les animaux performant et (2) les unités d'action sont spécifiées d'un point de vue dynamique et non pas, comme il est souvent supposé, en termes de variables cinématiques ou musculaires. Nous proposons ainsi, et cherchons d'élaborer ici, une structure de contrôle *invariante* qui est spécifiée dynamiquement selon les exigences de la tâche et qui donne naissance à des conséquences cinématiques diverses », page 4 [notre traduction].

clusion du corps de la langue. Un Geste va alors se caractériser par (1) des informations sur les articulateurs qui le forment, (2) des informations sur la constriction qui est son but (sa tâche), et (3) des paramètres dynamiques spécifiant comment cette constriction est faite »<sup>37</sup>. Dans ce modèle, un geste et un mouvement n'ont pas la même signification. En effet, tous les gestes articulatoires sont composés de mouvements articulatoires, toutefois, tous les mouvements ne peuvent pas être interprétés comme des gestes (Fowler et Saltzman, 1993). Par exemple, lors de la réalisation de la clôture bilabiale, le mouvement qui résulte des changements de la distance verticale entre la lèvre supérieure et la lèvre inférieure, provoqués par la coordination active entre les lèvres et la mâchoire, est considéré comme un geste articulatoire. Quand la distance entre les lèvres change, il est question d'une conséquence passive de la participation active de la mâchoire dans la réalisation des différents gestes (par exemple, le geste alvéolaire). Dans ce cas, le mouvement bilabial ne correspond pas à un geste articulatoire. *La coordination* est mise en œuvre par des contraintes qui lient les gestes phonétiques créant ainsi des dépendances entre les articulateurs. Celles-ci réduisent les degrés de liberté des articulateurs qui sont engagés à la réalisation du geste pour limiter les articulateurs à un ensemble de mouvements possibles accomplissant un geste phonétique. Par exemple, afin de réaliser la fermeture bilabiale, les locuteurs établissent des contraintes coordinatives entre la mâchoire et les deux lèvres restreignant ces articulateurs à un ensemble de mouvements qui assurent la clôture bilabiale.

Généralement, une tâche est un geste impliquant le contrôle d'un objet ou d'une représentation abstraite de ce qui est en train d'être contrôlé. Dans un sens abstrait, la finalité d'une action sur une cible est visualisée avant son exécution, comme la position de la main atteignant un objet ou la réalisation d'une constriction articulatoire dans une tâche de production de la parole. De tels objets, appelés **variables du conduit vocal** (*tract variables*), sont des parties du conduit vocal qui permettent aux constriction d'être spécifiées (Figure 2.17). Dans la production de la parole, la tâche est définie en termes de zones (lieux) et de domaines transversaux (degrés) d'une constriction idéale dans le conduit vocal. La tâche non-spécifiée est alors transformée en une constriction spécifique du conduit vocal. Par exemple, le degré d'ouverture bilabiale appropriée à une configuration des lèvres donnée est considéré dans le modèle *task-dynamics* comme une exigence pour achever une apertures labiale particulière et non comme une nécessité à ce que les lèvres et la mâchoire atteignent une certaine position. Dans ce cas, les lèvres sont considérées comme les dispositifs terminaux ou les effecteurs finaux parce que leur position définit directement l'aperture des lèvres. Les lèvres supérieure et inférieure et la mâchoire forment ensemble le système effecteur<sup>38</sup>, c'est-à-dire l'ensemble des organes qui comprend le dispositif terminal et le dirige vers le lieu exact au moment adéquat. Le *task-dynamics* commence donc

---

37. Page 268.

38. Le **système effecteur** est l'ensemble des membres du corps ou des organes phonatoires utilisés dans une action donnée. Un **dispositif terminal** ou un **effecteur final** est une partie d'un système effecteur qui réalise directement une action cible. Par exemple, dans une tâche de *tasse-à-la-bouche*, la tasse saisie est l'effectif terminal et le système arme-bras est le système effecteur ; dans une tâche de production d'une voyelle, la surface du corps de la langue est l'effectif terminal et le système mâchoire-langue est le système effecteur (Saltzman et Kelso, 1983).

par définir les mouvements en termes de tâches spatiales abstraites utilisant des coordonnées spatiales et des équations de mouvement qui sont naturels pour la tâche. Afin d'effectuer les transformations entre les espaces de la tâche et ceux de l'articulateur, le modèle *task-dynamics* traduit chaque équation d'une tâche spécifique en équations équivalentes. Une fois appliquées, ces dernières permettent d'obtenir les mouvements des parties spécifiques du conduit vocal (Kelso et al., 1986; Saltzman et Munhall, 1989). La première transformation est le passage de l'idée de la tâche à accomplir de l'espace abstrait à sa réalisation dans un espace du conduit vocal plus spécifique. L'équation d'une tâche spécifique est définie uniquement selon le type de tâche : le lieu et le degré d'une constriction non-spécifiée. La transformation à l'espace du conduit vocal spécifie la tâche actuelle : les variables du conduit vocal sont ainsi engagées pour déterminer une constriction spécifique comme l'aperture des lèvres ou le lieu d'une constriction formée par le corps de la langue. La seconde transformation correspond au changement d'un niveau à un autre, autrement dit de l'espace du conduit vocal à l'espace de l'articulateur (Figure 2.17). En effet pour mettre en action les variables du conduit vocal, les articulateurs engagés sont activés. Il existe donc plusieurs degrés de liberté ou plusieurs moyens pour que les articulateurs composants achèvent la même trajectoire.

Le modèle *task-dynamic* est basé sur trois questions principales. La première concerne la coordination inter-articulateurs durant un geste unique, par exemple, la coordination entre les lèvres et la mâchoire pendant la clôture bilabiale. La deuxième traite le mélange inter-gestuel lors du chevauchement de plusieurs gestes en temps (la coproduction) et lors du chevauchement des gestes partageant des articulateurs en commun. Par exemple, dans le cas du mélange dans les séquences /VbV/, un mélange d'influence existerait entre les contrôles bilabial et vocalique qui partagent la mâchoire. La troisième relève de la temporalité inter-gestuelle relative, à titre d'exemple, le temps relatif entre la fermeture-ouverture bilabiale et l'abduction-adduction laryngale pour /p/ (Saltzman, 1991). Des évidences sur *la coordination inter-articulateurs* peuvent être remarquées lors de la réorganisation des gestes qui subissent une perturbation afin d'accomplir un but articulatoire. Si une trajectoire qui mène à une cible a été bloquée, le système moteur trouve une trajectoire alternative. Un exemple de ce comportement a été étudié par Kelso et al. (1984) où les mouvements des lèvres supérieure et inférieure et de la mâchoire ainsi que l'élévation de la langue ont été mesurés. Une force de charge constante inattendue est appliquée pendant le mouvement vertical de la mâchoire pour la fermeture du /b/ final dans la séquence /bæb/. Une compensation quasi-immédiate<sup>39</sup> est révélée dans les lèvres par un abaissement compensatoire de la lèvre supérieure et une élévation compensatoire de la lèvre inférieure. Aucune interaction n'est remarquée pour la langue. La même perturbation appliquée pendant la réalisation de la séquence /bæz/ provoque une augmentation rapide de l'activité de la langue durant la production de /z/ et une absence de participation compensatoire des lèvres. Les auteurs expliquent que, tandis que la perturbation de la mâchoire représente un obstacle dans la réalisation des séquences, aucune distorsion perceptible de la parole n'est notée. Ces résultats

39. La réponse compensatoire est de 20 à 30 ms après la perturbation.

Variable du conduit vocal		Articulateur
<b>LP</b>	protrusion des lèvres	Lèvres inférieure et supérieure
<b>LA</b>	aperture des lèvres	Lèvres inférieure et supérieure, mâchoire
<b>TDCL</b>	position de la constriction du dos de la langue	Dos de la langue, mâchoire
<b>TDCD</b>	degré de la constriction du dos de la langue	Dos de la langue, mâchoire
<b>LTH</b>	élévation des dents inférieures	Mâchoire
<b>TTCL</b>	position de la constriction de l'apex de la langue	Apex et dos de la langue, mâchoire
<b>TTCD</b>	degré de la constriction de l'apex de la langue	Apex et dos de la langue, mâchoire
<b>VEL</b>	aperture du voile du palais	Voile du palais
<b>GLO</b>	aperture glottale	Glotte

FIGURE 2.17 – En haut, un tableau montrant la relation entre les variables du conduit vocal et les articulateurs. En bas, une représentation schématique sagittale du conduit vocal avec les degrés de liberté des variables du conduit vocal indiqués par des flèches (d'après Saltzman (1991)).

fournissent un soutien à l'ensemble flexible des structures coordinatives dans la production de la parole. La structure du modèle *task-dynamics* permet une compensation immédiate de manière efficace. Le contrôle est spécifié dans le niveau des variables du conduit vocal. À partir de ce niveau, il existe une modélisation des équations qui gouvernent les mouvements de chaque articulateur. Puisque la modélisation est mise à jour continuellement, les petites perturbations d'un articulateur seront immédiatement et automatiquement ajustées en répondant aux demandes des cibles des variables du conduit vocal qui gouvernent le système effecteur. Il n'est donc pas nécessaire d'utiliser de commandes explicites, la compensation se produira dans un intervalle de temps réduit.

Une des tâches majeures de la production de la parole est la création et le relâchement des constrictions de degrés différents dans diverses zones du conduit vocal. Dans le modèle *task-dynamic*, de telles constrictions sont contrôlées par un système dynamique à deux fonctionnalités distinctes mais à deux niveaux qui interagissent. L'un est le niveau de *la coordination inter-gestuelle* défini selon un ensemble de coordonnées d'activations gestuelles. L'autre correspond au niveau de *la coordination inter-articulateurs* caractérisé selon les coordonnées des

articulateurs et des variables du conduit vocal (Figure 2.18).

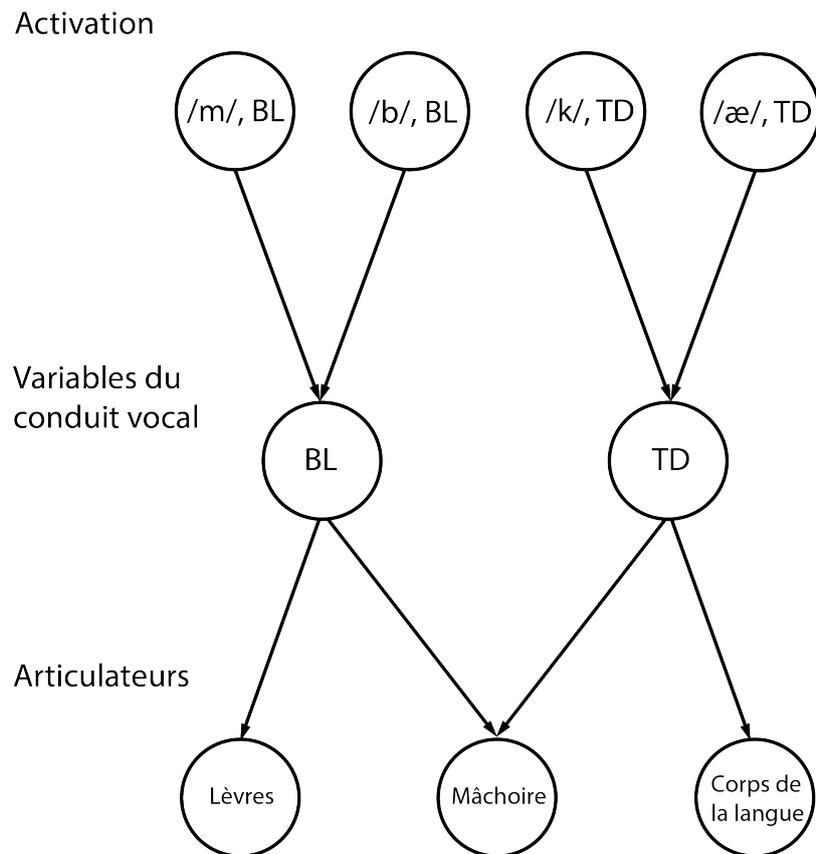


FIGURE 2.18 – Exemple des relations architecturales définies entre les articulateurs, les variables du conduit vocal et les systèmes des activations coordonnées. BL et TD représentent les variables du conduit vocal associées aux constrictions bilabiale et linguale, respectivement. Les gestes au niveau de l'activation sont désignés en termes d'identité linguistique (ex.  $/k/$ ) et ceux au niveau du conduit vocal sont nommés selon l'affiliation de leurs variables (ex. TD) (d'après Saltzman (1991)).

L'identité de linguistique de chaque unité gestuelle est définie indépendamment du contexte par ses coordonnées d'activation et celles de l'ensemble des variables du conduit vocal et par les articulateurs. Pour réaliser un geste, les activités de tous les articulateurs qui contribuent aux variables du conduit vocal sont coordonnées. Les composants des systèmes effecteurs fonctionnent comme des unités fonctionnelles et ils sont spécifiés par des équations dynamiques. L'aperture "zéro" des lèvres qui est nécessaire pour la production de la fermeture du  $/b/$  est accomplie par la combinaison de l'élévation de la mâchoire et de la lèvre inférieure avec l'abaissement de la lèvre supérieure. Dans une séquence comme  $/aba/$ , la mâchoire est abaissée lors de la prononciation des deux  $/a/$  et elle est élevée au cours de la réalisation du  $/b/$ . Dans une séquence comme  $/ibi/$ , elle sera plus élevée durant l'articulation des  $/i/$ . Par conséquent, les lèvres montrent des activités différentes de manière à achever l'aperture (zéro) du  $/b/$  dans les deux séquences. La tâche est donc identique pour  $/b/$  dans  $/aba/$  et dans  $/ibi/$  et les variables du conduit vocal sont les mêmes également. Toutefois, les détails physiques de comment la tâche est achevée sont différentes (Hawkins, 1992).

Les mouvements des articulateurs sont gouvernés par une série dynamique de gestes discrets et abstraits qui se chevauchent en temps et en espace. L'activation de ces gestes a une forme d'ondes avec un temps fixe. Les ondes de l'activation présentent une mise en œuvre graduelle montante jusqu'au sommet (activation maximale) suivie par une relaxation graduelle descendante jusqu'à l'état d'équilibre (activation minimale) (Figure 2.19). L'activation gestuelle est spécifiée dans ce modèle dans le niveau de la coordination inter-gestuelle comme un ensemble de fonctions simples dont les valeurs changent entre 0 (geste inactif) et 1 (geste activé au maximum). Les intervalles de coarticulation d'un geste sont définis pendant l'exécution (zone de coarticulation anticipatoire) ou pendant la relaxation (zone de coarticulation persévérante) des ondes d'activation des gestes. En conséquence, les gestes ne sont pas modifiés quand ils sont actualisés au cours de la production de la parole. Par contre, la structure temporelle intrinsèque des gestes accepte un chevauchement temporel lors de la parole et donc les gestes ne sont pas alternés par le contexte mais ils sont coproduits avec le contexte (Fowler et Saltzman, 1993).

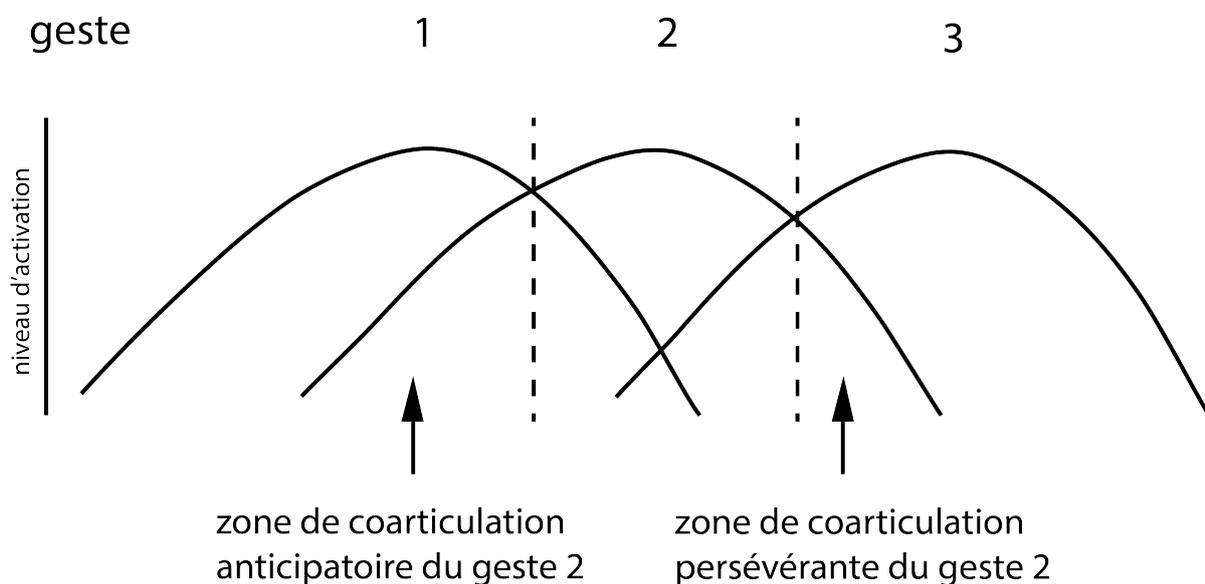


FIGURE 2.19 – Représentation schématique de l'activation des ondes pour trois gestes phonétiques chevauchés avec des zones montrant les coarticulations anticipatoire et persévérante (d'après Fowler et Saltzman (1993)).

La vision dynamique de ce modèle fournit une réflexion sur le mécanisme de la coarticulation. Elle représente une application concrète de la théorie de la coproduction. Ce modèle suppose que les trajectoires du mouvement continu peuvent être analysées en termes de gestes discrets. Il existe deux conséquences importantes de l'utilisation des gestes comme des éléments primitifs de la production de la parole. Les unités basiques de la parole sont modulées comme des mouvements vers des cibles abstraites au lieu des positions cibles statiques. Il est possible de travailler avec des unités abstraites et invariantes qui produisent de la variabilité. Alors que les gestes peuvent être invariants, les mouvements associés avec eux seraient affectés par d'autres gestes. De ce fait, ils varient selon le contexte. La coarticulation est donc modulée comme inhérente dans le processus de la production de la parole.

## 2.4 Conclusion

Ce chapitre présente une description générale du phénomène de la coarticulation. Il montre également le progrès théorique et expérimental reflété dans l'évolution continue des modèles et la méthodologie, de plus en plus, rigoureuse et sophistiquée. La définition de la notion de la coarticulation avec une description chronologique de celle-ci a été avancée. En effet, la coarticulation se produit parce que les mouvements des différents articulateurs se chevauchent et interagissent pendant la production de la parole. Par conséquent, la configuration du conduit vocal est influencée par plusieurs segments à n'importe quel moment durant l'articulation. Les impacts de la coarticulation attestent de la complexité de la production de la parole. Le degré de coarticulation entre les segments dépend en grande partie du degré de conflit entre les gestes qui les composent. Il est également lié aux intervalles spatial et temporel qui les séparent. La coarticulation est une caractéristique systématique et universelle de la parole. De ce fait, la production de la parole peut être considérée comme *un ensemble de segments coarticulés*. Les recherches physiologiques mettent en évidence que les commandes qui gouvernent les mouvements phonétiques (les mouvements en général aussi) ne correspondent pas à des commandes individuelles centrales pour chaque muscle mais à l'initialisation d'un certain groupe de muscles (Browman et Goldstein (1984), par exemple). Dans ces études, les assemblages musculaires fonctionnels ont été appelés les structures coordinatives. Ces dernières ont la propriété, essentielle, d'être capables de produire des classes de mouvement équivalent. L'invariant n'est donc pas absolu mais relatif.

Plusieurs approches théoriques différentes de la coarticulation ont été exposées. Elles révèlent le mécanisme et l'origine de la coarticulation avec des points de vue différents. L'une des différences principales entre les théories de la coarticulation est l'interprétation de la représentation temporelle. La théorie de propagation des traits considère que le segment diffuse son trait intrinsèque à tous les segments qui le précèdent n'ayant pas ce trait. Ce postulat suppose l'ignorance du facteur temporel intrinsèque des segments. Quant à la théorie de la coproduction, elle met en avant que chaque segment possède une durée relativement fixe. Un deuxième geste peut commencer avant que le premier ne finisse ; de cette manière, les gestes des différents segments sont coproduits et se chevauchent en temps et en espace. La théorie de la variabilité adaptative traite de la coarticulation d'un aspect communicatif et sociolinguistique. Le locuteur adapte son articulation (hypo- ou hyper- articule) en fonction des besoins de l'auditeur et de la situation communicative.

Les théories de la coarticulation exposent les grandes approches à partir desquelles les modèles de la coarticulation se basent. Ces derniers représentent une application pratique des approches théoriques. Ils permettent d'avoir une vision détaillée du fonctionnement de la coarticulation et de la production de la parole. La théorie du locus argumente l'existence d'un point fréquentiel virtuel invariant de chaque occlusive indépendamment de l'effet coarticulatoire de la voyelle adjacente. Le modèle VCV d'Öhman s'oppose à l'idée du locus, étant donné que celui-

ci change compte tenu de la coarticulation de voyelle à voyelle. Le modèle de l'*undershoot* explique que les changements temporels et/ou l'accent tonique influencent le degré de coarticulation. Ces changements peuvent provoquer une perte d'informations spectrales en raison de l'échec de l'articulateur à atteindre sa cible articuloire. La théorie H&H avance que les exigences perceptives de l'auditeur et la situation communicative impactent le degré de coarticulation. Les modèles de résistance coarticuloire et des équations de locus proposent une quantification du degré de coarticulation. Le modèle du DAC suggère une prédiction du degré de coarticulation. Le modèle *task-dynamics* offre une explication dynamique et systématique de la production de la parole. Il est considéré comme une application concrète de la théorie de coproduction.

Les variations temporelles influencent le degré de coarticulation. Les fréquences formantiques de la consonne et/ ou de la voyelle ainsi que les transitions formantiques sont impactées par ce phénomène (Lindblom, 1963a; Gay, 1978a; O'Shaughnessy, 1986; Krull, 1989). Un des objectifs de cette recherche est d'étudier l'étendue et les conséquences de ces variations temporelles. Dans le prochain chapitre, la notion des équations de locus sera développée. Ces dernières représentent un outil de mesure de coarticulation et est un bon indicateur du lieu d'articulation. Les équations de locus seront utilisées dans les analyses des données de cette étude.

# Chapitre 3

## Les équations de locus

La variabilité contextuelle des segments est considérée comme un problème majeur dans la recherche de la perception. Durant les trois dernières décennies, un grand intérêt a été porté aux équations de locus notamment au travers de deux problématiques parallèles. D'un côté, elles représentent un outil paramétrique permettant la mesure du degré de coarticulation entre la consonne et la voyelle. D'un autre côté, les équations de locus sont considérées comme des indices invariants qui caractérisent le lieu d'articulation des consonnes. Les études sur les équations de locus ont montré que ces dernières sont un phénomène reproductible de manière fiable. Leurs pentes et ordonnées à l'origine varient systématiquement en fonction du lieu d'articulation et du degré de coarticulation. L'objectif de ce chapitre est de définir la notion des *équations de locus* et d'exploiter son origine ainsi que son utilisation.

### 3.1 Qu'est-ce que les équations de locus ?

Les équations de locus sont considérées comme un marqueur de lieu d'articulation des occlusives et comme un indice numérique du degré de coarticulation (Nearey et Shammass, 1987; Krull, 1989; Sussman et al., 1991). Le terme "locus" ici a un sens concret, désignant une fréquence formantique mesurable à la frontière des séquences CV ou VC (Lindblom, 1963b; Krull, 1989). Il est différent du locus d'origine défini comme étant un point virtuel sur l'échelle de la fréquence situé à 50 ms avant le bruit de relâchement (Delattre et al., 1955). Le locus dans le sens classique (locus "virtuel") est une caractéristique acoustique fixe et invariante du lieu d'articulation des occlusives indépendamment de la voyelle suivante. Les équations de locus, quant à elles, sont calculées à partir du *locus* mesuré d'une consonne. L'étude expérimentale de Lindblom (1963a,b) a fourni un indice important sur les productions *occlusive + voyelle* qui peuvent être caractérisées acoustiquement d'une manière ordonnée donnant naissance aux équations de locus. Ces dernières sont des régressions linéaires du deuxième formant (F2) prélevé à son "*onset*" (le locus ou le début des transitions de F2, appelé  $F2_{onset}$ ) et à son "*midpoint*" (le milieu du F2, désigné par  $F2_{mid}$ ) (Figure 3.1 a)). Ces valeurs de fréquence sont tracées pour une seule

consonne prononcée avec plusieurs voyelles. Les  $F2_{onset}$  sont figurés sur l'axe des ordonnées (y) et les  $F2_{mid}$  sont alignés sur l'axe des abscisses (x) (Figure 3.1 b)). Par exemple, sur la Figure 3.1 b), les tracés de ces paramètres concernant l'occlusive /k/, produite par un locuteur de l'arabe jordanien avec plusieurs contextes vocaliques (/ki:s/, /ka:z/, /ku:ʔ/), forment trois nuages de points<sup>1</sup> avec une distribution positivement corrélée. Ces nuages de points peuvent correspondre à une droite de régression linéaire qui est exprimée par l'équation suivante :

$$F2_{onset} = k \times F2_{mid} + c \quad (3.1)$$

Où k et c sont les constants : la pente et l'ordonnée à l'origine, respectivement.

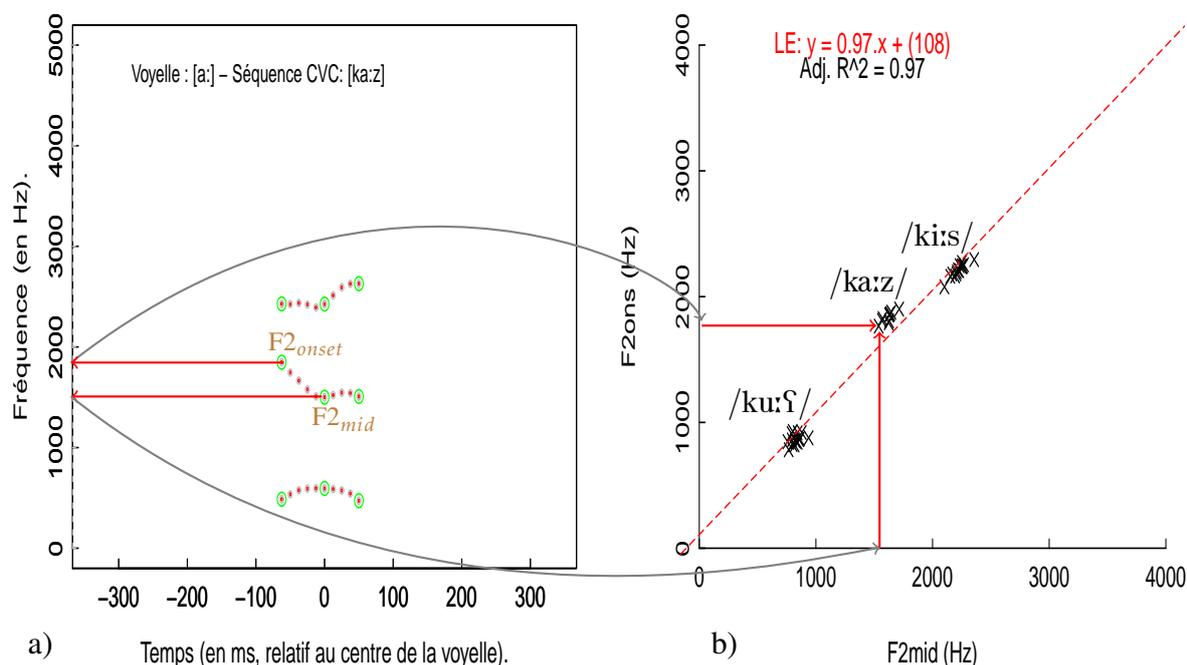


FIGURE 3.1 – Illustration de la méthode de construction des nuages de point pour le calcul des équations de locus. À gauche a), un spectrogramme schématisé dérivé automatiquement par une analyse LPC permet d'identifier les points correspondant à la mesure de  $F2_{mid}$  et  $F2_{onset}$  dans une séquence CV. À droite b), les deux valeurs de fréquence prélevées permettent de tracer un point unique dans un espace  $F2_{mid} / F2_{onset}$ .

## 3.2 Les premières études

Dans sa recherche sur la réduction vocalique, Lindblom (1963a,b) remarque une corrélation forte entre le début et le milieu du deuxième formant. Il constate que « [...] *the position of the second formant at the initial vowel segment boundary is correlated with its location in the vowel nucleus of the syllable* »<sup>2</sup>. Il a calculé les équations de locus à l'origine pour un seul locu-

1. Chaque nuage de points représente 15 répétitions du mot.

2. « [...] la position à la frontière initiale du deuxième formant de la voyelle est corrélée avec celle située au milieu », Lindblom 1963b, page 65 dans Sussman (1994), [notre traduction].

teur suédois qui a produit des séquences CV. Ces dernières sont composées de trois occlusives voisées /b, d, g/ et sept voyelles suédoises. La pente et l'ordonnée à l'origine de la droite de régression de chaque consonne sont affichées dans la Table 3.1. Lindblom (1963b) est le pre-

C	/b/		/d/		/g/	
	P	OO	P	OO	P	OO
	0.69	410	0.28	1225	0.95	360

TABLE 3.1 – Valeurs des pentes (P) et des ordonnées à l'origine (OO) des trois consonnes produites par un locuteur suédois (Lindblom (1963b)).

mier à suggérer que les pentes des équations de locus varient systématiquement en fonction du lieu d'articulation (Lindblom et Sussman, 2012). Pourtant, Lindblom ne propose pas de bases théoriques au phénomène des équations de locus et explique que « *Fitting straight lines serves only as a descriptive purpose and is not motivated by deeper considerations as to the underlying complex articulatory processes* »<sup>3</sup>. L'étude des équations de locus n'a pas été poursuivie ensuite ; elle ne sera reprise qu'à partir des années 1980 et 1990, notamment avec les travaux de Klatt (1979); Lacerda (1985); Krull (1987); Nearey et Shammas (1987); Sussman et al. (1991) et Fowler (1994).

Dans ses recherches sur la parole synthétique, Klatt (1979) trouve que la théorie du locus est une sursimplification des deux premiers formants. Il précise qu'une version modifiée du concept du locus est nécessaire afin de prendre en compte la nature dynamique et variée de la parole. Il demande à un locuteur américain de prononcer la vélaire /g/ avec 16 voyelles dans des séquences gV. L'auteur indique qu'un phénomène de locus se manifeste lorsque les fréquences des débuts des transitions formantiques ( $F1_{onset}$ ,  $F2_{onset}$ ,  $F3_{onset}$ ) sont combinées avec celles des milieux des voyelles ( $F1_{mid}$ ,  $F2_{mid}$ ,  $F3_{mid}$ ). Ces points de données ajustent des lignes droites (Figure 3.2), permettant la prédiction de  $F2_{onset}$  à partir de  $F2_{mid}$  par le biais de l'équation linéaire suivante :

$$F2_{onset} = F2_{locus} + k * [F2_{mid} - F2_{locus}] \quad (3.2)$$

Où  $F2_{locus}$  est la fréquence du locus et k est le degré de la coarticulation de la voyelle au moment du relâchement de la consonne. Ils correspondent aux paramètres à ajuster avec les données observées pour chaque consonne. Klatt explique que les transitions peuvent être rapides ou lentes compte tenu du relâchement rapide de l'obstruction qui est suivi d'un mouvement graduel du dos de la langue. Il émet l'hypothèse que les influences principales de la voyelle sur l'articulation de la consonne résident dans l'antériorité/postériorité du corps de la langue et l'arrondissement des lèvres. Il divise ainsi les voyelles anglaises en trois catégories : {+

3. « L'ajustement des lignes droites sert uniquement à des objectifs descriptifs et il n'est pas motivé par des considérations plus profondes pour ce qui est des processus articulatoires complexes sous-jacents », Lindblom 1963b, page 67 dans Sussman (1994), [notre traduction].

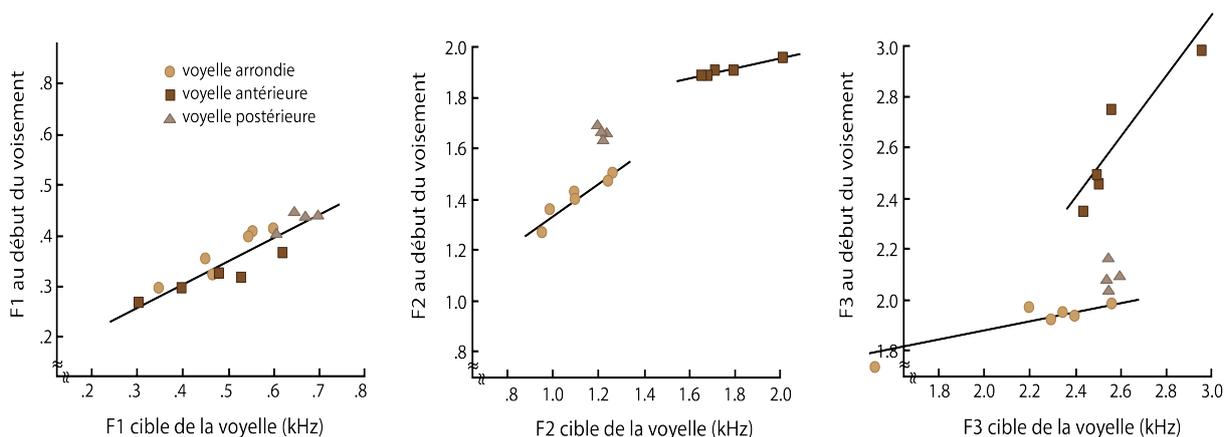


FIGURE 3.2 – Les fréquences des débuts des transitions formantiques ( $F1_{onset}$ ,  $F2_{onset}$ ,  $F3_{onset}$ ) sont affichées en fonction du milieu des 16 voyelles ( $F1_{mid}$ ,  $F2_{mid}$ ,  $F3_{mid}$ ) pour l’occlusive /g/. Chaque point est la moyenne de six répétitions d’un seul locuteur. Les cercles représentent les voyelles arrondies, les carrés marquent les voyelles antérieures et les triangles indiquent les voyelles postérieures (d’après Klatt (1979) dans Klatt (1987)).

antérieure}, {+ postérieure} et {– antérieure, – postérieure}. Il souligne qu’avec chacun de ces groupes, les données sont suffisamment régulières pour qu’elles soient décrites par des droites de régressions linéaires (Figure 3.2). Selon Klatt (1979), la linéarité des points de données de /g/ produit avec les voyelles (antérieures et postérieures) confirme l’existence de l’équation de la théorie du locus.

Lacerda (1985) utilise également les équations de locus dans l’étude de la parole synthétique. Il argumente que ces équations sont appropriées dans ce domaine puisqu’elles peuvent être utilisées pour calculer les loci actuels des consonnes en connaissant le contexte vocalique. Il précise que les pentes des équations de locus dépendent directement de la consonne et des voyelles qui la précèdent/suivent. Elles peuvent aussi être interprétées comme un indicateur de degré de coarticulation : une valeur de pente élevée (proche de 1) indique un degré de coarticulation maximale. Les valeurs de pentes et d’ordonnées à l’origine sont d’abord estimées à partir des données réelles ( $F2_{onset}$  et  $F2_{mid}$  issus de la parole naturelle). La valeur fréquentielle de la consonne (le début/la fin de transition de F2) peut donc être prédite pour chaque voyelle. Par exemple, dans la Figure 3.3, pour n’importe quelle valeur de F2 cible ( $F2_{mid}$ ) en abscisses, la valeur prédite du locus de la consonne est obtenue en lisant l’ordonnée du point correspondant à l’intersection avec le point en abscisses sur *la ligne droite de la consonne* correspondante.

Dans son étude, des séquences VCV ont été réalisées par un seul locuteur suédois, où C = [b, d, ɖ, g, f, s, ʃ, ɕ, h] et V = [ɪ, ε, a, ʊ, i:, ε:, æ:, α:, u:]. Le locus initial (CV) et celui final (VC) ont été évalués pour chaque consonne. Les résultats des équations de locus montrent que le degré de la coarticulation persévérante (VC) est important pour toutes les consonnes, au contraire du degré de coarticulation anticipatoire. Toutefois, les valeurs des pentes en position VC (coarticulation persévérante) et en position CV (coarticulation anticipatoire) sont similaires. Les pentes des occlusives sont élevées pour /b/ et basses pour /d/ avec les valeurs suivantes :

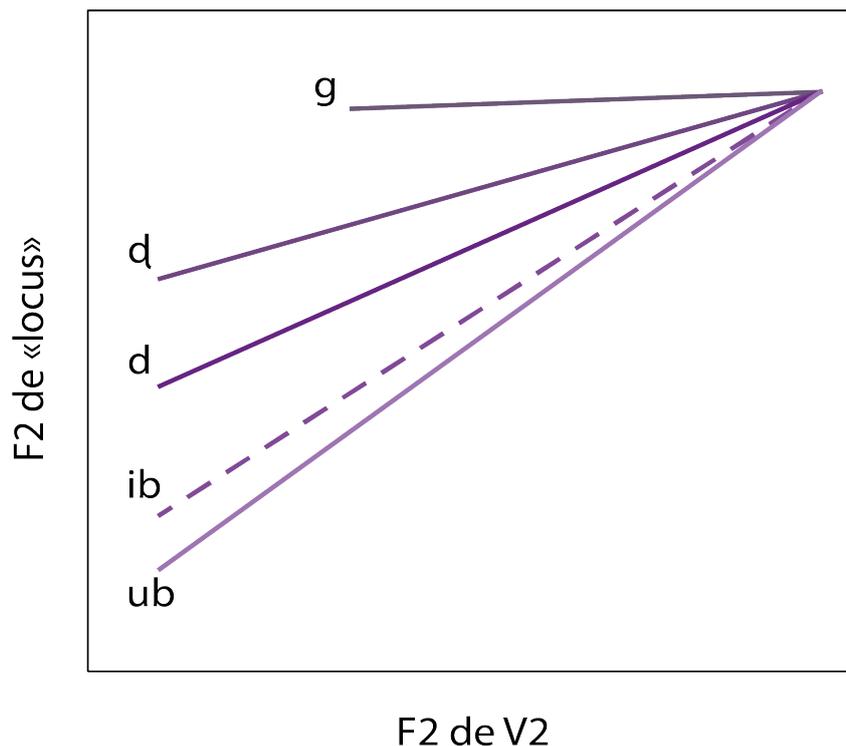


FIGURE 3.3 – Des équations de locus stylisées pour les occlusives étudiées dans l'article de Lacerda (1985) (d'après Lacerda (1985)).

/b/ (0.71), /d/ (0.35), /d/ (0.46) et /g/ (0.65). Quant aux fricatives, les pentes sont élevées pour /h/, /f/ et /s/ ( $\approx 1$ ) et basses pour /j/ (0.55). Il ajoute que l'engagement de la langue dans la production des fricatives influence le degré de coarticulation anticipatoire (+ d'engagement, – de valeur de pente). Les paramètres des équations de locus ont été employés dans l'attention de prédire le locus de F2 pour synthétiser des séquences VCV avec les occlusives citées plus haut. Lacerda (1985) souligne que le taux d'identification des occlusives est relativement élevé lorsque les paramètres des équations de locus estiment le locus de F2.

### 3.3 Un outil de mesure de coarticulation

Les trajectoires formantiques entre le moment du relâchement de la consonne et le milieu de la voyelle montrent une multitude de caractéristiques du lieu d'articulation de la consonne et de la voyelle. L'influence de la consonne est maximale au moment du relâchement de celle-ci alors qu'elle est faible ou absente au milieu de la voyelle. Krull (1988) indique que les équations de locus peuvent être utilisées comme un outil de mesure de la quantité de coarticulation entre la consonne et la voyelle adjacente. Lorsque la pente des équations de locus ( $k$  dans l'équation 3.1) est égale à 1, la fréquence du locus ( $F2_{onset}$ ) et celle de l'état stable de la voyelle ( $F2_{mid}$ ) sont équivalentes ( $F2_{onset} = F2_{mid}$ ). Dans ce cas, la valeur de  $F2_{onset}$  est fortement dépendante de la voyelle. La consonne atteste une résistance coarticulatoire minimale ou un degré de coarticulation maximal (en bas de la Figure 3.4). Quand la pente est égale à 0, la valeur du  $F2_{onset}$  est

indépendante du  $F2_{mid}$ . L'effet de la coarticulation est inexistant car la fréquence du locus reste

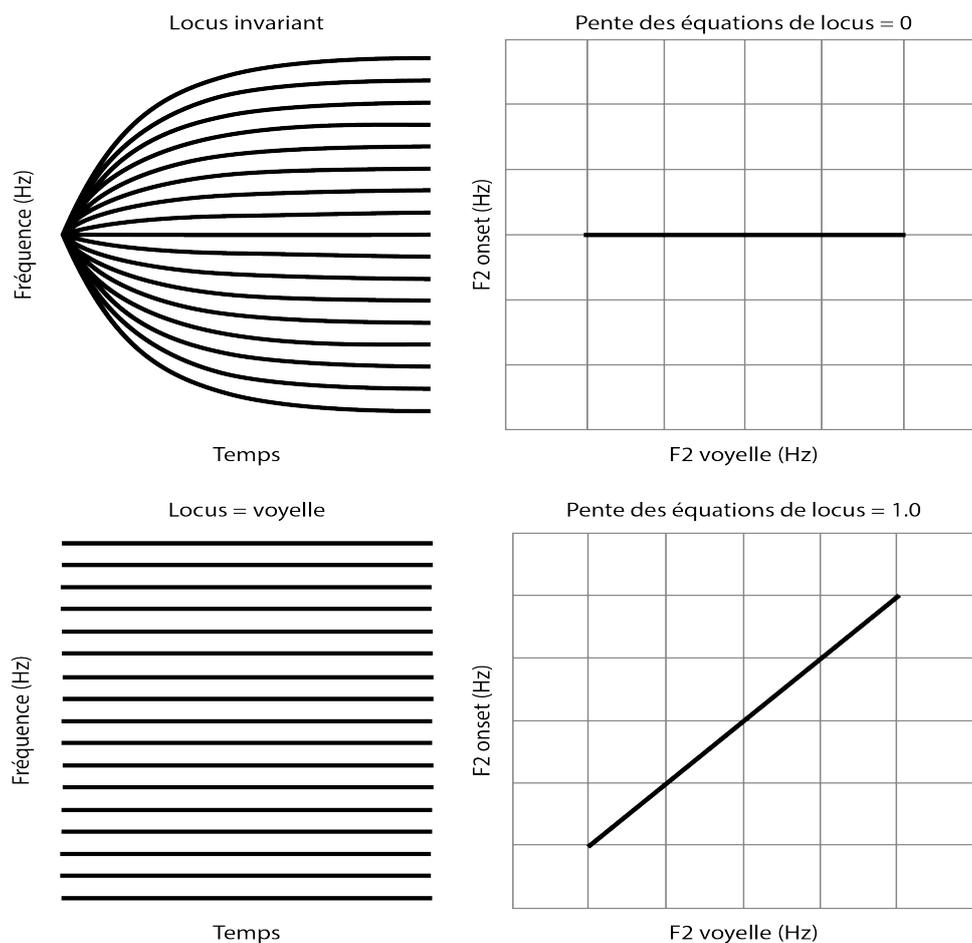


FIGURE 3.4 – Représentation schématique expliquant comment la pente des équations de locus reflète le degré de coarticulation entre une consonne et les voyelles suivantes. Une coarticulation minimale est remarquée quand la pente est égale à 0, le locus est ainsi invariant (en haut). Une coarticulation maximale est notée lorsque la pente est égale à 1, le locus et la voyelle ont alors des fréquences similaires (en bas) (d'après Lindblom et Sussman (2012)).

constante dans les différents contextes vocaliques (en haut de la Figure 3.4). La résistance co-articulaire de la consonne est donc maximale. La valeur de pente varie entre ces deux valeurs extrêmes (0 et 1).

Krull (1988) calcule les équations de locus pour trois premiers formants ( $F1$ ,  $F2$ ,  $F3$ ) afin d'étudier la relation coarticulaire anticipatoire et persévérante entre la consonne et la voyelle en suédois. Un locuteur suédois a produit quatre occlusives (/b, d, ɖ, g/) avec cinq voyelles (/i, ε, a, ɔ, u/) dans des séquences  $V_1CV_2$ . Elle explique que deux régressions linéaires sont nécessaires pour la vélaire /g/ puisqu'elle a deux lieux d'articulation en fonction de la voyelle adjacente (antérieure ou postérieure). Les équations de locus ont été calculées pour les /g/ produits avec les voyelles antérieures uniquement<sup>4</sup>. Les équations de locus du  $F2$  de l'occlusive en position initiale ( $CV_2$ ) désignent une valeur de pente basse pour le /g/ (avec les voyelles

4. Il n'existe pas assez de données pour les voyelles postérieures.

antérieures), moyenne pour /d/ et /d/ et élevée pour /b/ (Figure 3.5.a). Ces résultats révèlent que, pour les quatre lieux d'articulation, le degré de coarticulation est maximal pour la bilabiale, minimal pour la vélaire et intermédiaire pour la rétroflexe et l'alvéolaire.

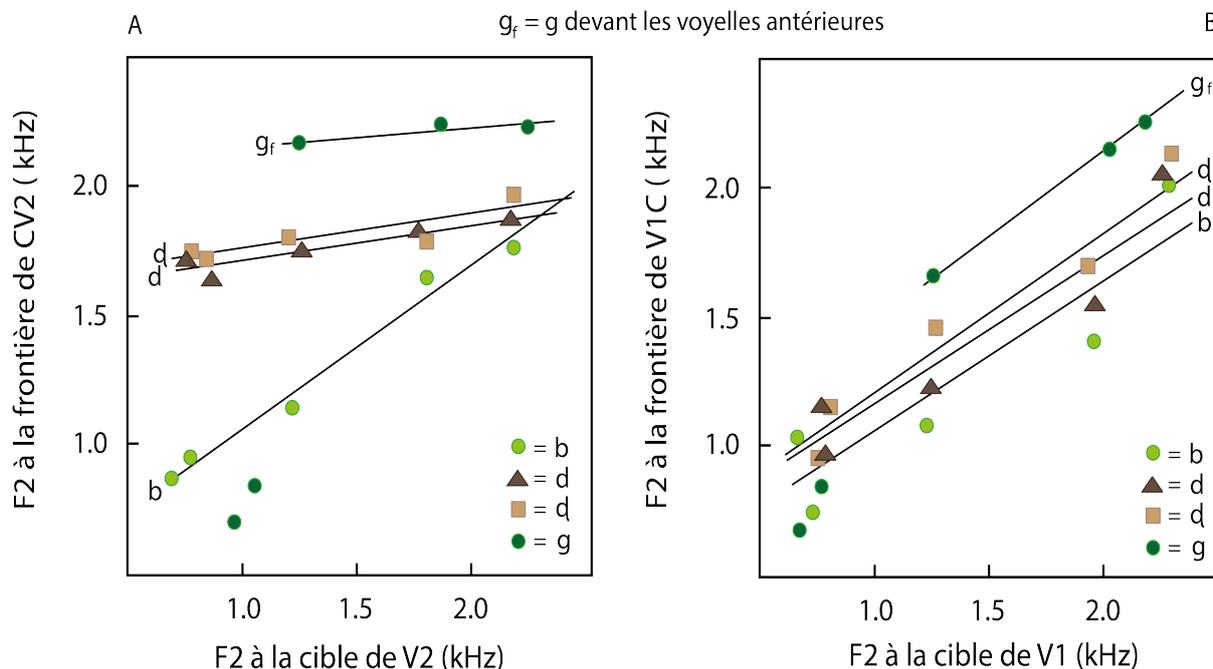


FIGURE 3.5 – Les droites de régression des équations de locus de F2 des quatre occlusives en position initiale (à gauche) et en position finale (à droite). La droite de régression de la vélaire /g/ a été calculée avec les voyelles antérieures (d'après Krull (1988)).

Les équations de locus des consonnes en position finale ( $V_1C$ ) ont des pentes relativement élevées (proche de 1). De plus, la variation entre les pentes des occlusives est réduite (Figure 3.5.b). Ce constat met en exergue que le locus en position finale est très dépendant de la voyelle précédente. Le degré de coarticulation est ainsi élevé pour toutes les occlusives en position finale. D'après Krull, la seule différence distinguant les lieux d'articulation de ces occlusives apparaît dans la valeur de l'ordonnée à l'origine de chaque droite de régression. Elle est élevée pour la vélaire et décroît dans l'ordre suivant pour la rétroflexe, l'alvéolaire et la bilabiale. En effet, les droites de régression estimées pour F3 et F1 montrent peu de variations en fonction du lieu d'articulation. Krull (1988) propose que chaque lieu d'articulation a un degré de coarticulation distinct. Cette différence est reflétée dans les valeurs de pente des équations de locus.

### 3.3.1 Mesurer la coarticulation dans différents styles de parole

Les études de Krull (1987, 1989) sont considérées parmi les premières recherches utilisant les équations de locus comme un outil de mesure du degré de coarticulation des différents styles de parole. Krull (1987) compare le degré de coarticulation en parole spontanée avec les mots en isolation (mots de référence) dans les séquences CV en position initiale de chaque mot ( $C=$

/b,d/, V=/a, e, i, o, u/). Deux locuteurs suédois ont participé à l'expérience de cette étude. La parole spontanée a été obtenue par deux manières : d'un côté, les locuteurs racontaient des histoires courtes qu'ils avaient déjà lues avant le début de l'enregistrement ; d'un autre côté, ils discutaient librement entre eux. Les mots cibles prononcés en parole spontanée ont été notés dans une liste et reproduits en isolation. La chercheuse s'est servie des équations de locus comme outil de mesure de la coarticulation. Les résultats montrent que les pentes des équations de locus de l'occlusive alvéolaire (dans les deux styles de parole) sont inférieures à celles de la bilabiale (Table 3.2). Elle explique que la valeur élevée de la pente de /b/ est provoquée par le non-engagement de la langue pendant la production de la bilabiale. Par ailleurs, les pentes des équations de locus en parole spontanée sont plus élevées que celles en isolation pour les deux locuteurs.

Lieu d'articulation	Bilabiale /b/				Alvéolaire /d/			
	Spontanée		Lue		Spontanée		Lue	
Locuteur/Style	P	OO	P	OO	P	OO	P	OO
PAT	0.96	30	0.81	210	0.45	810	0.25	1190
AV	0.75	320	0.61	570	0.47	830	0.43	900

TABLE 3.2 – Valeurs des pentes (P) et des ordonnées à l'origine (OO) des deux consonnes produites en parole spontanée et en isolation par les deux locuteurs (Krull, 1987).

Ces observations révèlent que le chevauchement articuloire entre la consonne et la voyelle est plus important en parole spontanée. Krull (1987) signale que ce phénomène pourrait être lié à deux facteurs de nature différente :

1. La durée des mots en parole spontanée est plus courte que celle des mots prononcés en isolation. La réduction de la durée de la séquence dans le premier cas implique un *undershoot* puisque les articulateurs n'atteignent pas leurs cibles articuloires entraînant un chevauchement plus important.
2. Elle ajoute que la clarté de la prononciation ou l'hypo-/hyper-articulation pourrait avoir un impact sur la coarticulation. Ce facteur est indépendant du débit de parole, par contre, il peut dépendre de facteurs sociaux ou communicatifs.

Krull (1989) étend son travail sur les variations spectrales engendrées par le changement de style de parole. Cinq locuteurs suédois ont produit de la parole spontanée et des mots en isolation comme dans l'expérience expliquée plus haut. Les mots contenant des séquences CV initiales (où C = /b, d, m, n, l/ et V n'importe quelle voyelle) ont été analysés dans les deux styles de parole. Les résultats mettent en évidence que les consonnes appartenant au même lieu d'articulation ont des pentes similaires. Ils sont en accord avec ceux présentés dans l'article de Krull (1987) où les bilabiales (/b, m/) ont des valeurs de pentes de régression plus élevées que celles des alvéolaires (/d, n, l/) dans les deux styles de parole. Le chevauchement coarticuloire

des consonnes est plus important en parole spontanée qu'avec les mots lus en isolation (Table 3.3). Des différences de pentes (et d'ordonnées à l'origine) entre les locuteurs sont notées. Par

Lieu d'articulation	Bilabiale /b,m/				Alvéolaire /d,n,l/			
	Spontanée		Lue		Spontanée		Lue	
Style de parole	P	OO	P	OO	P	OO	P	OO
Moyenne	0.76	250	0.71	382	0.43	849	0.35	1049
Écart-type	0.07	97	0.08	113	0.07	83	0.06	119

TABLE 3.3 – Les valeurs moyennes et les écart-types des pentes (P) et des ordonnées à l'origine (OO) des cinq locuteurs en parole spontanée et parole lue en isolation (Krull, 1989).

exemple, le locuteur AV a des pentes relativement élevées dans tous les contextes alors que le locuteur OE a des pentes basses. Ces variations inter-locuteurs peuvent être occasionnées par le propre débit de parole du locuteur.

Elle analyse également l'influence de l'accent tonique et de la longueur phonologique (voyelle longue vs. courte) sur la coarticulation<sup>5</sup>. L'opposition de longueur phonologique ne semble pas avoir d'impact sur la coarticulation (avec une moyenne de pente de 0.41 pour les voyelles longues et 0.45 pour les courtes). D'un autre côté, les mots accentués ont des valeurs de pente plus basses (0.73) que ceux non-accentués (0.45). Krull s'interroge sur l'origine spectrale de la variation des pentes. La distance fréquentielle entre le locus et le milieu de la voyelle (soit la différence entre  $F2_{onset}$  et  $F2_{mid}$ ) est plus petite en parole spontanée qu'en parole lue. Les voyelles avec un  $F2$  élevé, comme le /i/, provoquent une élévation de la fréquence du locus et lorsqu'elles ont un  $F2$  bas, la fréquence du locus est abaissée. Toutefois, le locus n'est pas stable, il peut varier en fonction de la voyelle adjacente dans un intervalle compris entre deux valeurs limites, une valeur minimale et une valeur maximale. Les milieux des voyelles connaissent une variation importante dans le contexte de /d/. Elle explique que ce phénomène est provoqué par l'existence de deux variants de ce phonème en position initiale en suédois. Il s'agit du /d/ produit dans les mots de fonction et celui produit dans les mots de contenu<sup>6</sup>. Dans le premier cas, le /d/ est non-accentué et plus réduit que dans le second. Krull révèle que malgré les différences de pentes et de fréquences observables entre les locuteurs, le locus et le milieu de la voyelle sont plus proches en parole spontanée qu'en parole lue en isolation pour tous les locuteurs, indiquant ainsi une coarticulation ou un *undershoot* plus important.

Duez (1989, 1991, 1992) réplique l'expérience de Krull (1987, 1989) sur le français avec les deux lieux d'articulation : bilabial /b, m/ et alvéolaire /d, n, l/. Cinq locuteurs français ont participé à cette étude afin de produire de la parole spontanée et de la parole lue (des mots

5. Cette analyse est réalisée uniquement pour la consonne /l/ dans les mots contenant la séquence /l/V pour trois locuteurs. La quantité des données des autres consonnes n'était pas suffisante pour effectuer le calcul.

6. Les mots de fonction sont les mots qui ont une signification lexicale réduite, ils sont utilisés pour exprimer les relations entre les mots dans une phrase (prépositions, articles, ...). Les mots de contenu, quant à eux, sont de type nom, verbe, adjectif et adverbe.

lus en isolation). L'impact de la proéminence syntactique<sup>7</sup> sur la coarticulation a également été analysé. Les résultats montrent des valeurs élevées de pentes et des valeurs basses des ordonnées à l'origine pour tous les locuteurs (Table 3.4).

Lieu d'articulation	Bilabiale /b,m/				Alvéolaire /d,n,l/			
	Spontanée		Lue		Spontanée		Lue	
Style de parole	P	OO	P	OO	P	OO	P	OO
Moyenne	0.81	166	0.75	229	0.67	553	0.51	852
Écart-type	0.07	89	0.06	74	0.03	62	0.04	88

TABLE 3.4 – Valeurs moyennes des pentes (P) et des ordonnées à l'origine (OO) en parole spontanée et les mots lus en isolation des 5 locuteurs (Duez, 1992).

Les pentes des bilabiales sont plus élevées que celles des alvéolaires. La durée des voyelles en parole spontanée est plus courte que celle des mots lus en isolation pour toutes les séquences. Des variations de pentes et d'ordonnée à l'origine sont aussi remarquées entre les locuteurs au sein d'une même catégorie. Duez (1992) souligne que ces variations peuvent être la conséquence directe des différences individuelles de l'organisation temporelle. Les syllabes prominentes ont été comparées avec celle non-prominentes<sup>8</sup>. Une syllabe prominente est plus longue que sa contrepartie non-prominente. Les consonnes avec des syllabes non-prominentes ont des pentes plus élevées que celles avec des syllabes prominentes. Ce phénomène met en avant une coarticulation plus importante dans le premier cas que dans le second. Les résultats de cette étude sont en accord avec ceux de Krull (1989). Duez (1992) attire l'attention sur le fait que les différences de pentes observées dans cette expérience peuvent être expliquées en grande partie par les variations temporelles. Le chevauchement coarticulatoire augmente lorsque la durée vocalique diminue. La différence fréquentielle est ainsi moins importante entre le début et le milieu des transitions formantiques.

D'autres phénomènes articulatoires ont incité l'utilisation des équations de locus pour mesurer le degré de chevauchement coarticulatoire, comme le débit de parole<sup>9</sup> et l'accent emphatique (Agwuele et al., 2007, 2008; Berry et Weismer, 2013; Lindblom et al., 2007). Selon ces études, les pentes des équations de locus augmentent lorsque le chevauchement coarticulatoire croît; ce qui est en accord avec les travaux de Krull (1989) et de Duez (1992). La plupart des recherches sur les équations de locus ont confirmé leur capacité à décrire le degré de coarticulation. Puisque chaque consonne révèle un degré de coarticulation distinct, les équations de locus ont été proposées comme un indicateur du lieu d'articulation. Un grand débat a été engagé à partir de cette proposition, il sera décrit dans la section suivante.

7. L'auteure considère qu'il n'existe pas d'accent tonique en français dans le sens traditionnel mais qu'un allongement de la syllabe peut se produire quelques fois à la fin des mots et toujours à la fin des groupes syntaxiques. Elle appelle ce phénomène syllabe proéminente ou proéminence syntactique.

8. Les équations de locus ont été uniquement calculées pour les alvéolaires produites par trois locuteurs.

9. L'impact du débit de parole sur les équations de locus sera détaillé dans le chapitre 3.

### 3.4 Un descripteur du lieu d'articulation

Il a été montré dans la littérature que les transitions formantiques fournissent aux auditeurs des informations importantes pour l'identification des occlusives en position initiale. Toutefois, l'interprétation de comment ces informations sont utilisées a toujours été un sujet de débat. La sensibilité de ces transitions en fonction du contexte vocalique adjacent a représenté un problème majeur dans le déchiffrement des informations invariantes qui permettent l'identification des occlusives (Stevens et Blumstein, 1981). Une première interprétation a été avancée par la théorie du locus de Delattre et al. (1955) où les transitions formantiques semblent se diriger vers un point fréquentiel virtuel spécifique pour chaque occlusive (Chapitre 2 Le locus). Öhman (1966) a trouvé que le locus n'est pas invariant puisqu'il est influencé par les voyelles précédentes. De plus, Kewley-Port (1982) a montré expérimentalement l'absence de l'invariant dans les transitions formantiques puisqu'elles sont sensibles aux contextes vocaliques. L'approche des équations de locus n'est pas basée sur la recherche d'un invariant dans les signaux acoustiques des séquences individuelles mais plutôt sur la recherche d'une propriété d'un ordre plus élevé, trouvée dans la relation entre deux points de la transition d'une consonne dans plusieurs contextes vocaliques. Les équations de locus sont ainsi un invariant relationnel et non un absolu comme l'ont imaginé des chercheurs comme Delattre et al. (1955) et Stevens et Blumstein (1981). La renaissance des équations de locus, notamment par Sussman et al. (1991), a apporté le développement d'un algorithme prometteur permettant d'interpréter le problème de la variabilité coarticulatoire dans des séquences CV en termes de catégorisation phonologique. La régularité des équations de locus peut être traduite comme un fait achevé malgré l'existence de la coarticulation ou au contraire, à cause de la coarticulation (Sussman et al., 1993). La distinction acoustique que les équations de locus mettent en lumière expose l'existence de stratégies coarticulatoires différentes pour chaque catégorie consonantique.

Shammas et Nearey (Shammas, 1985; Nearey et Shammas, 1987)<sup>10</sup> cherchent à quel point les informations formantiques les plus pertinentes peuvent servir à la classification des consonnes occlusives en anglais canadien. Des locuteurs (5 féminins et 5 masculins) ont prononcé des séquences de type CVd où C = /b, d, g/ et V = /i, ɪ, e, ε, æ, ʌ, ɔ, o, ʊ, u, ə, aɪ, aʊ, oɪ/.

Shammas (1985) observe une relation linéaire forte lorsque les valeurs des débuts des transitions du F2 ( $F2_i$ ) sont placées contre les états stables des voyelles ( $F2_v$ )<sup>11</sup> pour chacune des occlusives produite par un groupe de locuteurs (femmes et hommes)<sup>12</sup>. Un phénomène linéaire est aussi remarqué pour  $F3_i$  contre  $F3_v$  (Figure 3.6).

Les droites de régression réalisées sont distinctes pour chaque consonne (Figure 3.7). Les résultats montrent une pente élevée pour le /g/ (0.99), intermédiaire pour le /b/ (0.83) et basse pour le /d/ (0.50). Les données des deux groupes (femmes et hommes) suivent les mêmes

10. Les auteurs n'avaient pas connaissance du travail de Lindblom (1963b) sur les équations de locus.

11. Les valeurs de  $F2_v$  sont mesurées à 60 ms après le début de la transition du F2.

12. Les groupes femmes vs. hommes ont été traités séparément.

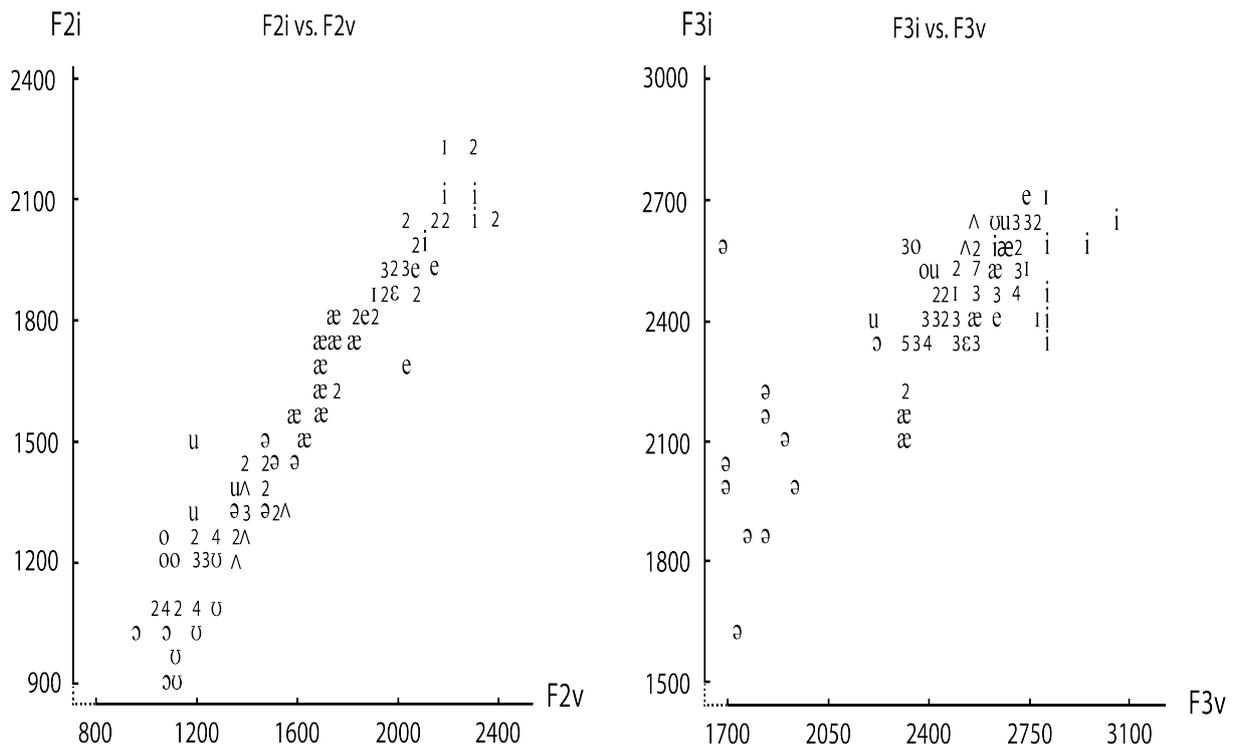


FIGURE 3.6 – Représentation de la linéarité des valeurs de  $F2_i$  en ordonnée affichées contre les valeurs de  $F2_v$  en abscisse (à gauche) et celles de  $F3_i$  contre  $F3_v$  (à droite) pour la consonne /b/ prononcée par les 5 locuteurs masculins (d'après Shammass (1985)). Les chiffres n'ont pas été définis par l'auteur.

tendances. Shammass (1985) indique que ces paramètres formantiques ne peuvent pas être suffisants pour une catégorisation parfaite puisqu'il existe des intersections entre les droites de régression des trois catégories consonantiques (Figure 3.7). Ce postulat signifie la présence de zones d'ambiguïté entre les trois catégories. Elle ajoute que les informations spectrales résidentes dans le bruit de relâchement sont indispensables pour enlever cette ambiguïté.

Toutefois, le but de son étude était d'analyser l'importance des informations formantiques indépendamment du bruit de relâchement. Plusieurs types de classification ont été utilisés<sup>13</sup> de manière à comprendre l'importance de ces paramètres dans la catégorisation des occlusives. Un taux de classification correcte relativement élevé des occlusives étudiées est noté lorsque les *débuts* et les *milieux* des deuxièmes et des troisièmes formants ont servi de prédicteurs (entre 66% et 74% en fonction de la méthode de classification adoptée). Les informations fréquentielles de F2 et de F3 ont été employées dans le but de réaliser de la parole synthétique pour une expérience de perception. En général, les réponses de catégorisation des auditeurs montrent qu'un  $F2_i$  avec une fréquence basse ou moyenne est entendu comme un /g/ et un  $F2_i$  avec une fréquence élevée est perçu comme un /d/ lorsque les séquences synthétiques ont une fréquence de  $F2_v$  basse. Cependant, quand les séquences ont un  $F2_v$  élevé, un  $F2_i$  bas est entendu comme un /b/, un  $F2_i$  moyen est discerné comme un /d/ et un  $F2_i$  est perçu comme un /g/.

13. Telles des analyses discriminantes linéaires, des analyses discriminantes quadratiques et une classification de régression à distance minimale (*Minimum Distance Regression Classification*).

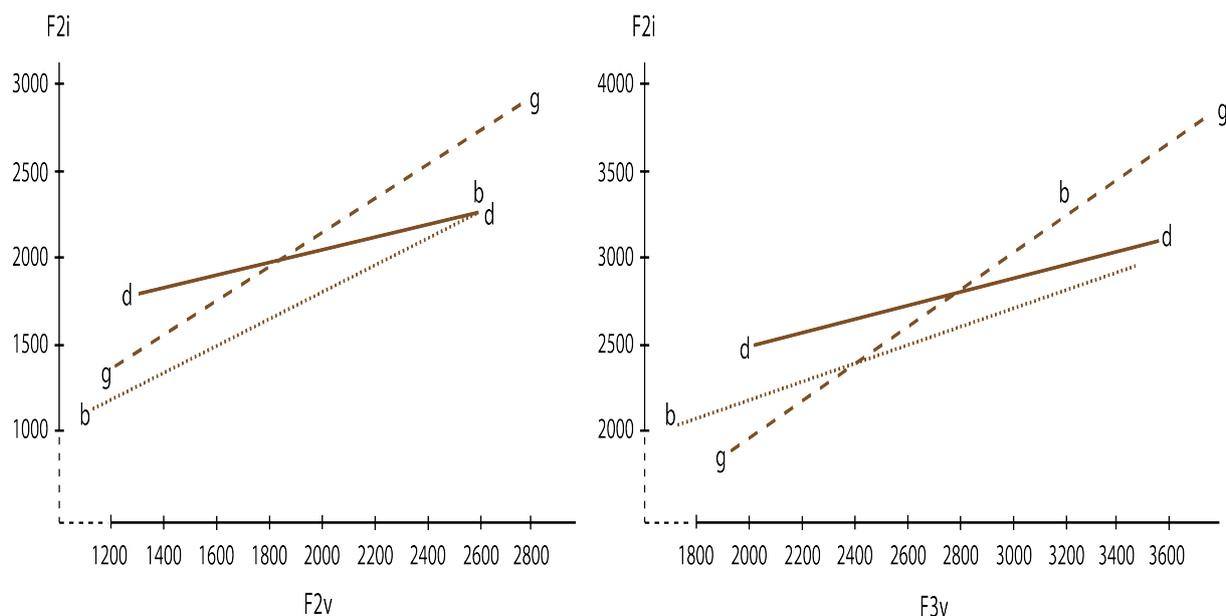


FIGURE 3.7 – Droites de régression linéaire des trois occlusives calculées pour les données de  $F2_i$  contre  $F2_v$  (à gauche) et de  $F3_i$  contre  $F3_v$  des 5 locutrices (d'après Shammass (1985)).

Le croisement entre le /d/ et le /g/ dans l'étude de perception correspond au chevauchement de ceux-ci trouvé dans les droites de régression des données de la parole naturelle (Figure 3.7). Shammass (1985) ajoute que les résultats de l'étude de perception révèlent que les prédictions basées sur les droites de régression sont liées. Nearey et Shammass (1987) exposent les résultats de l'étude de Shammass (1985) et décrivent la droite de régression issue de la relation entre  $F2_{onset}$  et  $F2_{mid}$  par : « la droite de régression représente une propriété relative invariante de la consonne, bien que les trajectoires ne soient pas suffisantes pour décrire les occlusives de façon non-ambigüe, puisque les droites de régression des trois consonnes se coupent (ce qui indique un chevauchement des catégories). Cependant, les pentes de ces droites et leurs points de rencontre avec les axes de coordonnées ont des valeurs distinctes et par conséquent représentent des propriétés invariantes partiellement distinctives ou invariants partiels »<sup>14</sup>. Ces conclusions suggèrent que les variations dépendantes de la voyelle dans les transitions de F2 et de F3 sont systématiques dans des séquences CV. Ces variations systématiques peuvent être représentées dans la relation linéaire existante entre les débuts et les milieux du F2.

Sussman (1989) propose que les équations de locus jouent un rôle important dans la perception du lieu d'articulation des occlusives. Il fait une analogie entre la perception du lieu d'articulation chez les êtres humains et la localisation du son chez les chouettes effraies. Les études neurologiques sur la localisation du son chez les chouettes effraies révèlent qu'elles arrivent à localiser le son grâce à la combinaison de deux informations dans son cerveau : la différence de phase interaurale<sup>15</sup> et la fréquence du son. La valeur de la différence temporelle interaurale (ITD, "Interaural Time Differences") peut être trouvée en associant ces deux infor-

14. Page 17.

15. Il s'agit de la différence de phase d'une onde qui atteint chaque oreille.

mations. Une valeur négative d'ITD indique que l'oreille gauche dirige et une valeur positive désigne que l'oreille droite domine. Sussman (1989) effectue une analogie de ce processus de combinaison d'informations chez les chouettes effraies avec la perception du son chez les êtres humains. Il explique que le cas caractéristique de "noninvariance" <sup>16</sup> est celui de la perception des occlusives suivies par des voyelles différentes. En se basant sur les travaux de Klatt (Klatt, 1979, 1987), Sussman (1989) argumente que les équations de locus issues de la combinaison de deux informations fréquentielles ( $F2_{onset}$  et  $F2_{mid}$ ) précisant le lieu d'articulation peuvent être analogiques aux ITD. L'auteur établit un protocole expérimental afin de vérifier son hypothèse. Deux locuteurs masculins de l'anglais américain ont produit des séquences CVC où la consonne initiale correspond aux occlusives /b, d, g/. Le premier locuteur a énoncé ces séquences avec les voyelles /i, ɪ, e, u, o, a/ et le second locuteur a réalisé les séquences avec 4 voyelles de plus /æ, ɑ, ɔ, ε/. Chaque séquence a été répétée 5 fois. Les débuts des transitions formantiques de F2 et de F3 ( $F2_{onset}$  et  $F3_{onset}$ ) ont été déterminés à la première pulsation de la voyelle suivante et les fréquences des cibles vocaliques de F2 et F3 ( $F2_{mid}$  et  $F3_{mid}$ ) ont été prélevées à 60-70 ms. après le début des transitions. Les équations de locus ont été calculées pour le deuxième et le troisième formants. Les résultats montrent des pentes de F2 distinctes en fonction du lieu d'articulation et des valeurs de pente similaires pour chaque locuteur (Table 3.5). La linéarité des pentes de F2 est élevée pour toutes les catégories consonantiques avec une moyenne de  $R^2$  de 0.91 <sup>17</sup>. Les pentes des F3 sont moins différenciées et moins linéaires ( $R^2 = 0.77$ ) que celles de

	F2						F3					
	/b/		/d/		/g/		/b/		/d/		/g/	
	P	R <sup>2</sup>										
L1	0.92	0.96	0.46	0.95	0.67	0.92	0.52	0.84	0.76	0.75	1.76	0.92
L2	0.90	0.95	0.49	0.87	0.72	0.78	0.39	0.59	0.38	0.66	1.27	0.88

TABLE 3.5 – Valeurs des pentes (P) des équations de locus et  $R^2$  du F2 et du F3 des trois occlusives avec 6 voyelles pour le premier locuteur (L1) et 10 voyelles pour le second (L2) (Sussman, 1989).

F2 et différentes entre les deux locuteurs (Table 3.5). Sussman expose que les pentes distinctives et contrastives qui caractérisent les trois catégories consonantiques peuvent être utilisées pour construire un tableau de réseau neuronal hypothétique de l'homme similaire à celui de l'ITD des chouettes effraies. Il suggère que la sortie ("*output*") de ce tableau signifierait le lieu de pro-

16. La notion de *noninvariance* peut être définie par le fait que les différents traits acoustiques dépendants du contexte signalent une même catégorie phonémique. Par exemple, le /b/ initial dans les mots : beau, bic, bac, bout, bain, bon, est catégorisé par les auditeurs par le phonème /b/. Pourtant dans chaque mot, le /b/ est physiquement différent.

17. Le  $R^2$  ou le coefficient de détermination « mesure l'intensité de la dépendance des deux variables [(x, y)] et il indique dans quelle mesure les variations d'une variable expliquent celles de l'autre [...]. Le coefficient de détermination linéaire retient la part de la variance de Y expliquée par l'ajustement de la variance totale de Y » (Baillly et Carrère 2015, page 165). La valeur de  $R^2$  varie entre 0 et 1. Plus le  $R^2$  est proche de 1, plus la part de la variance expliquée par la droite de régression est forte.

duction des occlusives. Il ajoute qu'un trait dérivé acoustiquement comme les coordonnées des équations de locus peut être : (1) systématiquement et orthogonalement planifié en substrat neuronal<sup>18</sup> et (2) combiné hiérarchiquement et interactivement avec d'autres traits indépendants et parallèles (comme les informations dans F2, F3 et dans le bruit de relâchement) et donc (3) construire des conditions efficaces pour un contraste perceptuel suffisant afin de résoudre l'ambiguïté physique au niveau des séquences CV. Une représentation schématique indiquant les

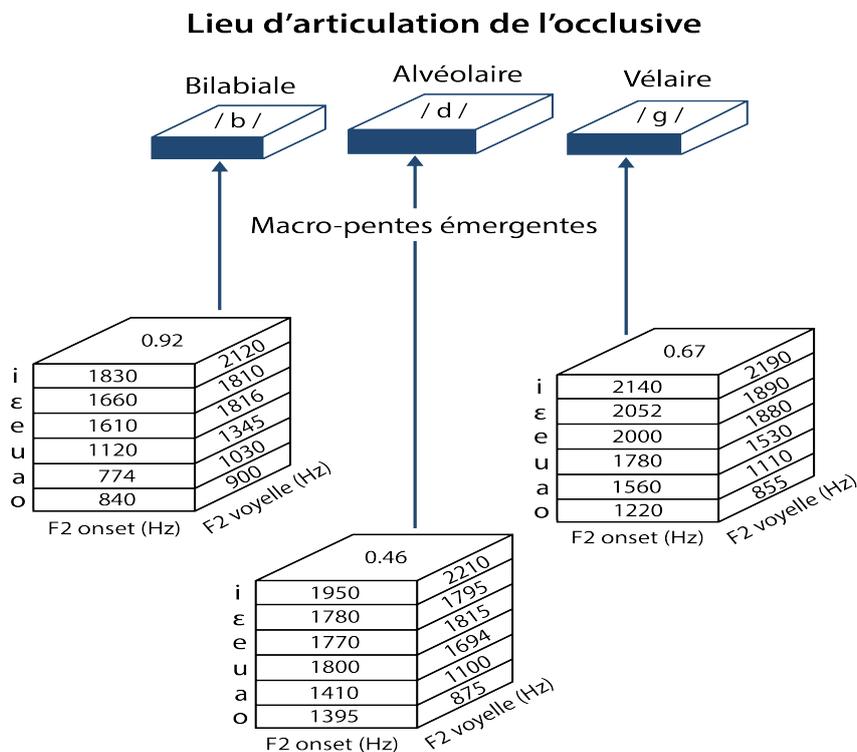


FIGURE 3.8 – Représentation schématique des tableaux des "macro-pentes" pour les occlusives /b, d, g/ dans différents contextes vocaliques. Chaque case expose la moyenne des cinq réalisations du premier locuteur (empruntée à Sussman (1989)).

équations de locus en tableau bidimensionnel est affichée dans la Figure 3.8. Chaque ligne de ce tableau contient deux informations combinées : la valeur de la fréquence de  $F2_{onset}$  et celle de  $F2_{mid}$  de chaque occlusive dans les différents contextes vocaliques. Sussman nomme les lignes des tableaux les "micropentes", elles décrivent les deux points de transition de F2 ( $F2_{onset}$  et  $F2_{mid}$ ) de chaque réalisation CV. Dans chaque tableau, la sortie de ce processus collectif, prenant en compte toutes les voyelles, est le ratio normalisé de l'occlusive appelé "macro-pente". L'apparition de ce paramètre représente un niveau plus abstrait du traitement intégratif qui n'est pas disponible dans le niveau "micropente" individuel. Chaque tableau peut être conceptualisé comme une unité fonctionnelle, analogique avec le tableau de l'ITD des chouettes effraies. Sussman souligne que l'architecture neuronale proposée pour la perception du lieu d'articulation des occlusives représente seulement un sous-ensemble de traits dans un système polymorphique. En

18. « Substrats neuronaux : Unités fonctionnelles du système nerveux central, souvent composées d'une série d'unités structurales qui interagissent pour soutenir les fonctions complexes du système nerveux » (Encyclopédie sur le développement des jeunes enfants, 2010)

effet, le signal de la parole est suffisamment riche ; si le sous-ensemble de traits contrastifs est planifié et combiné, la hiérarchie de projection résultante peut définir une distinction neuronale dans des zones phonémiques non-chevauchées.

Deux ans plus tard, Sussman et ses collègues mettent en place une expérience dans le but de fournir des arguments soutenant l'hypothèse du rôle des équations de locus dans la classification des consonnes (Sussman et al., 1991). Ils étudient les équations de locus sur un nombre important de locuteurs américains (10 féminins et 10 masculins). Des séquences CVt ont été produites et répétées cinq fois dans des phrases porteuses où C= /b, d, g/ et V=/i, ɪ, e<sup>y</sup>, ε, æ, a, o<sup>w</sup>, ʌ, ɔ, u/. Les valeurs de F2<sub>onset</sub> sont mesurées aux premières pulsations après le bruit de relâchement alors que celles de F2<sub>mid</sub> sont déterminées visuellement à l'état stable du formant (près du milieu). Au total, soixante équations de locus ont été générées (trois par sujets). Les résultats montrent des valeurs de pentes élevées pour la bilabiale, intermédiaires pour la vélaire et basses pour l'alvéolaire (Table 3.6). Une linéarité relativement importante a été remarquée pour toutes les pentes (moyenne des R<sup>2</sup> des pentes = 0.89). Il n'existe pas de différence si-

C	/b/		/d/		/g/	
	P	OO	P	OO	P	OO
Masculins	0.87	106	0.43	1073	0.66	807
Féminins	0.90	91	0.41	1349	0.75	777

TABLE 3.6 – Valeurs moyennes des pentes (P) et des ordonnées à l'origine (OO) des trois consonnes produites par les deux groupes de participants : masculin et féminin (Sussman et al. (1991)).

gnificative entre les pentes en fonction du sexe. Les analyses statistiques révèlent aussi une différence significative entre les valeurs de pente en fonction de l'occlusive. Les valeurs des ordonnées à l'origine augmentent dans l'ordre suivant : /b/ < /g/ < /d/ et signalent une variation importante, notamment dans les séquences contenant la vélaire. Elles attestent également une différence significative en fonction du lieu d'articulation.

Les paramètres des équations de locus (pente et ordonnée à l'origine) ont été comparés dans les deux groupes allophoniques de /g/ de manière à déterminer l'impact du lieu d'articulation de la voyelle sur la coarticulation de cette occlusive. Deux équations de locus par locuteur ont été estimées pour le /g/ : le /g/ suivi par les voyelles antérieures (/g/ palatal) et le /g/ prononcé avec les voyelles postérieures (/g/ vélaire). Pour tous les locuteurs, les pentes et les ordonnées à l'origine des deux allophones sont différentes et la moyenne des pentes du /g/ palatal est de 0.34, celle de /g/ vélaire de 0.87. Les deux variants allophoniques sont aussi caractérisés par deux ordonnées à l'origine distinctes : 1602 pour le /g/ palatal et 559 pour le /g/ vélaire. Les pentes et les ordonnées à l'origine des deux allophones sont statistiquement différentes. En effet, les valeurs de pente du /g/ vélaire et du /g/ palatal présentent un chevauchement avec celles de /b/ et du /d/, respectivement. Un effet global du lieu d'articulation est observé dans les pentes

et les ordonnées à l'origine des quatre catégories (/b/, /d/, /g/ palatal et /g/ vélaire). Des tests de moyennes n'ont pas signalé de différence de pentes entre /b/ vs. /g/ vélaire et entre /d/ vs. /g/ palatal. Toutefois, une différence significative des ordonnées à l'origine est notée pour les mêmes comparaisons. D'après ces résultats, Sussman et al. (1991) expliquent que « *a single [g] function contains greater statistical distinctiveness across place of articulation than each separate allophonic /g/ function* »<sup>19</sup>.

Des analyses discriminantes linéaires (LDA) ont été effectuées dans le but de quantifier la catégorisation du lieu d'articulation des occlusives dans les différents contextes vocaliques. Dans un premier temps, les fréquences du  $F2_{onset}$  et du  $F2_{mid}$  ont été utilisées comme des variables prédictives pour classer les trois lieux d'articulation dans les deux groupes (femmes et hommes). Chaque valeur de  $F2_{onset}$  ou de  $F2_{mid}$  adoptée dans les LDA correspond à la moyenne des cinq répétitions de chaque séquence. Au total, 300 mesures de  $F2_{onset}$  et de  $F2_{mid}$  ont été considérées par groupe de sexe. Le taux de classification correcte des trois classes (/b, d, g/) est de 78% pour les hommes et de 76% pour les femmes dans tous les contextes vocaliques. Des résultats similaires ont été obtenus en divisant les données en deux sous-ensembles en fonction du lieu d'articulation des voyelles (antérieures vs. postérieures).

Dans un deuxième temps, les valeurs de pente et d'ordonnée à l'origine ont été employées comme des variables prédictives. Une seule paire de pente et d'ordonnée à l'origine par catégorie consonantique par locuteur a été utilisée dans les LDA. Au total, 60 entrées de pente et d'ordonnée à l'origine des 20 locuteurs ont été incorporées<sup>20</sup>. Un taux de classification correcte de 100% a été obtenu en utilisant la pente et l'ordonnée à l'origine comme prédicteurs. Le calcul des LDA avec l'emploi de la pente uniquement comme prédicteur donne un taux de classification de 93%. Quant à l'ordonnée à l'origine, elle offre un taux de classification correcte de 88%. Sussman et al. (1991) expliquent que ces données soutiennent la proposition que les équations de locus sont des invariants relationnels du lieu d'articulation. Ils ajoutent que les représentations graphiques et les analyses de classification mettent en exergue que les catégories phonétiques peuvent être directement représentées acoustiquement.

Le protocole expérimental de Sussman et al. (1991) a été répliqué par MacLeod et al. (2001) dans l'intention de vérifier la stabilité des paramètres des équations de locus dans un autre échantillon de la population. Douze locuteurs (6 femmes et 6 hommes) de l'anglais américain ont prononcé les mêmes séquences utilisées dans l'étude de Sussman et al. (1991). En effet, les valeurs de pente et d'ordonnée à l'origine sont similaires à celles exploitées dans le travail de

19. « Une seule fonction du [g] contient une discriminabilité statistique plus importante du lieu d'articulation que deux fonctions pour les deux allophones de /g/ » page 1318, [notre traduction].

20. Il faut distinguer les analyses de classification sur les  $F2_{onset}$  et les  $F2_{mid}$  et celles sur les paramètres des équations de locus (pente et ordonnée à l'origine). Les premières sont les moyennes des répétitions des réalisations réelles. Chaque entrée ( $F2_{onset}$  et  $F2_{mid}$ ) correspond aux informations des transitions entre la consonne et la voyelle suivante. Les secondes sont des valeurs dérivées à partir des valeurs réelles issues de la prononciation d'une occlusive avec plusieurs voyelles. Selon Sussman et al. (1991), si les paramètres des équations de locus possèdent un rôle organisé dans la catégorisation du lieu d'articulation, le taux de classification correcte doit être plus élevé que celui avec les fréquences de F2.

Sussman et al. (1991). Toutefois, MacLeod et al. (2001) trouvent une différence significative de pente et d'ordonnée à l'origine en fonction du sexe. Le groupe des femmes montre des pentes moins élevées et des ordonnées à l'origine plus importantes que celles des hommes. Le même comportement des paramètres des équations de locus en fonction de sexe a été observé par Sussman et al. (1991) mais sans effet significatif. Les auteurs exposent l'existence de cette différence par « *this gender disparity in slope may be due to differences in participant samples. It is reasonable to suggest, for example, that dialectical or speaking rate variations could have contributed to gender differences observed between the present findings and those of Sussman et al. because dialectical variation and speaking rate were not controlled systematically in the present study [...] or the Sussman et al. study* »<sup>21</sup>.

La proposition de la fonction des équations de locus par Sussman et al. (1991) nécessite un approfondissement dans la recherche de nouveaux arguments qui soutiennent leurs propos. Sussman et ses collègues vont dès lors mener un ensemble de travaux visant à étayer l'hypothèse des équations de locus comme un descripteur du lieu d'articulation. En parallèle, un débat émerge par des chercheurs qui remettent en cause la capacité des équations de locus à classifier les lieux d'articulation.

### 3.4.1 Vers une validation des équations de locus

La validité des équations de locus comme un descripteur du lieu d'articulation a été examinée par Sussman et ses collègues dans plusieurs contextes : chez les enfants et dans différentes langues. Face aux critiques de Fowler (1994), Sussman et ses collègues publient un ensemble d'articles qui répondent à ses critiques et qui visent également à prouver la stabilité des équations de locus comme étant un invariant relationnel du lieu d'articulation. Ils mettent en exergue la constance de la capacité des équations de locus à décrire les lieux d'articulation dans divers modes d'articulation et dans l'articulation compensatoire. Les parties de cette sous-section a pour objectif d'explorer toutes ces questions.

#### 3.4.1.1 Chez les enfants

Sussman et al. (1992) examinent la capacité des équations de locus à décrire le lieu d'articulation ainsi que le développement de l'acquisition phonologique et coarticulatoire chez les enfants. Les auteurs expliquent que la caractérisation de la performance de l'acquisition d'un système phonologique constitue un vrai défi pour les phonéticiens. Il s'agit de créer un indice pouvant qualifier le niveau de la performance d'un enfant en comparaison de celle d'un adulte. Étant donnée que les équations de locus sont sensibles aux variations coarticulatoires, Sussman

21. « cette différence de pente en fonction du sexe peut être due aux différences dans les échantillons des participants. Il est raisonnable de suggérer, par exemple, que les variations dialectales ou de débit de parole peuvent contribuer aux distinctions observées dans les groupes de sexe dans les résultats de cette étude et ceux et dans l'étude de Sussman et ses collègues, parce que les variations dialectales et le débit de parole n'ont pas été contrôlés dans l'étude présente [...] ni dans celle de Sussman et ses collègues », page 103 [notre traduction].

et al. (1992) proposent que celles-ci peuvent signaler le niveau de la maturité articulatoire chez les enfants. Un protocole expérimental a été conçu dans le but de vérifier cette hypothèse. Seize enfants divisés en trois groupes d'âge ont participé à ce protocole : deux enfants de trois ans (féminins), sept enfants de quatre ans (4 féminins et 3 masculins) et sept enfants de cinq ans (1 féminin et 6 masculins). Les enfants ont produit des séquences CVt en imitant l'expérimentateur, où C = /b, d, g/ et V = /i, ɪ, æ, ʌ, u/. Chaque séquence a été répétée entre trois et six fois. Les équations de locus ont été calculées pour chaque groupe d'enfant. Les résultats affichent une linéarité élevée pour toutes les pentes des tous groupes (moyenne de  $R^2 = 0.93$ ). Les valeurs des pentes des enfants sont élevées pour les bilabiales, intermédiaires pour les vélaires et basses pour les alvéolaires (Table 3.7). Les analyses statistiques sur l'ensemble des trois groupes

	/b/		/d/		/g/	
	P	OO	P	OO	P	OO
3 ans	0.93 (0.13)	-90 (362)	0.42 (0.06)	1623 (219)	0.75 (0.16)	982 (426)
4 ans	0.80 (0.13)	121 (231)	0.40 (0.17)	1585 (445)	0.83 (0.13)	699 (290)
5 ans	0.85 (0.08)	132 (197)	0.54 (0.14)	1270 (415)	0.73 (0.10)	936 (249)

TABLE 3.7 – Valeurs moyennes (et écart-types) des pentes (P) et des ordonnées à l'origine des trois groupes d'enfants (Sussman et al., 1992).

montrent un effet global du lieu d'articulation sur les pentes. Les analyses de comparaison par paire indiquent que la différence de pentes entre /d/ vs. /g/ et /b/ vs. /d/ est significative mais elle ne l'est pas entre /b/ vs. /g/. Quant aux ordonnées à l'origine, un effet significatif du lieu d'articulation est remarqué. Les analyses par paires révèlent une différence significative de toutes les combinaisons. Les paramètres des équations de locus des deux groupes d'âge de quatre et cinq ans ont été comparés afin d'examiner s'ils sont suffisamment sensibles pour détecter les changements de progression avec l'âge<sup>22</sup>. Les pentes et les ordonnées à l'origine sont statistiquement similaires. Les analyses séparées de chaque groupe montrent qu'il existe un effet principal de pente en fonction du lieu d'articulation chez les enfants de quatre ans et chez les enfants de cinq ans. Dans le premier groupe, les analyses de comparaison par paires dévoilent les mêmes résultats qu'avec les analyses globales ; c'est-à-dire une différence de pentes significative entre /d/ vs. /g/ et /b/ vs. /d/ et non significative entre /b/ vs. /g/. Toutefois, les analyses de comparaison par paire du groupe des enfants de cinq ans présentent une différence significative dans toutes les combinaisons. D'après les auteurs, ces résultats mettent en évidence que les enfants de cinq ans contrastent plus les lieux d'articulation que les enfants de quatre ans. Ce postulat peut être considéré comme un indice de repérage de maturité phonologique signalé par les équations de locus. Les auteurs comparent les paramètres des équations de locus entre les enfants et les adultes (avec les données de Sussman et al. (1991)). Une similarité globale des valeurs de pente et d'ordonnée à l'origine a été observée entre les deux groupes.

22. Le groupe d'âge de trois ans n'a pas été inclus car le nombre de sujets était faible (2 enfants de 3 ans comparé à 7 pour les autres groupes).

Les moyennes de pentes sont identiques pour le /b/ dans les deux groupes (0.89), néanmoins elles diffèrent légèrement pour le /d/ (0.49 vs. 0.42) et pour /g/ (0.65 vs. 0.71) (Figure 3.9). Les analyses sur l'erreur type de l'estimation (SE)<sup>23</sup> révèlent que les enfants ont des pentes plus

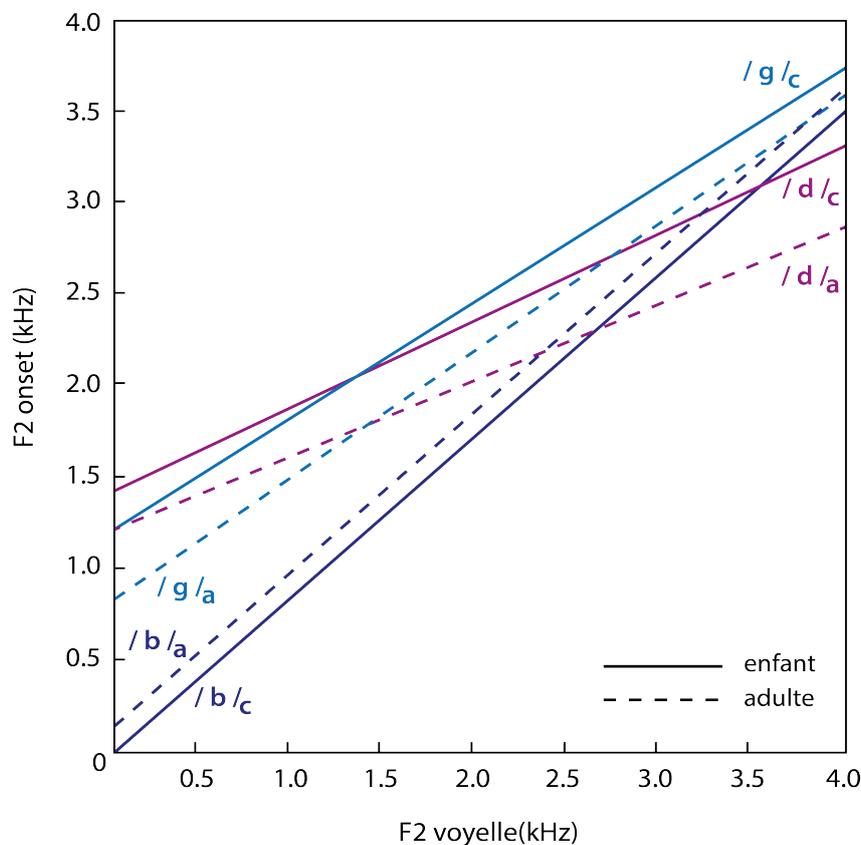


FIGURE 3.9 – Régressions des équations de locus des enfants (en lignes continues) et des adultes (en lignes discontinues) pour /b/, /d/ et /g/ (Sussman et al., 1992).

précises quand ils avancent dans l'âge et que leur performance se rapproche ainsi de celle des adultes. Selon Sussman et al. (1992), les équations de locus codifient l'effet de la coarticulation CV ainsi que le lieu d'articulation indépendamment de la taille du conduit vocal. La différence principale entre les équations de locus des enfants et des adultes est le degré de contrastivité du lieu d'articulation qui est plus marqué chez les adultes.

Dans une étude préliminaire, Sussman et al. (1996) étudient également la capacité des équations de locus à détecter l'évolution de l'acquisition phonétique chez les enfants. Le babillage d'une enfant a été enregistré lors d'une séance de 30 minutes lorsqu'elle était âgée de douze mois puis lors d'une seconde séance d'une durée équivalente, neuf mois plus tard, à l'âge de vingt-et-un mois. À l'âge de douze mois, les séquences jugées similaires à des /b/V, /d/V et /g/V ont été sélectionnées pour calculer les équations de locus (144 séquences CV). À l'âge de vingt-et-un mois, les mots contenant une des trois occlusives en position initiale ont été

23. L'erreur type de l'estimation ("The standard error of estimate", SE) est un outil qui permet de mesurer le degré de précision de l'estimation. Plus l'erreur type est petite, plus l'estimation est précise.

choisis pour les analyses (243 séquences CV)<sup>24</sup>. Les résultats des équations de locus à l'âge de douze et de vingt-et-un mois montrent une "tendance linéaire" des données de cette enfant (moyenne de  $R^2 = 0.55$ ). Toutefois, elle est moins importante que celle des adultes ( $R^2 > 0.90$ ). Les catégories des occlusives gardent des distributions distinctes en fonction du lieu d'articulation (Figure 3.10). Pourtant, les pentes ne suivent pas les mêmes comportements que celles des adultes (enfant /g/ > /b/ > /d/ vs. adulte /b/ > /g/ > /d/). Les valeurs des ordonnées à l'origine sont élevées pour l'enfant en comparaison de celle des adultes. Ces valeurs élevées sont dues aux valeurs élevées des fréquences de  $F2_{onset}$ . Les auteurs avancent que les valeurs

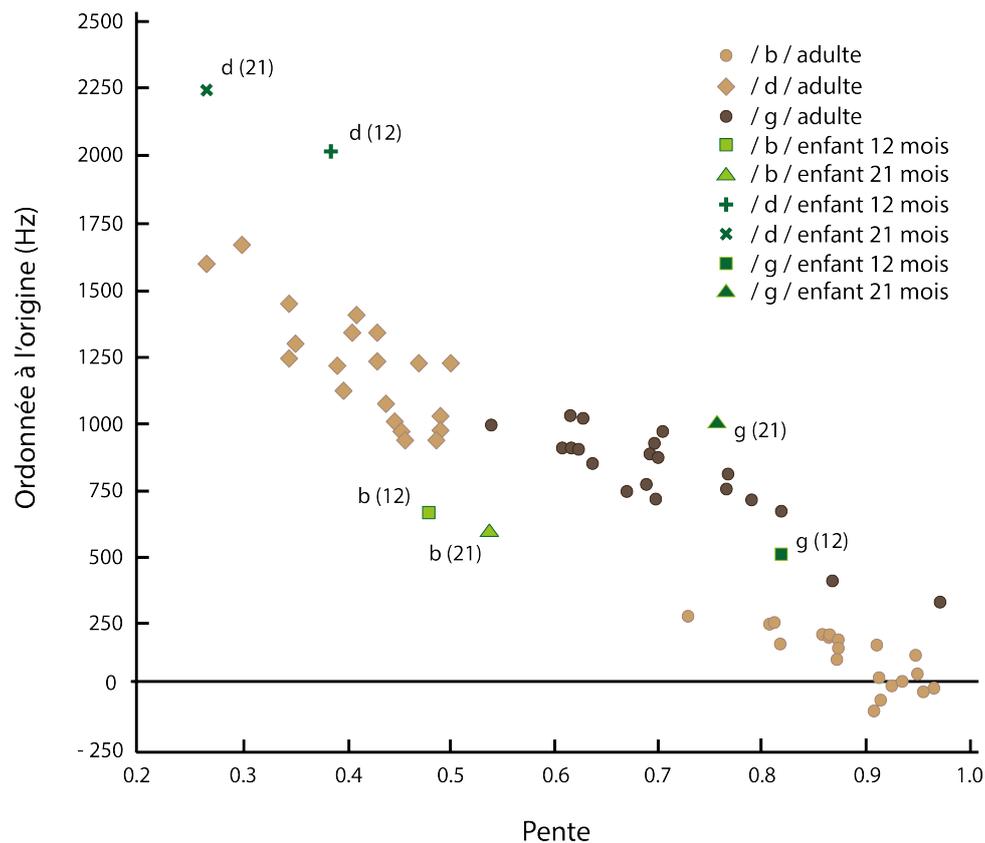


FIGURE 3.10 – Paramètres des équations de locus des données de Sussman et al. (1991) (pente en abscisse et ordonnée à l'origine en ordonnée) comparées à celles de l'enfant aux deux âges (Sussman et al., 1996).

basses de /b/ chez l'enfant comparées à celles des adultes révèlent une absence de coarticulation anticipatoire linguale pendant l'occlusion de la bilabiale. Les valeurs de pente et de  $R^2$  à l'âge de douze mois sont plus proches de celles des adultes en regard des pentes de vingt-et-un mois. Sussman et al. (1996) signalent que l'enfant a tendance à hyper-articuler les segments lors de l'acquisition des premiers mots (– de coarticulation), contrairement au babillage ; les valeurs de pentes s'éloignent alors des normes des adultes (Figure 3.10). Ils ajoutent que les pentes des babillages sont plus élevées (+ de coarticulation) car à ce stade, l'enfant a plus de liberté

24. Des tests de perception ont été effectués sur trois sujets dans le but de vérifier si les séquences sélectionnées dans les deux périodes ne portent pas d'ambiguïté.

pour choisir les voyelles coarticulées puisqu'il n'existe pas de sens à ces séquences. Selon les auteurs, malgré le peu de données (étude basée uniquement sur une enfant), les équations de locus ont fourni un outil fiable pour expliquer le développement de l'acquisition phonologique des enfants.

### 3.4.1.2 Dans d'autres langues

Sussman et al. (1993) exposent que le degré de coarticulation d'un segment pourrait être dépendant de la langue. Le degré de discriminabilité ou de contraste nécessaire pour distinguer les différents lieux d'articulation est lié aux exigences phonologiques de la langue. Les auteurs examinent la validité des équations de locus comme un descripteur du lieu d'articulation dans plusieurs langues : le thaï, l'arabe du Caire et l'ourdou. Toutes les occlusives voisées existantes dans chacune de ces langues ont été produites en position initiale dans des séquences CV ou CVC. En thaï, les occlusives /b, d/ ont été réalisées avec les voyelles /i, æ, u, eɪ, əɪ, æɪ, ɔɪ, a:/ par six locuteurs. En arabe du Caire, les occlusives /b, d, d<sup>ɕ</sup>, g/ ont été articulées avec les voyelles /i, a, u, ʊ, iɪ, æɪ, uɪ, eɪ, oɪ/ par trois locuteurs. En ourdou, les occlusives /b, d, ɖ, g/ ont été prononcées avec les voyelles /ɪ, a, ʊ, ε, iɪ, aɪ, uɪ/ par cinq locuteurs.

C	/b/		/d/		/d <sup>ɕ</sup> /		/ɖ/		/g/	
	P	OO	P	OO	P	OO	P	OO	P	OO
Thaï	0.70	228	0.30	1425	--	--	--	--	--	--
Arabe du Caire	0.77	206	0.25	1307	0.21	933	--	--	0.92	229
Ourdou	0.81	172	0.50	857	--	--	0.44	1070	0.97	212

TABLE 3.8 – Moyennes des valeurs de pente (P) et d'ordonnée à l'origine (OO) des équations de locus des trois langues étudiées (Sussman et al., 1993).

Les résultats présentent une linéarité élevée de la régression des équations de locus dans les langues étudiées ( $R^2$  global de toutes les équations de locus = 0.86). Les valeurs moyennes des pentes et des ordonnées à l'origine figurent dans la Table 3.8. Un effet significatif de pente et d'ordonnée à l'origine en fonction du lieu d'articulation est remarqué dans toutes les langues (avec des analyses multivariées MANOVA). En thaï, les valeurs des pentes de /b/ sont plus élevées que celles de /d/ et à l'inverse, les valeurs des ordonnées à l'origine de /b/ sont moins élevées que celles de /d/. La différence est significative entre les paramètres des équations des locus des deux lieux d'articulation. En arabe du Caire, les valeurs de pentes sont élevées pour /g/, intermédiaires pour /b/ et basses pour /d/ et /d<sup>ɕ</sup>/. Les valeurs des ordonnées à l'origine sont élevées pour /d/, intermédiaires pour /d<sup>ɕ</sup>/ et basses pour /b/ et /g/. Les analyses statistiques sur les pentes indiquent que chaque paire d'occlusives est significativement différente à l'exception des deux coronales : l'alvéolaire /d/ et l'alvéolaire pharyngalisée /d<sup>ɕ</sup>/. Quant aux ordonnées à l'origine, elles sont statistiquement différentes pour chaque paire d'occlusives à

l'exclusion de celles des bilabiale et vélaire. En ourdou, les tendances des valeurs des pentes sont équivalentes à celles observées en arabe pour /b, d, g/. Les analyses de comparaison de pentes entre chaque paire d'occlusives sont significatives pour toutes les comparaisons à l'exception des deux coronales : la rétroflexe /d/ et l'alvéolaire /d/. Les mêmes analyses sur les ordonnées à l'origine signalent des différences significatives de toutes les paires hors celles de /b/ et /g/. Sussman et al. (1993) exposent que les paires de consonnes coronales, en arabe (/d/ vs. /d<sup>ɾ</sup>/) et en ourdou (/d/ vs. /d/), sont les seules occlusives qui n'ont pas révélées de différences significatives entre les pentes des équations de locus. Toutefois, les deux ordonnées à l'origine sont significativement distinctes pour chaque paire d'occlusives. Les auteurs expliquent que cette différence d'ordonnée à l'origine est la conséquence des changements de F2<sub>onset</sub>. Selon eux, comme le degré de coarticulation, reflété par la pente, est similaire dans ces contrastes coronaux, les valeurs des ordonnées à l'origine sont capables d'indexer phonétiquement les contrastes des lieux d'articulation.

Sussman et al. (1993) cherchent l'existence des zones universelles du lieu d'articulation (ou des "*phonetic hot spots*") dans un domaine bidimensionnel des paramètres des équations de locus. Ces *hot spots* sont censés regrouper les mêmes lieux d'articulation des différentes langues dans les mêmes zones en fonction de la pente et de l'ordonnée à l'origine. Les données des équations de locus des études précédentes des deux autres langues ont été ajoutées aux analyses : les données de Lindblom (1963b) et de Krull (1989) pour le suédois et celles de Sussman et al. (1991) pour l'anglais<sup>25</sup>. Des analyses de comparaison des trois lieux d'articulation (bilabial, alvéolaire et vélaire) dans les cinq langues ont été effectuées<sup>26</sup>.

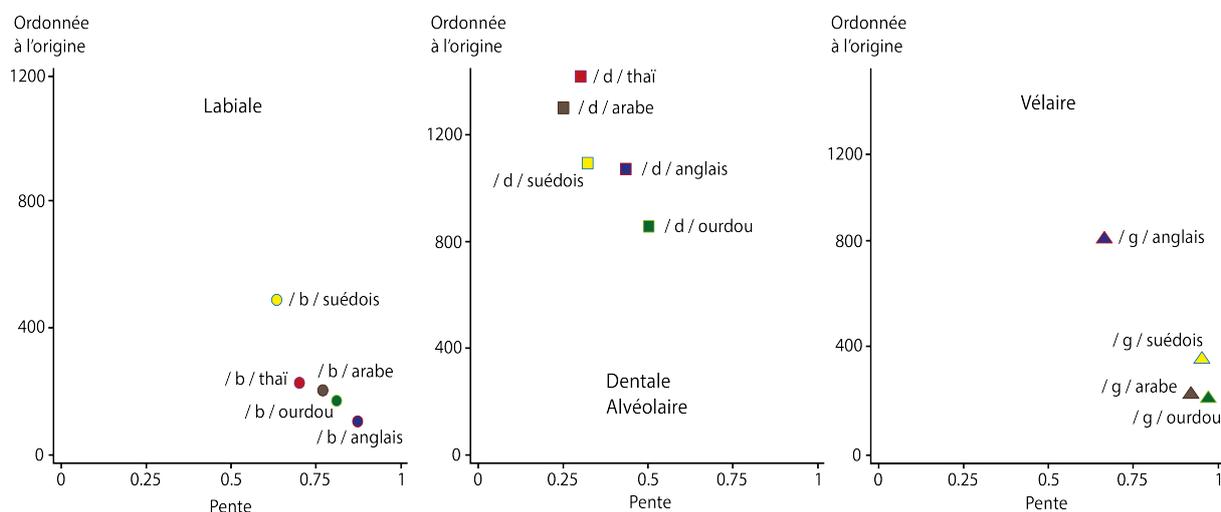


FIGURE 3.11 – Représentation translinguistique du lieu d'articulation en fonction des paramètres des équations de locus dans un domaine bidimensionnel (d'après Sussman et al. (1993)).

25. Ces données figurent dans la Table 3.1 et la Table 3.3 pour le suédois et dans la Table 3.6 pour l'anglais.

26. Il n'a été analysé que deux lieux d'articulation (bilabial et alvéolaire) de thaï et de suédois. Les raisons de cette sélection sont que le thaï ne contient pas d'occlusive vélaire voisée, que les données de la vélaire de l'étude de Lindblom (1963b) ne reposaient que sur un seul locuteur, elles ont donc été éliminées des analyses statistiques et que Krull (1989) n'a pas traité de vélaire dans son étude.

Les auteurs soulignent que les occlusives bilabiales des cinq langues se regroupent dans une zone relativement étroite dans l'espace des équations de locus (Figure 3.11). Les /b/ de l'anglais ou du thaï sont significativement similaires à ceux de l'arabe du Caire et de l'ourdou (absence d'effet de la langue sur la pente et sur l'ordonnée à l'origine). Un *hot spot* est donc identifié pour les bilabiales pour au moins trois langues. Selon les auteurs, la stabilité du /b/ trans-linguistique est due au lieu de constriction bilabiale homogène. Quant aux coronales et aux dorsales, elles ont un degré de liberté plus élevé lors de la formation des lieux de constriction. La bilabiale en suédois atteste un degré de coarticulation moins élevé que dans les autres langues. Sussman et al. (1993) suggèrent que ce phénomène peut être expliqué par l'existence d'un facteur de coarticulation relativement réduit pour les bilabiales suédoises. En effet, il n'existe pas de regroupement clair des coordonnées qui émerge en comparant les occlusives alvéolaires /d/ dans les cinq langues (Figure 3.11). Les /d/ des langues étudiées sont ainsi dispersés dans la zone centrale de l'espace des équations de locus. Aucun *hot spot* n'est trouvé pour ce lieu d'articulation puisque les tests statistiques montrent que toutes les alvéolaires sont significativement différentes au moins dans une seule dimension. Les vélaires de l'arabe du Caire et de l'ourdou sont statistiquement similaires. Quant à celle de l'anglais, elle est significativement différente des deux autres /g/ dans les deux dimensions. Les auteurs signalent que les valeurs des pentes et des ordonnées à l'origine du /g/ de l'anglais canadien, analysées par Nearey et Shamma (1987), sont plus proches de celles de l'arabe du Caire et de l'ourdou que de l'anglais américain (*pente* = 0.99, *ordonnée à l'origine* = 215). En anglais américain, Sussman et al. (1991) ont observé deux pentes différentes de /g/ en fonction des voyelles adjacentes antérieures ou postérieures. Le /g/vélaire<sup>27</sup> produit avec les voyelles postérieures a une pente et une ordonnée à l'origine comparables à celles des /g/ des langues citées plus haut (*pente* = 0.87 et *ordonnée à l'origine* = 559). Sussman et al. (1993) expliquent que les locuteurs américains ont une tendance forte à antérioriser les vélaires devant les voyelles antérieures et que cette alternance allophonique affecte d'une manière importante la régression globale de /g/. Ce phénomène conduit à une pente basse et à une ordonnée à l'origine élevée comparées aux autres langues. Les auteurs déclarent que les résultats ne sont pas favorables à l'existence de *hot spot* du lieu d'articulation des occlusives étudiées. Ils concluent que « [...] *the data of this study point to the conclusion that languages tend to use "b-like", "d-like", and "g-like" initial stops coarticulatorily tuned by the specific vowel inventory that follows* »<sup>28</sup>.

La capacité des équations de locus à caractériser les lieux d'articulation des occlusives a également été examinée dans la langue espagnole par Celdran et Villalba (1995). Dix locuteurs espagnols (5 hommes et 5 femmes) ont prononcé des séquences /Kan'CVna/ où C = /b, p, d, t, g, k/ et V = /i, a, u, e, o/. Les séquences ont été répétées cinq fois. Les équations de locus ont été calculées pour chacune des catégories (occlusives voisée + non voisée). Les résultats

27. Opposé au /g/palatal produit avec les voyelles antérieures. Ce dernier a une pente de 0.34 et une ordonnée à l'origine de 1602 dans l'étude de Sussman et al. (1991).

28. [...] les données de cette étude indiquent que les langues ont tendance à utiliser des "quasi-b", "quasi-d" et "quasi-g" en position initiale, ajustés coarticulatoirement par les voyelles qui les suivent.

montrent des pentes élevées pour les vélares, intermédiaires pour les bilabiales et basses pour les alvéolaires (Table 3.9). Les analyses statistiques ne signalent pas de différence significative

Catégorie	Bilabiale /b, p/		Alvéolaire /d, t/		Vélaire /g, k/	
	P	OO	P	OO	P	OO
Moyenne	0.83	126	0.58	746	1	29
Écart-type	0.05	79	0.07	161	0.09	139

TABLE 3.9 – Valeurs moyennes et écart-types des pentes (P) et des ordonnées à l'origine (OO) des trois catégories consonantiques des 10 locuteurs (Celdran et Villalba, 1995).

de pente ou d'ordonnée à l'origine en fonction du sexe du locuteur. Les pentes ainsi que les ordonnées à l'origine des trois classes de consonne sont significativement différentes. Les analyses discriminantes linéaires effectuées avec les pentes et les ordonnées à l'origine révèlent un taux de classification correcte de 100%. En comparant les valeurs de pente en espagnol avec celles de l'anglais canadien (Nearey et Shammass, 1987), de l'anglais américain (Sussman et al., 1991), du suédois (Krull, 1989), du thaï, de l'arabe du Caire et de l'ourdou (Sussman et al., 1993), Celdran et Villalba (1995) soulignent que les valeurs de pente des bilabiales et celles des vélares sont trans-linguistiquement cohérentes. Toutefois, les valeurs de pente des /d, t/ sont plus élevées que celles des autres langues. Les auteurs précisent que ces deux consonnes sont des dento-alvéolaires et non des alvéolaires comme dans la plupart de ces langues. Ils ajoutent que les pentes les plus proches sont celles en ourdou (0.50) car le /d/ dans cette langue est également dento-alvéolaire.

Les propositions de Sussman et ses collègues sur la capacité des équations de locus à caractériser les lieux d'articulation ont été remises en cause par plusieurs chercheurs, notamment par Fowler (1994).

### 3.4.1.3 Critiques de l'invariant relationnel

Fowler (1994) examine la fiabilité des équations de locus comme un indicateur du lieu d'articulation. L'auteure souligne que les variations interlinguistique et interlocuteur de la pente observées dans les études précédentes (notamment Sussman et al. (1991, 1993)) ne soutiennent pas l'hypothèse d'un invariant du lieu d'articulation. De plus, seules les occlusives voisées ont été analysées dans ces travaux. Selon elle, un invariant durable du lieu d'articulation doit expliquer la structure des caractéristiques phonétiques comme les utilisateurs de la langue la perçoivent. Ainsi, n'importe quel invariant du lieu d'articulation doit rester invariant globalement et indépendant du mode d'articulation. Fowler (1994) précise qu'une pente positive est obtenue lors du calcul des équations de locus parce que les locuteurs coarticulent. Par conséquent, quand la voyelle a une fréquence de F2 élevée, la fréquence de  $F2_{onset}$  sera aussi relativement élevée car la production de la voyelle commence avant le relâchement de la consonne. De même, lorsque la voyelle a une fréquence de F2 basse, la fréquence de  $F2_{onset}$  sera relativement basse également.

Les pentes des équations de locus reflètent ainsi le degré du chevauchement entre la consonne et la voyelle. Elle considère que les valeurs de pente obtenues par Sussman et al. (1991) peuvent être interprétées comme un indicateur de degré de résistance coarticulatoire. Elles fournissent - indirectement - des informations sur le lieu d'articulation puisque les variations du lieu d'articulation est une source importante de la variation de résistance coarticulatoire.

Fowler (1994) met en place un protocole expérimental afin d'examiner ces arguments. Dix locuteurs de l'anglais américain ont prononcé des séquences CVt (où C = /b, v, ð, d, z, ʒ, g/ et V = /iy, I, æ, ʌ, a, ɔ, uw/). Les résultats des équations de locus des occlusives voisées (/b, d, g/)<sup>29</sup> montrent des valeurs de pente moins élevées que celles observées dans l'étude de Sussman et al. (1991) (Table 3.10)<sup>30</sup>. Les analyses de comparaison révèlent que les pentes de chaque paire d'occlusives sont significativement différentes à l'exception de celle de /b/ vs. /g/ où la différence est marginale. Quant aux ordonnées à l'origine, elles sont significativement différentes dans toutes les combinaisons. L'impact de l'antériorité vs. postériorité des voyelles sur les pentes a été testé pour les trois occlusives. Les valeurs des pentes des occlusives en contexte de voyelles antérieures sont moins élevées que celles en contexte de voyelles postérieures. Un effet significatif du contexte vocalique sur les pentes a été observé. Cette constatation indique que les occlusives sont plus résistantes devant les voyelles antérieures que devant les voyelles postérieures. La différence de pente en fonction du contexte vocalique est particulièrement mar-

C	/b/		/d/		/g/		/v/		/ð/		/z/		/ʒ/	
	P	OO	P	OO	P	OO	P	OO	P	OO	P	OO	P	OO
	0.80	244	0.47	1120	0.68	815	0.73	336	0.50	902	0.42	1078	0.34	1407

TABLE 3.10 – Moyennes des valeurs de pente (P) et d'ordonnée à l'origine (OO) des équations de locus des sept consonnes étudiées (Fowler, 1994).

quée pour /g/. Par ailleurs, la pente de /g/ devant les voyelles antérieures est la plus élevée de toutes les consonnes alors qu'elle est la plus basse devant les voyelles postérieures. Une absence de différence significative a également été notée entre les pentes de /b/ vs. /g/ produits avec les voyelles postérieures. L'auteure précise qu'il n'est pas nécessaire de calculer deux équations de locus pour l'occlusive /g/ selon le contexte vocalique (antérieur vs. postérieur). Elle avance les arguments suivants pour justifier sa proposition :

1. Dans ses données et dans celles de Sussman et al. (1991), les pentes de /g/ avec les voyelles antérieures ne sont pas distinguées de celles de /b/ dans le même contexte.

29. Dans un premier temps, l'auteure analyse les trois occlusives afin de les comparer avec les données de Sussman et al. (1991). Dans un second temps, l'ensemble des sept consonnes est examiné dans le but de vérifier la fiabilité des équations de locus dans les différents modes d'articulation.

30. Il est important de remarquer que les calculs des équations de locus et les analyses discriminantes linéaires ne sont pas effectuées sur des valeurs moyennées, comme c'est le cas dans Sussman et al. (1991). Fowler (1994) précise que ce comportement ne représente pas la réalité car un auditeur entend une réalisation singulière à chaque fois et non la moyenne des cinq réalisations.

2. Les deux droites de régression séparées ne semblent pas améliorer l'ajustement des points de données dans la régression des équations de locus. Les moyennes des  $R^2$  sont égales à 0.74 et 0.71 pour les contextes des voyelles antérieures et postérieures respectivement. Le  $R^2$  d'une seule droite de régression (avec les deux contextes regroupés) de /g/ est égale à 0.88.
3. Il n'existe aucune preuve que les auditeurs entendent des consonnes distinctes pour le /g/ ; au contraire, il existe des preuves comme quoi ils n'entendent qu'un seul /g/.

Les résultats des analyses discriminantes sur  $F2_{onset}$  et  $F2_{mid}$  des trois occlusives désignent un taux de classification moins important que celui observé par Sussman et al. (1991) (en moyenne 66 % contre 77%).

En parallèle, des analyses statistiques des équations de locus ont été effectuées sur l'ensemble des consonnes. Un effet global de pente en fonction du lieu d'articulation a été relevé. Dans les analyses de comparaison par paires, les pentes sont significativement différentes pour la plupart des consonnes dont celles qui ont le même lieu d'articulation (/z/ vs. /d/). La fricative a une pente plus basse que celle de l'occlusive (Table 3.10). Selon l'auteure, cette différence de pente résulte de la résistance coarticulatoire élevée des fricatives en comparaison des occlusives. D'un autre côté, elle observe que des consonnes qui ne partagent pas le même lieu d'articulation (/g/ vs. /v/ et /z/ vs. /d/) ont des pentes statistiquement équivalentes. Fowler (1994) met en évidence que ce phénomène provient du fait que les consonnes dans chacune de ces paires ont une résistance coarticulation similaire. Les analyses discriminantes linéaires, avec  $F2_{onset}$  et  $F2_{mid}$  comme prédicteurs, ont été utilisées de manière à classifier les sept consonnes. Le taux de classification correcte est relativement faible (40 % pour les hommes et 33 % pour les femmes) mais il est supérieur au hasard<sup>31</sup>. L'auteure indique qu'il n'existe pas de lien entre les résultats des analyses discriminantes sur les  $F2_{onset}$  et  $F2_{mid}$  et ceux des équations de locus. Les consonnes /b/ et /v/ ont des pentes et des ordonnées à l'origine distinctes mais une grande partie des /b/ a été classifiée dans la catégorie du /v/ tandis qu'une petite partie a bien été catégorisée dans la classe /b/ (10.5 % comme /b/ vs. 53.5% comme /v/). Les consonnes /v/ et /g/, qui ont la même pente, sont rarement confondus dans les LDA. Fowler (1994) résume que les résultats de cette étude montrent que les équations de locus offrent peu d'informations sur le lieu d'articulation. Toutefois, elles semblent fournir une mesure fiable du degré de chevauchement coarticulatoire.

Sussman et ses collègues ont établi un ensemble d'études de façon à répondre aux critiques de Fowler (1994). Ces études ont pour objectif de tester la stabilité des équations de locus comme étant un indicateur du lieu d'articulation dans plusieurs modes d'articulation et dans une articulation compensatoire (Sussman, 1994; Sussman et al., 1995; Sussman et Shore, 1996).

---

31. Le hasard est égal à 14.3% (1/7) pour toutes les catégories à l'exception de /d/ et /z/ où le hasard est équivalent à 28.6% (2/7).

### 3.4.1.4 Dans différents modes d'articulation

Sussman (1994) cherche à déterminer si les pentes des équations de locus captureraient d'une façon uniforme le lieu d'articulation dans des modes d'articulation différents : occlusives, fricatives et approximantes. Autrement dit, le but est de savoir si les alvéolaires /d, t, n, s, z, l, r/ sont caractérisées par des pentes des équations de locus similaires parce qu'elles partagent toutes le même lieu d'articulation. Dans un premier temps, dix locuteurs américains (5 femmes et 5 hommes) ont prononcé des occlusives non voisées non-aspirées dans des séquences sCV (où C = /p, t, k/ et V = /i, ɪ, ε, e, æ, a, ʌ, ɔ, o, u/). Dans un second temps, quatre locuteurs américains (2 femmes et 2 hommes) ont produit des mots CVC qui commencent par /m, n, v, ð, z, ʒ, w, j, r, l/ et qui contiennent les mêmes voyelles citées plus haut. Chacune des séquences sCV et CVC a été répétée trois fois. Les valeurs de  $F2_{onset}$  sont toujours mesurées à la première pulsation de la voyelle après le relâchement sauf pour les approximantes où elles sont prélevées au début de la consonne. Les données des occlusives voisées /b, d, g/ de Sussman et al. (1991) ont été incluses dans cette étude (Table 3.6).

	/m/				/n/			
	P	OO	P	OO	P	OO	P	OO
Nasales	0.86	86			0.32	1421		
Occlusives non voisées	/p/		/t/		/k/			
	P	OO	P	OO	P	OO	P	OO
	0.79	202	0.44	1120	0.88	414		
Fricatives	/v/		/ð/		/z/		/ʒ/	
	P	OO	P	OO	P	OO	P	OO
	0.74	306	0.34	1179	28	1342	0.37	1295
Approximantes	/w/		/j/		/r/		/l/	
	P	OO	P	OO	P	OO	P	OO
	-0.01	751	0.05	2569	0.07	869	0.08	1252

TABLE 3.11 – Valeurs moyennes des pentes (P) et des ordonnées à l'origine (OO) pour chaque consonne (Sussman, 1994).

Les résultats indiquent des pentes élevées pour la nasale bilabiale /m/ spécifiant une coarticulation importante par comparaison avec la nasale alvéolaire /n/ (Table 3.11). Ce postulat est expliqué par le degré de l'engagement de la langue pendant la production de la consonne. Dans le premier cas (/m/), la langue n'est pas engagée directement dans la production de cette consonne alors que dans le second (/n/), elle l'est. Les pentes de /m/ sont semblables à celles de /b/ et les pentes de /n/ sont comparables à celles de /t/. Les valeurs des pentes des occlusives non voisées décroissent selon l'ordre suivant : la vélaire /k/ > la bilabiale /p/ > l'alvéolaire /t/ (Table 3.11). Ces résultats ne sont pas en accord avec ceux des occlusives voisées de l'anglais américain mentionnés dans l'article de Sussman et al. (1991) où l'ordre des valeurs de pentes

est : la bilabiale /b/ > la vélaire /g/ > l'alvéolaire /d/. Toutefois, cet ordre de pentes (des non voisées) est conforme avec celui trouvé en ourdou et en arabe du Caire (Sussman et al., 1993). Il semble donc qu'il existe un effet spécifique de la langue sur le degré de chevauchement coarticulatoire entre les bilabiales et vélaire et les voyelles suivantes. Les fricatives ont des pentes similaires (relativement basses) à l'exception de la labiodentale /v/ qui a des pentes plus élevées. Sussman (1994) explique que les catégories de coronales (dentale, alvéolaire et post-alvéolaire) ne peuvent pas être distinguées par les paramètres des équations de locus. Les valeurs des pentes des approximantes révèlent des valeurs proches de zéro, indiquant une absence de coarticulation. Ces valeurs très basses de pente peuvent être expliquées par le choix de mesure de  $F2_{onset}$  qui ne reflète pas la frontière entre la consonne et la voyelle. Le début de la consonne (le moment de la mesure de  $F2_{onset}$ ) est physiquement très loin du milieu de la voyelle ( $F2_{mid}$ ). De plus, les transitions formantiques entre les approximantes et les voyelles sont relativement lentes. Par conséquent, une absence d'influence mutuelle prédomine entre le début de la consonne et le milieu de la voyelle. Une représentation bidimensionnelle des paramètres des équations de locus est mise en place, où les pentes sont en abscisses et les ordonnées à l'origine sont en ordonnée de toutes les consonnes. Dans cette représentation, les consonnes qui partagent des lieux d'articulation identiques mais des modes d'articulation différents se regroupent ensemble. Trois groupes distincts se sont constitués : les coronales /ð, d, t, n, z, ʒ/, les vélaire /g, k/ et les labiales/labiodentales /b, p, m, v/. Quant aux approximantes, elles forment un nuage de points autour de la valeur zéro de la pente qui s'étend verticalement en fonction de la valeur de  $F2_{onset}$ . Sussman (1994) conclut que les résultats de cette étude soutiennent l'efficacité des équations de locus dans la catégorisation du lieu d'articulation indépendamment du mode d'articulation. Pourtant, il n'a pas effectué de comparaison statistique.

Sussman et Shore (1996) répliquent et étendent une partie de l'expérience de Fowler (1994) en comparant cinq alvéolaires qui ont des modes d'articulation différents (occlusion, friction, nasalité, voisement). L'objectif de l'étude est d'examiner la capacité des équations de locus à décrire les paramètres communs des consonnes qui partagent un même lieu d'articulation mais prononcées dans différents modes d'articulation. Vingt-deux locuteurs masculins de l'anglais américain ont produit des séquences CVt où C = /d, t, n, z, s/ et V = /i, I, e, ε, æ, a, A, o, u/. Chaque séquence a été réalisée dans une phrase porteuse et répétée cinq fois. Les auteurs prennent en compte la critique de Fowler (1994) et ne moyennent pas les répétitions lors du calcul des équations de locus. Ces dernières sont donc estimées avec 50 mesures pour chaque consonne (10 voyelles × 5 répétitions). Les  $F2_{onset}$  sont mesurés à la première pulsation de la voyelle pour /d/, /n/, /z/ et /s/. La position de prélèvement de  $F2_{onset}$  de /t/ est modifiée afin de prendre en considération la période d'aspiration longue qui suit le relâchement de la consonne. Sussman et Shore (1996) expliquent qu'une mesure traditionnelle de  $F2_{onset}$  pour /t/ à la première pulsation de la voyelle mènera à une valeur de  $F2_{onset}$  semblable à celle de  $F2_{mid}$ . De ce fait, les équations de locus calculées avec cette mesure ne reflètent pas les caractéristiques de la consonne et conduisent à des valeurs élevées de pente. L'influence de l'occlusive non

voisée sur le début des transitions formantiques est freinée par la durée importante du VOT. Les auteurs soulignent qu'une mesure de F2 plus proche de la barre du relâchement entraîne une fréquence de  $F2_{onset}$  plus représentative de la consonne. Le  $F2_{onset}$  du /t/ est prélevé à la première résonance visible de F2 suivant le bruit de relâchement (choisie par les auteurs). Ces derniers ont jugé que ce point, cette première résonance, est le prolongement de F2 de la voyelle (Figure 3.12). Cette mesure de  $F2_{onset}$  proche du bruit de relâchement est appelé  $F2_{@burst}$ . Les équations de locus de l'occlusive /t/ et de /d/ ont été calculées avec  $F2_{@burst}$  et  $F2_{mid}$  afin de comparer les mesures des deux consonnes. Les équations de locus de /d/ ont également été estimées avec les coefficients traditionnels ( $F2_{onset}$  et  $F2_{mid}$ ) pour les comparer avec le reste des consonnes.

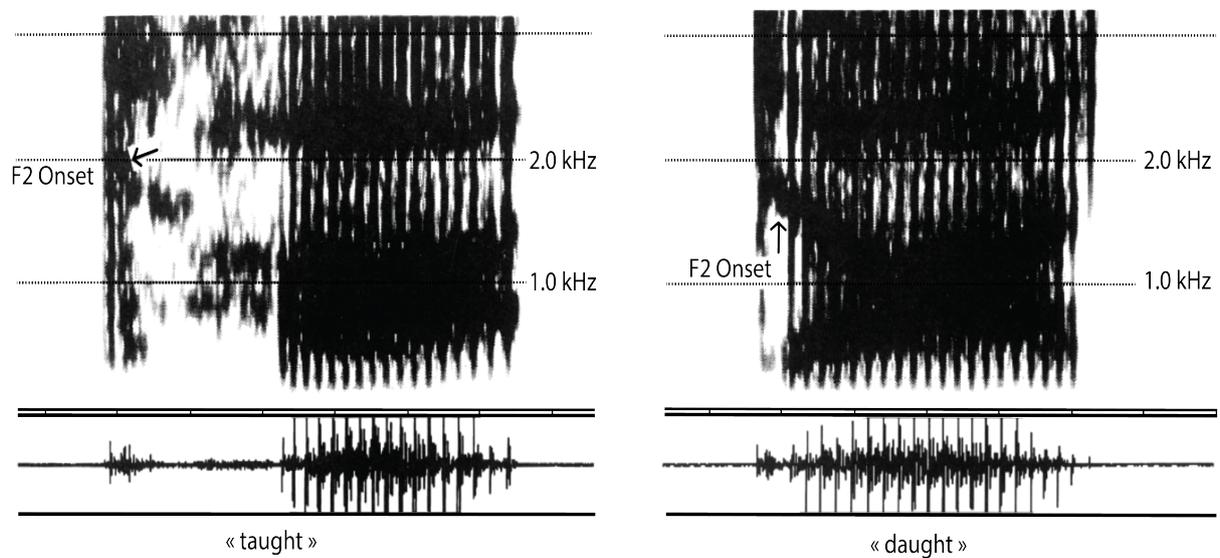


FIGURE 3.12 – À droite, le spectrogramme du mot "taught" de la même personne montrant l'apparence de la résonance de F2 dans l'intervalle de l'aspiration. La flèche indique l'endroit où la valeur du  $F2_{@burst}$  du /t/ est mesurée. À gauche, le spectrogramme du mot "daught" montrant la présence d'une résonance de F2 pendant l'intervalle court de VOT qui suit le relâchement de /d/. La flèche indique la position du prélèvement de  $F2_{@burst}$  du /d/ (empruntée à Sussman et Shore (1996)).

Les pentes des équations de locus des cinq modes d'articulation ont des valeurs relativement basses et des ordonnées à l'origine élevées (Table 3.12). Les analyses statistiques multivariées (manova) avec la pente et l'ordonnée à l'origine comme variables dépendantes montrent un effet significatif du mode d'articulation. Les analyses de comparaison par paire sur les pentes ne désignent aucune différence significative entre : /d/ vs. /n/, /d/ vs. /z/, /d/@burst vs. /t/ et /n/ vs. /s/. Les trois consonnes voisées /d, n, z/ ont des pentes statistiquement similaires. Les analyses de comparaison par paires sur les ordonnées à l'origine indiquent des différences non-significatives entre : /d/ vs. /z/, /d/@burst vs. /t/ et /n/ vs. /s/. Sussman et Shore (1996) mettent en exergue que les résultats de cette étude contredisent ceux de Fowler (1994) qui observe une différence statistique de pente entre le /d/ et le /z/ (0.47 et 0.42 respectivement). Cette différence n'a pas été constatée dans l'étude de Sussman et Shore (1996). Le /s/ est la

C	/d/		/d/@burst		/t/		/n/		/z/		/s/	
	P	OO	P	OO	P	OO	P	OO	P	OO	P	OO
Moyenne	0.40	1117	0.26	1418	0.23	1495	0.48	899	0.38	942	0.57	643
Écart-type	0.075	158	0.062	117	0.66	156	0.089	203	0.068	134	0.097	169

TABLE 3.12 – Valeurs moyennes et écart-types des pentes (P) et des ordonnées à l'origine (OO) des cinq classes des modes d'articulation (Sussman et Shore, 1996).

seule consonne qui témoigne d'une différence significative de pente et d'ordonnée à l'origine par rapport aux autres consonnes (à l'exception de la pente de /n/). Les auteurs déclarent que les pentes significativement élevées de /s/ peuvent être le résultat d'une limitation de mesure et non d'un phénomène d'origine phonétique. En prenant les valeurs de  $F2_{onset}$  à la première pulsation glottale de la voyelle qui suit l'intervalle de friction long du /s/, le mouvement de transition de F2 est déjà accompli. De ce fait, les fréquences de  $F2_{onset}$  sont fortement conditionnées par celles du milieu de la voyelle ( $F2_{mid}$ ) et non par le locus réel de la consonne. Cela conduit à des valeurs de pentes plus élevées que dans le cas où les "onsets" sont capables d'être évalués à un moment plus précoce dans l'intervalle de friction.

Sur une autre thématique, Sussman et Shore (1996) ont utilisé les paramètres des équations de locus des trois occlusives (/b, d, g/) présentés dans l'article de Sussman et al. (1991) dans le but de les comparer avec les données de cette étude. Une représentation bidimensionnelle (de second ordre), de pente en abscisse et d'ordonnée à l'origine en ordonnée, a été établie pour les deux ensembles de données. Dans cet espace bidimensionnel, chaque catégorie de lieu d'articulation est représentée par un nuage de point distinct. En ce qui concerne les alvéolaires, elles s'étendent sur une zone de pente relativement large (entre 0.1 pour une pente de /t/ et 0.72 pour une pente de /s/). Les auteurs expliquent que cette variation de pente dans la même classe de lieu d'articulation est la conséquence du choix de prélèvement de la valeur de  $F2_{onset}$  de /t/ et de /s/. Une valeur de  $F2_{onset}$  proche du bruit de relâchement ( $F2_{@burst}$ ) mène à un minimum de coarticulation comparée à une valeur mesurée à la première pulsation de la voyelle. Sussman et Shore (1996) soulignent que malgré la variation de pente notée dans les différents modes d'articulation, les alvéolaires ne se chevauchent pas avec les bilabiales et les vélares dans l'espace de second ordre des équations de locus. Des analyses discriminantes linéaires avec la pente et l'ordonnée à l'origine comme prédicteurs ont été utilisées dans l'attention de séparer les trois lieux d'articulation<sup>32</sup>. Les alvéolaires ont été classées correctement avec un taux de 87%. Les auteurs concluent : « *The overall results of this study can be taken to support the contention that locus equations can indeed be used as a general phonetic descriptor of place of articulation. As a class, alveolar obstruents, even across phonetically diverse manner classes, will have lower slopes (less CV coarticulation) characterized by more stable onset loci*

32. 172 pentes et ordonnées à l'origine ont été employées dans les LDA. La classe des alvéolaires regroupe 132 pentes et ordonnées à l'origine et les autres classes en contiennent 20 chacune.

than will labial and velar places of articulation ». <sup>33</sup>

Sur le même sujet d'étude, Yeou (1997) examine la stabilité des équations de locus comme un descripteur du lieu d'articulation dans les différents modes d'articulation en arabe standard. Neuf locuteurs masculins de l'arabe marocain ont produit des séquences CVCVC(VC) de l'arabe standard où les consonnes initiales correspondent à /f, ð, ð<sup>ʕ</sup>, s, s<sup>ʕ</sup>, d, d<sup>ʕ</sup>, t, t<sup>ʕ</sup>, ʃ, ħ, ʁ/. Chacune de ces consonnes a été suivie par les voyelles /i, æ, u, i:, æ:, u:/ . Au total, chaque locuteur a prononcé 360 mots (12 consonnes × 6 voyelles × 5 répétitions). Dans un premier temps,

	/ð/		/d/		/s/		/t/	
	P	OO	P	OO	P	OO	P	OO
moyenne	0.46	875	0.48	936	0.56	741	0.66	623
écart-type	0.08	137	0.05	135	0.08	105	0.05	96
	/ð <sup>ʕ</sup> /		/d <sup>ʕ</sup> /		/s <sup>ʕ</sup> /		/d <sup>ʕ</sup> /	
	P	OO	P	OO	P	OO	P	OO
moyenne	0.22	778	0.31	839	0.35	681	0.37	678
écart-type	0.07	91	0.09	187	0.06	64	0.07	93
	/f/		/ʃ/		/ʁ/		/ħ/	
	P	OO	P	OO	P	OO	P	OO
moyenne	0.92	61	0.62	694	0.90	106	0.68	587
écart-type	0.02	49	0.10	149	0.06	70	0.11	135

TABLE 3.13 – Valeurs moyennes et écart-types des pentes (P) et des ordonnées à l'origine (OO) des consonnes étudiées des 9 locuteurs (Yeou, 1997).

l'auteur analyse les paramètres des équations de locus des fricatives avec ceux des occlusives décrits dans les études de Sussman et al. (1991) et de Fowler (1994) qui partagent des lieux d'articulation similaires. Les fricatives labiale, dentale et uvulaire (/f, s, ʁ/) sont comparées avec les occlusives labiale, alvéolaire et vélaire (/b, d, g/). Les pentes des équations de locus des labiales sont plus élevées que celles des alvéolaires dans les expériences de Sussman et al. (1991) et Fowler (1994). Cet effet a été répliqué dans cette recherche (Table 3.13). Yeou (1997) reprend l'explication de Fowler (1994) sur le fait que ces valeurs de pente résultent du degré de l'engagement de la langue lors de la production de la consonne. Il est minimal pour les labiales et maximal pour les alvéolaires. Toutefois, l'uvulaire /ʁ/ qui implique également un engagement de la langue a des pentes plus élevées que le /s/. L'auteur signale que ce phénomène peut être lié à l'absence de contraste entre la vélaire /x/ et l'uvulaire /ʁ/ entraînant une coarticulation vocalique plus importante. Cette interprétation est aussi avancée par Fowler (1994) pour la vélaire /g/ qui a une valeur de pente élevée malgré la participation de la langue. Dans un second temps, les équations de locus des douze consonnes ont été comparées. Les analyses sta-

33. « Les résultats globaux de cette étude soutiennent l'hypothèse que les équations de locus peuvent être utilisées comme un descripteur phonétique général du lieu d'articulation. Comme catégorie, les obstruantes alvéolaires auront toujours des pentes plus basses (moins de coarticulation CV) caractérisées par des *onsets* plus stables que celles des bilabiales et des vélaire, même dans divers modes d'articulation », Sussman et Shore 1996, page 945, [notre traduction].

tistiques sur l'ensemble des consonnes révèlent un effet principal de la consonne sur les pentes et sur les ordonnées à l'origine. Des analyses de comparaison par paires sur les pentes et les ordonnées à l'origine montrent que la plupart des consonnes sont significativement différentes. Les pentes de l'alvéolaire /t/ sont statistiquement plus élevées que celles des deux autres alvéolaires /d, s/ (Table 3.13). L'auteur met en exergue que cette différence peut être provoquée par une limitation de mesure et non par une origine phonétique<sup>34</sup>. Il ajoute que la proposition de Sussman et Shore (1996) sur la mesure  $F2_{onset}$  au moment du relâchement de l'occlusive non voisée ( $F2_{@burst}$ ) peut représenter une solution pour le /t/. Il indique que les valeurs relativement élevées des pentes de la fricative /s/ seraient également dues à des problèmes de mesure, reprenant ainsi le raisonnement de Sussman et Shore (1996)<sup>35</sup>.

Yeou (1997) déclare que les paramètres des équations de locus ne permettent pas de distinguer certains lieux d'articulation dans ces données. Les pentes ainsi que les ordonnées à l'origine ne sont pas statistiquement distinctes entre les catégories de lieux d'articulation suivantes : la post-alvéolaire /ʃ/ et la pharyngale /ħ/, l'alvéolaire /d/ et la dentale /ð/ et la labiale /f/ et l'uvulaire /χ/. Selon l'auteur, les équations de locus ne représentent pas un bon indicateur du lieu d'articulation puisque les pentes des consonnes appartenant au même lieu d'articulation sont distinctes et celles possédant des lieux d'articulation différents ont des pentes similaires. Par contre, les équations de locus réussissent à différencier les consonnes pharyngalisées des consonnes non-pharyngalisées. Ces dernières ont des pentes plus élevées que celles des pharyngalisées. Les analyses de comparaison par paires indiquent que les pentes et une majorité des ordonnées à l'origine sont significativement différentes pour ces deux catégories. Ces résultats dévoilent que les pharyngales constituent une catégorie consonantique distincte des autres catégories. La stabilité du locus ( $F2_{onset}$ ) dans les divers contextes vocaliques entraîne des valeurs basses de pente des pharyngalisées et par conséquent une résistance coarticulatoire maximale. En effet, cette constatation est liée à la double articulation exigée pour les pharyngalisées, ce qui conduit à un engagement plus important de la langue que dans les autres catégories consonantiques.

### 3.4.1.5 Dans une articulation compensatoire

Dans la recherche de la validation de leur proposition, Sussman et ses collègues examinent la stabilité de la relation linéaire entre le début et le milieu du deuxième formant dans le cas de l'articulation compensatoire provoquée par un "bite-block" (Sussman et al., 1995). L'objectif de cette étude est également de vérifier la constance des équations de locus comme un indice

34. Pour les occlusives non voisées, le relâchement de la consonne et le début de la transition sont séparés par un intervalle temporel (VOT) important en comparaison des occlusives voisées. Ce phénomène implique une diminution de l'information consonantique portée au début de la transition. Selon Sussman et Shore (1996), la mesure classique de  $F2_{onset}$  à la première pulsation de la voyelle n'est pas valable pour les occlusives non voisées.

35. Sussman et Shore (1996) exposent que le bruit de friction relativement long diminue le transfert de l'information consonantique au début des transitions formantiques de  $F2_{onset}$ . Ils suggèrent que la position la plus représentative du  $F2_{onset}$  de /s/ se trouve dans le bruit de friction. Actuellement, aucune technique n'est proposée pour le détecter.

de lieu d'articulation. Les auteurs soulignent que la production de la parole nécessite un degré élevé de flexibilité adaptative afin de maintenir les cibles acoustiques. Ils ajoutent que, dans un cas extrême où l'articulation normale est empêchée (par un *bite-block* par exemple), les locuteurs réussissent à modifier leurs gestes articulatoires pour réaliser de la parole intelligible. Des changements spectraux sont engendrés par ces modifications articulatoires. Pourtant, les caractéristiques principales qui servent à identifier les catégories phonémiques semblent être maintenues. Sussman et al. (1995) expliquent que si les équations de locus sont un bon indicateur du lieu d'articulation, leurs paramètres (la pente et l'ordonnée à l'origine) doivent rester relativement stables dans les situations de *bite-block*. Autrement dit, les locuteurs gardent la relation linéaire entre le début et le milieu de la transition du deuxième formant de manière à produire des segments intelligibles dans les situations anormales. Six locuteurs (3 femmes et 3 hommes) de l'anglais américain ont prononcé des séquences CVt (où C = /b, d, g/ et V = /i, I, ε, æ, Λ, a, ɔ, o, u/). Ces séquences ont été réalisées dans deux conditions : normale et avec un *bite-block* mis dans la bouche. La comparaison entre parole normale et parole avec un *bite-block* montre des valeurs significativement similaires entre les pentes et les ordonnées à l'origine (Table 3.14). Les pentes des trois catégories consonantiques sont significativement différentes dans les deux

C	/b/		/d/		/g/	
	normale	bite-block	normale	bite-block	normale	bite-block
condition	P/OO	P/OO	P/OO	P/OO	P/OO	P/OO
Moyenne	0.85/175	0.81/217	0.39/1364	0.38/1376	0.74/750	0.74/770
Écart-type	0.03/76	0.04/46	0.06/199	0.08/213	0.04/96	0.04/122

TABLE 3.14 – Valeurs moyennes et écart-types des pentes (P) et des ordonnées à l'origine (OO) des six locuteurs dans les deux conditions de production : normale et avec un *bite-block* (Sussman et al., 1995).

conditions, le constat est le même pour les ordonnées à l'origine. Des analyses de comparaison par paires sur les pentes ainsi que sur les ordonnées à l'origine suggèrent que toutes les combinaisons des catégories sont significativement différentes. Les valeurs de  $R^2$  sont élevées pour les deux conditions (0.9 pour la parole normale et 0.89 pour la parole avec un *bite-block*). Ces valeurs signifient que la relation linéaire entre  $F2_{onset}$  et  $F2_{mid}$  est constante même si les locuteurs réorganisent leurs gestes articulatoires en fonction du *bite-block*. D'après Sussman et al. (1995), les résultats de cette étude laisseraient entendre que l'objectif quasi-constant du système articulatoire est de conserver l'intégrité de la relation entre le début et le milieu de F2. La stabilité des équations de locus dans les deux conditions est la conséquence de la stabilité de ces deux points. Les différences de fréquences sont négligeables entre les  $F2_{onset}$  et les  $F2_{mid}$  en parole normale et en parole avec *bite-block*. Les trajectoires entre ces deux points ( $F2_{onset}$  et  $F2_{mid}$ ) peuvent varier compte tenu de la réorganisation des gestes articulatoires. Les auteurs précisent que ces trajectoires semblent porter peu d'informations sur le lieu d'articulation. De ce fait, les locuteurs cherchent à atteindre les deux extrémités de la transition ( $F2_{onset}$  et  $F2_{mid}$ ),

peu importe la trajectoire.

### 3.4.1.6 Dans un modèle articulatoire

Les équations de locus sont dérivées des phénomènes acoustiques qui sont à leur tour issus des gestes articulatoires. Un moyen d'examiner les variations et la sensibilité des équations de locus est de simuler des séquences CV avec un système artificiel contrôlable. Les changements des paramètres des équations de locus seront ainsi expliqués directement par les manipulations effectuées par l'expérimentateur. Chennoukh et al. (1997) emploient le modèle *the Distictive Region Model* (DRM) pour comprendre les origines de la linéarité et des différentes variations des équations de locus. Ils l'utilisent dans le but de vérifier si la relation linéaire entre  $F2_{onset}$  et  $F2_{mid}$  peut être maintenue. Des séquences VCV sont obtenues à l'aide du DRM qui fournit une représentation des zones de fonction dans le conduit vocal. Il s'agit d'un tube divisé en huit zones "regions" (Figure 3.13). Les limites de ces zones émanent du contraste acoustique des sons, ce qui maximise la gamme de variation des fréquences des formants. Chaque zone

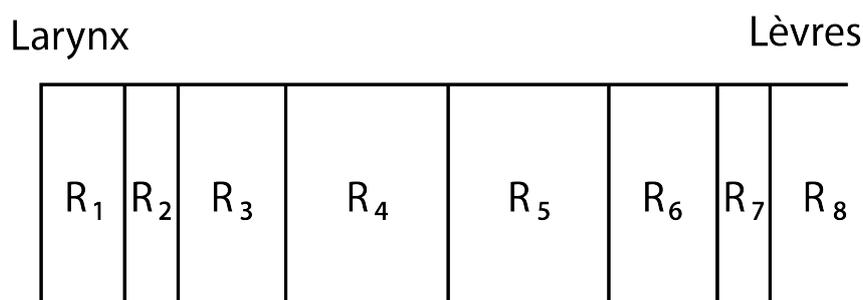


FIGURE 3.13 – Illustration du modèle DRM et de ces 8 zones ("**R**egions"). R<sub>1</sub> correspond à la cavité laryngale, R<sub>3</sub>, R<sub>4</sub>, R<sub>5</sub>, R<sub>6</sub> à la cavité orale, R<sub>7</sub> aux dents et R<sub>8</sub> aux lèvres. R<sub>2</sub> correspond à la transition entre les cavités laryngale et orale (d'après Chennoukh et al. (1997)).

cadre un lieu de constriction du conduit vocal d'un segment. Par exemple, pour la production des voyelles, les zones R<sub>3</sub> et R<sub>4</sub> sont dédiées aux constriction postérieures, celles des R<sub>5</sub> et R<sub>6</sub> aux constriction antérieures et celles des R<sub>4</sub> et R<sub>5</sub> aux constriction centrales. Des séquences V<sub>1</sub>CV<sub>2</sub> ont été simulées par le biais de la variation des zones de fonction du DRM dans le temps avec le principe de superposition du geste de la consonne dans les transitions voyelle-à-voyelle d'Öhman (1966). La consonne intervocalique C regroupe les consonnes /b, d, g/. Les voyelles V<sub>1</sub> et V<sub>2</sub> ont été choisies parmi /a, ø, o, u, y/. Trois phases d'articulation ont été générées pour chaque séquence dans le but de simuler trois degrés de coarticulation en changeant la temporalité inter-gestuelle (degré de coarticulation de la ph<sub>1</sub> = élevé, de la ph<sub>2</sub> = moyen et de la ph<sub>3</sub> = bas). Les trois zones R<sub>8</sub>, R<sub>6</sub> et R<sub>5</sub> ont été employées dans la simulation des trois consonnes /b, d, g/ respectivement. Toutes les combinaisons possibles de V<sub>1</sub> et V<sub>2</sub> ont été effectuées pour chaque occlusive réalisant ainsi 75 séquences dans les divers degrés de coarticulation (au total 225 séquences obtenues : 5 V<sub>1</sub> × 5 V<sub>2</sub> × 3 C × 3 degrés de coarticulation). Les séquences ont les mêmes durées : 100 ms pour V<sub>1</sub>, 50 ms pour la transition V<sub>1</sub>C, 50 ms pour la clôture de C,

40 ms pour la transition  $CV_2$  et 150 ms pour  $V_2$ . Les trajectoires des trois premiers formants ont été calculées pour chacune des séquences simulées. Les informations retenues de ces mesures ont servi à la synthèse des séquences. À partir de ces dernières, les valeurs de  $F2_{onset}$  et celles des  $F2_{mid}$  ont été prélevées respectivement à 10 ms et à 110 ms après le début des transitions  $CV_2$ . Les équations de locus ont été estimées pour chaque occlusive/zone dans chaque degré de coarticulation. Une relation linéaire forte a été remarquée pour les trois occlusives avec les degrés de coarticulation élevé et moyen ( $r > 0.84$ ). Quant aux occlusives en contexte coarticulatoire bas, la relation linéaire est faible ( $r < 0.50$ ). Les coefficients des équations de locus

	R <sub>8</sub> /b/		R <sub>6</sub> /d/		R <sub>5</sub> /g/	
	P	OO	P	OO	P	OO
ph <sub>1</sub>	0.85	94	0.51	1132	0.98	849
ph <sub>2</sub>	0.75	212	0.42	1362	0.86	500
ph <sub>3</sub>	0.33	770	0.22	1566	0.45	1198

TABLE 3.15 – Pentas (P) et ordonnées à l'origine (OO) des équations de locus des trois occlusives/zones issues du modèle DRM dans les trois phases de degré de coarticulation (Chennoukh et al., 1997).

sont modifiés en fonction du degré de coarticulation pour les trois occlusives. Les pentes sont proches de 1 quand le degré de coarticulation est maximal (ph<sub>1</sub>) et elles sont proches de 0 lorsqu'elles sont réalisées avec le degré de coarticulation minimal (ph<sub>3</sub>). Ces constatations mettent en avant la capacité des équations de locus à détecter les différents degrés de coarticulation dans divers contextes articulatoires (Krull, 1988). Dans chaque phase, les pentes des vélaires sont les plus élevées, celles des alvéolaires les plus basses et les bilabiales ont des pentes intermédiaires (Table 3.15). Les résultats en ph<sub>1</sub> et ph<sub>2</sub> montrent des valeurs similaires à ceux rapportés par Sussman et al. (1993) pour l'arabe et l'ourdou. Une autre série d'équations de locus a été calculée en fonction de  $V_1$  de manière à tester la stabilité des équations de locus dans les différents contextes vocaliques initiaux (5 équations de locus pour chaque occlusive dans chacun des degrés de coarticulation). Les résultats signalent des variations non-négligeables selon la voyelle initiale. Par exemple, les pentes de R<sub>8</sub> dans la ph<sub>1</sub> varient entre 0.7 ( $V_1 = /o/$ ) et 0.96 ( $V_1 = /u/$ ). D'après les auteurs, les équations de locus ne représentent pas un invariant relationnel direct du lieu d'articulation puisqu'elles sont sensibles au contexte vocalique initial ainsi qu'à la temporalité gestuelle entre la consonne et la voyelle. En revanche, ils expliquent qu'une corrélation constante qui résiste à ces variations apparaît lorsque des équations de locus de second ordre (SOLE)<sup>36</sup> sont estimées. Autrement dit, l'invariant relationnel surgit dans un niveau plus haut des équations de locus (SOLE). Dans ce niveau, les données de chaque catégorie consonantique s'ajustent parfaitement malgré les variations du degré coarticulatoire du

36. *Second Order Locus Equations* (SOLE) : il s'agit des équations calculées à partir des nuages de point issus des coefficients des équations de locus traditionnelles ou bien des équations de locus de premier ordre (FOLE), avec les pentes en abscisses et les ordonnées à l'origine en ordonnée.

contexte vocalique initial.

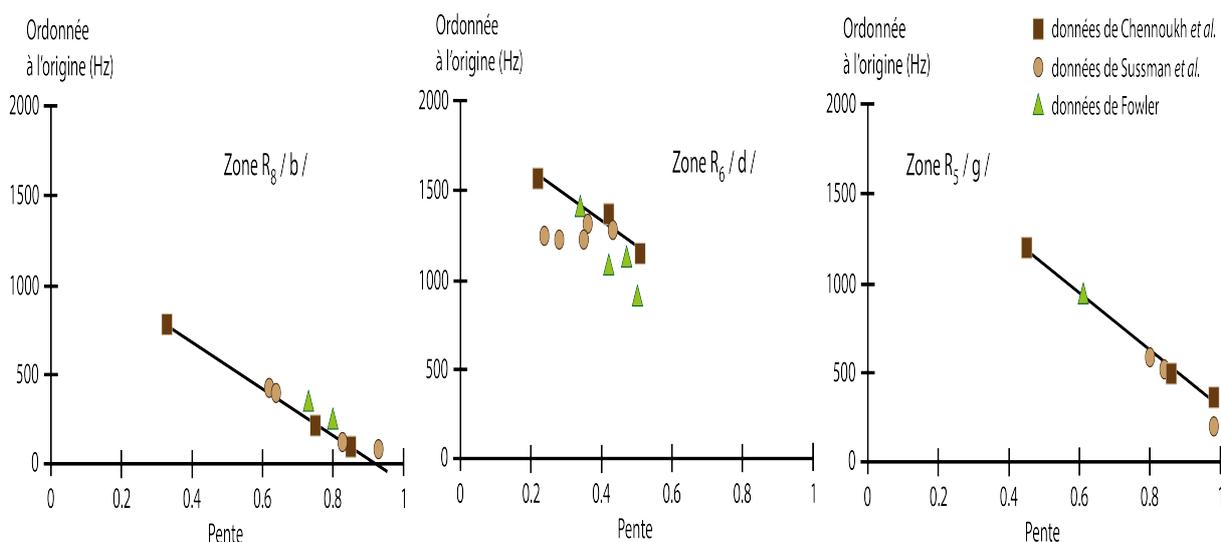


FIGURE 3.14 – Représentation schématique des paramètres des équations de locus de second ordre (SOLE) des trois degrés de coarticulation (représentés par des carrés) des trois zones R<sub>5</sub>, R<sub>6</sub>, R<sub>8</sub>, qui sont affichés avec les données de Fowler (1994) (représentées par des triangles) et celles de Sussman et al. (1991) (représentées par des cercles) (d'après Chenoukh et al. (1997)).

Les auteurs précisent que des droites de régression linéaires sont remarquées lors de la combinaison des SOLE de cette étude avec celles de la parole naturelle de l'anglais issues des travaux de Fowler (1994) et de Sussman (1994) (Figure 3.14). Une corrélation forte est également observée dans les données translinguistiques (thaï, arabe du Caire et ourdou) de Sussman et al. (1993). Selon les auteurs, chaque catégorie consonantique appartient à une droite de régression linéaire de second ordre nonobstant l'instabilité constatée dans les relations linéaires de premier ordre. Toutefois, Chenoukh et al. (1997) n'offrent aucune interprétation ou comparaison des valeurs de pente et d'ordonnée à l'origine des SOLE.

D'une manière similaire à celle de Chenoukh et al. (1997), Lindblom (1998) a montré comment les pentes des équations de locus capturent le degré de coarticulation anticipatoire à l'aide d'un modèle articulaire de conduit vocal (APEX). Dans un premier temps, une simulation de la configuration du corps de la langue a été réalisée de façon à ce que la langue soit déjà mise en position pour produire la voyelle suivante pendant la fermeture orale de la consonne avec des séquences /dV/. Dans un second temps, une simulation du corps de la langue a été effectuée dans le cas de figure où la langue est restée en position neutre durant la clôture de l'occlusive alvéolaire. La pente résultante de la première simulation est de 0.94 et celle de la seconde de 0.07. Le grand changement des valeurs de pente dans les deux situations souligne que la coarticulation anticipatoire a été maximisée dans le premier cas et minimisée dans le second. Les études de Chenoukh et al. (1997) et Lindblom (1998) révèlent que les modèles articulaires qui simulent les différentes conditions coarticulatoires reflètent systématiquement les valeurs attendues des équations de locus.

### 3.5 La position syllabique

La majorité des études sur les équations de locus a examiné les consonnes en position syllabique initiale. Une question se pose alors : est-ce que les paramètres des équations de locus d'une consonne changent en fonction de sa position syllabique ? Dans le travail original sur les équations de locus de Lindblom (1963b), les occlusives /b, d, g/ d'un locuteur suédois ont été analysées en position initiale CV et en position finale VC. Les valeurs de pentes des occlusives /b, d, g/ en position initiale sont différentes de celles en position finale (respectivement 0.69, 0.28, 0.95 et 0.57, 0.55, 1.02). Ces valeurs indiquent une coarticulation plus importante pour le /b/ initial comparé au /b/ final et une coarticulation moins marquée pour le /d/ et le /g/ en position initiale par rapport au /d/ et au /g/ en position finale. Lacerda (1985) a évalué les équations de locus en position initiale (CV) et finale (VC) pour des consonnes suédoises. Les résultats des équations de locus montrent que le degré de la coarticulation persévérante (VC) est plus important que celui de la coarticulation anticipatoire (CV). Krull (1988) a également confronté les paramètres des équations de locus dans les deux positions. Elle explique que les pentes des équations de locus des consonnes initiales affichent des valeurs distinctes en fonction du lieu d'articulation. Par contre, les pentes des consonnes en position finale sont similaires et proches de 1. Les occlusives en position finale attestent donc un degré de coarticulation plus important qu'en position initiale. Krull souligne que les catégories consonantiques en position finale peuvent être distinguées par les ordonnées à l'origine.

Sussman et al. (1997) examinent la capacité des équations de locus à distinguer phonétiquement les catégories du lieu d'articulation des occlusives en position syllabique initiale, centrale et finale. Dix locuteurs (5 femmes et 5 hommes) de l'anglais américain ont produit des séquences CVC et VCV (où C = /b, d, g/ et V = /i, ɪ, e, ε, æ, α, o, ʌ, ɔ, u/). Les locuteurs ont été incités à réaliser les CVC avec le relâchement de l'occlusive finale. Les séquences CVC ont été utilisées pour calculer les équations de locus pour les occlusives en position initiale CV et en position finale VC<sup>37</sup>. Les V<sub>1</sub>CV<sub>2</sub> ont été employées afin d'estimer les équations de locus pour les occlusives centrales. Les voyelles en position initiale correspondent à une des trois voyelles /i, α, u/. Les équations de locus ont été dérivées uniquement pour les occlusives avec les voyelles post-consonantiques CV<sub>2</sub>. Les voyelles en position initiale servent à tester l'influence des voyelles pré-consonantiques sur les équations de locus<sup>38</sup>. En position initiale (CVC) et centrale (VCV), les valeurs des F<sub>2<sub>onset</sub></sub> sont déterminées à la première pulsation de la voyelle suivante. Dans la position finale (CVC), les valeurs de F<sub>2<sub>offset</sub></sub> sont déterminées à la dernière pulsation de la voyelle précédente. Les valeurs de F<sub>2<sub>mid</sub></sub> ont été prélevées au milieu de la voyelle suivante (pour les CVC et VCV) ou précédente (pour les CVC). Les équations de locus ont été calculées pour les trois occlusives dans les trois positions. Deux équations de

37. Chaque séquence CVC a été répétée 3 fois soit un total de 30 mesures prononcées par locuteur par position syllabique (1 occlusive × 10 voyelles × 3 répétitions).

38. Chaque séquence VCV a été répétée 5 fois. Au total, 50 mesures par locuteur par voyelle initiale pour chaque occlusive ont été obtenues pour calculer les équations de locus (50 × /i/CV, 50 × /α/CV et 50 × /u/CV).

locus ont été estimées pour l'occlusive /g/ avec les voyelles antérieures et postérieures afin de prendre compte les deux allophones de ce phonème.

C		/b/						/d/					
Position	Initiale		Centrale		Finale		Initiale		Centrale		Finale		
	P	OO	P	OO	P	OO	P	OO	P	OO	P	OO	
Moyenne	0.77	272	0.77	270	0.50	428	0.37	1294	0.39	1269	0.32	1232	
Écart-type	0.08	95	0.05	82	0.16	183	0.07	106	0.08	148	0.09	141	

C		/g/ <i>palatal</i>						/g/ <i>vélair</i> e					
Position	Initiale		Centrale		Finale		Initiale		Centrale		Finale		
	P	OO	P	OO	P	OO	P	OO	P	OO	P	OO	
Moyenne	0.26	1711	0.26	1813	0.43	1456	1.12	199	1.24	26	0.76	394	
Écart-type	0.12	425	0.08	178	0.184	404	0.18	253	0.28	353	0.50	596	

TABLE 3.16 – Valeurs moyennes et écart-types des pentes (P) et des ordonnées à l'origine (OO) dans les trois positions syllabiques (initiale, centrale et finale) des /b/ et /d/ (en haut) et les /g/*palatal* et /g/*vélair*e (en bas) (Sussman et al., 1997).

Les résultats soulignent que les pentes des quatre catégories consonantiques (/b/, /d/, /g/*palatal* et /g/*vélair*e) sont significativement différentes (Table 3.16). Les analyses de comparaison par paires signalent des différences significatives pour toutes les combinaisons de pente (à l'exception d'une paire, /d/ vs. /g/*palatal*). Les pentes en position initiale et centrale sont significativement similaires. Celles en position finale sont statistiquement distinctes des deux autres positions à l'exception de l'alvéolaire /d/ où les trois positions ont des pentes similaires. Les analyses sur les ordonnées à l'origine présentent un effet principal du lieu d'articulation. Les comparaisons par paires dévoilent des différences significatives de toutes les combinaisons des ordonnées à l'origine (à l'exception de la paire /b/ vs. /g/*vélair*e). Les ordonnées à l'origine des occlusives en position initiale et centrale sont statistiquement similaires tandis que celles en position finale sont significativement différentes des autres positions. L'alvéolaire /d/ quant à elle n'est statistiquement pas différente dans les trois positions syllabiques. Les valeurs de  $R^2$  en position initiale et centrale sont statistiquement similaires (moyennes de  $R^2$  en initiale = 0.72 et en centrale = 0.73). Toutefois, elles sont significativement distinctes de celles en position finale (moyenne de  $R^2$  en finale = 0.51). Ces valeurs de  $R^2$  indiquent que les pentes des occlusives en position finale sont moins linéaires que dans les deux autres positions. L'influence des contextes consonantique ( $C_2$  sur  $C_1V$  ou  $C1$  sur  $VC_2$  dans les  $C1VC_2$ ) et vocalique ( $V_1$  dans  $V_1CV_2$ ) a également été examinée. Les analyses statistiques montrent une absence d'effet du contexte consonantique en position initiale ou finale sur la pente. L'impact du contexte vocalique initial sur les pentes est faible, il est significatif dans trois contextes seulement (/i/bV (0.79) vs. /u/bV (0.72), /i/dV (0.36) vs. /u/dV (0.42) et /i/g<sub>pa</sub>V (0.22) vs. /æ/g<sub>pa</sub>V (0.28)).

Les variations des valeurs de pente en position initiale et centrale vs. position finale sont provoquées par le type d'interaction articulatoire entre la consonne et la voyelle. Sussman et al.

(1997) soulignent que l'interaction anticipatoire de la voyelle vers la consonne (V-à-C<sub>1</sub>) est une séquence articulatoire planifiée qui demande et doit être réalisée avec une précision articulatoire plus importante comparée à la coarticulation VC<sub>2</sub> qui est conçue comme un effet mécanique contextuel. Le rôle de la transition du F2 dans le codage du lieu d'articulation dans les séquences VC est clairement diminué par rapport aux séquences CV. Ils ajoutent que la rupture acoustique de l'ordre des équations de locus des occlusives en position finale comparée à la position initiale/centrale peut être interprétée en termes de déclinaison articulatoire. Les auteurs expliquent que les équations de locus révèlent que la coarticulation CV possède un processus articulatoire contrôlé d'une façon "élégante" qui contribue à encoder et à contraster le lieu d'articulation des occlusives. Ils résument que « [...] CV and VC entities appear to be phonetically and hence perhaps, representationally, fundamentally dis-similar. Locus equation data strongly suggest that the CV F2 transition is more precisely controlled than the VC transition to bring about a lawful and emergentlike property—the vowel-normalization of the variable F2 transitions encoding stop place contrasts »<sup>39</sup>.

Modarresi et al. (2004a) examinent également l'influence de la position syllabique sur les équations de locus. Ils testent la stabilité de la pente des occlusives lors la coarticulation anticipatoire /CV/, persévérante /VC/ et anticipatoire trans-syllabique /C#V/. Cinq locuteurs de l'anglais américain et deux du persan ont produit des non-mots avec des syllabes ouvertes /CV<sub>1</sub>.CV<sub>2</sub>/ et avec des syllabes fermées /tV<sub>1</sub>C#V<sub>2</sub>t/ où C = /b, d, g/ et V = /i, æ, u, ɔ/ pour l'anglais et /i, ε, æ, u, o, a/ pour le persan. Dans les syllabes ouvertes, les séquences CV<sub>2</sub> ont été étudiées alors que dans les syllabes fermées, les séquences V<sub>1</sub>C et C#V<sub>2</sub> ont été analysées. Les locuteurs ont été informés qu'ils devaient réaliser le relâchement de la consonne finale (VC, C#V). Les valeurs de F2<sub>onset</sub> pour l'occlusive en position initiale (CV) ont été déterminées à la première pulsation de la voyelle suivante tandis qu'en position finale (VC) et en position trans-syllabique (C#V), elles ont été mesurées à la première résonance de F2 observable après le relâchement (F2<sub>@burst</sub>). Selon les auteurs, cette position est la plus proche de la consonne et donc la plus représentative en positions finale et trans-syllabique. Les valeurs de F2<sub>mid</sub> sont toujours fixées au milieu de la voyelle cible. Les résultats des deux langues ont été combinés puisqu'ils sont significativement similaires. Les pentes de chaque occlusive diffèrent en fonction de la position syllabique. Celles des occlusives en position VC ou C#V sont toujours plus basses avec des valeurs de R<sup>2</sup> réduites qu'en position CV. De plus, l'erreur standard est plus importante dans VC et C#V que dans CV. Les analyses statistiques entre les occlusives de chaque position dévoilent que les pentes des /b/ vs. /d/, /b/ vs. /g/ et /d/ vs. /g/ sont significativement différentes en position CV ainsi qu'en position C#V. De même, les pentes des occlusives en position VC sont statistiquement différentes, à l'exception de /b/ vs. /d/. La distance eu-

39. « [...] les entités CV et VC semblent être phonétiquement et par conséquent peut-être, représentativement, fondamentalement différentes. Les données des équations de locus suggèrent fortement que le contrôle de la transition du F2 de CV est plus précis que celui de VC. Ce postulat entraîne une propriété réglemентаire et émergente qui est la normalisation-vocalique de la variable des transitions du F2. Cette dernière encode les contrastes des lieux d'articulation des occlusives », Sussman 1997, page 2837, [notre traduction].

	P	OO	R <sup>2</sup>	SE
bV	0.84 (0.04)	183 (68)	0.95 (0.01)	92 (11)
Vb	0.16 (0.3)	1054 (93)	0.20 (0.20)	180 (49)
b#V	0.29 (0.10)	917 (185)	0.48 (0.20)	138 (13)
dV	0.43 (0.07)	1075 (132)	0.75 (0.07)	107 (14)
Vd	0.17 (0.05)	1445 (93)	0.33 (0.21)	115 (45)
d#V	0.14 (0.07)	1488 (146)	0.32 (0.18)	115 (32)
gV	0.93 (0.11)	267 (215)	0.90 (0.03)	150 (19)
Vg	0.39 (0.17)	1196 (313)	0.29 (0.21)	321 (94)
g#V	0.47 (0.16)	1084 (254)	0.47 (0.20)	269 (72)

TABLE 3.17 – Valeurs moyennes (et écart-types) des pentes (P), des ordonnées à l’origine (OO), des R<sup>2</sup> et de l’erreur standard (SE) dans les trois positions syllabiques pour tous les locuteurs (Modarresi et al., 2004a).

clidienne a été estimée entre les trois occlusives dans l’espace pente-ordonnée à l’origine dans chaque position pour les sept locuteurs. L’objectif de cette mesure est d’examiner la séparabilité des trois catégories consonantiques au sein de chaque position syllabique. Comme attendu, les occlusives en position CV sont les plus dispersées suivies par celles en position C#V. Au final, les occlusives en position VC sont les moins séparées. D’après Modarresi et al. (2004a), lorsque l’occlusive est en position finale, les équations de locus ne peuvent pas capturer la coarticulation persévérante (VC) et par conséquent les lieux d’articulation des occlusives. Modarresi et al. (2004b) utilisent une autre technique<sup>40</sup> que les équations de locus pour mesurer le degré de coarticulation avec les mêmes données. Ils observent que le chevauchement coarticulatoire est plus important en position finale qu’en position initiale (CV vs. VC). Toutefois, selon Modarresi et al. (2004a), les équations de locus ont classifié les occlusives en position trans-syllabique (C#V) d’une manière similaire à celle remarquée en position initiale (CV). Ils ajoutent qu’il existe deux facteurs principaux influençant la coarticulation des occlusives en position C#V : la voyelle initiale (V<sub>1</sub>C#V<sub>2</sub>) et l’intervalle de pause dû à la séparation syllabique entre la consonne et la voyelle. Les valeurs de pentes changent en fonction de la voyelle précédente. Elles sont négativement corrélées avec l’intervalle de pause (+ de pause, – de pente).

L’influence de la voyelle précédente a été notée sur les pentes des consonnes suivantes dans l’étude de Sussman et al. (1997) et celle de Modarresi et al. (2004a). Lindblom et al. (2007) proposent une équation de locus modifiée afin de capter le degré de coarticulation de la voyelle précédent et celle suivant la consonne dans des séquences V<sub>1</sub>CV<sub>2</sub>. Cette équation a la forme suivante :

$$F2_{onset}(V_2) = a + b * F2_{mid}(V_2) + c * F2_{mid}(V_1) \quad (3.3)$$

40. Modarresi et al. (2004b) mesurent le degré de coarticulation en calculant la différence de fréquence de  $F_{onset}/F_{offset}$  de chaque CV/VC dans deux contextes vocaliques (iCV – ɔCV et VCi – VCɔ). Une grande différence de fréquence souligne un maximum de coarticulation.

Cette formule indique que dans des séquences  $V_1CV_2$ , le  $F2_{onset}$  de la seconde voyelle est une combinaison linéaire pondérée des valeurs de  $F2_{mid}$  de  $V_1$  et de  $V_2$  (Lindblom et al., 2010). Comme dans l'équation 3.1, "a" et "b" sont les coefficients représentant la pente et l'ordonnée à l'origine et "c" est la pente soulignant le degré de coarticulation de  $V_1$ . Cette équation a été dérivée dans le but de quantifier la coarticulation anticipatoire et celle persévérante des voyelles sur la consonne centrale. Elle permet ainsi de dissocier l'impact de la voyelle précédente sur les fréquences de  $F2_{onset}$  des autres effets comme l'accent ou le débit de parole (Lindblom et al., 2007; Agwuele et al., 2008; Lindblom et al., 2009; Lindblom et Sussman, 2012).

Les études présentées dans les deux dernières sections ont exposé la capacité des équations de locus à déchiffrer les lieux d'articulation à l'aide d'outils mathématiques et statistiques. Deux questions peuvent alors être posées : quel est le lien entre les équations de locus et la perception des consonnes ? Est-ce que les êtres humains perçoivent les occlusives de la même manière que ce que proposent les équations de locus ? Dans la section suivante, le lien des équations de locus et sa pertinence à la perception sera discuté .

### 3.6 La relation perceptuelle des équations de locus

Les travaux innovateurs de Lacerda (1985) et de Shammass (1985) sont considérés comme les premières études qui ont testées la pertinence perceptuelle des paramètres des équations de locus. Ces études soulignent que les auditeurs arrivent à identifier une catégorie consonantique lorsque les valeurs de  $F2_{onset}$  et  $F2_{mid}$  synthétiques correspondent à la régression linéaire des équations de locus de cette catégorie. Les zones d'intersection entre les pentes des classes consonantiques correspondent aux phénomènes les plus confondus par les auditeurs (*cf.* sections 3.2 et 3.4). En effet, avant l'apparition des équations de locus (Lindblom, 1963b), dans une étude de perception, Liberman et al. (1954) ont manipulé les fréquences de  $F2_{onset}$  de 480 à 720 Hz avec les F2 de sept voyelles en présence de deux formants synthétiques. Il concluent que « *the results of this experiment show that the direction and degree of second-formant transitions can serve as cues for the aurally perceived distinctions among the stop consonants* »<sup>41</sup>. Fruchter et Sussman (1997) reprennent les données de l'étude précédente dans l'objectif de les relier avec les paramètres des équations de locus. Fruchter et Sussman (1997) transforment les fréquences des sept voyelles ( $F2_{mid}$ ), celles des onze débuts de transition par voyelle ( $F2_{onset}$ ) et les valeurs d'identification en coordonnées (x, y, z) de manière à les comparer avec les données acoustiques des équations de locus des cinq locuteurs de l'étude de Sussman et al. (1991). La Figure 3.15 affiche les résultats obtenus. Les données des identifications sont exposées dans une échelle de gris ; un taux élevé d'identification de la consonne est marqué par les zones les plus foncées. Les données acoustiques à comparer sont désignées par des cercles blancs. Les auteurs

41. « les résultats de cette expérience montrent que la direction et le degré des transitions du second formant peuvent servir de traits pour les distinctions perçues oralement dans les consonnes occlusives », Liberman et al. 1954, page 9, [notre traduction].

signalent que la correspondance entre les résultats de perception et les données acoustiques sont "assez raisonnables". Les identifications de /b/ et de /d/ sont en accord avec les données acoustiques et elles les superposent. Quant au /g/, les résultats sont plus faibles puisque les zones de dominances perceptuelles (zones foncées) sont moins couvertes de données acoustiques.

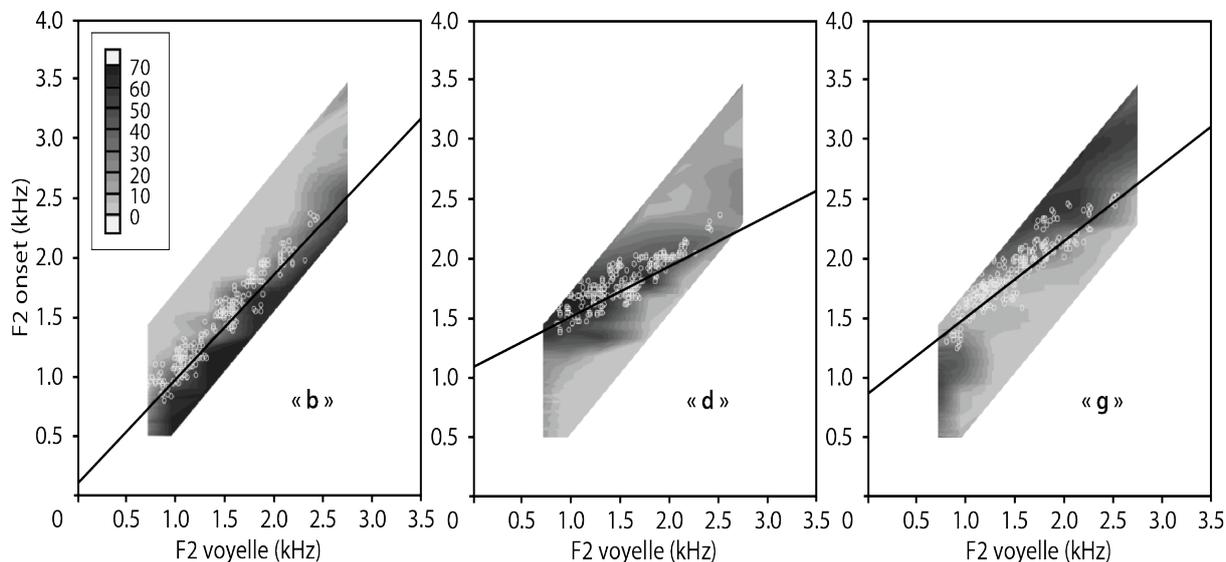


FIGURE 3.15 – Surfaces d'identification de /b, d, g/ avec 7 contextes vocaliques. Les données de perception ont été empruntées à Liberman et al. (1954) et exposées dans l'espace des coordonnées des équations de locus. Les zones les plus foncées signifient un taux d'identification élevée. Les données acoustiques (cercles blancs) proviennent de l'étude de Sussman et al. (1991) (d'après Fruchter et Sussman (1997)).

Eek et Meister (1995) examinent également la pertinence perceptuelle des équations de locus en estonien. Un locuteur masculin de l'estonien a produit des séquences de type CVV où C = /p, t, k/ et V correspond aux voyelles longues /i, a, u, e, o, ä, ü, ö, ɔ/. Les fréquences des pics les plus élevés dans le bruit de relâchement de chaque séquence ont été prélevées ainsi que les informations formantiques. Dans une première expérience, 27 séquences CVV originales avec et sans le bruit de relâchement ont été présentées quatre fois dans un ordre aléatoire à 13 auditeurs. Leurs tâches était d'identifier la consonne au début des séquences. Les syllabes avec le bruit de relâchement ont été identifiées correctement à 100%. Dans les séquences sans le bruit de relâchement, la bilabiale /p/ a été perçue comme un /p/ à 100%. L'alvéolaire initiale /t/ a été entendue comme un /t/ uniquement avec les voyelles /ɔ/ et /a/. Elle a été perçue comme un /p/ à 50% avec les voyelles /o/ et /u/ et comme un /k/ devant la voyelle /ü/. Dans les autres cas, les consonnes ont été jugées comme des voyelles longues, aucune consonne n'a été identifiée (Figure 3.16). Les auteurs justifient ces résultats par l'absence de transition (dite aussi "transitions négligeables") du deuxième formant de ces séquences. Concernant l'occlusive vélaire, elle a été entendue comme un /p/ devant les voyelles antérieures. Les auteurs mettent en évidence que l'omission du bruit de relâchement a réduit considérablement la perception des occlusives. Les mêmes CVV sans bruit de relâchement de la première expérience ont été utilisées

comme des séquences de base pour réaliser des syllabes CVV synthétiques. Les transitions du F2 peuvent être montantes ou descendantes en fonction des valeurs de  $F2_{onset}$  variant entre des valeurs maximales (F3) et minimales (F1). Les transitions montantes ont été identifiées comme un /p/, celle descendantes ont été entendues comme un /t/ devant les voyelles postérieures et comme un /k/ devant les voyelles antérieures.

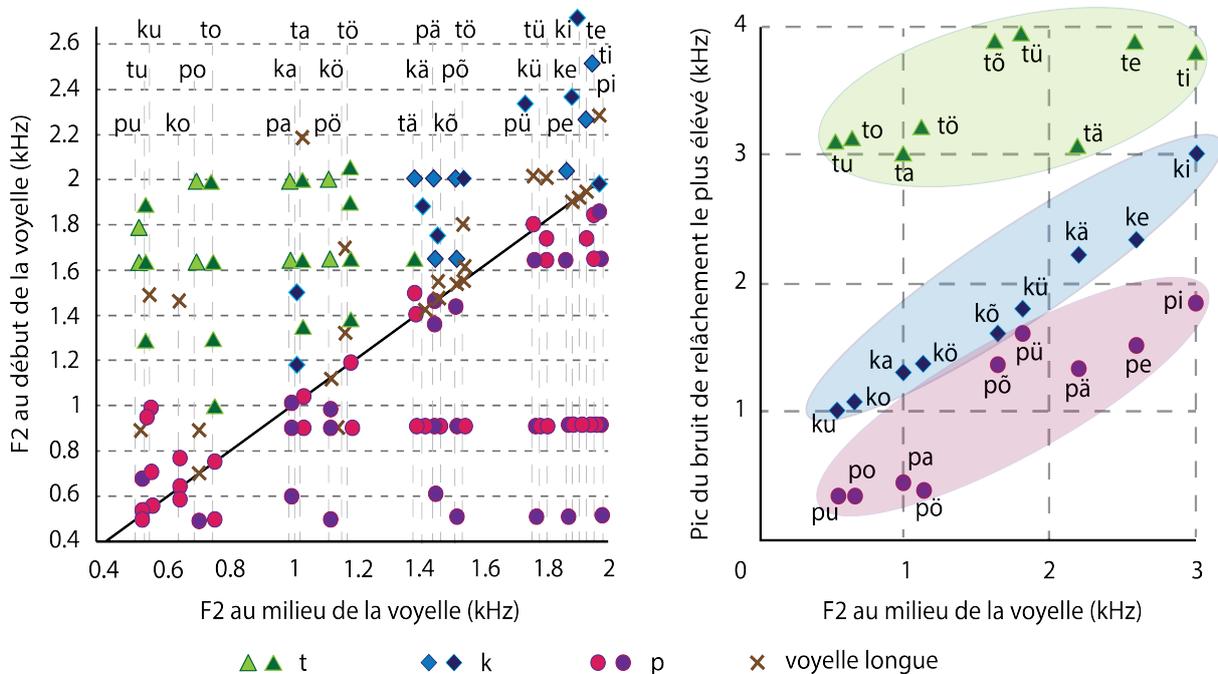


FIGURE 3.16 – À gauche, représentation de la perception des occlusives dans des syllabes CVV sans le bruit de relâchement. Chaque stimulus a été identifié comme un /p/, un /t/, un /k/ ou comme une voyelle longue. La ligne diagonale expose les cas où le début de la transition du F2 ( $F2_{onset}$ ) est égal à son milieu ( $F2_{mid}$ ). À droite, illustration de l'espace de perception des occlusives avec la fréquence du pic le plus élevé en fonction de  $F2_{mid}$  (Eek et Meister, 1995).

Les auteurs soulignent que « *F2 transitions did not provide the primary context-independent cues for place of articulation. Locus equations showed relational invariance for stop categorization in the production space but they had not the same relevant role in perception* »<sup>42</sup>. Une troisième expérience a été conçue afin d'examiner l'importance du bruit de relâchement. Des séquences sans transition formantique ont d'abord été générées, puis des séquences avec 50 ms de transitions de F2 en direction de la fréquence du pic le plus élevé du bruit de relâchement ont été produites. Les deux types de séquences ont été précédés par le bruit de relâchement original. Le taux de l'identification est situé entre 75 et 100 % dans tous les cas. Il n'existe pas de différence de réponse entre les séquences avec et sans transition du F2. Eek et Meister (1995) concluent que la connexion entre le pic le plus élevé du bruit de relâchement et l'état stable de la voyelle suivante constitue un trait fiable de la catégorisation des occlusives dans l'espace

42. « les transitions de F2 n'ont pas fourni de traits sur le lieu d'articulation indépendamment du contexte. Les équations de locus ont montré une invariance relationnelle mais elles n'ont pas le même rôle pertinent dans la perception », Eek et Meister 1995, page 20, [notre traduction].

de perception. Les auditeurs combinent ces deux informations plutôt que les coefficients des équations de locus ( $F2_{onset}$  et  $F2_{mid}$ ).

Dans leur étude de perception, Fruchter et Sussman (1997) étudient la pertinence perceptuelle des équations de locus. Des séquences CV ont été synthétisées sans le bruit de relâchement avec dix voyelles (/ɔ, o, a, ʌ, u, æ, ε, ɪ, e, i/) basées sur les données de Sussman et al. (1991). Plusieurs variables ont été maintenues constantes : la durée des séquences à 300 ms, la durée des transitions de F2 et de F3 à 48 ms et la durée de VOT à 6 ms. L'état stable de chaque voyelle a été combiné à onze valeurs de  $F2_{onset}$ . Les fréquences de  $F3_{onset}$  ont été variées avec trois valeurs afin d'obtenir des transitions qui ressemblent à celles de /b, d, g/. Au total, 330 stimuli ont été recueillis ( $11 F2_{onset} \times 10$  voyelles  $\times 3 F3_{onset}$ ). Six sujets (3 féminins et 3 masculins) ont participé à l'expérience de perception. Ils ont écouté les stimuli huit fois pendant quatre jours pour éviter l'effet de la fatigue (2 répétitions de stimuli par jour,  $2 \times 330$ ). La tâche des sujets était d'identifier chaque stimulus qui ressemble le plus à un "b", un "d", un "g", un "w" ou à "pas de consonne". Les deux derniers choix ont été proposés pour ne pas forcer les sujets à donner obligatoirement une identification. Les résultats ont été présentés comme des surfaces d'identification superposées par les données acoustiques des équations de locus de l'étude de Sussman et al. (1991).

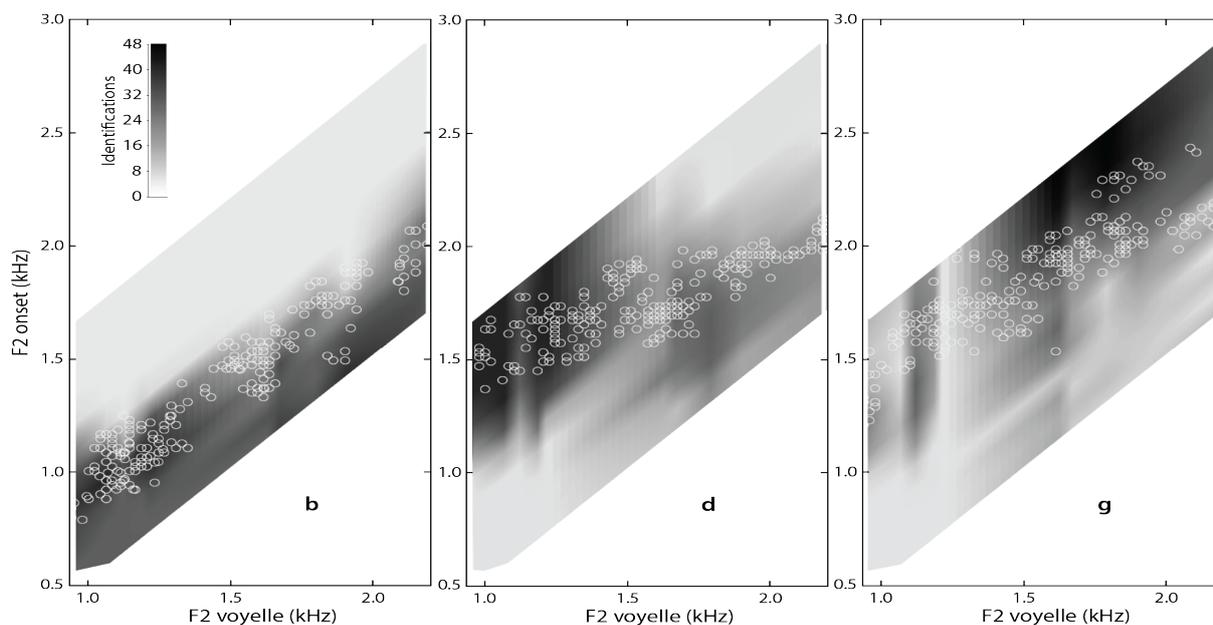


FIGURE 3.17 – Surfaces d'identification de /b, d, g/ avec les 10 contextes vocaliques. Les réponses des sujets ont été exposées dans l'espace des coordonnées des équations de locus. Les zones les plus foncées signifient un taux d'identification élevé. Les données acoustiques (cercles blancs) proviennent de l'étude de Sussman et al. (1991) (d'après Fruchter et Sussman (1997)).

Les auteurs indiquent que par le biais ce genre d'affichage, les correspondances des réponses perceptuelles avec les données acoustiques des équations de locus peuvent être jugées au mieux. Au moyen d'une représentation tri-dimensionnelle ( $x, y, z$ ),  $F2_{onset}$  et  $F2_{mid}$  sont placés sur les axes des abscisses et des ordonnées ( $x, y$ ). Quant à la troisième dimension ( $z$ ), elle correspond au

degré de gris dont le foncé signifie un taux d'identification élevé (Figure 3.17). Un seul niveau de F3 est utilisé dans chaque représentation correspondant à chaque catégorie d'occlusive.

Les résultats de cette expérience ressemblent à ceux de Liberman et al. (1954) affichés dans la Figure 3.15. Ils montrent que les réponses de "b" et de "d" s'accordent aux données acoustiques, les zones les plus foncées sont les plus couvertes avec les cercles blancs. La correspondance entre les réponses de "g" et les données acoustiques dans la zone des voyelles antérieures est relativement bonne. Toutefois, elle est problématique dans la zone des voyelles postérieures où la dispersion des cercles blancs sont à l'extérieur des zones foncées. Les auteurs expliquent que ces résultats de /g/ peuvent être dus à plusieurs raisons dont l'absence du bruit de relâchement, la durée courte de VOT et les transitions du F2 et F3 - qui sont plus courtes que dans le /g/ naturel. L'effet de F3 est élevé dans les zones d'intersection des droites de régression des équations de locus. Dans ces zones de chevauchement, le taux d'identification augmente d'une

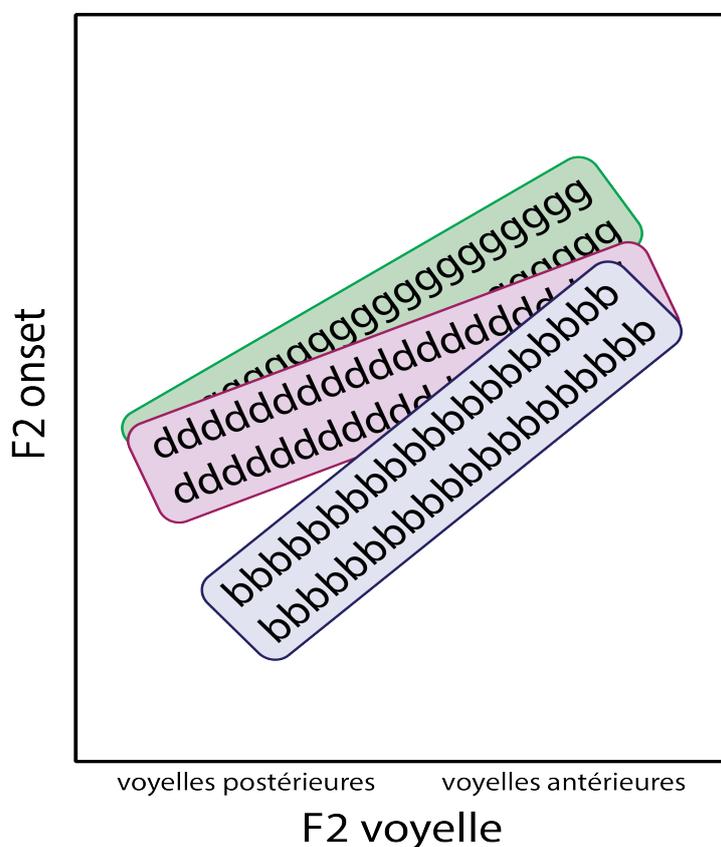


FIGURE 3.18 – Représentation schématique de l'hypothèse de la hiérarchie de dominance perceptuelle montrant les /b/ > /d/ avec les voyelles antérieures et /d/ > /g/ avec les voyelles postérieures. Le /b/ avec les voyelles postérieures et le /g/ avec les voyelles antérieures représentent deux zones de non-compétition (Fruchter et Sussman, 1997).

catégorie consonantique quand le F3 ressemble le plus à cette catégorie. Fruchter et Sussman (1997) proposent une hypothèse de dominance hiérarchique afin de conceptualiser les relations acoustique et perceptuelle des séquences sans bruit de relâchement de cette expérience et de celle de Eek et Meister (1995). Les auteurs offrent une illustration schématique de cette hypo-

thèse avec les trois occlusives dans trois cadres (Figure 3.18). Chaque cadre représente un nuage de points d'une occlusive dans l'espace acoustique des équations de locus. Il s'agit d'une interprétation abstraite de la combinaison entre les régressions linéaires et l'effet de la dominance perceptuelle des zones de chevauchement. Les occlusives /d/ et /g/ se chevauchent dans l'espace des voyelles postérieures et les occlusives /b/ et /d/ se croisent dans la zone des voyelles antérieures. Les auteurs soulignent que ces zones sont les plus confondues par les auditeurs. Ils indiquent que ces zones de perception peuvent être interprétées en termes de dominance hiérarchique, c'est-à-dire  $b > d$  et  $d > g$ . Les locuteurs ont tendance à identifier un /b/ quand le stimulus est situé dans la zone de chevauchement de /b/ et de /d/ dans le contexte des voyelles antérieures. De même, une perception de /d/ prévaut lorsque le stimulus tombe dans la zone de chevauchement entre /d/ et /g/ avec les voyelles postérieures. Les consonnes /b/ et /g/ ne se chevauchent pas, la relation de dominance est donc non pertinente. Cette dominance est remarquée dans l'étude de Eek et Meister (1995) ainsi que dans celle de Liberman et al. (1954). Les auteurs résument que « *it seems that one key to understanding the patterns of both burst and F3 effects on stop place perception is to view them in relation to the pattern of locus equation data. The way in which burst perceptual weight and F3 perceptual weight both correlate with the overlap versus distinctiveness of locus equation acoustic data argues strongly for a locus equation interpretation of F2 transitional cues* »<sup>43</sup>.

Les études menées par Sussman et ses collègues sur les équations de locus ont permis la suggestion d'une hypothèse perceptuelle. L'objectif de celle-ci est d'interpréter les résultats de ces études d'un point de vue perceptif. Elle est présentée dans la section suivante.

### 3.6.1 *The orderly output constraint*

L'approche de l'invariant relationnel est impliquée dans la construction d'un modèle de perception des consonnes basé sur les équations de locus. L'aspect systématique observé dans les équations de locus n'est pas explicitement contenu dans une seule séquence. C'est pourquoi un type de système de représentation est exigé pour que la séquence puisse être associée à d'autres (non présentes) afin de détecter les propriétés pertinentes (Brancazio et Fowler, 1998). D'après Sussman et al. (1998b), le système de la perception de la parole est créé au niveau neuronal à partir d'un plan bidimensionnel dans un espace acoustique bivarié dans lequel les fonctions linéaires représentent des catégories. Dans cet espace, les  $F2_{onset}$  et les  $F2_{mid}$  sont les paramètres acoustiques qui établissent ces fonctions linéaires. Dans une étude antérieure, Sussman a proposé l'existence de tableaux liés à des neurones de combinaison sensitive qui réagissent à des paires de  $F2_{onset}$  et  $F2_{mid}$  extraites du signal en entrée ((Sussman, 1989), explorée dans

43. « il semble qu'une des clés pour comprendre les effets du bruit de relâchement et du F3 sur la perception du lieu d'articulation est de les considérer en relation avec les données des équations de locus. La manière dont le poids perceptuel du bruit de relâchement et celui du F3 corrélient avec le chevauchement *versus* la discriminabilité des données acoustiques des équations de locus argumente fortement pour l'interprétation des équations de locus des traits transitionnels du F2 », Fruchter et Sussman 1977, page 3007, [notre traduction].

la section 3.4). Ces neurones de combinaison sensitive spécifiques à des syllabes CV particulières sont rangées dans une colonne qui correspond à une consonne (Figure 3.8). Dans le tableau, les paires de  $F2_{onset}$  et  $F2_{mid}$  correspondent à la zone dans l'espace de F2 capturée par les équations de locus d'une consonne donnée. La sortie des neurones, quant à elle, active des neurones d'un niveau plus élevé qui représente la consonne. Sussman et al. (1998b) cherchent à constituer une conceptualisation de la catégorisation des consonnes et de sa représentation neurobiologique dans le cerveau. Les auteurs émettent l'hypothèse que les mécanismes neuro-naux du traitement de la parole humaine peuvent exploiter les stratégies du traitement auditif conservées dans l'évolution et trouvées chez les non-humains. Ils argumentent alors que le traitement du signal acoustique chez les êtres humains peut être comparé à celui des autres espèces animales. Ils justifient ce raisonnement par les arguments suivants : 1) le cerveau humain est un produit de l'évolution avec une conception et une structure généralement conservées, 2) les grandes similarités de structure et de fonction des substrats neuraux existent dans les exigences de traitement de stimulus dans les différentes espèces, 3) l'évolution a tendance à produire des solutions similaires à des problèmes similaires. Les études neuroéthologiques révèlent une stratégie suivie par les différentes espèces lors du traitement du signal. Le principe de la combinaison sensitive de deux paramètres acoustiques conduit à une information d'un ordre élevé, faisant émerger une propriété de portée biologique à l'organisme (Figure 3.19). Ces paramètres

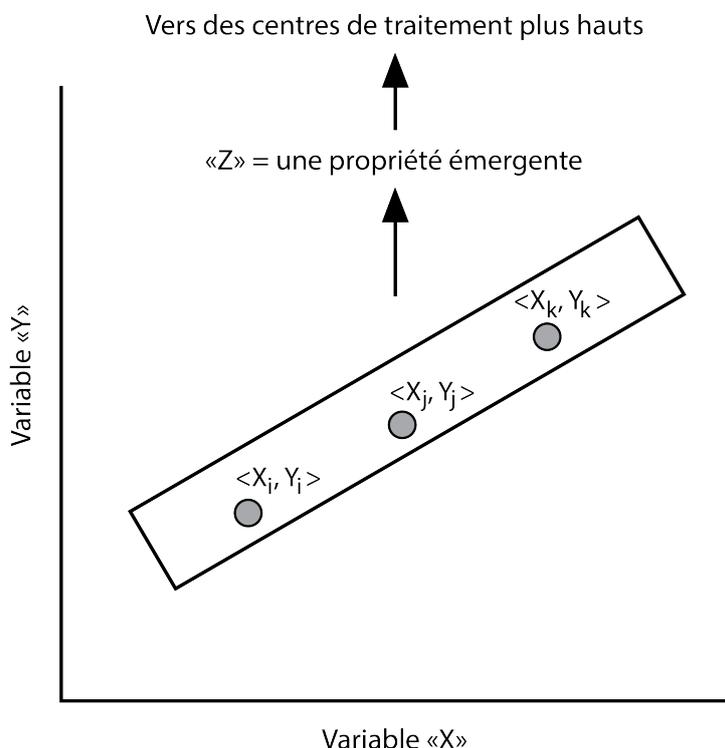


FIGURE 3.19 – Plan schématique de deux stimuli indépendants combinés dans des paires ordonnées ( $\langle x_i, y_i \rangle, \langle x_j, y_j \rangle, \dots$ ) produisant une propriété émergente et dérivée ( $Z$ ), qui représente une catégorie ou une classe donnée (d'après Sussman et al. (1998b)).

établissent une relation linéaire inhérente forte. Les chercheurs argumentent que cette linéarité

de données remarquée chez les animaux est comparable à celle rencontrée dans les équations de locus. D'après les auteurs, la relation linéaire entre les deux composantes clés du signal de la séquence CV est une adaptation coévolutionnaire du système de la production de la parole des êtres humains à une stratégie évolutionnaire conservée du traitement auditif. L'organisation du système de la perception de la parole, qui forme un plan auditif dans l'espace de  $F2_{onset}$  et  $F2_{mid}$ , est facilitée par un signal ordonné. Par conséquent, les locuteurs se sont adaptés pour co-articuler de façon à produire la relation linéaire capturée par les équations de locus. Autrement dit, les locuteurs sont contraints à réaliser des valeurs  $F2_{onset}$  et  $F2_{mid}$  linéairement ordonnées dans des séquences CV afin de permettre un système de représentation efficace. Cette idée est interprétée comme une sortie contrainte ordonnée ou "*orderly output constraint*" (OOC). Selon l'OOC, la relation linéaire entre les éléments acoustiques critiques d'un signal complexe améliore le traitement et la représentation éventuelle de ces entrées par une extraction catégorielle de traits dans un tableau bidimensionnel de neurones auditifs de combinaison sensitive. Cette hypothèse a été inspirée par les mécanismes présents dans les systèmes auditifs des chauves-souris et des chouettes.

Une variété de neurones de combinaison sensitive a été observée chez les chauves-souris de *pteronotus parnellii*. Certains de ces neurones répondent aux composants de la pulsation de l'écholocalisation et de son écho. La pulsation et son écho sont composés de quatre harmoniques (de 30 à 120 KHz) dont chacune a une fréquence constante (CF) et une fréquence modulée (FM). L'écho est retardé temporellement et la fréquence est décalée (l'effet Doppler<sup>44</sup>) au regard de la pulsation. En effet, les neurones CF/CF encodent la vitesse d'une cible grâce à la détection de l'effet Doppler entre les appariements des différents CF des composantes harmoniques de la pulsation émise et de son écho (Figure 3.20). Les neurones de combinaison sensitive détectent la relation linéaire créée entre les fréquences de CF1/CF2 et CF1/CF3 (CF1 de la pulsation et CF2 et CF3 de l'écho). Ces combinaisons constituent des plans de vitesse présentés systématiquement dans un espace neuronal CF/CF. Les chauves-souris se servent de ces plans pour déterminer la vitesse de leurs cibles. Le traitement de l'écholocalisation utilise des paires de composantes du signal qui sont linéairement liées.

Les neurones de combinaison sensitive ont également été repérés chez les chouettes effraie. De la même façon, le système neuronal des chouettes effraie combinent deux composantes du signal dans un domaine bidimensionnel (x, y) afin de localiser la source du son. La localisation est déterminée grâce à la différence temporelle inter-orale (ITD). Des études ont montré que l'ITD est obtenue à partir de la relation linéaire entre la fréquence du signal et la phase de cette fréquence (Figure 3.20). Les auteurs suggèrent l'existence de plans à deux dimensions dans un espace acoustique bivarié dans lequel les fonctions linéaires représentent des catégories. En comparaison avec la détection de la vitesse chez les chauves-souris et les fonctions de

44. L'effet Doppler ou le décalage Doppler : « Il s'agit de l'augmentation de la fréquence d'un son lorsque sa source, par exemple un train (qui se met à siffler) approche un récepteur (nos oreilles), et la diminution de la fréquence lorsque la source s'éloigne », Thornton et Rex 2009, page 53.

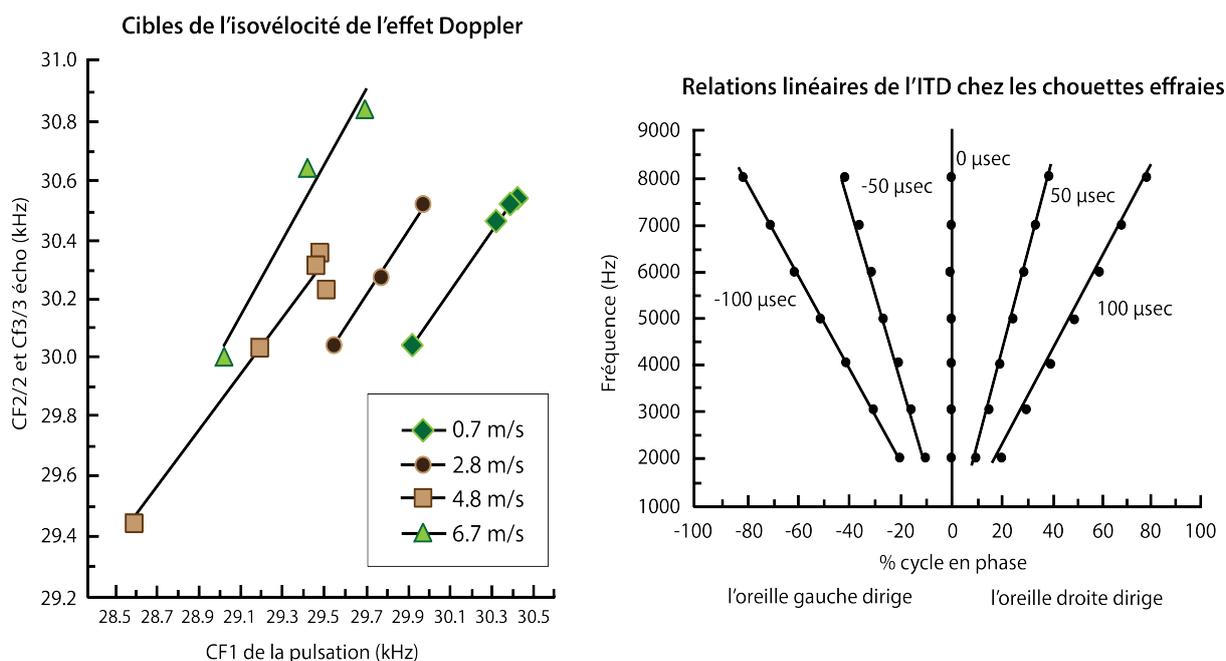


FIGURE 3.20 – À gauche, exemples de quatre fonctions d'isovélocité des chauves-souris de *pteronotus parnellii* obtenues par l'affichage des CF1 (KHz) de la pulsation en abscisse en fonction des deuxième (CF2) et troisième (CF3) harmoniques de l'effet Doppler et de l'écho en ordonnée (données empruntées à Suga et al. (1983)). À droite, exemples des cinq catégories linéaires de l'ITD des chouettes effraie obtenues par l'affichage des différences de phases en abscisse contre la fréquence en ordonnée (données empruntées à Wagner et al. (1987)), (d'après Sussman et al. (1998b)).

l'ITD chez les chouettes effraie, les équations de locus peuvent être conceptualisées comme des fonctions de détection des lieux d'articulation.

Sussman et al. (1998b) expliquent que les équations de locus représentent une variance réglementaire dans la manifestation acoustique de la coarticulation qui est stable dans la catégorisation du lieu d'articulation des occlusives. Le degré de coarticulation, indexé par les équations de locus, devient alors un paramètre de catégorisation. Les auteurs soulignent qu'un plan auditif employant les informations de  $F2_{onset}$  et  $F2_{mid}$  aide à dériver les catégories du lieu d'articulation des occlusives. Ils indiquent que plusieurs arguments soutiennent cette idée :

1. Les équations de locus sont un phénomène linguistique universel dans la parole des adultes et elles sont considérées comme un indice important du contraste phonémique (Sussman et al., 1993; Celdran et Villalba, 1995).
2. Quand la forme typique des équations de locus est préservée face à la perturbation articulaire avec un *bite-block*, une équivalence perceptuelle est également conservée. Toutefois, lorsque cette forme typique est largement influencée (linéarité faible, variation importante de pente) comme dans le cas de l'apraxie développementale de la parole<sup>45</sup>,

45. L'apraxie développementale de la parole (DAS de l'anglais "*Developmental Apraxia Speech*") est un trouble congénital qui concerne l'habilité à programmer les mouvements de la parole en l'absence d'une pathologie neuromusculaire (Sussman et al., 1998b).

l'intelligibilité diminue (Sussman et al., 1998b).

3. Il existe un lien entre la valeur du trait du lieu d'articulation de l'occlusive et le degré de relation de la transition entre le début et le milieu du F2. Cette transition est réputée dans la littérature pour encoder des traits importants du lieu d'articulation des occlusives.
4. Des expériences ont été effectuées sur les réseaux de neurones dans des séquences CV dans le but de classifier les consonnes par lieu d'articulation. Ces expériences ont révélées que les parties les plus informatives du signal pour la catégorisation des consonnes sont les fréquences des  $F2_{onset}$  et  $F2_{mid}$  (Hinton et Lang, 1988; McDermott et Katagiri, 1988).
5. Des substrats neuronaux dédiés aux tâches de traitement par combinaison des fréquences modulées à deux moments dans le signal ont déjà été démontrés chez les animaux (Fitzpatrick et al., 1993).

Les auteurs exposent que tous ces arguments fournissent la preuve que le début et la fin de la transition du F2 ( $F2_{onset}$  et  $F2_{mid}$ ), en combinaison, sont des indices majeurs pour décrire le lieu d'articulation des occlusives. Ces composantes du signal de la parole semblent être connectées ensembles et sont interprétées comme une caractéristique du lieu d'articulation pendant la perception de la parole. Les auteurs admettent l'importance d'autres indices qui participent à la perception des occlusives tels que le troisième formant, le bruit de relâchement et le VOT. Selon les auteurs, ces informations se combinent avec les équations de locus dans un niveau plus élevé que celui des équations de locus pour former la perception d'une occlusive. Les équations de locus ne sont donc pas le spécifieur unique du lieu d'articulation mais elles constituent un indice important pour sa perception.

L'article source où est développée l'hypothèse de l'OOC est suivi d'une série de critiques, globalement négatives, émises par un ensemble de spécialistes<sup>46</sup> sous forme de commentaires publiés dans le même volume. Les critiques se concentrent majoritairement sur le problème de chevauchement, l'inadéquation de l'analogie neuroethologique, la linéarité et la stabilité des équations de locus :

- Plusieurs critiques ont concerné la question de l'absence d'une séparabilité complète des catégories consonantiques lorsque les équations de locus sont utilisées. Par comparaison, il existe au contraire une séparation parfaite de l'ITD chez les chouettes et des fonctions de vélocité chez les chauves-souris. Sussman et al. (1998b) expliquent que le chevauchement est le résultat du degré de liberté élevé des articulateurs, de l'assimilation des phonèmes à cause de la coarticulation et de la variabilité des stratégies adaptées par les locuteurs lors de la production de la parole.
- Le manque de correspondance du signal entre les animaux et les êtres humains a été également évoqué. Autrement dit, les compositions du signal de l'entrée sont dissimilaires entre la chouette, la chauve-souris et l'être humain. Sussman et ses collègues soulignent

---

46. Au total, 39 linguistes, phonéticiens, psycholinguistes, psychologues, neurologues, neurobiologistes, biologistes, physiologistes, anthropologues, zoologues et informaticiens ont critiqué l'hypothèse de l'OOC dans des commentaires qui sont ensuite complétés par les réponses des auteurs.

que ce n'est pas l'exactitude de l'adéquation de tous les détails de l'analogie entre les systèmes des humains et des non-humains qui est examinée puisque cette tâche est irréaliste et impertinente pour leur cas d'étude. En effet, leur préoccupation principale porte sur les points communs calculables qui peuvent être identifiés entre les espèces.

- Une question a été posée par plusieurs chercheurs sur la nécessité de la relation linéaire pour la perception : est-ce que la linéarité est un prérequis pour établir les catégories susceptibles d'être distinguées au niveau perceptuel ? Sussman et ses collègues répondent qu'en termes de nécessité neurobiologique, il est probablement prudent de conclure que la linéarité n'est pas un prérequis pour élaborer les représentations neuronales des signaux de l'entrée des différentes classes. Toutefois, parmi les indices de la perception, les régularités statistiques sont extrêmement utiles à la compréhension des phénomènes perceptuels pour des raisons d'apprentissage ("*learnability*"). Ils ajoutent que les relations linéaires sont les marqueurs les plus efficaces sur le plan mathématique.
- La stabilité des équations de locus a également été mise en cause dans les critiques. Les équations de locus sont souvent affectées par des facteurs phonétiques divers comme le contexte vocalique précédent, l'accent tonique, le débit de parole et d'autres influences coarticulatoires. Sussman et ses collègues signalent que les variations des paramètres des équations de locus provoquées par ces facteurs sont minimales et n'affectent pas la séparabilité relative des catégories consonantiques.

Sussman défend théoriquement l'idée de neurone de combinaison sensitive dans plusieurs recherches (Sussman, 2002, 2013). Sussman (2013) explique que ce phénomène a été observé chez les chouettes, les chauves-souris et chez d'autres animaux comme les singes et les chat ; toutefois, aucune technique actuelle ne permet de tester ce phénomène sur les êtres humains. Le débat sur la pertinence des équations de locus dans la perception du lieu d'articulation a été poursuivi, notamment en ce qui concerne la capacité des équations de locus à résister à certains phénomènes phonétiques comme les changements de style et de débit de parole.

### 3.6.2 L'influence du degré de coarticulation sur la catégorisation

Il a été constaté depuis les premières études sur les équations de locus qu'elles reflètent le chevauchement coarticulatoire provoqué par les changements des différentes conditions de parole (style, débit, accent, etc.). Pour une consonne donnée, une augmentation du chevauchement articulaire est exprimée par une élévation de la valeur de pente alors qu'une diminution du chevauchement articulaire est désigné par un abaissement de la valeur de pente de la même consonne (Krull, 1989; Duez, 1992). Puisque les paramètres des équations de locus d'une consonne peuvent varier en fonction de ces conditions, est-ce qu'une bonne séparabilité entre les différentes catégories consonantiques est conservée, comme Sussman et al. (1998b) l'ont indiquée ? Une série d'études a examiné l'influence des changements de condition de parole sur la catégorisation des équations de locus et son implication sur la perception.

Crowther (1994) expose que si les auditeurs arrivent à percevoir des occlusives dans des conditions coarticulaires extrêmes, les équations de locus doivent marquer des paramètres distincts permettant la séparation entre les lieux d'articulation. Une locutrice de l'anglais a produit des syllabes de type CV dans trois conditions coarticulaires (où C = /b, d, g/ et V = /i, ɪ, eɪ, ɛ, æ, o, ʌ, ɔ, u/). Les séquences ont été réalisées avec un EPG pour avoir des comparaisons articulatoires avec les données acoustiques. Chaque séquence a été répétée 5 fois. Dans la condition avec un degré de coarticulation élevé (CoÉ), la réalisation de la voyelle cible (V) est maintenue puis interrompue par la consonne cible (C) afin de garantir une coarticulation maximale. La séquence devient  $V_1CV_2$  où  $V_1=V_2$ . Dans la condition avec un degré de coarticulation moyen (CoM), les séquences CV ont été réalisées dans une phrase porteuse. Cette condition est similaire à celle utilisée dans les études antérieures. Quant à la condition avec un degré de coarticulation bas (CoB), le but est de minimiser l'influence de la voyelle qui précède la consonne. Ceci a été accompli en produisant les C et les V cibles dans la syllabe /íC<sup>j</sup>V/, avec la C palatalisée et l'accent sur la voyelle initiale. Cette manipulation aide à bloquer l'influence coarticulaire de la voyelle cible. Selon l'auteur, la production du /i/ accentué avant la consonne cible et la palatalisation de la consonne cible offrent un bon contrôle sur la position linguale pour C. Par conséquent, le contact linguo-palatal de la consonne cible doit être similaire dans la réalisation de toutes les voyelles cibles. Crowther (1994) met en place trois moyens de mesures de manière à vérifier que les trois conditions conduisent bien à trois degrés de coarticulation différents. La première mesure indexant le degré de coarticulation est la dispersion de  $F2_{onset}$  dans les différents contextes vocaliques obtenue par l'écart-type. Une dispersion élevée signifie une dépendance de la consonne sur la voyelle suivante et une dispersion basse montre que la consonne n'est pas influencée par la voyelle adjacente. Les valeurs de l'écart-type des trois occlusives augmentent dans l'ordre CoB < CoM < CoÉ. Le deuxième moyen de mesure est basé sur le calcul de la différence absolue entre  $F2_{onset}$  et  $F2_{mid}$  ( $|F2_{onset} - F2_{mid}|$ ). Le degré de coarticulation est maximal lorsque la différence est basse et minimal quand elle est élevée. Les valeurs de la différence absolue pour les trois occlusives baissent dans l'ordre CoB > CoM > CoÉ. Le troisième moyen de mesure est effectué à l'aide des données de l'EPG. Ce dernier a été utilisé afin de mesurer le contact linguo-palatal dans les moments où  $F2_{onset}$  et  $F2_{mid}$  ont été mesurés. Un palais artificiel avec 96 électrodes a été conçu pour la locutrice. Quand la langue touche une électrode du palais artificiel avec une pression qui dépasse un certain seuil, l'électrode est activée. Le degré de coarticulation est indiqué par le degré de correspondance entre les électrodes activées au moment où  $F2_{onset}$  est mesuré et celles au moment où  $F2_{mid}$  est déterminé. Lorsque une électrode est activée au moment de  $F2_{onset}$  mais qu'elle ne l'est pas au moment de  $F2_{mid}$ , la correspondance est dite faible. Une correspondance faible entre le contact au moment de la mesure de  $F2_{onset}$  et celui de  $F2_{mid}$  est interprété comme une coarticulation minimale. Les résultats dévoilent que les données articulatoires sont en accord avec les données acoustiques. En effet, les trois moyens de mesure ont mis en valeur l'efficacité de la manipulation dans les trois conditions et confirment l'obtention de trois degrés de coarticu-

lation distincts. Les équations de locus ont été calculées pour les trois occlusives dans les trois conditions. Comme attendu, pour les trois occlusives, la pente augmente quand la coarticulation croît. Les pentes en condition CoB ont des valeurs proches de zéro pour les trois occlusives. En condition CoM, la vélaire a la pente la plus élevée (0.82), l'alvéolaire la pente la plus basse (0.26) et la bilabiale une pente intermédiaire (0.77). Dans la condition CoÉ, les valeurs de pente sont plus élevées que dans les deux autres conditions. Les pentes de /b/ et /g/ sont presque identiques avec une valeur de 1 pour /b/ et 0.98 pour /g/. Quant au /d/, il garde une valeur de pente relativement basse (0.48). Les valeurs des ordonnées à l'origine varient dans le sens inverse. Elles sont plus élevées dans la condition CoB et basses dans celle de CoÉ. Une représentation bidimensionnelle avec les pentes en abscisse et les ordonnées à l'origine en ordonnée montre la capacité des équations de locus à discriminer les lieux d'articulation (Figure 3.21). Sur cette représentation, les catégories consonantiques connaissent un chevauchement impor-

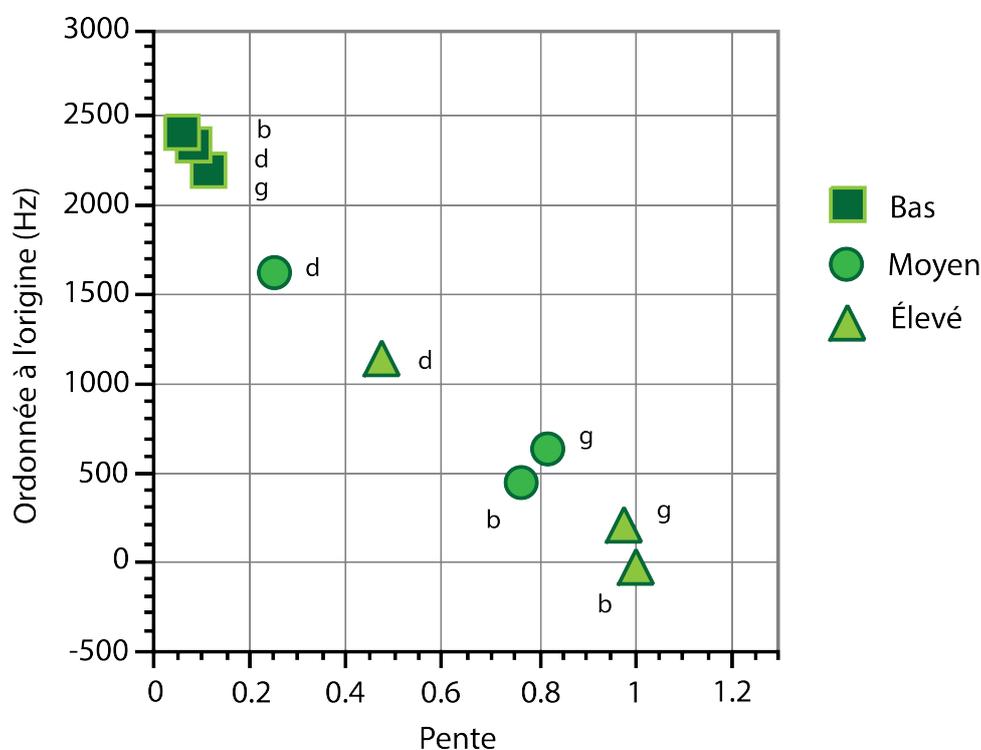


FIGURE 3.21 – Représentation de l'espace pente-ordonnée à l'origine des trois occlusives dans les trois conditions (d'après Crowther (1994)).

tant empêchant la création de classes distinctes. Les trois occlusives produites dans la condition CoB se regroupent dans la même zone révélant une discrimination faible des équations de locus du lieu d'articulation dans cette condition.

Une expérience de perception a été mise en place dans le but de déterminer la pertinence perceptuelle des équations de locus, notamment la relation entre la similarité dans l'espace physique (pente et ordonnée à l'origine) et celle dans l'espace perceptuel. Une prédiction raisonnable peut être faite sur le lien entre les espaces physique et perceptuel : les consonnes qui ont des points proches dans l'espace pente-ordonnée à l'origine peuvent être proches dans

l'espace perceptuel et par conséquent, elles seraient confondues dans des tâches de perception. Cinq auditeurs ont écouté deux répétitions des séquences /iCjV/ en condition CoB des trois occlusives avec les dix voyelles. Au total, soixante séquences ont été identifiées par auditeur. Les résultats affichent un niveau de performance élevé des cinq sujets : deux ont identifié 60 séquences sur 60, un sujet 59 séquences sur 60 et les deux derniers 58 sur 60. Contrairement aux attentes, les auditeurs ont parfaitement identifié les séquences qui se confondent dans l'espace acoustique. Crowther (1994) souligne que deux interprétations possibles peuvent expliquer la relation inattendue entre l'espace physique des équations de locus et celui perceptuel : 1) malgré l'ambiguïté des informations dans les  $F2_{onset}$  et  $F2_{mid}$ , les auditeurs utilisent davantage l'information du bruit de relâchement dans la catégorisation des lieux d'articulation, 2) il se peut également que les auditeurs encodent les informations des transitions formantiques d'une autre manière que celle proposée par les équations de locus.

D'un autre côté, Bakran et Mildner (1995) examine la stabilité des paramètres des équations de locus avec l'alternance "naturelle" du débit de parole d'une seule occlusive (/d/). Quatre locuteurs du croate ont produit des mots de longueur différente (entre 1 et 9 syllabes) contenant l'alvéolaire /d/ dans cinq contextes vocaliques. L'occlusive cible est réalisée dans des syllabes accentuées (débit lent) et non-accentuées (débit rapide). Les différences de débit se manifestent dans la durée de fermeture de la consonne. La durée de la clôture consonantique est maximale en débit lent et minimale en débit rapide, elle varie entre 58 et 131 ms. En débit rapide, les valeurs de pente sont plus élevées qu'en débit lent. La moyenne des différences des pentes des deux débits est de 0.158. L'équation de locus de second ordre (SOLE) avec les pentes et les ordonnées à l'origine des deux débits des quatre locuteurs atteste une relation linéaire élevée. Cette constatation indique que la stabilité des équations de locus apparaît dans un ordre plus élevée (SOLE) que celles-ci (Chennoukh et al., 1997).

Brancazio et Fowler (1998) conçoivent deux protocoles expérimentaux afin d'étudier l'influence de la variation coarticulatoire sur la catégorisation des occlusives et sur leur perception. L'objectif de la première expérience est de faire varier le chevauchement coarticulatoire par une manipulation artificielle de l'articulation. Trois locuteurs de l'anglais américain ont produit des séquences VCV ou CV dans trois conditions de parole (où C = /b, d, g/ et  $V_{cible} = /i, ɪ, e, æ, a, ʌ, ɔ, o/$ ). Dans la première condition, les auteurs ont essayé d'augmenter le chevauchement coarticulatoire (et par conséquent la pente) en demandant aux participants d'accélérer - autant que possible - le geste de la voyelle après le relâchement de la consonne et de prononcer les séquences  $V_1CV_2$  avec un débit rapide en suivant la technique de Crowther (1994) où  $V_1 = V_2$ . Dans la deuxième condition, les chercheurs ont visé à diminuer le chevauchement coarticulatoire en précisant aux locuteurs de réaliser les séquences VCV avec un /ʌ/ initial de façon à maintenir une configuration centrale de la langue avant le relâchement de la consonne et de les prononcer en débit lent. Dans la troisième condition, les locuteurs ont produit des séquences CV avec un débit normal. Les équations de locus ont été calculées dans les trois conditions de chevauchement. Les résultats dévoilent une augmentation de pente et une diminution d'ordonnée

à l'origine quand le chevauchement coarticulatoire croît (Table 3.18). Les analyses statistiques

	élevé			normal			bas		
	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
/b/	0.90	11	0.98	0.82	145	0.98	0.76	219	0.98
/d/	0.61	804	0.94	0.48	1016	0.94	0.46	1074	0.93
/g/	0.99	169	0.95	0.88	340	0.90	0.87	388	0.91

TABLE 3.18 – Valeurs moyennes des pentes et des ordonnées à l'origine des trois locuteurs dans les trois conditions de chevauchement (Brancazio et Fowler, 1998).

sur les pentes et également sur les ordonnées à l'origine révèlent une différence significative des trois catégories dans chaque condition. Des analyses de contraste sur le degré de chevauchement soulignent que les pentes des occlusives en débit rapide sont significativement plus élevées que celles des autres débits. Toutefois, les pentes avec un débit lent ne sont pas significativement moins élevées que celles en débit normal. Les analyses de contraste sur les ordonnées à l'origine montrent que /b/ et /d/ sont statistiquement différents mais pas /g/ et /b/. Les tests sur le degré de chevauchement signalent une différence significative entre les occlusives en débit rapide et en débit normal mais pas entre le débit normal et celui lent. Brancazio et Fowler (1998) expliquent que les changements de pente et d'ordonnée à l'origine en fonction de la condition coarticulatoire sont peu importants. Ils ajoutent qu'il n'existe pas de chevauchement de pente et d'ordonnée à l'origine entre l'alvéolaire /d/ et les autres occlusives dans les trois conditions. De plus, la manipulation du degré de coarticulatoire n'a pas influencé la linéarité des régressions, les R<sup>2</sup> restent quasiment stables (Table 3.18). L'examen de l'influence des trois conditions coarticulatoires dans l'espace pente-ordonnée à l'origine met en évidence que les trois occlusives de tous les locuteurs tombent dans des zones non-chevauchées nonobstant les variations observées dans les pentes et les ordonnées à l'origine (Figure 3.22). Les auteurs exposent que même si les pentes et les ordonnées à l'origine n'ont pas réussi individuellement à distinguer les occlusives /b/ et /g/, elles le peuvent ensemble. Les points de chaque occlusive sont bien ajustés dans une régression linéaire lors du calcul des équations de locus de second ordre (SOLE). Les coefficients des trois SOLE sont distinctes, avec une linéarité élevée pour le /b/ (0.97) et le /g/ (0.98) et plus basse pour le /d/ (0.80). D'après les auteurs, les résultats de cette expérience, en adéquation avec ceux d'autres études (Krull, 1989; Crowther, 1994; Bakran et Mildner, 1995), mettent en avant que malgré la discrimination des lieux d'articulation des occlusives par la pente et l'ordonnée à l'origine, les équations de locus ne constituent pas un invariant. Ils précisent que l'élaboration des SOLE illustre un phénomène fiable qui sert également à distinguer les lieux d'articulation avec une linéarité et une stabilité importantes. La variabilité de pente et d'ordonnée à l'origine trouvée dans l'espace des équations de locus est absente dans l'espace des SOLE. Ce raisonnement va à l'encontre du modèle de l'OOC qui prédit la perception d'une occlusive à l'aide de  $F2_{onset}$  et  $F2_{mid}$  par un plan linéaire de cette occlusive.

Brancazio et Fowler (1998) mettent en exergue que les SOLE représentent un exemple d'une

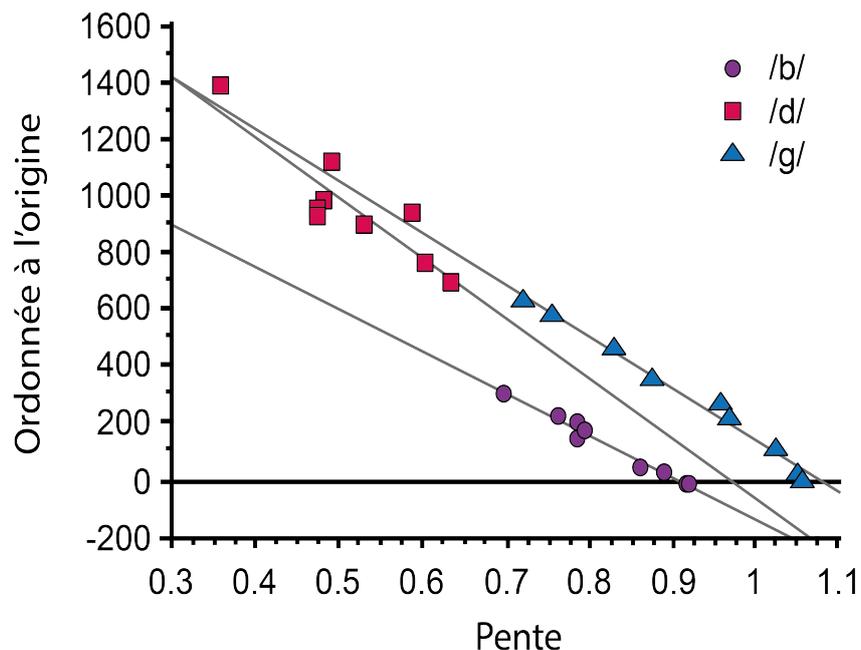


FIGURE 3.22 – Illustration de l'espace pente-ordonnée à l'origine des trois occlusives dans trois conditions de chevauchement des locuteurs (d'après Brancazio et Fowler (1998)).

linéarité qui est le résultat de plusieurs fonctions linéaires et qui ne peut donc pas être utilisée comme fonction linéaire. Par conséquent, elles ne peuvent pas refléter l'existence d'une contrainte de production qui a une base perceptuelle. Ils ajoutent qu'il existe, en effet, des variations systématiques de pente et d'ordonnée à l'origine en fonction du lieu d'articulation. Pourtant, les équations de locus ne sont pas le résultat des contraintes de production dérivé par les demandes du système perceptuel comme le modèle de l'OOC le propose. Les auteurs suggèrent une interprétation alternative autre que celle de l'OOC concernant la linéarité et des changements systématiques des équations de locus. Ils expliquent que la pente des équations de locus indexe le degré de la résistance coarticulatoire. La résistance est maximale quand la consonne et la voyelle sont en conflit sur un articulateur, notamment le corps de la langue (Recasens, 1985). Une relation entre  $F2_{onset}$  et  $F2_{mid}$  "proche de linéaire" signifie qu'une variation dans  $F2_{mid}$  est associée à un degré de changements dans  $F2_{onset}$  qui reste le même tout au long de la fonction. Les auteurs soulignent que leur hypothèse implique que la résistance coarticulatoire est généralement la même entre une consonne et l'ensemble des voyelles, puisqu'elles utilisent toutes le corps de la langue. Le degré du chevauchement doit ainsi être probablement identique pour une consonne donnée. Selon cette hypothèse (également citée par Fowler (1994)), les équations de locus indiquent directement le degré de résistance coarticulatoire et indirectement le lieu d'articulation car chaque consonne a un degré de résistance coarticulatoire spécifique.

La deuxième expérience effectuée par Brancazio et Fowler (1998) a pour objectif de tester la pertinence perceptuelle des équations de locus. Les auteurs comparent trois classifications entre elles : la classification des auditeurs dans deux conditions coarticulatoires (normale et élevée), celle des équations de locus et celle d'un autre modèle non linéaire ("*exemplar model*"). Ce

dernier est construit avec les mêmes informations employées pour l'estimation des équations de locus (les  $F2_{onset}$  et  $F2_{mid}$ ). Le but de la comparaison avec ce modèle est de vérifier la nécessité de la linéarité des équations dans la perception, comme le modèle de l'OOC le suggère. Les séquences en condition coarticulaire normale et élevée d'un locuteur (M2) ont été écoutées par trente-deux sujets pour une tâche de perception. Le bruit de relâchement de toutes les séquences a été enlevé ainsi que la voyelle initiale en condition élevée pour n'obtenir que des séquences CV sans bruit de relâchement dans les deux conditions. Il a été demandé aux auditeurs d'identifier la consonne initiale de chaque séquence parmi /b, d, g/. D'un autre côté, la distance euclidienne a été utilisée pour la classification des occlusives pour les équations de locus et pour l'*exemplar model*. Pour ce faire, les données CV ( $F2_{onset}$  et  $F2_{mid}$ ) du locuteur M2 et celles des cinq locuteurs de l'étude de Fowler (1994) ont été employées (un total de 720 stimuli). Pour les équations de locus, la distance euclidienne est mesurée entre chaque point des séquences CV et le point le plus proche sur la droite de régression de chaque consonne. Pour l'*exemplar model*, les distances ont été calculées entre chaque séquence. Ces distances ont été transformées pour que les distances les plus petites aient un effet proportionnellement plus important que les distances longues. En effet, les auteurs estiment les distances euclidiennes dans quatre modèles : les équations de locus, les équations de locus avec deux régressions de  $/g/^{47}$ , l'*exemplar model* transformé à la puissance 5 et l'*exemplar model* transformé à la puissance 10<sup>48</sup>. Les résultats de l'identification des auditeurs montrent un taux total de classification correcte de 66%. Les analyses statistiques révèlent une absence d'effet principal du chevauchement coarticulaire par sujet ou par item. Les classifications par modèle signalent un taux de classification correcte légèrement plus élevé que celui des auditeurs. Le taux de classification correcte des équations de locus est identique à celui de l'*exemplar model 5* (71%), pourtant leurs erreurs ne se sont pas produites avec les mêmes séquences. Les modèles des équations de locus avec deux régressions de  $/g/$  et l'*exemplar model 10*, quant à eux, ont classifié correctement les occlusives avec un taux identique de 77%. La précision de ces modèles a été comparée avec les identifications des auditeurs afin d'examiner la force de la prédiction des quatre modèles. Le modèle *exemplar 10* a la meilleure performance (62%) suivi par les équations de locus avec deux régressions (59%), l'*exemplar 5* (58%) et les équations de locus (56%). Une grande quantité des erreurs n'a pas été prédite par les modèles, même si ces derniers ont prédit entre 56% et 62% des réponses des sujets (au-dessus du hasard de 33%). Brancazio et Fowler (1998) exposent que l'utilisation des  $F2_{onset}$  et  $F2_{mid}$  comme paramètres de classification peut repérer un aspect systématique des réponses des sujets, mais elle échoue à expliquer une grande proportion des données. Selon les auteurs, les résultats dévoilent que le succès dans la classification des occlusives ne doit pas être compris comme une preuve en faveur de leur pertinence perceptuelle parce que les erreurs

47. Les auteurs précisent que deux fonctions linéaires de  $/g/$ , en fonction de la voyelle antérieure ou postérieure, conduisent à deux régressions plus ajustées et de ce fait, à une meilleure classification.

48. Le modèle avec un exposant plus élevé (10) mènerait à une classification plus importante que celle avec un exposant plus petit (5) parce que l'exposant le plus élevé atténue au maximum les contributions des séquences ayant des grandes distances dans la classification.

des modèles coïncident rarement avec celles des sujets. De plus, la majorité des séquences mal identifiées par les sujets a été correctement prédite par les modèles. Les auteurs concluent que les résultats des deux expériences soutiennent la possibilité de l'existence de zones spécifiques pour les différentes consonnes dans l'espace du F2 qui sont importantes pour leur perception. En revanche, ces résultats ne démontrent pas une supériorité du modèle des équations de locus sur un modèle comparable qui n'utilise pas de représentation linéaire. Par conséquent, leurs observations n'appuient pas l'argument que les équations de locus aient une pertinence perceptuelle.

Sussman et al. (1998a) mettent également en place un protocole expérimental dans le but d'étudier la capacité des équations de locus à discriminer certaines catégories consonantiques dans différentes conditions de parole. Vingt-deux locuteurs de l'anglais américain (11 femmes et 11 hommes) ont produit des séquences de type CVC où C initial = /b, d, g/ et V = /i, ɪ, e, ε, æ, a, o, u, ʌ, ɔ/ dans deux styles de parole : citée et spontanée. Dans la parole citée, les locuteurs devaient lire lentement et devaient hyperarticuler des séquences de type CV/t/. Chaque séquence a été répétée 3 fois, donnant un total de 1980 séquences obtenues (3 répétitions × 3 consonnes × 10 voyelles × 22 locuteurs). Dans la parole spontanée, les locuteurs devaient lire un texte contenant des mots qui commencent par /b/V, /d/V et /g/V. Les locuteurs se sont d'abord familiarisés avec ce texte et l'ont lu plusieurs fois avant le début de l'enregistrement. Ensuite, lors de celui-ci, il leur a été demandé de lire deux fois le texte comme s'ils avaient une conversation avec un ami et en hyperarticulant. Au total, 15180 mots ont été analysés en parole spontanée (288 /b/ + 252 /d/ + 150 /iəʊ/ multipliés par 22 locuteurs). Plusieurs variables dépendantes ont été utilisées pour examiner le caractère distinctif et la stabilité des équations de locus comme descripteur du lieu d'articulation dans les différents styles de parole. La stabilité coarticulatoire a été évaluée par l'analyse des pentes, le maintien de la linéarité par le R<sup>2</sup> et l'erreur standard (SE). La précision de la classification catégorielle a été étudiée par les analyses discriminantes et la séparation relative des catégories des occlusives par le calcul des distances euclidiennes entre les trois paires d'occlusives (/b/ vs. /d/, /b/ vs. /g/, /d/ vs. /g/) dans l'espace pente-ordonnée à l'origine. Deux équations de locus ont été calculées pour les deux allophones de /g/ (palatal avec les voyelles antérieures et vélaire avec les voyelles postérieures) de manière à neutraliser toutes les variables qui pourraient influencer les valeurs des pentes dans les deux styles de parole. Les résultats montrent de légères variations provoquées par le changement du style de parole (Table 3.19). Les pentes en parole spontanée sont globalement plus élevées que celles en parole citée. Les pentes de la bilabiale /b/ et celles du /g/ vélaire ne changent pas significativement en fonction du style de parole alors que les pentes de l'alvéolaire /d/ et celles du /g/ palatal sont statistiquement influencées par les deux styles de parole. Des analyses multivariées (manova), avec les pentes et les ordonnées à l'origine comme variables dépendantes, indiquent une différence significative des pentes et des ordonnées à l'origine en fonction du lieu d'articulation (/b/, /d/, /g/<sub>p</sub>, /g/<sub>v</sub>). Les comparaisons par paires révèlent une différence significative des moyennes des pentes de toutes les combinaisons possibles. Les analyses sur les R<sup>2</sup>

	Citation				Spontanée			
	P	OO	R <sup>2</sup>	SE	P	OO	R <sup>2</sup>	SE
/b/	0.85(0.07)	175(106)	0.96(0.02)	89(16)	0.85(0.04)	155(75)	0.92(0.03)	104(18)
/d/	0.43(0.08)	1182(147)	0.87(0.06)	70(19)	0.50(0.07)	980(162)	0.74(0.07)	94(10)
/g/p	0.30(0.13)	1657(372)	0.54(0.23)	68(28)	0.47(0.13)	1212(248)	0.55(0.22)	83(20)
/g/v	0.98(0.26)	334(379)	0.68(0.2)	118(35)	0.94(0.21)	310(291)	0.60(0.15)	158(36)

TABLE 3.19 – Valeurs moyennes et écart-types (entre parenthèses) des pentes (P), des ordonnées à l’origine (OO), des R<sup>2</sup> et des SE des 22 locuteurs pour les trois occlusives (avec deux régressions de /g/ : palatal et vélaire) en paroles citée et spontanée (Sussman et al., 1998a).

dévoilent une moyenne de 0.76 en parole citée et une moyenne légèrement plus basse en parole spontanée (0.70). Quant à la moyenne du SE, elle augmente de 86 Hz en parole citée à 110 Hz en parole spontanée. Selon les auteurs, ces changements relativement réduits du pourcentage de la variance expliquée (R<sup>2</sup>) et de l’ajustement des points sur la droite de régression (SE) ont atteint un niveau significatif. Néanmoins, ils n’ont pas diminué l’ordre linéaire des équations de locus. Les analyses discriminantes ont été effectuées pour les trois occlusives sur les pentes et les ordonnées à l’origine dans les deux styles de parole sur deux groupes : un groupe avec une seule régression du /g/ (/b/, /d/, /g/) et un autre avec deux régressions de /g/ (/b/, /d/, /g/p, /g/v). Les moyennes des taux de classification correcte attestent une similarité dans les deux styles de parole des deux groupes : 97% vs. 98.5% avec une régression de /g/ (hasard = 33.3%) et 84% vs. 82% avec deux régressions de /g/ (hasard = 25%). Les analyses de la distance euclidienne révèlent la réduction des séquences dans un espacement acoustique entre les catégories des occlusives (Figure 3.23). La moyenne totale des distances euclidiennes séparant

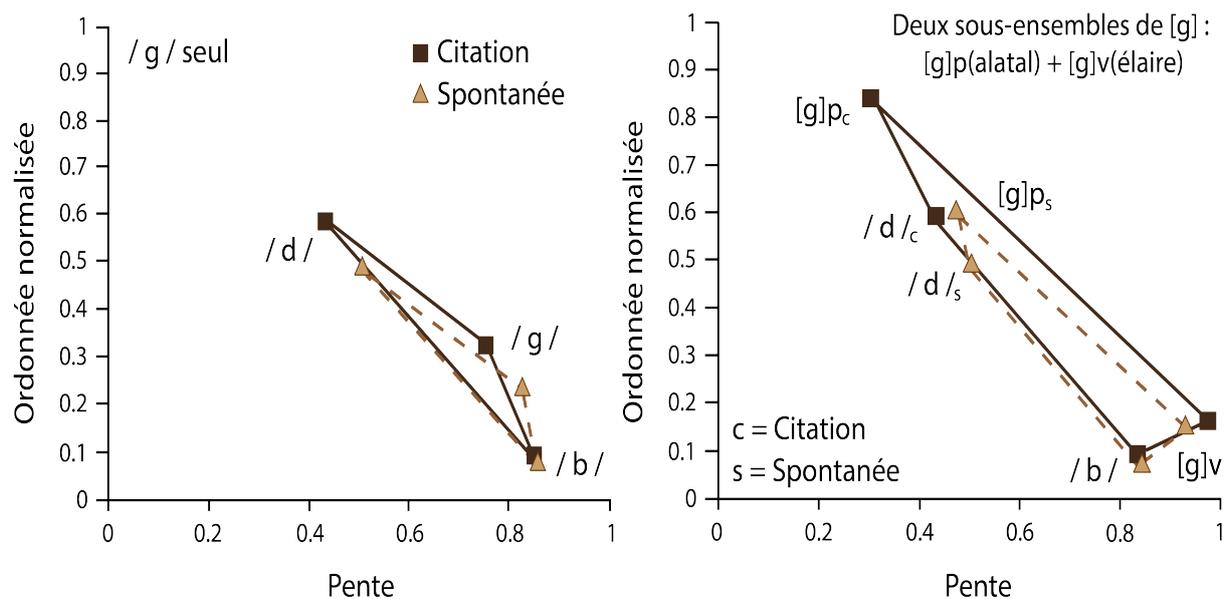


FIGURE 3.23 – Illustration de la distance euclidienne connectant /bV/, /dV/ et /gV/ dans les coordonnées des équations de locus. À gauche, une seule régression et à droite, deux régressions de /g/ ont été traitées (d’après Sussman et al. (1998a)).

les catégories /b-d-g/ en parole citée est de 1.36 comparée à une valeur plus réduite de 1.16 pour la parole spontanée. La différence entre les deux moyennes des distances euclidiennes est significative. Sussman et al. (1998a) exposent que ces résultats mettent en avant la sensibilité des équations de locus à détecter les différentes variations dues à des changements coarticulaires. De plus, ils mentionnent que malgré ces variations, la séparabilité des équations de locus n'a pas été affectée. Les auteurs expliquent que les équations de locus n'éliminent pas la variabilité du signal mais elles la traitent directement. En effet, quand les séquences sont codées par des paramètres acoustiques appropriés, la variabilité devient réglementée d'une manière automatique et organisée par les équations de locus. Selon les auteurs, les équations de locus mènent à l'hypothèse que les systèmes neuronaux de l'être humain développent des sensibilités à des régularités statistiques réglementées qui codent la structure phonétique du signal de la parole.

### 3.7 L'origine articulaire des équations de locus

Les équations de locus sont basées sur le calcul des informations acoustiques afin d'indexer le degré de coarticulation et les lieux d'articulation des consonnes dans des séquences CV. Dès les premières interprétations des équations de locus, un lien a été établi entre les variations systématiques des paramètres des équations de locus et les contraintes articulaires des consonnes. Par exemple, Krull (1987) explique que l'engagement faible de la langue pendant la production des bilabiales conduit à un chevauchement important avec les voyelles suivantes expliquant ainsi leurs valeurs de pente élevées. L'étude de Chennoukh et al. (1997), qui utilise un modèle de simulation articulaire, a montré que les changements des mouvements articulaires sont reflétés dans les variations des pentes et des ordonnées à l'origine. L'objectif de cette section est d'explorer la relation entre les mesures des équations de locus et celles des mouvements articulaires.

Löfqvist (1999) souligne que les études sur les équations de locus laissent suggérer l'existence d'une relation directe entre la pente et le degré de coarticulation, pourtant aucune vérification avec des mesures articulaires n'a été réalisée afin d'examiner cette suggestion. L'auteur expose qu'un changement important dans la forme du conduit vocal amène à une grande variation acoustique alors qu'un changement articulaire minimal conduit à une petite modification acoustique. Il met alors en exergue que la relation articulaire-acoustique est codifiée par la forme du conduit vocal et le signal acoustique. Selon lui, cette relation n'est pas linéaire. Löfqvist (1999) effectue alors un protocole expérimental pour examiner la relation des équations de locus avec des mesures articulaires de la coarticulation. Il essaie de vérifier si les observations acoustiques du degré de coarticulation obtenues par les équations de locus ont des origines articulaires. Quatre locuteurs (2 féminins et 2 masculins) de l'anglais américain ont produit des séquences  $V_1CV_2$  répétées 10 fois, où  $V_1$  et  $V_2$  sont l'une des voyelles /i, a, u/ et C correspond à /b, d, g/. Les mouvements des lèvres, de la mâchoire et de la langue ont été enregistrés à l'aide d'un EMMA. Les récepteurs de l'EMMA ont été placés sur les lèvres, sur les incisives in-

férieures et sur quatre points de la langue (apex, lame, dos, racine). Trois mesures articulatoires différentes ont été réalisées pour les comparer avec les pentes des équations de locus.

La première mesure consiste à calculer l'intervalle temporel entre le début du mouvement de la fermeture des lèvres de l'occlusive bilabiale et le début du mouvement de la langue de la première à la seconde voyelle dans des séquences VCV. Elle est obtenue par l'addition des distances euclidiennes entre le début et la fin de la fermeture pour chaque récepteur. Lorsque le chevauchement entre les mouvements de la langue et des lèvres augmente, un accroissement de la valeur de la pente des équations de locus est attendu. Une augmentation du chevauchement articulatoire pourrait être marquée par un début plus rapide du mouvement lingual que celui du mouvement des lèvres. La deuxième mesure correspond à la magnitude du mouvement lingual pendant la fermeture orale de l'occlusive. Le mouvement global de la langue est acquis par le calcul de la moyenne des mouvements des quatre récepteurs durant la clôture orale. Un grand mouvement lingual lors la fermeture signifie un degré de coarticulation élevé entre la consonne et la voyelle suivante. Par conséquent, la consonne est moins résistante et une pente élevée est prévue. La troisième mesure repose sur la magnitude du mouvement lingual pendant la voyelle qui suit la consonne. Un mouvement lingual réduit entre le relâchement de la consonne et la voyelle cible implique un grand degré de chevauchement articulatoire, car une grande partie du mouvement lingual se serait produite pendant la consonne. Dans ce cas, une pente élevée est également attendue. La même procédure que celle de la deuxième mesure a été utilisée pour évaluer la moyenne du mouvement lingual de la fin de la première voyelle jusqu'à la fin de la seconde.

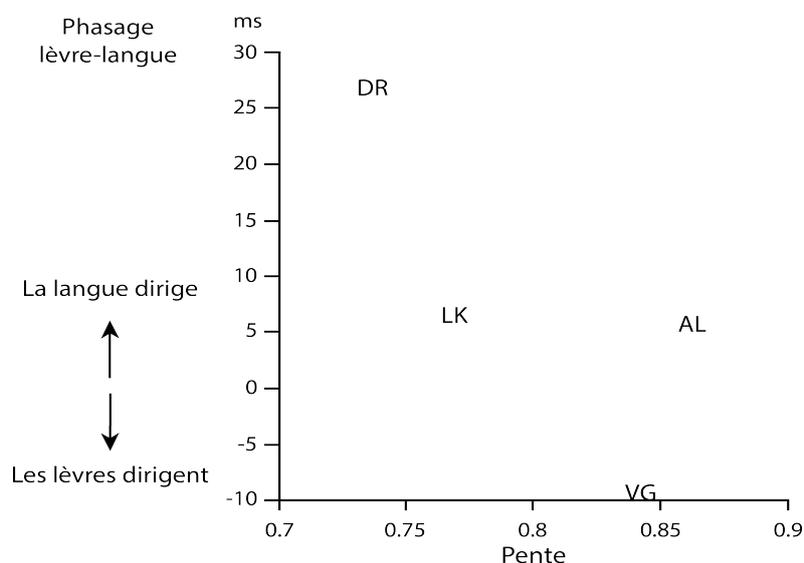


FIGURE 3.24 – Illustration des pentes en abscisse en fonction de l'intervalle entre le début du mouvement lingual de la première voyelle à la seconde et le début de la fermeture de la bilabiale pour chaque locuteur (DR, LK, VG, AL) (d'après Löfqvist (1999)).

Les équations de locus ont ensuite été calculées pour les trois occlusives afin de comparer les

pentés avec les mesures articulatoires<sup>49</sup>. Les mesures articulatoires du phasage lèvres-langue ont été moyennées pour toutes les séquences de chaque locuteur et affichées (sur l'axe des ordonnées) en fonction des valeurs de pente (sur l'axe des abscisses) de façon à illustrer la relation entre ces mesures et les pentes (Figure 3.24). Le mouvement de la langue qui commence avant celui des lèvres montre un chevauchement articulatoire élevé. La relation attendue est la suivante : lorsque la langue domine, la valeur de la pente doit augmenter ; or, cette relation n'a pas été observée (Figure 3.24). Les mesures de la magnitude du mouvement lingual pendant la fermeture orale des occlusives ont été affichées en fonction des pentes. Les pentes des vélaires sont plus élevées que celles des alvéolaires pour tous les locuteurs. Pourtant, le mouvement lingual des alvéolaires est plus important que celui des vélaires pour deux locuteurs et similaires pour les deux autres. Les pentes du /b/ dans les contextes vocaliques asymétriques sont plus importantes que celles dans les contextes vocaliques symétriques, toutefois, le mouvement de la langue est plus grand pour les /b/ asymétriques pour tous les locuteurs. Les mesures de la magnitude du mouvement lingual durant la seconde voyelle sont également affichées par rapport aux valeurs des pentes des occlusives. Dans ce cas, quand le mouvement de la langue est minimal pendant la seconde voyelle, une pente élevée des équations de locus est attendue. Ce postulat est vrai puisque la majorité du mouvement lingual a été anticipée et est produite pendant la consonne. La bilabiale (contexte asymétrique) a les plus petites valeurs du mouvement lingual pendant la production de la seconde voyelle, pourtant, elle n'a pas toujours les pentes les plus élevées (à l'exception d'un locuteur). Aucune relation n'a été constatée entre les pentes de la vélaire et de l'alvéolaire avec le mouvement lingual. Löfqvist (1999) explique que les résultats de cette étude contredisent l'idée que la pente des équations de locus est associée au degré de coarticulation entre la consonne et la voyelle. Il indique que la relation en faveur des équations de locus observée dans le travail de Chennoukh et al. (1997) est provoquée par le contrôle des paramètres de simulation. Par exemple, le mouvement de la langue a été fixé comme condition de simulation, ce qui n'est pas le cas en parole naturelle. De plus, la parole naturelle subit des variations des amplitudes de la vélocité et de la durée du mouvement lingual alors que ces paramètres ont été maintenus constants dans les données simulées. L'auteur conclut que « [...] *it appears that caution is necessary in making claims that locus equation slope can be used to assess differences in coarticulation and applied to studies of developmental issues [...] or deviant speech production [...]* »<sup>50</sup>.

Tabain (2000) examine également la relation entre les mesures articulatoires et la pente des équations de locus. Elle utilise des données électropalatographiques (EPG) dans le but de

49. Les équations de locus de la bilabiale ont été estimées avec des contextes vocaliques symétriques ( $V_1=V_2$ ) et asymétriques ( $V_1 \neq V_2$ ) pour vérifier si la pente est influencée par le contexte vocalique. Les pentes de /b/ sont plus élevées dans les contextes vocaliques symétriques (moyenne = 0.86) que dans ceux asymétriques (moyenne = 0.80). Quant à la vélaire et à l'alvéolaire, trois pentes ont été calculées pour chaque consonne en fonction de la voyelle initiale.

50. « [...] il apparaît nécessaire d'être prudent lorsque l'on prétend que les équations de locus peuvent être utilisées pour décrire les différents degrés de coarticulation dans les études appliquées dans les questions de l'acquisition développementale ou dans les troubles du langage [...] », page 2029 [notre traduction].

déterminer s'il existe une corrélation entre les données articulatoires et celles des équations de locus sur la coarticulation. Il a été avancé dans la littérature que les équations de locus déduisent le degré de l'engagement de la langue au cours de la production des séquences CV (Krull, 1987, 1989). L'objectif de l'étude de Tabain (2000) est de tester cette hypothèse en comparant les données des équations de locus avec celles obtenues à l'aide d'un EPG. Quatre locutrices de l'anglais australien ont prononcé des syllabes CV où C = /ð, z, ʒ, d, g, n, ŋ, l, r/ et V = /ɪ, e, æ, ɐ, ɔ, ʊ, i:, e:, ɛ:, o:, ɜ:, ɜ:/. Les locutrices portaient des électropalatographes contenant 62 électrodes pendant la production des séquences. Les coefficients des équations de locus ont été prélevés au début de la transition du F2 ( $F2_{onset}$ ) et à la cible vocalique de F2 ( $F2_{cible}$ )<sup>51</sup>. Les données acoustiques ( $F2_{onset}$  et  $F2_{cible}$ ) ont été exploitées pour calculer les équations de locus. Tabain (2000) emploie deux points de mesure articulatoires déterminés au moment de la réalisation de la consonne et de la voyelle afin de calculer une régression linéaire similaire à celle des équations de locus. Les données articulatoires de l'EPG correspondent à la somme des contacts des électrodes (entre 0 et 62) au milieu acoustique de la consonne et à celle à la cible acoustique de la voyelle. L'auteure explique que la méthode des mesures de l'EPG est basée sur le fait que le total des contacts des électrodes d'une consonne varie selon le contexte vocalique. Cette relation peut donc être considérée comme un indice de coarticulation. En effet, dans les différents contextes vocaliques, les consonnes présentant peu de coarticulation avec les voyelles suivantes auront une variabilité minimale du total des contacts tandis que les consonnes qui exposent une influence coarticulatoire plus importante que celle des voyelles suivantes révéleront une variabilité du nombre de contacts total. Par exemple, si une consonne implique un geste coronal et qu'elle n'est pas résistante à la coarticulation, il existera plus de contact dans la moitié postérieure du palais en contexte de voyelles hautes que dans celui des voyelles basses. Par contre, si la consonne coronale est résistante à la coarticulation, le nombre de contacts dans la zone arrière du palais reste constant dans les différents contextes vocaliques.

Les mesures acoustiques et articulatoires ont servi au calcul des deux types de régression. Les pentes de la régression des données de l'EPG ont été comparées à celles des équations de locus. Les résultats montrent une corrélation forte des pentes des équations de locus avec un coefficient de corrélation ( $r$ ) moyenne de 0.86, contrairement aux régressions de l'EPG qui ont une moyenne de  $r$  de 0.44. En accord avec la littérature, les valeurs de pentes des équations de locus sont élevées pour les dorsales (/g, ŋ/) et basses pour les coronales (/d, n, z, r, l, ð, ʒ/). Cette constatation a été également remarquée avec les pentes des régressions des données de l'EPG. L'auteure souligne que la corrélation entre les pentes des régressions des deux types de données est relativement claire pour les occlusives (orales et nasales) mais elle ne l'est pas pour les fricatives. Les données de l'EPG et celles des équations de locus témoignent d'une incohérence, particulièrement pour /z/ et /ʒ/ (Figure 3.25). Une grande variabilité entre les locutrices

51. La cible vocalique correspond à la valeur maximale du F2 s'il avait une forme de "n" ou minimale s'il avait une forme de "u". Quand le formant est relativement plat ou en cas d'ambiguïté, la cible est déterminée au milieu du F2.

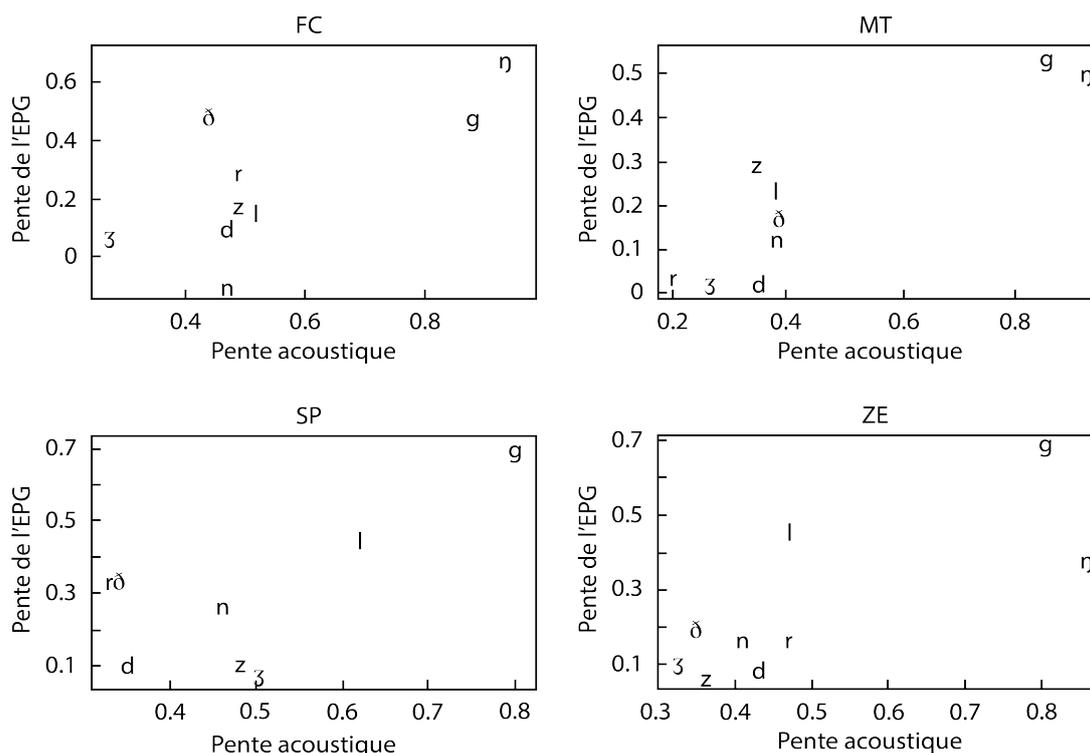


FIGURE 3.25 – Représentations des valeurs de pentes de l'EPG en fonction de celles des équations de locus pour chacune des locutrices (d'après Tabain (2000)).

est également notée. L'auteure déclare que ces résultats suggèrent que les équations de locus ne fournissent pas d'information précise sur le degré de coarticulation pour les fricatives. Elle propose deux explications à ces observations : 1) la corrélation réduite entre les pentes de l'EPG et celles des équations de locus des fricatives est le résultat du masquage de la transition de F2 par le bruit de friction de la consonne qui conduit à des mesures de  $F2_{onset}$  imprécises ; 2) les équations de locus peuvent capturer les différents degrés de résistance coarticulatoire entre les consonnes qui impliquent deux articulateurs différents (le corps de la langue pour les vélaire et l'apex pour les alvéolaires), mais elles ne sont pas assez précises pour capter les différences coarticulatoires entre les consonnes qui partagent le même articulateur.

Tabain (2002) étend son travail exploité dans son précédent article (Tabain, 2000) en examinant la pertinence de l'utilisation des équations de locus avec les consonnes non voisées. Quatre locutrices de l'anglais australien ont réalisé des syllabes CV où  $C = /t, k, θ, s, f/$  et  $V = /i, e, æ, ɐ, ɔ, u, ɪ, eɪ, ɐɪ, oɪ, ɛɪ, ɜɪ/$ . Les données articulatoires ont été collectées à l'aide des électropalato-graphes portés par les locutrices lors de l'enregistrement. Les mêmes procédures d'analyse dans l'étude de Tabain (2000) ont été suivies. Deux équations de locus ont été estimées pour chacune des occlusives non voisées  $/t, k/$  selon le moment de prélèvement du  $F2_{onset}$  : au moment du relâchement de la consonne  $F2_{@burst}$  et à la première pulsation de la voyelle suivante  $F2_{onset}$ . Les résultats indiquent des pentes des équations de locus basses pour les fricatives coronales, intermédiaires pour l'occlusive alvéolaire et élevées pour l'occlusive vélaire (Table 3.20). Les pentes des occlusives estimées avec  $F2_{@burst}$  révèlent des pentes très basses pour le  $/t/_{@burst}$  et

	équations de locus			régressions articulatoires		
	P	OO	r	P	OO	r
/θ/	0.50	924	0.91	0.33	5.33	0.36
/s/	0.54	885	0.91	0.19	19.59	0.36
/ʃ/	0.50	1107	0.89	0.02	24.22	0.05
/t/	0.73	610	0.95	–	–	–
/k/	0.89	323	0.95	–	–	–
/t/@burst	0.10	1859	0.31	0.03	29.63	0.04
/k/@burst	0.64	834	0.84	0.63	11.67	0.63

TABLE 3.20 – Moyennes des pentes (P) et des ordonnées à l'origine (OO) des équations de locus et des régressions avec les données articulatoires des quatre locutrices (Tabain, 2002).

moyennes pour /k/@burst. La comparaison de ces résultats avec ceux des consonnes voisées de Tabain (2000) met en avant que les non voisées sont moins résistantes que leurs contreparties voisées. Au contraire, les pentes des occlusives mesurées au moment du relâchement montrent que les occlusives non voisées sont plus résistantes que leurs équivalentes voisées. Les résultats des pentes des données de l'EPG signalent des valeurs basses avec une corrélation faible. L'auteure explique que ces valeurs basses de pente et de r suggèrent que la relation dans les données articulatoires est moins monotone que celle dans les données des équations de locus. D'après ces résultats, la variable indépendante (la voyelle) contribue très peu à la variable dépendante (la consonne). Par conséquent, ces consonnes sont très résistantes à la coarticulation. La comparaison entre les pentes des deux ensembles de données témoigne d'une faible corrélation. Cette faiblesse de corrélation entre les pentes des données articulatoires et acoustiques des consonnes non voisées est plus importante que celle observée avec les consonnes voisées. Toutefois, les pentes des équations de locus des occlusives mesurées au moment du relâchement marquent des pentes plus précises puisque les pentes acoustiques (/t/@burst et /k/@burst) vs. articulatoires (/t/ et /k/) sont similaires avec une corrélation élevée. En parallèle, les analyses statistiques sur les pentes des données articulatoires attestent une absence d'effet de voisement. Autrement dit, les paires des fricatives voisées vs. non voisées et des occlusives voisées vs. non voisées ne sont pas significativement différentes. Néanmoins, une différence significative est observée avec les pentes des équations de locus entre les consonnes voisées vs. non voisées. Tabain (2002) déclare que les équations de locus ne sont pas des mesures fiables de la différence coarticulatoire quand une paire voisée vs. non voisée est comparée. Le VOT important des consonnes non voisées donne l'illusion d'un degré de coarticulation élevé dans les analyses des équations de locus. Des différences significatives de pente en fonction du mode d'articulation (/s/ vs. /t/) ont été remarquées pour toutes les locutrices avec les données acoustiques tandis que les pentes avec les données articulatoires attestent les valeurs non significatives pour deux locutrices. Selon l'auteure, ces résultats exposent que les équations de locus ne doivent pas être utilisées pour comparer les différents modes d'articulation parce que les VOT peuvent exagérer les différences lorsqu'elles existent. Elle conclut que « [...] it is clear that a locus equation analysis is even less

*accurate for voiceless consonants than for voiced consonants, and that it performs more poorly on English fricatives than on English stops* »<sup>52</sup>.

Suivant les travaux de Recasens (1984a,b, 1985) sur la résistance coarticulatoire, Fowler et Brancazio (2000) examinent les différences dans la résistance coarticulatoire des consonnes de l'anglais américain. Les auteurs précisent que l'engagement de la langue est le principal responsable du degré de résistance coarticulatoire des segments. Ils utilisent donc des données articulatoires de la langue ainsi que celles acoustiques du F2 pour estimer le degré de la coarticulation anticipatoire entre la consonne et la voyelle suivante. La corrélation est analysée entre la résistance coarticulatoire mesurée avec des données articulatoires et les pentes des équations de locus. Les auteurs mettent en évidence l'existence d'une corrélation forte entre les mesures de la résistance coarticulatoire réalisées par Recasens (1985)<sup>53</sup> et les pentes des équations de locus mesurées dans l'étude de Sussman et al. (1991) et celle de Sussman (1994). Lorsque la valeur des pentes augmente, celle de l'écart-type croît également. Fowler et Brancazio (2000) cherchent alors les origines articulatoires de cette résistance coarticulatoire observée dans le domaine acoustique. Les données articulatoires ont été obtenues à l'aide d'un EMMA. Plusieurs électrodes ont été placées sur la langue pour détecter les mouvements horizontaux et verticaux de celle-ci. Les auteurs ont choisi l'électrode située sur le corps de la langue à analyser dans cette étude puisqu'elle fournit des informations directes sur le lieu de la constriction des voyelles. Deux locuteurs américains ont réalisé des séquences /ə/CV où C = /b, v, ð, d, z, ʒ, g/ et V = /i, a, ʌ/. Les informations acoustiques et articulatoires ont été prélevées à neuf moments de chaque séquence (T1-T9). Ces points de mesures ont été sélectionnés en précisant le début et la fin acoustique du schwa (T1 et T4) et de la voyelle finale (T6 et T9). T2 et T3 sont déterminés à un tiers et deux tiers, respectivement, de la distance entre T1 et T4 ; T7 et T8 ont été fixés d'une façon similaire entre T6 et T9. Quant à T5, il a été déterminé au milieu de la distance entre la fin du schwa (T4) et le début de la voyelle finale (T6).

Les résultats désignent que les consonnes labiales /b, v/ et la vélaire /g/ ont été associées aux grandes variations des mouvements vertical et horizontal du corps de la langue pendant leur constriction (T5). Les autres consonnes linguales sont moins marquées par les variations de la position de la langue en fonction de la voyelle suivante. Les variations acoustiques au début des transitions de F2 (T6) pour les labiales et pour le /g/ sont plus grandes que pour les autres consonnes linguales. Les deux locuteurs ont montré des effets de la coarticulation anticipatoire voyelle-à-voyelle. Le corps de la langue est plus avancé et plus élevé dans les séquences avec un /i/ final que celles avec un /ʌ/ durant le schwa. Les fréquences de F2 du

52. « [...] il est clair que les analyses des équations de locus sont moins précises pour les consonnes non voisées que celles voisées et la précision est encore plus faible avec les fricatives que les occlusives en anglais », page 36 [notre traduction].

53. Recasens (1985) estime la résistance coarticulatoire d'une consonne par le calcul de l'écart-type de  $F2_{onset}$  de la consonne produite avec plusieurs contextes vocaliques. Une valeur élevée de l'écart-type souligne une variation conséquente de la consonne en fonction de la voyelle et une résistance coarticulatoire réduite alors qu'une valeur basse de l'écart-type révèle une stabilité de la consonne dans les différents contextes vocaliques et une résistance coarticulatoire élevée (cf. Chapitre 2 section 2.3.4).

schwa sont également plus hautes dans le contexte de /i/ que celui de /Λ/. Il est à remarquer que les consonnes résistantes freinent l'influence de la seconde voyelle sur la première. Autrement dit, le schwa est impacté au maximum par la seconde voyelle devant les consonnes les moins résistantes (Figure 3.26)<sup>54</sup>. Les résultats des équations de locus dévoilent des valeurs de

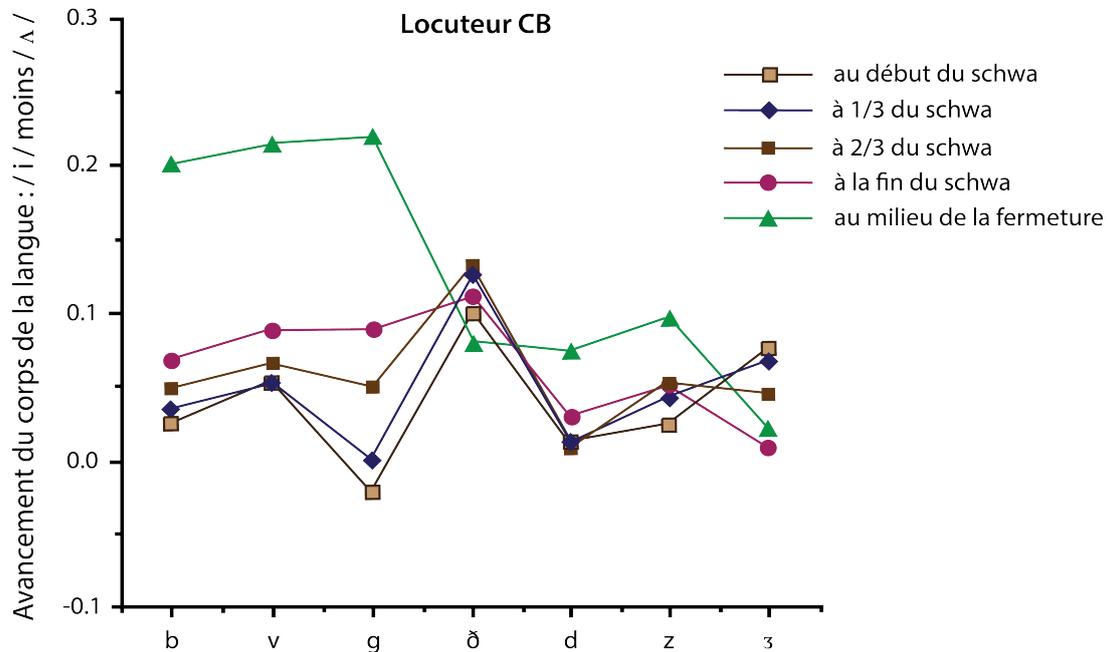


FIGURE 3.26 – Les différences de voyelles (/i/ – /Λ/) dans le mouvement horizontal du corps de la langue pris à 5 moments (du début du schwa à la fermeture consonantique) des séquences produites par le locuteur CB (Fowler et Brancazio (2000)).

pentés élevées pour les labiales et basses pour les linguales. Une corrélation forte a été observée entre les valeurs de pente et la mesure du degré de résistance coarticulatoire appliquées aux données acoustiques et articulatoires. Cette corrélation est en accord avec celle rapportée dans les données de Recasens (1985) et celles de Sussman et al. (1991) et Sussman (1994). Fowler et Brancazio (2000) calculent deux équations linéaires, avec TBx et TBy, qui prennent comme coefficients, les positions du corps de la langue au relâchement de la consonne (T6) et au milieu de la voyelle (T8) dans les deux dimensions (horizontale et verticale). Ces équations ont pour objectif de trouver une relation directe entre les données acoustiques interprétées par les équations de locus et celles articulatoires. Les valeurs des pentes des équations de TBx et de TBy sont relativement similaires à celles notées dans les pentes des équations de locus, soit des pentes élevées pour les consonnes labiales et basses pour les consonnes linguales. Les auteurs soulignent que le nombre réduit de voyelles dans cette étude ne permet pas d'aborder la ques-

54. Le degré de la résistance coarticulatoire de la consonne est calculé par la formule suivante : /i/ – /Λ/. Pour les données acoustiques, les valeurs de  $F2_{onset}$  (T6) sont utilisées dans la formule pour une consonne donnée (ex. : la résistance coarticulatoire pour la consonne /d/ =  $F2(d)/i/ - F2(d)/\Lambda/$ ). Quant aux données articulatoires, les mouvements linguaux horizontal (TBx) et vertical (TBy) sont mesurés pendant la clôture de la consonne (T5) (ex. : la résistance coarticulatoire (sur le plan horizontal) pour la consonne /v/ =  $TBx(v)/i/ - TBx(v)/\Lambda/$ ). Dans les deux types de données, une valeur élevée de /i/ – /Λ/ indique une résistance coarticulatoire réduite et une valeur basse signifie une résistance importante.

tion de la détermination de la linéarité des équations de locus par des variables dans le domaine articulatoire. Pourtant, ces analyses préliminaires soutiennent l'existence d'une relation étroite entre les pentes des équations de locus et celles des régressions de TBx et TBy.

Iskarous et al. (2010) s'inspirent du travail de Fowler et Brancazio (2000) et examinent les bases articulatoires des équations de locus. Les auteurs mentionnent que le mouvement horizontal du corps de la langue a été associé au F2 dans la littérature (Joos, 1948; Delattre, 1951). Quand le corps de la langue est avancé, F2 est généralement élevé et lorsqu'il est rétracté, F2 est souvent bas. Par conséquent, des régressions linéaires ont été estimées avec les positions horizontales du corps de la langue après le relâchement de la consonne et au milieu de la voyelle pour une consonne dans plusieurs contextes vocaliques. Ces régressions ont pour objectif d'évaluer l'origine articulatoire des équations de locus avec le mouvement horizontal du corps de la langue. Deux ensembles de données ont été analysés ; le premier correspond à des données articulatoires (et acoustiques) issues d'un EMMA et le second contient des données articulatoires (et acoustiques) provenant des images de rayon X *microbeam* (XRMB). Les données d'EMMA ont été obtenues d'une locutrice de l'anglais américain. Des séquences əCV ont été prononcées par la locutrice où C = /b, d, g, v, ð, z/ et V = /i, ei, ε, ʌ, u, a/. Les séquences ont été répétées quinze fois. Les bobines d'EMMA ont tracé les mouvements des lèvres, de la mâchoire et de quatre points sur la langue. Les données acoustiques ont été segmentées pour obtenir les valeurs de F2 au relâchement de la consonne et au milieu de la seconde voyelle. Pour chaque séquence, les positions des articulateurs aux deux mêmes moments ont été extraites. Quant aux données du XRMB, des syllabes CV ont été prélevées du corpus de *Wisconsin X-ray Microbeam* produit par trente-huit locuteurs (Westbury, 1994). Les consonnes sont les occlusives bilabiales /b, p/, alvéolaires /t, d/ et vélaire /k/ et la fricative alvéolaire /s/. Chaque catégorie des lieux d'articulation a été réalisée avec huit voyelles différentes. L'avantage d'utiliser les données de cette base de données est le nombre important des locuteurs à étudier, ce qui est relativement rare dans les expériences articulatoires de production de la parole.

Les auteurs émettent l'hypothèse que si les résultats de la locutrice de l'EMMA est en accord avec ceux du XRMB, les résultats sont fiables. L'extraction des données du XRMB est comparable à celle de l'EMMA ; les positions des lèvres, de la mâchoire et du corps de la langue ont été tracées au relâchement de la consonne et au milieu de la voyelle pour chaque séquence. Deux types de régressions linéaires ont été estimés. La première régression a été calculée à partir des valeurs de  $F2_{onset}$  et de  $F2_{mid}$  (équations de locus traditionnelles). La seconde a été estimée en prenant la mesure de la position horizontale de la langue TBx à deux moments : au relâchement de la consonne et au milieu de la voyelle (régressions articulatoires). Les résultats des régressions des données de l'EMMA affichent des pentes des équations de locus similaires à celles rapportées dans la littérature avec une linéarité élevée (moyenne de  $R^2 = 0.95$ ). Les régressions articulatoires de TBx montrent également une linéarité élevée (moyenne de  $R^2 = 0.92$ ). Les valeurs de pente de TBx ne sont pas identiques à celles des équations de locus pourtant elles suivent les même tendances : 1) les pentes sont élevées pour les labiales, intermé-

diaires pour les vélares et basses pour les alvéolaires ; 2) les ordonnées à l'origine sont élevées pour les alvéolaires, intermédiaires pour les vélares et basses pour labiales (Figure 3.27). De

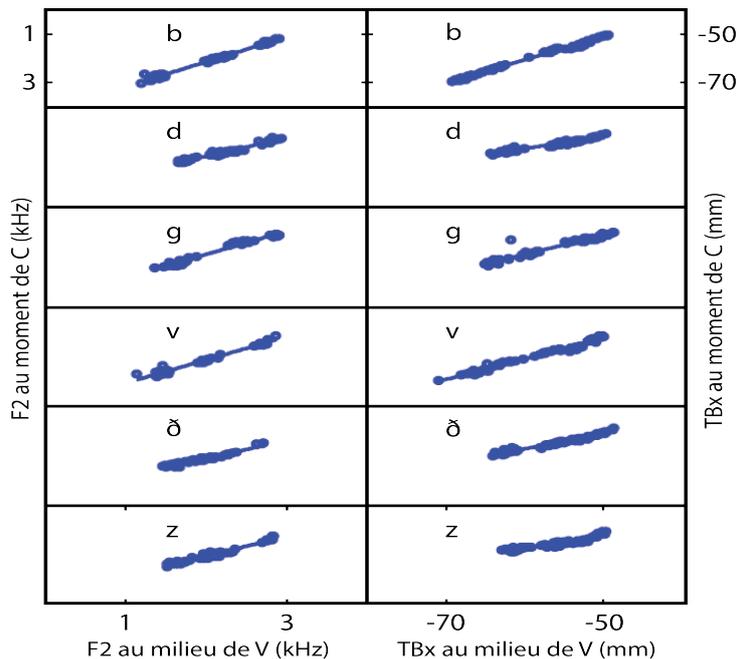


FIGURE 3.27 – Illustration des deux fonctions linéaires acoustique et articuloire des données de l'EMMA d'une locutrice. À droite, la position horizontale du dos de la langue (TBx) au relâchement de la consonne (en ordonnée) en fonction de celle au milieu de la voyelle (en abscisse), à gauche, les valeurs de F2 au relâchement de la consonne contre celles au milieu de la voyelle (Iskarous et al., 2010).

plus, une corrélation forte a été observée entre les pentes ( $r = 0.87$ ) et les ordonnées à l'origine ( $r = -0.89$ ) des deux types de régressions. Les données du XRMB confirment que les résultats des pentes et des ordonnées à l'origine, articuloires et acoustiques, sont similaires à ceux des données de l'EMMA (pentes : labiales > vélares > alvéolaires, ordonnées à l'origine : alvéolaires > vélares > labiales). Des analyses statistiques utilisant les modèles mixtes linéaires ont été effectuées avec deux variables indépendantes : la modalité (avec 2 niveaux : articuloire et acoustique) et le segment (avec 4 niveaux : p/b, t/d, k, s). Les pentes et les ordonnées à l'origine étaient les variables dépendantes. L'objectif de ces analyses est de vérifier si les deux régressions articuloires et acoustiques peuvent distinguer les différentes catégories consonantiques. Les auteurs soulignent que si les régressions différencient les lieux d'articulation d'une manière équivalente, l'argument de l'existence de la linéarité des équations de locus au niveau articuloire est validé. Les résultats révèlent que les pentes de toutes les catégories sont significatives à l'exception de /s/ vs. /d-t/ et /b-p/ vs. /g-k/. De plus, les alvéolaires dans les différents modes d'articulation ne sont pas distinguées par la pente, soutenant l'hypothèse que les pentes peuvent être une propriété du lieu d'articulation. Les résultats des ordonnées à l'origine affichent des différences significatives de toutes les combinaisons de paires à l'exception de /d-t/ vs. /s/. Ces résultats sont statistiquement équivalents dans les données acoustiques et articuloires. D'après ces résultats, les équations de locus interprètent le mouvement hori-

zontal du dos de la langue qui est, en l'occurrence, lié au lieu d'articulation. Le mouvement du dos de langue (TBx) tracé par l'EMMA pour chaque consonne à partir du début acoustique de la transition F2 varie en fonction du lieu d'articulation (Figure 3.28). Dans le cas de l'alvéo-

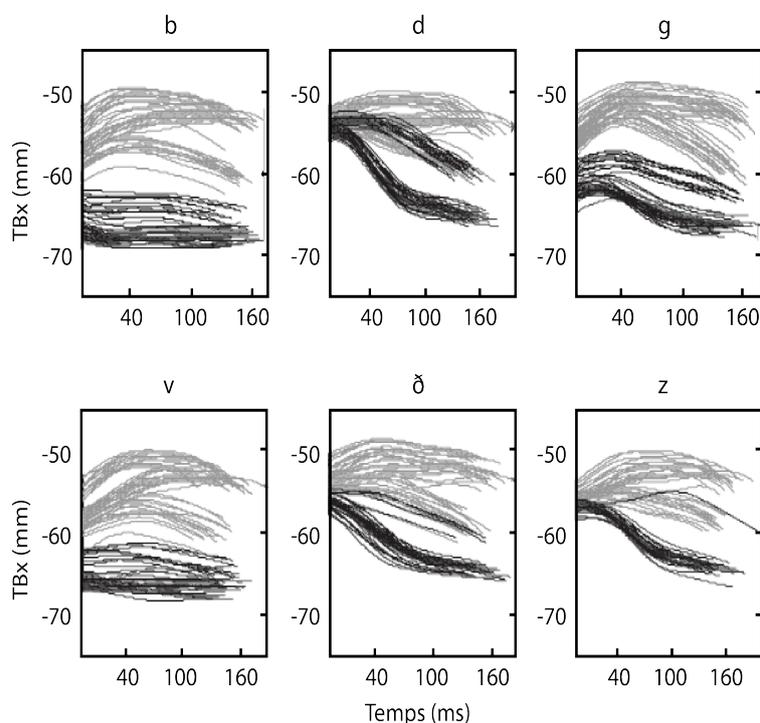


FIGURE 3.28 – Trajectoires de TBx pour les consonnes des données de l'EMMA. Les débuts des trajectoires correspondent aux débuts acoustiques des transitions formantiques de chaque consonne. Les trajectoires en noir représentent le contexte avec les voyelles postérieures et celles en gris illustrent le contexte des voyelles antérieures. L'avancement croît verticalement (Iskarous et al., 2010).

laire /d/ au début des transitions, le dos de la langue est avancé si la voyelle est antérieure ou postérieure. Par conséquent, la moyenne de TBx au moment de la réalisation de la consonne /d/ est plus élevée que pour les autres consonnes (langue + avancée). Quant à la vélaire /g/, l'avancement du dos de la langue est moins présent et ainsi, la moyenne de TBx est moins importante au début des transitions vers la voyelle. Pour les labiales, la position horizontale du dos de la langue est complètement dépendante de la voyelle suivante. Des différences en fonction du mode d'articulation peuvent également être notées. Le dos de la langue est plus postérieur pour /z/ que pour /d/ avec les voyelles postérieures. Ce phénomène résulte des contraintes plus importantes du dos de la langue pour les fricatives que pour les occlusives (Recasens et Espinosa, 2009). D'ailleurs, la résistance coarticulatoire mesurée avec l'écart-type de TBx et de  $F2_{onset}$  est fortement corrélée avec les valeurs des pentes acoustique et articulatoire. Selon les résultats de cette étude, les mesures articulatoires de la production de la parole fournissent une preuve de l'origine articulatoire des équations de locus. La fonction qui relie le mouvement horizontal du corps de la langue de la consonne à la voyelle est linéaire et l'ensemble des analyses statistiques décrivant les régressions sont les mêmes que celles des équations de locus. Il

a été observé que la linéarité des équations de locus est déjà présente dans la relation linéaire de TBx. Cette constatation implique la présence de contraintes sur le corps de la langue qui entraînent diverses résistances coarticulatoires et par conséquent, la formation des différentes consonnes. Les équations de locus offrent alors une mesure de la synergie du corps de la langue durant la production de la parole. Par exemple, si la pente des équations de locus est basse pour une consonne, l'interprétation est que le dos de la langue est crucial pour la réalisation de la consonne. Cela explique pourquoi la parole hypo-articulée mène à des pentes élevées. Si l'utilisation synergique du dos de la langue qui aide l'apex ou la lame est réduite dans le cas de la parole en débit rapide ou non accentuée, les pentes deviennent automatiquement plus hautes. En revanche, Iskarous et al. (2010) déclarent que « *The present study generated no evidence that supports the use of locus equations as invariant or relationally invariant cues for stop location. There is of course a relation between the place of primary constriction of a consonant and whether the tongue back assists in the achievement of that constriction, but this relation is not direct enough to allow for the use of locus equations as a direct measure of place* »<sup>55</sup>.

Plus récemment, Noiray et al. (2013) calculent les équations de locus acoustiques ( $F2_{onset} \sim F2_{mid}$ ) et articulatoires (TBx ~ TBy) dans l'objectif de comparer le développement coarticulatoire chez les enfants (4-5 ans) avec la coarticulation chez les adultes en français canadien. Le mouvement horizontal de la langue (TB) a été obtenu grâce aux images ultrasons. Les résultats de cette étude sont en accord avec ceux observés dans le travail d'Iskarous et al. (2010). Les pentes et les ordonnées à l'origine des régressions linéaires représentant les deux positions horizontales de la langue (au relâchement et au milieu de la voyelle) lors de la réalisation de chaque occlusive sont similaires à celles calculées avec des équations de locus pour les deux populations (enfants et adultes).

Lindblom et Sussman (2012) proposent une autre approche méthodologique que celle présentée par Iskarous et al. (2010) pour expliquer la source articulatoire des équations de locus. Ils abordent la question de l'origine de la linéarité des équations de locus par l'analyse des changements temporels du contour de la langue quand elle se déplace de  $V_1$ -à-C-à- $V_2$ . Les données sont issues de 20 secondes d'images de rayons X d'un locuteur suédois qui a produit des séquences / $\epsilon$ 'CV/ où C = /b, d, g/ et V = /i:, e:, a,  $\alpha$ :,  $\text{ɔ}$ , u:/ (Lindblom (2003), Lindblom et Sussman (2004)). Les trajectoires horizontale et verticale de la langue ont été tracées toutes les 20 ms pour chacune des séquences (Figure 3.29)<sup>56</sup>.

Les points représentant la fermeture de la consonne (cercles blancs) s'écartent systématiquement de la position hypothétique de la langue (fermeture cible, lignes bleues) selon la distance entre la cible consonantique et celle vocalique de  $V_2$ . Autrement dit, plus la distance

55. « La présente étude n'a pas généré de preuves qui soutiennent l'utilisation des équations de locus comme un invariant ou un invariant relationnel du lieu d'articulation des occlusives. Il existe bien sûr une relation entre le lieu de constriction principale d'une consonne et la participation du dos de la langue à l'achèvement de cette constriction. Toutefois, cette relation n'est pas assez directe pour permettre l'emploi des équations de locus comme une mesure directe du lieu d'articulation », page 2031 [notre traduction].

56. Les trajectoires verticales montrent les mêmes effets que ceux observés sur l'axe horizontal mais ils sont moins marqués. Les auteurs se sont donc plus intéressés aux mouvements horizontaux.

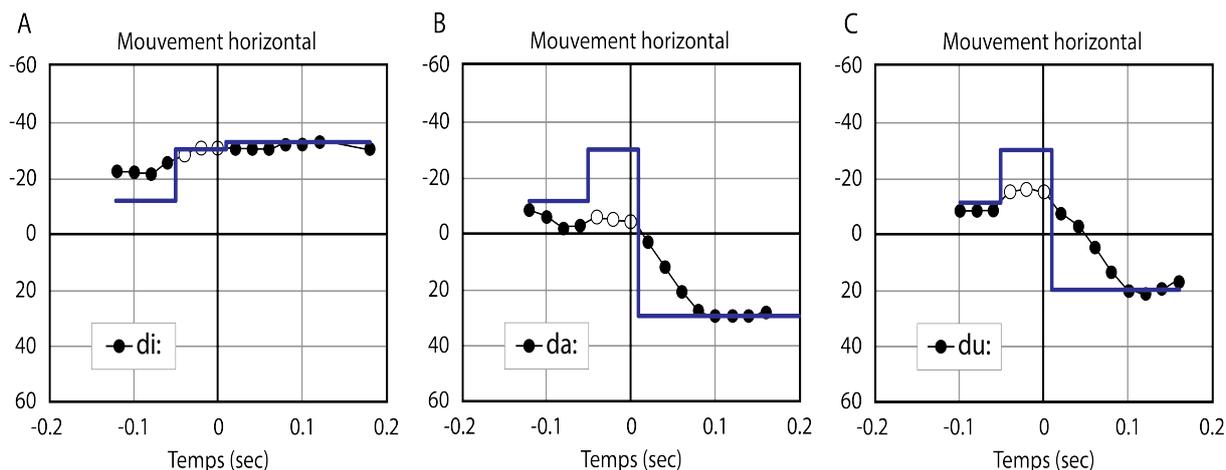


FIGURE 3.29 – Trajectoires horizontales de la langue prélevées toutes les 20 ms lors de la réalisation des séquences /ɛ'di:/, /ɛ'dɑ:/ et /ɛ'du:/. Le mouvement horizontal de la langue est représenté par des cercles noirs pour les voyelles et par des cercles blancs pour la consonne. Les lignes horizontales (bleues) correspondent à la position hypothétique de la langue durant /ɛ/, /d/ et V (/i:, ɑ:, u:/). La valeur zéro sur l'axe des ordonnées sépare les mouvements antérieur (valeurs négatives) vs. postérieur (valeurs positives) de la langue (Lindblom et Sussman, 2012).

entre les positions hypothétiques de la consonne et de la voyelle suivante est importante, plus la déviation est conséquente. Lindblom et Sussman (2012) soulignent que ce phénomène de déviations est en accord avec le modèle de l'*undershoot* ; les mouvements vers la cible de /d/ sont souvent non-atteints. Une autre information importante peut également être remarquée dans la Figure 3.29. Il s'agit de la stabilité temporelle de l'intervalle entre le relâchement de la consonne C et la cible de  $V_2$ . Quelque soit la distance traversée entre les deux cibles, l'intervalle de transition est de 100 ms environ. Les auteurs suggèrent ainsi que les régularités temporelles exposées dans les mouvements linguaux entre  $V_1CV_2$  peuvent être un facteur qui contribue à la linéarité des équations de locus. Des indices provenant des données acoustiques normalisées<sup>57</sup> soutiennent l'existence d'un constant temporel lors de la transition entre  $F2_{onset}$  et  $F2_{mid}$  d'une consonne produite avec plusieurs voyelles (Lindblom et al., 2009). La stabilité temporelle du mouvement de la langue observée dans les données articulatoires est donc reflétée dans les transitions formantiques de F2. Les auteurs déclarent que le temps constant uniforme des transitions de F2 a des implications importantes sur la forme de la représentation des équations de locus. Ils illustrent deux scénarios afin d'expliquer l'importance du constant temporel montré dans la Figure 3.30 :

- Le premier (à gauche) dévoile l'effet du constant temporel uniforme des transitions du F2 sur les équations de locus. Les constants temporels uniformes indiquent qu'après un temps  $t$ , toutes les transitions de F2 ont accomplies le même pourcentage de distance entre le locus et la cible vocalique, quelque soit la distance. En haut, toutes les transitions du F2 partent d'un point de fréquence fixe, similaire au locus virtuel de Delattre et al.

57. Ces données acoustiques ont été normalisées à l'aide d'une équation exponentielle (cf. Lindblom et al. (2009)).

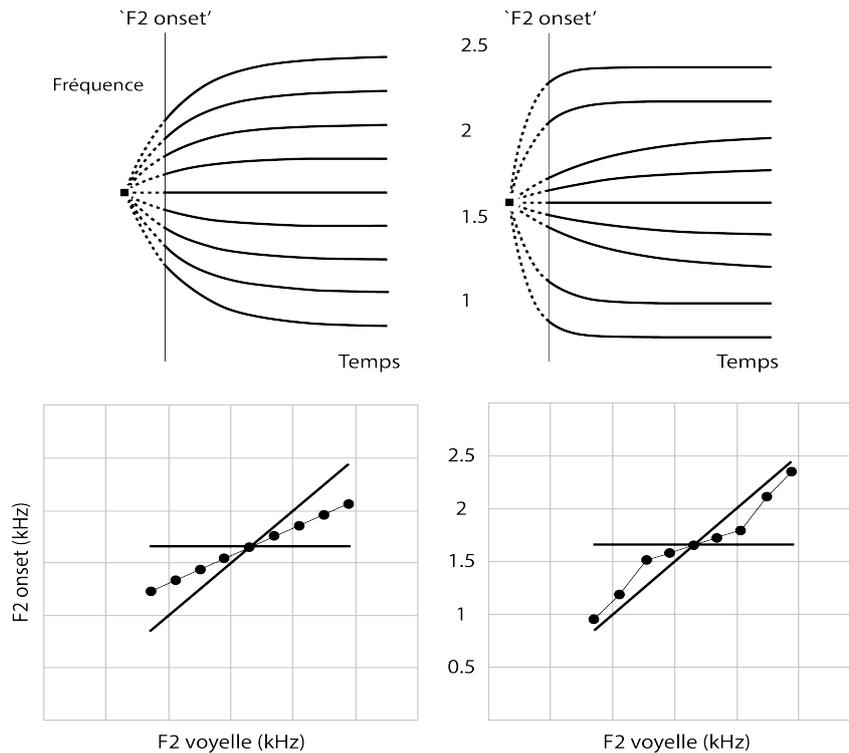


FIGURE 3.30 – Illustration de l’effet sur les équations de locus quand les transitions formantiques enfreignent le principe du constant temporel uniforme. En haut : à gauche, des transitions formantiques du F2 générées mathématiquement respectant le principe du temps constant uniforme partant d’un point commun ; à droite, des transitions formantiques générées d’une façon similaire mais avec des constantes temporelles variées. En bas, une représentation des effets du constant temporel uniforme (à gauche) et non-uniforme (à droite) sur les équations de locus (d’après Lindblom et Sussman (2012)).

(1955), et se déplacent vers une variété large de  $F2_{\text{voyelle}}$  cibles. Les transitions ont été générées mathématiquement et formées par un lissage exponentiel amorti déterminé par un constant temporel uniforme. Les  $F2_{\text{onset}}$  ont été choisis à 30 ms après le point de fréquence fixe. En bas, une représentation des fréquences de  $F2_{\text{onset}}$  en fonction de  $F2_{\text{mid}}$ . Les points de données s’ajustent pour former une régression linéaire située entre deux lignes représentant les deux régressions des équations de locus avec deux valeurs de pente extrêmes (0 et 1).

- Le second (à droite) signale l’impact du constant temporel non-uniforme sur les équations de locus. Les constantes temporelles non-uniformes montrent qu’il existe une variabilité des transitions en fonction de la distance parcourue entre le locus et la cible, peu importe le pourcentage. En haut, les transitions commencent également à partir d’un point fixe et se déplacent de ce point à la cible vocalique. Toutefois, le lissage exponentiel amorti des trajectoires de F2 est calculé selon un constant temporel varié ou non-uniforme. Les mouvements vers les cibles avec des fréquences élevées et basses sont plus rapides que celles centrales. Les points des données sont aussi localisés entre les deux lignes avec des valeurs des pentes maximale et minimale mais ne forment pas

une série linéaire.

Cette démonstration a permis d'exprimer l'importance de la stabilité du constant temporel dans la contribution de la linéarité des équations de locus. Les auteurs résument que « *Such a uniform temporal property has the consequence of making all CVs equally long and all CV velocities proportional to the distance to the V target. The constancy of the transition shapes implies that they are scaled linearly by variations in  $[F2_{onset} - F2_{vowel}]$  extent. As a result, for each stop place  $F2_{onset}$  vs.  $F2_{vowel}$  emerges as a linear scatterplot* »<sup>58</sup>.

Plusieurs recherches examinant les origines articulatoires des équations de locus, avec différents outils de mesure, ont observé l'existence d'une relation directe de la linéarité et de l'aspect systématique des équations de locus. Lorsque ces dernières sont calculées, elles risquent d'être perturbées par plusieurs facteurs liés à la position temporelle de la mesure acoustique de  $F2_{onset}$  et  $F2_{mid}$ , aux erreurs de mesure et au nombre de voyelles utilisées. La section suivante développe ces différents facteurs.

### 3.8 Facteurs influençant le calcul des équations de locus

Les équations de locus sont mesurées à partir des valeurs fréquentielles déterminées à deux moments du deuxième formant ( $F2_{onset}$  et  $F2_{mid}$ ). D'un côté,  $F2_{mid}$  est toujours prélevé à l'état d'équilibre de la voyelle, déterminé souvent au milieu temporel du F2. Le moment de la mesure n'est donc pas fixe de manière à prendre en compte les variations temporelles intrinsèques et extrinsèques des voyelles. Il est également à remarquer que Shammass (1985) a estimé  $F2_{mid}$  à un moment temporel fixe de 60 ms après le début des transitions et qu'il a obtenu des résultats similaires à ceux où  $F2_{mid}$  est déterminé au milieu du F2. D'un autre côté, la position de mesure de  $F2_{onset}$  est précisée traditionnellement à la première pulsation de la voyelle suivante. Pourtant, elle ne semble pas être adéquate pour les occlusives non voisées. Une position de mesure plus proche du bruit de relâchement est alors proposé (Sussman et Shore, 1996).

Par conséquent, le choix de la position de la mesure des coefficients des équations de locus ( $F2_{onset}$  et  $F2_{mid}$ ) peut influencer les résultats des pentes et des ordonnées à l'origine. D'autres facteurs peuvent également impacter les valeurs des équations de locus comme les erreurs de l'estimation des fréquences par le logiciel d'analyse acoustique utilisé ou le nombre de voyelles employées dans le calcul.

58. « Une telle propriété temporelle uniforme induit que toutes les séquences CV sont équitablement longues et que toutes les vitesses de CV sont proportionnelles à la distance de la cible V. La stabilité des formes des transitions implique leur organisation de façon linéaire en fonction de l'étendue de la variation dans  $[F2_{onset} - F2_{mid}]$ . Par conséquent, lorsque  $F2_{onset}$  est affiché contre  $F2_{mid}$ , une représentation linéaire émerge pour chaque lieu d'articulation des occlusives », Lindblom et Sussman 2012, page 18 [notre traduction].

### 3.8.1 Les mesures $F2_{onset}$

Le point de mesure de  $F2_{onset}$  a été déterminé au début de la transition du deuxième formant dans la littérature. Cette position de mesure est considérée comme la partie formantique contenant le maximum d'informations sur la consonne car elle est le point le plus proche temporellement de la consonne. La distance temporelle entre le début des transitions formantiques et la consonne devient plus importante avec les occlusives non voisées. De ce fait, les valeurs de  $F2_{onset}$  attestent une grande variation en fonction du voisement pour un même lieu d'articulation. Par exemple, Fant (1973) a observé des différences importantes de  $F2_{onset}$  entre  $/bV/ > /pV/$  et  $/dV/ > /tV/$  et des différences faibles entre  $/gV/$  et  $/kV/$  en suédois. Les équations de locus sont alors modifiées pour un même lieu d'articulation en fonction du voisement de la consonne. Molis et al. (1994) sont parmi les premiers à examiner l'influence du voisement sur la stabilité des équations de locus lorsqu'elles sont utilisées comme descripteurs du lieu d'articulation. Les auteurs comparent les pentes des équations de locus du  $/b/$  vs.  $/p/$  en français, en anglais américain et en suédois. Trois locuteurs (un par langue) ont prononcé des séquences  $/bV/$  et  $/pV/$  où  $V = /i, a, u/$ . Les valeurs de  $F2_{onset}$  ont été prélevées à un moment près du bruit de relâchement. Les résultats montrent que les pentes des  $/p/$  en anglais et en suédois sont plus basses que celles des  $/b/$  en anglais et en suédois (respectivement 0.49 et 0.25 pour  $/p/$  vs. 0.77 et 0.81 pour  $/b/$ ). Les bilabiales françaises, elles, ont des pentes similaires (0.89 pour  $/p/$  vs. 0.91 pour  $/b/$ ). Les auteurs expliquent que la grande différence de pente en anglais et en suédois est la conséquence du phénomène de l'aspiration qui a lieu lors de la production des bilabiales dans ces langues. Autrement dit, l'intervalle qui sépare la consonne et la voyelle est élargi à cause de l'aspiration, menant à des valeurs de  $F2_{onset}$  erronées. Au contraire, les bilabiales françaises ne sont pas aspirées.

Sussman et Shore (1996) ont également analysé l'impact du voisement sur les équations de locus pour les alvéolaires  $/d, t/$  (exploité dans la Section 3.4.1.4). Les valeurs de  $F2_{onset}$  ont été déterminées à deux positions différentes pour le  $/d/$  : à la première pulsation de la voyelle suivante et au moment proche du bruit de relâchement, appelé *@burst*. Les valeurs de  $F2_{onset}$  pour les séquences de  $/t/$  ont été prélevées *@burst*. Les résultats des pentes sont similaires pour les deux alvéolaires quand les valeurs de  $F2_{onset}$  sont extraites *@burst* (0.26 pour  $/d/$  et 0.23 pour  $/t/$ ). Les pentes de  $/d/$  mesurées avec  $F2_{onset}$  à la première pulsation de la voyelle affichent des valeurs significativement plus élevées (0.40). Les auteurs ont souligné que l'estimation des équations de locus des non voisées est plus précise avec la mesure de  $F2_{onset}$  *@burst*.

Engstrand et Lindblom (1997) étudient l'influence de l'aspiration sur les pentes des équations de locus en suédois. Trois locuteurs ont prononcé des séquences contenant des occlusives voisées et non voisées  $/b, d, g, p, t, k/$  suivies des voyelles  $/i:, \emptyset:, a, \alpha:, u:/$ . Deux points de mesure de  $F2_{onset}$  sont pris en compte dans le calcul des équations de locus : un point proche du bruit de relâchement et un autre à l'*onset* de la transition du F2. Les mots contenant les séquences CV ont été choisis de manière à faire varier la durée de VOT (position des CV dans le

mot, longueur du mot et niveau d'accent). La durée de VOT a été mesurée afin de tester l'existence d'une corrélation entre l'intervalle de l'aspiration et les pentes. Les résultats révèlent que les valeurs des pentes calculées avec  $F2_{@burst}$  et  $F2_{mid}$  baissent lorsque la durée de VOT augmente tandis que les valeurs des pentes estimées avec  $F2_{onset}$  et  $F2_{mid}$  décroissent quand la durée de VOT croît.

Modarresi et al. (2005) testent l'influence de l'aspiration sur les équations de locus dans les trois occlusives (bilabiale, alvéolaire et vélaire) et l'importance de la position de mesure de  $F2_{onset}$  en anglais américain et en persan. Cinq locuteurs de l'anglais et deux de persan ont prononcé des non-mots contenant des syllabes CV où C = /b, d, g, p, t, k/ et V = /i, e, æ, u, ɔ/ pour l'anglais et V = /i, ε, æ, u, o, α/ pour le persan. Les valeurs de  $F2_{onset}$  des voisées ont été déterminées à deux moments distincts : à la première pulsation de la voyelle suivante ( $F2_{onset}$ ) et à la première résonance visible de F2 au moment le plus proche après le bruit de relâchement ( $F2_{@burst}$ )<sup>59</sup>. Les  $F2_{onset}$  des non voisées ont été prélevés uniquement à la première résonance de F2 après le bruit de relâchement ( $F2_{@burst}$ ) (cf. Figure 3.12 de Sussman et Shore (1996) pour une démonstration). Quant aux valeurs de  $F2_{mid}$ , elles ont été fixées au milieu de F2. Les pentes issues des données des deux langues ont été traitées dans un seul groupe puisqu'elles sont significativement similaires.

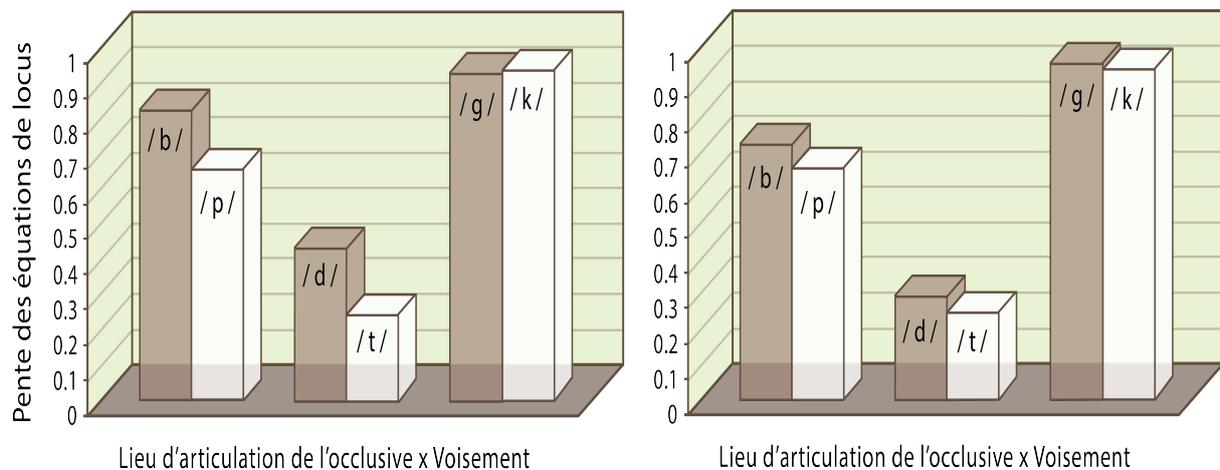


FIGURE 3.31 – Valeurs des pentes des équations de locus montrant les différences entre les positions de mesure de  $F2_{onset}$ . À gauche, les  $F2_{onset}$  sont déterminées à la première pulsation de la voyelle pour les voisées et au bruit de relâchement pour les non voisées (@burst). À droite, les  $F2_{onset}$  sont prélevées au bruit de relâchement pour les voisées et les non voisées (Modarresi et al. (2005)).

Les résultats indiquent que les pentes des voisées estimées avec  $F2_{onset}$  sont significativement plus élevées que celles des non voisées avec  $F2_{@burst}$  pour /b/ vs. /p/ et /d/ vs. /t/. Quant

59. Les auteurs ont employé les procédures adoptées par Sussman et Shore (1996) pour préciser la position de  $F2_{onset}/@burst$  présentés dans la section 3.4.1.4. Il est également à remarquer que la détermination de  $F2_{@burst}$  pour les voisées n'est pas toujours possible car les résonances de F2 peuvent de ne pas être visibles après le bruit de relâchement (62% pour les /b/ et 30% pour les /d/ des  $F2_{@burst}$  n'ont pas pu être identifiés). Dans ces cas, les valeurs traditionnelles de  $F2_{onset}$  ont été utilisées.

aux vélaires, elles ont des pentes significativement semblables (Figure 3.31). Lorsque les pentes sont calculées avec  $F2_{@burst}$  pour les voisées et les non voisées, elles sont statistiquement équivalentes pour les trois lieux d'articulation. Dans les deux catégories, un effet principal du lieu d'articulation est observé. Selon les résultats, quand  $F2_{onset}$  est mesuré le plus proche possible du bruit de relâchement de /b/ et de /d/ et par comparaison avec la position traditionnelle, les valeurs des pentes sont considérablement basses. Modarresi et al. (2005) expliquent que cette baisse résulte de la position de mesure de  $F2_{burst}$  qui a des fréquences plus stables que celle observées à  $F2_{onset}$ . Autrement dit, les valeurs de fréquence de  $F2_{@burst}$  des occlusives sont moins impactées par les voyelles suivantes que celle de  $F2_{onset}$ . Les auteurs soulignent que l'absence de différence significative entre les pentes des vélaires en fonction du voisement dans les deux positions de mesure est liée à la difficulté du maintien du voisement lors de la production des occlusives vélaires. L'occlusion postérieure des vélaires est accompagnée par une augmentation du niveau de la pression de l'air intra-oral et donc les différences de pression de l'air trans-glottal sont réduites pour soutenir le voisement dans les séquences /gV/.

Comme il a déjà été précisé, les équations de locus exposent la relation entre une consonne et des voyelles suivantes. Les valeurs de  $F2_{onset}$  sont considérées comme la partie qui représente la consonne. Les études exposées dans cette section mettent en avant que la position temporelle de mesure de  $F2_{onset}$  affecte largement les valeurs des pentes résultantes. Celles-ci sont plus représentatives de la consonne quand elles sont mesurées au moment du relâchement de la consonne, notamment pour les occlusives non voisées. Les valeurs de pente des voisées et des non voisées baissent quand le point de prélèvement de  $F2_{onset}$  s'approche du bruit de relâchement. Ce phénomène indique que la consonne est moins influencée par la voyelle.

Quant aux valeurs de  $F2_{mid}$ , elles sont considérées comme la partie qui représente la voyelle. Elles sont souvent fixées au milieu de F2. Une question peut être posée sur les  $F2_{mid}$  : un changement de position d'extraction de  $F2_{mid}$  modifiera-t-il les valeurs des équations de locus ? Cette question sera examinée dans la section suivante.

### 3.8.2 Les mesures de $F2_{mid}$

Rhone et Jongman (2012) expliquent que la majorité des études sur les équations de locus ont utilisé une position de  $F2_{mid}$  fixée au milieu de la voyelle suivant la consonne cible. Cette position peut varier approximativement entre 60 et 110 ms après le début des transitions formantiques en anglais. Pourtant, des études de perception ont montré que les auditeurs identifient les lieux d'articulation des occlusives avec une période du signal beaucoup plus courte (20 ms) (Tekieli et Cullinan, 1979). Autrement dit, les auditeurs n'ont pas besoin d'attendre le milieu de la voyelle pour percevoir la consonne. Les auteurs soulignent que si les équations de locus sont un descripteur fiable du lieu d'articulation, une position temporelle de  $F2_{mid}$  plus précoce que celle traditionnelle doit différencier les lieux d'articulation des occlusives c'est-à-dire des pentes et des ordonnées à l'origine distinctes en fonction du lieu d'articulation. Vingt

locuteurs de l'anglais américain ont prononcé des séquences CVt où C = /b, d, g/ et V = /i, I, ei, ε, æ, α, ʌ, ɔ, ou, u/. Les valeurs de  $F2_{onset}$  ont été déterminées à la première pulsation de la voyelle et les valeurs de  $F2_{mid}$  ont été prélevées à 9 (T1), 18 (T2) et 27 (T3) ms de  $F2_{onset}$  et au milieu de la voyelle pour chaque séquence. Quatre équations de locus ont été calculées pour chaque catégorie consonantique selon la position temporelle de  $F2_{mid}$ . Les résultats des équations de locus estimées avec  $F2_{mid}$  au milieu de F2 révèlent que les pentes et les ordonnées à l'origine des occlusives sont significativement différentes. Les analyses par paires sur les pentes dévoilent des différences significatives entre /b/ vs. /d/ et entre /d/ vs. /g/ mais pas entre /b/ vs. /g/. Quant aux ordonnées à l'origine, elles sont toutes statistiquement distinctes. Des analyses linéaires discriminantes (LDA) ont été effectuées, avec les pentes et les ordonnées à l'origine comme prédicteurs. Le taux de classification correcte est de 87%. Lorsque les  $F2_{onset}$  et  $F2_{mid}$  (au milieu) sont utilisés comme prédicteurs, un taux de classification correcte de 71% est noté. Les résultats des équations de locus estimées avec les trois autres positions

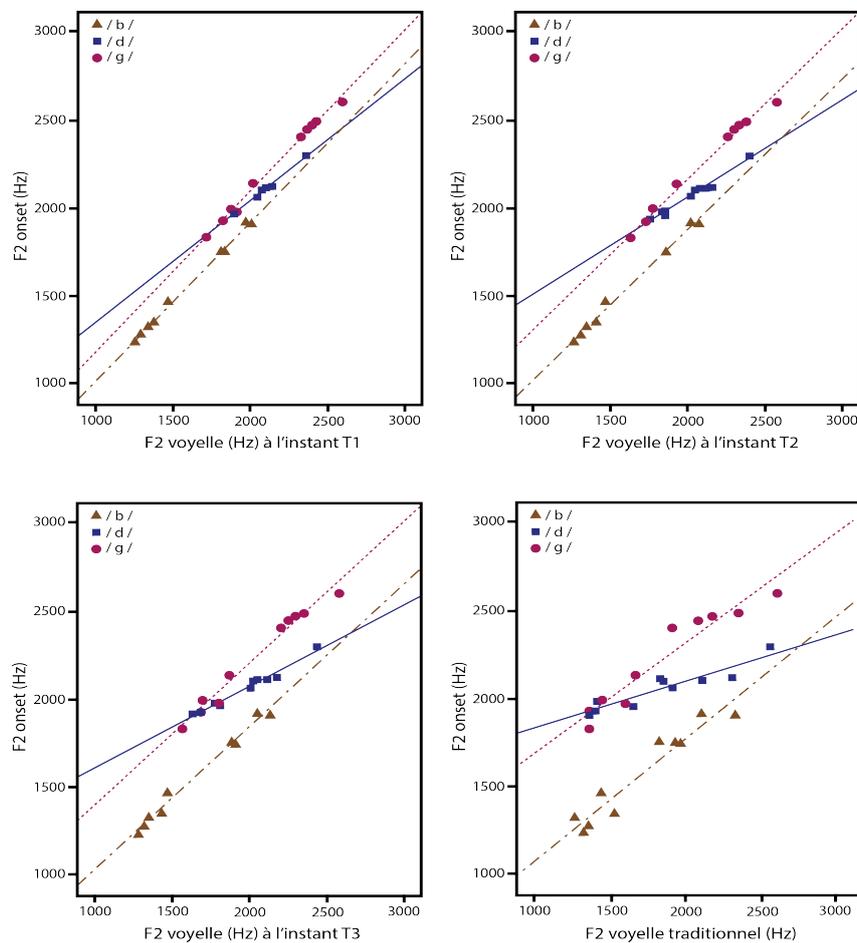


FIGURE 3.32 – Illustration des moyennes des quatre équations de locus calculées selon la position temporelle de  $F2_{mid}$  : respectivement de gauche à droite, à 9, 18 et 27 ms après  $F2_{onset}$  et au milieu de F2 (d'après Rhone et Jongman (2012)).

temporelles de  $F2_{mid}$  T1, T2 et T3 suivent les mêmes tendances que ceux avec le  $F2_{mid}$  traditionnel (/b/ > /g/ > /d/). Les valeurs de pente augmentent quand la distance temporelle entre

$F2_{onset}$  et  $F2_{mid}$  se réduit (Figure 3.32). Ce constat est la conséquence de l'influence importante de la consonne et du chevauchement coarticulatoire qui est élevée entre les zones réduites et qui est traduite par une similarité de fréquences entre  $F2_{onset}$  et  $F2_{mid}$  et pentes des pentes élevées. Les pentes ainsi que les ordonnées à l'origine des trois consonnes sont significativement différentes dans chacune des équations de locus des trois positions temporelles de  $F2_{mid}$ . En adéquation avec les résultats de  $F2_{mid}$  traditionnel, les analyses de comparaison par paires sur les pentes affichent des différences significatives entre /b/ vs. /d/ et entre /d/ vs. /g/ mais pas entre /b/ vs. /g/. Une différence significative est remarquée pour toutes les combinaisons des ordonnées à l'origine. Les LDA sur chacune des trois équations de locus montrent également des résultats similaires avec celle traditionnelle. Quand les pentes et les ordonnées à l'origine sont utilisées comme prédicteurs, les taux de classification correcte sont de 81%, 83% et 83% respectivement pour les équations de locus avec T1, T2 et T3. Les LDA avec  $F2_{onset}$  et  $F2_{mid}$  (avec T1, T2 et T3) comme prédicteurs présentent des taux de classification correcte de 76%, 76% et 75% respectivement. Une comparaison statistique entre les équations de locus estimées avec  $F2_{mid}$  au milieu et celles avec chacune des trois positions temporelles indique des différences significatives de pentes et d'ordonnées à l'origine. Selon Rhone et Jongman (2012), les résultats de cette étude mettent en avant que les équations de locus dérivées avec des points de mesures qui représentent des durées relatives à celles de la perception des auditeurs peuvent être utilisées pour catégoriser les lieux d'articulation des occlusives. Les premières 10 ms de la transition vocalique offrent des informations suffisantes pour les auditeurs afin d'identifier les occlusives. La méthode de l'estimation des équations de locus modifiées, présentées ci-dessus, signale également un taux de classification correcte élevé avec des durées comparables.

Cette étude dévoile qu'un changement de position de mesure de  $F2_{mid}$  conduit à des variations des équations de locus. Toutefois, malgré ces variations, les équations de locus classifient toujours les lieux d'articulation des occlusives. D'autres facteurs, autres que les positions de mesure de  $F2_{onset}$  et  $F2_{mid}$ , peuvent également influencer les valeurs des pentes et des ordonnées à l'origine tels que les erreurs de mesures issues du logiciel ou le nombre réduit des voyelles utilisées dans le calcul.

### 3.8.3 Les erreurs de mesures

Les fréquences des  $F2_{onset}$  et  $F2_{mid}$  utilisées dans le calcul des équations de locus peuvent subir des erreurs de détection induites par le logiciel d'analyse phonétique employé comme Praat (Boersma et Weenink, 2017) ou WaveSurfer (Sjolander et Beskow, 2013). Ces erreurs peuvent être provoquées par la mauvaise qualité de l'enregistrement, par des fréquences fondamentales élevées ou par d'autres raisons. Montgomery et al. (2014) expliquent que lors de la mesure des valeurs fréquentielles des formants, il existe une quantité - souvent inconnue - d'erreurs de détection des fréquences qui varient selon la méthode d'analyses, le logiciel, etc. Les auteurs examinent alors l'étendue de l'influence de ces erreurs sur les pentes des équations

de locus. Ils testent également l'impact de la réduction du nombre des voyelles utilisées dans le calcul des équations de locus en le variant systématiquement de dix jusqu'à trois voyelles. Dans un premier temps, 75 équations de locus ont été estimées pour les consonnes /b, d, g/ produites avec dix voyelles. Les valeurs de  $F2_{onset}$  et de  $F2_{mid}$  (= 10 + 10 valeurs pour chaque équation de locus) proviennent des données recueillies dans la littérature. Dans un deuxième temps, 2000 équations de locus ont été simulées pour chacune des 75 équations de locus d'origine à l'aide de la méthode de Monte-Carlo<sup>60</sup>. Dans un troisième temps, les valeurs de  $F2_{onset}$  et  $F2_{mid}$  des 2000 équations de locus ont été perturbées avec 50, 100 et 200 Hz d'erreurs aléatoires. Les trois niveaux d'erreur aléatoire ont été choisis pour représenter les quantités des erreurs réduite, moyenne et extrême. Dans un quatrième temps, le nombre de voyelles employées dans le calcul des équations de locus a été réduit avec 10, 8, 6, 4 et 3 voyelles pour les 2000 équations de locus<sup>61</sup>. Les résultats signalent une stabilité des valeurs de pente face à une quantité réduite

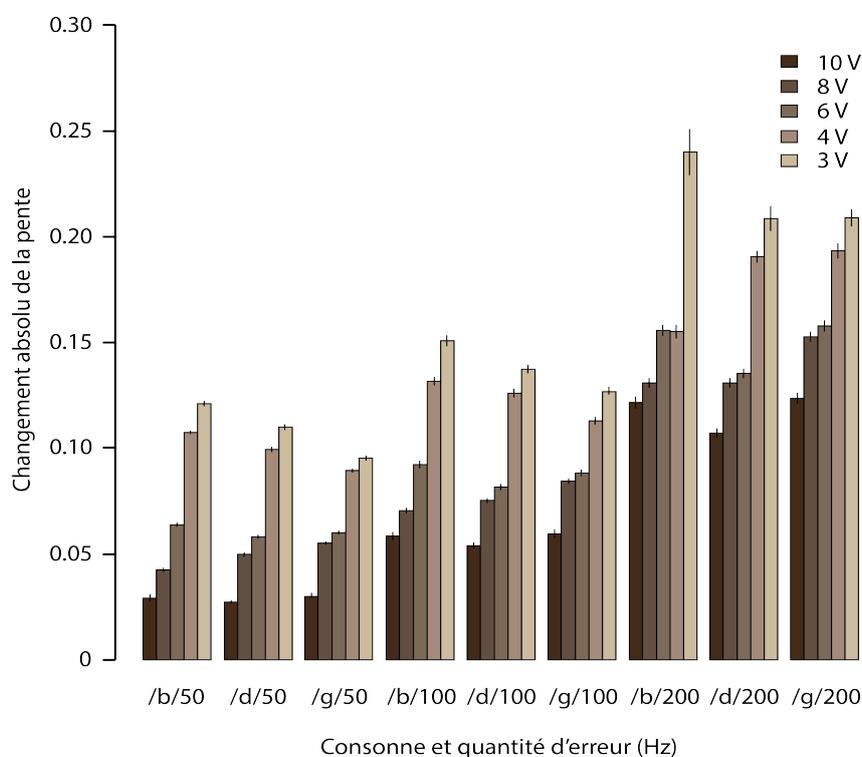


FIGURE 3.33 – Moyennes des changements des pentes des équations de locus divisées par le nombre de voyelles utilisées et par le niveau d'erreur aléatoire de  $F2_{onset}$  et  $F2_{mid}$ . Les changements des pentes sont calculés par la soustraction de chaque pente dans la condition de sa pente d'origine (Montgomery et al. (2014)).

d'erreurs (50 Hz) avec les équations de locus de dix, huit et six voyelles. En revanche, l'ajout des erreurs moyennes (100 Hz) ou élevées (200 Hz) conduit à des changements importants comparé

60. La méthode de Monte-Carlo désigne des techniques probabilistes visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires. À partir d'un petit échantillon, un nombre important d'équations de locus peut être estimé (Noiray et al., 2013).

61. Les groupes de voyelles sont : 8 V = /i, e, ε, æ, a, ɔ, o, u/, 6 V = /i, e, æ, a, o, u/, 4 V = /i, æ, a, u/ et 3 V = /i, a, u/.

aux pentes d'origine. Par exemple, dans le groupe avec huit voyelles, l'ajout de 50 Hz d'erreur aléatoire produit des moyennes de la différence entre les pentes d'origine et celles simulées de 0.043, 0.050 et 0.055 pour /b/, /d/ et /g/ respectivement<sup>62</sup> ; toutefois, 100 Hz d'erreur a provoqué 0.071, 0.75 et 0.84 de différence. L'utilisation de trois ou quatre voyelles avec 50 Hz (ou plus) d'erreur conduit à des variations importantes de pentes. Ces dernières sont traduites par de grandes différences de pente et des valeurs d'écart-types élevées en comparaison des autres groupes de voyelles (Figure 3.33).

Cette étude a permis d'explorer l'importance du nombre de voyelles utilisées lors du calcul des équations de locus et de la quantité d'erreur de mesure que celles-ci tolèrent. Quand les équations de locus sont estimées avec un nombre réduit de voyelles, il faut que le taux d'erreur soit inférieur à 50 Hz. Montgomery et al. (2014) précisent qu'il est nécessaire d'être prudent lors de l'utilisation de trois ou quatre voyelles pour mesurer les équations de locus. Toutefois, les auteurs n'ont pas testé si l'emploi de trois ou de quatre voyelles avec un certain nombre de répétitions pourrait conduire à des pentes plus précises. Une solution de ce problème sera présentée dans la partie expérimentale.

### 3.9 Conclusion

Ce chapitre décrit l'utilisation des équations de locus comme un outil de mesure de coarticulation et comme un descripteur du lieu d'articulation. Il présente les différents aspects théoriques et expérimentaux employés pour leur estimation. L'apparition des équations de locus a offert la possibilité de quantifier numériquement l'étendue globale de la coarticulation, évaluée pour une consonne occlusive suivie de plusieurs voyelles. Cette mesure de coarticulation est réalisée pour chaque lieu d'articulation malgré le fait que les voyelles exercent des quantités non-uniformes de coarticulation sur la consonne (Modarresi et al., 2004a). Autrement dit, chaque voyelle a des impacts différents sur la consonne précédente. De ce fait, il a été proposé que les équations de locus fournissent des informations importantes sur le lieu d'articulation et elles sont présentées comme un invariant relationnel du lieu d'articulation. Elles donnent alors une description du lieu d'articulation grâce à la relation entre deux valeurs fréquentielles situées au deuxième formant ( $F2_{onset}$  et  $F2_{mid}$ ).

Les études sur les équations de locus ont montré que les bilabiales sont toujours caractérisées par des pentes plus élevées que les alvéolaires. Cela implique l'existence d'une influence coarticulatoire plus importante de la voyelle suivante sur les fréquences de  $F2_{onset}$  de /b/ comparé à /d/. Les alvéolaires résistent fortement aux effets coarticulatoires de la voyelle suivante. Les vélaires, quant à elles, sont désignées par deux pentes distinctes. Une vélaire produite avec les voyelles antérieures possède une pente basse indiquant ainsi un effet minimal de la coarticulation CV alors qu'une vélaire prononcée avec les voyelles postérieures a des pentes élevées

62. Les moyennes des différences des pentes sont estimées par la soustraction de chaque pente simulée avec sa pente d'origine.

soulignant un impact maximal. Lorsque les équations de locus de /g/ sont calculées avec les voyelles postérieures et antérieures, les valeurs des pentes peuvent être supérieures ou inférieures à celles des bilabiales selon la langue, le dialecte ou l'échantillon. Toutefois, elles sont toujours plus élevées que les valeurs des pentes des alvéolaires. Le calcul des équations de locus sur d'autres consonnes révèlent que les segments qui ont des lieux d'articulation proches ont des pentes similaires. Il a également été observé que les équations de locus décrivent les occlusives mieux que les autres modes d'articulation.

L'utilisation des équations de locus comme un moyen de mesure du degré de coarticulation a été validée dans la plupart des études. En revanche, leur emploi comme un descripteur du lieu d'articulation a ouvert un grand débat. Ce dernier porte en particulier sur le degré de la participation des équations de locus à décrire les lieux d'articulation et sur l'origine de leurs régularité et linéarité. Dans le premier cas, Sussman et ses collègues indiquent que les paramètres des équations de locus contiennent la majorité de l'information pour décrire les lieux d'articulation des occlusives ; d'autres informations sont importantes en cas d'ambiguïté, comme le troisième formant, le bruit de relâchement et le VOT (Sussman et al., 1998b). D'autres chercheurs, comme Fowler et ses collègues, sont d'un avis différent. Selon leurs études, les équations de locus ne représentent qu'un indice parmi d'autres pour préciser le lieu d'articulation. Elles ne peuvent donc pas à elles seules décrire sans ambiguïté les lieux d'articulation (Fowler, 1994; Brancazio et Fowler, 1998; Fowler et Brancazio, 2000; Iskarous et al., 2010).

Dans le second cas sur l'origine de la linéarité et de la régularité des équations de locus, Sussman et ses collègues argumentent que la linéarité et l'aspect systématique des équations de locus sont des contraintes perceptuelles. Les locuteurs organisent leurs articulateurs de façon à ce qu'ils respectent les coefficients des équations de locus de chaque occlusive. Fowler et ses collègues expliquent que cette régularité est le résultat des contraintes articulaires des occlusives. La production d'un segment exige certaines contraintes qui varient en fonction du lieu et du mode d'articulation.

Les origines articulaires des équations de locus ont été examinées de manière à avoir une meilleure compréhension de leur comportement régulier et linéaire. Des études ont dévoilé qu'il existe des liens directs entre la position horizontale de la langue et les équations de locus. Une régression linéaire similaire à celle des équations de locus est obtenue lorsque les positions de la langue sont combinées lors de la réalisation de la consonne et de la voyelle (Iskarous et al., 2010; Noiray et al., 2013). De plus, la linéarité des équations de locus est expliquée par la régularité temporelle du mouvement horizontal de la langue (Lindblom et Sussman, 2012).

Les équations de locus ont montré leur capacité à capter l'augmentation du chevauchement coarticulaire provoquée par plusieurs phénomènes comme le style de parole, l'accent emphatique et le débit de parole. Cette hausse est étroitement liée à la réduction temporelle de la réalisation de la séquence. Selon le modèle de l'*undershoot*, lorsque le chevauchement coarticulaire est élevé, les valeurs formantiques de la voyelle et de la consonne sont modifiées ; par conséquent, des informations spectrales sont perdues. Sussman et al. (1998b) déclarent que

les variations spectrales induites par les changements temporels affectent peu les paramètres des équations de locus. Ces derniers séparent les catégories consonantiques malgré ses variations. Les prochains chapitres seront alors consacrés à explorer la relation entre la temporalité et les équations de locus et à la question de l'impact du degré des variations temporelles sur les équations de locus et sur la catégorisation des occlusives.

## **Deuxième partie**

### **Problématiques et aspects expérimentaux**



# Chapitre 4

## Problématiques

L'un des objectifs principaux de l'étude des sons de la parole est d'identifier des paramètres qui forment leur identité et permettent leur perception. Les caractéristiques les plus pertinentes dans la perception des occlusives et des voyelles ont été mises en exergue au cours de la première partie de ce manuscrit (Chapitre I). Dans les recherches antérieures, ces caractéristiques ont été traitées initialement sans prendre en compte l'impact des segments voisins. Or, ces dernières semblent être influencées systématiquement par le contexte adjacent dans la parole naturelle lors de la coarticulation des différents sons (Chapitre II). La compréhension de la relation co-articulaire dynamique entre les occlusives et l'environnement vocalique a ouvert la voie vers l'élaboration des modèles et des théories qui peuvent mesurer et prédire le degré de l'influence d'un segment sur un autre. Par conséquent, certaines caractéristiques acoustiques des occlusives peuvent être décrites de manière dynamique grâce à des outils proposés par les modèles de coarticulation comme les équations de locus (Chapitre III). Celles-ci ont montré une capacité élevée à offrir un degré de coarticulation entre les occlusives et les différents contextes vocaliques. Elles ont également dévoilé une bonne fiabilité, souvent polémique, dans la distinction des lieux d'articulation.

Deux aspects généraux ont été observés lors de l'examen des études exploitées dans la littérature sur la coarticulation et sur la classification des occlusives. Tout d'abord, ces études se concentrent principalement sur l'analyse de trois lieux d'articulation et les cas où le nombre de lieux d'articulation est supérieur à trois sont très peu développés dans la littérature. Ensuite, elles ne mettent pas en avant l'importance de l'information des variations temporelles (intrinsèque et extrinsèque) des voyelles. De même, l'influence de la temporalité sur la classification des occlusives a rarement été traitée. C'est pourquoi ces thématiques encore peu abordées nous ont incité à concevoir ce travail sur l'impact des variations temporelles sur les transitions formantiques en arabe jordanien et en français.

La présente recherche est ainsi motivée par deux problématiques principales. La première porte sur l'analyse de l'efficacité des paramètres des équations de locus à désigner le degré de coarticulation et le lieu d'articulation dans un système consonantique plus riche que ce qui a été décrit dans la littérature. La seconde étudie l'impact des différentes variations temporelles

sur la coarticulation et sur la classification des lieux d'articulation à l'aide des équations de locus. Parallèlement, d'autres questions peu développées dans la littérature telles que l'impact de l'opposition de la position syllabique (initiale *vs.* finale) et du voisement (occlusives voisées *vs.* non voisées) sur les paramètres des équations de locus seront également traitées. Le contexte de ces problématiques sera présenté dans ce chapitre.

## 4.1 Un espace consonantique riche

Il a été montré dans les chapitres précédents que les transitions formantiques contiennent des informations importantes pour la perception des occlusives. Les fréquences au début et à la fin des transitions du deuxième formant permettent une mesure fiable du degré de coarticulation entre une occlusive et plusieurs voyelles grâce à la pente de régression linéaire (équation de locus) calculée à partir de ces paramètres ( $F2_{onset}$ ,  $F2_{mid}$ ). Ces derniers catégorisent les lieux d'articulation avec un taux autour de 70 %. Puisqu'il existe un degré de coarticulation différent pour chaque lieu d'articulation, l'utilisation de la pente et de l'ordonnée à l'origine des équations de locus conduit à une classification parfaite (Sussman et al., 1991). En effet, la plupart des études menées, dévoilées dans la littérature et résumées dans ce manuscrit, se sont concentrées sur trois lieux d'articulation des occlusives : bilabial, alvéolaire et vélaire. Ces trois lieux d'articulation existent dans un grand nombre de langues (Ladefoged, 2005), cependant, certaines langues possèdent un nombre plus important de lieux d'articulation des occlusives.

La première problématique de notre étude consiste alors à examiner si cette séparabilité des lieux d'articulation reste intacte si le nombre de ceux-ci augmente. Autrement dit, les paramètres des équations de locus seraient-ils aussi performants pour classer des occlusives dans une langue qui contient quatre ou cinq lieux d'articulation (comme l'arabe) que dans celles décrites dans les recherches citées précédemment (comme l'anglais). Dans le premier cas, la distribution des points d'articulation s'étend sur un espace plus large dans le conduit vocal pouvant entraîner des chevauchements acoustiques et/ou des degrés de coarticulation similaires entre les catégories consonantiques. En effet, un intérêt particulier a été porté à l'arabe jordanien dans notre étude puisqu'il a un système consonantique riche contenant cinq lieux d'articulation des occlusives : bilabial, alvéolaire, vélaire, uvulaire et glottal dont cinq occlusives non voisées (/t, t<sup>h</sup>, k, q, ʔ/) et quatre voisées (/b, d, d<sup>h</sup>, g/). Cette richesse consonantique est opposée à la pauvreté du système vocalique, l'arabe jordanien ne comprenant que cinq timbres vocaliques avec trois voyelles courtes et cinq longues (/i, a, u, iː, aː, uː, eː, oː/). Celui-ci représente alors un cas intéressant car il montre une situation inverse des langues examinées dans la littérature où le nombre de voyelles est plus important (entre 9 et 11) et où celui des lieux d'articulation est limité à trois. La section suivante propose quelques caractéristiques acoustiques des sons de l'arabe jordanien.

---

1. D'après l'UNESCO.

### 4.1.1 Caractéristiques acoustiques de l'arabe jordanien

L'arabe est la langue officielle dans 25 pays (dont les 22 pays membres de la ligue arabe et l'Israël, le Tchad et l'Érythrée) et est parlé par plus de 400 millions de locuteurs natifs<sup>1</sup>. Deux variétés de la langue parlée aujourd'hui coexistent. Il s'agit de l'arabe classique standard appelé *فُصْحَى* /fusʰħa:/ et l'arabe dialectal ou vernaculaire dit *عَامِيَة* /ʕa:mi:ya/. La langue standard est utilisée dans la parole écrite et dans les conversations officielles et religieuses alors que la dialectale est employée dans toute autre situation (Versteegh, 2014). L'arabe dialectal est ainsi la langue maternelle des locuteurs de l'arabe et la langue standard est apprise à l'école. Il est également possible d'ajouter à ces deux variétés de l'arabe une variété intermédiaire, nommée variété dialectale médiane, qui est constituée d'un mélange de l'arabe standard classique et de l'arabe dialectal (Versteegh, 2014). Cette variété intermédiaire est parlée dans les milieux politique, universitaire et intellectuel.

Concernant les dialectes arabes, ils peuvent être divisés en fonction de la zone géographique d'où ils proviennent, comme l'a proposé Versteegh (2014) : 1) la péninsule arabique, 2) la Mésopotamie, 3) le Levant, 4) l'Égypte et 5) le Maghreb. Il est également possible de les classer par catégorie socio-culturelle comme dialectes 1) citadins, 2) villageois et 3) bédouins. L'arabe jordanien fait partie des dialectes levantins et est parlé par 9.5 millions de locuteurs<sup>2</sup>. Il contient les trois catégories socio-culturelles. Chaque région de la Jordanie se caractérise par un parler spécifique marqué par un accent particulier ou des expressions propres. Tous ces parlers se rapprochent plus ou moins les uns des autres sans qu'il n'y ait un blocage de communication. L'arabe jordanien, comme les autres dialectes arabes, comprend une version modifiée des inventaires consonantique et vocalique.

#### 4.1.1.1 Le système vocalique de l'arabe jordanien

Le système vocalique de l'arabe jordanien est composé des six voyelles de l'arabe classique /i, a, u, i:, a:, u:/ (correspondant à trois timbres vocaliques s'opposant phonémiquement en durée) et des deux voyelles longues /e:/ et /o:/ qui étaient les deux anciennes diphtongues de l'arabe classique /aj/ et /aw/<sup>3</sup> (entre autres, Barkat (2000); Al-Tamimi (2007); Embarki (2008); Nawafleh (2012)). Les voyelles en arabe ne sont jamais produites en isolation ce qui provoque

2. Selon le rapport de "Population and Social Statistics Directorate of the Hashemite Kingdom of Jordan (2015)".

3. cf. Embarki (2008) pour une revue historique sur ces deux diphtongues.

une difficulté aux locuteurs pour les prononcer en isolation (Al-Tamimi, 2007; Nawafleh, 2012).

Les voyelles longues se différencient principalement de leurs contreparties courtes par leur durée physique. Par exemple, Barkat (2000) observe que la durée des voyelles /i:, a:, u:/, réalisées par deux locuteurs de l'arabe jordanien, est 2.1 fois plus grande que celle de leurs correspondantes courtes. Il existe également des différences de qualité qui distinguent les deux catégories de voyelles notamment entre /i:/ vs. /i/ et /u:/ vs. /u/.

Al-Tamimi (2007) examine les propriétés acoustiques des voyelles de l'arabe jordanien dans trois situations de production : en isolation, en syllabe et en mot. Les voyelles /i, a, u, i:, a:, u:, e:, o:/ ont été prononcées avec les occlusives /b, d, d<sup>ɬ</sup>, k/ par dix locuteurs jordaniens. Ces derniers devaient lire les parties colorées des mots commençant par une des occlusives suivie d'une voyelle cible. Les parties colorées correspondent à des mots entiers, à des syllabes CV ou à une voyelle V. Les résultats montrent que les voyelles ne sont pas influencées par la situation de production (Figure 4.1). L'auteur explique que cette observation est due à la difficulté de

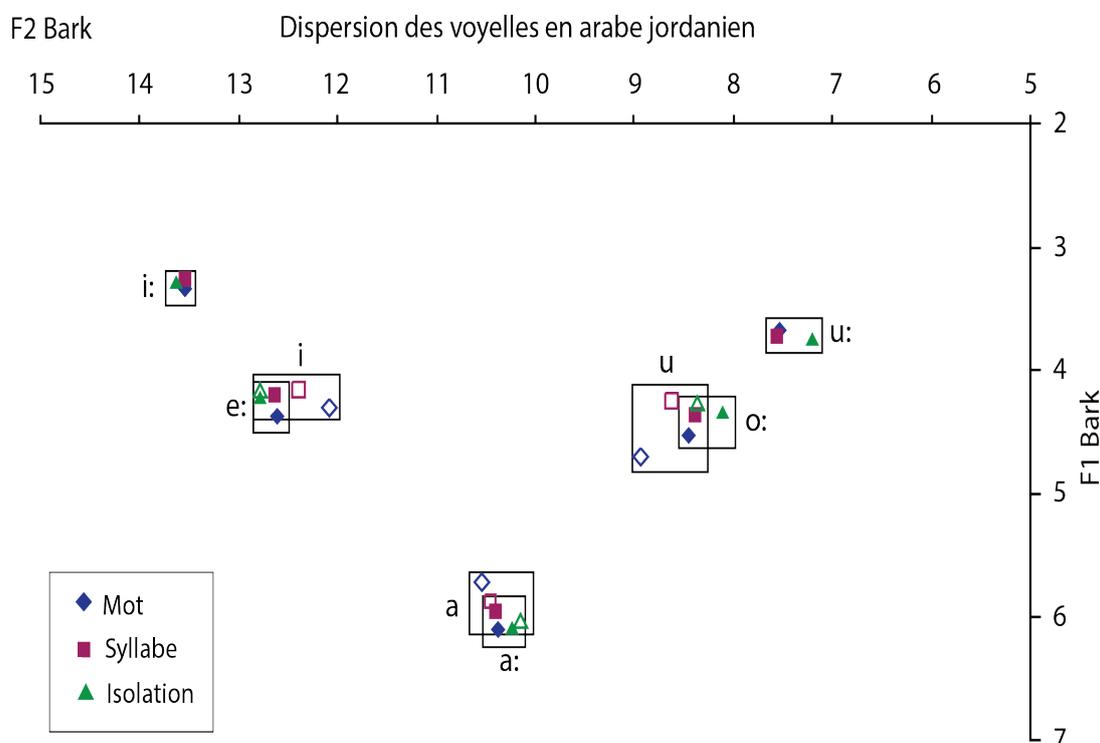


FIGURE 4.1 – Voyelles de l'arabe jordanien en fonction de la situation de production : en mot, en syllabe et en isolation (d'après Al-Tamimi (2007)).

la tâche pour les locuteurs jordaniens à prononcer les voyelles en isolation puisqu'« il leur a été difficile de produire les voyelles de leur système dans les deux situations de production Syllabe et Isolation, car c'était une tâche très artificielle pour eux »<sup>4</sup>. Toutefois, les voyelles sont impactées par les différents contextes consonantiques. L'examen de l'opposition entre voyelles longues vs. courtes révèle que les voyelles /i:/ et /u:/ sont significativement différentes de leurs contreparties courtes concernant le F1 et le F2. Le /i/ court est plus ouvert et plus postérieur

4. Page 236. Comme mentionné précédemment, les voyelles en arabe n'existe pas en isolation.

que son correspondant long ce qui conduit à un rapprochement entre l'espace de la voyelle /e:/ celui de /i/. Quant au /u/ court, il est plus ouvert et plus antérieur que sa contrepartie longue, se rapprochant alors de l'espace vocalique du /o:/ (Figure 4.1). Au contraire, les voyelles /a:/ et /a/ sont statistiquement similaires.

Les analyses discriminantes avec les fréquences de F1, F2 et F3 comme paramètres de classification avec et sans le paramètre de la durée vocalique ont été effectuées. Les résultats indiquent que l'ajout de la durée vocalique améliore le taux total de classification de 10 % (73 % vs. 63 %). Les voyelles longues /i:, u:/ ont rarement été confondues avec leurs contreparties courtes et la moyenne de leur taux de classification correcte dépasse les 90 % lorsque la durée est incluse. En effet, les voyelles courtes /i, a, u/ sont souvent confondues avec les trois voyelles longues /e:, a:, o:/ respectivement en raison de leur proximité (/i/ avec /e:/, /a/ avec /a:/ et /u/ avec /o:/). Selon Al-Tamimi (2007), ces résultats laissent suggérer que le système vocalique de l'arabe jordanien est composé des huit voyelles /i:, a:, u:, e:, o:, e, a, o/ plutôt que celui proposé dans la littérature avec les voyelles /i:, a:, u:, e:, o:, i, a, u/. Cette proposition est motivée par le fait que les deux voyelles brèves /i, u/ sont produites de façon différente par rapport à leurs correspondantes longues et sont positionnées dans l'espace acoustique des voyelles /e:, o:/ respectivement dans tous les contextes étudiés (à l'exception du contexte de /d/).

En résumé, le système vocalique de l'arabe jordanien est composé de cinq timbres vocaliques (/i, a, u, e, o/) qui comprennent huit voyelles. Parmi ces voyelles, six s'opposent phonologiquement en durée ; il s'agit de /i:, a:, u:, i, a, u/ (ou /i:, ɪ, u:, e, a, o/ selon Al-Tamimi (2007)). La durée vocalique joue alors un rôle principal dans la distinction des voyelles en arabe jordanien. Les trois timbres /i, a, u/ de l'arabe jordanien témoignent de propriétés similaires à celles observées dans la thèse d'Allatif (2008), présentée dans le chapitre I, sur les mêmes voyelles en arabe syrien. Les voyelles /i, u/ et /i:, u:/ se différencient par la durée et par les fréquences des deux formants. Quant aux voyelles /a/ et /a:/, elles sont discriminées principalement par la durée vocalique.

La section précédente a mis en avant deux aspects qui caractérisent le système vocalique de l'arabe jordanien : le nombre restreint des timbres et l'importance de la durée vocalique. La section suivante décrit les aspects généraux des occlusives de l'arabe jordanien.

#### 4.1.1.2 Les occlusives de l'arabe jordanien

Contrairement au système vocalique de l'arabe jordanien, le système consonantique est relativement riche et divers. Il inclut principalement le système consonantique de l'arabe classique avec l'ajout de l'occlusive vélaire /g/ (Table 4.1). D'autres consonnes peuvent également être utilisées telles que la fricative labiodentale /v/ dans les mot empruntés (/vajru:s/ = virus) et l'affriquée /tʃ/ dans certains parlars jordaniens (/tʃe:f/ au lieu de /ke:f/) (Al-Tamimi, 2007).

Comme souligné dans la Table 4.1, l'arabe jordanien contient neuf consonnes occlusives avec cinq lieux d'articulation différents. Il s'agit des lieux d'articulation bilabial, alvéolaire, vélaire, uvulaire et glottal. Parmi les neuf occlusives, deux sont réalisées avec deux lieux d'ar-

	Bilabiale	Labiodentale	Dentale	Alvéolaire	Postalvéolaire	Pharyngalisée	Palatale	Vélaire	Uvulaire	Pharyngale	Glottale
Occlusive	b			t d		t <sup>ʕ</sup> d <sup>ʕ</sup>		k g	q		ʔ
Nasale	m			n							
Trille				r							
Fricative		f	θ ð	s z	ʃ ʒ	s <sup>ʕ</sup> d <sup>ʕ</sup>			χ ʁ	ħ ʕ	h
Approximante							j	w			
Latérale				l							

TABLE 4.1 – Système consonantique en Alphabet Phonétique International (IPA, pour *International Phonetic Alphabet*) de l'arabe jordanien en soulignant les différentes catégories des occlusives (Watson (2002), Barkat (2000), Al-Tamimi (2001), Al-Tamimi (2007)).

ticulation (principal et secondaire). Elles sont nommées les alvéolaires pharyngalisées /t<sup>ʕ</sup>, d<sup>ʕ</sup>/. Les occlusives en arabe jordanien peuvent alors être divisées en six catégories : bilabiale, alvéolaire, vélaire, uvulaire, glottale et pharyngalisée. Celles-ci interagissent avec les voyelles voisines de différentes façons. Cette interaction est reflétée par les changements du début et du milieu des formants vocaliques et peut être mesurée à l'aide des équations de locus. Les propriétés de ces catégories consonantiques et leur interaction avec les voyelles adjacentes seront présentées dans les sections suivantes.

### La bilabiale /b/

L'arabe jordanien contient une seule occlusive orale bilabiale (/b/) comme c'est le cas dans les autres dialectes de l'arabe. Autrement dit, l'occlusive bilabiale non voisée /p/ n'est pas dans l'inventaire consonantique de cette langue. Comme développé dans le premier chapitre, les bilabiales ont un locus bas conduisant à une baisse des fréquences au début des deuxièmes formants des voyelles avec des fréquences de F2 élevées (/i:, i, e:/) et moyennes (/a:, a/). Les voyelles avec des F2 relativement bas (/u:, u, o:/) sont quant à elles peu impactées (Al-Ani, 1970; Delattre et al., 1955).

L'occlusive bilabiale a un degré de coarticulation qui varie entre élevé et moyen selon la langue. Par exemple, les pentes des équations de locus sont de 0.69 en suédois (Lindblom, 1963b), 0.63 en arabe jordanien (Al-Tamimi, 2004), 0.89 en anglais américain (Sussman et al., 1991) et 0.87 en perse (Modarresi et al., 2004a). Le degré de la résistance coarticulatoire du /b/ est alors moyen ou bas car la langue n'est pas directement engagée dans la réalisation de cette occlusive et, par conséquent, le geste de la voyelle est libre pendant la fermeture (Fowler, 1994; Recasens et al., 1997).

**Les alvéolaires /t, d/**

Il existe deux occlusives alvéolaires en arabe jordanien (en plus des deux alvéolaires pharyngalisées) qui s'opposent en voisement (voisées /d/ vs. non voisées /t/). Les fréquences au début du F2 des voyelles /i:, i, e:/ ont légèrement baissé comparées à celles de leurs états stables lorsque ces voyelles sont produites avec les alvéolaires. Au contraire, les fréquences du F2<sub>onset</sub> des voyelles /u:, u, o:/ augmentent brusquement vers des fréquences plus élevées que celles de leurs états stables. Quant aux alvéolaires devant les voyelles /a:, a/, elles ont des fréquences de F2<sub>onset</sub> légèrement supérieures à celles des F2<sub>mid</sub> (Al-Ani, 1970; Delattre et al., 1955).

Les alvéolaires ont un degré de coarticulation bas avec les pentes des équations de locus inférieures à 0.6 dans la plupart des langues étudiées dans la littérature. En effet, les alvéolaires sont considérées comme les occlusives les plus résistantes à la coarticulation puisque leur prononciation implique le mouvement de la langue (Fowler, 1994; Recasens et al., 1997).

**Les pharyngalisées /t<sup>ħ</sup>, d<sup>ħ</sup>/**

La pharyngalisation est un phénomène articulatoire qui engage deux lieux d'articulation, principal et secondaire. La constriction principale est située typiquement dans la zone dentale/alvéolaire et celle secondaire se trouve dans la partie postérieure du conduit vocal (zone uvulaire ou pharyngale). Quatre consonnes pharyngalisées (dites aussi emphatiques) existent en arabe classique qui sont aussi rencontrées dans les dialectes arabes. Ces consonnes comptent deux occlusives alvéolaires pharyngalisées (/t<sup>ħ</sup>, d<sup>ħ</sup>/), une fricative alvéolaire pharyngalisée (/s<sup>ħ</sup>/) et une fricative dentale pharyngalisée (/ð<sup>ħ</sup>/) qui s'opposent phonologiquement à leurs contreparties non pharyngalisées /t, d, s, ð/.

Les caractéristiques articulatoires principales accompagnant la production des consonnes pharyngalisées en arabe jordanien ont été examinées par plusieurs chercheurs (Zawaydeh, 1999; Al-Tamimi et al., 2009; Al-Tamimi et Heselwood, 2011). Au moyen d'images endoscopiques d'une seule locutrice jordanienne (l'auteure elle-même), Zawaydeh (1999) décrit que l'articulation secondaire des consonnes emphatiques se produit par la rétraction de la partie postérieure de la langue contre la luette. Selon l'auteure, les consonnes emphatiques en arabe jordanien sont ainsi uvularisées et non pharyngalisées. Toutefois, Al-Tamimi et al. (2009) critiquent la méthodologie de Zawaydeh (1999) et conçoivent un protocole expérimental utilisant des images vidéoendoscopiques de quatre locuteurs jordaniens (2 femmes et 2 hommes) prononçant les consonnes emphatiques et leurs contreparties non emphatiques. Les auteurs observent que des mécanismes articulatoires différents ont lieu simultanément lors de l'articulation des emphatiques. En effet, la racine rétractée de la langue s'approche de l'oropharynx dans la cavité orale (zone appropriée pour la réalisation des pharyngalisées) et, par conséquent, l'os hyoïde et le larynx s'élèvent. Aucun engagement de la luette n'est noté. D'après ces résultats, Al-Tamimi et al. (2009) indiquent que les emphatiques en arabe jordanien sont pharyngalisées.

Plus tard, Al-Tamimi et Heselwood (2011) examinent des images nasoendoscopiques et

vidéofluoroscopiques de six locuteurs de l'arabe jordanien (3 femmes et 3 hommes) afin de préciser le lieu de la constriction secondaire et de comparer les gestes articulatoires des consonnes pharyngalisées avec ceux de leurs correspondantes non pharyngalisées. Les résultats des deux types de données sont en accord avec ceux de l'étude précédente d'Al-Tamimi et al. (2009). Les données nasoendoscopiques révèlent qu'une constriction se produit à l'oropharynx inférieur par le mouvement de l'épiglotte. Cette constriction épiglottopharyngale est similaire à celle observée lors de la prononciation des consonnes pharyngales. Les images vidéofluoroscopiques présentent les mêmes résultats que ceux des images vidéoscopiques de l'étude de 2009. Un résumé référant les différences entre les mouvements articulatoires des consonnes pharyngalisées et leurs contreparties non pharyngalisées est montré dans la Table 4.2.

Structure anatomique	Non pharyngalisées	Pharyngalisées
Corps de la langue	plus élevé	plus bas
Os hyoïde	plus bas	plus élevé
Racine de la langue	avancée	rétractée
Mur pharyngal postérieur	moins courbé	plus courbé
Épiglotte	avancée	rétractée
Sinus piriforme	plus ouvert	moins ouvert
Aperture laryngale	plus large	plus étroite
Larynx	plus bas	plus élevé

TABLE 4.2 – Comparaison entre les ajustements articulatoires relatifs aux consonnes emphatiques et non emphatiques (Al-Tamimi et Heselwood, 2011)

Les conséquences coarticulatoires acoustiques des emphatiques se manifestent par une légère élévation du F1, une baisse importante du F2 et une élévation moyenne du F3 des voyelles adjacentes (Zawaydeh, 1997, 1999; Zawaydeh et de Jong, 2002, 2011; Jongman et al., 2011; Al-Tamimi et Heselwood, 2011). Jongman et al. (2011) étudient l'impact de l'opposition emphatique *vs.* non emphatique en arabe jordanien en examinant plusieurs facteurs (position syllabique, quantité et qualité vocalique). Six locuteurs jordaniens ont produit les quatre consonnes pharyngalisées et leurs correspondantes non pharyngalisées dans des séquences CVC en position initiale et finale où V = /i, a, u, i, a, u/. Les fréquences des trois premiers formants ont été prélevées au début, au milieu et à la fin de chaque séquence. Les résultats montrent que les fréquences de F1 et de F3 des voyelles réalisées avec les consonnes emphatiques, dans les trois points de mesures (début, milieu et fin), sont significativement plus élevées que quand les mêmes voyelles sont prononcées avec leurs contreparties non emphatiques. Concernant les fréquences du F2 avec les pharyngalisées, elles sont statistiquement inférieures à celles des F2 avec les consonnes non pharyngalisées dans les trois positions de mesures (Figure 4.2). Les effets des consonnes emphatiques sont similaires lorsqu'elles sont en position initiale ou finale. Les trois timbres vocaliques mesurés au milieu ne sont pas influencés au même degré. La voyelle /a/ est la plus impactée des trois voyelles par les consonnes emphatiques sur les trois formants. La voyelle /i/ est plus affectée que /u/ en F1 et en F2 mais pas en F3. Il est également à noter

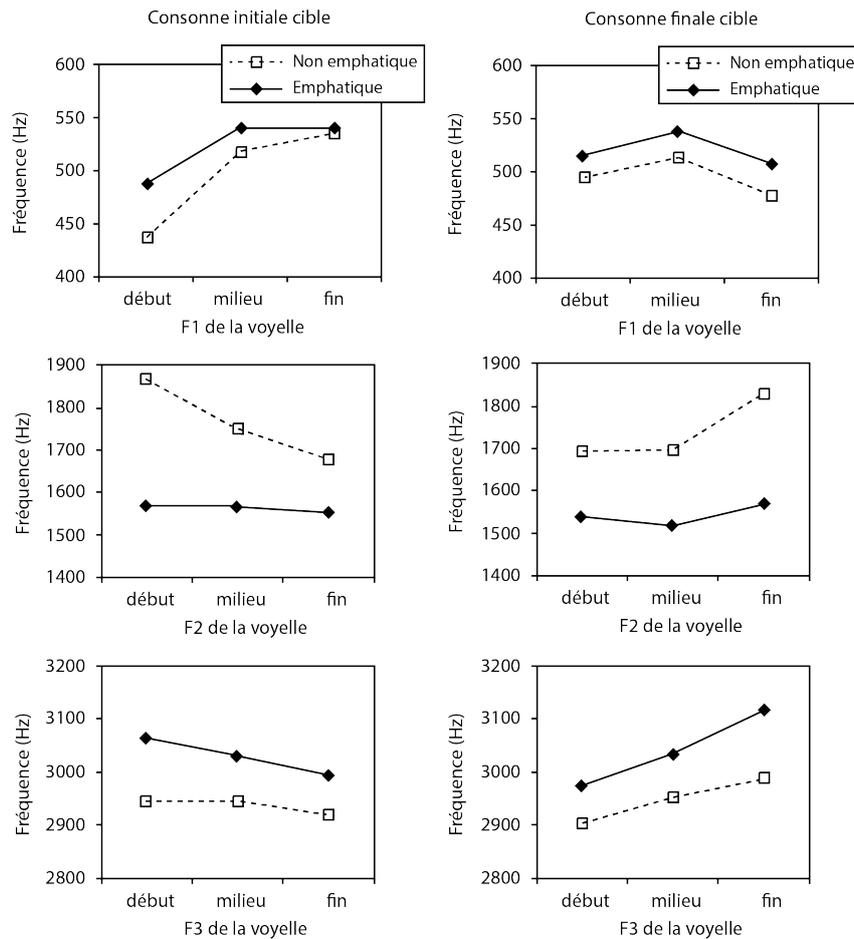


FIGURE 4.2 – Fréquences des formants (F1 en haut, F2 au milieu et F3 en bas) mesurées au début, au milieu et à la fin des voyelles dans les séquences CVC avec la consonne cible (emphatique vs. non emphatique) en position initiale (à gauche) et finale (à droite) (d'après Jongman et al. (2011)).

que l'effet de la consonne emphatique continue durant toute la voyelle /a/ mais il décroît rapidement pour les voyelles /i, u/ lorsque les mesures sont prises loin du début de la consonne cible.

D'un autre côté, un effet principal de la longueur vocalique (voyelles longues vs. courtes) a été observé lors de l'examen de l'impact des consonnes pharyngalisées sur les voyelles. Pour toutes les voyelles, celles courtes ont des F1 et des F3 significativement plus élevés et des F2 plus bas que les voyelles longues dans les trois points de mesures (à l'exception du point final du F3). Ces résultats mettent en avant que les voyelles courtes sont plus influencées que leurs contreparties longues par les consonnes pharyngalisées.

Le degré de coarticulation des consonnes pharyngalisées a été calculé à l'aide des équations de locus et examiné par plusieurs chercheurs (Sussman et al. (1993) en arabe égyptien, Yeou (1997) et Embarki et al. (2006) en arabe standard, Al-Tamimi (2004) en arabe jordanien, Embarki et al. (2007); Emnarki et al. (2011) en arabe standard et en arabe dialectal). Toutes ces études dévoilent un degré de coarticulation moindre pour les pharyngalisées en comparaison de

leurs contreparties non pharyngalisées (la pente est plus basse). Autrement dit, les consonnes pharyngalisées sont plus résistantes que leurs correspondantes non pharyngalisées.

### Les vélares /k, g/

Les vélares en arabe jordanien sont composées de deux consonnes, la voisée /g/ et la non voisée /k/. La vélaire voisée /g/ n'est pas présente dans l'inventaire consonantique de l'arabe classique mais elle est une variante du /q/. Elle apparaît pourtant dans plusieurs dialectes de l'arabe dont le jordanien pour remplacer l'uvulaire /q/ et dans les mots empruntés (Bahloul, 2011; Embarki, 2011).

Les vélares se distinguent des autres occlusives par l'existence de deux allophones qui changent en fonction de la voyelle suivante. Le lieu d'articulation devient palatal lorsqu'elles sont coarticulées avec les voyelles antérieures alors qu'il est vélaire quand elles sont produites avec les voyelles postérieures.

Les vélares ont des fréquences de  $F2_{onset}$  légèrement plus basses que celles du F2 des voyelles /i, i, e:/ et légèrement plus élevées devant les voyelles /u, u, o:/. Quant aux fréquences des  $F2_{onset}$  des vélares devant les voyelles /a:, a/, elles sont nettement plus élevées que les  $F2_{mid}$  (Al-Ani, 1970; Delattre et al., 1955). Les vélares ont un degré de coarticulation variant entre moyen et élevé selon la langue et l'échantillon avec des pentes des équations de locus qui s'étendent entre 0.65 et 1 (Sussman et al., 1991; Fowler, 1994; Modarresi et al., 2004a). Lors de la séparation des deux allophones (palatal et vélaire) en fonction de la voyelle suivante, ces occlusives prononcées avec les voyelles antérieures sont les plus résistantes avec des pentes des équations de locus inférieures à 0.4 et celles produites avec les voyelles antérieures sont les plus moins résistantes avec des pentes proches de 1 (Sussman et al., 1991; Fowler, 1994).

### L'uvulaire /q/

L'occlusive uvulaire non voisée est articulée par le contact de la partie supérieure-postérieure du dos de la langue contre la luette (Ghazali, 1977). Cette occlusive représente un cas particulier en arabe jordanien puisqu'elle existe en trois variantes : la forme standard [q] (uvulaire), la forme bédouine/rurale [g] (vélaire) et la forme urbaine/palestinienne [ʔ] (glottale). En effet, les continueurs de l'occlusive uvulaire, [g] et [ʔ], reflètent une diversité considérable concernant plusieurs paramètres : standard vs. dialectale, classe sociale et urbain vs. non-urbain (Al-Tamimi, 2001). Globalement, le /q/ est prononcé comme [ʔ] par les jordaniens d'origine palestinienne et comme [g] par les villageois et les bédouins jordaniens. L'augmentation du niveau de l'éducation académique accroît l'utilisation du [q] standard dans les deux groupes (Al-Tamimi, 2001)<sup>5</sup>. Al-Tamimi (2001) expose que les locuteurs jordaniens utilisent la forme standard du /q/ (examinée dans ce manuscrit) pour exprimer des mots lorsque l'énoncé/la phrase est associé-e à un contexte littéraire, religieux ou technique, indépendamment de leur niveau

5. Cf. Al-Wer et Herin (2011) pour une présentation du cycle de vie du /q/ en arabe jordanien.

d'éducation. Toutefois, les variantes dialectales ([g] ou [ʔ]) peuvent être employées pour indiquer d'autres significations (par exemple : /qasaman billa:h/ "je jure par Dieu" vs. /gasam/ "il a divisé").

L'inclusion de l'occlusive uvulaire dans la présente étude permet d'examiner le degré de coarticulation d'un lieu d'articulation postérieur à celui des vélaires à l'aide des équations de locus. En effet, cette occlusive étant peu présente dans les langues (Ladefoged, 2005), la plupart des recherches dans la littérature ont analysé les lieux d'articulation situés entre les bilabiales et les vélaires.

L'engagement de la racine de la langue représente un point commun entre l'occlusive uvulaire et les emphatiques. Par conséquent, des impacts acoustiques similaires sont observés sur les voyelles adjacentes (Al-Ani, 1970; Zawaydeh, 1997). Zawaydeh (1997) remarque que le /q/ en arabe jordanien a tendance à uvulariser les voyelles adjacentes. Autrement dit, les fréquences des F2 sont plus basses et celles de F1 sont légèrement plus élevées que celles des voyelles voisines. Toutefois, cette influence est moins forte que celle exercée par les consonnes emphatiques. De plus, elle ne dépasse pas la syllabe comme c'est le cas pour les emphatiques. En l'état actuel des recherches, cette occlusive n'a été examinée à l'aide des équations de locus qu'en arabe algérien (Bouferroum et Boudraa, 2015). Les auteurs ont observé que la pente de cette occlusive en arabe algérien est relativement moyenne avec une valeur de 0.75. La rareté de l'analyse de cette occlusive a motivé l'intérêt de l'étudier dans notre recherche.

### La glottale /ʔ/

Le lieu d'articulation de l'occlusive glottale non voisée (/ʔ/) représente le point le plus extrême dans le conduit vocal. Elle est produite avec une fermeture complète de la glotte suivie d'un relâchement brusque (Ladefoged, 2005). D'autres événements articulatoires peuvent accompagner la réalisation de cette occlusive. Dans leur étude nasoendoscopique et vidéofluoroscopique des consonnes laryngales et pharyngales en arabe jordanien, Heselwood et Al-Tamimi (2011) observent un déplacement de l'épiglotte vers l'arrière pendant la réalisation des glottales /h, ʔ/. Ce déplacement est plus important dans le contexte de la voyelle /a/ que dans celui de /i/. Dans les deux cas, aucune constriction pharyngale n'est remarquée.

Le /ʔ/ apparaît sur un spectrogramme comme un *burst* suivi d'un intervalle de silence court (entre 15 et 20 ms) ou comme un début de *glide* court amorçant les formants des voyelles suivantes (Al-Ani, 1970). De ce fait, les fréquences au début des transitions formantiques de /ʔ/ sont instables et sont complètement dépendantes des voyelles adjacentes (Al-Ani, 1970; Heselwood et Al-Tamimi, 2011). Par conséquent, la fréquence au début de la transition ( $F_{onset}$ ) ou à la fin ( $F_{offset}$ ) est similaire à celle au milieu du formant ( $F_{mid}$ ). Autrement dit, il n'existe pas de transitions formantiques particulières (Bin-Muqbil, 2006). Ce phénomène est expliqué par l'absence de constriction orale ou pharyngale lors de la réalisation de cette occlusive. D'après ces informations, il est attendu à ce que les pentes des équations de locus de l'occlusive glottale soient proches de 1 puisque la coarticulation est maximale.

En résumé, les sections précédentes ont présenté une description générale des voyelles et des occlusives de l'arabe jordanien en montrant la pauvreté du système vocalique et la richesse des lieux d'articulation des occlusives. Il a également été mis en avant que la prononciation des occlusives pharyngalisées, uvulaire et glottale entraînent des constrictions postérieures. L'examen de ces occlusives, souvent peu étudiées dans la littérature, fournissent alors une vue globale des différents degrés de coarticulation sur un espace consonantique large contenant les deux points d'articulation les plus extrêmes (bilabial et glottal), un point médian (vélaire), un point avant ce point médian (alvéolaire) et un autre après (uvulaire).

L'étude sur l'arabe jordanien permet aussi d'examiner un autre phénomène qui est l'influence de la variation temporelle intrinsèque des voyelles (longues vs. courtes) sur la coarticulation et sur la classification des consonnes. Autrement dit, le degré de coarticulation d'une consonne produite avec des voyelles longues serait-il impacté si cette consonne est prononcée avec des voyelles courtes ? De même, la classification des consonnes serait-elle affectée par cette opposition ? La présence de l'information temporelle dans ces cas-là est-elle importante pour la classification ? Une autre question peut également être posée résultant de celles précédentes : la variation temporelle dans une langue qui ne contient pas d'opposition phonologique de durée, tel que le français, aurait-elle les mêmes conséquences acoustiques ou spectrales que celle provoquée par l'opposition phonologique en arabe jordanien ? Dans le premier cas, la variation temporelle est extrinsèque à la voyelle (comme les variations du débit de parole). Afin de répondre à cette question, le français et l'arabe jordanien ont été choisis comme langues de référence dans notre étude.

## 4.2 Variations temporelles intrinsèques et extrinsèques

Les variations temporelles d'une voyelle sont associées à l'opposition de durée phonologique dans certaines langues. Cette opposition (voyelle longue vs. courte) est intrinsèquement liée à la voyelle, autrement dit, les voyelles longues sont physiquement plus longues que les voyelles courtes. Les variations temporelles d'une voyelle peuvent aussi dépendre des variations du débit de parole. Ce phénomène extrinsèque à la voyelle change également la durée ; les voyelles réalisées à débit lent sont physiquement plus longues que les mêmes voyelles prononcées à débit rapide. L'un des objectifs de ce travail est d'étudier l'impact de ces différents types de variations temporelles sur les qualités spectrales de la voyelle et de la consonne, sur la relation coarticulatoire entre la consonne et la voyelle ainsi que les conséquences de ces variations sur la classification des consonnes.

Stevens (1998) décrit l'importance de l'organisation temporelle entre les articulateurs durant la production de la parole. Il déclare qu'une modification de mouvement pour toute structure laryngale ou supra-laryngale exige un intervalle de temps qui varie entre 150 et 300 ms, selon la structure et l'amplitude du mouvement. Cette étendue de durée peut être comparée à la moyenne des durées des syllabes (qui s'étend entre 200 et 250 ms environ) pendant la lecture d'un texte

en débit normal. Selon lui, si l'on admet que chaque syllabe contient 2.8 phones en moyenne, la durée moyenne d'un phone serait entre 70 et 90 ms. Il est donc clair que les mouvements articulatoires produisant un son donné doivent être intercalés avec les mouvements nécessaires pour réaliser les phones adjacents. La production de la parole spontanée exige alors une coordination complexe des différents articulateurs qui ont chacun, à leur tour, des limitations dans leur débit de mouvement. Dans notre cas d'étude, deux situations temporelles s'opposent. La première concerne un timbre vocalique réalisé dans deux durées différentes conduisant à une opposition phonologique. Dans la seconde, la durée du timbre vocalique est modifiée pour une raison non-inhérente à la voyelle et n'implique pas d'opposition phonologique. La vitesse des mouvements des articulateurs changent alors dans les deux situations mais pour des raisons différentes. De plus, dans la seconde situation, l'ensemble des mouvements articulatoires des sons de l'énoncé sont accélérés en passant d'un débit lent à un débit rapide alors que dans la première, seule la durée des mouvements de la voyelle est réduite en allant d'une voyelle longue à sa contrepartie courte.

Plusieurs études ont été menées dans le but de comprendre les différences existantes entre les voyelles longue et courte au niveau quantitatif (durée) et qualitatif (fréquence) (cf. section 1.2.2). Par exemple, Hadding-Koch et Abramson (1964) ont révélé que le critère de la quantité des voyelles n'est pas suffisant pour percevoir la distinction entre voyelles longue et courte en suédois. Dans la plupart des cas, c'est la qualité des voyelles qui est indispensable pour les différencier. Dans une étude translinguistique sur l'arabe, le japonais et le thaï, Tsukada (2009) a observé que l'influence de la quantité de la voyelle /a/ sur sa qualité est faible dans toutes les langues étudiées. Selon lui, les auditeurs utilisent principalement l'information de la durée acoustique de la voyelle afin de choisir entre /a/ et /a:/. Allatif (2008) quant à lui met en évidence que l'importance de la quantité de la voyelle dans sa perception n'est pas similaire pour tous les timbres vocaliques. La durée vocalique est principalement utilisée pour distinguer /a/ et /a:/, moyennement pour /u/ et /u:/ et très peu pour /i/ et /i:/. L'importance décroît alors de /a/ (élevée), /u/ (moyenne) et /i/ (basse).

Il existe par ailleurs un débat à propos de l'impact du débit de parole sur les propriétés spectrales. Comme précisé précédemment (cf. section 2.3.2), Lindblom (1963a) a remarqué que la qualité des voyelles (en suédois) est modifiée quand elles subissent des changements de débit. Il a ainsi proposé la notion d'*undershoot* selon laquelle les voyelles n'atteignent pas leur cible articulatoire / acoustique quand elles sont confrontées à certains phénomènes dont la variation du débit de parole. Au contraire, dans l'étude de Gay (1978a), l'impact des variations du débit sur les propriétés spectrales des voyelles de l'anglais semble négligeable. Ce sont plutôt les consonnes (les  $F_{onsets}$ ) qui seraient influencées par ce phénomène. Sur le même sujet d'étude, O'Shaughnessy (1986) a étendu le travail de Gay (1978a) au français québécois en mettant en évidence que les variations temporelles attachées au débit de parole ne jouent pas directement un rôle sur la composition spectrale des voyelles et des consonnes. Selon lui, les indices spectraux restent relativement stables sous l'effet des variations temporelles. Il conclut

que c'est essentiellement la vitesse des transitions formantiques qui varie en fonction du débit de parole. Harrington (2010) commente ce débat en soulignant que « *The rate effects on the vowel space are not all consistent [...] not only because speakers do not all increase rate by the same factor, but also because there can be articulatory reorganization with rate changes* »<sup>6</sup>.

Les anciennes études sur les relations articulatoire / acoustique associées aux occlusives ont cherché à identifier des propriétés acoustiques invariantes de manière à décrire le lieu d'articulation (Delattre et al., 1955; Öhman, 1966; Stevens et Blumstein, 1978; Kewley-Port, 1982) (cf. section 1.3). Le locus formantique est l'un des concepts créés pour caractériser ces relations (Delattre et al., 1955). Toutefois, ce paramètre étant sensible aux contextes vocaliques (Öhman, 1966; Kewley-Port, 1982), il a fallu chercher des indices plus fiables qui prendraient en compte les aspects dynamiques de la production de la parole. Les équations de locus ont ainsi été proposées comme étant des invariants relationnels du lieu d'articulation (Sussman et al., 1991; Sussman et Shore, 1996; Sussman et al., 1998b; Lindblom et Sussman, 2012). Elles sont aussi de bons indicateurs du degré de coarticulation entre les consonnes et les voyelles (Lindblom, 1963b; Krull, 1989; Duez, 1992; Fowler, 1994). Dans ces études, les équations de locus ont été mesurées sans tenir compte des phénomènes temporels que sont la longueur et la durée des voyelles, à l'exception des études de Krull (1987) et Duez (1992). Ces dernières montrent alors un effet des durées sur les paramètres des équations de locus en comparant la parole spontanée et la parole contrôlée.

#### 4.2.1 Équations de locus et temporalité de la parole

Il apparaît que la représentation qui est fournie par les équations de locus fait abstraction d'un paramètre central de la parole : la temporalité. En effet, les valeurs qui servent à construire le nuage de points servant de support à la régression linéaire (voir l'illustration Figure 3.1) sont des mesures de fréquence prises à deux instants dans la réalisation : le moment initial correspondant à la consonne<sup>7</sup> et le milieu (ou zone de stabilité maximale) de la voyelle. L'impact possible de la temporalité sur les trajectoires formantiques ne semble donc pas crucial dans l'expression des équations de locus puisque ces deux « événements » peuvent se répartir différemment dans le temps. Par exemple, lorsque Rhone et Jongman (2012) ont calculé les équations de locus avec un  $F2_{mid}$  extrait de trois positions temporelles différentes du F2 (entre  $F2_{onset}$  et  $F2_{mid}$ ), ils ont observé que la position temporelle du  $F2_{mid}$  n'a pas d'impact sur la séparabilité des catégories consonantiques des équations de locus. Cela pourrait mettre en évidence une relative indépendance entre les paramètres des équations de locus et les propriétés temporelles des signaux de parole. Néanmoins, il serait intéressant d'intégrer la notion de temporalité pour représenter la

6. « L'influence du débit sur l'espace vocalique n'est pas toujours claire [...], pas seulement parce que les locuteurs n'augmentent pas le débit par le même facteur, mais aussi parce qu'il pourrait y avoir des réorganisations articulatoires accompagnant les variations de débit », page 92 [notre traduction].

7. La fréquence est prise au moment du bruit de friction résultant du relâchement ou lors des premiers cycles de vibration laryngale faisant suite à l'aspiration consonantique.

dynamicité temporelle de la parole.

Le mode de calcul des équations de locus impose aux trajectoires formantiques une forme de « discrétisation » ignorant les propriétés de continuité du signal de parole (O’Shaughnessy (1986); Al-Tamimi (2007)). Pourtant, Krull (1989) a observé un effet lié au style de parole dans son travail en comparant les degrés de coarticulation entre parole spontanée et contrôlée (cf. section 3.3). Selon elle, cet effet pourrait être dû aux différences de débit ainsi qu’à des phénomènes de réduction vocalique. Parallèlement, la réplique du travail de Krull (1989) sur des locuteurs natifs du français par Duez (1992) a conduit aux mêmes résultats. Malgré cela, Sussman et al. (1998b) semblent considérer que les effets mis en évidence par Krull (1989) sont relativement limités car « *mean slope difference between speaking styles across all consonants was only 0.06. Most important, speaking style variation did not perturb locus equation slopes in their role as phonetic descriptors of consonant place* »<sup>8</sup>.

Il apparaît effectivement que les valeurs moyennes notées diffèrent peu en termes de taille observée de l’effet et que les valeurs moyennes des pentes restent relativement ordonnées les unes aux autres entre labiales et dentales sans changement majeur. Or, il n’en reste pas moins que ce que met en évidence le travail de Krull (1989), c’est bien l’impact statistiquement significatif que peuvent avoir des influences stylistiques sur les paramètres des équations de locus. Par ailleurs, même si le travail de Krull (1989) illustre réellement un maintien de la différence entre labiales et dentales quel que soit le style de parole du point de vue de la pente des équations, il n’apporte pas de preuve concernant l’impact potentiel de cette variation en termes perceptifs. Pour cela, il conviendrait d’analyser la contribution de ces paramètres acoustiques à leur classification. Krull (1989) fait référence à deux phénomènes qui pourraient être à l’origine de ces variations de pente. L’un est un phénomène d’origine principalement spectrale concentré sur la voyelle où la réduction vocalique plus marquée en parole spontanée peut conduire à des changements de fréquence des formants plus importants au niveau de la cible vocalique. L’autre est un phénomène impactant la relation entre les segments où les changements de pente pourraient s’expliquer par des variations dans la relation temporelle qui s’établit entre la consonne et la voyelle. Cet effet du style de parole pourrait par exemple influencer le degré de coarticulation et contribuer alors à modifier simultanément les propriétés spectrales de la consonne et de la voyelle.

Plus récemment, deux études ont testé directement l’influence des variations temporelles issues du changement du débit de parole sur les propriétés spectrales et sur les pentes des équations de locus (Agwuele et al., 2007, 2008; Berry et Weismer, 2013). Agwuele et al. (2008) conçoivent un protocole expérimental sur l’anglais américain dans lequel six locuteurs produisent des mots CVC où la consonne initiale C correspond à /b, d, g/ et est prononcée avec dix voyelles /i, e, ε, æ, ai, u, o, ɔ, ʌ, au/. Les locuteurs lisaient les mots cibles dans une phrase

---

8. « la différence moyenne des pentes entre les styles de parole sur l’ensemble des consonnes était seulement de 0.06. Plus important, les variations du style de parole ne perturbaient pas les pentes des équations de locus dans leur rôle de descripteurs phonétiques de la place [d’articulation] des consonnes », page 248 [notre traduction].

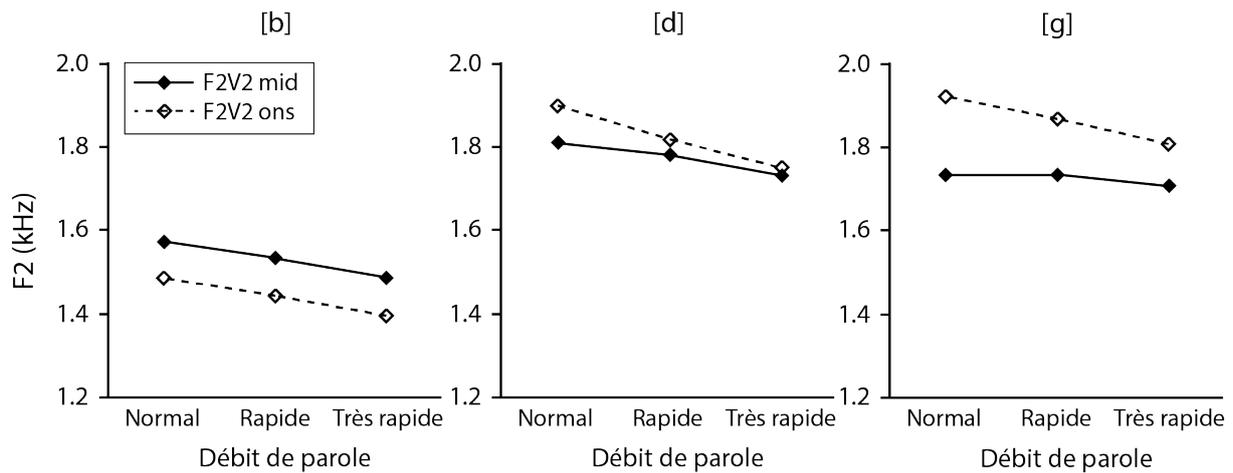


FIGURE 4.3 – Illustration de l’influence du débit de parole sur les fréquences des  $F2_{onset}$  et des  $F2_{mid}$  (emprunté à Agwuele et al. (2008)).

porteuse dans trois débits différents (normal, rapide et très rapide). Chaque participant déterminait les vitesses de son propre débit de parole. Les résultats indiquent que la durée vocalique diminue significativement en allant du débit lent au débit rapide. Les points de F1-F2 se centralisent en débit rapide dans l’espace vocalique tandis qu’ils s’éloignent vers des positions périphériques en débit lent. Les équations de locus ont ensuite été calculées pour chaque débit de parole. D’après les résultats, les pentes augmentent systématiquement lorsque le débit de parole accroît et inversement, elles diminuent quand le débit ralentit (Table 4.3). Les équations

	Normal		Rapide		Très rapide	
	P	OO	P	OO	P	OO
/b/	0.77	271	0.81	200	0.84	142
/d/	0.58	852	0.70	571	0.78	395
/g/	0.85	436	0.96	201	0.99	112

TABLE 4.3 – Valeurs moyennes des pentes (P) et des ordonnées à l’origine (OO) des six locuteurs pour les trois occlusives dans les trois débits de parole (Agwuele et al., 2008).

tions de locus peuvent ainsi refléter l’augmentation du chevauchement coarticulatoire lors des changements du débit de parole. Afin de comprendre l’origine acoustique de ces variations de pente, les auteurs ont analysé les coefficients des équations de locus ( $F2_{onset}$  et  $F2_{mid}$ ). Ils observent une déclinaison des fréquences des deux coefficients en fonction du débit de parole, qui se révèle plus importante pour les  $F2_{onset}$  (Figure 4.3). De plus, bien que les deux mesures aient tendance à se rapprocher lors de l’accélération du débit de parole, ce phénomène est moins prononcé pour la bilabiale. Selon les auteurs, ces modifications spectrales - notamment celle du  $F2_{onset}$  - sont responsables de l’élévation des valeurs de pente des équations de locus. La déviation des  $F2_{onset}$  de leurs valeurs en débit normal est essentiellement provoquée par les variations temporelles car le débit rapide laisse moins de temps aux articulateurs pour atteindre les configurations des consonnes et des voyelles. La fermeture des occlusives devient

ainsi moins appuyée en débit rapide qu'en débit normal conduisant à l'adaptation des gestes des articulateurs pour qu'ils ressemblent plus à ceux de la voyelle adjacente. Ils ajoutent que le cas contraire peut être rencontré avec les syllabes accentuées où les articulateurs ont plus de temps pour réaliser une fermeture plus appuyée ce qui entraîne une différence importante entre les gestes consonantiques et ceux vocaliques. Dans cette situation, reflétée par une baisse de la pente des équations de locus, les valeurs de  $F2_{onset}$  et  $F2_{mid}$  se dirigent vers des zones périphériques (– de coarticulation) (Lindblom et al., 2007). Agwuele et al. (2008) suggèrent alors que l'effet de l'augmentation du débit de parole et celui de l'accent tonique contribue à la variation de  $F2_{onset}$  sur les deux mêmes échelles corrélées. Ces dernières représentent d'un côté la durée (longue *vs.* courte selon la vitesse de l'articulation) et de l'autre, la force d'articulation (élaboration *vs.* réduction des mouvements articulatoires).

Berry et Weismer (2013) étudient également l'impact des variations du débit de parole sur le degré de coarticulation de l'occlusive bilabiale /b/ à l'aide des équations de locus. Quatre locuteurs de l'anglais américain ont prononcé des séquences bVt où V = /i, ɪ, ʌ, æ, ɑ, e, ε, o, ɔ, u/ dans une phrase porteuse. Les participants devaient lire les phrases en continu dans différents débits de leurs choix sans imposer de catégorie de débit comme dans l'étude de Agwuele et al. (2008). Chaque phrase a été répétée 80 fois par les locuteurs. L'occlusive bilabiale /b/ a été choisie dans cette étude pour maximiser l'indépendance articulatoire entre la consonne et la voyelle suivante provoquant moins de variant articulatoire. Dans ce cas, la langue est libre d'assurer le geste vocalique pendant la fermeture de la consonne. Les réalisations de chaque participant ont été séparées en quatre quartiles en fonction de la durée des voyelles constituant quatre sous-ensembles de données. Chacun de ces sous-ensembles représente un débit de parole (très lent, lent, rapide et très rapide). Les équations de locus ont été calculées pour chaque quartile/débit.

Les résultats montrent une grande variation dans les durées vocaliques entre les quatre locuteurs. Le degré de l'impact du débit de parole sur les  $F2_{onset}$  et les  $F2_{mid}$  dépend du locuteur et de la voyelle. Autrement dit, la distance entre ces deux mesures n'est pas comparable pour tous les locuteurs ou pour toutes les voyelles. Berry et Weismer (2013) ont effectué trois séries d'analyses d'équations de locus en fonction de groupes vocaliques. La première série correspond au calcul des équations de locus pour toutes les voyelles, la deuxième pour cinq sous-ensembles de voyelles et la dernière pour chaque voyelle séparée. Les résultats des pentes des équations de locus avec toutes les voyelles, en accord avec ceux de l'étude de Agwuele et al. (2008), révèlent une augmentation des valeurs de pente et une diminution de celles de l'ordonnée à l'origine au cours de l'accroissement du débit de parole pour trois des quatre locuteurs.

Les auteurs ont ensuite divisé les données en cinq groupes de voyelles dans le but d'examiner l'influence du contexte vocalique ainsi que le débit de parole sur les équations de locus. Ces groupes correspondent aux voyelles cardinales /i, u, æ, ɑ/, aux voyelles antérieures /i, e, ε, æ/, aux voyelles postérieures /ɑ, ɔ, o, u/, aux voyelles basses /ε, æ, ɑ, ɔ/ et aux voyelles hautes /i, e, o, u/. Les résultats mettent en avant que les groupes de voyelles cardinales et de voyelles

hautes ont des valeurs de pentes qui s'approchent le plus de celles obtenues dans la condition avec toutes les voyelles (Figure 4.4). Les autres conditions dévoilent des changements qui diffèrent en fonction du locuteur et du débit de parole. En effet, pour les pentes des trois locuteurs

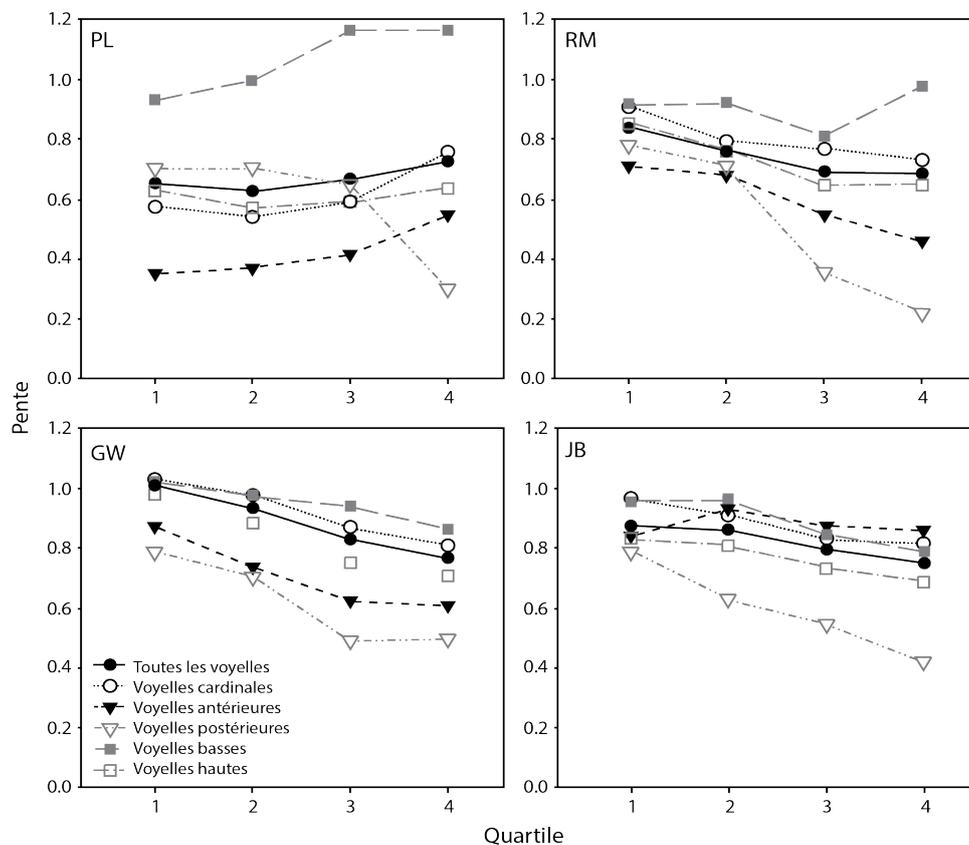


FIGURE 4.4 – Représentation des pentes des équations de locus des six groupes de voyelles (toutes les voyelles, cardinales, antérieures, postérieures, basses et hautes) dans les quatre quartiles/débits (très rapide, rapide, lent et très lent) pour les quatre locuteurs (d'après Berry et Weismer (2013)).

(RM, GW, JB) qui ont présenté des variations systématiques en fonction du débit de parole dans la condition avec toutes les voyelles ont les mêmes tendances dans la plupart des conditions vocaliques. La variation des valeurs de pente en fonction des groupes vocaliques indique que les voyelles ne sont pas toutes influencées au même degré par les changements du débit de parole. Les voyelles basses ont la moyenne de pente la plus élevée et sont les moins impactées par le débit de parole. Au contraire, les voyelles postérieures ont tendance à avoir la moyenne la plus basse et sont les plus influencées par le débit de parole.

La troisième série de calculs des équations de locus a été réalisée pour chaque voyelle soit dix équations de locus pour chaque locuteur. Aucune tendance n'a pu être remarquée pour les valeurs des dix pentes dans les quatre débits de parole puisque les valeurs de pente varient de façon presque aléatoire dans ces conditions. En effet, le calcul des équations de locus avec une seule voyelle représente un problème technique puisque la pente dépend d'un seul nuage de points et est facilement influencée par les points extrêmes.

L'étude de Berry et Weismer (2013) confirme les observations dans la littérature concernant l'influence des variations temporelles sur la coarticulation selon lesquelles une diminution de durée entraîne une augmentation du chevauchement articulatoire/acoustique et par conséquent, un degré de coarticulation plus élevé. Toutefois, cette tendance n'est pas toujours remarquée dans les différents groupes vocaliques car certaines catégories de voyelles sont moins impactées par les variations temporelles.

En résumé, les études dans la littérature ont mises en exergue que les variations temporelles extrinsèques modifient le temps de production d'un son. Cette modification augmente ou diminue le chevauchement articulatoire entre les segments adjacents engendrant ainsi des variations spectrales. L'impact des changements temporels sur la relation consonne-voyelle est détecté par les paramètres des équations de locus malgré l'absence de la dimension temporelle dans le domaine des équations de locus (pente et ordonnée à l'origine). En effet, les fréquences des  $F2_{onset}$  et des  $F2_{mid}$  sont modifiées par les variations temporelles et ont tendance à se rapprocher les unes des autres lorsque la durée diminue. D'après les études présentées ci-dessus, les consonnes et les voyelles sont sensibles à des degrés différents aux variations temporelles. L'un des objectifs de ce manuscrit est d'examiner l'importance de la présence de l'information temporelle dans la reconnaissance de la parole. Autrement dit, l'ajout de l'information temporelle améliore-t-il la classification des occlusives dans les différents débits de parole ?

L'impact des variations temporelles intrinsèques à la voyelle sur le degré de la coarticulation consonne-voyelle a rarement été évoqué dans la littérature. Dans ce cas, la durée temporelle d'une voyelle peut être longue ou courte sans subir une accélération du débit de parole ou un ralentissement d'accentuation. En effet, les études sur l'arabe ont montré que les voyelles longue vs. courte peuvent être spectralement différentes (selon le timbre) et que l'ajout de l'information temporelle améliore significativement la classification des voyelles (Al-Tamimi, 2007; Allatif, 2008). Cependant, en l'état actuel des recherches, l'impact de l'ajout de la durée vocalique intrinsèque sur la classification des consonnes n'a pas encore été traité. Il serait alors possible d'examiner la relation consonne-voyelle dans deux variations temporelles de nature différente. La première est la longueur phonologique en arabe jordanien et la seconde les changements de débits de parole en français. Il s'agira alors de comparer ces deux types de variations sur la perception des consonnes pour voir si l'information temporelle serait aussi importante dans les deux situations.

### 4.3 Conclusion

Dans ce chapitre, les problématiques principales ont été présentées avec leurs contextes. L'arabe jordanien constitue un cas particulier grâce à la richesse de son inventaire consonantique avec cinq lieux d'articulation et neuf occlusives dont deux avec un double points d'articulation. Cette richesse de consonnes est confrontée à la pauvreté du système vocalique avec cinq timbres dont trois qui s'opposent en longueur phonologique. En effet, l'examen des occlusives de cette

langue permet l'analyse de la relation consonne-voyelle dans deux situations intéressantes. La première correspond aux occlusives qui sont produites avec un déplacement maximal de la langue vers l'avant pour les alvéolaires et vers l'arrière pour l'uvulaire. Au contraire, la seconde situation atteste d'une absence de déplacement (ou d'un déplacement très faible) de la langue lors de la réalisation des deux occlusives ayant les points d'articulation les plus extrêmes dans le conduit vocal (bilabial et glottal).

D'un autre côté, l'opposition phonologique des voyelles en arabe jordanien donne également l'occasion d'étudier l'impact de la variation temporelle intrinsèque de la voyelle sur la relation entre la consonne et la voyelle. Les études dans la littérature ont indiqué que les voyelles longues ne diffèrent pas seulement en quantité mais elles peuvent également être distinguées par leur qualité. On peut alors s'interroger sur les conséquences que ces variations de quantité et/ou de qualité ont sur la coarticulation. Cette influence temporelle sera alors testée sur des lieux d'articulation assez étendus pour avoir une vision plus globale du phénomène.

Les changements du débit de parole représentent un autre type de variations temporelles qui est cette fois extrinsèque à la voyelle. Il a été montré que ce type de variations temporelles est lié à des modifications acoustiques et, par conséquent, à une perte d'informations spectrales (Lindblom, 1963a). De ce fait, la relation coarticulation entre les consonnes et les voyelles adjacentes témoigne d'une augmentation du chevauchement articulatoire et acoustique lorsque le débit de parole croît. Cette augmentation du degré de coarticulation a été mesurée à l'aide des équations de locus dans deux recherches récentes (Agwuele et al., 2008; Berry et Weismer, 2013). Les pentes des équations de locus s'élèvent et les ordonnées à l'origine décroissent lorsque le débit de parole augmente. Toutefois, l'impact de ces variations temporelles sur la classification des occlusives n'a pas été abordé. La perte de l'information spectrale engendrée par ces variations temporelles serait-elle compensée par l'ajout de l'information temporelle ? La langue française a été choisie ici comme langue de référence pour examiner cette question.

L'analyse des deux sujets cités plus haut permettra de comparer l'influence des variations temporelles de deux natures différentes (intrinsèque et extrinsèque) sur la coarticulation ainsi que l'importance de l'information temporelle dans ces deux cas dans la classification des occlusives. Autrement dit, si les occlusives sont prononcées avec les voyelles courtes (– de durée), auraient-elles un degré de coarticulation plus élevé que quand elles sont articulées avec les voyelles longues (+ de durée), étant donné que ce phénomène a déjà été observé dans les débits rapide (– de durée) vs. lent (+ de durée) ? De plus, est-ce que l'information temporelle dans les deux situations a le même poids perceptuel ?

Trois protocoles expérimentaux ont été conçus pour appréhender ces questions (deux en arabe jordanien et un en français). D'autres questions discutées dans la littérature seront également revues dans ces expériences telles que la position de  $F2_{onset}$  vs.  $F2_{@burst}$ , les différences entre occlusives voisée vs. non voisée et la position syllabique initiale vs. finale. Les chapitres suivants seront consacrés à la description des variations temporelles et spectrales et leurs impacts potentiels sur la coarticulation en arabe jordanien (dans le chapitre 5) et en français (dans

le chapitre 6). L'influence de ces variations temporelles/spectrales sur la classification des occlusives et l'importance de l'ajout de l'information temporelle sur l'amélioration du taux de classification seront examinés dans le chapitre 7. Les équations de locus seront utilisées comme un outil de mesure pour estimer le degré de coarticulation et les analyses linéaires discriminantes pour évaluer la classification des occlusives.



# Chapitre 5

## La longueur vocalique et la coarticulation

Ce chapitre consiste à décrire la méthodologie et énonce les résultats de deux expériences menées sur la production de la parole en arabe jordanien. La première se concentre sur l'examen des propriétés spectrales et coarticulatoires des occlusives non voisées alors que la seconde se focalise sur les occlusives voisées et sur l'opposition voisée *vs.* non voisée. L'objectif de ces expériences est d'offrir une analyse acoustique de l'impact des variations temporelles issues de l'opposition phonologique des voyelles sur les fréquences des formants et sur la relation entre les consonnes et les voyelles dans plusieurs contextes consonantiques.

### 5.1 Expérience I

Cette expérience a été conçue dans le but de construire une base de données contenant les propriétés acoustiques (fréquences des formants et durées vocaliques) des séquences CV et VC produites par des locuteurs de l'arabe jordanien. Cette base de données permettra la mise en place des analyses statistiques et celle du calcul des équations de locus qui fourniront les informations liées à l'influence temporelle sur les occlusives et sur les voyelles ainsi que sur leur relation coarticulatoire. Elle sera également utilisée pour effectuer les tests de classification de cinq occlusives non voisées étudiées *via* les analyses discriminantes linéaires (présentées dans le chapitre 7).

#### 5.1.1 Participants

Dix locuteurs jordaniens (5 femmes et 5 hommes) ont participé à cette expérience. Les locuteurs masculins proviennent de trois villes différentes en Jordanie dont deux locuteurs, nommés Ab et Ah<sup>1</sup>, sont originaires de la ville de Ma'an (y compris l'expérimentateur), deux d'Amman (Ko et Am) et un d'Irbid (Os). L'enregistrement s'est déroulé en France puisque tous étaient étudiants à l'Université de Nantes au moment de l'expérimentation. Les cinq locutrices féminines, quant à elles, sont originaires de la ville de Ma'an (Iq, An, Ib, Ha, Da) et ont passé le

---

1. Les deux lettres correspondent aux deux premières lettres du nom ou du prénom du locuteur.

protocole expérimental en Jordanie. Les dix locuteurs ne présentent aucun trouble du langage et étaient âgés de 20 à 38 ans lors de l'expérience.

### 5.1.2 Stimuli

Les locuteurs devaient lire à un rythme soutenu mais confortable des séquences CVC enchâssées dans une phrase porteuse (/ħaka CVC marte:n/ : « Il a dit CVC deux fois »). Les séquences apparaissaient l'une après l'autre en alphabet arabe vocalisé sur un écran d'ordinateur où les diacritiques représentant les voyelles étaient affichés (Figure 5.1). Un intérêt particulier a été porté aux cinq occlusives non voisées de l'arabe jordanien /t, t<sup>h</sup>, k, q, ʔ/ (alvéolaire, alvéolaire pharyngalisée, vélaire, uvulaire et glottale). Chacune de ces occlusives figurait dans un mot de la langue ayant une structure CVC. Ces cinq occlusives cibles ont été combinées avec les huit voyelles de l'arabe jordanien (trois courtes /i, a, u/ et cinq longues /i:, a:, u:, e:, o:/) dans deux positions lexicales (initiale C<sub>cible</sub>VC et finale CVC<sub>cible</sub>). L'autre consonne du mot était sélectionnée aléatoirement en fonction des mots existant dans le vocabulaire de l'arabe soit un total de 80 mots différents, tous en usage dans le dialecte jordanien. La fréquence des mots n'a pas été contrôlée.

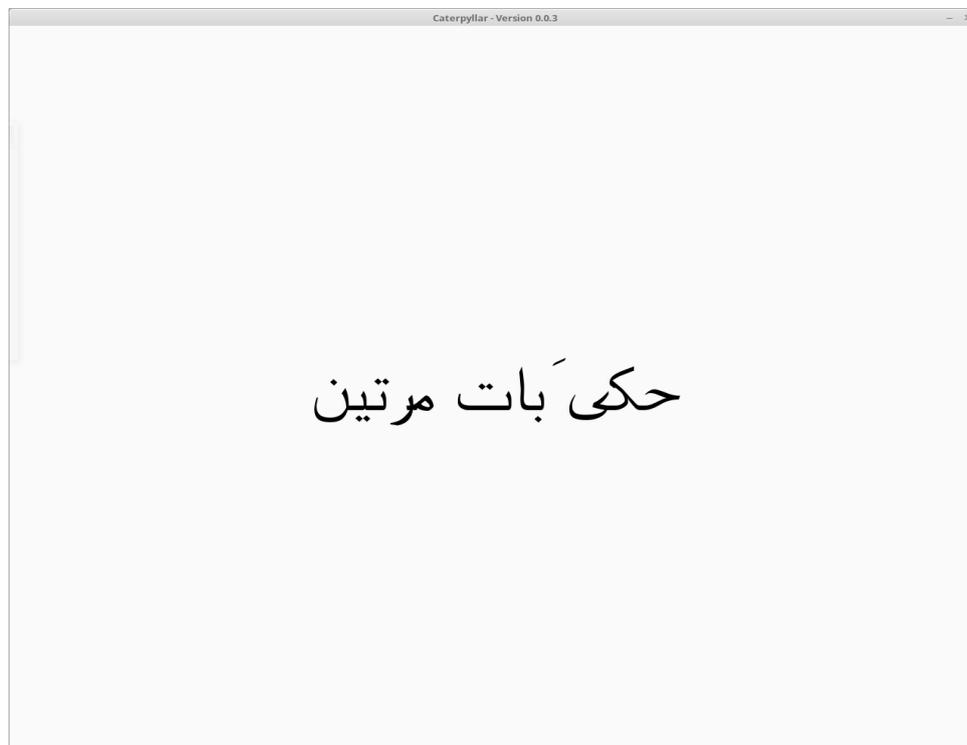


FIGURE 5.1 – Capture d'écran représentant un exemple des stimuli que les locuteurs devaient lire.

### 5.1.3 Procédure

Pour l'enregistrement, les locuteurs étaient placés dans une pièce calme devant un microphone électrostatique Rode S1 relié à un magnétophone portable Tascam Teac Professionnal DR-100. Chaque séance d'enregistrement durait approximativement une heure. Une pause automatique était programmée toutes les 40 phrases, toutefois, les locuteurs pouvaient interrompre l'enregistrement à tout moment quand ils le voulaient. Les locuteurs avaient pour consigne de lire les séquences « naturellement » sans chercher à « parler correctement ». Un programme écrit en Python<sup>2</sup> avec les bibliothèques PyGame<sup>3</sup> et PyFribidi<sup>4</sup> contrôlait le déroulement de l'expérience (durée d'affichage, ordre de présentation des séquences, interruption et reprise lors des pauses). Les séquences (mot en contexte avec la phrase porteuse) étaient présentées sur l'écran d'ordinateur dans un ordre aléatoire et à un rythme normal se rapprochant du débit de parole naturelle de chaque locuteur (une phrase toutes les 2500 ms) (Figure 5.1).

Les participants ne connaissaient pas les séquences avant le début de l'enregistrement. Chaque mot a été répété 15 fois soit 1200 séquences enregistrées par locuteur (5 consonnes × 2 positions syllabiques × 8 voyelles × 15 répétitions) totalisant 12 000 séquences CVC à analyser.

Les enregistrements ont été échantillonnés à 16 kHz et à 16 bits en mono puis les séquences CVC ont été segmentées et transcrites à la main avec le logiciel Praat (Boersma et Weenink, 2017). Deux tiers ont été utilisés dans la segmentation de chaque séquence avec le premier représentant le mot et le second la voyelle. Dans le tier du mot, les barres de segmentation ont été placées au moment du relâchement de l'occlusive et à la fin ou au début des formants vocaliques - selon la position de la consonne cible - alors que dans le tier de la voyelle, elles ont été placées au début et la fin des formants vocaliques (Figure 5.2). Toutes les séquences ont été transcrites en SAMPA (*Speech Assessment Methods Phonetic Alphabet*) qui est un alphabet phonétique lisible par les logiciels de programmation. Il est plus adapté que l'Alphabet Phonétique International (API) lors du traitement des données avec les logiciels de programmation<sup>5</sup>.

Les fréquences des trois premiers formants de chaque séquence CV/VC cible et leurs coordonnées temporelles ont été extraites automatiquement grâce à un script Praat<sup>6</sup>. L'algorithme d'extraction *Burg* (analyse LPC par auto-corrélation) a été employé avec une fenêtre d'analyse de 0.025 seconde et un pas de 0.00625 seconde. Prélevées par ce script et sauvegardées dans un tableur Calc, les données brutes correspondent aux tracés formantiques des F1, F2 et F3 avec leurs positions temporelles dans le tier de la voyelle et la position temporelle du bruit de relâchement dans tier du mot. Ces données ont par la suite été traitées par un script R (R Core Team, 2012) chargé de procéder à l'analyse des données. Dans un premier temps, les tracés formantiques ont été lissés par moyennage glissant sur toutes les suites de deux valeurs successives afin

---

2. <http://www.python.org>

3. <http://www.pygame.org>

4. <http://pyfribidi.sourceforge.net/>

5. <http://www.phon.ucl.ac.uk/home/sampa/index.html>

6. Les scripts Praat et R ont été écrits par Olivier Crouzet et ont ensuite été modifiés par Mohammad Abuoudeh.

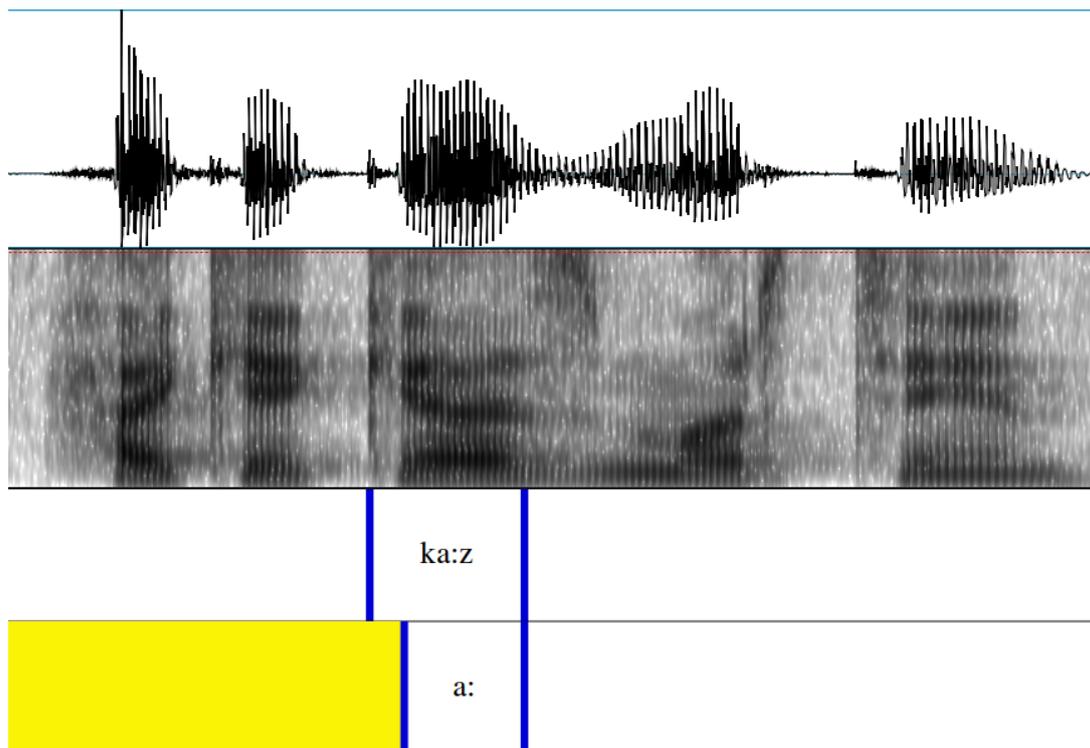


FIGURE 5.2 – Exemple de la segmentation de la séquence /ka:z/ dans Praat avec le tier du mot (en haut) et le tier de la voyelle (en bas).

de compenser les variations trop rapides des fréquences formantiques (filtrage passe-bas). Dans un second temps, les fréquences initiales ou finales (en fonction de la position de l'occlusive étudiée dans le mot) et médianes des trois premiers formants de chaque tracé formantique ont été sélectionnées et associées aux variables contrôlées dans l'expérience (occlusive cible, longueur vocalique, timbre vocalique, position de l'occlusive dans le mot, locuteur). Les positions au début et à la fin de chaque voyelle ainsi que la durée associée ont été déterminées à partir du début de l'apparition des résonances formantiques des deux premiers formants en incluant les transitions formantiques *on-glide* et *off-glide*. La Figure 5.3 représente la segmentation de la séquence /be:t/ dans Praat et la même séquence après le traitement dans R.

### 5.1.3.1 Les anomalies et les valeurs extrêmes

Avant le début des analyses statistiques et du calcul des équations de locus, les réalisations non conformes aux attentes de l'expérience ont été éliminées telles que la production d'un [g] au lieu de [q] et /ay/ au lieu de /e:/ et les séquences trop courtes (soit 5 séquences)<sup>7</sup>. Les locuteurs ont prononcé 348 fois le /q/ comme un [g] sur les 2400 réalisations totales de cette occlusive (10 locuteurs × 2 positions syllabiques × 8 voyelles × 15 répétitions). La majorité de ces séquences proviennent des locuteurs Os, Ha et Am (respectivement, 137, 73 et 72 fois). La réalisation de ces anomalies peut être due au fait que la tendance générale en arabe jordanien est de produire

7. C'est-à-dire, les séquences qui ont deux ou trois points de mesures. C'est le cas quand le locuteur n'a pas prononcé correctement le mot : bafouillage, bégaiement, etc.

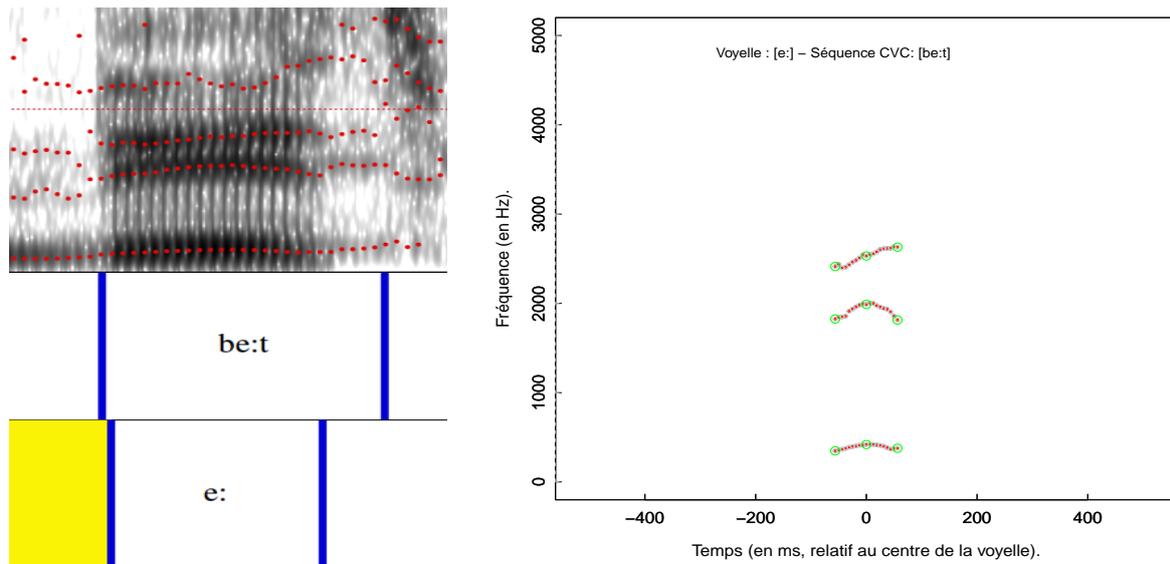


FIGURE 5.3 – Illustration du spectrogramme et de la segmentation de la séquence /be:t/ produite par le locuteur Ab dans Praat (à gauche) et résultat de la même séquence après le traitement dans R.

systématiquement le /q/ en /g/ dans un grand nombre de mots. Sachant que l'expérimentateur n'a pas influencé la lecture des locuteurs pour qu'ils réalisent les prononciations voulues (/q/), certains des mots choisis ont été prononcés avec un /g/. De plus, certaines séquences ont été lues dans une version non attendue. Par exemple, le mot /qad/ (emprunt de l'arabe standard) signifiant "peut-être" a été produit en /gad/ signifiant "il a déchiré", les deux mots ayant la même écriture. Quant à la voyelle /e:/, elle a été réalisée dans sa version de l'arabe standard 6 fois (par exemple, prononcer /kayf/ au lieu de /ke:f/). Au total, 359 séquences (348 + 5 + 6 sur les 12 000) ont été supprimées avant le début du traitement des données.

Les fréquences des trois premiers formants peuvent subir des erreurs de détection provoquées par le logiciel Praat (cf. 1.8.3). Puisque ces erreurs peuvent influencer les résultats globaux de cette expérience, il a fallu les éliminer. Ces valeurs ont été détectées automatiquement par le script R cité précédemment. Pour déterminer ces valeurs extrêmes, des régressions linéaires ont été produites à partir des points de  $F2_{onset}$  et  $F2_{mid}$  de manière à calculer les paramètres de chacune des 200 équations de locus (5 consonnes  $\times$  2 longueurs vocaliques  $\times$  2 positions syllabiques  $\times$  10 locuteurs). Les mesures dont le résidu (l'erreur) est compris entre la valeur observée dans l'espace  $F2_{onset} \sim F2_{mid}$  et celle prédite par la régression et qui était supérieur à la moyenne des valeurs absolues des résidus +1.5 fois l'écart-type ont été identifiées comme *outlier* et retirées de l'analyse dans le but d'éviter l'impact de ces valeurs extrêmes sur les pentes des équations de locus et sur les calculs statistiques. Suite à cette première phase d'analyse, une seconde phase de régressions linéaires a été réalisée sans ces valeurs. Ces valeurs extrêmes représentent 7 % des données (807 séquences).

La Figure 5.4 montre un exemple de la procédure adaptée pour la suppression des points extrêmes dans cette expérience. Dans un premier temps, une équation de locus est calculée à

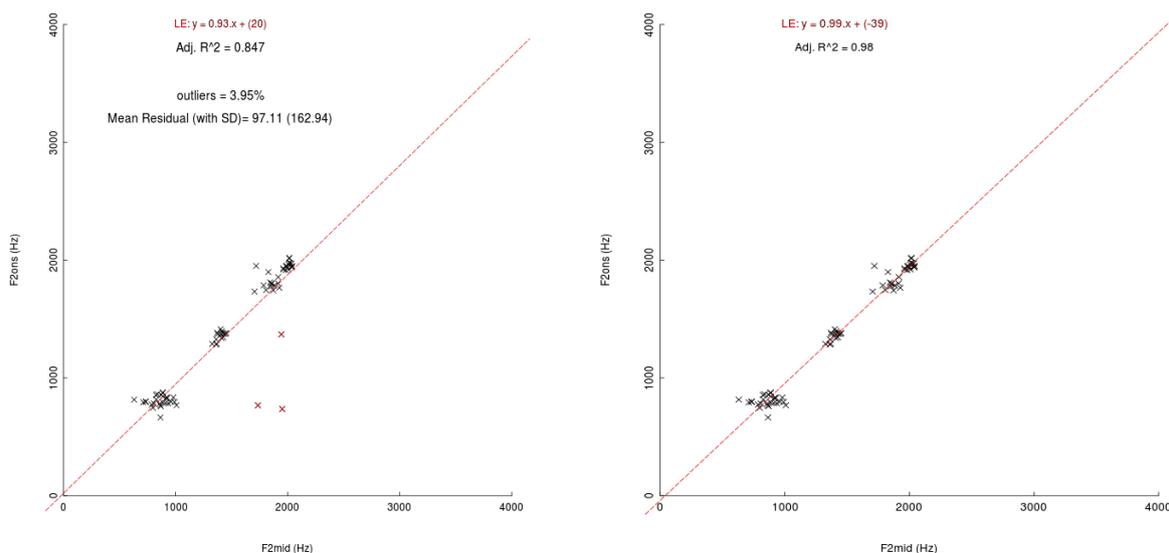


FIGURE 5.4 – Représentation de la détection des valeurs extrêmes par le biais du calcul des équations de locus. L'équation de locus de la glottale /ʔ/ produite par le locuteur Ah avec les *outliers* (les x en rouges) est affichée à gauche et celle sans à droite.

partir des valeurs de  $F2_{onset}$  et de  $F2_{mid}$  de l'occlusive /ʔ/ (à gauche). Dans un deuxième temps, le système de détection, cité plus haut, a repéré trois points extrêmes (en rouge) de cette équation qui ont ensuite été éliminés des données. Dans un troisième temps, l'équation de locus a ainsi été estimée sans ces trois points *outliers* (à droite). Il est à remarquer que l'omission de ces points a conduit à une modification de la valeur de la pente et de celle de l'ordonnée à l'origine. En effet, la pente sans les *outliers* est plus ajustée que celle avec. L'indication de cet ajustement est visible par l'augmentation de la valeur du  $R^2$  puisqu'elle est de 0.85 lorsque la pente est calculée avec les *outliers* et de 0.98 quand elle est calculée sans ces valeurs. Cette amélioration de la valeur de  $R^2$ , dans cet exemple et dans la plupart des autres équations de locus, met en exergue l'efficacité de la méthode de détection des valeurs extrêmes. D'autres valeurs extrêmes non détectées par cette méthode automatique ont été repérées et corrigées manuellement.

### 5.1.3.2 L'estimation de $F2_{@burst}$

De nombreuses études (par exemple Modarresi et al. (2005)) ont montré que la fréquence du deuxième formant au moment du relâchement ( $F2_{@burst}$ ) porte plus d'informations sur la consonne que celle de  $F2_{@onset}$ , notamment pour les occlusives non voisées. Par conséquent, les deux points de fréquences ont été pris en compte dans nos analyses. Sussman et Shore (1996) ont proposé de déterminer manuellement la fréquence de  $F2_{@burst}$  en prélevant la première résonance visible de F2 après le bruit de relâchement. Les auteurs ont jugé que cette première résonance est le prolongement de F2 de la voyelle (cf. 3.4.1.4, page 154). En effet, cette méthode est problématique car les résonances sont souvent cachées par le bruit de relâchement et/ou ont une intensité faible empêchant leur visibilité. En conséquence, un prélèvement automatique de ces mesures a été favorisé face à la quantité de données importante.

Les fréquences de  $F2_{@burst}$  en position initiale ont été estimées par la technique d'extrapolation<sup>8</sup> linéaire dans le logiciel R. Celles en position finale n'ont pas été retenues car la barre du bruit de relâchement n'est pas souvent trouvée (occlusive non relâchée) ou parce qu'elle est souvent trop éloignée de la voyelle conduisant à des valeurs de prédictions erronées et négatives. Deux méthodes d'extrapolation ont été utilisées pour estimer la valeur de  $F2_{@burst}$ . La première consiste à calculer la régression linéaire qui passe entre le  $F2_{@onset}$  et le  $F2_{mid}$ , la valeur de  $F2_{@burst}$  est ensuite prédite au moment du relâchement de l'occlusive. Cette opération est réalisée grâce à la fonction « `approxExtrap` » de la librairie `Hmisc`<sup>9</sup>. La seconde méthode est basée sur le calcul d'une régression linéaire des deux premiers points seulement ( $F2_{onset}$  et le point suivant) et permettant ensuite la prédiction de la valeur de  $F2_{@burst}$  au moment du relâchement (**extrapolation 2** ou  $F2_{@burst2}$ ). La différence entre les deux méthodes réside dans l'inclusion des variations spectrales entre les  $F2_{onset}$  et les  $F2_{mid}$  dans la première alors que la seconde ne considère que les variations les plus proches de la consonne. Autrement dit, l'ex-

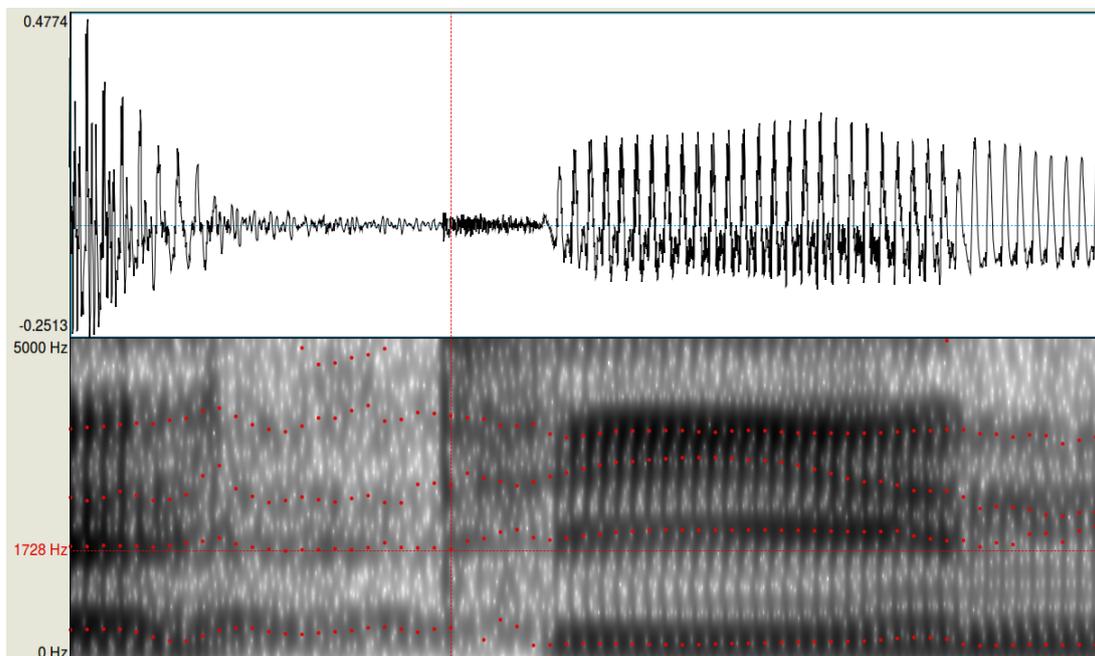


FIGURE 5.5 – Spectrogramme de la séquence /ti:n/ prononcée par le locuteur Ah montrant la valeur de  $F2_{@burst}$  (1728 Hz) suivant les consignes de Sussman et Shore (1996).

trapolation 1 considère les variations relativement lentes de la transition du deuxième formant (entre son début et son milieu) par comparaison avec l'extrapolation 2 qui ne considère que les changements rapides de fréquences au début de la transition.

Par exemple, en suivant les consignes de Sussman et Shore (1996), l'estimation de la valeur de  $F2_{@burst}$  de la séquence /ti:n/, affichée dans la Figure 5.5, est de 1728 Hz. Pour appliquer les méthodes d'extrapolation citées précédemment sur cette séquence, le tracé formantique du

8. « L'extrapolation est une estimation de la variable dépendante pour des valeurs des variables indépendantes qui se situent en dehors de l'ensemble des observations », page 201, Dodge (1988).

9. <https://CRAN.R-project.org/package=Hmisc>

F2 (les point de fréquence entre  $F2_{onset}$  et  $F2_{offset}$ ) est d'abord extrait<sup>10</sup> puis, une régression linéaire est calculée pour  $F2_{onset}$  et  $F2_{mid}$  (extrapolation 1) et pour les deux premiers points ( $F2_{onset}$  et le point suivant) (extrapolation 2). Le point de fréquence de  $F2_{@burst}$  est prédit au moment du relâchement dans les deux méthodes ce qui correspond à 46 ms (le VOT) avant le  $F2_{onset}$  pour cet exemple (Figure 5.6).

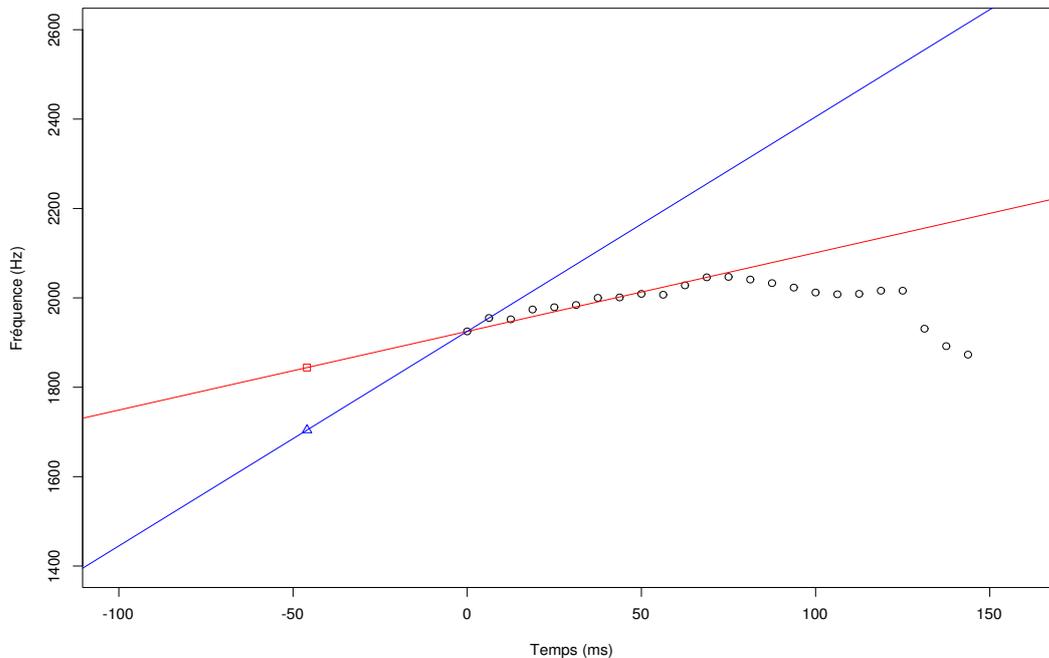


FIGURE 5.6 – Illustration du tracé formantique du F2 de la Figure 5.5 montrant la différence entre les deux méthodes de prédiction linéaire. Le carré rouge affiche la position du  $F2_{@burst}$  prédite par la méthode extrapolation 1 alors que le triangle bleu représente la valeur de  $F2_{@burst}$  estimée par l'extrapolation 2.

La Figure 5.6 illustre le tracé formantique du F2 affiché dans la Figure 5.5 (points entre  $F2_{onset}$  et  $F2_{offset}$ ). La droite de régression en rouge est estimée pour le premier et le dernier points ( $F2_{onset}$  et  $F2_{mid}$ ) (extrapolation 1) et le point de fréquence qui passe par cette droite à 46 ms avant le  $F2_{onset}$  est de 1844 Hz ( $F2_{@burst1}$ ). Quant à la droite de régression en bleu, elle est calculée pour les deux premiers points de fréquence du F2 (extrapolation 2) et le point de fréquence que la droite traverse à 46 ms est de 1702 Hz ( $F2_{@burst2}$ ). En effet, le  $F2_{@burst2}$  est plus proche de celui déterminé en suivant les consignes de Sussman et Shore (1996) (Figure 5.5) dans cet exemple. Toutefois, les valeurs de  $F2_{@burst}$  des deux méthodes seront prises en compte dans les analyses des équations de locus afin d'obtenir une comparaison globale entre les deux prédictions. Lorsque les variations des fréquences du F2 sont trop rapides (entre les deux premiers points et/ou entre  $F2_{onset}$  et  $F2_{mid}$ ) et/ou la durée du VOT est importante, des valeurs négatives de  $F2_{@burst}$  peuvent être trouvées dans les deux méthodes. Ces valeurs

10. Représenté par les points rouges entre le début et le fin de F2 de la voyelle dans la Figure 5.5.

extrêmes/négatives ont été détectées et supprimées par la même procédure expliquée dans la section précédente.

Il a été précisé plus haut que les valeurs de  $F2_{@burst}$  exposées dans cette expérience ont été prédites au moment (réel) du bruit de relâchement. Il faut toutefois savoir que lors de la phase préliminaire de nos analyses, les valeurs réelles du moment du relâchement n'étaient pas encore déterminées. Par conséquent, les  $F2_{@burst}$  ont été estimés à un moment fixe (à 30 ms) avant le début des transitions formantiques (voir Abuoudeh et Crouzet (2015)). Ces estimations préliminaires du  $F2_{@burst}$  à un moment fixe ont aidé à la mise en place du script R en attendant les valeurs réelles. De plus, elles ont fourni une idée générale des valeurs de  $F2_{@burst}$  et des équations de locus lorsqu'elles sont prédites à un moment proche du bruit de relâchement.

Dans cette expérience, seules les valeurs calculées au moment réel du bruit de relâchement ont été considérées.

### 5.1.4 Résultats

Cette section présente les résultats et les analyses des données de cette expérience en fonction des trois points principaux suivants : la durée vocalique, l'impact de la durée vocalique sur les fréquences des formants et l'influence de l'opposition phonologique sur les équations de locus. Concernant les deux derniers points, les analyses seront divisées et traitées séparément en fonction de la position syllabique de l'occlusive cible (initiale ou finale) et en fonction du sexe des locuteurs.

#### 5.1.4.1 La durée vocalique

Comme attendu, les voyelles longues ont bien des durées plus longues que les voyelles courtes. Les deux groupes de voyelles constituent deux ensembles temporels distincts avec une zone de chevauchement (Figure 5.7). La Figure 5.7 représente la distribution de la densité de la durée des voyelles courtes avec une moyenne globale de la durée de 67 ms et celle des voyelles longues avec une moyenne globale de 134 ms. La durée moyenne des voyelles courtes est ainsi égale à la moitié de celle des voyelles longues pour ces données (soit 50%). Il a également été constaté que les variations des durées des voyelles longues sont plus grandes que celles des voyelles courtes où l'écart-type global est de 22 ms pour les voyelles courtes et de 40 ms pour les longues. En effet, les voyelles longues sont plus étendues sur l'axe vertical (la durée) que les voyelles courtes, ces dernières se concentrant dans une zone plus réduite (densité plus élevée).

Il est à remarquer que malgré la différence importante de durée entre les deux catégories de voyelles, un intervalle de chevauchement non négligeable est observé entre les deux longueurs. Celui-ci s'étend globalement entre les deux moyennes des voyelles et le dépasse légèrement du côté des voyelles longues et de façon plus importante pour les voyelles courtes. De ce fait, une voyelle longue peut être réalisée avec une durée égale à celle d'une voyelle courte et inversement. Autrement dit, le passage d'une longueur à une autre se fait de manière continue et non

catégorielle.

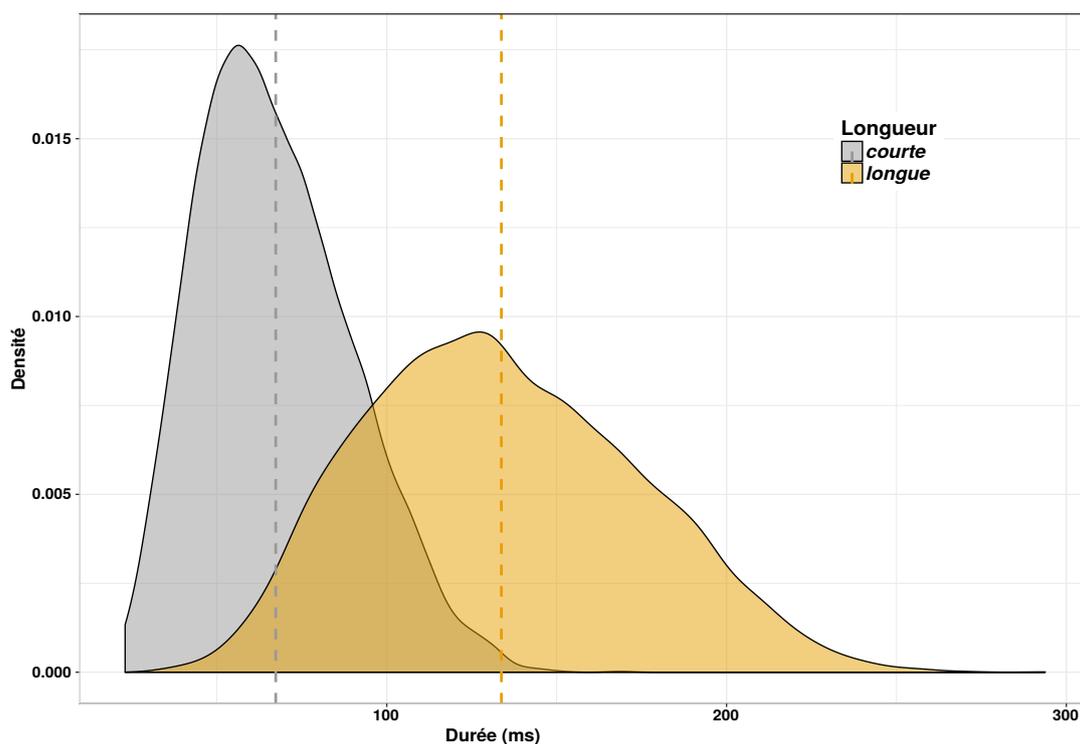


FIGURE 5.7 – Représentation de la densité et des moyennes (lignes en pointillé verticales) des durées des voyelles (en ms) en fonction des deux longueurs phonologiques.

L'analyse de la durée en fonction de la voyelle dévoile l'existence d'une croissance de la durée selon l'ouverture de la bouche pendant la réalisation de la voyelle (+ ouverte = + de durée au sein de chaque catégorie de longueur) (Table 5.1). En moyenne, la voyelle /a:/ est la plus longue et la voyelle /i:/ est la plus courte pour les voyelles longues. De même, la voyelle /a/ est la plus longue et la voyelle /i/ est la plus courte pour les voyelles courtes.

Sexe	Longueur/Timbre	/i/	/a/	/u/	/e/	/o/
M	longue	123 (39)	136 (39)	129 (35)	135 (41)	130 (39)
	courte	64 (23)	69 (22)	67 (23)	—	—
F	longue	127 (40)	156 (35)	127 (38)	141 (40)	132 (42)
	courte	65 (22)	71 (21)	68 (23)	—	—
Tous	longue	125 (39)	146 (38)	128 (37)	138 (41)	131 (40)
	courte	65 (23)	70 (22)	67 (23)	—	—

TABLE 5.1 – Valeurs moyennes et écart-types (entre parenthèses) des durées des cinq timbres vocaliques pour les locuteurs masculins (M) et féminins (F) et les moyennes des deux (Tous) selon la longueur vocalique.

D'importantes variations inter-locuteurs ont également été notées. Par exemple, les durées moyennes les plus courtes sont produites par le locuteur Am (45 ms pour les courtes vs. 89 ms pour les longues) et celles les plus longues par le locuteur Ah (94 ms pour les courtes vs. 172 ms

pour les longues), indiquant que la moyenne des voyelles courtes d'un locuteur peut dépasser celle des voyelles longues d'un autre locuteur (Figures 5.7 et 5.8). Ces variations inter-locuteurs sont dues aux différents débits de parole adaptés par chaque locuteur.

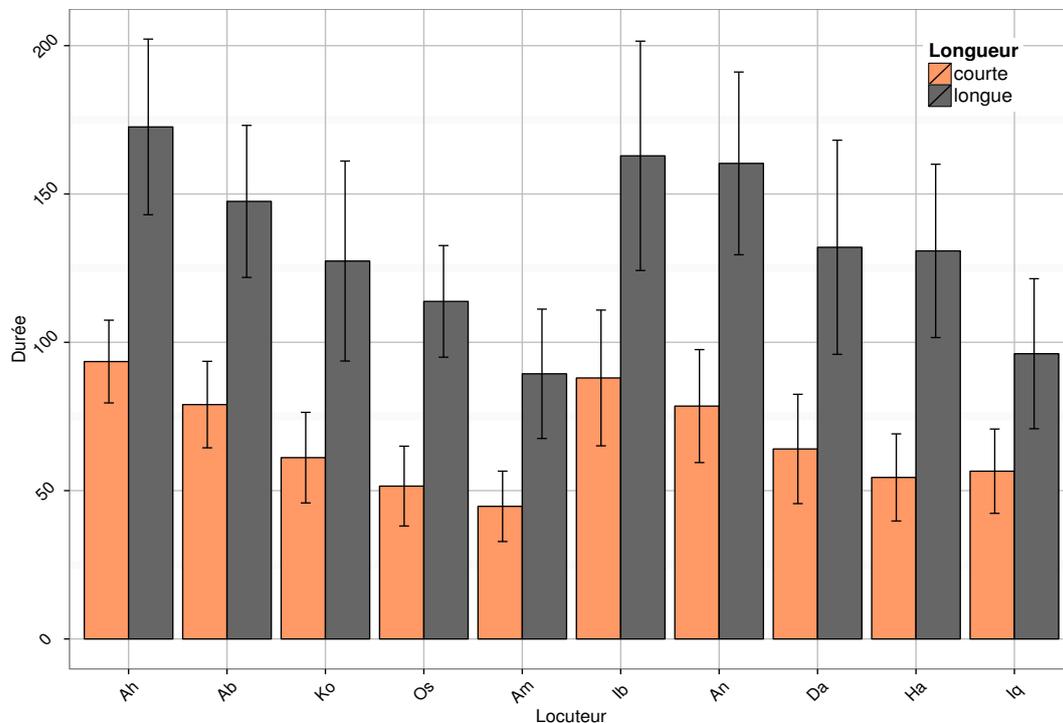


FIGURE 5.8 – Valeurs moyennes avec les barres de l'écart-type des durées vocaliques pour les locuteurs masculins (de Ah à Am) et féminins (de Ib à Iq).

Les données obtenues pour les voyelles longues dont le timbre n'est pas présent dans l'inventaire des voyelles courtes de la langue (/e:/, o:/) n'ont pas été retenues pour l'analyse de statistique inférentielle de la durée afin de ne pas déséquilibrer l'analyse.

Des analyses de variance (*anova*) de type III à mesures répétées (avec deux facteurs intra-locuteurs - longueur, sexe et timbre - et la variable locuteur comme facteur aléatoire) ont été effectuées pour comparer la durée des voyelles longues vs. courtes<sup>11</sup>. Les voyelles longues sont significativement plus longues que leurs contreparties courtes soit  $F_{(1,5)} = 57.72, p < 0.001$ . Aucune interaction entre la longueur et le timbre n'est observée, ce qui sous-entend que l'effet significatif de la longueur est valable pour les trois timbres. La Table 5.1 indique que les valeurs moyennes des deux groupes de locuteurs sont globalement similaires. Cette similarité est appuyée par l'absence d'effet d'interaction entre les variables "sexe" et "longueur" dans l'*anova* précédente.

11. Le plan d'analyse n'est pas équilibré à cause de l'élimination des valeurs extrêmes qui a entraîné un nombre différent d'effectifs pour chaque catégorie étudiée. Les *anovas* ont alors été effectuées sur les moyennes des quinze répétitions de manière à garantir l'équilibre du plan d'analyse. Ces *anovas* ont d'ailleurs été réalisées sur des données de pente et d'ordonnée à l'origine transformées numériquement (pentes élevées au carré, racine carrée des ordonnées à l'origine) afin de s'assurer qu'elles respectaient bien les conditions de validité de l'ANOVA (distribution normale, homogénéité des variances).

En résumé, les analyses de la durée vocalique des voyelles étudiées montrent l'existence de deux catégories de longueur distinctes pour tous les locuteurs malgré la présence d'un intervalle de chevauchement et la grande variation inter-locuteurs. La section suivante présente les propriétés spectrales pour chaque catégorie de longueur dans le but d'examiner si cette opposition de quantité est accompagnée d'une différence de qualité.

#### 5.1.4.2 Propriétés spectrales des formants

Les propriétés spectrales des formants (les  $F_{onsets}/F_{offsets}$  et les  $F_{mids}$ ) ont été examinées selon la position syllabique de la consonne cible et selon le sexe des participants. Par conséquent, les données ont été divisées en deux grands sous-ensembles selon la position syllabique (positions initiale et finale) et traitées par groupe de locuteur m et f (m pour masculin et f pour féminin). La division des données en fonction de la position syllabique permet d'étudier la co-articulation anticipatoire et persévérante entre les consonnes et les voyelles, alors que celle en fonction du sexe vise à éviter l'influence des différences anatomiques entre les deux groupes de locuteurs. Avant la division des données en fonction de la position syllabique, le plan F1-F2 a été analysé pour toutes les séquences (division m/f). Cette analyse permet de représenter les positions des voyelles longues vs. courtes étudiées et leurs chevauchements dans l'espace vocalique.

#### Plan F1-F2

L'espace vocalique des huit voyelles a été examiné selon le sexe des locuteurs (Figure 5.9). Les données affichées dans cette figure représentent les valeurs de  $F1_{mid}$  et  $F2_{mid}$  des voyelles dans toutes les séquences pour les locuteurs masculins (Figure 5.9.(a)) et féminins (Figure 5.9.(b)). Ces figures ont été réalisées à l'aide du package `phonR` (McCloy, 2016) dans R. Les ellipses permettent de repérer plus facilement la concentration des points de données, elles montrent l'intervalle de confiance dans la zone de la moyenne des réalisations de chaque voyelle. La valeur par défaut de l'intervalle de confiance utilisée correspond à 68% du niveau de confiance soit  $\pm 1$  d'écart-type du contour de la densité normale (McCloy, 2016). Plusieurs tendances globales peuvent être observées à partir de ces figures. La première est que la comparaison entre les deux groupes de locuteurs révèle une tendance générale de la distribution des voyelles similaire. La deuxième montre que les voyelles longues sont plus périphériques que leurs correspondantes courtes. La troisième concerne les écarts spectraux entre les voyelles longues et courtes. Par exemple, les zones de concentration des points semblent être superposées pour les voyelles /a, a:/ (écart minimal), contrairement au degré de chevauchement entre les voyelles /i, i:/ très faible (écart maximal). Toutefois, la voyelle /i/ se chevauche principalement avec la voyelle /e:/, tandis que la voyelle /u/ se chevauche avec le /o:/ et le /u:/ (écart moyen).

Ces observations sont en accord avec celles exploitées dans les études de Al-Tamimi (2007)

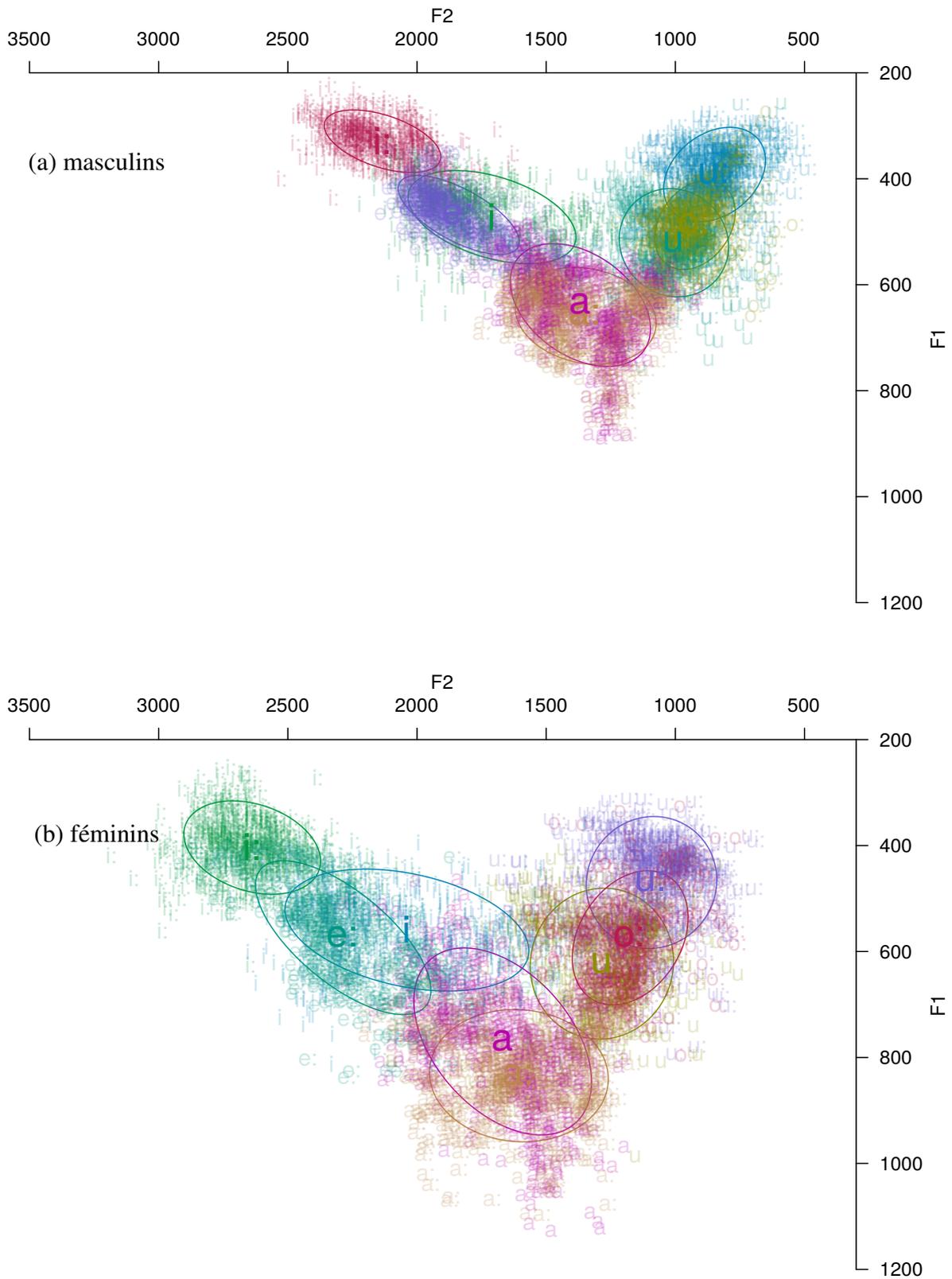


FIGURE 5.9 – Illustrations du plan F1-F2 (mesurés au milieu du formant) des 8 voyelles de l'arabe jordanien produites par les locuteurs masculins (a) et féminins (b).

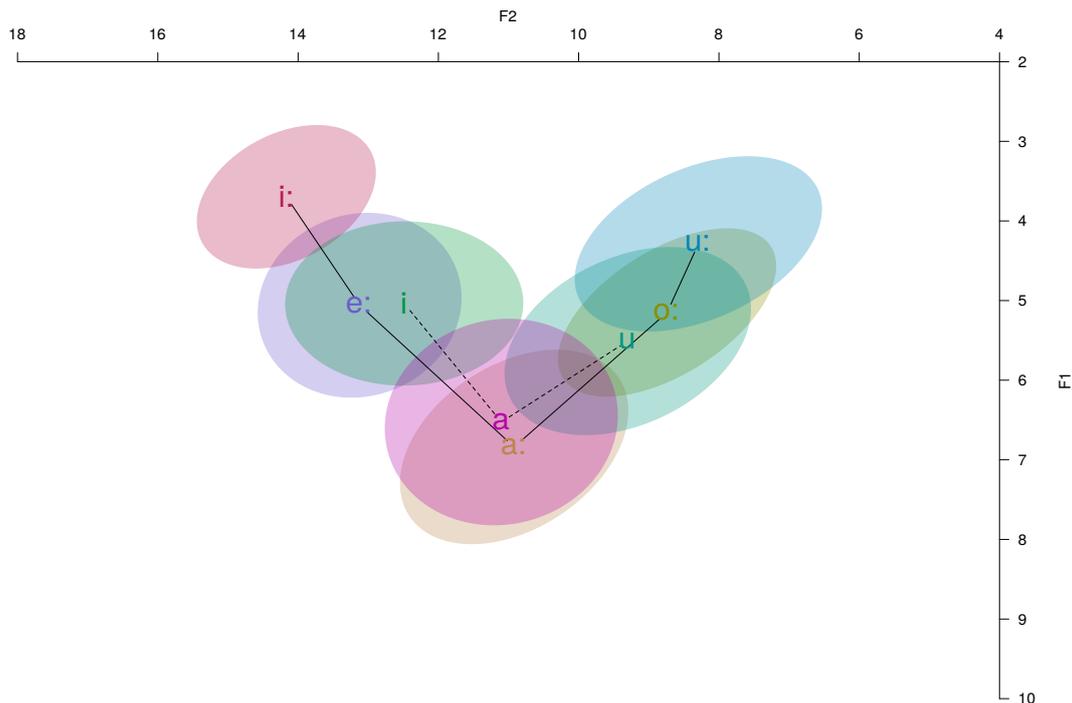


FIGURE 5.10 – Illustration de l'espace vocalique (F1-F2) de tous les locuteurs (m et f) en Bark.

et Allatif (2008) où l'écart spectral entre les voyelles longues *vs.* courtes augmente de la façon suivante : faible pour /a/-/a:/, moyen pour /u/-/u:/ et élevé pour /i/-/i:/. De ce fait, les voyelles courtes /i/ et /u/ sont plus proches acoustiquement des voyelles longues /e:/ et /o:/ que les voyelles /i:/ et /u:/, respectivement. Ces constatations soutiennent la proposition d'Al-Tamimi (2007) sur l'existence des /e/ et des /o/ courts plutôt que des /i/ et des /u/ (cf. 4.1.1.1).

La Figure 5.10 affiche l'espace vocalique de l'ensemble des données (m et f) après la transformation en Bark<sup>12</sup>. Elle représente une vision plus claire des tendances expliquées plus haut. Afin d'évaluer l'impact spectral de la longueur vocalique (courte *vs.* longue), deux séries d'analyses statistiques ont été effectuées selon le groupe de voyelles examinées, soit /i, i:, a, a:, u, u:/ d'un côté et /i, e:, a, a:, u, o:/ de l'autre. Cette procédure, appliquée sur chacun des sous-ensembles f et m, a été réalisée de manière à obtenir trois catégories de timbre comparables et afin ne pas déséquilibrer l'analyse car la variable longueur n'a qu'une seule catégorie pour chacun des timbres /e/ et /o/. Elle permet également de vérifier si les regroupements vocaliques observés visuellement (/i-e:/ et /u-o:/ dans les Figures 5.9 et 5.10) sont significativement similaires ou non. La Table 5.2 affiche les moyennes et les écart-types des F1 et F2 des deux sous-ensembles pour les deux groupes (m et f).

Les analyses de variance (*anovas*) à mesures répétées de type II ont été utilisées dans le but

12. La conversion de Hz à Bark est réalisée à l'aide de la formule de Traunmüller (1990) :

$$Bark = \frac{26.81 \times F}{1960 + F} - 0.53 \quad (5.1)$$

F1 :		Sous-ensemble 1			Sous-ensemble 2		
Longueur/V		i-i:	a-a:	u-u:	i-e:	a-a:	u-o:
M	courte	472 (58)	638 (76)	521 (67)	472 (58)	638 (76)	521 (67)
	longue	329 (39)	659 (60)	392 (59)	468 (49)	659 (60)	487 (55)
F	courte	560 (58)	770 (76)	623 (67)	560 (58)	770 (76)	623 (67)
	longue	404 (39)	834 (60)	470 (58)	574 (49)	834 (60)	574 (55)
F2 :							
M	courte	1710 (214)	1368 (179)	1006 (139)	1710 (214)	1368 (179)	1006 (139)
	longue	2134 (149)	1347 (179)	848 (130)	1839 (154)	1347 (179)	924 (102)
F	courte	2039 (214)	1668 (179)	1284 (139)	2039 (214)	1668 (179)	1284 (139)
	longue	2638 (149)	1606 (179)	1092 (130)	2285 (154)	1606 (179)	1175 (102)

TABLE 5.2 – Valeurs moyennes et écart-types (entre parenthèses) des F1 et des F2 pour les deux sous-ensembles étudiés pour les deux groupes de participants (Masculins et Féminins). Les trois voyelles courtes ainsi que le /a:/ sont présentes dans les deux sous-ensembles.

d'examiner l'impact de la longueur phonologique sur la fréquence des F1 et F2. La variable "locuteur" a été déterminée comme facteur aléatoire, les variables "timbre" et "longueur" comme variables indépendantes et "F1<sub>mid</sub>" ou "F2<sub>mid</sub>"<sup>13</sup> comme variable dépendante. Des analyses de comparaisons multiples des moyennes *Tukey Contrasts*<sup>14</sup> (analyses *post hoc*) ont ensuite été effectuées afin d'obtenir des détails sur la comparaison de chaque paire de voyelles au sein de chaque groupe de locuteurs (m et f) et de chaque variable dépendante (F1 et F2).

Les résultats des *anovas* sur le premier sous-ensemble de voyelles (/i, a, u, i:, a:, u:/) montrent que les fréquences des F1 et des F2 des voyelles longues *vs.* courtes sont significativement distinctes pour les locuteurs masculins, avec  $F_{(1,2)} = 100.96, p < 0.01$  pour F1 et  $F_{(1,2)} = 26.49, p < 0.05$  pour F2, et pour les locuteurs féminins, avec  $F_{(1,3)} = 58.70, p < 0.01$  pour F1 et  $F_{(1,3)} = 90.73, p < 0.01$  pour F2. Quant au second sous-ensemble de voyelles (/i, a, u, e:, a:, o:/), aucune différence significative n'est observée entre les voyelles longues et courtes pour les locuteurs masculins, avec  $F_{(1,3)} = 0.543, p = 0.515$  pour F1 et  $F_{(1,3)} = 3.24, p = 0.170$  pour F2, et féminins, avec  $F_{(1,3)} = 0.41, p = 0.569$  pour F1 et  $F_{(1,3)} = 0.90, p = 0.413$  pour F2. Il a également été constaté qu'une interaction du timbre existe avec la longueur vocalique sur F1 et sur F2 dans les analyses des deux sous-ensembles. Cette observation signifie que l'effet de la longueur est influencé par le timbre de la voyelle et que, par conséquent, les différentes paires de voyelles ne sont pas impactées aux mêmes degrés au sein de chaque sous-ensemble. Il faudra alors évaluer la différence entre chaque paire de voyelles et la taille de l'effet dans les deux sous-ensembles.

Les analyses *post hoc*, avec les combinaisons qui nous intéressent<sup>15</sup>, mettent en exergue

13. Deux tests ont été appliqués, l'un pour F1<sub>mid</sub> comme variable dépendante et l'autre pour F2<sub>mid</sub> comme variable dépendante.

14. Ces analyses ont été réalisées à l'aide de la fonction "lme" du package `nlme` et de la fonction "glht" du package `multcomp` dans R.

15. /a-a:/, /i-i:/, /u-u:/ pour le premier sous-ensemble et /a-a:/, /i-e:/, /u-o:/ pour le second.

que les fréquences de F2 des trois paires de voyelles dans chacun des deux sous-ensembles sont significativement différentes à l'exception de celles du F2 de la paire /a-a:/ des locuteurs masculins (dans les deux sous-ensembles) et celles du F1 de la paire /i-e:/ des locuteurs féminins. La Table 5.3 présente les résultats des analyses *post hoc* (Tukey) pour les cinq paires de voyelles longues vs. courtes étudiées. Elle indique que toutes les paires de voyelles sont statistiquement différentes (à part les deux exceptions citées ci-dessus). Un point important à remarquer est que les tailles d'effet pour chaque paire sont assez distinctes, c'est-à-dire que certaines paires ont des tailles d'effet plus petites et d'autres plus grandes. En effet, la taille d'effet permet de définir « [...] *the degree to which the phenomenon is present in the population or the degree to which the null hypothesis is false* »<sup>16</sup>. Donc, « [...] la taille de l'effet décrit dans quelle mesure l'hypothèse nulle est fautive : plus la taille de l'effet est grande, plus il est justifié de rejeter l'hypothèse nulle »<sup>17</sup>.

La comparaison entre les tailles d'effet des paires /i-i:/ et /i-e:/ d'un côté et entre les paires /u-u:/ et /u-o:/ de l'autre donne une indication de l'étendue de la différence ou de la similarité entre les voyelles longues et courtes au sein de chaque paire au niveau du F1 et du F2. Les tailles d'effet reflètent alors les différences remarquées visuellement sur l'espace vocalique dans la Figure 5.10. Au niveau du F1, la voyelle /i:/ a des fréquences significativement inférieures à celles du /i/ avec une taille d'effet élevée. Cependant, la voyelle /e:/ a des fréquences qui ne sont pas significativement différentes avec la voyelle /i/ pour les deux groupes de locuteurs (Table 5.3). Les voyelles /u:/ et /o:/, quant à elles, ont des fréquences du F1 significativement inférieures à celles du /u/. Toutefois, la taille d'effet de la comparaison entre /u/ et /u:/ est nettement supérieure à celle entre /u/ et /o:/. Au niveau du F2, les voyelles /i:/ et /e:/ ont des fréquences significativement plus élevées que celles du /i/, néanmoins la taille de l'effet est plus réduite pour la paire /i-e:/. De la même manière, les voyelles /u:/ et /o:/ ont des fréquences de F2 significativement plus basses que celles du /u/ mais la taille de l'effet est plus élevée pour la paire /u-u:/ que celle de /u-o:/.

Ces observations révèlent que les voyelles longues /e:/ et /o:/ sont plus proches spectralement que les voyelles /i:/ et /u:/ de leurs correspondantes courtes /i/ et /u/ respectivement, malgré les différences significatives de moyennes. En ce qui concerne la paire /a-a:/, présente dans les deux sous-ensembles, les analyses montrent que les fréquences du F1 de la voyelle longue sont statistiquement plus élevées que celles de la courte. De plus, les fréquences du F2 de la voyelle longue sont légèrement plus basses que celles de la voyelle courte avec une différence non significative pour le groupe des hommes. Bien que la taille de l'effet soit globalement limitée, elle est plus importante pour le groupe des femmes (approximativement, le triple pour F1 et le double pour F2).

Ces analyses statistiques mettent en avant que les voyelles longues /i:, a:, u:/ se différencient

16. « [...] le degré auquel le phénomène est présent dans la population ou le degré auquel l'hypothèse nulle est fautive » Cohen (1988), page 9 [notre traduction].

17. Bourque et al. (2009), page 216.

		F1			F2		
	sexe	taille de l'effet	valeur z	valeur p	taille de l'effet	valeur z	valeur p
i-i:	m	144.95	46.82	p<.001	426.05	47.09	p<.001
	f	156.43	14.49	p<.001	595.44	51.70	p<.001
u-u:	m	129.23	48.43	p<.001	156.81	16.79	p<.001
	f	154.02	33.40	p<.001	193.39	16.29	p<.001
i-e:	m	3.30	1.04	<b>p=.904</b>	127.89	15.87	p<.001
	f	0.56	2.43	<b>p=.145</b>	248.52	21.76	p<.001
u-o:	m	33.96	10.23	p<.001	81.97	9.21	p<.001
	f	48.94	10.09	p<.001	110.08	9.09	p<.001
a-a:	m	21.28	6.98	p<.001	21.72	2.43	<b>p=.144</b>
	f	64.80	14.49	p<.001	61.34	5.33	p<.001

TABLE 5.3 – Résultats des tests *post hoc* des paires de voyelles étudiées sur F1 et sur F2 pour les locuteurs masculins et féminins.

globalement des voyelles courtes (/i, a, u/) au niveau spectral (F1 et F2) et que les voyelles longues /e:, a:, o:/ sont globalement similaires aux voyelles courtes (/i, a, u/). Les résultats sont semblables pour les deux groupes de locuteurs (m et f). Les comparaisons par paires révèlent bien qu'il existe une différence de moyennes entre la plupart des combinaisons des deux sous-ensembles mais la taille de l'effet est beaucoup plus réduite pour les paires du premier sous-ensemble. Ces résultats laissent suggérer que les voyelles en arabe jordanien peuvent être divisées en deux sous-ensembles contenant des voyelles longues et courtes. Dans le premier, les voyelles courtes se distinguent au niveau temporel et spectral de leurs correspondantes longues alors que dans le second, les voyelles courtes s'opposent uniquement au niveau temporel et non spectral. Ces deux sous-ensembles permettent d'étudier la relation coarticulatoire des cinq occlusives avec les voyelles dans deux contextes différents. L'un contient une opposition de longueur et spectrale et l'autre comprend une opposition de longueur uniquement. L'exploration des deux sous-ensembles vont permettre de voir si les deux contextes impactent la coarticulation au même degré et si la durée vocalique seule peut influencer la coarticulation.

Dans la section suivante, les données seront divisées en deux grands sous-ensembles selon la position syllabique de l'occlusive cible. Cette division permet d'évaluer l'impact réciproque de la voyelle et de la consonne lorsque celle-ci est en position initiale et finale. Parallèlement, l'impact de la longueur vocalique sur la fréquence des consonnes sera également examiné.

### 5.1.4.3 La coarticulation en position initiale (CV)

Les fréquences des trois premiers formants au début ( $F_{onset}$ ) et au milieu ( $F_{mid}$ ) de la voyelle ont été analysées pour chaque séquence afin d'observer l'impact de la longueur vocalique sur la consonne initiale et sur la voyelle. Les deux sous-ensembles de voyelles (/i, i:, a, a:, u, u:/ et /i, e:, a, a:, u, o:/) ont été évalués séparément pour chaque groupe de locuteurs (m et f). Les analyses statistiques indiquent que pour le premier sous-ensemble de voyelles, les fréquences au milieu

des trois formants sont significativement différentes en fonction de la longueur vocalique pour les données des hommes, avec  $F_{(1,2)} = 46.93, p < 0.05$ ;  $F_{(1,2)} = 179.27, p < 0.01$ ;  $F_{(1,2)} = 48.85, p < 0.05$ , et dans celles des femmes, avec  $F_{(1,3)} = 12.78, p < 0.05$ ;  $F_{(1,3)} = 439.73, p < 0.001$ ;  $F_{(1,3)} = 10.70, p < 0.05$ , respectivement pour  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$ . Un effet principal de la longueur vocalique a également été observé pour les fréquences au début des trois premiers formants dans les données des hommes – bien que l'effet soit marginal pour les  $F3_{onset}$  – avec  $F_{(1,2)} = 32.67, p < 0.05$ ;  $F_{(1,2)} = 756.1, p < 0.01$ ;  $F_{(1,2)} = 16.55, p = 0.554$  et pour celles des femmes, avec  $F_{(1,3)} = 145.84, p < 0.01$ ;  $F_{(1,3)} = 123.37, p < 0.01$ ;  $F_{(1,3)} = 24.12, p < 0.05$ , respectivement pour  $F1_{onset}$ ,  $F2_{onset}$  et  $F3_{onset}$ . Quant au second sous-ensemble de voyelles, la longueur vocalique n'a pas d'effet significatif sur les fréquences au milieu des trois formants à l'exception de  $F1_{mid}$  pour les femmes et de  $F2_{mid}$  pour les hommes, soit  $F_{(1,3)} = 2.23, p = 0.232$ ;  $F_{(1,3)} = 10.76, p < 0.05$ ;  $F_{(1,3)} = 0.047, p = 0.105$  pour les données des hommes et soit  $F_{(1,3)} = 11.77, p < 0.05$ ;  $F_{(1,3)} = 1.92, p = 0.26$ ;  $F_{(1,3)} = 0.464, p = 0.545$  pour celles des femmes, respectivement pour  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$ . L'influence de la longueur vocalique sur les débuts des formants n'est pas significative pour les données des hommes et celles des femmes, à l'exception de  $F2_{onset}$  des données des hommes, avec  $F_{(1,3)} = 0.937, p = 0.405$ ;  $F_{(1,3)} = 72.05, p < 0.01$ ;  $F_{(1,3)} = 0.217, p = 0.673$  pour les données des hommes et avec  $F_{(1,3)} = 0.006, p = 0.943$ ;  $F_{(1,3)} = 5.09, p = 0.109$ ;  $F_{(1,3)} = 1.70, p = 0.284$  pour celles des femmes, respectivement pour  $F1_{onset}$ ,  $F2_{onset}$  et  $F3_{onset}$ .

Les résultats globaux de l'impact de la longueur vocalique sur le début et le milieu des formants sont en accord avec les analyses des F1-F2 dans la section précédente. Autrement dit, la différence de qualité entre les voyelles longues et courtes est plus importante dans le premier sous-ensemble par comparaison à celle dans le second. Ces analyses laissent penser que, quand il existe une différence significative de qualité vocalique ( $F_{mids}$ ) entre les voyelles courtes et les voyelles longues (sous-ensemble 2), l'impact de la longueur est également significative sur les consonnes ( $F_{onsets}$ ). De même, lorsque la qualité vocalique est globalement similaire entre les voyelles longues et courtes (sous-ensemble 2), l'impact de la longueur est faible sur les consonnes. Il est à remarquer que dans les analyses des deux sous-ensembles, un effet d'interaction du timbre avec la longueur a été observé dans toutes les *anovas* effectuées, à l'exception de ceux du  $F1_{onset}$  et du  $F3_{onset}$  des données des hommes et de ceux du  $F3_{mid}$  et du  $F1_{onset}$  des données des femmes. En d'autres termes, les voyelles ne sont pas influencées au même degré par la longueur phonologique. Les analyses *post hoc* montrent également que les différences entre les voyelles longues vs. courtes au début et au milieu des formants sont plus importantes entre les paires du premier sous-ensemble de voyelles.

### Analyses de dispersion

Un intérêt particulier a été porté sur les fréquences de  $F2_{mid}$  et de  $F2_{onset}$  afin d'examiner l'impact de la longueur vocalique, de la consonne et de la voyelle sur la coarticulation CV. La Figure 5.11 illustre la distribution des changements de la qualité spectrale des voyelles

( $F_{mid}$ ) en association avec les variations de la durée vocalique dans les contextes consonantiques étudiés pour les hommes (en haut) et les femmes (en bas). Les voyelles avec des cercles transparents représentent les voyelles longues et celles sans cercles les voyelles courtes. Les droites de régression indiquent la tendance globale de la fréquence de chaque timbre lorsqu'il passe de voyelle longue à courte et/ou d'une durée vocalique longue à une autre courte au sein de la même longueur vocalique. Cette représentation permet également d'observer l'influence des consonnes sur les voyelles longues et courtes ; autrement dit, la résistance coarticulatoire des voyelles et des consonnes peut être remarquée visuellement grâce à cette figure.

Les deux groupes de locuteurs affichent les mêmes tendances générales des changements spectraux en fonction du temps dans tous les contextes consonantiques à l'exception de la voyelle /e:/ avec la pharyngalisée /t<sup>h</sup>/ et l'uvulaire /q/. Cette voyelle semble être moins impactée par la durée vocalique et/ou par la pharyngalisation et l'uvularisation chez les femmes que chez les hommes. Cette constatation est attendue puisque les hommes ont tendance à pharyngaliser ou à uvulariser les voyelles plus que les femmes en arabe jordanien (Al-Tamimi, 2001; Al-Wer et Herin, 2011).

Pour les analyses numériques, l'écart-type a été utilisé dans le but d'obtenir des mesures numériques de l'observation visuelle qui offrent le degré de dispersion des valeurs des fréquences des voyelles. Par conséquent, un écart-type élevé d'une voyelle (+ de dispersion) signifie que cette dernière est sensible aux contextes consonantiques et donc qu'elle a un degré de résistance coarticulatoire faible. De la même façon, si l'écart-type d'une voyelle est bas, cela indique que la voyelle résiste aux influences des segments consonantiques et qu'elle garde une stabilité élevée malgré l'influence du contexte consonantique (Recasens, 1985). Généralement, les voyelles longues sont plus étendues verticalement (spectralement) et horizontalement (temporellement) que celles courtes. L'écart-type global de la durée est de 22 et 40 ms pour les voyelles courtes et longues respectivement et celui du  $F2_{mid}$  de 397 et 608 Hz. Autrement dit, les voyelles courtes sont plus concentrées que celles longues sur les deux axes. Pour les deux sous-ensembles de voyelles, celles courtes sont plus impactées que leurs correspondantes longues par le contexte consonantique. Elles ont une moyenne des écart-types de 177 Hz pour les hommes et 240 Hz pour les femmes<sup>18</sup>. Pour le premier sous-ensemble, la moyenne des écart-types des voyelles longues est de 153 Hz pour hommes et de 189 Hz pour les femmes. Quant au second sous-ensemble, les voyelles longues ont une moyenne d'écart-type de 145 Hz pour les hommes et de 200 Hz pour les femmes. De ce fait, la dispersion plus importante des voyelles longues signifie qu'elles sont plus résistantes aux différentes influences consonantiques que leurs correspondantes courtes.

Les voyelles postérieures /u, u:, o:/ sont moins impactées par le contexte consonantique que les voyelles centrales et antérieures (Table 5.4). Cette observation peut être expliquée par le fait que lors de la production des voyelles postérieures, le geste du corps de la langue n'est

18. L'écart-type de chaque voyelle produite dans tous les contextes a d'abord été calculé (Table 5.4) puis, le total de ces écart-types a été moyenné pour chaque catégorie de longueur des deux sous-ensembles.

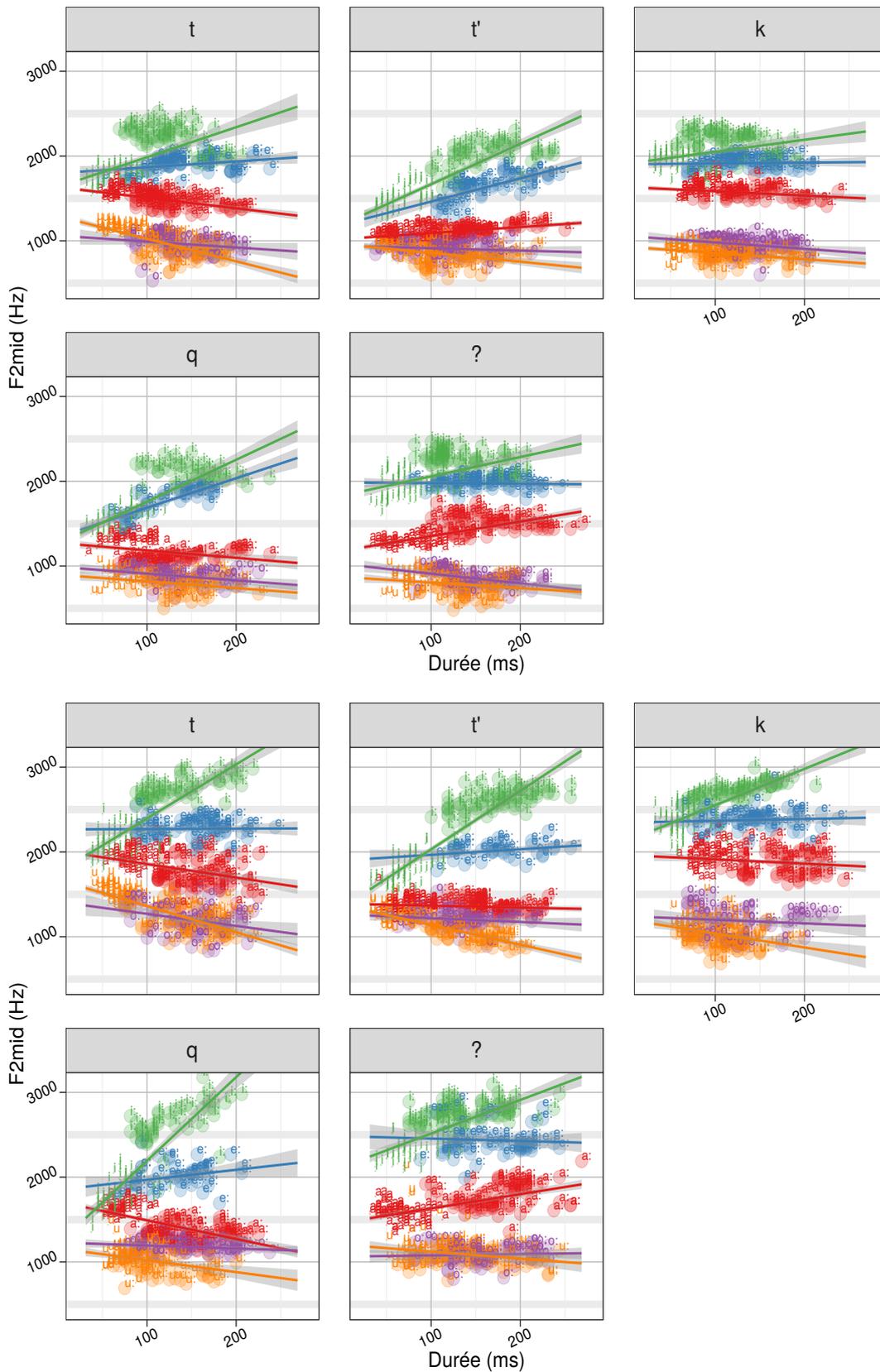


FIGURE 5.11 – Représentation de la variation du  $F2_{mid}$  en fonction du temps dans chaque contexte consonantique initial pour les locuteurs masculins (en haut) et féminins (en bas).

	i	i:	a	a:	u	u:	e:	o:
M	214	149	179	179	139	130	154	102
F	311	174	227	228	183	166	224	148
Moyenne	263	162	203	204	161	148	189	125

TABLE 5.4 – Valeurs des écart-types du  $F2_{mid}$  des 8 voyelles pour les participants masculins (M) et féminins (F) et la moyenne des deux groupes.

pas en conflit avec les gestes des consonnes pharyngalisée et uvulaire alors qu'il l'est dans la réalisation des voyelles antérieures et centrales (Recasens et al., 1997; Recasens et Espinosa, 2009) (*cf.* 2.3.5). Ce phénomène provoque un abaissement de la fréquence du F2 plus important pour les voyelles antérieures par comparaison avec celui des voyelles postérieures.

Le degré de l'impact de la consonne sur le milieu des voyelles peut également être estimé par le calcul des écart-types de  $F2_{mid}$  des voyelles dans chaque contexte consonantique. Une valeur élevée de l'écart-type (signalant une dispersion maximale) indique une influence faible de la consonne sur  $F2_{mid}$  et inversement. Pour les deux groupes de locuteurs, l'occlusive pharyngalisée /tʰ/ est la consonne la plus influente des cinq occlusives (Table 5.5). Par contre, comme cité précédemment, une différence de pharyngalisation et d'uvularisation est attendue selon le sexe des locuteurs. L'uvulaire /q/ et l'alvéolaire /t/ n'occupent pas le même classement de degré de résistance dans les deux groupes. Pour les locuteurs masculins, l'uvulaire est plus dominante que l'alvéolaire tandis que pour les femmes, c'est l'alvéolaire qui est la plus puissante. Comme attendu, l'occlusive glottale /ʔ/ est la consonne la moins influente des cinq pour les deux groupes de locuteurs avec les écart-types les plus élevés. Quant à la vélaire /k/, le degré de dispersion la classe en quatrième position pour les deux groupes de locuteurs.

	$F2_{mid}$					$F2_{onset}$				
	tʰ	t	q	ʔ	k	tʰ	t	q	ʔ	k
M	387	442	423	510	462	196	270	349	493	523
F	489	523	540	612	572	269	323	456	585	642
Moyenne	438	483	482	561	517	233	297	403	539	583

TABLE 5.5 – Valeurs des écart-types des  $F2_{mid}$  et des  $F2_{onset}$  dans le contexte des cinq occlusives pour les participants masculins (M) et féminins (F).

Le début du deuxième formant représente la partie qui porte le plus d'informations et qui est la plus influencée par les consonnes. De ce fait, lorsque les fréquences du  $F2_{onset}$  d'une consonne produite avec plusieurs voyelles sont concentrées dans une zone de fréquences réduite, cela signifie que cette consonne influence au maximum le début de la voyelle. Autrement dit, la consonne "oblige" les différentes voyelles à se concentrer près de son locus. La Figure 5.12 affiche les fréquences du  $F2_{onset}$  des huit voyelles en fonction de la durée vocalique dans les cinq contextes consonantiques pour les locuteurs masculins (en haut) et féminins (en bas).

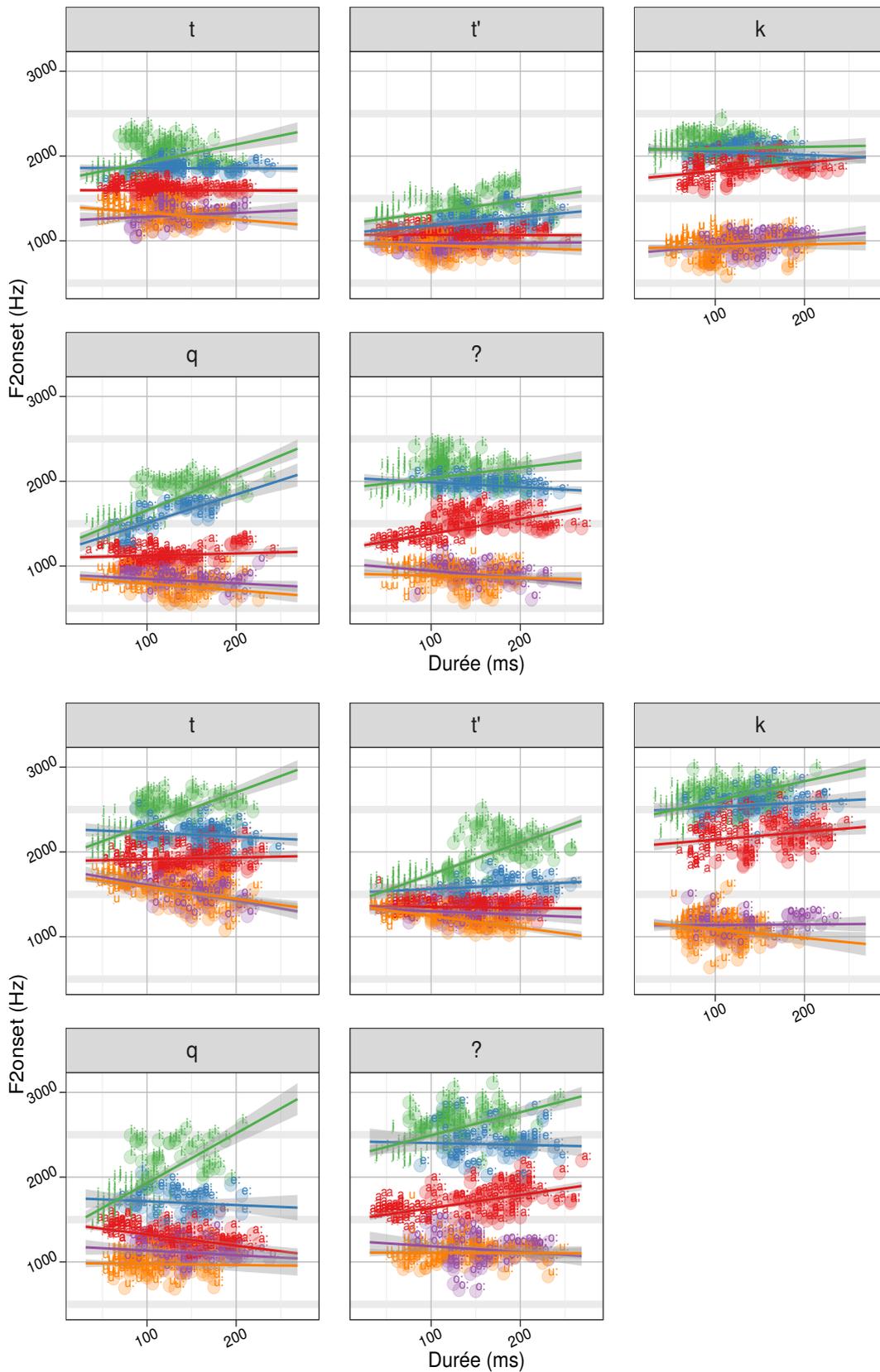


FIGURE 5.12 – Représentation de la variation du  $F2_{onset}$  en fonction du temps dans chaque contexte consonantique initial pour les locuteurs masculins (en haut) et féminins (en bas).

L'observation de cette figure révèle que les occlusives peuvent être classées en fonction de la concentration des  $F2_{onset}$  des voyelles. En première place, l'occlusive alvéolaire pharyngalisée (/t<sup>h</sup>/) est la consonne la plus puissante avec une concentration de points très élevée dans les deux groupes de locuteurs reflétée par une dispersion (écart-type) faible (Table 5.4). L'alvéolaire /t/ est en deuxième place au niveau de la concentration de points avec un écart-type de 270 Hz pour les hommes et de 323 Hz pour les femmes. Il apparaît donc que les alvéolaires sont les occlusives les plus résistantes<sup>19</sup> et par conséquent, qu'elles ont les degrés de coarticulation les plus bas. L'occlusive qui occupe la troisième place au niveau de la concentration des points est l'uvulaire /q/ avec un écart-type de 349 Hz pour les hommes et de 456 Hz pour les femmes. La glottale /ʔ/, quant à elle, a un écart-type de 493 Hz pour les hommes et de 585 Hz pour les femmes la plaçant en quatrième position au niveau de la concentration des points. Cela veut dire que cette occlusive est complètement dépendante de la voyelle suivante. Toutefois, l'occlusive qui a les  $F2_{onset}$  les plus dispersés est la vélaire /k/ avec un écart-type de 523 Hz pour les hommes et de 642 Hz pour les femmes. Cette dispersion élevée est provoquée par l'existence des deux allophones (palatal et vélaire) de cette occlusive en fonction de la voyelle suivante (antérieure et postérieure). Les fréquences de la vélaire /k/ ( $F2_{onset}$ ) produite avec les voyelles antérieures et centrales (/i, i:, a, a:, e:/) sont concentrées dans une zone de fréquences élevées alors que celle prononcée avec les voyelles postérieures a des fréquences de  $F2_{onset}$  qui se concentrent dans une zone de fréquences basses. Ce grand écart entre les deux zones de concentration de points entraîne les valeurs élevées de dispersion (les deux nuages de points affichés dans la Figure 5.12). D'après ces résultats, la résistance coarticulatoire des consonnes avec les voyelles suivantes diminue dans cet ordre : t<sup>h</sup> > t > q > ʔ > k. En d'autres mots, le degré de coarticulation augmente de t<sup>h</sup> < t < q < ʔ < k.

En résumé, les résultats présentés dans cette section soutiennent l'existence de deux sous-ensembles de voyelles où le premier s'oppose au niveau spectral et temporel et le second au niveau temporel seulement. Pour le premier groupe, les débuts (et les milieux) des trois formants sont significativement influencés par la longueur phonologique de la voyelle alors que pour le second, les débuts (et la plupart des milieux) du deuxième formant sont statistiquement similaires pour les voyelles longues et courtes. L'examen de la distribution des points de  $F2_{mid}$  en fonction du temps a mis en avant que les voyelles courtes sont les plus impactées par le contexte consonantique et que les voyelles postérieures sont les moins influencées par l'ensemble des consonnes étudiées. Les analyses de dispersion sur les  $F2_{onset}$  ont révélé que les occlusives alvéolaires résistent le plus au contexte vocalique, contrairement à la vélaire et à la glottale qui résistent le moins. En effet, les occlusives glottale et vélaire ont le degré de coarticulation le plus élevé (dispersion maximale) mais pour deux raisons différentes. La grande dispersion des vélaire est provoquée par l'engagement du dos de la langue lors de sa réalisation. Elle est donc influencée - à différents degrés - par la direction et le degré de l'engagement du corps de la langue de la voyelle suivante. Par contre, la dispersion élevée de la glottale est

19. La plus résistante des deux est celle réalisée avec une double articulation.

due au fait que la réalisation de cette occlusive n'entraîne pas de mouvement lingual et par conséquent, la langue est libre pendant la production des voyelles suivantes.

En général, les analyses de dispersion sur les fréquences de  $F2_{mid}$  et  $F2_{onset}$  permettent de prédire les valeurs des équations de locus. Selon les résultats présentés plus haut, les valeurs de pentes doivent être élevées pour la vélaire et la glottale (proche de 1), intermédiaires pour l'uvulaire et basses pour les alvéolaires (proche de 0).

#### 5.1.4.4 Les équations de locus (CV)

Les équations de locus ont été estimées selon trois groupes de voyelles. La première estimation, basée sur la réalisation des huit voyelles (approximativement 108 points de mesures par locuteur pour chaque pente), offre l'opportunité d'avoir une idée globale du degré de coarticulation des cinq occlusives sans la prise en compte de l'impact temporel (voyelles courtes vs. voyelles longues). La deuxième porte d'un côté sur les voyelles longues et de l'autre sur celles courtes du premier sous-ensemble (/i, a, u/ vs. /i:, a:, u:/). Cette estimation permet de comparer l'effet de la longueur phonologique sur la coarticulation pour des voyelles qui s'opposent en durée et au niveau spectral. La troisième concerne également la comparaison entre les voyelles longues et les voyelles courtes mais cette fois-ci du second sous-ensemble (/i, a, u/ vs. /e:, a:, o:/). Elle permet d'évaluer l'impact de la longueur phonologique sur la coarticulation pour des voyelles qui diffèrent au niveau temporel mais similaires au niveau spectral.

Pour chacune de ces estimations, les équations de locus ont été générées trois fois en fonction de la position du  $F2_{onset}$ . La première est calculée avec le début du F2 ( $F2_{onset}$ ) et les deux autres ont été estimées avec les  $F2_{onset}$  prédits au moment du relâchement de la consonne ( $F2_{@burst}$ ) selon les deux méthodes d'estimation expliquées dans la méthodologie.

#### Première estimation : avec toutes les voyelles

Dans les analyses précédentes (sur les fréquences), les deux groupes de locuteurs ont été traités séparément afin d'éviter les variations liées au sexe des locuteurs. Dans les analyses suivantes sur les équations de locus, les données des deux groupes de locuteurs (femmes et hommes) ont été combinées puisque la différence entre les pentes et les ordonnées à l'origine des deux groupes est faible (Figure 5.13). Des *anovas* de type 1 à mesures répétées ont été effectuées de manière à vérifier si cette similarité est statistique ou non. Dans ces tests, le facteur "locuteur" a été fixé comme la variable aléatoire, le facteur "sexe" comme la variable indépendante et ceux "pente" et "ordonnée à l'origine" comme des variables dépendantes. Ces tests ont montré une absence d'effet lié au sexe sur les valeurs des pentes soit  $F_{(1,8)} = 0.113, p = 0.746$ ;  $F_{(1,8)} = 0.214, p = 0.656$  respectivement pour la pente et l'ordonnée à l'origine<sup>20</sup>. En effet, ces

20. Ces tests ont également été effectués sur les équations de locus calculées avec les  $F2_{@burst1}$  et les  $F2_{@burst2}$ . Les mêmes résultats ont été obtenus pour les deux calculs (différence non significative) soit  $F_{(1,8)} = 0.084, p = 0.779$ ;  $F_{(1,8)} = 1.184, p = 0.308$  pour  $F2_{@burst1}$  et  $F_{(1,8)} = 0.086, p = 0.777$ ;  $F_{(1,8)} = 1.219, p = 0.302$  pour  $F2_{@burst2}$ , respectivement pour la pente et l'ordonnée à l'origine.

résultats justifient la combinaison des équations de locus des deux groupes.

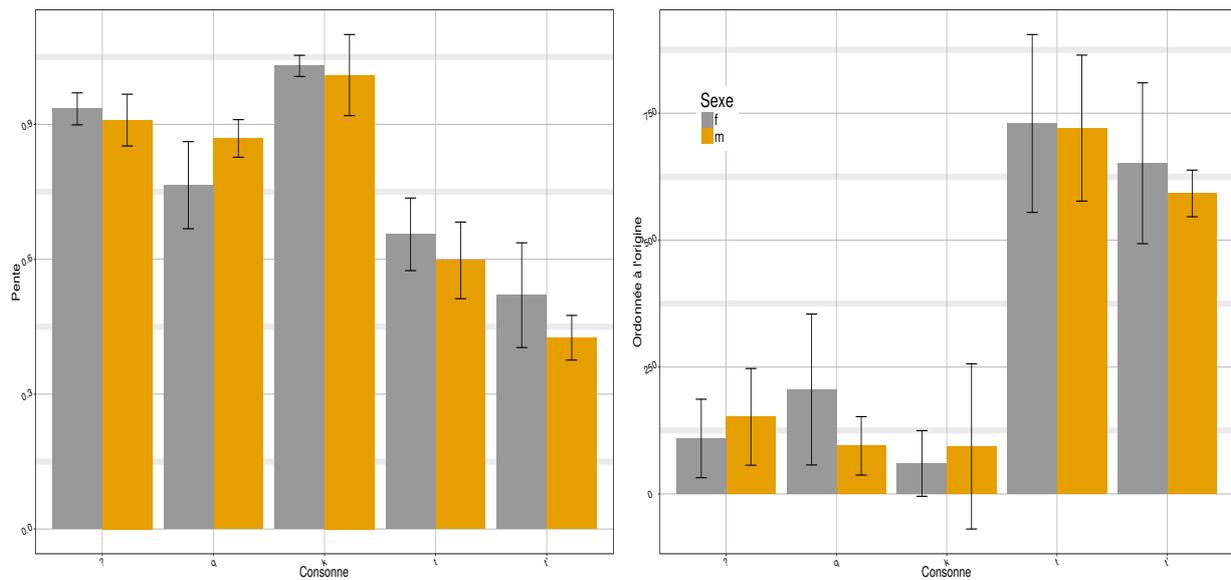


FIGURE 5.13 – Graphiques des moyennes des pentes (à gauche) et des ordonnées à l'origine avec les barres de l'écart-type des équations de locus calculées avec les  $F2_{onset}$  des cinq occlusives pour les locuteurs féminins (f) et masculins (m).

La Table 5.6 affiche les moyennes des pentes, des ordonnées à l'origine et des  $R^2$  des équations de locus calculées avec  $F2_{onset}$ ,  $F2_{@burst1}$  et  $F2_{@burst2}$ <sup>21</sup> des dix locuteurs. Les résultats globaux des équations de locus (avec  $F2_{onset}$  et  $F2_{@burst}$ ) sont en accord avec ceux des analyses de dispersion présentées plus haut, c'est-à-dire que les pentes augmentent de la manière suivante :  $t^h < t < q < k < ʔ$  (Figure 5.13 et Table 5.6). Cet ordre est le même pour les trois estimations des équations de locus ( $F2_{onset}$ ,  $F2_{@burst1}$  et  $F2_{@burst2}$ ). Cela signifie que l'alvéolaire pharyngalisée est l'occlusive qui permet le moins la coarticulation (la plus résistante) et la vélaire est l'occlusive qui permet le plus la coarticulation (la moins résistante). Concernant les ordonnées à l'origine, elles ont les valeurs les plus élevées pour les alvéolaires et les moins élevées pour la vélaire. Il est à remarquer d'ailleurs que les écart-types des ordonnées à l'origine sont relativement élevées notamment pour les occlusives /ʔ, q, k/ qui, parfois, dépassent leur moyenne. Cette observation montre que l'ordonnée à l'origine est très sensible aux variations inter-locuteurs. À propos de la qualité de la prédiction des équations de locus, les  $R^2$  révèlent des valeurs relativement élevées dans la plupart des cas (majoritairement  $> 0.85$ ), indiquant un ajustement élevé des nuages de points pour les pentes des équations de locus. Les valeurs de  $R^2$  sont plus élevées pour les équations de locus calculées avec les  $F2_{onset}$  (moyenne = 0.94). Quant aux  $R^2$  des équations de locus estimées avec les  $F2_{@burst}$ , ils sont légèrement plus élevés pour  $F2_{@burst2}$  que ceux de  $F2_{@burst1}$  (moyenne = 0.84 vs. 0.80).

Comme cité plus haut, la comparaison des trois types d'équations de locus dans la Table 5.6 montre une similarité de la tendance globale des pentes et des ordonnées à l'origine. Pourtant,

21. Pour rappel, le  $F2_{@burst1}$  est estimé avec la régression linéaire qui passe par  $F2_{onset}$  et  $F2_{mid}$  alors que le  $F2_{@burst2}$  est prédit avec la régression linéaire qui passe par  $F2_{onset}$  et le point suivant.

C	F2 <sub>onset</sub>			F2 <sub>@burst1</sub>			F2 <sub>@burst2</sub>		
	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
t <sup>ɿ</sup>	0.47 (0.10)	623 (114)	0.90 (0.05)	0.33 (0.13)	793 (154)	0.69 (0.23)	0.44 (0.15)	660 (181)	0.78 (0.19)
t	0.63 (0.08)	730 (152)	0.94 (0.03)	0.42 (0.12)	1123 (234)	0.71 (0.19)	0.50 (0.13)	998 (233)	0.79 (0.13)
q	0.81 (0.09)	121 (144)	0.94 (0.03)	0.80 (0.14)	135 (196)	0.84 (0.11)	0.83 (0.10)	102 (180)	0.86 (0.10)
ʔ	0.92 (0.05)	141 (94)	0.98 (0.02)	0.89 (0.06)	171 (108)	0.92 (0.08)	0.91 (0.05)	145 (101)	0.92 (0.07)
k	1.02 (0.06)	87 (133)	0.94 (0.03)	1.06 (0.12)	97 (230)	0.83 (0.07)	1.05 (0.10)	101 (191)	0.86 (0.06)

TABLE 5.6 – Moyennes et écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus calculées avec F2<sub>onset</sub>, F2<sub>@burst1</sub> et F2<sub>@burst2</sub> des cinq occlusives.

les équations de locus calculées avec les F2<sub>@burst</sub> ont tendance à avoir des pentes plus basses et des ordonnées à l'origine plus élevées que celles estimées avec les F2<sub>onset</sub> (à l'exception des pentes de /ʔ/ et de /k/). En effet, cette constatation est attendue car l'influence de la consonne est plus importante au moment du relâchement (F2<sub>@burst</sub>) qu'au début de la voyelle (F2<sub>onset</sub>), ce qui explique les valeurs des pentes des F2<sub>@burst</sub> plus basses par comparaison à celles des F2<sub>onset</sub>. Les différences les plus marquées entre ces trois types se manifestent clairement dans les occlusives les plus résistantes /t<sup>ɿ</sup>/ et /t/. Ainsi, lorsque les voyelles influencent au maximum les consonnes précédentes, le calcul des équations de locus au moment du relâchement ou au début du deuxième formant révélera des degrés de coarticulation similaires (toujours élevés) car l'effet de la consonne sur la voyelle suivante ne change pas de façon importante. Par contre, quand les voyelles sont impactées au maximum par les consonnes précédentes, le calcul des équations de locus au moment du relâchement conduit à des valeurs plus basses que celui au début des transitions puisque l'effet de la consonne est maximale sur les voyelles suivantes dans ce cas.

En résumé, les résultats des équations de locus estimées avec les huit voyelles (pour les F2<sub>onset</sub> et les F2<sub>@burst</sub>) présentent des valeurs de pentes (degré de coarticulation) élevées pour /k/ et /ʔ/, moyennes pour /q/ et basses pour /t<sup>ɿ</sup>/ et /t/. L'examen de l'influence de la durée phonologique de la voyelle sur la coarticulation CV entraîne la division des données selon la longueur de la voyelle (courte vs. longue). Deux séries d'estimation des équations de locus ont alors été effectuées avec des voyelles longues et courtes qui sont spectralement différentes dans la première et similaires dans la seconde.

**Deuxième estimation : /i, a, u/ vs. /i:, a:, u:/**

Dans la deuxième série d'estimation, les équations de locus ont été calculées en fonction de la longueur des voyelles du premier sous-ensemble pour chaque consonne (/i, a, u/ vs. /i:, a:, u:/). Autrement dit, chaque équation de locus est estimée à partir de 40 points de mesure en moyenne (3 voyelles × 15 répétitions – anomalies – *outliers* = 40). Les résultats de ces équations indiquent que les consonnes produites avec les voyelles courtes ont des pentes plus élevées que quand elles sont réalisées avec les voyelles longues dans les trois méthodes (à l'exception de celles des /t<sup>h</sup>/ et /t/ de F2<sub>@burst2</sub>) (Table 5.7). Au contraire, les ordonnées à

C	Longueur	F2 <sub>onset</sub>			F2 <sub>@burst1</sub>			F2 <sub>@burst2</sub>		
		P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
t <sup>h</sup>	courte	0.59	501	0.91	0.41	725	0.68	0.43	710	0.68
	longue	0.49	598	0.93	0.35	759	0.77	0.49	598	0.89
t	courte	0.69	584	0.94	0.50	970	0.66	0.50	959	0.64
	longue	0.63	761	0.96	0.40	1203	0.69	0.50	1027	0.84
q	courte	0.93	5	0.95	0.90	-9	0.80	0.90	-2	0.78
	longue	0.80	179	0.93	0.79	190	0.87	0.84	141	0.91
ʔ	courte	0.98	50	0.97	0.97	73	0.94	0.98	51	0.93
	longue	0.89	168	0.96	0.87	208	0.94	0.87	199	0.93
k	courte	1.16	-22	0.97	1.19	-53	0.91	1.19	-46	0.91
	longue	1.04	252	0.90	0.87	406	0.77	0.90	326	0.84

TABLE 5.7 – Moyennes des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus (calculées avec F2<sub>onset</sub>, F2<sub>@burst1</sub> et F2<sub>@burst2</sub>) des cinq occlusives réalisées avec les voyelles /i, a, u/ vs. /i:, a:, u:/.

l'origine des occlusives produites avec les voyelles courtes ont des valeurs plus basses que celles prononcées avec les voyelles longues. L'ordre des valeurs de pente dans chaque longueur reste identique à celui des pentes estimées avec toutes les voyelles, c'est-à-dire que t<sup>h</sup> < t < q < ʔ < k. Cet ordre met en avant la stabilité des degrés de coarticulation des consonnes au sein de chaque longueur vocalique. Les pentes des F2<sub>@burst</sub> sont plus basses pour /t<sup>h</sup>/ et /t/, semblables pour /q/ et /ʔ/ et plus élevées pour le /k/ par rapport à celles des F2<sub>onset</sub>. Les valeurs du R<sup>2</sup> sont élevées pour les équations de locus avec F2<sub>onset</sub> (moyenne = 0.94) et moins élevées pour celles avec F2<sub>@burst1</sub> (moyenne = 0.80) et F2<sub>@burst2</sub> (moyenne = 0.84). Similairement à la première estimation, les variations entre les pentes des locuteurs sont plus importantes pour les F2<sub>@burst</sub> avec une moyenne d'écart-type des pentes de 0.07 pour F2<sub>onset</sub> et 0.13 pour chacun des F2<sub>@burst1</sub> et F2<sub>@burst2</sub> (cf. Table A.1 dans les annexes pour plus de détails sur les valeurs des écart-types).

Des *anovas* de type II à mesures répétées ont été effectuées avec les facteurs "longueur" et "consonne" comme variables dépendantes et "locuteur" comme variable aléatoire de façon à examiner la significativité des différences de longueur phonologique. Les résultats montrent que les pentes ainsi que les ordonnées à l'origine des occlusives réalisées avec les voyelles longues

sont significativement différentes de celles produites avec les voyelles courtes avec  $F_{(1,9)} = 45.15, p < 0.001$ ;  $F_{(1,9)} = 55.18, p < 0.001$ , respectivement pour les pentes et les ordonnées à l'origine des équations de locus avec les  $F2_{onset}$ . Il n'existe pas d'interaction entre la longueur et la consonne, signifiant que l'effet ne dépend pas de la consonne. Autrement dit, les pentes des occlusives produites avec les voyelles courtes sont significativement plus élevées que celles produites avec les voyelles longues et inversement pour les ordonnées à l'origine.

Quant aux équations de locus calculées avec les  $F2_{@burst1}$  et avec les  $F2_{@burst2}$ , un effet principal de la longueur a également été remarqué pour les deux soit  $F_{(1,9)} = 26.95, p < 0.001$ ;  $F_{(1,9)} = 26.91, p < 0.001$  pour  $F2_{@burst1}$  et  $F_{(1,9)} = 6.16, p < 0.05$ ;  $F_{(1,9)} = 7.36, p < 0.05$  pour  $F2_{@burst2}$ , respectivement pour les pentes et les ordonnées à l'origine. Toutefois, une interaction entre la longueur et la consonne a été observée stipulant ainsi que l'effet dépend de la consonne, c'est-à-dire que les pentes des voyelles courtes peuvent être plus élevées ou plus basses que celles des voyelles longues selon les consonnes.

Les résultats de cette série d'analyses mettent en avant que le chevauchement coarticulaire augmente lorsque les occlusives sont prononcées avec des voyelles courtes par comparaison avec celles longues. Cette différence coarticulaire en fonction de la longueur vocalique peut être la conséquence de l'opposition spectrale, temporelle ou des deux ensemble entre ces deux catégories vocaliques. La série d'analyses suivante étudie le degré de coarticulation de ces occlusives lorsqu'elles sont réalisées avec deux groupes de voyelles qui s'opposent en durée et sont spectralement similaires. Cette série permet d'examiner l'impact de la variation de la durée vocalique sur la coarticulation indépendamment de l'opposition spectrale.

### Troisième estimation : /i, a, u/ vs. /e:, a:, o:/

Les équations de locus ont été estimées pour les cinq occlusives étudiées prononcées avec les voyelles longues /e:, a:, o:/ afin de les comparer à celles réalisées avec les voyelles courtes /i, a, u/. En accord avec la deuxième estimation, les résultats des pentes exposent que les occlusives articulées avec les voyelles longues ont des pentes plus élevées et des ordonnées à l'origine plus basses que celles produites avec des voyelles courtes à l'exception de l'occlusive vélaire (Table 5.8)<sup>22</sup>. Cette tendance est remarquée dans les trois méthodes de calcul (avec  $F2_{onset}$  et les  $F2_{@burst}$ ). De même, un ordre identique du degré de coarticulation a été observé pour ce groupe de voyelles longues. Les pentes des occlusives estimées avec ces dernières révèlent un degré de coarticulation plus élevé que celles du groupe /i:, a:, u:/ pour les occlusives /t<sup>h</sup>, t, q/ et plus basses pour les occlusives /ʔ, k/. Cette constatation signifie que le groupe des voyelles /e:, a:, o:/ est moins résistant aux influences des occlusives /t<sup>h</sup>, t, q/ que celui des voyelles /i:, a:, u/. Les variations entre les pentes des locuteurs sont similaires à celles observées dans la deuxième estimation avec des moyennes d'écart-types de 0.09, 0.14 et 0.13 respectivement pour

22. Il est à noter que les valeurs des équations de locus calculées avec les voyelles courtes sont les mêmes que celles exploitées dans la deuxième estimation (Table 5.7). Elles ont été gardées dans la Table 5.8 à titre de comparaison.

C	Longueur	F2 <sub>onset</sub>			F2 <sub>@burst1</sub>			F2 <sub>@burst2</sub>		
		P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
t <sup>f</sup>	courte	0.59	501	0.91	0.41	725	0.68	0.43	710	0.68
	longue	0.35	758	0.76	0.22	919	0.52	0.31	811	0.61
t	courte	0.69	584	0.93	0.50	960	0.66	0.50	959	0.64
	longue	0.60	779	0.94	0.41	1159	0.75	0.47	1049	0.75
q	courte	0.93	5	0.95	0.91	-9	0.80	0.90	-2	0.78
	longue	0.74	257	0.87	0.75	214	0.89	0.77	195	0.89
ʔ	courte	0.98	50	0.97	0.97	73	0.94	0.98	-46	0.93
	longue	0.92	145	0.99	0.90	176	0.98	0.94	-247	0.97
k	courte	1.09	-22	0.97	1.19	-53	0.91	1.19	51	0.91
	longue	1.17	-132	0.92	1.29	-248	0.84	1.27	123	0.87

TABLE 5.8 – Moyennes des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus calculées avec les voyelles /i, a, u/ (courtes) vs. /e:, ai, o:/ (longues) des dix locuteurs.

les pentes des F2<sub>onset</sub>, F2<sub>@burst1</sub> et F2<sub>@burst2</sub> (cf. annexes Table A.2).

Les analyses statistiques (*anova* de type II à mesures répétées) montrent que les pentes et les ordonnées à l'origine des occlusives produites avec les voyelles longues sont significativement différentes de celles prononcées avec les voyelles courtes pour les équations estimées avec F2<sub>onset</sub> soit  $F_{(1,9)} = 10.16, p < 0.05$ ;  $F_{(1,9)} = 9.36, p < 0.05$ , respectivement. Une interaction entre les variables "consonne" et "longueur" a été observée pour les pentes et les ordonnées à l'origine. Cette information indique que l'effet de la longueur dépend de la consonne, comme c'est le cas de l'occlusive vélaire /k/ par rapport aux autres occlusives /t<sup>f</sup>, t, q, ʔ/.

Concernant les équations de locus estimées avec les deux F2<sub>@bursts</sub>, elles ont signalé des résultats différents. Aucun effet principal de la longueur sur les pentes et sur les ordonnées à l'origine n'a été mis en évidence pour F2<sub>@burst1</sub> soit  $F_{(1,9)} = 4.22, p = 0.07$ ;  $F_{(1,9)} = 2.21, p < 0.171$  et pour F2<sub>@burst2</sub> soit  $F_{(1,9)} = 1.125, p = 0.32$ ;  $F_{(1,9)} = 0.503, p = 0.496$ , respectivement. Une intersection des variables "longueur" et "consonne" a également été remarquée pour les deux calculs. L'effet principal non significatif des pentes et des ordonnées à l'origine peut être dû aux variations importantes de la direction de l'effet entre les locuteurs. Autrement dit, les pentes avec les voyelles courtes sont plus élevées que celles avec les voyelles longues pour certains locuteurs et pour d'autres, le contraire est vrai. Ces variations positives vs. négatives des pentes mais également des ordonnées à l'origine cacheraient l'effet principal et conduiraient à une absence d'effet.

En effet, les points de mesures dans l'espace des équations de locus (F2<sub>@burst</sub> en fonction de F2<sub>mid</sub>) sont plus concentrés que ceux du F2<sub>onset</sub> pour les occlusives produites avec ces voyelles, notamment les occlusives /t<sup>f</sup>, t, q/. Pour certains locuteurs, les points de mesures de F2<sub>@burst</sub> et F2<sub>mid</sub> sont regroupés dans un seul ou deux nuages de points pour chaque pente (au lieu de trois). Ce phénomène indique une grande influence de la consonne sur les voyelles mais conduit à des valeurs de R<sup>2</sup> basses et des pentes moins fiables.

En résumé, les résultats globaux des trois séries d'analyses des équations de locus sur les consonnes en position initiale mettent en exergue que les cinq occlusives ont des degrés de coarticulation différents correspondant aux résultats des analyses de dispersion sur les  $F2_{onset}$  et les  $F2_{mid}$  (résistance coarticulatoire). L'impact de la longueur phonologique de la voyelle sur la relation coarticulatoire entre la consonne et les voyelles suivantes a été observé avec et sans l'existence de l'opposition spectrale entre les deux groupes de voyelles (voyelles longues et courtes). À partir de ces résultats, deux questions peuvent être posées. En premier lieu, les degrés de coarticulation des consonnes étudiées augmentent-ils ou baissent-ils lorsque celles-ci sont réalisées en position finale ? En second lieu, l'impact temporel sur la relation coarticulatoire entre la consonne et les voyelles est-il également présent dans cette position ? Dans la section suivante, le degré de coarticulation et l'impact temporel sur la relation coarticulatoire des cinq occlusives en position finale seront ainsi examinés.

#### 5.1.4.5 La coarticulation en position finale (VC)

Il a été avancé dans la littérature que la position syllabique des occlusives peut modifier la qualité de son interaction avec les voyelles (Fougeron, 1998; Keating et al., 2001). En conséquence, les degrés de coarticulation des occlusives en position initiale présentés ci-dessus peuvent être différents de ceux en position finale. L'objectif de cette section est d'évaluer la relation coarticulatoire entre les cinq occlusives et les huit voyelles ainsi que l'impact de l'opposition temporelle sur cette relation en position finale. Les analyses de dispersion (des  $F2_{onset}$  et  $F2_{mid}$ ) et celles des équations de locus réalisées sur les séquences CV ont également été effectuées sur celles VC.

Tout d'abord, les fréquences des trois premiers formants mesurées au milieu ( $F2_{mid}$ ) et à la fin ( $F2_{offset}$ ) des voyelles longues ont été comparées à leurs contreparties courtes dans les deux sous-ensembles de voyelles au sein de chaque groupe de locuteurs (masculins et féminins) où le sous-ensemble 1 est constitué de /i-i:/, a-a:/, u-u:/ et le sous-ensemble 2 est composé de /i-e:/, a-a:/, u-o:/.

Pour le premier sous-ensemble, les *anovas* montrent que les voyelles longues et celles courtes sont statistiquement différentes pour les hommes où  $F_{(1,3)} = 186.33, p < 0.001$  ;  $F_{(1,3)} = 12.71, p < 0.05$  ;  $F_{(1,3)} = 27.52, p < 0.05$  et pour les femmes où  $F_{(1,4)} = 174.4, p < 0.001$  ;  $F_{(1,4)} = 26.3, p < 0.01$  ;  $F_{(1,4)} = 16.51, p < 0.05$ , respectivement pour  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$ . Les fréquences des trois formants mesurées à la fin de leurs transitions sont statistiquement distinctes en fonction de la longueur, à l'exception de celles des  $F2_{offset}$  et  $F3_{offset}$  des données des femmes, soit  $F_{(1,3)} = 51.00, p < 0.001$  ;  $F_{(1,3)} = 11.59, p < 0.05$  et  $F_{(1,3)} = 66.77, p < 0.01$  pour les hommes et  $F_{(1,4)} = 31.7, p < 0.01$  ;  $F_{(1,4)} = 1.565, p = 0.279$  et  $F_{(1,4)} = 7.30, p = 0.054$  pour les femmes, respectivement pour  $F1_{offset}$ ,  $F2_{offset}$  et  $F3_{offset}$ . Les analyses *post hoc* ont été effectuées sur les  $F2_{mid}$  et les  $F2_{offset}$  pour les deux groupes de locuteurs (Table 5.9). La comparaison des trois paires de voyelles révèlent que la différence au sein de chacune des paires /a:-a/ et /u:-u/ passe de significative à  $F2_{mid}$  à non significative à  $F2_{offset}$  dans le cas

des locuteurs féminins. La différence entre /i:i-/ reste significative dans les deux positions de

V	F2	Femmes			Hommes		
		t. effet	valeur-z	valeur-p	t. effet	valeur-z	valeur-p
i:i	mid	533	33.92	<.001	415.15	36.77	<.001
	offset	253	10.62	<.001	251.82	13.92	<.01
a:a	mid	61.37	3.85	<.01	16.75	1.51	<b>=.655</b>
	offset	68.47	2.84	<b>=.052</b>	46.23	2.7	<b>=.095</b>
u:u	mid	205	12.75	<.001	191.64	16.66	<.001
	offset	65.74	2.70	<b>=.08</b>	92.02	4.99	<.01

TABLE 5.9 – Résultats des analyses statistiques *post-hoc* (Tukey) effectuées sur les  $F2_{mid}$  et les  $F2_{offset}$  des voyelles longues vs. courtes avec les valeurs de la taille de l'effet, la valeur z et la valeur p pour les deux groupes de locuteurs.

mesures mais la taille de l'effet est réduite de plus de la moitié à  $F2_{offset}$  par comparaison à  $F2_{mid}$ . Quant aux locuteurs masculins, la différence est non significative entre /a:a/ au milieu et à la fin du deuxième formant. Les voyelles au sein de chaque paire /i:i-/ et /u:u-/ restent statistiquement distinctes mais la taille de l'effet est diminuée de façon considérable à  $F2_{offset}$  par comparaison avec celle à  $F2_{mid}$  pour les deux paires. Ces résultats laissent suggérer que les occlusives en position finale exercent une influence sur les voyelles longues et/ou sur les voyelles courtes conduisant à ce que ces deux catégories de longueur soient spectralement similaires.

En ce qui concerne le second sous-ensemble de voyelles, il n'existe pas de différence significative entre les voyelles longues et courtes au milieu des trois premiers formants, à l'exception du  $F3_{mid}$  des données des hommes et du  $F1_{mid}$  des données des femmes, soit  $F_{(1,3)} = 9.82, p = 0.0519$ ;  $F_{(1,3)} = 0.55, p = 0.512$  et  $F_{(1,3)} = 22, p < 0.05$  pour les hommes et  $F_{(1,4)} = 25.63, p < 0.05$ ;  $F_{(1,4)} = 0.72, p = 0.46$  et  $F_{(1,4)} = 2.27, p = 0.229$  pour les femmes, respectivement pour  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$ . À propos des  $F_{offsets}$ , la longueur vocalique a un effet non significatif sur la fin des trois formants, à l'exception de  $F1_{offset}$  et  $F3_{offset}$  des données des hommes et  $F1_{offset}$  de celles des femmes, où  $F_{(1,3)} = 27.13, p < 0.05$ ;  $F_{(1,3)} = 5.84, p = 0.095$  et  $F_{(1,3)} = 13.47, p < 0.05$  pour les hommes et  $F_{(1,4)} = 81.66, p < 0.01$ ;  $F_{(1,4)} = 0.035, p = 0.863$  et  $F_{(1,4)} = 4.94, p = 0.113$  pour les femmes, respectivement pour  $F1_{offset}$ ,  $F2_{offset}$  et  $F3_{offset}$ . Les analyses *post hoc* révèlent les mêmes tendances observées dans le premier sous-ensemble<sup>23</sup>. Les résultats statistiques des deux sous-ensembles pour les occlusives finales sont en accord avec ceux des occlusives initiales. Autrement dit, la différence entre les voyelles longues vs. courtes est plus grande dans le premier sous-ensemble par rapport à celle dans le second.

23. La différence spectrale entre les voyelles longues et courtes est plus petite à la fin des formants.

### Analyses de dispersion

Les analyses de dispersion ont été réalisées dans le but d'examiner l'impact de la consonne sur les voyelles et de la longueur phonologique sur la relation coarticulatoire entre la consonne et les voyelles précédentes en position finale. La Figure 5.14 représente les valeurs des  $F2_{mid}$  des huit voyelles (cinq timbres) dans les cinq contextes consonantiques en fonction du temps pour les locuteurs masculins (en haut) et féminins (en bas). Les lignes colorées indiquent le comportement de chaque timbre vocalique en fonction du temps.

Lorsqu'une ligne penche vers le bas, cela signifie que les voyelles courtes et/ou les voyelles longues avec une durée relativement courte ont des fréquences de F2 plus basses que celles des voyelles longues (par exemple, le /i/ dans les cinq contextes consonantiques). Quand elle se dirige vers le haut, cela désigne que les voyelles courtes et/ou les voyelles longues avec une durée relativement courte ont des fréquences de F2 plus élevées que celles des voyelles longues (comme la voyelle /i/ dans tous les contextes, à l'exception de celui du /k/). Une ligne horizontale signale que les voyelles courtes et/ou celles longues avec une durée courte ont des fréquences de F2 similaires à celles des voyelles longues (par exemple, le /a/ dans la plupart des contextes).

Comme à la position initiale, les voyelles courtes sont généralement plus concentrées sur les axes vertical (spectral) et horizontal (temporel) que les voyelles longues. La voyelle courte /i/ est plus influencée par les différents contextes consonantiques que ses correspondantes longues /i:/ ou /e:/ pour les deux groupes de locuteurs. Cette observation est reflétée par un écart-type supérieur à ceux des deux voyelles longues (Table 5.10). Au contraire, la voyelle courte /u/ a globalement un écart-type inférieur à ceux des voyelles longues /u:/ et /o:/, à l'exception de ceux des /u/ et /o:/ des locuteurs masculins qui sont similaires. Cette voyelle courte est

	i	i:	a	a:	u	u:	e:	o:
M	212	134	141	144	108	132	144	105
F	327	175	205	204	150	167	226	165
Moyenne	269	154	173	174	129	150	185	135

TABLE 5.10 – Valeurs des écart-types des  $F2_{mid}$  des huit voyelles des locuteurs masculins (M) et féminins (F) et les moyennes des deux groupes.

donc moins impactée par le contexte. Quant aux voyelles /a/ et /a:/, elles ont des écart-types semblables indiquant qu'elles sont influencées aux mêmes degrés par les consonnes suivantes.

À propos de l'influence consonantique sur le milieu des voyelles précédentes, l'occlusive pharyngalisée est la consonne la plus influente des cinq pour les deux groupes de locuteurs (classée première au niveau de l'influence sur les voyelles) (Figure 5.14, Table 5.11). Ce résultat expose que pour les locutrices, l'effet de la pharyngalisation sur la voyelle est plus importante en position finale qu'en position initiale<sup>24</sup>. L'uvulaire /q/ influence les voyelles précédentes

24. Pour rappel, le degré de dispersion des  $F2_{mid}$  des huit voyelles dans les contextes des occlusives en position

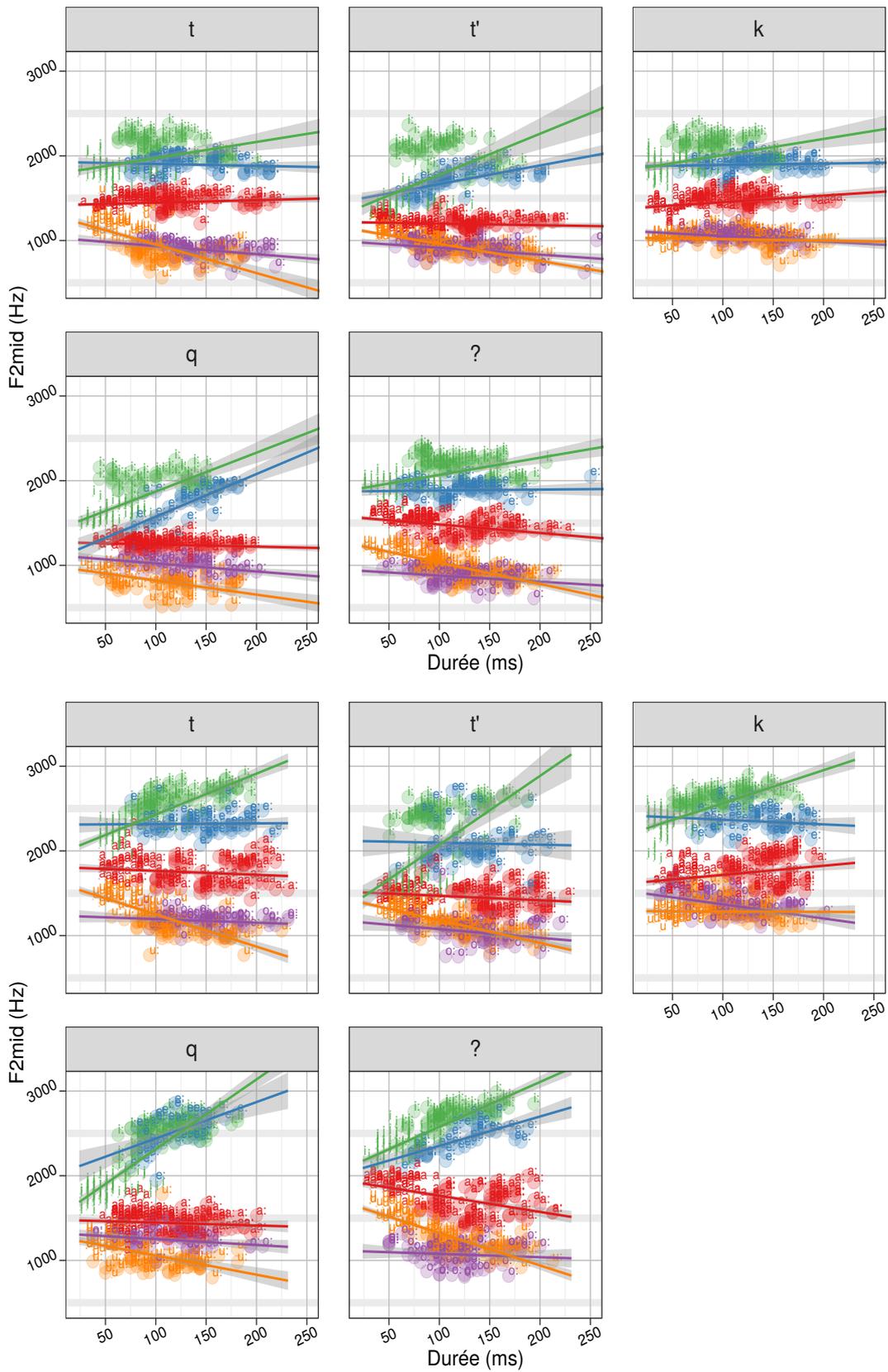


FIGURE 5.14 – Représentation de la variation du  $F2_{mid}$  en fonction du temps dans chaque contexte consonantique final pour les locuteurs masculins (en haut) et féminins (en bas).

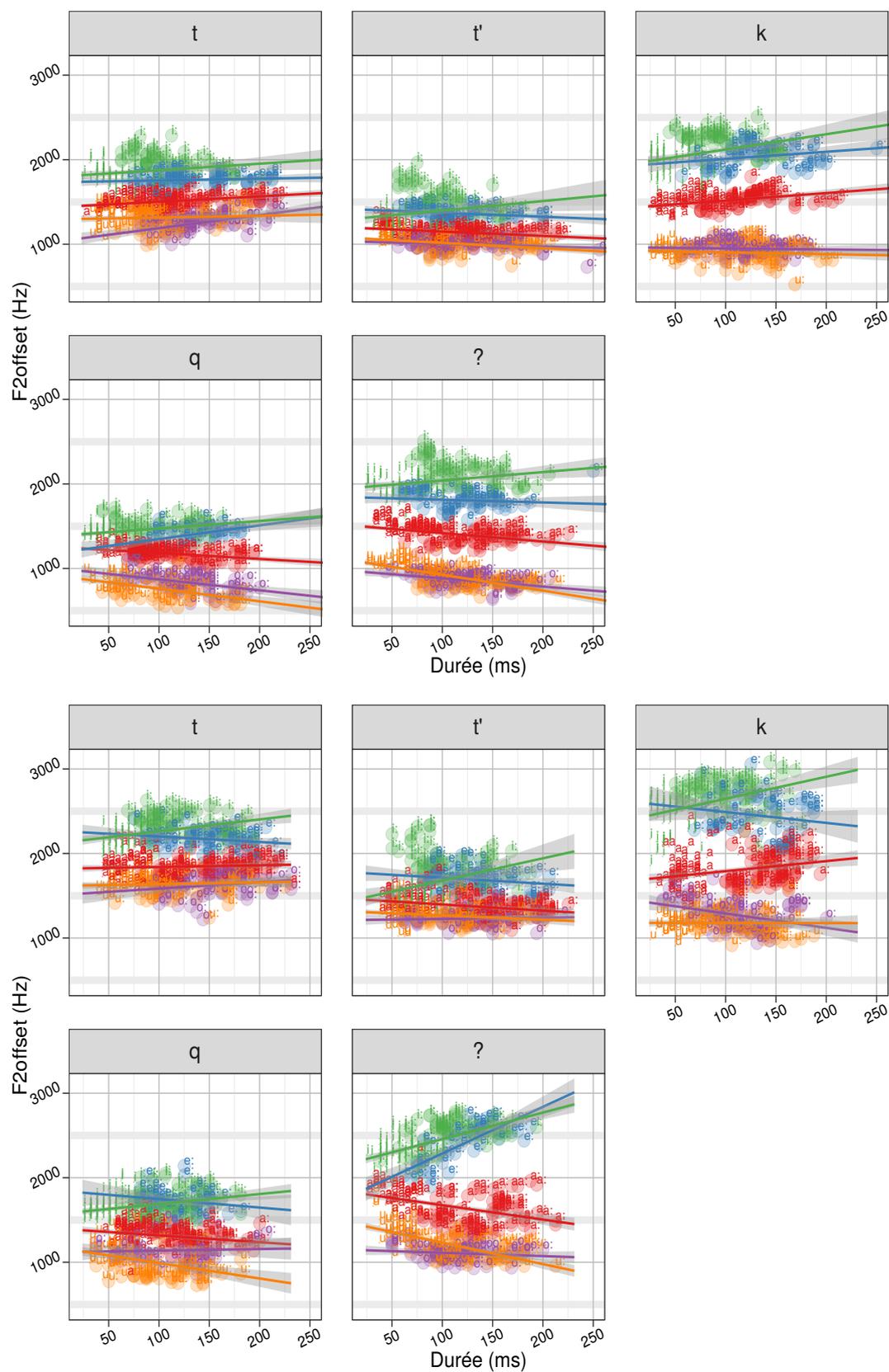


FIGURE 5.15 – Représentation de la variation du  $F2_{offset}$  en fonction du temps dans chaque contexte consonantique final pour les locuteurs masculins (en haut) et féminins (en bas).

de façon importante pour les hommes (classée deuxième) et moins importante pour les femmes (classée quatrième). Contrairement à la position initiale, la vélaire /k/ semble ici exercer un impact important sur les voyelles, ce qui la classe en deuxième position dans le groupe des femmes et en troisième pour les hommes. Concernant l'occlusive alvéolaire, elle a un effet similaire à celui en position initiale puisque les écart-types sont proches dans les deux positions. Pourtant, elle est classée quatrième pour les locuteurs masculins et troisième pour ceux féminins alors qu'en position initiale, elle est classée troisième et quatrième respectivement. L'occlusive glottale /ʔ/ reste toujours la consonne la moins influente des cinq.

Ces résultats mettent en avant que l'impact des occlusives en position finale peut être similaire (/k/), plus important (/t<sup>h</sup>/ et /t/) ou moins important (/q/ des femmes) qu'en position initiale selon le lieu d'articulation. De ce fait, le classement de ces cinq occlusives en position finale de la plus à la moins influente selon le degré de dispersion est : /t<sup>h</sup>/ > /q/ > /k/ > /t/ > /ʔ/ pour les hommes et /t<sup>h</sup>/ > /k/ > /t/ > /q/ > /ʔ/ pour les femmes.

	F2 <sub>mid</sub>					F2 <sub>offset</sub>				
	t <sup>h</sup>	t	q	ʔ	k	t <sup>h</sup>	t	q	ʔ	k
M	377	442	398	468	415	195	259	287	478	514
F	466	535	558	571	517	248	293	305	552	613
Moyenne	422	488	478	519	467	222	276	296	515	564

TABLE 5.11 – Valeurs des écart-types des F2<sub>mid</sub> et des F2<sub>offset</sub> pour les locuteurs masculins (M) et féminins (F).

La Figure 5.15 représente la distribution des valeurs de F2<sub>offset</sub> des huit voyelles dans chaque contexte consonantique en fonction du temps. La comparaison de cette représentation avec celle des occlusives en position initiale (Figure 5.12) montre que l'uvulaire /q/ est plus puissante en position finale, notamment avec l'abaissement des F2<sub>offset</sub> des voyelles antérieures /i:/ et /e:/ pour les deux groupes de locuteurs. Les voyelles centrales /a, a:/ sont moins impactées par la vélaire car elles sont moins élevées et séparées des voyelles antérieures. Par ailleurs, la résistance coarticulatoire de chaque consonne est également calculée par le biais du degré de dispersion. La consonne finale qui résiste le plus à l'influence des différentes voyelles est alors celle qui a l'écart-type du F2<sub>offset</sub> le moins élevé (dispersion minimale). Il a été constaté que l'ordre de la résistance coarticulatoire des cinq occlusives est identique à celui en position initiale, c'est-à-dire que la résistance baisse selon l'ordre suivant : /t<sup>h</sup>/ < /t/ < /q/ < /ʔ/ < /k/ pour les deux groupes de participants. Le degré de coarticulation entre la consonne et les voyelles précédentes devrait donc avoir le même ordre que celui observé en position initiale. Pour répondre à ce postulat, des analyses sur les équations de locus ont été réalisées sur les occlusives en position finale.

En résumé, les résultats des analyses de dispersion soulignent que l'impact des occlusives initiale les classe dans un ordre décroissant selon leur influence : /t<sup>h</sup>/ > /q/ > /t/ > /k/ > /ʔ/ pour les hommes et /t/ > /t<sup>h</sup>/ > /q/ > /k/ > /ʔ/ pour les femmes (Table 5.5).

sur les voyelles peut être différent selon la position syllabique de la consonne. En effet, le degré de l'agression consonantique sur les voyelles en position initiale n'est pas le même que celui en position finale et cela se manifeste dans les différences de classement dans les deux positions. Par contre, les degrés de résistance coarticulatoire des occlusives en position finale sont similaires à ceux en position initiale avec un classement identique. Il est donc prévu que les degrés de coarticulation en position finale mesurés à l'aide des équations de locus soient dans le même ordre qu'en position initiale.

#### 5.1.4.6 Les équations de locus (VC)

Les équations de locus ont été calculées trois fois pour les cinq occlusives en position finale en fonction des voyelles : une fois avec toutes les voyelles, une autre fois avec les voyelles longues *vs.* courtes du premier sous-ensemble (/i, a, u/ *vs.* /i:, a:, u:/) et une dernière avec celles du second sous-ensemble (/i, a, u/ *vs.* /e:, a:, o:/). Comme expliqué dans la méthodologie, les équations de locus des occlusives en position finale ont uniquement été estimées avec  $F2_{mid}$  et  $F2_{offset}$ . En effet, les  $F2_{@burst}$  en position finale sont souvent difficiles à prédire car la barre de relâchement de l'occlusive n'est pas visible sur le spectrogramme (relâchement trop faible ou absence de relâchement). De plus, dans certains cas, la durée des VOT est très longue avec une distance conséquente entre  $F2_{offset}$  et la barre de relâchement conduisant à des valeurs trop élevées ou trop basses de  $F2_{@burst}$ .

#### Première estimation : avec toutes les voyelles

Les équations de locus ont été calculées pour chaque occlusive produite avec l'ensemble des huit voyelles. Les données des locuteurs masculins et féminins ont également été combinées puisqu'il n'existe pas de différence significative en fonction du sexe selon une *anova* de type I à mesures répétées, soit  $F_{(1,8)} = 4.23, p = 0.0737$ . Les résultats montrent que le classement des degrés de coarticulation est identique aux analyses de dispersion présentées ci-dessus, ce qui signifie qu'il est en accord avec celui des occlusives en position initiale (Table 5.12). Autrement dit, l'alvéolaire pharyngalisée est l'occlusive la plus résistante (avec les pentes les

C	P	OO	R <sup>2</sup>
t <sup>h</sup>	0.49 (0.14)	616 (188)	0.88 (0.07)
t	0.52 (0.09)	885 (211)	0.87 (0.05)
q	0.57 (0.13)	439 (206)	0.82 (0.09)
ʔ	0.98 (0.07)	-18 (106)	0.97 (0.01)
k	1.18 (0.06)	-270 (88)	0.95 (0.02)

TABLE 5.12 – Moyennes des valeurs des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus calculées avec toutes les voyelles.

moins élevées) suivie par l'alvéolaire /t/, l'uvulaire, la glottale et la vélaire. Les ordonnées

à l'origine sont élevées pour les alvéolaires, intermédiaires pour la bilabiale et basses pour la vélaire et la glottale. Quant aux  $R^2$ , ils sont généralement élevés avec une moyenne de 0.90.

La comparaison de ces résultats avec ceux des occlusives en position initiale montre, d'une part, que l'influence de l'occlusive pharyngalisée reste globalement similaire dans les deux positions et, d'autre part, que l'alvéolaire /t/ et l'uvulaire /q/ sont plus résistantes en position finale. Cette dernière observation est signalée par la diminution importante de leurs pentes dans cette position par comparaison avec celles de ces occlusives en position initiale. Au contraire, les occlusives glottale et vélaire révèlent des pentes plus élevées en position finale indiquant une dépendance maximale de celles-ci aux contextes vocaliques précédents.

Quant à l'influence de la longueur vocalique sur la relation coarticulatoire entre les voyelles et la consonne en position finale, elle a été évaluée dans les deuxième et troisième estimations. Similairement aux calculs des équations de locus pour les occlusives en position initiale, la première estimation inclut les consonnes produites avec les voyelles /i, a, u, i:, a:, u:/ et la seconde contient celles réalisées avec les voyelles /i, a, u, e:, a:, o:/.

#### Deuxième estimation : i, a, u vs. i:, a:, u:

Les résultats des équations de locus estimées pour le premier sous-ensemble indiquent que les voyelles longues ont des pentes plus basses et des ordonnées à l'origine plus élevées que leurs correspondantes courtes (Table 5.13). L'ordre des degrés de coarticulation reste identique à celui des équations de locus calculées avec toutes les voyelles (et à celui de la position initiale) à l'exception des deux alvéolaires qui montrent des moyennes de pentes identiques lorsqu'elles sont produites avec les voyelles longues. De plus, la valeur de pente de l'uvulaire est très proche

C	Longueur	P	OO	$R^2$
t <sup>h</sup>	courte	0.56 (0.15)	521 (202)	0.68 (0.15)
	longue	0.50 (0.15)	600 (183)	0.92 (0.04)
t	courte	0.72 (0.10)	523 (214)	0.92 (0.03)
	longue	0.50 (0.11)	960 (221)	0.90 (0.04)
q	courte	0.81 (0.14)	146 (191)	0.89 (0.08)
	longue	0.52 (0.13)	453 (190)	0.84 (0.13)
ʔ	courte	1.17 (0.06)	-368 (107)	0.97 (0.03)
	longue	0.98 (0.08)	-27 (153)	0.99 (0.01)
k	courte	1.24 (0.08)	-310 (96)	0.96 (0.03)
	longue	1.16 (0.11)	-275 (211)	0.95 (0.03)

TABLE 5.13 – Valeurs moyennes et écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des  $R^2$  des cinq occlusives en position finale dans le contexte des voyelles courtes opposé à celui des voyelles longues du premier sous-ensemble.

de celle des alvéolaires, ce qui signifie que ces trois occlusives sont moins influencées par les voyelles longues. Le  $R^2$  est relativement élevé (> à 0.80) pour toutes les pentes à l'exception

de celles de /tʰ/ réalisées avec les voyelles courtes. Cette valeur basse de R<sup>2</sup> est expliquée par la grande concentration des points de mesures dans un ou deux nuages de points dans l'espace des équations de locus. Cette constatation reflète l'impact important que cette occlusive exerce sur les voyelles précédentes au milieu et à la fin du deuxième formant.

Les analyses statistiques ont été effectuées à l'aide des *anovas* de type II à mesures répétées où la pente puis l'ordonnée à l'origine sont les variables dépendantes, la longueur et la consonne sont les variables indépendantes et le facteur "locuteur", la variable aléatoire. Les résultats exposent que les pentes ainsi que les ordonnées à l'origine des occlusives produites avec les voyelles courtes sont significativement différentes de celles des consonnes prononcées avec les voyelles longues, soit  $F_{(1,9)} = 57.19, p < 0.001$  ;  $F_{(1,9)} = 44.01, p < 0.001$ , respectivement pour les pentes et les ordonnées à l'origine. Un effet d'interaction avec les consonnes a été observé indiquant que l'effet de la longueur diffère en fonction de la consonne.

### Troisième estimation : i, a, u vs. e:, a:, o:

Les résultats des équations de locus du deuxième sous-ensemble sont similaires à ceux du premier. Les pentes des occlusives prononcées avec les voyelles longues sont moins élevées et les ordonnées à l'origine sont plus grandes que celles des occlusives réalisées avec les voyelles courtes (Table 5.14). De même, les R<sup>2</sup> sont globalement semblables dans les estimations des deux sous-ensembles. Le classement des pentes des occlusives réalisées avec les

C	Longueur	P	OO	R <sup>2</sup>
tʰ	courte	0.56 (0.15)	521 (202)	0.68 (0.15)
	longue	0.47 (0.14)	648 (180)	0.87 (0.04)
t	courte	0.72 (0.10)	523 (214)	0.92 (0.03)
	longue	0.47 (0.07)	955 (193)	0.88 (0.07)
q	courte	0.81 (0.14)	146 (191)	0.89 (0.08)
	longue	0.65 (0.26)	325 (356)	0.79 (0.19)
ʔ	courte	1.17 (0.06)	-368 (107)	0.97 (0.03)
	longue	0.93 (0.11)	78 (169)	0.97 (0.02)
k	courte	1.24 (0.08)	-310 (96)	0.96 (0.03)
	longue	1.20 (0.12)	-288 (129)	0.96 (0.03)

TABLE 5.14 – Valeurs moyennes et écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des cinq occlusives en position finale dans le contexte des voyelles courtes opposé à celui des voyelles longues du second sous-ensemble.

voyelles longues est identique à celui de la même catégorie de longueur de la deuxième estimation. Les deux alvéolaires, possédant les mêmes valeurs moyennes (avec un écart-type plus important pour la pharyngalisée), sont les occlusives les moins coarticulées. L'uvulaire est en deuxième position avec une moyenne de pentes plus élevée que celle observée pour les voyelles longues du premier sous-ensemble. La valeur de cette moyenne est expliquée principalement par la pente trop élevée de la locutrice Ha (1.19). Si on enlève la pente de Ha, la moyenne des

pentences devient alors 0.59<sup>25</sup>, ce qui veut dire que la valeur moyenne se rapproche de celle de la deuxième estimation (différence de 0.07 au lieu de 0.13). Quant aux occlusives glottale et vélaire, elles sont les consonnes qui ont les degrés de coarticulation les plus importantes avec des pentes similaires à celles du premier sous-ensemble.

L'examen des résultats des équations de locus des deux sous-ensembles souligne que les degrés de coarticulation des occlusives restent globalement similaires lorsqu'elles sont prononcées avec les deux groupes de voyelles longues (/i:, a:, u:/ et /e:, a:, o:/). Cette constatation est appuyée par les analyses statistiques qui révèlent une différence significative de pentes et d'ordonnées à l'origine en fonction de la longueur phonologique des voyelles pour le second sous-ensemble, soit  $F_{(1,9)} = 144.4, p < 0.001$  ;  $F_{(1,9)} = 83.83, p < 0.001$ , respectivement. Un effet d'interaction avec la variable "consonne" a également été notée pour les pentes et les ordonnées à l'origine.

En résumé, les analyses de dispersion et celles des équations de locus des occlusives en position finale sont généralement en accord avec celles des occlusives en position initiale avec un classement identique du degré de coarticulation/résistance coarticulatoire. De plus, il existe également une influence significative de la longueur vocalique sur le degré de coarticulation dans cette position. En effet, le degré de l'influence de la consonne sur les voyelles semble être plus important en position finale comme dans le cas de la vélaire et de l'uvulaire. Le degré de cet impact est modifié à la fin de la transition du deuxième formant pour devenir proche de celui de la position initiale. Dans les deux positions syllabiques, ces variations des degrés de l'influence ont conduit à des pentes plus basses pour l'uvulaire et plus élevées pour la vélaire et la glottale alors que les degrés de coarticulation des alvéolaires restent globalement similaires.

### 5.1.5 Discussion

Les objectifs principaux de cette expérience étaient d'évaluer le degré de coarticulation des cinq occlusives non voisées de l'arabe jordanien et l'influence de la longueur vocalique sur la relation coarticulatoire entre chacune de ces consonnes et les voyelles adjacentes. Les analyses spectrales ont révélé que les voyelles longues et courtes peuvent être divisées en deux sous-ensembles en fonction de leur similarité acoustique. Le premier sous-ensemble est composé des trois voyelles courtes et des trois longues qui sont spectralement et temporellement différentes (/i, a, u/ et /i:, a:, u:/) et le second correspond aux trois voyelles courtes et aux trois longues qui sont spectralement similaires mais temporellement distinctes (/i, a, u/ et /e:, a:, o:/). Cette division a permis de comparer l'influence de l'opposition temporelle sur la relation coarticulatoire entre la consonne et la voyelle dans les deux contextes. Les résultats des équations de locus mettent en avant que le degré de coarticulation des occlusives est influencé de manière comparable dans les deux sous-ensembles. La production des occlusives avec les

25. La pente de /q/ réalisé avec les voyelles longues du second sous-ensemble par la locutrice Ha est calculée avec 19 points de mesures qui sont concentrés dans un seul nuage de points expliquant la valeur élevée de pente.

voyelles courtes amène toujours à des pentes plus élevées (+ de coarticulation) que celles des voyelles longues lorsque les deux catégories sont opposées spectralement (correspondant au premier sous-ensemble) ou non (second sous-ensemble). Autrement dit, l'aspect temporel des voyelles qui s'opposent phonologiquement joue un rôle important sur la relation coarticulatoire entre les consonnes et les voyelles, peu importe le degré de différence spectrale entre les voyelles longues et courtes. Cependant, le classement des degrés de coarticulation reste identique dans chaque longueur et dans chaque sous-ensemble de voyelles.

En effet, le classement des degrés de coarticulation de la vélaire (le plus élevé) et des alvéolaires (les plus bas) est en accord avec celui exploré dans la littérature. Comme prévu, la glottale présente un degré de coarticulation élevé mais il est toujours plus bas que celui de la vélaire. La réalisation de la première n'engage pas de mouvement lingual ou d'autres mouvements qui soient en conflit avec les voyelles adjacentes. L'impact de la voyelle sur la glottale est ainsi maximal. Toutefois, la production de la vélaire entraîne un engagement de la langue dominé par le mouvement lingual vocalique d'où un degré de coarticulation maximal. L'uvulaire, quant à elle, a des valeurs de pentes toujours intermédiaires par rapport à celles des alvéolaires (basses) et celles de la vélaire et de la glottale (hautes). Le degré de coarticulation de cette occlusive varie de façon importante en fonction du locuteur, de la position syllabique (+ élevée en position finale), de la longueur phonologique (+ basses avec les voyelles longues) et du groupe vocalique (+ élevées avec le second sous-ensemble). De ce fait, le degré de la résistance et de l'agression coarticulatoire de cette occlusive mais également celui de l'alvéolaire pharyngalisée dépendent du locuteur car les locuteurs pharyngalisent et uvularisent les voyelles voisines à des degrés différents. Cette distinction est plus visible entre les deux sexes de locuteurs. En effet, les locuteurs féminins ont tendance à avoir des pentes plus basses que celles des locuteurs masculins pour l'uvulaire et, au contraire, les femmes ont des pentes relativement plus élevées que celles des hommes pour l'occlusive pharyngalisée. Au sein de chaque groupe de locuteurs, les pentes varient également, désigné par les écart-types relativement importants des pentes de ces occlusives par comparaison avec les autres pentes dans la plupart des conditions.

Les deux positions syllabiques des cinq occlusives mettent en évidence des degrés de coarticulation généralement similaires. Des *anovas* de type II<sup>26</sup> ont été effectuées sur les paramètres des équations de locus de la première estimation (avec toutes les voyelles) afin d'examiner statistiquement l'impact global de la position syllabique sur le degré de coarticulation. Les résultats montrent qu'il n'existe pas de différence significative de pentes ou d'ordonnées à l'origine en fonction de la position syllabique, soit  $F_{(1,9)} = 0.94, p = 0.358$ ;  $F_{(1,9)} = 0.08, p = 0.79$ , respectivement. Une interaction entre la consonne et la position a été observée pour la pente et l'ordonnée à l'origine signifiant que l'impact de la position syllabique est différent en fonction de la consonne. En effet, la comparaison entre les Tables 5.6 et 5.12 montre que la pharyngalisée a des pentes et des ordonnées à l'origine globalement similaires alors que le /t/ et le /q/

26. Avec la pente puis l'ordonnée à l'origine comme variables dépendantes, la position et la consonne comme variables indépendantes et le locuteur comme facteur aléatoire.

ont des pentes plus élevées et des ordonnées à l'origine plus basses en position initiale qu'en position finale. Au contraire, la glottale et la vélaire ont des pentes plus basses et des ordonnées à l'origine plus élevées en position initiale. Deux autres *anovas* de type III (avec position, consonne et longueur comme variables indépendantes) ont également été réalisées sur les équations de locus des consonnes produites avec les deux sous-ensembles de voyelles. Les résultats signalent que la position syllabique dans les deux sous-ensembles n'influence pas significativement les valeurs des pentes, où  $F_{(1,9)} = 0.39, p = 0.546$  ;  $F_{(1,9)} = 1.323, p = 0.28$ , respectivement pour le premier et le second sous-ensemble. Quant aux ordonnées à l'origine, elles sont statistiquement distinctes en fonction de la position syllabique pour les deux sous-ensembles, avec  $F_{(1,9)} = 7.96, p < 0.05$  ;  $F_{(1,9)} = 12.47, p < 0.01$  respectivement. Une interaction entre la consonne et la position a également été notée pour les pentes et les ordonnées à l'origine. Cependant, aucune interaction de la pente n'est remarquée entre la position et la longueur dans les deux sous-ensembles, c'est-à-dire que l'effet de la longueur sur la pente n'est pas influencé par celui de la position syllabique. Cette absence d'interaction n'est présente que dans le premier sous-ensemble pour les ordonnées à l'origine.

En ce qui concerne les équations de locus estimées au moment du relâchement de la consonne, les deux méthodes de prédiction de  $F2_{@burst}$  ( $F2_{@burst1}$  et  $F2_{@burst2}$ ) révèlent des valeurs de pentes et d'ordonnées à l'origine semblables. Ces équations de locus ont montré des pentes généralement plus basses pour les consonnes /t<sup>h</sup>, t/, similaires pour /q, ʔ/ et plus hautes pour /k/ comparées à celles calculées avec les  $F2_{onset}$ . Cette variation de pente en fonction du lieu d'articulation signifie que l'impact de la consonne peut être plus élevé, similaire ou plus bas au moment du relâchement par rapport au début du voisement de la voyelle. Comme expliqué précédemment, la prédiction des valeurs de  $F2_{onset}$  des occlusives non voisées au moment du relâchement ( $F2_{@burst}$ ) amène à réduire l'effet de l'intervalle de silence situé entre le moment du relâchement et la première pulsation de la voyelle suivante. L'information du lieu d'articulation de l'occlusive non voisée serait alors plus proche de sa contrepartie voisée. Les résultats de cette expérience mettent en avant que cet intervalle de silence n'a pas le même effet sur les cinq occlusives.

D'après ces arguments, les paramètres des équations de locus estimées avec les  $F2_{@burst}$  des occlusives non voisées seraient comparables à celles de leurs correspondantes voisées lorsqu'elles sont calculées avec des  $F2_{onset}$  et/ou des  $F2_{@burst}$ . Une seconde expérience sur l'arabe jordanien a été mise en place afin d'examiner ce postulat. Cette expérience permet d'évaluer les méthodes de prédiction du  $F2_{onset}$  au moment du relâchement des consonnes non voisées en comparant les paramètres des équations de locus calculées avec les  $F2_{@burst}$  prédits à ceux estimés avec les  $F2_{onset}$  des occlusives voisées. Autrement dit, pour les occlusives qui partagent un même lieu d'articulation, il est attendu à ce que ce lieu d'articulation possède un degré de coarticulation similaire peu importe le voisement.

Le second protocole expérimental a également pour objectif d'analyser l'impact de la longueur vocalique sur la relation coarticulaire entre des occlusives voisées de l'arabe jordanien

et des voyelles adjacentes, étant donné que la première expérience n'analyse que les occlusives non voisées. En d'autres mots, les effets de la longueur vocalique observés dans le cas des occlusives non voisées sont-ils présents pour celles voisées ? De plus, un cinquième lieu d'articulation (occlusive bilabiale) non abordé dans la première expérience sera étudié. Cependant, en arabe jordanien, pour ce dernier, il n'existe qu'une seule occlusive qui est voisée ; par conséquent, aucune comparaison voisée/non voisée ne peut être réalisée (*cf.* Chapitre 4).

## 5.2 Expérience II

La seconde expérience sur l'arabe jordanien a pour but d'examiner l'influence des variations temporelles provoquées par l'opposition de la longueur phonologique des voyelles sur les occlusives voisées et non voisées. En effet, les résultats de la première expérience ont révélé un chevauchement coarticulatoire plus important entre les voyelles courtes et les occlusives non voisées comparé à celui entre les voyelles longues et les mêmes consonnes. L'expérience actuelle évalue alors l'étendue de ce phénomène dans le cas des occlusives voisées où l'effet du voisement sur l'impact temporel des voyelles sera mis en avant.

Le deuxième objectif de cette expérience est - comme cité plus haut - l'examen des équations de locus des occlusives voisées vs. non voisées pour aider à tester l'efficacité des deux méthodes de prédiction de  $F2_{@burst}$  exploitées dans la première expérience. En effet, il a été précisé précédemment que le moment du relâchement des occlusives voisées est généralement plus proche du début des transitions formantiques que celui des occlusives non voisées. Le début des transitions formantiques des non voisées contiennent ainsi moins d'informations sur le lieu d'articulation que les voisées à cause de l'intervalle - relativement long - qui sépare le moment du relâchement des formants vocaliques. De ce fait, dans la première expérience, l'intérêt des deux méthodes de prédiction était de réduire l'impact de cet intervalle en extrapolant une valeur de  $F2_{onset}$  au moment du relâchement. La nouvelle valeur prédite de  $F2_{onset}$  ( $F2_{@burst}$ ) est sensée conduire à des équations de locus plus représentatives du lieu d'articulation que celle du  $F2_{onset}$  mesurée. Dans cette expérience, les équations de locus estimées avec  $F2_{onset}$  et celles avec  $F2_{@burst}$  des occlusives non voisées seront comparées à celles de leur contreparties voisées. Il est prévu que les paramètres des équations de locus calculés avec  $F2_{onset}$  et/ou ceux avec  $F2_{@burst}$  des voisées soient comparables aux paramètres des équations de locus estimés avec  $F2_{@burst}$  des occlusives non voisées.

Cette expérience peut alors être considérée comme une expérience contrôle dans laquelle il serait possible de vérifier l'efficacité des méthodes d'estimation du  $F2_{@burst}$  et de comparer les résultats de la première expérience dans une autre population.

### 5.2.1 Participants

Sept locuteurs jordaniens masculins âgés de 27 à 35 ans au moment de l'enregistrement ont participé à cette expérience. Ne présentant aucun trouble du langage, ils sont originaires de trois villes différentes de Jordanie dont trois locuteurs venant de Ma'an (Ah, Ab et As), deux d'Amman (Al, Ta) et deux de Jerash (Su, Ma). Deux de ces locuteurs ont participé à l'expérience I dont l'expérimentateur (Ab et Ah).

### 5.2.2 Stimuli

Les locuteurs devaient lire des séquences CVC enchâssées dans une phrase porteuse (/ħaka CVC hassa/ : « Il a dit CVC à l'instant »). Similairement à la première expérience, les séquences apparaissaient l'une après l'autre en alphabet arabe vocalisé (avec les diacritiques représentant les voyelles) sur un écran d'ordinateur.

Trois occlusives voisées et deux occlusives non voisées ont été étudiées, /b, d, g, t, k/. Chacune de ces occlusives apparaissait dans un mot de la langue ayant une structure CVC. Ces cinq occlusives cibles ont été combinées avec six voyelles : trois courtes (/i, a, u/) et trois longues (/i:, a:, u:/) dans deux positions lexicales (initiale  $C_{cible}VC$  vs. finale  $CVC_{cible}$ ). L'autre consonne du mot était sélectionnée au hasard en fonction des mots existants dans le vocabulaire de l'arabe jordanien, soit un total de 80 mots différents dont la fréquence des mots n'a pas été contrôlée.

### 5.2.3 Procédure

Les mêmes procédures d'enregistrement, de segmentation, d'extraction et de traitement des données acoustiques mises en place dans la première expérience ont été utilisées pour ce protocole expérimental. Les participants ne connaissaient pas les séquences avant le début de l'enregistrement mais ils étaient prévenus que la consonne /q/ dans ces mots devait être prononcée comme un /g/. Chaque mot a été répété dix fois soit 600 séquences par locuteur (5 consonnes  $\times$  2 positions syllabiques  $\times$  6 voyelles  $\times$  10 répétitions) totalisant 4200 séquences CVC à analyser. Le système de détection des valeurs extrêmes a repéré 255 séquences comme des *outliers* qui ont ensuite été éliminés des analyses. Ces dernières ont alors été réalisées sur un total de 3945 séquences.

### 5.2.4 Résultats

Les résultats de cette expérience seront présentés de la même façon que ceux exposés dans la première. Autrement dit, les analyses de la durée vocalique et des propriétés spectrales de l'espace vocalique sont effectuées sur l'ensemble des données (3945 séquences). Les analyses de coarticulation (analyses de dispersion et des équations de locus) sont ensuite traitées selon la position syllabique.

#### 5.2.4.1 La durée vocalique

Les résultats globaux de la durée mettent en exergue que, tout d'abord, la distribution de la densité de la durée des voyelles longues vs. courtes affiche deux catégories distinctes de voyelles avec une zone de chevauchement (Figure 5.16). Ensuite, la moyenne de la durée des voyelles courtes (77 ms) représente 55.8 % de celle des voyelles longues (138 ms). De plus, les variations des voyelles courtes sont plus petites que celles des voyelles longues avec un écart-type total

de 23 ms vs. 40 ms. Ces différences de variations sont visibles sur la Figure 5.16 avec une concentration plus grande des voyelles courtes sur l'axe des abscisses (durée) par comparaison avec celle des voyelles longues.

L'intervalle de chevauchement de la durée entre les deux longueurs semble également être comparable avec celui remarqué dans la première expérience. Il est compris entre les deux moyennes et s'étend même au-delà (Figure 5.16). Les voyelles longues peuvent ainsi être réalisées avec des durées égales à celles des voyelles courtes.

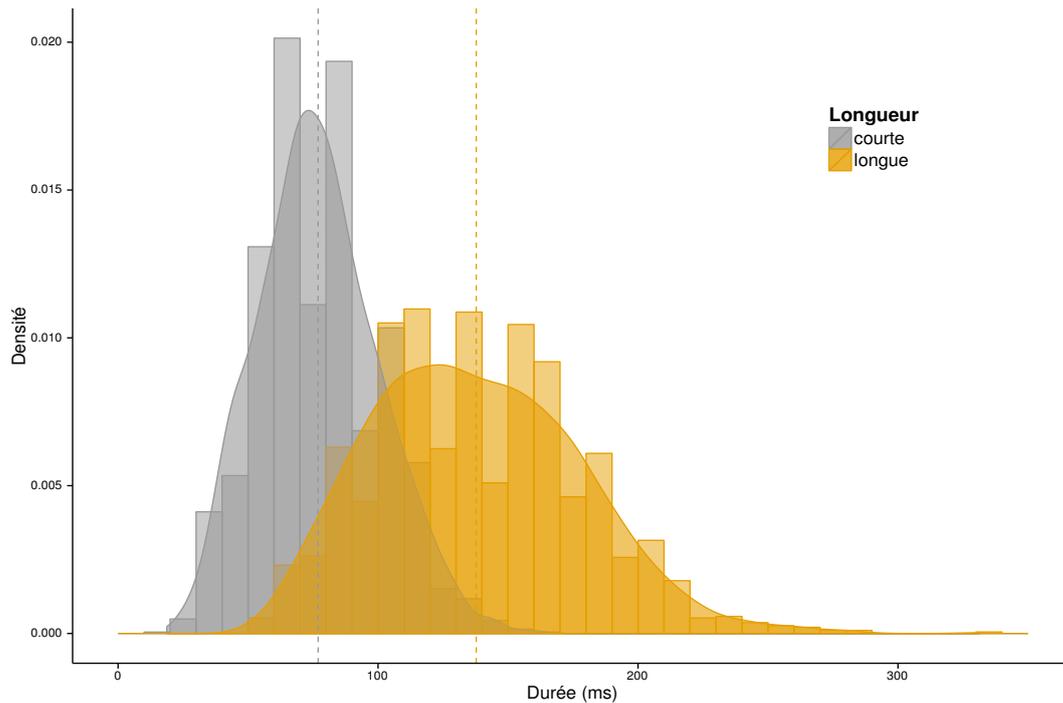


FIGURE 5.16 – Distribution de la densité des voyelles en fonction de la longueur phonologique avec la barre de la moyenne pour chaque catégorie.

L'examen des moyennes de la durée vocalique pour chaque voyelle indique que les voyelles ouvertes /a, a:/ sont les plus longues et les voyelles fermées /i, i:/ sont les plus courtes au sein de chaque catégorie de longueur (Table 5.15). Une *anova* de type II à mesures répétées (avec longueur et timbre comme variables indépendantes et variable "locuteur" comme facteur aléatoire) a été effectuée pour comparer la durée des voyelles longues avec celle des courtes. Les

	i	a	u
courte	70 (23)	84 (22)	77 (21)
longue	127 (36)	154 (44)	131 (34)

TABLE 5.15 – Valeurs des moyennes et des écart-types des durées des trois voyelles courtes et des trois longues étudiées.

résultats montrent que les voyelles courtes sont significativement plus courtes que leur contreparties longues, soit  $F_{(1,5)} = 74.52, p < 0.001$ . Aucun effet d'interaction n'a été remarqué entre

la longueur et le timbre sous-entendant que l'effet principal de la longueur n'est pas influencé par le timbre.

En résumé, les résultats des analyses de statistiques descriptives et inférentielles sur la longueur vocalique sont en accord avec ceux observés dans la première expérience. La section suivante présente les propriétés spectrales des six voyelles étudiées.

#### 5.2.4.2 Propriétés spectrales des formants

Les valeurs de fréquence mesurées au milieu et au début (pour les occlusives en position initiale) ou à la fin (pour les occlusives en position finale) des trois premiers formants ont été analysées. Dans un premier temps, l'espace vocalique (plan F1-F2) des six voyelles a été évalué avec toutes les données. Dans un second temps, les analyses de coarticulation ont été examinées en fonction de la position syllabique de l'occlusive cible.

#### Plan F1-F2

L'espace vocalique des six voyelles analysé pour les sept locuteurs révèle des caractéristiques similaires à celles observées dans la première expérience. La Figure 5.17 affiche les valeurs de  $F1_{mid}$  et de  $F2_{mid}$  de toutes les réalisations en Hz (en haut) et en Bark (en bas). Les voyelles longues sont plus périphériques et plus concentrées (ellipses plus petites = – de dispersion) que leur contreparties courtes. Les voyelles longues vs. courtes sur la Figure 5.17 peuvent être classées en fonction de leur distance ou chevauchement acoustique. Les voyelles /a, a:/ sont les plus proches avec un chevauchement important suivies par /u, u:/ avec un chevauchement moindre et les /i, i:/ viennent en dernière position avec très peu de chevauchement.

La Table 5.16 indique les moyennes et les écart-types des F1 et des F2 des sept locuteurs donnant les valeurs numériques de l'observation visuelle de la Figure 5.17. Les écart-types des voyelles courtes sont plus élevés que ceux des longues pour les deux formants (à l'exception du F2 du timbre /i/). Quant aux moyennes des deux formants, l'écart entre les voyelles longues et courtes reflète l'ordre suivant : |a-a:| < |u-u:| < |i-i:| pour les F2 et |a-a:| < |i-i:| < |u-u:| pour le F1.

Timbre	Longueur	F1	F2
i	courte	401 (57)	1833 (133)
	longue	318 (33)	2081 (142)
a	courte	553 (74)	1525 (162)
	longue	623 (61)	1469 (150)
u	courte	454 (65)	1081 (131)
	longue	354 (39)	938 (119)

TABLE 5.16 – Valeurs des moyennes et des écart-types (entre parenthèses) des F1 et des F2 des six voyelles de l'ensemble des données.

Les *anovas* ont été utilisées afin d'évaluer statistiquement la différence entre les voyelles

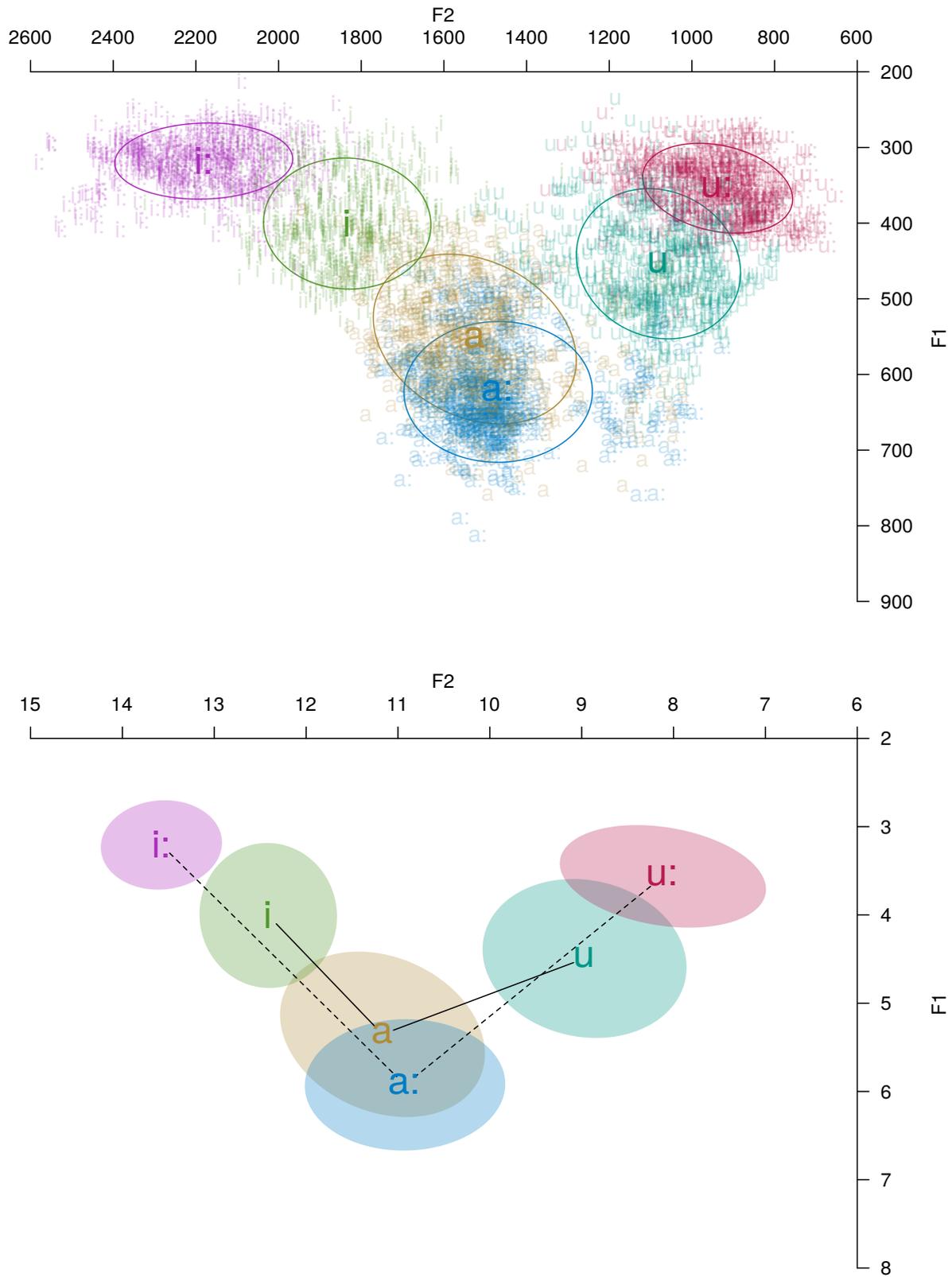


FIGURE 5.17 – Illustration du plan des fréquences des F1-F2 (mesurées au milieu du formant) des six voyelles de l’arabe jordanien produites par les sept locuteurs masculins en Hz (en haut) et en Bark (en bas).

longues et courtes. Les fréquences des deux premiers formants des voyelles longues sont significativement distinctes de celles des voyelles courtes, soit  $F_{(1,5)} = 15.91, p < 0.05$  ;  $F_{(1,5)} = 11.87, p < 0.05$ , respectivement pour F1 et F2. Une interaction entre le timbre et la longueur a été observée. Elle peut être expliquée par les différentes directions que l'écart peut prendre puisque les F1 et les F2 des voyelles longues peuvent avoir des valeurs supérieures ou inférieures à leurs correspondantes courtes selon le timbre vocalique. En complément, les analyses *post hoc* (Tukey) sur les deux formants montrent que les moyennes des trois paires de voyelles (longues vs. courtes) sont significativement distinctes.

Pour conclure, les résultats des analyses de l'espace vocalique des trois voyelles sont identiques à ceux du premier sous-ensemble de la première expérience. Autrement dit, il existe deux catégories de voyelles (longues et courtes) qui s'opposent au niveau temporel et au niveau spectral. La comparaison de la durée et des propriétés spectrales entre les deux expériences est importante pour s'assurer que les deux échantillons de locuteurs proviennent de la même population, ce qui légitime la comparaison des résultats des analyses de la coarticulation entre les deux expériences.

#### 5.2.4.3 Coarticulation en position initiale (CV)

Les analyses de coarticulation ont d'abord été traitées lorsque les occlusives cibles sont en position initiale. En premier lieu, les fréquences des trois premiers formants mesurées au début et au milieu ont été examinées. Les analyses statistiques révèlent que les  $F1_{mid}$ , les  $F2_{mid}$  et les  $F3_{mid}$  des voyelles longues sont significativement différentes de celles des voyelles courtes, soit  $F_{(1,6)} = 9.21, p < 0.05$  ;  $F_{(1,6)} = 6.505, p < 0.05$  et  $F_{(1,6)} = 16.78, p < 0.01$ , respectivement. Quant aux fréquences des  $F1_{onset}$ ,  $F2_{onset}$  et  $F3_{onset}$ , un effet principal de la longueur a également été remarqué pour les trois mesures, soit  $F_{(1,6)} = 29.42, p < 0.01$  ;  $F_{(1,6)} = 43.95, p < 0.001$  et  $F_{(1,6)} = 8.77, p < 0.05$ , respectivement. Ces analyses sont en accord avec celles du F1-F2 indiquant l'existence d'une différence spectrale entre les deux longueurs phonologiques.

#### Analyses de dispersion

Les analyses de dispersion ont été effectuées sur les  $F2_{mid}$  et les  $F2_{onset}$  de manière à examiner le degré de la résistance coarticulatoire des cinq occlusives ainsi que le degré de l'agression coarticulatoire que les consonnes exercent sur le milieu des voyelles. La Figure 5.18 représente les fréquences du deuxième formant mesurées au milieu (en haut) et au début (en bas) des six voyelles en fonction de la durée dans les cinq contextes consonantiques pour les sept locuteurs. Il apparaît que les  $F2_{mid}$  des voyelles longues (en cercles transparents) sont plus étendues verticalement (fréquence) et horizontalement (durée) que leurs correspondantes courtes. Les droites de régression, elles, représentent les tendances générales du mouvement formantique en fonction du temps pour chaque timbre vocalique.

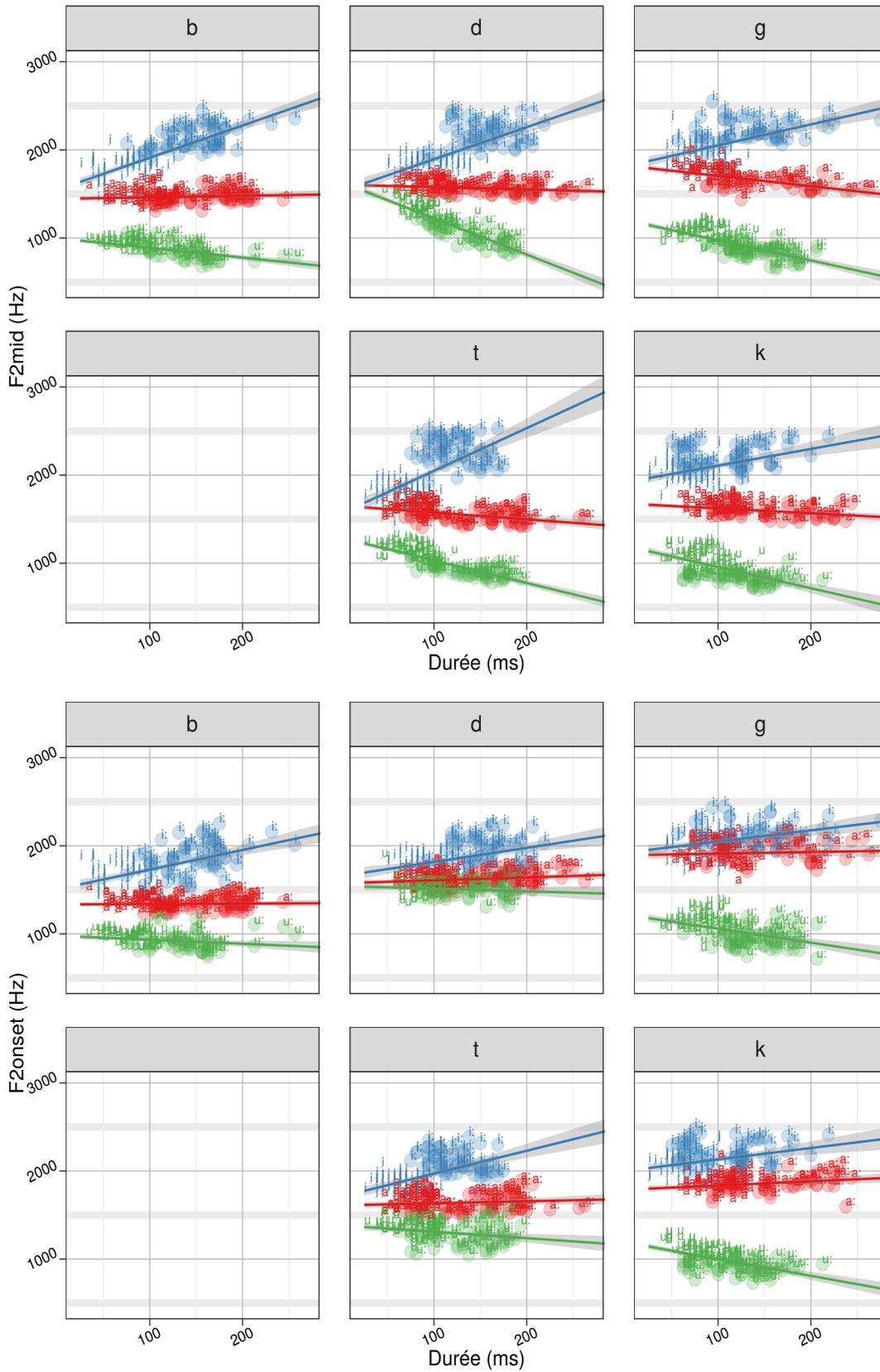


FIGURE 5.18 – Représentation de la distribution des  $F2_{mid}$  (en haut) et des  $F2_{onset}$  (en bas) en fonction du temps dans chaque contexte consonantique lorsque les consonnes cibles sont en position initiale.

Les voyelles /a, a:/ semblent être les moins influencées par la durée vocalique dans les cinq contextes consonantiques avec des pentes relativement plates. Les  $F2_{mid}$  des voyelles /u, u:/ semblent augmenter lorsque la durée diminue, signalées par des pentes descendantes. La vitesse de cette augmentation varie en fonction du contexte consonantique. Quant aux voyelles /i, i:/, elles sont les voyelles les plus influencées par la durée.

Globalement, lorsque la durée croît, les fréquences du  $F2_{mid}$  augmentent avec des pentes montantes. Le degré de cette tendance varie selon le contexte consonantique avec une tendance moins visible pour les alvéolaires. Dans les contextes, les points de  $F2_{mid}$  du timbre /i/ semblent former deux nuages de points séparés avec très peu de chevauchement en fonction de la longueur. En effet, ces deux nuages de points sont moins écartés dans le contexte de l'alvéolaire voisée en comparaison de celui de sa contrepartie non voisée. Les deux voyelles subissent alors une influence plus importante dans le premier contexte que dans le second.

L'examen des écart-types des deux catégories pour chaque locuteur indique que les voyelles courtes sont plus dispersées par le contexte consonantique que leurs correspondantes longues avec une moyenne de l'écart-type de 118 Hz pour les courtes et 84 Hz pour les longues. Ces dernières sont ainsi plus résistantes aux influences consonantiques que les voyelles courtes. Les deux timbres vocaliques /i/ et /u/ ont tendance à se diriger vers une zone de fréquences centrales proche de celle du /a/ lorsque la durée diminue. Le débit de ce mouvement dépend du lieu d'articulation et du voisement de la consonne cible. Il est plus élevé dans le contexte des alvéolaires et dans celui des occlusives voisées.

Les analyses de dispersion sur les  $F2_{mid}$  au sein de chaque contexte consonantique mettent en avant que les alvéolaires /d, t/ sont les occlusives les plus influentes (avec des écart-types bas) suivies par la bilabiale /b/ (écart-type moyen) et les vélaires /g, k/ (écart-types élevés) (Table 5.17). Il a été constaté que l'alvéolaire voisée montre un écart-type plus petit que celui de la non voisée mais l'écart-type de la vélaire non voisée est légèrement plus élevé que celui de sa contrepartie voisée. Par conséquent, le voisement ne joue pas de rôle sur l'impact des vélaires sur les voyelles.

	b	d	g	t	k
$F2_{mid}$	453	359	498	437	489
$F2_{onset}$	355	183	483	292	498

TABLE 5.17 – Valeurs des écart-types des  $F2_{mid}$  et des  $F2_{onset}$  dans les cinq contextes consonantiques en position initiale.

À propos de la représentation des points de mesures du  $F2_{onset}$  en fonction du temps dans les cinq contextes consonantiques, elle affiche une concentration plus importante sur l'axe des ordonnées (fréquence) que celle du  $F2_{mid}$  à cause de l'influence consonantique sur le début des voyelles (Figure 5.18). De plus, le degré de la résistance coarticulatoire des voyelles longues vs. courtes est différent pour chaque timbre. En effet, la moyenne des écart-types du /i/ est plus

importante que celle du /i:/ (175 vs. 155 Hz) désignant un degré de résistance coarticulatoire plus élevé pour la voyelle longue par rapport à celle courte. Au contraire, la moyenne des écart-types du /u/ est plus basse que celle du /u:/ (210 vs. 247 Hz) signifiant que la voyelle courte est plus résistante que sa correspondante longue. Quant aux moyennes des écart-types des voyelles /a, a:/, elles montrent des valeurs proches (214 vs. 216 Hz) signalant des degrés de résistance semblables. Ces résultats révèlent que les degrés de la résistance coarticulatoire des trois timbres vocaliques diminuent de la façon suivante :  $i > a > u$ .

La résistance coarticulatoire des occlusives aux différentes voyelles peut être soulignée par le degré de dispersion des  $F2_{onset}$  au sein de chaque contexte consonantique (+ de dispersion, – de résistance), affichés dans la Figure 5.18 (en bas). L'observation des  $F2_{onset}$  de la Figure 5.18 montre que la concentration des points est élevée pour les alvéolaires, intermédiaire pour la bilabiale et basse pour les vélaires. Par ailleurs, les  $F2_{onset}$  des occlusives voisées sont moins dispersés que ceux de leurs correspondantes non voisées. Ces observations sont confirmées par les mesures de l'écart-type des  $F2_{onset}$  pour les cinq occlusives rapportées dans la Table 5.17. Ces valeurs montrent également que le non voisement des occlusives affaiblit l'influence du lieu d'articulation sur le début du deuxième formant.

En résumé, les résultats des analyses de dispersion mettent en exergue que la résistance coarticulatoire des voyelles courtes est plus petite que celle des voyelles longues. Cette observation laisse supposer que les pentes des équations de locus seront plus élevées dans le contexte des voyelles courtes que dans celui des longues. De ce fait, les degrés de la résistance coarticulatoire augmentent de la manière suivante :  $k < g < b < t < d$ , signifiant qu'il est prévu que les valeurs de pentes aient l'ordre inverse ( $k > g > b > t > d$ ). D'après ces résultats, les occlusives non voisées sont toujours moins résistantes (+ coarticulées) que leurs contreparties voisées en raison de l'intervalle de silence entre le bruit de relâchement et le début du voisement. Ces occlusives représentent alors moins leur lieu d'articulation que leurs contreparties voisées. Comme expliqué précédemment, la prédiction de  $F2_{@burst}$  a pour objectif de réduire la différence provoquée par le non voisement pour les occlusives qui partagent un même lieu d'articulation. Par conséquent, il est attendu que les valeurs des pentes des occlusives non voisées estimées avec les  $F2_{@burst}$  soient plus proches de celles de leurs correspondantes voisées que quand elles sont calculées avec les  $F2_{onset}$ .

#### 5.2.4.4 Les équations de locus (CV)

Dans le but d'examiner le degré de coarticulation des occlusives étudiées ainsi que l'impact de la longueur sur la coarticulation, les équations de locus ont été calculées pour les cinq occlusives dans deux séries d'estimation en fonction du groupe de voyelles utilisées. La première estimation correspond aux équations de locus estimées avec toutes les voyelles et la seconde aux équations de locus calculées avec les voyelles longues et avec celles courtes séparément. Dans chaque série d'estimation, les équations de locus des cinq occlusives ont été réalisées trois fois selon son premier coefficient, soit avec  $F2_{onset}$ , avec  $F2_{@burst1}$  et avec  $F2_{@burst2}$  (le deuxième

coefficient restant toujours  $F2_{mid}$ ).

### Première estimation

Dans cette estimation, les équations de locus ont été calculées pour les cinq occlusives produites avec les six voyelles. Autrement dit, chaque équation de locus est réalisée avec 60 points de mesures (6 voyelles  $\times$  10 répétitions pour chaque occlusive) auxquels ont été soustraits les *outliers* (4 points de mesures en moyenne). Les équations de locus ont été générées trois fois soit une avec  $F2_{onset}$  et deux avec les  $F2_{@burst}$  prédits par les deux méthodes d'extrapolation expliquées dans la première expérience ( $F2_{@burst1}$  et  $F2_{@burst2}$ ). Au total, 35 équations de locus ont été obtenues pour chaque calcul (5 occlusives  $\times$  7 locuteurs).

Les résultats des équations de locus estimées avec  $F2_{onset}$ ,  $F2_{@burst1}$  et  $F2_{@burst2}$  sont en accord avec les analyses de dispersion (Table 5.18). En effet, les valeurs de pentes augmentent dans l'ordre suivant  $/d/ < /t/ < /b/ < /g/ < /k/$  et celles des ordonnées à l'origine croissent dans l'ordre inverse ( $/k/ < /g/ < /b/ < /t/ < /d/$ ). Ces ordres de pentes et d'ordonnées à l'origine sont valables pour les trois calculs des équations de locus (avec  $F2_{onset}$ ,  $F2_{@burst1}$  et  $F2_{@burst2}$ ) où les consonnes vélaires sont les occlusives les plus coarticulées, les alvéolaires les moins coarticulées et la bilabiale a un degré de coarticulation intermédiaire. Les variations inter-locuteurs des pentes et des ordonnées à l'origine sont réduites, désignées par les valeurs basses des écart-types. Quant aux mesures de  $R^2$ , elles sont majoritairement élevées pour les pentes estimées avec les  $F2_{onset}$  (moyenne = 0.90) et avec les  $F2_{@burst2}$  (moyenne = 0.86) alors que celles avec les  $F2_{@burst1}$  sont moins élevées notamment celles des alvéolaires (moyenne = 0.79). Ces valeurs indiquent que les points de  $F2_{@burst2}$  prédits sont mieux ajustés que ceux de  $F2_{@burst1}$ .

	$F2_{onset}$			$F2_{@burst1}$			$F2_{@burst2}$		
	P	OO	$R^2$	P	OO	$R^2$	P	OO	$R^2$
b	0.75 (0.08)	268 (100)	0.94 (0.03)	0.71 (0.09)	309 (113)	0.92 (0.05)	0.79 (0.06)	214 (76)	0.94 (0.05)
d	0.42 (0.08)	995 (126)	0.81 (0.02)	0.31 (0.09)	1192 (149)	0.62 (0.11)	0.43 (0.05)	1005 (101)	0.82 (0.05)
g	0.92 (0.07)	233 (116)	0.90 (0.03)	0.90 (0.08)	284 (127)	0.86 (0.04)	0.97 (0.06)	197 (109)	0.89 (0.03)
t	0.64 (0.04)	656 (96)	0.93 (0.02)	0.39 (0.06)	1090 (137)	0.63 (0.17)	0.49 (0.04)	922 (103)	0.77 (0.10)
k	0.98 (0.02)	130 (21)	0.93 (0.02)	0.97 (0.06)	221 (79)	0.82 (0.04)	1.00 (0.05)	155 (67)	0.86 (0.05)

TABLE 5.18 – Valeurs des moyennes et des écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des  $R^2$  des équations de locus des cinq occlusives estimées avec  $F2_{onset}$ ,  $F2_{@burst1}$  et  $F2_{@burst2}$ .

Il est à remarquer que l'effet du voisement sur les paramètres des équations de locus esti-

mées avec les  $F2_{onset}$  est plus important pour les alvéolaires que pour les vélaires puisque l'écart entre les moyennes des pentes est de 0.22 *vs.* 0.06 et celui entre les moyennes des ordonnées à l'origine est de 339 *vs.* 103 respectivement. Cette constatation laisse supposer que la quantité d'informations perdues sur le lieu d'articulation est plus conséquente pour les alvéolaires que pour les vélaires. Autrement dit, les alvéolaires sont plus influencées par l'intervalle de silence du non voisement. Comme avancé précédemment, la prédiction par extrapolation des valeurs de  $F2_{onset}$  au moment du relâchement a pour objectif de réduire l'impact de cet intervalle de silence. De ce fait, la comparaison entre les coefficients des équations de locus (pentes et ordonnées à l'origine) estimées avec les  $F2_{@burst}$  des occlusives voisées *vs.* non voisées pour chaque lieu d'articulation permet de vérifier si cet objectif est atteint ou non. La réduction de l'écart de ces coefficients par rapport à celui observé dans les équations de locus estimées avec les  $F2_{onset}$  (écart de référence) signifie que les méthodes d'extrapolation utilisées diminuent l'effet de l'intervalle du silence provoqué par le non voisement.

L'examen de l'écart entre les pentes et celui entre les ordonnées à l'origine au sein de chaque lieu d'articulation (alvéolaire et vélaire) des catégories voisée *vs.* non voisée révèle que l'écart des moyennes des pentes des alvéolaires (voisées *vs.* non voisées) a été réduit à 0.08 et à 0.06 et celui de la moyenne des ordonnées à l'origine à 102 et à 83, respectivement pour les équations de locus estimées avec  $F2_{@burst1}$  et celles avec  $F2_{@burst2}$ . Quant à l'écart des moyennes entre les vélaires, il a été réduit pour les pentes des  $F2_{@burst2}$  mais pas pour celles des  $F2_{@burst1}$ , soit 0.03 et 0.07, et il a été diminué pour les moyennes des ordonnées à l'origine des  $F2_{@burst1}$  et des  $F2_{@burst2}$  soit 63 et 42 respectivement.

D'après ces résultats, les deux méthodes de prédiction de  $F2_{@burst}$  conduisent à une réduction de l'effet du non voisement menant à des coefficients des équations de locus plus similaires et, par conséquent, à des degrés de coarticulation semblables pour les occlusives qui partagent un même lieu d'articulation. Toutefois, les équations de locus estimées avec  $F2_{@burst2}$  (la seconde méthode d'extrapolation) semblent être plus précises car les valeurs de  $R^2$  sont plus élevées et les écarts des moyennes sont plus petits par comparaison avec celles estimées avec  $F2_{@burst1}$  (la première méthode d'extrapolation).

En résumé, les coefficients des équations de locus sont en accord avec les analyses de dispersion qui mettent en évidence un degré de coarticulation maximal pour les vélaires, intermédiaire pour la bilabiale et bas pour l'alvéolaire. Les équations de locus calculées avec les  $F2_{@burst}$  prédits par les méthodes d'extrapolation réduisent l'effet de l'intervalle de silence et conduisent à des paramètres comparables.

### Seconde estimation

La seconde estimation a pour objectif d'étudier l'effet de la longueur phonologique sur le degré de coarticulation des occlusives voisées et non voisées en position initiale. Les équations de locus ont alors été calculées pour les occlusives lorsqu'elles sont produites avec les voyelles courtes et quand elles sont prononcées avec les voyelles longues. Chaque équation de locus est

réalisée avec 30 points de mesures (3 voyelles × 10 répétitions) auxquels ont été enlevés les *outliers* (2 en moyenne). Similairement à la première estimation, les équations de locus ont été calculées trois fois (avec  $F2_{onset}$ , avec  $F2_{@burst1}$  et avec  $F2_{@burst2}$ ). Au total, 70 équations de locus ont été générées par calcul (5 occlusives × 2 longueurs × 7 locuteurs).

Les résultats montrent que les pentes des occlusives produites avec les voyelles courtes sont toujours plus élevées et les ordonnées à l'origine plus basses que lorsqu'elles sont prononcées avec les voyelles longues dans les trois calculs (à l'exception de la bilabiale dans  $F2_{@burst2}$ ) (Table 5.19). De ce fait, le chevauchement coarticulatoire CV est plus important dans le contexte

		$F2_{onset}$			$F2_{@burst1}$			$F2_{@burst2}$		
	Longueur	P	OO	$R^2$	P	OO	$R^2$	P	OO	$R^2$
b	courte	0.77	236	0.95	0.73	282	0.92	0.77	235	0.93
	longue	0.73	291	0.95	0.71	308	0.93	0.79	213	0.95
d	courte	0.54	774	0.86	0.44	947	0.67	0.54	795	0.77
	longue	0.39	1081	0.84	0.28	1268	0.66	0.40	1081	0.86
g	courte	1.03	60	0.95	1.03	104	0.91	1.08	39	0.94
	longue	0.87	305	0.88	0.85	348	0.84	0.92	279	0.86
t	courte	0.70	545	0.94	0.48	947	0.65	0.52	881	0.72
	longue	0.62	692	0.93	0.37	1120	0.69	0.48	927	0.82
k	courte	1.09	-50	0.94	1.14	-62	0.82	1.18	-154	0.86
	longue	0.94	205	0.94	0.89	354	0.83	0.92	274	0.87

TABLE 5.19 – Valeurs moyennes des pentes (P), des ordonnées à l'origine (OO) et des  $R^2$  des équations de locus des cinq occlusives avec les voyelles longues et courtes dans les trois calculs ( $F2_{onset}$ ,  $F2_{@burst1}$  et  $F2_{@burst2}$ ).

des voyelles courtes que dans celui des longues. Concernant l'ordre des valeurs des coefficients des équations de locus dans chaque catégorie de longueur, il est identique à celui observé dans la première estimation (vélares > bilabiale > alvéolaires pour les pentes et alvéolaires > bilabiale > vélares pour les ordonnées à l'origine). Les valeurs des  $R^2$  sont également similaires à celles de la première estimation avec des moyennes de 0.92, de 0.79 et de 0.86 pour les calculs avec  $F2_{onset}$ , avec  $F2_{@burst1}$  et  $F2_{@burst2}$ , respectivement. L'examen des écart-types des paramètres des équations de locus soulignent qu'ils sont plus élevés dans le contexte des voyelles courtes que celui des longues dans la plupart des cas (Annexe A, Table A.3). Cette observation indique que le degré de coarticulation varie de façon plus marquée entre les locuteurs quand ils produisent les occlusives avec les voyelles courtes.

Des *anovas* de type II à mesures répétées - avec la longueur et la consonne comme variables indépendantes et le locuteur comme variable aléatoire - ont été effectuées sur les pentes et sur les ordonnées à l'origine afin d'évaluer la différence de ces paramètres en fonction de la longueur phonologique. Les résultats exposent qu'il existe un effet principal de la longueur phonologique sur les pentes et sur les ordonnées à l'origine avec  $F_{(1,6)} = 10.95, p < 0.05$  et  $F_{(1,6)} = 12.53, p < 0.05$ , respectivement. Un effet d'interaction a également été observé entre

les variables "consonne" et "longueur" signalant que l'impact de la longueur phonologique est différent sur les cinq occlusives. En effet, les analyses *post hoc* (Tukey) désignent que la différence des pentes et des ordonnées à l'origine en fonction de la longueur n'est pas significative pour le /p/ et pour le /t/.

Les écarts des moyennes des pentes et ceux des ordonnées à l'origine entre les occlusives voisées vs. non voisées ont été analysés pour chaque longueur dans chaque calcul des équations locus de manière à examiner l'efficacité des méthodes d'extrapolation dans cette série d'estimations. L'écart entre les occlusives voisées et non voisées a d'abord été calculé pour les paramètres estimés avec les  $F2_{onset}$  (écart de référence) puis comparé avec les écarts des moyennes des paramètres obtenues avec les deux  $F2_{@burst}$ . Les résultats de ces calculs mettent en avant que les écarts entre les paramètres des occlusives voisées vs. non voisées ont été réduits lorsqu'ils sont générés avec les  $F2_{@burst}$  par rapport à ceux calculés avec les  $F2_{onset}$  (à l'exception de ceux des vélares avec les voyelles courtes) (Table 5.20). La réduction de l'écart

	Alvéolaires						Vélares					
	$F2_{onset}$		$F2_{@burst1}$		$F2_{@burst2}$		$F2_{onset}$		$F2_{@burst1}$		$F2_{@burst2}$	
	P	OO	P	OO	P	OO	P	OO	P	OO	P	OO
courte	<b>0.16</b>	<b>229</b>	0.04	0	0.02	86	<b>0.06</b>	<b>110</b>	0.11	166	0.10	193
longue	<b>0.23</b>	<b>389</b>	0.09	148	0.08	154	<b>0.07</b>	<b>100</b>	0.04	6	0	5

TABLE 5.20 – Valeurs des écarts entre les moyennes des pentes (P) et entre celles des ordonnées à l'origine (OO) des occlusives voisées vs. non voisées dans chaque longueur lorsqu'elles sont générées avec les  $F2_{onset}$  (valeurs de référence), avec  $F2_{@burst1}$  et avec  $F2_{@burst2}$ .

entre les paramètres d'un même lieu d'articulation signifie qu'ils indiquent globalement le degré de coarticulation de leur lieu d'articulation. Ces résultats - et ceux de la première estimation - montrent l'importance de la prédiction des  $F2_{onset}$  au moment du relâchement pour obtenir un degré de coarticulation commun d'un même lieu d'articulation, peu importe le voisement.

En résumé, les résultats des analyses des cinq occlusives en position initiale sont en accord avec ceux de la première expérience. La longueur vocalique impacte le degré de coarticulation des occlusives de façon importante. De plus, l'effet de l'intervalle de silence peut être réduit grâce aux méthodes d'extrapolation. La section suivante évalue le degré de coarticulation et l'effet de la longueur sur les mêmes occlusives lorsqu'elles sont prononcées en position finale.

#### 5.2.4.5 Coarticulation en position finale (VC)

Les mêmes analyses sur les séquences des cinq occlusives en position initiale ont été effectuées sur les occlusives en position finale. Les analyses statistiques indiquent que les fréquences des  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$  des voyelles longues sont significativement différentes de celles des courtes, soit  $F_{(1,5)} = 22.69, p < 0.01$ ,  $F_{(1,5)} = 16.94, p < 0.01$  et  $F_{(1,5)} = 28.74, p < 0.01$ , respectivement. Quant aux  $F2_{offset}$ , les fréquences des voyelles longues sont statistiquement

distinctes de celles courtes pour les  $F2_{offset}$  et  $F3_{offset}$ , mais pas pour les  $F1_{offset}$ , avec  $F_{(1,5)} = 1.976, p = 0.219$ ,  $F_{(1,5)} = 113.02, p < 0.001$  et  $F_{(1,5)} = 15.36, p < 0.05$  respectivement pour  $F1_{offset}$ ,  $F2_{offset}$  et  $F3_{offset}$ . Comme pour les résultats en position initiale, ces analyses désignent que les deux groupes de voyelles (longues et courtes) sont spectralement distinctes. Les analyses sur les influences mutuelles entre ces voyelles et les contextes consonantiques finaux sont présentées ci-dessous.

### Analyses de dispersion

Les valeurs des fréquences des  $F2_{mid}$  et des  $F2_{offset}$  ont été affichées en fonction du temps dans chaque contexte consonantique de façon à visualiser l'influence de l'occlusive, de la voyelle et de la durée sur la relation coarticulatoire VC (Figure 5.19). La Figure 5.19 (en haut) représente la distribution des valeurs de  $F2_{mid}$  des six voyelles lorsqu'elles sont coarticulées avec les cinq occlusives étudiées en position finale. Il ressort que les fréquences du timbre /i/ ont tendance à augmenter lorsque la durée croît (pentes montantes) alors que celles du timbre /u/ décroît quand la durée augmente (pentes descendantes). Il est à remarquer que la vitesse de ces tendances dépend du contexte consonantique. Quant au timbre /a/, il reste relativement stable dans tous les contextes (pentes plates). Néanmoins, les fréquences ce timbre dans le contexte de l'occlusive /d/ sont abaissées de façon importante. En effet, cet abaissement des deux voyelles /a, a:/ est provoqué par la présence de la consonne pharyngalisée /s<sup>ʕ</sup>/ en position initiale (non contrôlée) dans les séquences /s<sup>ʕ</sup>ad/ et /s<sup>ʕ</sup>a:d/<sup>27</sup>.

Les voyelles courtes /i, u/ sont plus influencées par le contexte consonantique que leurs correspondantes longues avec des moyennes des écart-types (de tous les locuteurs) plus élevées, soit 120 vs. 85 Hz pour le /i/ et 92 vs. 84 Hz pour le /u/. Par contre, le /a/ court est moins impacté que celui long par l'occlusive suivante avec des moyennes des écart-types de 154 vs. 162 Hz. Ces résultats indiquent également que /a/ est le timbre le plus dispersé des trois. Les valeurs élevées de l'écart-type de ces deux voyelles (/a, a:/) sont provoquées par l'effet de l'agression coarticulatoire de l'occlusive pharyngalisée initiale /s<sup>ʕ</sup>/ dans le contexte de /d/.

En ce qui concerne l'influence des occlusives finales sur les voyelles précédentes, elle a été examinée par le calcul de l'écart-type des  $F2_{mid}$  des voyelles au sein de chaque contexte consonantique (Table 5.21). Les degrés de dispersion sont globalement proches dans les cinq contextes signalant que ces occlusives finales influencent peu le milieu des voyelles précédentes. Toutefois, l'alvéolaire /t/ est la plus influente suivie par la bilabiale et les vélaires. L'alvéolaire /d/ est la moins puissante des cinq occlusives à cause de l'effet de la pharyngalisation cité ci-dessus.

La Figure 5.19 (en bas) représente la distribution des fréquences des  $F2_{offset}$  des voyelles en fonction du temps dans le contexte des cinq occlusives. L'observation de cette figure montre que les points des  $F2_{offset}$  dans les contextes des alvéolaires et de la bilabiale sont plus concentrés

27. Ces séquences sont d'ailleurs les seules à contenir des occlusives pharyngalisées dans le corpus de la deuxième expérience (cf. Annexe B, Table B.2).

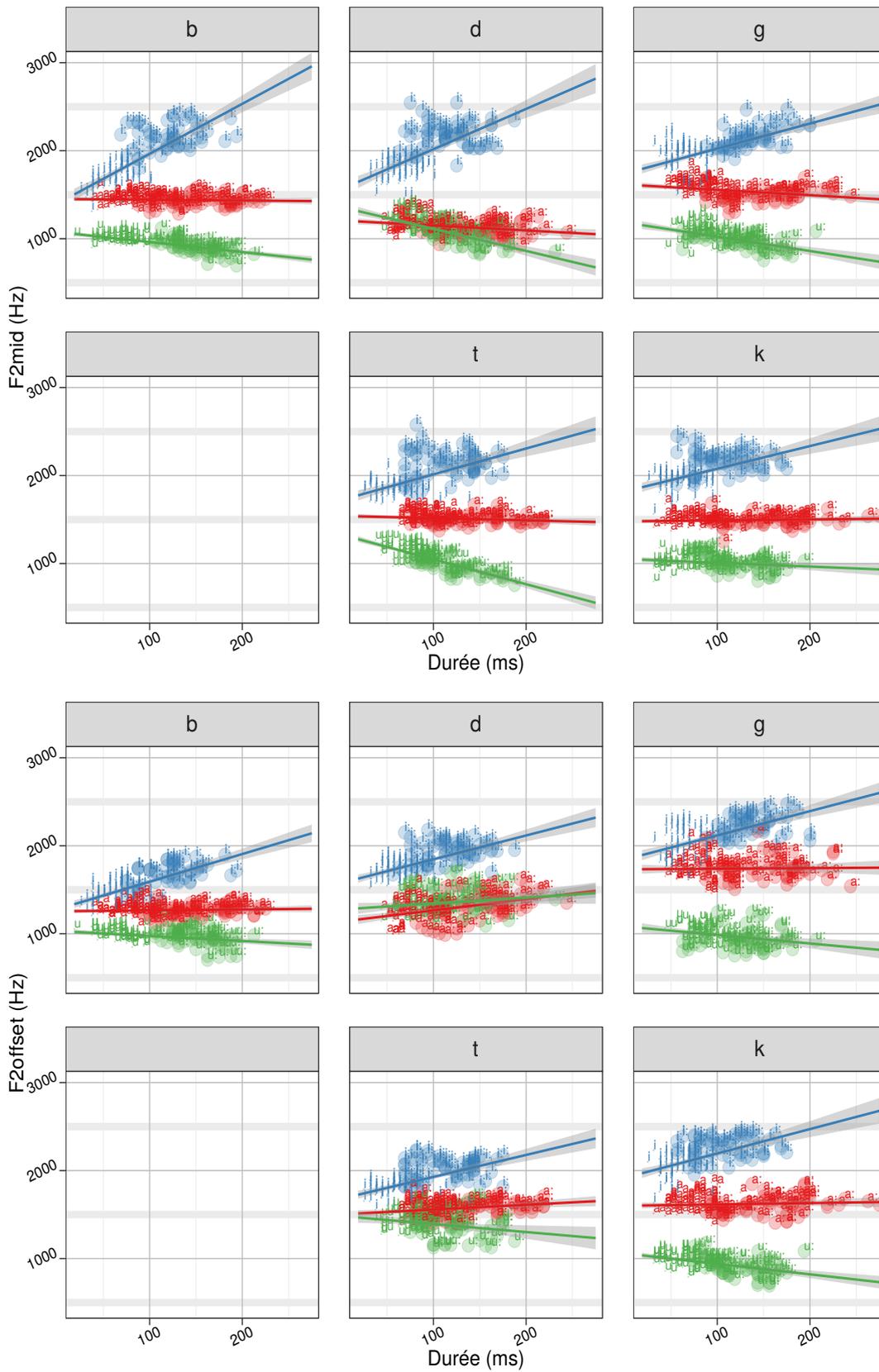


FIGURE 5.19 – Représentation de la variation des  $F2_{mid}$  (en haut) et des  $F2_{offset}$  (en bas) en fonction du temps dans chaque contexte consonantique final.

que ceux des  $F2_{mid}$  alors qu'ils ne le sont pas dans le contexte des deux vélaires. En effet, les /i/ et les /a/ subissent un abaissement de fréquence à leurs  $F2_{offset}$  alors que les /u/ restent généralement stables par comparaison avec leurs  $F2_{mid}$ . L'alvéolaire voisée semble abaisser les fréquences des  $F2_{offset}$  des /i/ et augmenter celles des /a/ - qui ont déjà été abaissées par l'effet de la pharyngalisation - et des /u/. De façon similaire, l'alvéolaire non voisée exerce un impact sur les  $F2_{offset}$  des /i/ et des /u/ pour qu'ils soient proches de ceux des /a/ (zone centrale) qui semblent être globalement stables. Quant aux vélaires, les  $F2_{offset}$  des /a/ sont élevés (élévation moins marquée pour la non voisée) et les deux autres timbres sont légèrement modifiés par rapport à ceux des  $F2_{mid}$ . De ce fait, les valeurs des écart-types de  $F2_{offset}$  dans le contexte de chaque occlusive reflète ces observations visuelles (Table 5.21). Selon ces valeurs, le degré de résistance (- de dispersion) de chaque occlusive peut être classé dans l'ordre suivant :  $t > b > d > g > k$ . Ce classement est ainsi différent de celui noté pour les occlusives en position initiale. Ce constat souligne que l'alvéolaire voisée est moins résistante lorsqu'elle est réalisée avec des /a/ pharyngalisés à cause de l'abaissement du deuxième formant.

	b	d	g	t	k
$F2_{mid}$	418	435	430	400	434
$F2_{offset}$	261	281	494	247	509

TABLE 5.21 – Valeurs des écart-types des  $F2_{mid}$  et des  $F2_{offset}$  dans les cinq contextes consonantiques en position finale.

Dans cette expérience, il est prévu que le degré de coarticulation de l'alvéolaire en position finale soit différent de celui en position initiale car son degré de résistance a été abaissé par la pharyngalisation du /a/ par le /s<sup>h</sup>/ . En d'autres termes, l'alvéolaire voisée n'aura plus les pentes les plus basses des cinq occlusives. Cette hypothèse est développée dans les analyses des équations de locus.

#### 5.2.4.6 Les équations de locus (VC)

Similairement aux analyses des occlusives en position initiale, les équations de locus ont été estimées deux fois selon l'ensemble de voyelles utilisé dans le calcul : la première estimation étant avec toutes les voyelles et la seconde étant avec les voyelles courtes et avec les voyelles longues séparément. Les équations de locus générées avec les  $F2_{@burst}$  n'ont pas été prises en compte en raison de l'absence de relâchement de l'occlusive ou d'un VOT trop grand provoquant des pentes erronées, comme expliqué dans la première expérience.

#### Première estimation

Les résultats des équations de locus des cinq occlusives effectuées avec les six voyelles exposent que les pentes de l'alvéolaire non voisée sont les plus basses suivies par celles de

sa contrepartie voisée et de la bilabiale (Table 5.22). Ces trois occlusives ont des moyennes de pentes étroitement proches signalant des degrés de coarticulation relativement bas. Quant aux vélaires, elles ont toujours les pentes les plus élevées (celle de la voisée étant toutefois plus basse). Les ordonnées à l'origine montrent l'ordre inverse, c'est-à-dire que leurs valeurs augmentent de la façon suivante :  $g < k < b < d < t$ . Les écart-types des pentes et des ordonnées à l'origine sont relativement élevés comparés à ceux de la position initiale indiquant que la variation du degré de coarticulation entre les locuteurs est plus importante en position finale qu'en position initiale. Concernant les valeurs des  $R^2$ , elles sont relativement élevées avec une moyenne de 0.89.

	P	OO	$R^2$
b	0.59 (0.10)	416 (131)	0.92 (0.05)
d	0.58 (0.11)	672 (219)	0.87 (0.05)
g	1.08 (0.11)	-41 (191)	0.90 (0.06)
t	0.55 (0.11)	787 (198)	0.83 (0.08)
k	1.15 (0.07)	-162 (102)	0.95 (0.02)

TABLE 5.22 – Valeurs moyennes et écart-types des pentes (P), des ordonnées à l'origine (OO) et des  $R^2$  des cinq occlusives en position finale produites avec les six voyelles.

Les résultats des équations de locus de cette première estimation sont globalement en accord avec ceux des analyses de dispersion sauf que la différence entre le /b/ et le /d/ est moins prononcée ici. En effet, les paramètres des équations de locus de /d/ sont également perturbés par l'effet de la pharyngalisation des voyelles /a, a:/. Par conséquent, les comparaisons de l'effet du voisement (voisée vs. non voisée) et celui de la position syllabique (initiale vs. finale) ne sont plus légitimes pour cette occlusive.

### Seconde estimation

Les équations de locus ont été calculées pour les occlusives finales lorsqu'elles sont prononcées avec les voyelles longues et avec les voyelles courtes séparément. Les résultats montrent que les pentes des occlusives produites avec les voyelles courtes sont plus élevées et les ordonnées à l'origine sont plus basses que quand elles sont réalisées avec les voyelles longues à l'exception des pentes de /t/ (Table 5.23). En effet, pour cette occlusive, seulement deux locuteurs (As et Ta) ont des pentes plus basses avec les voyelles courtes que celles avec les longues ainsi que des valeurs basses de  $R^2$  (0.49 et 56). Ces variations de pentes et de  $R^2$  bas conduisent à des valeurs élevées d'écart-types des pentes, des ordonnées à l'origine et des  $R^2$  (Table 5.23). Concernant les valeurs moyennes des  $R^2$ , elles sont généralement élevées (à l'exception de celles citées précédemment) avec une moyenne globale de 0.91.

Les analyses statistiques mettent en avant que les pentes ainsi que les ordonnées à l'origine des occlusives produites avec les voyelles longues sont significativement différentes de celles

	Longueur	P	OO	R <sup>2</sup>
b	courte	0.67 (0.08)	307 (108)	0.93 (0.05)
	longue	0.56 (0.10)	454 (119)	0.94 (0.03)
d	courte	0.71 (0.07)	458 (145)	0.92 (0.05)
	longue	0.51 (0.11)	814 (244)	0.89 (0.03)
g	courte	1.14 (0.18)	-122 (322)	0.92 (0.06)
	longue	1.08 (0.10)	-26 (166)	0.91 (0.06)
t	courte	0.53 (0.18)	779 (314)	0.78 (0.19)
	longue	0.54 (0.11)	852 (193)	0.87 (0.08)
k	courte	1.18 (0.10)	-190 (162)	0.95 (0.04)
	longue	1.08 (0.05)	-174 (81)	0.96 (0.02)

TABLE 5.23 – Valeurs moyennes et écart-types des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des cinq occlusives en position finale produites avec les voyelles courtes vs. longues.

réalisées avec les courtes soit  $F_{(1,6)} = 13.3, p < 0.05$  et  $F_{(1,6)} = 13.39, p < 0.05$  respectivement. De même, un effet d'interaction a été observé entre les facteurs "consonne" et "longueur" signifiant que l'effet principal est influencé par la consonne.

En résumé, les résultats des analyses de dispersion et des équations de locus sur les occlusives en position finale ont été influencés par l'agression coarticulation exercée par l'occlusive initiale (/s<sup>h</sup>/) sur les voyelles /a, a:/ dans le contexte de l'occlusive /d/. Le degré de coarticulation est ainsi plus élevé pour cette occlusive que pour celui de sa contrepartie non voisée, ce qui ne coïncide pas avec les résultats en position initiale. Cependant, l'impact de la longueur phonologique sur le degré de coarticulation est toujours présent en position finale.

### 5.2.5 Discussion

Un des intérêts principaux de la seconde expérience sur l'arabe jordanien était d'examiner l'impact de la longueur vocalique sur le degré de coarticulation dans le contexte des occlusives voisées et non voisées. Les résultats de cette expérience montrent que le degré de coarticulation des occlusives augmente de manière significative lorsqu'elles sont prononcées avec les voyelles courtes comparé à celui des mêmes occlusives réalisées avec les voyelles longues dans les deux positions syllabiques (initiale et finale) pour les voisées et les non voisées. Le voisement n'a donc pas d'influence sur l'effet de la longueur mais il en a une sur le degré de coarticulation.

En effet, il a été précisé que pour un lieu d'articulation, l'occlusive non voisée a un degré de coarticulation plus bas que sa contrepartie voisée (d'où + de dispersion des  $F2_{onset}$  et des pentes + élevées). Cette différence de degré de coarticulation est provoquée par l'impact moins important de l'occlusive non voisée sur le début des transitions par rapport à celui exercé par sa contrepartie voisée en raison du VOT relativement élevé des occlusives non voisées<sup>28</sup>. La mé-

28. Pour rappel, le VOT est l'intervalle séparant le bruit de relâchement (moment le plus proche de la configuration de la consonne) et le début des transitions formantiques.

thode d'extrapolation a été utilisée dans la première expérience afin de prédire la fréquence du  $F2_{onset}$  au moment du relâchement ( $F2_{@burst}$ ) pour réduire l'effet de cet intervalle. L'examen de l'opposition voisée vs. non voisée a permis de tester cette méthode de prédiction en comparant les paramètres des équations de locus estimées avec les  $F2_{@burst}$  des deux occlusives (voisée et non voisée) qui partagent le même lieu d'articulation.

Les résultats exposent que l'écart entre les paramètres des équations de locus des occlusives voisées et ceux des non voisées ayant en commun un même lieu d'articulation est réduit de façon importante quand ils sont calculés avec les  $F2_{@burst}$  (provenant des méthodes d'extrapolation) au regard des équations de locus générées avec  $F2_{onset}$ . Une série d'*anovas* de type II à mesures répétées<sup>29</sup> a été effectuée sur les pentes et les ordonnées à l'origine de manière à évaluer cette différence entre voisée et non voisée dans chacune des trois estimations (avec  $F2_{onset}$ , avec  $F2_{@burst1}$  et avec  $F2_{@burst2}$ ). Les paramètres des équations de locus de l'occlusive /b/ ont été éliminés afin de ne pas déséquilibrer les analyses étant donné qu'elle n'a pas de contrepartie non voisée. Pour les équations de locus estimées avec les  $F2_{onset}$ , les résultats<sup>30</sup> mettent en évidence que les pentes et les ordonnées à l'origine des occlusives voisées sont significativement différentes de celles de leurs contreparties non voisées soit  $F_{(1,6)} = 35.93, p < 0.001$  et  $F_{(1,6)} = 20.91, p < 0.01$  respectivement. Par contre, pour les équations de locus calculées avec les  $F2_{@burst}$ , cette différence n'est pas significative, soit  $F_{(1,6)} = 0, p = 0.994$  et  $F_{(1,6)} = 0.214, p = 0.66$  avec  $F2_{@burst1}$  et soit  $F_{(1,6)} = 0.186, p = 0.681$  et  $F_{(1,6)} = 0.17, p = 0.695$  avec  $F2_{@burst2}$ , respectivement pour les pentes et les ordonnées à l'origine.

Comme décrit précédemment, les occlusives voisées et non voisées qui partagent un même lieu d'articulation devraient - normalement - avoir des degrés de coarticulation proches au moment du relâchement puisque la configuration des deux occlusives sont globalement similaires. Cette similarité disparaît au début de la voyelle à cause des caractéristiques différentes de l'intervalle temporel qui sépare le moment du relâchement et le début de la voyelle pour chacune des deux catégories. Effectivement, le début de la voyelle est plus proche de la configuration du lieu d'articulation dans le cas des occlusives voisées que dans celui des non voisées. C'est pourquoi le début des transitions formantiques contient plus d'informations sur le lieu d'articulation quand l'occlusive est voisée que lorsqu'elle est non voisée. Le VOT plus long de cette dernière entraîne qu'au moment des premières pulsations de la voyelle, la configuration du conduit vocal ressemble moins à celle du lieu d'articulation de la consonne que pour sa contrepartie voisée. Dans le cas de l'occlusive voisée, la configuration du conduit vocal n'a pas eu le temps de se modifier autant face à un VOT plus court. Cette occlusive influence donc plus le début de la voyelle.

29. Avec le facteur "locuteur" comme variable aléatoire et les variables "voisement" et "consonne" comme variables indépendantes. En effet, une nouvelle variable "voisement" associée à chaque consonne a été créée dans le but d'examiner la dépendance du degré de coarticulation sur l'opposition voisée vs. non voisée.

30. Les résultats des analyses sur l'effet du voisement présentés sont issus des données de la première estimation, c'est-à-dire les équations de locus estimées avec toutes les voyelles. Des résultats similaires ont également été obtenus avec les équations de locus calculées en fonction de la longueur vocalique (seconde estimation).

Selon les résultats de cette expérience, la prédiction de  $F2_{onset}$  au moment du relâchement, quant à elle, a conduit à des degrés de coarticulation proches des occlusives voisée et non voisée du même lieu d'articulation. Cette observation laisse suggérer que la configuration du conduit vocal au moment du relâchement des occlusives voisée et non voisée du même lieu d'articulation est similaire.

En ce qui concerne les deux méthodes d'extrapolation utilisées, elles offrent des résultats semblables. Toutefois, ces valeurs prédites entraînent des équations de locus moins ajustées ( $R^2$ ) que celles avec les  $F2_{onset}$ . Ce phénomène peut être expliqué par les différentes variations provoquées par l'usage de l'extrapolation linéaire mais également par trois facteurs distincts. Tout d'abord, les valeurs extrapolées sont sensibles aux variations rapides de fréquence au début (et au milieu) du F2 et également au VOT lorsqu'il est relativement long. Enfin, lorsque l'occlusive a un degré de résistance coarticulatoire élevé, les points extrapolés se concentrent dans une zone limitée de fréquences dans l'espace  $F2_{@burst}-F2_{mid}$  (dans un ou deux nuages de points). Des méthodes d'extrapolation plus précises (polynomiale par exemple) ou d'autres moyens de prédiction pourraient peut-être permettre d'obtenir des points plus ajustés.

Quant à la position syllabique, les occlusives bilabiale et alvéolaires semblent être moins coarticulées en position finale qu'en position initiale (à l'exception du /d/ pour les raisons citées plus haut). Au contraire, les vélaires ont un degré de coarticulation plus élevé en position initiale qu'en position finale. Cependant, les analyses statistiques indiquent que la différence entre les deux positions syllabiques est significative pour les ordonnées à l'origine seulement et non pour les pentes avec  $F_{(1,6)} = 7.69, p < 0.05$  et  $F_{(1,6)} = 2.40, p = 0.172$  respectivement. De plus, les résultats présentés dans cette expérience soulignent que les occlusives peuvent être plus ou moins coarticulées en position finale qu'en position initiale selon le lieu d'articulation.

### 5.3 Conclusion

Les résultats globaux des deux expériences apportent des réponses similaires au sujet des questions principales posées sur les occlusives en arabe jordanien. En effet, les occlusives prononcées avec les voyelles courtes sont moins résistantes que quand elles sont réalisées avec les voyelles longues. Les voyelles courtes sont également moins résistantes aux influences consonantiques que leurs contreparties longues. Ces phénomènes révèlent la présence d'un chevauchement articulatoire important dans le contexte des voyelles courtes par comparaison à celles longues et par conséquent, des degrés de coarticulation élevés. Ces constatations sont valables pour les occlusives voisées et non voisées ainsi qu'en position syllabique initiale et finale. Selon les résultats de la première expérience, l'augmentation du degré de coarticulation des occlusives avec les voyelles courtes par rapport aux longues est liée à la durée courte de celles-ci et non simplement à cause d'une opposition spectrale. La division des voyelles longues vs. courtes en fonction de la similarité spectrale a ainsi conduit à des paramètres distincts des équations de locus. Autrement dit, les occlusives ont un degré de coarticulation plus élevé dans le contexte

des voyelles courtes que dans celui des longues peu importe la distance spectrale entre les deux catégories vocaliques (longue et courte). Ce postulat peut être expliqué par la distance temporelle plus grande qui sépare le milieu de la voyelle longue et le relâchement de la consonne par rapport à celle de la voyelle courte. Grâce à ce délai, la configuration de la voyelle a un temps plus long pour se former et pour s'éloigner de celle de la consonne précédente.

Dans les deux expériences, les paramètres des équations de locus ont offert des résultats en accord avec les analyses de dispersion<sup>31</sup>. Les deux méthodes présentent des mesures de degré de coarticulation ou de degré de résistance coarticulatoire de la consonne dans le contexte de plusieurs voyelles. D'après les résultats des deux expériences, les huit occlusives étudiées peuvent alors être classées en fonction de leur degré de résistance coarticulatoire en trois catégories. Les trois occlusives alvéolaires sont les consonnes les plus résistantes avec des pentes des équations de locus basses et des dispersions réduites de  $F2_{onset}$ . Au contraire, les vélaires et la glottale sont les moins résistantes des huit voyelles. Quant à la bilabiale et à l'uvulaire, elles ont des degrés de résistance coarticulatoire intermédiaire. De plus, il a été précisé que l'occlusive produite avec une double articulation, la pharyngalisée, est toujours la plus puissante et que la voisée est plus influente que la non voisée.

À propos de la différence du degré de coarticulation en fonction du voisement d'un lieu d'articulation (occlusives voisée vs. non voisée), les résultats des équations de locus montrent que cette différence disparaît au moyen des valeurs de  $F2_{onset}$  extrapolées ( $F2_{@burst}$ ). De ce fait, la prédiction par extrapolation linéaire a offert une solution pour automatiser l'estimation de  $F2_{onset}$  au moment du relâchement dans les grands corpus qui peut remplacer l'estimation visuelle utilisée auparavant. Il faut également signaler que le calcul des équations de locus avec les  $F2_{@burst}$  ne perturbe pas l'ordre de classement du degré de coarticulation des cinq lieux d'articulation cité plus haut.

Concernant la position syllabique (initiale ou finale) des occlusives, malgré un ordre des degrés de coarticulation identique dans les deux positions, la position finale atteste d'une augmentation de degré de coarticulation pour les vélaires et pour la glottale et d'une diminution pour les alvéolaires, la bilabiale et l'uvulaire par comparaison à celui des mêmes occlusives en position initiale d'après les résultats. Cependant, selon les tests statistiques, la différence entre les degrés de coarticulation des deux positions syllabiques n'est pas significative.

Les différences temporelles examinées dans ce chapitre sont issues d'une opposition phonologique entre les voyelles longues et courtes. Autrement dit, les variations temporelles sont intrinsèquement liées aux voyelles. Par ailleurs, les voyelles subissent d'autres types de variations temporelles qui pourraient modifier le degré de coarticulation des occlusives avec les voyelles adjacentes. Ces variations temporelles extrinsèques à la voyelle (non phonologique) peuvent se produire en changeant le débit de parole. Ces changements temporels entraînent une réduction (débit rapide) ou un rallongement de la durée (débit lent) de la réalisation des segments prononcés. Le chapitre suivant présente un protocole expérimental qui étudie si ces

31. À l'exception de ceux en position finale de la seconde expérience.

variations temporelles extrinsèques impliquent des degrés de coarticulation différents d'une façon similaire à ce qui a été remarqué pour les variations temporelles intrinsèques (c'est-à-dire + de temps, - de pentes et + de dispersion). Au final, l'impact des variations temporelles intrinsèques et extrinsèques sur la classification des occlusives sera évalué dans le septième chapitre de ce manuscrit.

## Chapitre 6

# Le débit de parole et la coarticulation

Il a été exposé précédemment que les variations temporelles d'une voyelle peuvent être associées à l'opposition de la durée phonologique dans certaines langues dont l'arabe jordanien. Cette opposition (voyelle longue vs. courte) est intrinsèquement liée à la voyelle, autrement dit, les voyelles longues nécessitent un temps de réalisation plus long que celui de leurs contreparties courtes. Les variations temporelles d'une voyelle peuvent également dépendre des variations du débit de parole. Ce phénomène extrinsèque à la voyelle change également la durée ; les voyelles produites en débit lent sont physiquement plus longues que les mêmes voyelles produites en débit rapide. L'objectif de ce chapitre est d'étudier l'impact de ces variations temporelles extrinsèques sur les qualités spectrales de la voyelle et de la consonne ainsi que sur la relation coarticulatoire entre la consonne et la voyelle en français.

Dans la première expérience, les résultats ont montré que les voyelles de l'arabe jordanien peuvent être divisées en deux sous-ensembles où l'opposition de quantité (durée) peut être accompagnée ou non d'une opposition de qualité (fréquence). D'après ces résultats, le chevauchement coarticulatoire augmente quand l'occlusive est prononcée avec les voyelles courtes (- de durée) par rapport à la même occlusive réalisée avec les voyelles longues (+ de durée), avec ou sans opposition spectrale entre les voyelles longues et courtes. Quant aux variations du débit de parole, elles conduisent à des variations de la durée vocalique et peuvent engendrer ou non des modifications spectrales sur les voyelles (Lindblom (1963a); Gay (1978a); O'Shaughnessy (1986) ; cf. section *undershoot*2.3.2). Un protocole expérimental a été conçu de manière à examiner si les variations temporelles extrinsèques liées aux changements du débit de parole amènent à des conséquences comparables à celles observées avec les variations temporelles intrinsèques présentées dans le chapitre précédent. De ce fait, la durée réduite des voyelles - en débit rapide - provoque-t-elle un chevauchement coarticulatoire plus élevé avec les occlusives adjacentes que quand elle est allongée ? Par la suite, les conséquences des deux types de variations temporelles intrinsèques et extrinsèques sur la classification des occlusives évalués dans le chapitre précédent et dans ce chapitre seront analysées dans le dernier chapitre.

## 6.1 Méthodologie

Le protocole expérimental sur le français a été créé de façon à permettre une comparaison avec les deux expériences sur l'arabe jordanien. Par conséquent, des occlusives voisées et non voisées ont été réalisées avec trois voyelles du français en variant le débit de parole afin d'obtenir des voyelles physiquement longues et courtes.

### 6.1.1 Participants

Cinq locuteurs français masculins ont participé à cette expérience âgés de 22 à 28 ans lors de l'enregistrement (Jere, Jul, Luc, Mae et Quen). Ils étaient étudiants à l'Université de Nantes originaires de la région Pays de la Loire (ouest de la France) et ne présentent aucun trouble du langage.

### 6.1.2 Stimuli

Les locuteurs devaient lire des séquences CVC enchâssées dans une phrase porteuse (« Il a dit CVC huit fois »). Les phrases à prononcer apparaissaient sur un écran d'ordinateur dans trois débits différents en trois blocs successifs (rapide puis moyen et enfin lent). Une barre de progression s'affichait progressivement sous les phrases pour donner une indication aux participants du débit voulu, avec une vitesse « cible » de 70, 140 et 210 ms par syllabe respectivement pour les débits rapide, moyen et lent. Les locuteurs ont été invités à suivre le plus possible la vitesse de la barre de progression lors de la lecture des phrases. Chaque mot était répété sept fois dans un ordre aléatoire à l'intérieur du bloc.

Un intérêt particulier a été porté sur les trois occlusives voisées et les trois non voisées du français (/b, d, g, p, t, k/). Chacune de ces occlusives apparaissait dans un mot de la langue ayant une structure CVC. Ces six occlusives cibles ont été combinées avec trois voyelles /i, a, u/ dans deux positions lexicales (initiale  $C_{cible}VC$  et finale  $CV C_{cible}$ ). L'autre consonne du mot était sélectionnée au hasard en fonction des mots existant dans le vocabulaire français. Au total, 48 mots (produits dans trois débits) ont été prononcés (*cf.* Table B.3). La fréquence des mots n'a pas été contrôlée.

### 6.1.3 Procédure

Les mêmes procédures de segmentation et d'extraction des données spectrales ont été employées que celles mises en place dans les deux expériences de l'arabe. L'enregistrement de l'expérience s'est déroulé dans une salle calme du Laboratoire de Linguistique de Nantes (LLING) à l'Université de Nantes avec le matériel utilisé dans les deux expériences de l'arabe jordanien. Au total, 3780 séquences CVC ont été segmentées dans ce corpus (6 occlusives × 3 voyelles × 2 positions syllabiques × 7 répétitions × 5 locuteurs × 3 débits). Le script R mis en œuvre dans

les deux premières expériences a été modifié pour traiter les données du français en fonction des nouvelles variables ("débit" au lieu de "longueur" et "voyelle" au lieu de "timbre"). Le système de détection des *outliers* a repéré 287 valeurs extrêmes qui ont ensuite été éliminées des analyses faisant 3493 séquences à étudier. Les valeurs extrêmes non détectées automatiquement ont été repérées et corrigées à la main.

### 6.1.4 Résultats

Les résultats de ce protocole expérimental sont présentés de manière similaire à ceux des deux premiers. Dans un premier temps, les analyses de durée ont été réalisées dans le but d'examiner l'existence de trois catégories distinctes de durée en fonction du débit de parole. Dans un deuxième temps, les analyses de l'espace vocalique des trois voyelles ont été effectuées afin de tester si les trois catégories de débit entraînent des modifications spectrales sur les voyelles. Dans un troisième temps, les deux types d'analyses de coarticulation (analyses de dispersion et équations de locus) ont été mises en place pour évaluer l'impact de ces variations temporelles sur la relation coarticulatoire CV et VC.

#### 6.1.4.1 Analyses de durée

La durée vocalique a été analysée sur l'ensemble des données des cinq locuteurs. L'examen de la distribution de la durée des trois voyelles en fonction du débit de parole désigne la présence de trois distributions qui se chevauchent de manière importante (Figure 6.1). Les moyennes globales des durées des trois voyelles sont de 91, 65 et 48 ms pour les débits lent, moyen et rapide respectivement. Cette constatation signifie que la durée moyenne des voyelles en débit rapide correspond à 52.7 % de celle des voyelles en débit lent. Ce pourcentage indique que la variation temporelle entre les voyelles réalisées en débits lent et rapide est aussi importante que celle observée entre les voyelles longues et courtes dans les deux expériences précédentes (50 % et 55.8 %). Le débit moyen, quant à lui, est entièrement chevauché par les deux autres débits. Il est également à remarquer que la durée des voyelles en débit lent est plus étendue horizontalement que celle des voyelles en débit moyen et rapide. Ce postulat peut être confirmé par les écart-types globaux de la durée dans les trois débits qui sont de 28, 19 et 16 respectivement.

L'analyse des moyennes de la durée de chaque voyelle dans les trois débits de parole signalent que le /a/ est la voyelle la plus longue au sein de chaque débit suivi par le /u/ et le /i/ (Table 6.1). Les écart-types montrent également que les durées des voyelles en débit rapide varient moins que celles des deux autres débits pour les trois voyelles.

Des variations inter-locuteurs concernant les durées ont été notées dans les trois débits et sont présentées dans la Figure 6.2. Les écarts entre les moyennes des durées des trois débits diffèrent selon le locuteur ainsi que selon la vitesse de chacun dans ces débits. Par exemple le locuteur "Luc" produit les voyelles les plus longues en débit rapide mais il réalise les voyelles les

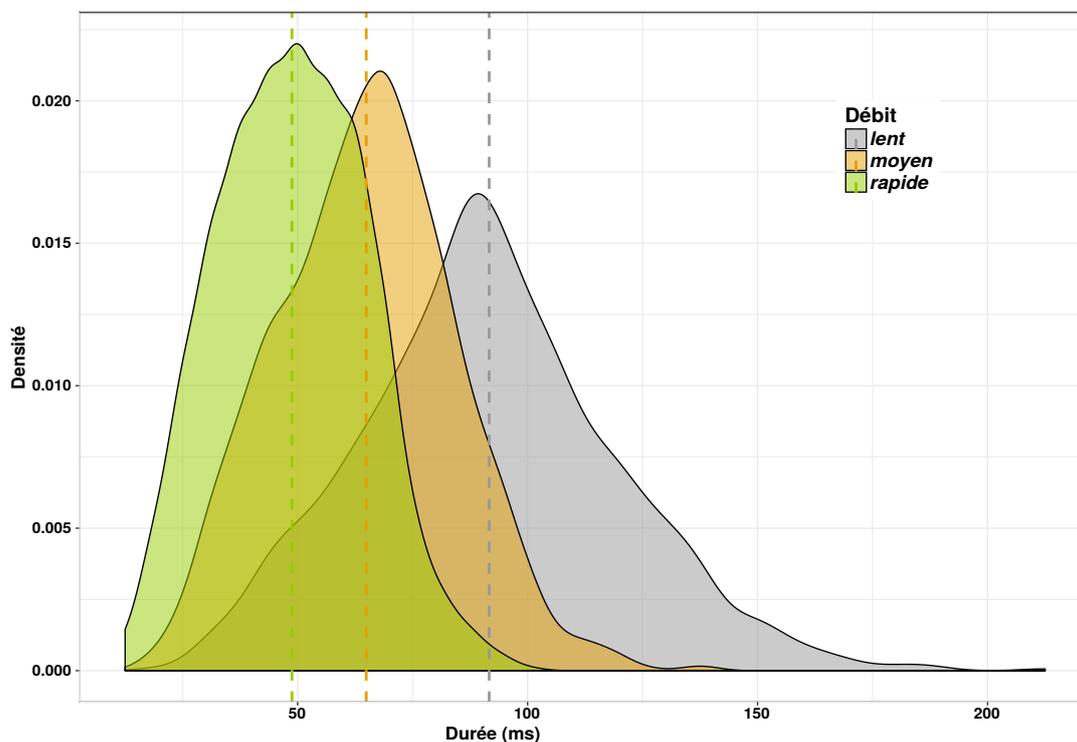


FIGURE 6.1 – Distribution de la densité de la durée des trois voyelles /i, a, u/ dans les trois débits de parole (lent, moyen, rapide) avec les barres verticales des moyennes.

Débit/V	i	a	u
lent	82 (28)	105 (22)	87 (27)
moyen	58 (19)	77 (15)	60 (17)
rapide	43 (16)	58 (13)	43 (14)

TABLE 6.1 – Valeurs des moyennes et des écart-types (entre parenthèses) des durées des trois voyelles étudiées dans les trois débits de parole.

plus courtes en débit lent alors que le locuteur "Mae" prononce les voyelles les plus longues en débits lent et moyen mais pas en débit rapide. Ces changements de durée inter-locuteurs peuvent être expliqués par les variations naturelles du débit de chaque locuteur et par leur capacité à suivre la barre de progression affichée sur l'écran lors de l'enregistrement des séquences. En effet, dans certains cas, les locuteurs vont plus vite que la barre de progression et dans d'autres, ils parlent plus lentement. Toutefois, les résultats des statistiques descriptives de la durée dans les trois débits laissent suggérer la possibilité d'obtenir trois catégories temporelles distinctes issues de la variation du débit de parole.

Une *anova* de type II à mesures répétées a été effectuée de manière à examiner la différence de durée dans les trois débits (où les variables "voyelle" et "débit" sont les variables dépendantes et le facteur "locuteur" la variable aléatoire). Les résultats indiquent que les durées vocaliques dans les trois débits sont significativement différentes avec  $F_{(2,8)} = 20.66, p < 0.001$ . Une interaction entre le débit et la voyelle a été observée signalant que l'étendue de l'effet du débit

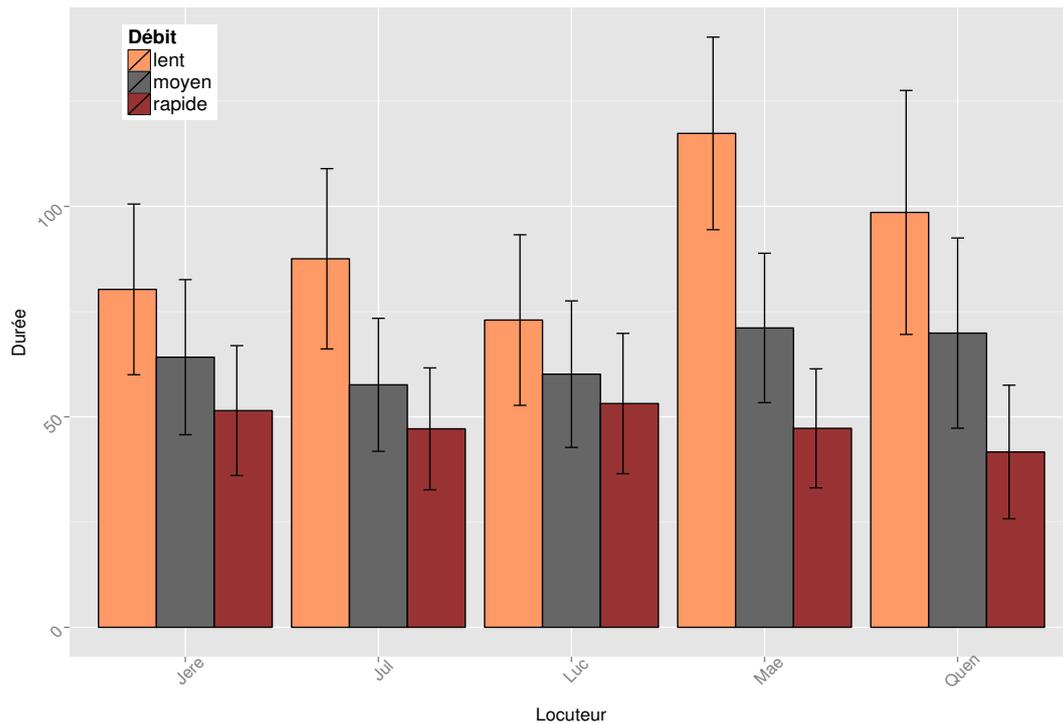


FIGURE 6.2 – Histogramme montrant les valeurs moyennes avec les barres des écart-types des trois débits de parole pour chaque locuteur.

dépend de la voyelle. De plus, les analyses *post hoc* (Tukey) exposent que la comparaison entre les durées des paires de "débit" <sup>1</sup> au sein de chaque catégorie vocalique révèle des différences statistiques dans toutes les combinaisons possibles. Cette constatation met en avant que la durée de chaque voyelle est significativement modifiée en passant d'un débit à un autre.

En résumé, la variation du débit de parole par les locuteurs a conduit à trois catégories de durée qui se chevauchent mais qui sont distinctes. L'impact de ces changements de durée sur l'espace vocalique est présenté dans la section suivante.

#### 6.1.4.2 Plan F1-F2

Les valeurs des fréquences mesurées au milieu des deux premiers formants ( $F1_{mid}$  et  $F2_{mid}$ ) des trois voyelles produites dans les trois débits par les cinq locuteurs ont été analysées (Figure 6.3). L'espace vocalique F1-F2 des voyelles /i, a, u/ affiche trois nuages de points (un pour chaque voyelle) qui se distinguent par une concentration de points différente pour chaque voyelle dont le degré peut être visualisé par la taille de l'ellipse (ellipse + grande = + de dispersion). En effet, les points de mesures de la voyelle /i/ semblent être les plus concentrés et ceux de la voyelle /a/ les plus dispersés. Les points de mesures de la voyelle /u/, quant à eux, semblent avoir un niveau de dispersion intermédiaire entre les deux autres voyelles. Le niveau de concentration des points de chaque voyelle indique son degré de stabilité face aux

1. Par exemple, pour la voyelle /i/, la comparaison de la durée entre les paires de débit correspond à i-débit lent vs. i-débit moyen, i-débit lent vs. i-débit rapide et i-débit moyen vs. i-débit rapide.

différentes influences qu'elle peut subir (débit, consonnes, variations inter-locuteurs, etc.). Le niveau de dispersion peut également être estimé en calculant l'écart-type pour chaque voyelle pour les deux axes (F1 et F2). Les résultats avancent que les F2 du /a/ sont les plus dispersés (axe horizontal) suivis par ceux du /u/ puis du /i/ avec des valeurs d'écart-types de 175, de 166 et de 140 Hz respectivement. Concernant la dispersion des points du F1, les résultats mettent

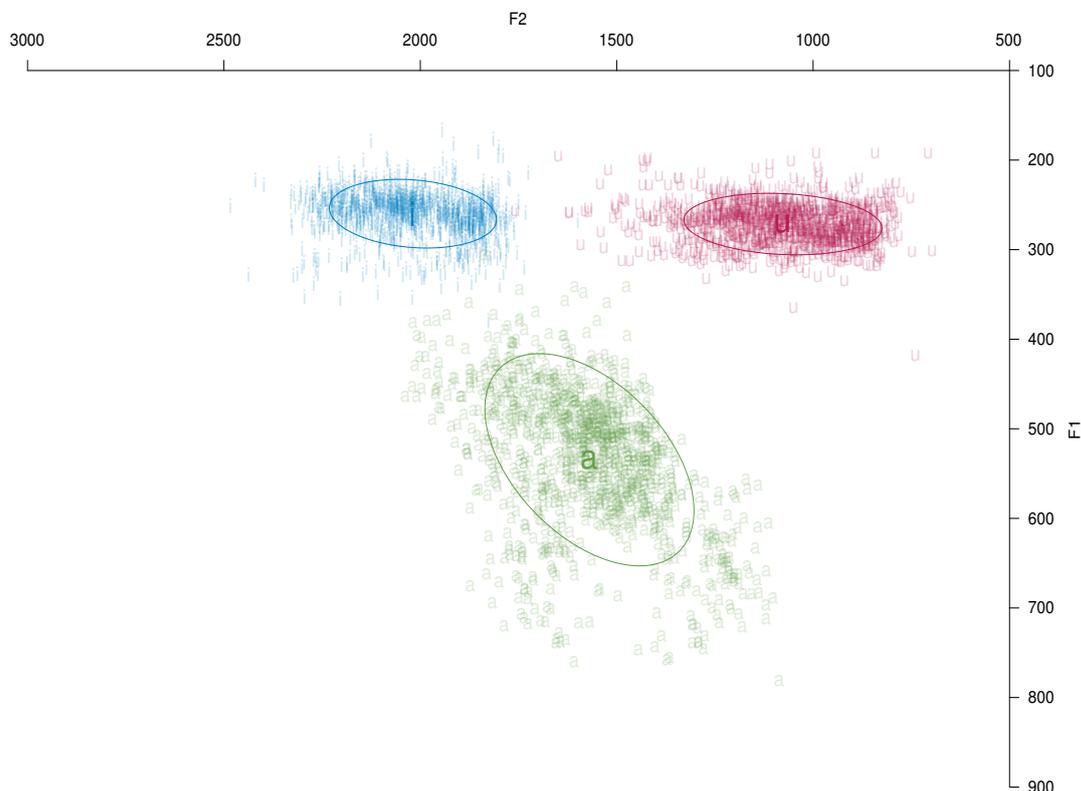


FIGURE 6.3 – Illustration de l'espace vocalique F1-F2 en Hz des trois voyelles produites par les cinq locuteurs dans les trois débits de parole.

également en évidence que le /a/ est le plus dispersé alors que le /u/ et le /i/ ont des degrés de dispersion presque semblables avec des écart-types de 78, de 23 et de 25 Hz respectivement.

Afin de visualiser l'impact des trois débits de parole sur l'espace vocalique, les fréquences des F1 et des F2 ont d'abord été transformées en Bark pour diminuer les variations inter-locuteurs et ont ensuite été affichées en fonction du débit de parole dans la Figure 6.4. Cette représentation montre que les trois voyelles ne sont pas influencées au même degré par le débit de parole. La voyelle /i/ semble rester globalement stable sur les deux axes et par conséquent, elle subit une influence faible du débit de parole. La voyelle /u/ semble être influencée horizontalement (sur l'axe des F2) mais pas verticalement (sur l'axe des F1) puisque les fréquences du F2 de cette voyelle augmentent lorsque le débit augmente. Au contraire, la voyelle /a/ semble être plus impactée verticalement (sur l'axe des F1) qu'horizontalement (sur l'axe des F2) correspondant à une baisse des fréquences du F1 lorsque le débit augmente. Cette baisse des fréquences du F1 peut être liée à l'ouverture plus réduite de la bouche en débit rapide. De ce fait, la

tendance générale des trois voyelles souligne une centralisation de l'espace vocalique quand la durée des voyelles augmente. Toutefois, cette centralisation n'a pas le même degré en fonction des trois voyelles. Cette tendance peut être observée dans la Figure 6.4 signalé par la diminution du triangle vocalique selon le débit de parole.

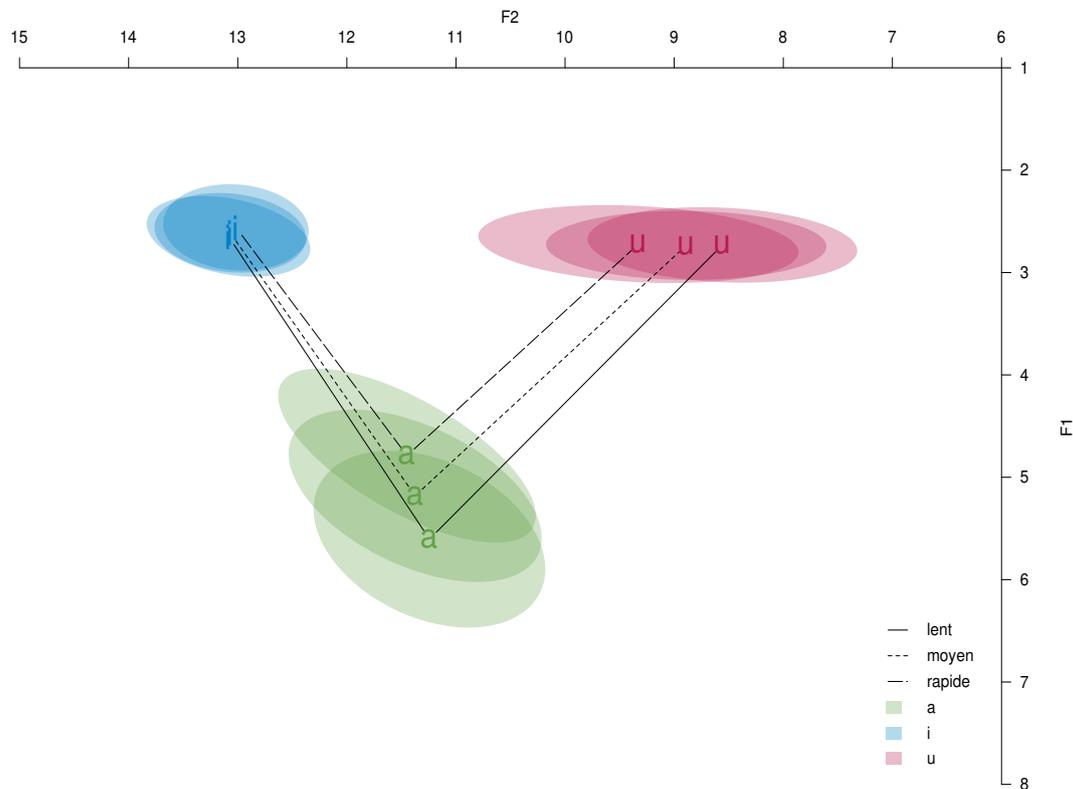


FIGURE 6.4 – Représentation de l'espace vocalique F1-F2 en Bark montrant l'influence des variations des débits de parole sur les trois voyelles.

Les analyses de variance ont été effectuées sur les fréquences des F1 et des F2 de manière à évaluer l'impact des variations du débit de parole sur les trois voyelles (avec les facteurs "voyelles" et "débit" comme variables dépendantes, le facteur "locuteur" comme variable aléatoire et les variables  $F1_{mid}$  puis  $F2_{mid}$  comme variables dépendantes). Les résultats révèlent un effet principal du débit sur les F1 et sur les F2 des voyelles étudiées, avec  $F_{(2,8)} = 21.84, p < 0.001$  et  $F_{(2,8)} = 6.69, p < 0.05$  respectivement. De plus, un effet d'interaction entre les variables "débit" et "voyelle" a été noté dans les deux tests, signifiant que l'impact du débit dépend de la voyelle. En effet, les analyses *post hoc* (Tukey) avancent que la différence entre les moyennes en fonction du débit n'est pas significative pour les voyelles /i/ et /u/ (toutes les paires V-débit) sur le F1 et pour le /i/ (toutes les paires) et le /a/ (pour la paire a-rapide vs. a-moyen) sur le F2.

En résumé, les résultats de cette section mettent en exergue que les fréquences des F1 et des F2 sont influencées par les changements du débit de parole. Néanmoins, l'étendue de cet impact diffère selon la voyelle puisque le /i/ est très peu influencé au niveau des deux formants,

le /u/ est très impacté au niveau du F2 mais très peu au niveau du F1 et le /a/ est influencé au niveau du F1 et moyennement au niveau du F2. Ces observations peuvent refléter les degrés de résistance des trois voyelles qui peuvent être classées dans l'ordre décroissant :  $i > u > a$ . Les sections suivantes présentent les degrés de résistance des voyelles et des occlusives de façon plus approfondie.

### 6.1.4.3 Coarticulation en position initiale (CV)

Les analyses de coarticulation (dispersion et équations de locus) ont été effectuées sur les séquences CVC en fonction de la position syllabique de l'occlusive cible (initiale et finale). Cette section met en avant les résultats des analyses de coarticulation en position initiale. Dans un premier temps, les fréquences mesurées au début et au milieu des trois premiers formants ont été étudiées de manière à tester l'impact des variations du débit de parole sur les voyelles et sur les consonnes. Les analyses de dispersion réalisées sur les  $F2_{mid}$  et les  $F2_{onset}$  ont ensuite permis d'évaluer le degré de résistance coarticulatoire des occlusives et des voyelles dans les trois débits. Enfin, les équations de locus ont été estimées pour les six occlusives en position initiale afin d'examiner l'influence du débit de parole sur leurs degrés de coarticulation.

Les *anovas* de type II à mesures répétées ont été générées pour le début et le milieu des trois premiers formants ( $F_{onsets}$  et  $F_{mids}$ ) de manière à explorer l'impact du débit de parole sur la voyelle et la consonne en position initiale<sup>2</sup>. Les résultats indiquent que le début des trois premiers formants n'est pas significativement influencé par le débit de parole, avec  $F_{(2,8)} = 0.14, p = 0.869$ ,  $F_{(2,8)} = 3.41, p = 0.085$  et  $F_{(2,8)} = 1.65, p = 0.251$ , respectivement pour  $F1_{onset}$ ,  $F2_{onset}$  et  $F3_{onset}$ . De plus, un effet d'interaction existe entre le débit et la voyelle dans les trois tests. Cette constatation laisse supposer que l'effet non significatif du débit est influencé par ce facteur (la voyelle). Quant au milieu des trois premiers formants, un effet significatif du débit a été noté sur  $F1_{mid}$  et sur  $F2_{mid}$ , mais pas sur  $F3_{mid}$  soit  $F_{(2,8)} = 21.77, p < 0.001$ ,  $F_{(2,8)} = 7.97, p < 0.05$  et  $F_{(2,8)} = 0.486, p = 0.632$  respectivement. Dans tous les tests, un effet d'interaction a été observé entre le débit et la voyelle, signalant que l'effet principal dépend de la voyelle.

En résumé, les analyses statistiques des trois formants mettent en exergue que les voyelles sont significativement influencées par les changements du débit de parole alors que les consonnes ne le sont pas dans le contexte des occlusives initiales.

### Analyses de dispersion (CV)

Les fréquences des  $F2_{mid}$  et des  $F2_{onset}$  ont été affichées de façon à montrer les variations spectrales des trois voyelles en fonction du temps dans les six contextes consonantiques initiaux (Figure 6.5). À propos des  $F2_{mid}$  (Figure 6.5 en haut), les fréquences de la voyelle /i/

2. Les variables indépendantes "débit" et "voyelle" ont été associées au facteur "locuteur" (variable aléatoire) pour évaluer les fréquences des  $F1_{onset}$ ,  $F2_{onset}$ ,  $F3_{onset}$ ,  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$  (variables dépendantes).

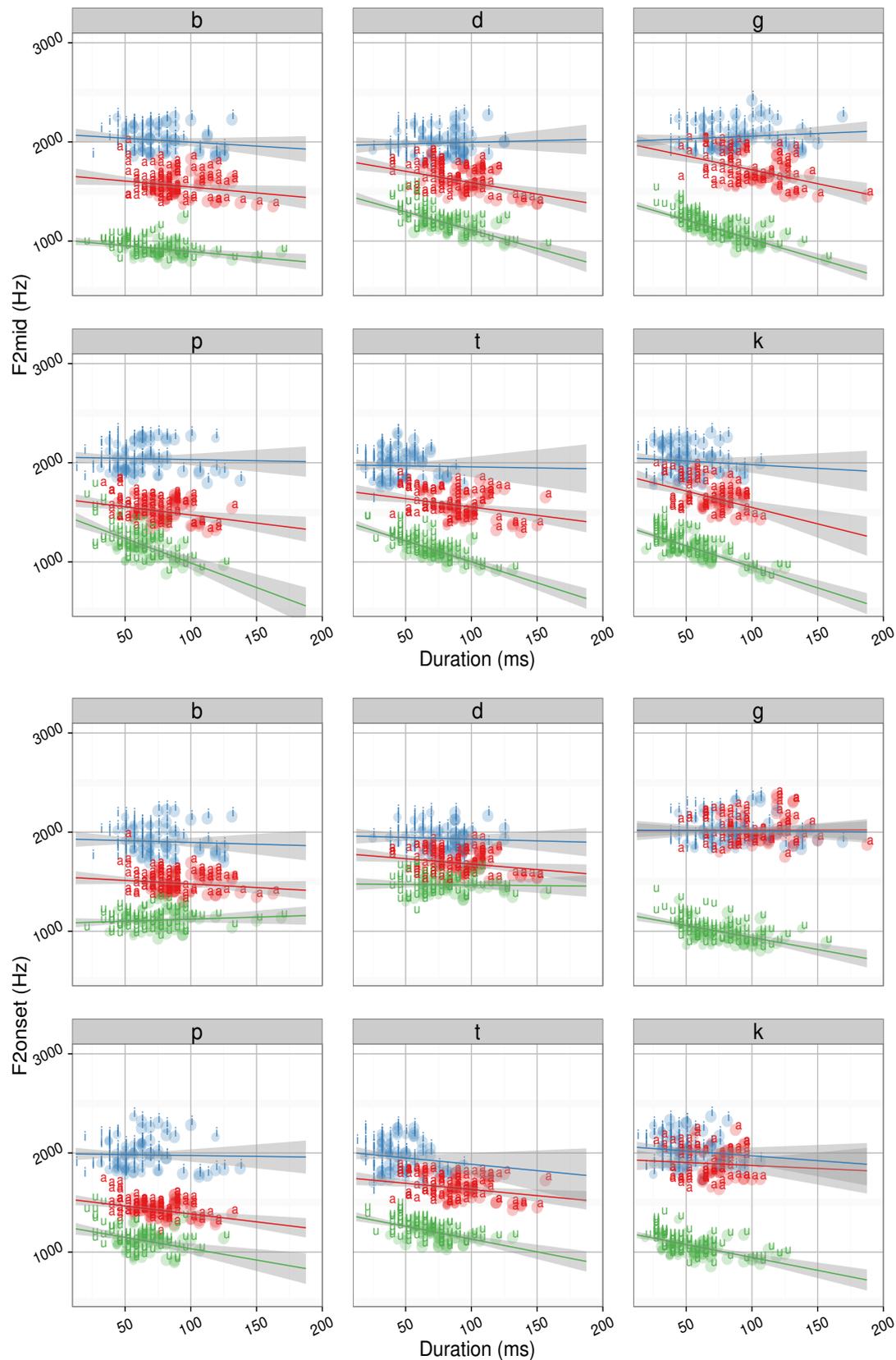


FIGURE 6.5 – Représentation de la variation des  $F2_{mid}$  (en haut) et des  $F2_{onset}$  (en bas) en fonction du temps dans chaque contexte consonantique initial pour les cinq locuteurs. Les cercles transparents de tailles différentes entourant les voyelles désignent qu’elles sont réalisées dans un débit lent (grands cercles) et moyen (petits cercles). L’absence de cercles indiquent que les voyelles sont prononcées dans un débit rapide.

ne semblent pas être impactées par les changements de durée. Cette tendance est reflétée par des pentes quasi-plates dans tous les contextes consonantiques. Les  $F2_{mid}$  des voyelles /a/ et /u/, quant à eux, ont tendance à augmenter lorsque la durée vocalique diminue dans le contexte des six occlusives. Cette observation est plus marquée pour les vélares et les alvéolaires des  $F2_{mid}$  du /a/ et moins prononcée dans le contexte du /b/ pour le /u/. Ces constatations sont en accord avec celles exposées dans les analyses de l'espace vocalique dans lesquelles l'influence du débit (durée) sur les  $F2_{mid}$  diminue dans l'ordre /u/ > /a/ > /i/ (Figure 6.4). Par ailleurs, il est à remarquer que les points de mesures des trois voyelles sont plus étendus horizontalement (durée) dans le contexte des occlusives voisées que dans celui des non voisées. Ce phénomène est dû au fait que les voyelles sont généralement plus longues dans le contexte des occlusives voisées (Hillenbrand et al., 2000).

Les écart-types pour chacune des trois voyelles soulignent que les degrés de dispersion des  $F2_{mid}$  du /a/ et du /u/ sont plus élevés en débit rapide qu'en débit lent avec 161 vs. 136 Hz et 159 vs. 135 Hz respectivement. Ces résultats signifient que les deux voyelles sont moins résistantes aux différents contextes consonantiques quand elle sont prononcées en débit rapide. Autrement dit, ces voyelles se laissent influencer plus facilement par les occlusives en débit rapide créant ainsi une dispersion plus grande. Toutefois, l'effet inverse est observé pour la voyelle /i/ où la dispersion des  $F2_{mid}$  est plus basse en débit rapide qu'en débit lent soit 126 vs. 143 Hz. D'autre part, l'écart-type global de l'ensemble des trois voyelles peut également révéler le degré de résistance des /i, a, u/ dans les trois débits de parole où un écart-type global élevé signifie une résistance vocalique importante. Les résultats montrent que les trois voyelles sont plus dispersées en débit lent et plus concentrées en débit rapide avec 416, 396, 355 Hz respectivement pour les débits lent, moyen et rapide.

Concernant l'influence coarticulatoire des six occlusives sur les voyelles, les écart-types des  $F2_{mid}$  dans chaque contexte consonantique indiquent que les alvéolaires /d, t/ sont les occlusives les plus influentes suivies par la bilabiale /p/, les vélares /g, k/ et la bilabiale /b/ (Table 6.2). L'influence moins importante de la bilabiale /b/ peut être expliquée par l'impact exercé par l'occlusive finale (non cible) /k/ sur les  $F2_{mid}$  dans les trois séquences CVC « bic, bac, bouc ». Il est à remarquer que dans les six contextes consonantiques, l'impact de l'occlusive

Débit	b	d	g	p	t	k
lent	460	374	512	398	389	432
moyen	477	338	426	372	366	390
rapide	440	307	365	349	300	351
<b>tous</b>	<b>454</b>	<b>340</b>	<b>411</b>	<b>374</b>	<b>353</b>	<b>392</b>

TABLE 6.2 – Valeurs des écart-types des  $F2_{mid}$  des trois voyelles dans le contexte des six occlusives lorsqu'elles sont prononcées en débits lent, moyen, rapide et avec tous les débits confondus (tous).

est maximal lorsqu'elle est réalisée en débit rapide (dispersion minimale). Cette constatation est

en accord avec les analyses de dispersion effectuées sur les voyelles où ces dernières sont plus sensibles (dispersées) en débit rapide qu'en débit lent. Autrement dit, une diminution de la durée entraîne une résistance coarticulatoire de la voyelle plus faible et une augmentation de l'influence consonantique sur le  $F2_{mid}$ .

Quant aux  $F2_{onset}$  (Figure 6.5 en bas), les voyelles semblent être moins impactées par l'accélération du débit de parole. En effet, les voyelles /i/ et /a/ dans tous les contextes ainsi que le /u/ avec le /b/ et le /d/ restent globalement stables quand la durée augmente avec des pentes presque plates. Au contraire, les fréquences des  $F2_{onset}$  du /u/ dans les autres contextes (/g, p, t, k/) augmentent lorsque la durée diminue de manière similaire (mais moins prononcée) à celles des  $F2_{mid}$ . Les concentrations des points semblent être plus élevées dans le contexte des alvéolaires, intermédiaires dans celui des bilabiales et basses dans celui des vélares.

Les valeurs des écart-types des  $F2_{onset}$  affichées dans la Table 6.3 reflètent les dispersions observées visuellement dans la Figure 6.5. Les valeurs des écart-types du /b/ sont très proches dans les trois débits de parole, comme c'est le cas également pour le /d/. Ces observations sont visibles sur la Figure 6.5 par des pentes globalement plates des trois voyelles (influence faible du débit). L'influence des quatre autres consonnes augmente quand le débit de parole s'accélère (– de dispersion des  $F2_{onset}$ ). Cette diminution des valeurs des écart-types quand le débit augmente est due principalement à l'élévation des fréquences des  $F2_{onset}$  de la voyelle /u/ dans ces contextes (Figure 6.5).

Débit	b	d	g	p	t	k
lent	339	219	532	390	349	484
moyen	343	220	509	372	325	455
rapide	344	224	458	354	209	418
<b>tous</b>	<b>341</b>	<b>220</b>	<b>500</b>	<b>372</b>	<b>322</b>	<b>452</b>

TABLE 6.3 – Valeurs des écart-types des  $F2_{onset}$  des trois voyelles dans le contexte des six occlusives lorsqu'elles sont prononcées en débits lent, moyen, rapide et avec tous les débits confondus (tous).

L'examen des valeurs des écart-types des  $F2_{onset}$  dans tous les débits signalent que les alvéolaires sont les occlusives les plus influentes suivies par les bilabiales puis les vélares (Table 6.3). De plus, les occlusives voisées sont plus résistantes que celles non voisées au sein de chaque lieu d'articulation à l'exception des vélares. Le classement du degré de résistance prend donc l'ordre suivant :  $d > t > b > p > k > g$ .

En résumé, les changements du débit de parole influencent le degré de résistance des voyelles et des consonnes à des niveaux différents. Les voyelles ont tendance à être moins résistantes quand les séquences sont produites en débit rapide, notamment les voyelles /u/ et /a/. Au contraire, le degré de résistance des occlusives augmente en débit rapide par rapport à celui en débit lent à l'exception des occlusives /b/ et /d/. Concernant les degrés de résistance en fonction du lieu d'articulation, ils sont élevés pour les alvéolaires, intermédiaires pour les bilabiales

et faibles pour les vélares. D'après ces résultats, il est ainsi prévu que les valeurs des pentes des équations de locus prennent l'ordre inverse de celui-ci (vélares > bilabiales > alvéolaires).

#### 6.1.4.4 Équations de locus (CV)

Deux séries d'estimation des équations de locus ont été calculées pour les occlusives en position initiale. La première correspond aux équations de locus générées pour chaque occlusive produite avec les trois débits de parole confondus. Cette estimation a pour objectif d'évaluer le degré de coarticulation des occlusives indépendamment du débit de parole. Pour la seconde, les équations de locus ont été effectuées pour chaque occlusive prononcée avec les trois voyelles dans chaque débit de parole de façon à examiner l'impact des variations du débit sur la coarticulation. Par conséquent, la première estimation est calculée avec 63 points de mesures auxquels les *outliers* sont éliminés (3 voyelles × 3 débits × 7 répétitions – 6 *outliers* en moyenne = 56 points de mesures pour chaque équation). La seconde estime, quant à elle, les équations de locus avec 21 points de mesures auxquels les *outliers* sont soustraits (3 voyelles × 1 débit × 7 répétitions – 2 *outliers* en moyenne = 19 points de mesures pour chaque équation).

Dans chacune de ces séries, les équations de locus ont été estimées avec les  $F2_{onset} \sim F2_{mid}$  et avec les  $F2_{@burst} \sim F2_{mid}$ . Les équations de locus générées avec les  $F2_{@burst}$  ont révélées des pentes trop basses (proches de zéro) ou trop élevées (supérieures à 1.5) et des  $R^2$  très bas dans certains cas. Ces valeurs extrêmes de pentes sont le résultat d'un nombre non négligeable de prédictions erronées dans les deux méthodes d'extrapolation des  $F2_{@burst}$  causé par les durées des VOT relativement longues<sup>3</sup>. Pour ces raisons, les résultats des équations de locus estimées avec les  $F2_{@burst}$  ne seront pas présentées dans le texte (voir la Table A.4 qui représente un exemple de ces valeurs et la section 6.2 pour la discussion).

#### Première estimation

Les équations de locus ont été générées pour les six occlusives /b, d, g, p, t, k/ produites avec les trois voyelles /i, a, u/ dans tous les débits de parole de manière à étudier le degré de coarticulation pour chacune de ces occlusives en position initiale. Dans le cas des occlusives voisées, les résultats montrent que les pentes ont des valeurs basses pour l'alvéolaire, intermédiaires pour la bilabiale et élevées pour les vélares tandis que les ordonnées à l'origine indiquent l'ordre inverse (Table 6.4). Les pentes et les ordonnées à l'origine des occlusives non voisées ont un ordre identique à celles de leurs contreparties voisées (alvéolaire < bilabiale < vélaire pour les pentes et vélaire < bilabiale < alvéolaire pour les ordonnées à l'origine). Toutefois, les valeurs des pentes des consonnes /p/ et /t/ sont beaucoup plus élevées et celles de leurs ordonnées à l'origine sont plus basses que leurs correspondantes voisées soulignant des degrés de coarticulation plus grands. Cette augmentation du degré de coarticulation est due à l'intervalle relativement long qui sépare le moment du relâchement et le début des transitions formantiques et donc à un

3. Par comparaison à celles de l'arabe jordanien.

	P	OO	R <sup>2</sup>
b	0.73 (0.05)	397 (60)	0.95 (0.01)
d	0.58 (0.08)	771 (124)	0.89 (0.04)
g	1.14 (0.10)	-216 (169)	0.87 (0.04)
p	0.98 (0.06)	-51 (109)	0.96 (0.02)
t	0.88 (0.10)	219 (161)	0.97 (0.02)
k	1.10 (0.10)	-139 (188)	0.91 (0.03)

TABLE 6.4 – Valeurs moyennes et écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine et des R<sup>2</sup> des occlusives initiales produites par les cinq locuteurs avec les trois voyelles dans tous les débits.

impact plus faible de la consonne sur le F2<sub>onset</sub> (moins de résistance). À propos des valeurs des R<sup>2</sup>, elles sont élevées pour toutes les consonnes avec une moyenne totale de 0.93 signalant un ajustement élevé des points dans les équations de locus des six occlusives.

### Seconde estimation

Les équations de locus ont également été estimées pour les six occlusives prononcées avec les trois voyelles mais cette fois-ci dans chaque débit de parole de manière à évaluer l'impact du changement des débits sur les paramètres des équations de locus (pente et ordonnée à l'origine). Les résultats révèlent que les pentes augmentent légèrement et les ordonnées à l'origine diminuent lorsque le débit de parole s'accélère (Table 6.5)<sup>4</sup>. Cette observation signifie que le degré

Débit	C	P	OO	R <sup>2</sup>	C	P	OO	R <sup>2</sup>
lent		0.68 (0.11)	479 (157)	0.91 (0.06)		0.97 (0.12)	0 (195)	0.95 (0.01)
moyen	<b>b</b>	0.71 (0.05)	426 (83)	0.96 (0.03)	<b>p</b>	0.99 (0.08)	-59 (126)	0.96 (0.03)
rapide		0.76 (0.05)	335 (59)	0.96 (0.03)		1.03 (0.09)	-139 (188)	0.97 (0.02)
lent		0.54 (0.05)	850 (105)	0.93 (0.04)		0.87 (0.11)	232 (179)	0.96 (0.02)
moyen	<b>d</b>	0.59 (0.13)	748 (202)	0.89 (0.07)	<b>t</b>	0.86 (0.09)	264 (132)	0.97 (0.03)
rapide		0.66 (0.15)	615 (251)	0.90 (0.09)		0.93 (0.09)	137 (144)	0.96 (0.02)
lent		1.13 (0.15)	-152 (257)	0.81 (0.06)		1.07 (0.17)	-79 (257)	0.87 (0.04)
moyen	<b>g</b>	1.15 (0.15)	-225 (273)	0.88 (0.05)	<b>k</b>	1.14 (0.08)	-192 (156)	0.90 (0.04)
rapide		1.20 (0.08)	-321 (153)	0.89 (0.07)		1.16 (0.09)	-249 (188)	0.93 (0.04)

TABLE 6.5 – Valeurs des moyennes et des écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus des occlusives initiales produites dans les trois débits de parole par les cinq locuteurs.

de coarticulation des occlusives est maximal en débit rapide. De plus, l'ordre du classement des degrés de coarticulation des six occlusives reste identique au sein de chaque débit de parole.

4. Il est à remarquer que cette tendance a été observée pour quatre locuteurs sur les cinq puisque les pentes du locuteur "Luc" baissent quand le débit s'accélère pour les occlusives /d, g, p, k/ et reste stable pour l'occlusive /t/.

Concernant les  $R^2$ , ils sont globalement élevés avec une moyenne globale de 0.92 signalant que les points de mesures sont bien ajustés aux pentes des équations de locus.

Des *anovas* de type II à mesures répétées ont été réalisées sur les pentes et les ordonnées à l'origine de façon à évaluer l'impact du débit de parole sur le degré de coarticulation (avec le débit et la consonne comme variables indépendantes, le locuteur comme variable aléatoire et la pente puis l'ordonnée à l'origine comme variables dépendantes). Les résultats indiquent qu'il existe une différence marginale entre les pentes et une différence significative entre les ordonnées à l'origine en fonction du débit de parole soit  $F_{(2,8)} = 4.15, p = 0.0579$  et  $F_{(2,8)} = 4.87, p < 0.05$ . Le résultat du test des pentes est influencé par le locuteur "Luc" car lorsque ce test est effectué sans les données de ce locuteur, l'effet du débit passe de marginal à significatif soit  $F_{(2,6)} = 32.62, p < 0.001$ . Par ailleurs, aucune interaction entre la "consonne" et le "débit" n'est observée dans les deux tests suggérant que la consonne n'impacte pas l'effet principal. Ces résultats mettent en exergue que les changements des débits de parole influencent marginalement le degré de coarticulation.

En résumé, les analyses des équations de locus estimées pour les occlusives en position initiale exposent que les degrés de coarticulation sont en accord avec ceux de la résistance coarticulatoire traités dans les analyses de dispersion (alvéolaire < bilabiale < vélaire). En outre, l'accélération du débit de parole conduit à des pentes légèrement plus élevées et des ordonnées à l'origine plus basses (+ de coarticulation).

#### 6.1.4.5 Coarticulation en position finale (VC)

Les mêmes analyses de coarticulation (dispersion et équations de locus) effectuées sur les séquences des occlusives initiales ont été réalisées sur celles des occlusives finales. Les *anovas* de type II à mesures répétées ont été générées sur le milieu et sur la fin des trois premiers formants ( $F_{mids}$  et  $F_{offsets}$ ) pour examiner l'impact du débit de parole sur la voyelle et la consonne en position finale<sup>5</sup>. Les résultats révèlent un effet significatif du débit sur le milieu du premier et du troisième formants et un effet marginal pour le deuxième, soit  $F_{(2,8)} = 18.36, p < 0.01$ ,  $F_{(2,8)} = 4.448, p = 0.0503$  et  $F_{(2,8)} = 8.797, p < 0.01$  respectivement pour  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$ . Aucun effet d'interaction n'a été observé entre le débit et la voyelle dans tous les tests signalant un effet principal indépendant de la voyelle. Quant aux fréquences prises à la fin des trois premiers formants, les résultats indiquent qu'elles ne sont pas significativement influencées par le débit de parole pour  $F2_{offset}$  et  $F3_{offset}$ , mais elles le sont pour  $F1_{offset}$  avec  $F_{(2,8)} = 0.608, p = 0.59$ ,  $F_{(2,8)} = 3.39, p = 0.0857$  et  $F_{(2,8)} = 9.76, p < 0.01$  respectivement. De plus, il existe un effet d'interaction entre le débit et la voyelle dans les trois tests, suggérant que l'effet non significatif du débit est impacté par ce facteur (la voyelle).

En résumé, similairement aux consonnes initiales, les analyses statistiques des trois formants dans le contexte des occlusives finales mettent en avant que les voyelles sont signifi-

5. Les variables indépendantes "débit" et "voyelle" ont été associées au facteur "locuteur" (variable aléatoire) pour évaluer les fréquences du  $F1_{offset}$ ,  $F2_{offset}$ ,  $F3_{offset}$ ,  $F1_{mid}$ ,  $F2_{mid}$  et  $F3_{mid}$  (variables dépendantes).

tivement influencées par les changements des débits de parole tandis que les consonnes ne le sont pas.

### Analyses de dispersion (VC)

Les fréquences des  $F2_{mid}$  et des  $F2_{offset}$  des voyelles /i, a, u/ ont été affichées en fonction du temps dans les six contextes consonantiques finaux afin de visualiser l'impact du débit de parole et celui de l'occlusive sur les variations fréquentielles au milieu et à la fin du deuxième formant. Concernant les  $F2_{mid}$ , les fréquences du /i/ semblent être peu influencées par les variations temporelles et par le contexte consonantique. Au contraire, la voyelle /a/ semble être impactée par ces deux facteurs puisque les fréquences ont tendance à augmenter quand la durée diminue dans les contextes des alvéolaires et de la bilabiale non voisée et restent généralement stables dans les autres contextes. Il est également à remarquer que les fréquences des  $F2_{mid}$  du /a/, élevées dans le contexte de la bilabiale non voisée /p/, sont abaissées dans le contexte de sa contrepartie voisée /b/. L'élévation est provoquée par l'influence de l'occlusive vélaire initiale non cible dans la séquence  $CVC_{cible}$  « cape ». Pareillement, l'abaissement dans le contexte du /b/ est également dû à l'impact de la consonne uvulaire initiale non cible (/ʁ/) dans la séquence  $CVC_{cible}$  « rab ». Quant à la voyelle /u/, les fréquences du  $F2_{mid}$  augmentent lorsque la durée de la voyelle décroît et la vitesse de cette augmentation dépend du contexte consonantique soit faible dans le contexte des vélares, intermédiaire dans celui des bilabiales et élevée dans celui des alvéolaires.

Les écart-types des  $F2_{mid}$  des voyelles /i, a, u/ ont été calculés pour chacun des débits de parole dans les six contextes consonantiques de manière à évaluer l'impact du débit sur les trois voyelles dans le contexte de chaque occlusive (Table 6.6). Les degrés de dispersion des

Débit	b	d	g	p	t	k
lent	502	410	512	455	389	490
moyen	477	356	506	421	369	506
rapide	423	304	511	411	294	479
<b>tous</b>	<b>467</b>	<b>357</b>	<b>508</b>	<b>429</b>	<b>355</b>	<b>490</b>

TABLE 6.6 – Valeurs des écart-types des  $F2_{mid}$  des trois voyelles lorsqu'elles sont produites avec les six contextes consonantiques finaux dans les trois débits de parole et dans tous les contextes.

$F2_{mid}$  (écart-types) dans chaque contexte consonantique diminuent lorsque le débit de parole augmente. Cette diminution est très faible pour les vélares<sup>6</sup>, intermédiaire pour les bilabiales et élevées pour les alvéolaires. De plus, les écart-types globaux des trois débits dans chaque contexte soulignent que les alvéolaires sont les occlusives les plus influentes suivies par les bilabiales puis les vélares (Table 6.6).

6. Le degré de dispersion reste globalement stable dans le contexte des deux occlusives avec moins de stabilité pour le /k/.

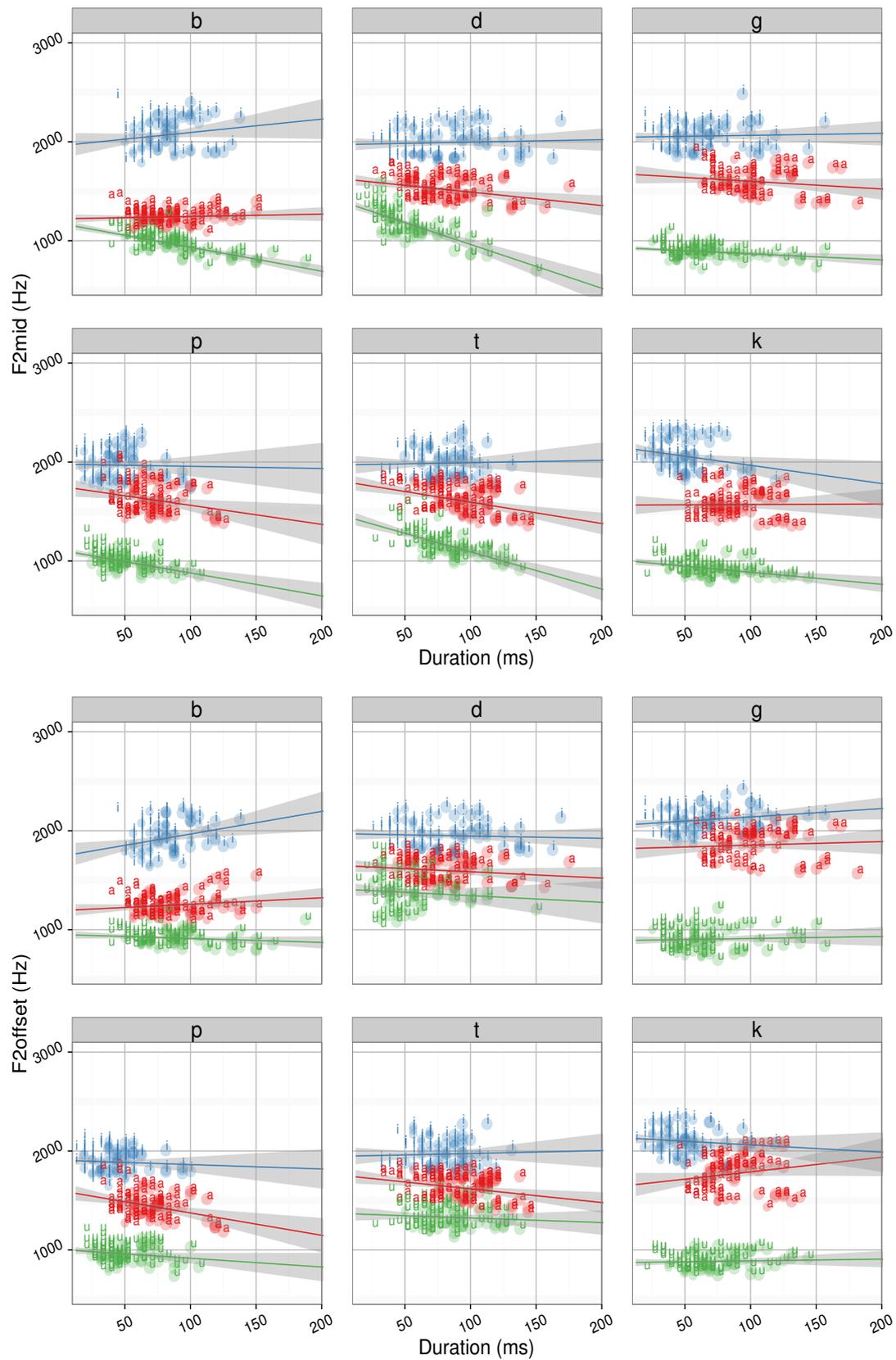


FIGURE 6.6 – Représentation de la variation des  $F2_{mid}$  (en haut) et des  $F2_{offset}$  (en bas) en fonction du temps dans chaque contexte consonantique final pour les cinq locuteurs.

Quant à l'écart-type de chaque voyelle dans l'ensemble des contextes consonantiques dans les trois débits, les résultats montrent que la résistance coarticulatoire des voyelles /a/ et /u/ baisse en passant du débit lent au débit rapide avec 176 vs. 193 Hz et 127 vs. 175 Hz respectivement (– d'écart-type = + de résistance). Au contraire, la voyelle /i/ semble être moins résistante en débit lent qu'en débit rapide avec 159 vs. 135 Hz respectivement. Par ailleurs, les valeurs de l'écart-type global de chaque voyelle dans tous les débits confondus signalent que le /i/ est la voyelle la plus résistante suivi par le /u/ puis le /a/ soit 146, 158, 187 Hz.

À propos des  $F2_{offset}$  (Figure 6.6 en bas), les fréquences des trois voyelles présentent une stabilité en fonction du temps avec des pentes plates dans la plupart des cas. Toutefois, une légère augmentation des fréquences du  $F2_{offset}$  en fonction du temps peut être notée dans les contextes du /p/ et du /t/ (Figure 6.6). Ces observations sont reflétées par des écart-types proches pour chaque voyelle dans l'ensemble des contextes consonantiques selon le débit soit 151, 156 et 150 Hz pour /i/, 235, 226 et 238 Hz pour /u/ et par des écart-types qui baissent légèrement avec 261, 239 et 224 Hz pour le /a/, respectivement pour le débit lent, moyen et rapide. En outre, les valeurs de l'écart-type global de chaque voyelle mettent en évidence que l'ordre du degré de la résistance coarticulatoire à la fin des trois voyelles ( $F2_{offset}$ ) est identique à celui de leur milieu ( $F2_{mid}$ ) soit 152, 233, 242 Hz pour /i/, /u/ et /a/ respectivement.

L'examen des écart-types des  $F2_{offset}$  dans chaque contexte consonantique en débit lent, moyen et rapide met en avant que l'influence du débit de parole reste globalement faible sur les degrés de résistance coarticulatoire des occlusives (concentration des  $F2_{offset}$ ) (Table 6.7). Toutefois, une tendance générale vers un impact plus grand des occlusives est remarquée quand le débit de parole s'accélère (– de dispersion). Elle est plus prononcée pour la bilabiale /b/ et absente pour la vélaire /k/. Les écart-types globaux des  $F2_{offset}$  dans le contexte de chaque

Débit	b	d	g	p	t	k
lent	443	282	560	413	288	534
moyen	436	281	534	397	297	553
rapide	396	269	547	396	269	537
<b>tous</b>	<b>424</b>	<b>276</b>	<b>546</b>	<b>401</b>	<b>284</b>	<b>539</b>

TABLE 6.7 – Valeurs des écart-types des  $F2_{offset}$  des trois voyelles lorsqu'elles sont réalisées avec les six occlusives finales dans les débits lent, moyen et rapide ainsi que dans l'ensemble des trois débits (tous).

consonne finale dans les trois débits confondus exposent que les alvéolaires sont les occlusives les plus résistantes suivies par les bilabiales puis les vélaire (Table 6.7). Il est à noter que les occlusives voisées sont légèrement moins résistantes que leurs contreparties non voisées. Cette constatation sur occlusives bilabiales et alvéolaires en position finale s'oppose à celle observée en position initiale où les voisées sont plus résistantes que leurs correspondantes non voisées (Table 6.3). Ce phénomène laisse supposer que la différence entre occlusives voisée vs. non voisée est largement réduite en position finale pour ces occlusives.

En résumé, les trois voyelles étudiées sont influencées à des degrés différents par les changements des débits de parole. En effet, le /i/ - qui est la voyelle la plus résistante des trois - est moins puissante en débit lent qu'en débit moyen alors que le /a/ et le /u/ présentent la situation inverse. Les occlusives finales, quant à elles, sont légèrement plus résistantes en débit rapide qu'en débit lent. Cette tendance est moins marquée que celle des mêmes occlusives en position initiale. Par ailleurs, l'ordre des degrés de résistance coarticulatoire des occlusives finales est identique à celui des occlusives initiales (alvéolaires > bilabiales > vélaires).

#### 6.1.4.6 Équations de locus (VC)

Similairement aux occlusives en position initiale, les équations de locus pour les occlusives finales ont été estimées deux fois, une première avec tous les débits confondus et une seconde avec chaque débit séparé.

##### Première estimation

Les équations de locus ont été calculées pour les occlusives /b, d, g, p, t, k/ prononcées avec les voyelles /i, a, u/ dans les trois débits de parole pour examiner le degré de coarticulation de ces occlusives en position finale. Les résultats montrent que les pentes sont basses pour les alvéolaires, intermédiaires pour les bilabiales et élevées pour les vélaires alors que les ordonnées à l'origine indiquent l'ordre inverse (Table 6.8). Les occlusives voisées ont toujours des pentes

	P	OO	R <sup>2</sup>
b	0.89 (0.09)	88 (132)	0.96 (0.03)
d	0.69 (0.11)	656 (176)	0.89 (0.10)
g	1.05 (0.05)	17 (95)	0.95 (0.02)
p	0.93 (0.08)	2 (117)	0.96 (0.02)
t	0.78 (0.04)	401 (79)	0.94 (0.01)
k	1.08 (0.05)	-66 (51)	0.95 (0.02)

TABLE 6.8 – Valeurs moyennes et écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus des occlusives finales prononcées par les cinq locuteurs dans tous les débits de parole.

plus basses et des ordonnées à l'origine plus élevées que leurs contreparties non voisées. Néanmoins, l'écart entre les occlusives voisée vs. non voisée est plus réduit que celui observé en position initiale. D'autre part, les valeurs des R<sup>2</sup> sont élevées pour toutes les équations avec une moyenne globale de 0.94.

##### Seconde estimation

Pour la seconde estimation, les équations de locus ont été calculées pour les six occlusives produites avec les trois voyelles dans chaque débit de parole de façon à étudier l'impact des

changements des débits de parole sur la coarticulation de ces occlusives en position finale. Les résultats soulignent que, dans la majorité des cas, les valeurs des pentes augmentent et celles des ordonnées à l'origine baissent quand le débit de parole s'accélère signalant une augmentation du degré de coarticulation (Table 6.8). Cette tendance est d'ailleurs plus prononcée pour les alvéolaires que pour les deux autres lieux d'articulation. Les occlusives voisées sont toujours plus basses que leurs contreparties non voisées mais avec un écart plus réduit que celui noté pour les occlusives en position initiale. De plus, l'ordre des degrés de coarticulation reste identique

Débit	C	P	OO	R <sup>2</sup>	C	P	OO	R <sup>2</sup>
lent		0.87 (0.12)	143 (167)	0.96 (0.04)		0.90 (0.11)	63 (159)	0.95 (0.03)
moyen	<b>b</b>	0.90 (0.10)	79 (148)	0.95 (0.04)	<b>p</b>	0.95 (0.10)	-28 (136)	0.96 (0.02)
rapide		0.92 (0.10)	32 (128)	0.96 (0.02)		0.95 (0.04)	-17 (47)	0.96 (0.02)
lent		0.60 (0.18)	728 (286)	0.83 (0.21)		0.69 (0.06)	557 (99)	0.92 (0.04)
moyen	<b>d</b>	0.74 (0.08)	475 (130)	0.96 (0.03)	<b>t</b>	0.78 (0.02)	413 (59)	0.96 (0.01)
rapide		0.78 (0.11)	395 (217)	0.87 (0.14)		0.89 (0.07)	186 (130)	0.96 (0.02)
lent		1.06 (0.08)	43 (120)	0.92 (0.05)		1.08 (0.11)	-8 (124)	0.92 (0.05)
moyen	<b>g</b>	1.03 (0.04)	48 (86)	0.95 (0.02)	<b>k</b>	1.08 (0.04)	-65 (57)	0.95 (0.03)
rapide		1.05 (0.05)	-29 (85)	0.97 (0.2)		1.11 (0.03)	-142 (41)	0.97 (0.01)

TABLE 6.9 – Valeurs des moyennes et des écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) de des R<sup>2</sup> des équations de locus des occlusives finales réalisées par les cinq locuteurs dans les trois débits de parole.

dans chaque débit de parole, c'est-à-dire que les alvéolaires sont les moins coarticulées suivies par les bilabiales et les vélaires ( $d < t < b < p < g < k$  pour les pentes et  $d > t > b > p > g > k$  pour les ordonnées à l'origine). Enfin, les équations de locus semblent être bien ajustées dans la plupart des cas avec une moyenne de R<sup>2</sup> de 0.94.

Des *anovas* de type II à mesures répétées ont été effectuées sur les pentes et sur les ordonnées à l'origine de manière à tester l'impact du débit de parole sur ces paramètres (avec le débit et la consonne comme variables indépendantes, le locuteur comme variable aléatoire et la pente puis l'ordonnée à l'origine comme variables dépendantes). Les résultats exposent que le débit de parole modifie significativement les pentes et les ordonnées à l'origine avec  $F_{(2,8)} = 23.36, p < 0.001$  et  $F_{(2,8)} = 34.31, p < 0.001$  respectivement. De plus, un effet d'interaction existe entre le débit et la consonne pour le test des ordonnées à l'origine mais pas pour celui des pentes, ce qui signifie que l'effet du débit dépend de la consonne pour les ordonnées à l'origine mais pas pour les pentes.

En résumé, les analyses des équations de locus des occlusives en position finale mettent en avant que le degré de coarticulation est élevé pour les vélaires, intermédiaire pour les bilabiales et bas pour les alvéolaires. Les consonnes voisées sont également plus résistantes que leurs contreparties non voisées bien que la différence entre ces deux catégories soit plus réduite en position finale qu'en position initiale. En outre, l'accélération du débit de parole augmente le degré de coarticulation des occlusives finales. Cette observation est caractérisée par l'augmen-

tation des pentes et la diminution des ordonnées à l'origine quand l'occlusive est prononcée en débit rapide comparée à la même occlusive produite en débit plus lent.

## 6.2 Discussion

### 6.2.1 Chevauchement temporel

Cette expérience a évalué l'impact des changements des débits de parole sur les occlusives voisées et non voisées /b, d, g, p, t, k/, sur les voyelles /i, a, u/ et sur le degré de coarticulation CV et VC. La technique utilisée pour faire varier le débit de parole des locuteurs a bien conduit à trois catégories de débit malgré le grand chevauchement temporel entre elles, notamment du débit moyen qui est complètement chevauché par les deux autres débits. La différence de l'influence temporelle sur les informations spectrales est alors supposée être maximale entre le débit lent et le débit rapide car l'écart entre les durées moyennes de ces deux débits est plus grand (– de chevauchement temporel) qu'entre le débit moyen vs. le débit rapide ou le débit moyen vs. le débit lent (+ de chevauchement temporel). En effet, cette hypothèse a été constatée dans la plupart des résultats des analyses de dispersion et ceux des équations de locus. Les séquences prononcées en débit moyen ont alors montré des résultats souvent intermédiaires entre les séquences des deux débits.

### 6.2.2 Position syllabique de l'occlusive

Les résultats des analyses de dispersion et des équations de locus sont globalement en accord quand l'occlusive est produite en position initiale ou finale. Cependant, quelques variations du degré de coarticulation peuvent être remarquées entre les occlusives dans les deux positions. Par exemple, les consonnes /b/ et /d/ sont plus coarticulées en position finale qu'en position initiale (pentes plus élevées). Les occlusives finales semblent aussi être moins influencées par les changements des débits de parole que celles initiales. Des analyses statistiques ont été effectuées sur les paramètres des équations de locus dans le but d'évaluer la différence entre les degrés de coarticulation des occlusives dans les deux positions syllabiques dans la première et la seconde estimations.

Pour les données de la première estimation, des *anovas* de type II à mesures répétées ont été réalisées sur les pentes et sur les ordonnées à l'origine où la position et la consonne ont été définies comme des variables indépendantes, le locuteur comme une variable aléatoire et la pente puis l'ordonnée à l'origine comme des variables dépendantes. Les résultats montrent que la différence n'est pas significative entre les pentes et les ordonnées à l'origine dans les deux positions soit  $F_{(1,4)} = 0.001, p = 0.973$  et  $F_{(1,4)} = 0.002, p = 0.969$  respectivement. Néanmoins, un effet d'interaction a été noté entre la position et la consonne signalant que l'effet principal est dépendant de cette dernière.

Pour les données de la seconde estimation, des *anovas* de type III à mesures répétées ont été effectuées sur les pentes et sur les ordonnées à l'origine avec les variables "position", "consonne" et "débit" comme variables indépendantes, le facteur "locuteur" comme variable aléatoire et les variables "pente" puis "ordonnée à l'origine" comme variables dépendantes. Les résultats mettent en avant que la différence n'est pas non plus significative entre les pentes et les ordonnées à l'origine dans les deux positions soit  $F_{(1,4)} = 0.006, p = 0.94$  et  $F_{(1,4)} = 0.024, p = 0.886$  respectivement. De plus, aucun effet d'interaction n'est observé entre le débit et la position signifiant que l'effet du débit n'est pas influencé par la position syllabique.

### 6.2.3 Opposition voisée vs. non voisée

Les occlusives voisées ont toujours révélé un degré de coarticulation plus bas que leurs contreparties non voisées (soit + de résistance coarticulatoire). Les méthodes d'extrapolation - utilisées dans les expériences de l'arabe jordanien - ont été employées afin de réduire l'effet du non voisement des occlusives sur le degré de coarticulation du lieu d'articulation. Les résultats des équations de locus estimées avec les  $F2_{@burst}$  ont signalé des valeurs de pentes trop élevées ou trop basses dans plusieurs cas. La discussion qui suit a pour objectif de soumettre quelques hypothèses qui pourraient expliquer l'origine de ces valeurs non attendues.

Parmi les quatre calculs des équations de locus avec les  $F2_{@burst}$  ( $F2_{@burst1}$  et  $F2_{@burst2}$  de la première et de la seconde estimations), les meilleurs résultats caractérisés par des  $R^2$  élevés et des pentes moins extrêmes ont été obtenus dans la première estimation avec les  $F2_{@burst2}$  (Table 6.10). La comparaison entre les équations de locus générées avec les  $F2_{@burst2}$  et celles

	P	OO	$R^2$
b	0.64 (0.06)	546 (83)	0.89 (0.06)
d	0.40 (0.11)	1074 (150)	0.69 (0.10)
g	1.26 (0.20)	-399 (350)	0.74 (0.09)
p	1.00 (0.15)	-167 (272)	0.83 (0.08)
t	0.78 (0.24)	392 (373)	0.69 (0.13)
k	1.44 (0.42)	-674 (798)	0.64 (0.09)

TABLE 6.10 – Valeurs moyennes et écarts-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des  $R^2$  des équations de locus estimées avec les  $F2_{@burst2}$ .

calculées avec les  $F2_{onset}$  met en avant que la moyenne des pentes a baissée pour les occlusives /b, d, t/, est quasiment similaire pour le /p/ et a augmentée pour les /g, k/ (Tables 6.10 et 6.4). La prédiction des  $F2_{@burst}$  ne semble pas fonctionner pour les occlusives non voisées /p, t, k/ parce qu'il est prévu que les pentes baissent significativement pour la bilabiale et l'alvéolaire et augmentent légèrement pour la vélaire quand elles sont calculées avec les  $F2_{@burst}$ . Au contraire, pour les occlusives voisées, la prédiction des  $F2_{@burst}$  semble fonctionner car les résultats sont comparables à ceux des deux expériences de l'arabe.

Les valeurs des équations de locus estimées avec les  $F2_{@burst}$ <sup>7</sup> non conformes aux attentes de l'expérience peuvent être expliquées par les VOT longs, par les variations très rapides de fréquences au début et au milieu du F2 et par le nombre réduit de séquences et de voyelles utilisées dans le calcul de chaque équation. Il a été avancé que les méthodes d'extrapolation sont sensibles à la longueur de l'intervalle séparant le moment du relâchement et le début des transitions formantiques. Des intervalles longs conduisent à des prédictions erronées et moins fiables. L'examen des durées de l'intervalle dans les trois expériences révèle qu'elles sont plus longues et qu'elles varient plus (+ d'écart-types) pour les occlusives du français (Table 6.11)<sup>8</sup>. Ces in-

Expérience	b	d	g	p	t	k	tʰ	q	ʔ
arabe 1	—	—	—	—	37 (17)	42 (14)	18 (8)	23 (13)	14 (9)
arabe 2	9 (6)	14 (7)	15 (8)	—	37 (14)	43 (13)	—	—	—
français	9 (5)	21 (11)	24 (15)	36 (21)	58 (26)	63 (25)	—	—	—

TABLE 6.11 – Valeurs des moyennes et des écart-types (entre parenthèses) des intervalles entre le bruit de relâchement et le début des transitions formantiques des occlusives étudiées dans les trois expériences.

tervalles relativement longs et variés ont pu alors conduire à des valeurs prédites moins fiables. Le nombre de réalisations et le nombre des voyelles employées dans le calcul des équations de locus jouent également un rôle important car lorsqu'une équation de locus est générée avec peu de voyelles et/ou de réalisations, sa pente est facilement influençable par les valeurs extrêmes. Par exemple, les valeurs des  $R^2$  des équations de locus calculées avec approximativement 60 réalisations sont plus élevées que celles estimées avec 20 réalisations environ (Tables 6.10, A.4). Il faut alors prendre ces facteurs en considération lors de l'application de ces méthodes de prédiction de  $F2_{@burst}$ .

### 6.3 Conclusion

L'objectif principal de ce chapitre était d'étudier l'impact des changements temporels extrinsèques sur la relation coarticulatoire entre les occlusives voisées et non voisées et les voyelles adjacentes dans les deux positions syllabiques initiale et finale. Les analyses statistiques ont montré que les voyelles sont significativement influencées par les variations du débit de parole alors que les occlusives ne sont pas impactées par ce facteur. L'influence du débit sur les voyelles est reflétée par un rétrécissement de l'espace vocalique comme affiché dans la Figure 6.4. La relation coarticulation entre les consonnes et les voyelles a ensuite été examinée par le biais des analyses de dispersion et des équations de locus. Les analyses de dispersion ont indiqué que les voyelles sont moins résistantes aux influences consonantiques et que les occlusives sont plus influentes quant le débit de parole s'accélère. Quant aux équations de locus, les résultats ont

7. De la première et de la seconde estimations (voir Table A.4)

8. Par rapport aux deux autres expériences sur l'arabe jordanien.

signalé que le chevauchement coarticulatoire augmente en passant d'un débit lent à un débit rapide.

D'après les résultats de cette expérience, les variations temporelles extrinsèques influencent de façon similaire la relation coarticulatoire entre les consonnes et les voyelles. Autrement dit, la diminution de la durée vocalique entraîne une augmentation du degré de coarticulation quelque soit l'origine de cette diminution, intrinsèque ou extrinsèque. Il serait alors intéressant d'étudier l'impact de ces deux types de variations temporelles sur la classification des occlusives en répondant aux deux questions suivantes : existe-t-il une influence des variations temporelles (des deux types) sur la classification des consonnes ? L'ajout de l'information temporelle améliore-t-il la classification des occlusives ? Ces questions seront discutées dans le chapitre suivant à l'aide des analyses discriminantes linéaires.



## Chapitre 7

# L'impact des variations temporelles sur la classification des occlusives

Les trois expériences précédentes ont offert des arguments sur l'existence d'influences temporelles intrinsèques et extrinsèques sur les propriétés spectrales de la consonne, de la voyelle et de leur relation coarticulatoire. Toutefois, il n'est pas possible de préciser si ces changements du degré de coarticulation peuvent entraver la classification du lieu d'articulation des occlusives en termes de mécanismes perceptuels. En effet, si les changements des moyennes des pentes et des ordonnées à l'origine ainsi que des degrés de dispersion dépendant de la longueur vocalique et du débit de parole confirment que les paramètres des équations de locus et le degré de dispersion sont liés au degré de coarticulation/résistance coarticulatoire, il n'est pas nécessaire que de telles variations conduisent à des difficultés dans la classification des catégories consonantiques. L'objectif de ce chapitre est d'examiner l'impact des variations temporelles intrinsèques (expériences de l'arabe jordanien) et extrinsèques (expérience du français) sur la classification des occlusives.

Les analyses discriminantes linéaires (LDA) ont été utilisées sur chacune des données des trois expériences de façon à étudier l'étendue de l'influence temporelle sur la classification des occlusives.

### 7.1 Qu'est-ce que les LDA ?

Les analyses discriminantes linéaires (LDA, de l'anglais "*Linear Discriminant Analysis*") sont une méthode de classification ayant pour but « d'expliquer et de prédire les valeurs d'une variable qualitative  $Y$  à partir de variables quantitatives et / ou qualitatives  $X = (X_1, \dots, X_p)$  »<sup>1</sup>. Cette méthode d'analyses classe donc les variables catégorielles (qualitatives) à partir des variables continues (quantitatives) qui y sont liées appelées également prédicteurs. En effet, les LDA peuvent être appréhendées selon deux approches différentes :

---

1. Cornillon et al. (2012), page 176.

1. L'estimation du taux de classification correcte ("*group membership probability*") : les tests des LDA cherchent à préciser la contribution des variables qui expliquent l'appartenance des individus (observations/variables dépendantes) à des groupes (variables catégorielles). Ils permettent ainsi de déterminer dans quelles mesures les variables quantitatives conduisent à procéder à une classification des observations correspondant à la variable catégorielle.
2. La prédiction du taux de classification correcte ("*training-test prediction*") : les LDA contribuent également à prédire la classification de nouvelles observations à partir des observations disponibles. Cette approche est composée de deux phases dont la première correspond à la phase d'apprentissage où un échantillon entraîne le modèle des LDA à reconnaître les critères appartenant aux différentes catégories. Dans la seconde (phase de test), ce modèle ("entraîné") est testé sur un autre échantillon afin d'évaluer les prédicteurs et le modèle utilisé (Baayen, 2008; Cornillon et al., 2012).

En d'autres mots, la première approche sert à classifier les observations qui se ressemblent le plus dans les catégories disponibles et la seconde prédit l'appartenance des observations aux catégories à partir d'un autre échantillon qui a entraîné le modèle.

## 7.2 La classification et la temporalité

De différents tests des LDA avec la procédure "*Leave-One-Out Cross Validation*" a été conçue sur les données acoustiques présentées dans ce manuscrit de manière à évaluer l'impact de l'aspect temporel sur la classification des catégories consonantiques. Par ailleurs, cet aspect temporel peut être vu comme catégoriel (longueur : longue & courte, débit : lent, moyen & rapide) ou continu (durée en ms). Deux hypothèses ont été formées à partir des résultats des analyses de dispersion et de ceux des équations de locus :

1. L'utilisation des  $F2_{mid}$  et des  $F2_{onset}$  comme prédicteurs du lieu d'articulation des occlusives amène généralement à un taux de classification correcte relativement élevé (par exemple, autour de 70 % dans l'étude de Sussman et al. (1991)). Ces deux prédicteurs sont alors des indices adéquats dans la classification des occlusives. Puisque la longueur vocalique et le débit de parole impactent les paramètres des équations de locus et les degrés de dispersion, l'ajout d'une information temporelle continue à un modèle des LDA pourrait augmenter le taux de la classification correcte par comparaison avec l'emploi des informations spectrales uniquement. Deux informations liées au temps peuvent être utilisées comme prédicteurs : la durée vocalique en ms et la vitesse<sup>2</sup> de la transition du F2. La première approche des LDA a été employée pour répondre à cette hypothèse.

---

2. La vitesse correspond à la vitesse couplée à la direction du déplacement. La vitesse de la transition du F2 peut être calculée par le biais de la formule suivante :

$$\frac{F2_{onset} - F2_{mid}}{\text{durée vocalique} \times 0.5} \quad (7.1)$$

2. Il a été précisé ci-dessus qu'il est possible d'entraîner un modèle des LDA sur un sous-ensemble d'observations et de prédire, en se basant sur celui-ci, la classification sur un autre sous-ensemble. Tout d'abord, il est envisageable d'entraîner un tel modèle pour la catégorisation des occlusives sur un premier sous-ensemble de voyelles longues ou courtes dans les expériences de l'arabe et sur un premier sous-ensemble de séquences réalisées en débit lent, moyen ou rapide pour celle du français. Puis, ce modèle d'apprentissage peut être évalué (testé) par la prédiction de la classification des occlusives sur le second sous-ensemble de voyelles longues ou courtes restant et sur le second des séquences prononcées en débit lent, moyen ou rapide. Si l'impact de la longueur vocalique et du débit de parole - détecté par les équations de locus et les analyses de dispersion - est fondamental pour la classification, le test de prédiction d'un sous-ensemble à *l'extérieur de son domaine d'apprentissage* devrait atteindre un niveau de précision de classification inférieur à celui appliqué à *l'intérieur de son domaine d'apprentissage*. Par exemple, lorsqu'un modèle emploie le sous-ensemble "voyelles longues" dans la phase d'apprentissage et que le test est effectué sur l'autre sous-ensemble "voyelles longues" (c'est-à-dire intérieur à son domaine d'apprentissage), le résultat doit révéler un taux de classification correcte plus grand que celui du même modèle entraîné sur le sous-ensemble "voyelles courtes" puis testé sur le sous-ensemble "voyelles longues" (c'est-à-dire extérieur à son domaine d'apprentissage). La seconde approche des LDA a été utilisée afin de vérifier cette hypothèse.

Les deux hypothèses s'interrogent ainsi sur l'effet de la temporalité dans ses deux versions, continue (première hypothèse) et catégorielle (seconde hypothèse). Ces interrogations ne permettent pas seulement d'évaluer l'importance du facteur "durée" dans la classification mais elles servent également à analyser le niveau des informations spectrales contenues dans chaque catégorie vocalique sur l'occlusive. La catégorie vocalique (voyelles longues ou courtes, débits lent, moyen ou rapide) conduisant à un taux de classification élevé indique qu'elle contient un niveau d'informations consonantiques plus grand que l'(es) autre(s) catégorie(s).

Deux séries de LDA ont été conçues dans l'objectif de répondre aux questions de temporalité exposées plus haut. Dans un premier temps, les LDA de la première approche ont été effectuées sur chacune des trois expériences de manière à examiner l'impact des variations de durée dans sa forme continue (durée physique et vitesse) sur la classification, issues de deux origines différentes : intrinsèque (première et deuxième expériences) et extrinsèque (troisième expérience). Dans un second temps, les LDA de la deuxième approche ont été réalisées sur les trois expériences de façon à tester l'impact temporel dans sa forme catégorielle (longueur et débit) sur la classification des classes consonantiques.

---

Le résultat peut être négatif indiquant que la transition est montante ( $F2_{mid}$  plus élevé que  $F2_{onset}$ ) ou positif signifiant que la transition est descendante ( $F2_{mid}$  plus bas que  $F2_{onset}$ ). La vitesse constitue alors une source importante qui représente les changements spectraux en fonction du temps.

## 7.2.1 L'impact de la durée physique sur la classification

La fonction "*lda*" de la librairie "*MASS*"<sup>3</sup> dans R a été utilisée pour effectuer les tests des LDA sur les données des trois expériences. Les variables " $F2_{onset}$ ", " $F2_{mid}$ " avec et sans la variable "durée" ou "vélocité"<sup>4</sup> comme prédicteurs ont été utilisées dans les modèles des LDA pour les trois bases de données. Par ailleurs, trois catégories consonantiques ont été testées dans chaque modèle afin de garder une vision comparative entre les trois expériences.

### 7.2.1.1 Première expérience

Les tests des LDA ont été appliqués sur les occlusives /t, t<sup>h</sup>, k/ dans le but d'évaluer la participation de l'information temporelle dans leur classification. Cette sélection a pris en compte le nombre total des réalisations et le degré de l'influence de l'occlusive sur les voyelles. C'est pourquoi le /q/ ayant moins de réalisations que les autres occlusives et le /ʔ/ n'impactant quasiment pas l'occlusive adjacente n'ont pas été traités. Les tests ont été réalisés sur l'ensemble des données et sur plusieurs sous-ensembles pour tester l'impact de l'information temporelle dans plusieurs contextes. Ces sous-ensembles ont été divisés selon le groupe de locuteurs (tous les locuteurs, les locuteurs femmes et les locuteurs hommes), la position syllabique de l'occlusive cible (les deux positions, position initiale et position finale) et la longueur vocalique (toutes les voyelles, voyelles longues et voyelles courtes). Pour chacune de ces divisions, trois modèles ont été utilisés en fonction des prédicteurs choisis : 1)  $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid}$  ; 2)  $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid} + \text{durée}$  ; 3)  $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid} + \text{vélocité}$ .

Les résultats montrent que l'ajout de la durée ou de la vitesse avec les deux prédicteurs spectraux n'améliore pas ou très peu la classification dans la majorité des conditions (Table 7.1)<sup>5</sup>. Il est à remarquer que les taux de classification des occlusives sont globalement similaires dans les trois ensembles de locuteurs et sont plus élevés lorsque les occlusives sont prononcées en position initiale. Cette différence du taux de classification selon la position syllabique baisse dans le contexte des voyelles longues. De plus, les taux de classification dans ce dernier sont toujours plus bas que ceux dans le contexte des voyelles courtes suggérant que les voyelles courtes contiennent plus d'informations sur les catégories consonantiques que leurs contreparties longues. Ce phénomène laissent supposer que l'influence exercée par les occlusives en position initiale sur les voyelles courtes suivantes est plus grande que celle sur leurs correspondantes longues. Toutefois, l'impact de ces occlusives en position finale est plus faible sur les voyelles courtes et devient plus proche de celui sur les voyelles longues.

L'examen des matrices de confusion<sup>6</sup> souligne que, pour toutes les conditions, l'occlusive

3. Ripley et al. (2017)

4. Cette nouvelle variable a été créée pour les données des trois expériences en appliquant l'équation 7.1.

5. Puisque les résultats de l'ajout de la vitesse sont similaires à ceux de la durée, ils n'ont pas été présentés dans la Table 7.1 pour faciliter la lecture du tableau.

6. Cette matrice sert à examiner le nombre des valeurs correctement et incorrectement classées dans chaque classe. Les colonnes de la matrice représentent le nombre d'occurrences de la catégorie estimée et les lignes indiquent le nombre d'occurrences de la catégorie de référence. Par exemple, les matrices de confusion ci-dessous

Locuteurs	Position	Prédicteurs	Contexte vocalique		
			V. longues	V. courtes	Toutes
Tous	Les deux	$F2_{mid}+F2_{onset/offset}$	<b>49%</b>	<b>60%</b>	<b>49%</b>
		$F2_{mid}+F2_{onset/offset}+durée$	<b>52%</b>	<b>60%</b>	<b>51%</b>
	Initiale	$F2_{mid}+F2_{onset}$	<b>50%</b>	<b>70%</b>	<b>53%</b>
		$F2_{mid}+F2_{onset}+durée$	<b>53%</b>	<b>70%</b>	<b>57%</b>
	Finale	$F2_{mid}+F2_{offset}$	<b>49%</b>	<b>53%</b>	<b>50%</b>
		$F2_{mid}+F2_{offset}+durée$	<b>51%</b>	<b>59%</b>	<b>51%</b>
Femmes	Les deux	$F2_{mid}+F2_{onset/offset}$	<b>50%</b>	<b>63%</b>	<b>50%</b>
		$F2_{mid}+F2_{onset/offset}+durée$	<b>54%</b>	<b>63%</b>	<b>55%</b>
	Initiales	$F2_{mid}+F2_{onset}$	<b>50%</b>	<b>74%</b>	<b>70%</b>
		$F2_{mid}+F2_{onset}+durée$	<b>56%</b>	<b>70%</b>	<b>62%</b>
	Finale	$F2_{mid}+F2_{offset}$	<b>51%</b>	<b>53%</b>	<b>52%</b>
		$F2_{mid}+F2_{offset}+durée$	<b>55%</b>	<b>62%</b>	<b>54%</b>
Hommes	Les deux	$F2_{mid}+F2_{onset/offset}$	<b>52%</b>	<b>59%</b>	<b>51%</b>
		$F2_{mid}+F2_{onset/offset}+durée$	<b>52%</b>	<b>59%</b>	<b>51%</b>
	Initiale	$F2_{mid}+F2_{onset}$	<b>56%</b>	<b>76%</b>	<b>52%</b>
		$F2_{mid}+F2_{onset}+durée$	<b>56%</b>	<b>75%</b>	<b>52%</b>
	Finale	$F2_{mid}+F2_{offset}$	<b>48%</b>	<b>58%</b>	<b>51%</b>
		$F2_{mid}+F2_{offset}+durée$	<b>48%</b>	<b>55%</b>	<b>50%</b>

TABLE 7.1 – Pourcentages des taux de classification correcte des trois occlusives /t, t<sup>ɿ</sup>, k/ dans le contexte des voyelles longues, des voyelles courtes et de toutes les voyelles avec les prédicteurs "F2<sub>mid</sub>" + "F2<sub>onset</sub>/F2<sub>offset</sub>" avec et sans le prédicteur "durée". Ces données ont été séparées en fonction des groupes de locuteurs : tous les locuteurs, les femmes et les hommes, ainsi qu'en fonction de la position syllabique de l'occlusive cible : dans les deux positions (initiale et finale), dans la position initiale et dans celle finale.

/t<sup>ɿ</sup>/ est la mieux classifiée des trois alors que la vélaire /k/ est souvent confondue avec les deux autres catégories consonantiques (cf. l'exemple dans la note 6). Ces classifications incorrectes peuvent être expliquées par l'existence de deux allophones de la vélaire et par le grand chevauchement spectral entre cette occlusive et les deux autres dans le domaine F2<sub>onset</sub>-F2<sub>mid</sub> affiché dans la Figure 7.1. Cette figure illustre les données des occlusives initiales prononcées par les locuteurs féminins (f) et masculins (m) et reflète la distribution des catégories consonantiques et leurs zones de fréquences chevauchées. Contrairement au /t/ et au /t<sup>ɿ</sup>/, les données du /k/ affichent deux nuages de points séparés et non continus en fonction des voyelles adja-

représente les résultats de classification des trois occlusives en position initiale produites par les locuteurs féminins (à gauche) et masculins (à droite). Les colonnes montrent le nombre d'occurrences des classes estimées (est) /k, t, t<sup>ɿ</sup>/ et les lignes désignent le nombre d'occurrences des classes de référence (réf) /k, t, t<sup>ɿ</sup>/. Les chiffres en gras représentent le nombre d'occurrences des valeurs correctement classées dans chaque catégorie.

réf/est	k	t	t <sup>ɿ</sup>	réf/est	k	t	t <sup>ɿ</sup>
k	<b>310</b>	169	0	k	<b>33</b>	252	18
t	59	<b>352</b>	38	t	341	<b>306</b>	0
t <sup>ɿ</sup>	185	45	<b>510</b>	t <sup>ɿ</sup>	191	5	<b>532</b>

centes, postérieures ou antérieures. Par ailleurs, les points des données de l'occlusive vélaire se chevauchent principalement dans la zone des fréquences basses (voyelles postérieures) avec ceux de l'occlusive pharyngalisée /t<sup>h</sup>/. Ce chevauchement est plus prononcé dans les données des hommes. Une seconde zone de chevauchement peut aussi être observé dans les fréquences moyennes et hautes (voyelles centrales et antérieures) entre le /t/ et le /k/. Quant aux deux alvéolaires, elles sont rarement chevauchées dans ce domaine où les fréquences des pharyngalisées sont toujours plus basses que celles des non pharyngalisées.

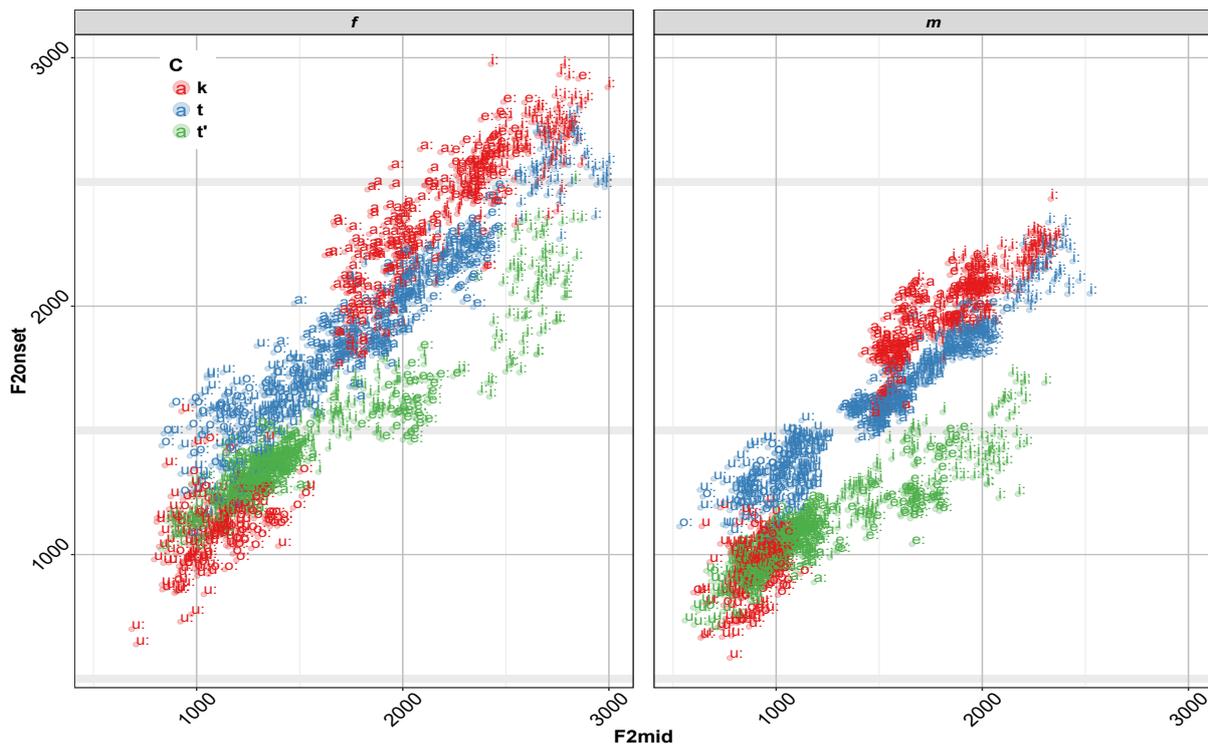


FIGURE 7.1 – Illustration des zones du chevauchement spectral dans le domaine  $F2_{onset} - F2_{mid}$  des occlusives initiale /k, t, t<sup>h</sup>/ prononcées par les locuteurs féminins (f) et masculins (m).

Concernant la représentation des occlusives en position finale ( $F2_{offset} - F2_{mid}$ ), elle montre que les trois classes consonantiques sont plus chevauchées que celles en position initiale pour les deux groupes de locuteurs (Figure 7.2). Ce chevauchement plus grand explique les taux de classification moins élevés des occlusives en position finale dans la plupart des conditions. Contrairement à la position initiale, les points de mesures de la vélaire semblent former un seul nuage de points continu qui superpose les deux autres nuages de points des /t, t<sup>h</sup>/. De ce fait, la zone de fréquences la plus chevauchée se situe au niveau des fréquences centrales (voyelles centrales /a, a:/) suivie par celle des fréquences basses (voyelles postérieures /u, u:, o:/). Pourtant, ces voyelles semblent être moins chevauchées lorsqu'elles sont produites avec les occlusives finales qu'avec celles initiales. Les voyelles antérieures (zones de fréquences hautes), quant à elles, paraissent être plus stables dans les deux positions syllabiques que les deux autres catégories vocaliques (centrales et postérieures). Ces observations indiquent un effet de la position syllabique dans la classification des occlusives désigné également par une

classification incorrecte plus grande dans les trois colonnes de la matrice de confusion des données des femmes et des hommes.

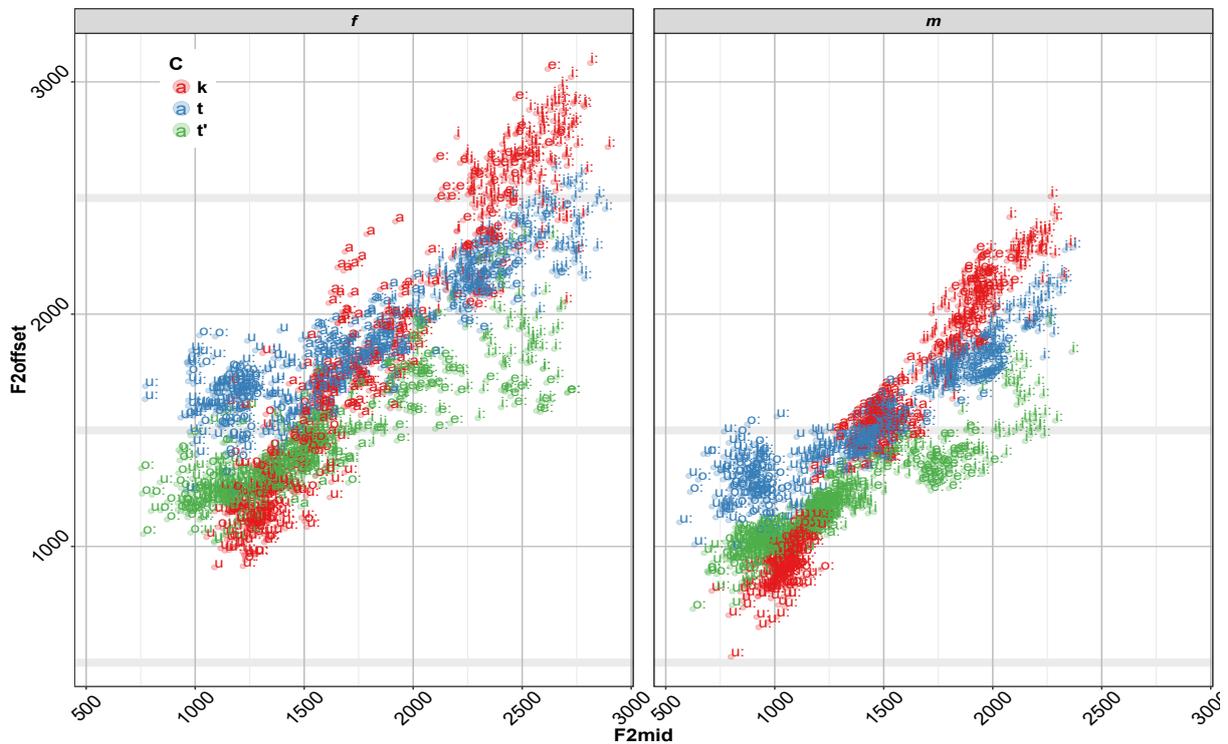


FIGURE 7.2 – Illustration du domaine  $F2_{offset} - F2_{mid}$  des occlusives /k, t, t<sup>i</sup>/ prononcées par les locuteurs féminins (f) et masculins (m) montrant les zones du chevauchement spectral.

Les résultats des LDA mettent en avant ces différences en fonction de la position syllabique de l'occlusive cible. Néanmoins, ils n'offrent pas d'arguments en faveur d'un impact temporel car l'ajout de l'information temporelle aux deux prédicteurs spectraux pour construire un domaine tri-dimensionnel (temps -  $F2_{onset}$  -  $F2_{mid}$ ) n'améliore pas la séparation entre les trois catégories. Autrement dit, l'information temporelle ajoutée ne permet pas un meilleur regroupement (appartenance) des points pour chaque classe consonantique.

Les mêmes tests des LDA ont été générés sur les données des occlusives initiales mais en utilisant cette fois-ci les  $F2_{@burst}$  au lieu des  $F2_{onset}$  comme prédicteurs. Ces nouveaux tests ont pour but d'examiner si l'effet de la durée n'a pas été caché par l'intervalle temporel relativement long des consonnes non voisées séparant le début des transitions formantiques et le relâchement de la consonne. Les résultats sont similaires à ceux des  $F2_{onset}$  (position initiale) où l'information temporelle (durée ou vélocité) n'améliore pas ou très peu le taux de classification (Table 7.2). De plus, les taux de classification correcte n'ont pas beaucoup changé et restent globalement semblables<sup>7</sup> par rapport à ceux des tests avec les  $F2_{onset}$ .

7. Il est à remarquer que dans la majorité des tests avec les  $F2_{@burst}$ , ces derniers suffisent à eux seuls à atteindre le meilleur taux de classification (sans les autres prédicteurs) alors que ce phénomène n'est pas observé avec les  $F2_{onset}$ . Cette constatation laisse supposer que les  $F2_{@burst}$  estimés contiennent plus d'informations sur l'identité de l'occlusive que les  $F2_{onset}$ .

Locuteur	Prédicteurs	Contexte vocalique		
		V. longues	V. courtes	Toutes
Tous	$F2_{mid}+F2_{@burst}$	53%	70%	68%
	$F2_{mid}+F2_{@burst}+durée$	55%	69%	60%
	$F2_{mid}+F2_{@burst}+vélocité$	54%	69%	62%
Femmes	$F2_{mid}+F2_{@burst}$	52%	70%	65%
	$F2_{mid}+F2_{@burst}+durée$	56%	69%	60%
	$F2_{mid}+F2_{@burst}+vélocité$	58%	70%	66%
Hommes	$F2_{mid}+F2_{@burst}$	56%	78%	60%
	$F2_{mid}+F2_{@burst}+durée$	56%	77%	55%
	$F2_{mid}+F2_{@burst}+vélocité$	56%	77%	55%

TABLE 7.2 – Pourcentages des taux de classification correcte des trois occlusives /t, t<sup>ʕ</sup>, k/ en position initiale dans le contexte des voyelles longues, des voyelles courtes et de toutes les voyelles avec les prédicteurs "F2<sub>mid</sub>" + "F2<sub>@burst</sub>" avec et sans le prédicteur "durée" ou "vélocité". Ces occlusives sont séparées en fonction des locuteurs : tous les locuteurs, les femmes et les hommes.

L'observation des matrices de confusion signalent également des résultats similaires à ceux notés pour les F2<sub>onset</sub>. En effet, la consonne pharyngalisée /t<sup>ʕ</sup>/ est toujours l'occlusive la mieux classifiée et sa colonne contient très peu d'occurrences erronées. Au contraire, la colonne de la vélaire contient la majorité des occurrences erronées et peu de classification correcte. Quant à l'alvéolaire /t/, le nombre des occurrences correctes et incorrectes se situe entre ceux des deux autres classes. À propos du domaine F2<sub>@burst</sub> - F2<sub>mid</sub>, il affiche une distribution de points globalement semblable à celle du domaine de F2<sub>onset</sub> - F2<sub>mid</sub> avec un degré de dispersion plus élevé dans chaque catégorie (cf. Figure C.1 dans les annexes pour illustration).

En résumé, les facteurs temporels continus (durée ou vélocité) ne jouent pas de rôle important dans l'amélioration du taux de classification des occlusives non voisées. La section suivante présente les mêmes types d'analyses sur les données de la deuxième expérience pour évaluer l'impact de ces facteurs sur les occlusives voisées.

### 7.2.1.2 Deuxième expérience

Les mêmes tests LDA avec F2<sub>onset</sub> et / ou F2<sub>offset</sub> + F2<sub>mid</sub>, F2<sub>onset</sub> et / ou F2<sub>offset</sub> + F2<sub>mid</sub> + durée et F2<sub>onset</sub> et / ou F2<sub>offset</sub> + F2<sub>mid</sub> + vélocité comme prédicteurs ont été appliqués sur les occlusives voisées /b, d, g/ produites avec les voyelles /i, a, u, i:, a:, u:/ par les sept locuteurs masculins de l'arabe jordanien. Les données traitées ont été divisées en fonction de la position syllabique de l'occlusive cible (les deux positions, la position initiale et la position finale) et en fonction du contexte vocalique (les voyelles longues, les voyelles courtes et toutes les voyelles).

Les résultats montrent que l'information temporelle (durée ou vélocité) n'augmente pas non plus le taux de classification des occlusives voisées dans toutes les conditions (Table 7.3). De

ce fait, le taux de classification est quasiment identique avec et sans les prédicteurs temporels signifiant une absence d'effet de ces derniers sur la catégorisation des occlusives<sup>8</sup>. Les taux

Position	Prédicteurs	Contexte vocalique		
		V. longues	V. courtes	Toutes
Les deux	$F2_{mid}+F2_{onset/offset}$	52%	66%	64%
	$F2_{mid}+F2_{onset/offset}+durée$	51%	66%	63%
	$F2_{mid}+F2_{onset/offset}+vélocité$	50%	67%	63%
Initiale	$F2_{mid}+F2_{onset}$	46%	64%	52%
	$F2_{mid}+F2_{onset}+durée$	49%	64%	51%
	$F2_{mid}+F2_{onset}+vélocité$	46%	64%	55%
Finale	$F2_{mid}+F2_{offset}$	68%	68%	68%
	$F2_{mid}+F2_{offset}+durée$	63%	68%	68%
	$F2_{mid}+F2_{offset}+vélocité$	62%	69%	68%

TABLE 7.3 – Pourcentages des taux de classification correcte des occlusives /b, d, g/ dans le contexte des voyelles longues, des voyelles courtes et de toutes les voyelles avec les prédicteurs " $F2_{mid}$ " et " $F2_{onset}$ " / " $F2_{offset}$ " avec et sans le prédicteur "durée" ou "vélocité". Les tests sont divisés en fonction de la position syllabique de l'occlusive cible : les deux positions, la position initiale et celle finale.

de classification des occlusives dans le contexte des voyelles courtes sont majoritairement plus élevés que ceux dans le contexte des voyelles longues. Cette différence en fonction du contexte est absente en position finale (avec les deux prédicteurs spectraux seulement). Ce phénomène laisse suggérer que les voyelles courtes contiennent des informations spectrales sur les occlusives en position initiale plus importantes que celles des voyelles longues. Autrement dit, les occlusives initiales transmettent leurs identités aux voyelles courtes suivantes plus efficacement qu'aux voyelles longues. Les occlusives en position finale, quant à elles, semblent transmettre une quantité similaire d'informations aux deux catégories vocaliques (longue et courte).

Les matrices de confusion des tests indiquent que la consonne bilabiale /b/ est l'occlusive la mieux classifiée dans toutes les conditions où elle a rarement été classée comme un /g/ et très peu comme un /d/. Le nombre d'occurrences des réponses correctes des deux autres occlusives dépend de la position syllabique ; il est plus élevé en position initiale qu'en position finale. Cette observation signale que les classes consonantiques sont plus chevauchées en position initiale qu'en position finale. Ce postulat est mis en avant dans l'illustration du domaine  $F2_{onset} - F2_{mid}$  où les points de mesures des trois classes sont plus proches les uns des autres en position initiale qu'en position finale (Figure 7.3).

Les tests des LDA ont également été appliqués sur les  $F2_{@burst}$  de manière à étudier l'im-

8. L'ajout d'une information temporelle a conduit à une très faible amélioration de 3% pour la classification des occlusives en position initiale dans le contexte des voyelles longues et dans celui de toutes les voyelles. Une légère régression de 5% peut également être remarquée pour les occlusives en position finale dans le contexte des voyelles longues.



informations temporelles dans ses versions continues (durée et vitesse) issues de l'opposition temporelle intrinsèque n'ont pas de conséquence sur le taux de classification des classes consonantiques. La section suivante est consacrée à l'évaluation de l'impact de ces informations dans le cas des variations temporelles extrinsèques.

### 7.2.1.3 Troisième expérience

Les tests des LDA ont été appliqués sur les données du français avec, tout d'abord,  $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid}$ , puis  $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid} + \text{durée}$  et enfin avec  $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid} + \text{vitesse}$  comme prédicteurs sur les occlusives voisées /b, d, g/ et non voisées /p, t, k/ produites avec les voyelles /i, a, u,/ par les cinq locuteurs masculins dans les trois débits de parole. Les données ont été divisées en fonction de la position syllabique de l'occlusive cible (soit les deux positions, la position initiale et la position finale) et en fonction du débit de parole (soit le débit lent, le débit moyen, le débit rapide et tous les débits).

Les résultats des tests sur les occlusives voisées révèlent que l'ajout de la vitesse ou de la durée aux deux prédicteurs spectraux améliore le taux de classification dans toutes les conditions dans le cas de l'emploi de la vitesse et uniquement pour les occlusives en position finale lors de l'ajout de la durée (Table 7.5). Pour les deux prédicteurs temporels, le degré de cette

Position	Prédicteurs	b d g			
		Débit de parole			
		Lent	Moyen	Rapide	Tous
Les deux	$F2_{mid}+F2_{onset}/offset$	38%	41%	49%	42%
	$F2_{mid}+F2_{onset}/offset+durée$	42%	47%	48%	45%
	$F2_{mid}+F2_{onset}/offset+vitesse$	47%	51%	53%	48%
Initiale	$F2_{mid}+F2_{onset}$	43%	41%	43%	44%
	$F2_{mid}+F2_{onset}+durée$	37%	35%	42%	42%
	$F2_{mid}+F2_{onset}+vitesse$	47%	56%	57%	48%
Finale	$F2_{mid}+F2_{offset}$	52%	49%	49%	52%
	$F2_{mid}+F2_{offset}+durée$	52%	57%	62%	55%
	$F2_{mid}+F2_{offset}+vitesse$	57%	58%	56%	57%

TABLE 7.5 – Pourcentages des classifications correctes des occlusives /b, d, g/ produites en débits lent, moyen, rapide et avec tous les débits où les prédicteurs sont les  $F2_{mid}$ , les  $F2_{onset}$  et / ou les  $F2_{offset}$  avec et sans les prédicteurs "durée" ou "vitesse". Les résultats sont séparés en fonction de la position syllabique de l'occlusive cible : les deux positions, la position initiale et celle finale.

amélioration semble dépendre du débit de parole car il est plus élevé dans le cas des séquences prononcées en débit moyen et rapide que celles réalisées en débit lent. Ces observations laissent suggérer que les facteurs temporels fournissent des indices qui permettent de compenser la "perte" des informations spectrales provoquée par les changements des débits de parole. L'im-

fact de la vélocité et de la durée sur la classification semble également être lié à la position syllabique de l'occlusive puisque l'amélioration du taux de classification est plus grande en position initiale qu'en position finale lors de l'utilisation du prédicteur "vélocité". Le prédicteur "durée", quant à lui, perturbe la classification lorsqu'il sert à classifier les occlusives en position initiale étant donné que le pourcentage diminue dans le contexte des trois débits dans cette position syllabique. Cette constatation indique que le taux baisse pour les occlusives initiales mais augmente pour celles finales quand elles sont réalisées dans les débits moyen et rapide.

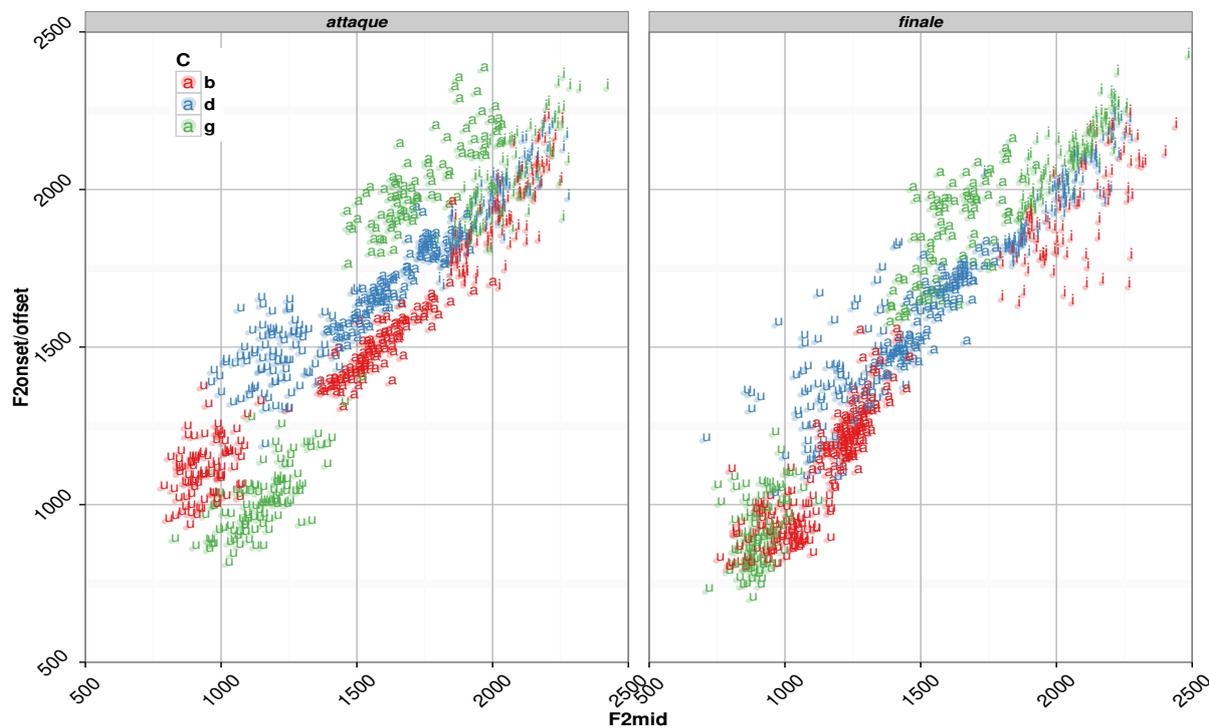
Concernant les occlusives non voisées, les résultats montrent que l'ajout du prédicteur "durée" augmente le taux de classification dans la plupart des conditions (Table 7.6). Cette augmentation est plus importante en position finale qu'en position initiale. À propos du prédicteur "vélocité", il améliore très peu la classification en position finale et ne l'influence pas positivement en position initiale.

Position	Prédicteurs	p t k			
		Lent	Moyen	Rapide	Tous
Les deux	$F2_{mid}+F2_{onset/offset}$	48%	48%	46%	42%
	$F2_{mid}+F2_{onset/offset}+durée$	55%	59%	55%	55%
	$F2_{mid}+F2_{onset/offset}+vélocité$	47%	48%	49%	48%
Initiale	$F2_{mid}+F2_{onset}$	49%	54%	54%	52%
	$F2_{mid}+F2_{onset}+durée$	53%	59%	57%	52%
	$F2_{mid}+F2_{onset}+vélocité$	45%	55%	54%	51%
Finale	$F2_{mid}+F2_{offset}$	48%	45%	53%	49%
	$F2_{mid}+F2_{offset}+durée$	60%	66%	62%	60%
	$F2_{mid}+F2_{offset}+vélocité$	50%	49%	55%	51%

TABLE 7.6 – Pourcentages des classifications correctes des occlusives /p, t, k/ produites dans les débits lent, moyen, rapide et dans tous les débits avec les prédicteurs " $F2_{mid}$ ", " $F2_{onset}$ " et / ou " $F2_{offset}$ " avec et sans le prédicteur "durée" ou "vélocité". Les résultats sont séparés en fonction de la position syllabique de l'occlusive cible : les deux positions, la position initiale et celle finale.

L'examen des matrices de confusion des tests des occlusives voisées et non voisées signalent que les bilabiales /b, p/ sont les consonnes les mieux classifiées avec des nombres de réponses correctes élevés suivies par les alvéolaires /d, t/ puis les vélares /g, k/. De plus, les occurrences des réponses incorrectes sont plus nombreuses pour les vélares et les alvéolaires que pour les bilabiales. Ces réponses incorrectes peuvent être expliquées par les degrés de chevauchement importants entre les trois classes consonantiques observés dans les domaines  $F2_{onset} - F2_{mid}$  et  $F2_{offset} - F2_{mid}$  (Figure 7.4). En effet, les points de mesures des occlusives voisées paraissent être plus chevauchés en position initiale qu'en position finale alors que l'effet inverse semble être observé dans le cas des occlusives non voisées. En outre, les zones de concentration des nuages de points sont modifiées en fonction de la position syllabique de l'occlusive pour

(a) Occlusives voisées



(b) Occlusives non voisées

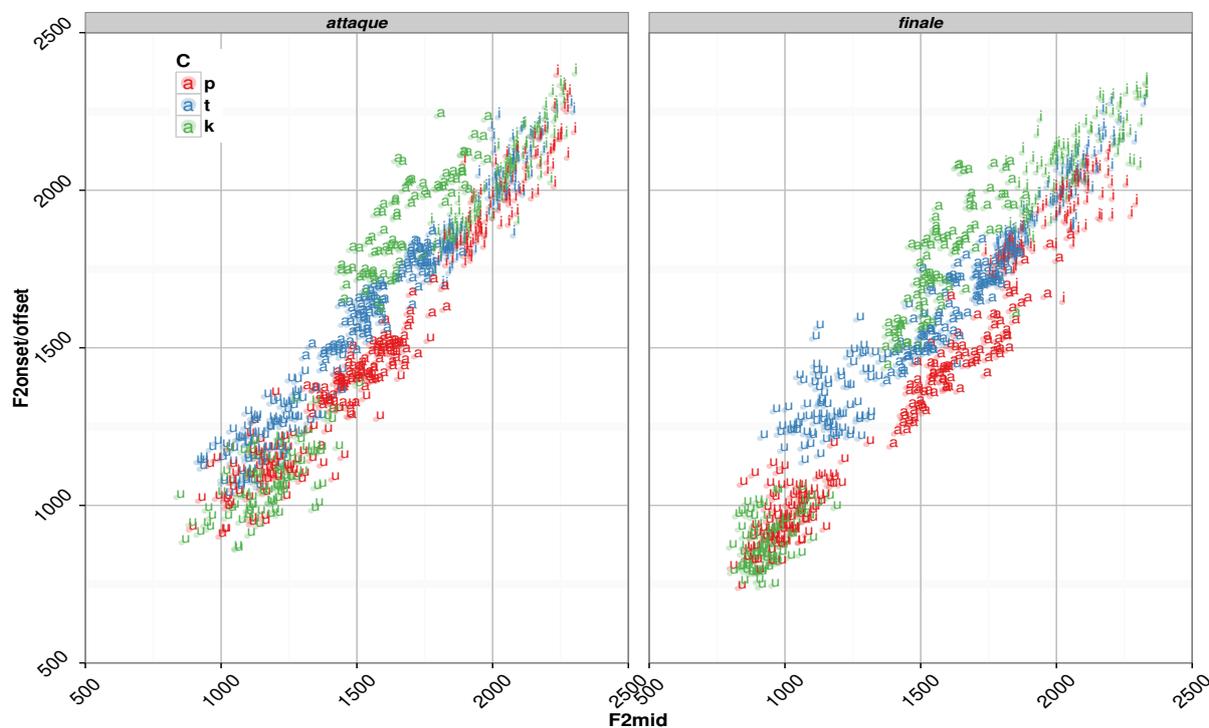


FIGURE 7.4 – Illustrations des domaines  $F2_{onset} - F2_{mid}$  (position initiale) et  $F2_{offset} - F2_{mid}$  (position finale) des occlusives voisées (a) et non voisées (b) prononcées par les 5 locuteurs français dans les trois débits de parole.

les voisées et pour les non voisées. Ces variations de fréquences dépendantes de la position syllabique mais également du voisement soulignent que le type d'information spectrale consonantique contenue dans les voyelles adjacentes est différent. C'est pourquoi certains contextes sont plus sensibles que d'autres à l'ajout de l'information temporelle.

En résumé, les résultats des LDA donnent des arguments en faveur d'un impact positif des informations temporelles extrinsèques dans la classification des occlusives voisées et non voisées. Toutefois, cet impact dépend de la position syllabique et du voisement. De ce fait, la vitesse augmente le taux de classification des occlusives voisées mais pas celui des non voisées, tandis que la durée améliore le pourcentage de classification pour les occlusives non voisées mais pas pour les non voisées. L'effet de la vitesse est plus grand et celui de la durée plus petit pour les occlusives en position initiale que pour celles en position finale.

## 7.2.2 L'impact des catégories temporelles sur la classification

Une série de tests des LDA a été construite dans l'objectif d'évaluer l'impact des catégories temporelles intrinsèque (longueur) et extrinsèque (débit) sur la classification des occlusives. Dans un premier temps, la fonction "*lda*" a été utilisée pour créer les modèles d'apprentissage avec les  $F2_{mid}$ , les  $F2_{onset}$  et / ou les  $F2_{offset}$  comme prédicteurs. Pour les tests de chaque expérience, un échantillon aléatoire a été sélectionné pour la phase d'apprentissage correspondant à un sous-ensemble prélevé dans une moitié des données. Dans un second temps, la fonction "*predict*" a été employée pour prédire le taux de classification correcte sur un sous-ensemble de l'autre moitié restante des données.

### 7.2.2.1 Catégories intrinsèques

Pour chacune des deux premières expériences, le modèle a été entraîné sur les données des voyelles courtes de l'échantillon d'apprentissage pour les tester : d'abord, 1) sur les données des voyelles courtes puis 2) sur celles des voyelles longues de l'échantillon du test. Le modèle a ensuite été entraîné sur les données des voyelles longues afin de les tester : 1) sur les données des voyelles courtes puis 2) sur celles des voyelles longues. Les résultats des LDA de la première expérience mettent en avant que quand l'entraînement est effectué sur le sous-ensemble des voyelles courtes et testé sur celui des voyelles courtes (sous-ensemble intérieur à son domaine d'apprentissage), les taux de classification sont plus élevés que lorsque le modèle est entraîné sur le sous-ensemble des voyelles longues et testé sur celui des voyelles courtes (sous-ensemble extérieur à son domaine d'apprentissage) (Table 7.7). Cette élévation est très prononcée lorsque les occlusives sont réalisées en position initiale et très faible lorsqu'elles sont articulées en position finale. Cette observation signifie que les voyelles courtes contiennent plus d'informations consonantiques que celles longues. C'est pourquoi le niveau d'informations consonantiques contenues dans les voyelles longues (sous-ensemble d'apprentissage des voyelles longues) n'est pas suffisant pour atteindre le taux de classification des voyelles courtes

(sous-ensemble d'apprentissage des voyelles courtes) lors de la classification des occlusives en contexte des voyelles courtes.

Apprentissage	Test	Tous			Femmes			Hommes		
		2 pos	initi	fin	2 pos	initi	fin	2 pos	initi	fin
courte	courte	<b>60%</b>	<b>70%</b>	<b>52%</b>	<b>62%</b>	<b>73%</b>	<b>54%</b>	<b>60%</b>	<b>75%</b>	<b>55%</b>
courte	longue	<b>49%</b>	<b>57%</b>	<b>48%</b>	<b>54%</b>	<b>60%</b>	<b>47%</b>	<b>50%</b>	<b>60%</b>	<b>48%</b>
longue	courte	<b>47%</b>	<b>43%</b>	<b>51%</b>	<b>50%</b>	<b>55%</b>	<b>53%</b>	<b>50%</b>	<b>47%</b>	<b>51%</b>
longue	longue	<b>50%</b>	<b>51%</b>	<b>49%</b>	<b>51%</b>	<b>56%</b>	<b>52%</b>	<b>52%</b>	<b>53%</b>	<b>50%</b>

TABLE 7.7 – Taux de classification correcte des occlusives /t, t<sup>ç</sup>, k/ lorsque le modèle des LDA est entraîné sur les voyelles courtes puis longues et testé sur les occlusives produites avec les voyelles courtes puis longues. Les résultats sont divisés en fonction du groupe de locuteurs (tous les locuteurs, les locuteurs féminins et les locuteurs masculins) et selon la position syllabique (les deux positions, la position initiale et la position finale).

Quand le modèle est entraîné sur le sous-ensemble des voyelles longues puis testé sur le sous-ensemble des voyelles longues (sous-ensemble intérieur à son domaine d'apprentissage), les taux de classification correcte sont légèrement plus bas en position initiale et très peu élevés en position finale que ceux issus du modèle entraîné avec le sous-ensemble des voyelles courtes et testé sur le sous-ensemble des voyelles longues (sous-ensemble extérieur à son domaine d'apprentissage). Ces résultats signalent que les voyelles courtes contiennent un niveau d'informations consonantiques assez élevé permettant au modèle de prédire un sous-ensemble extérieur à son domaine avec un taux de classification similaire à celui intérieur à son domaine d'apprentissage.

Concernant les tests des LDA sur la deuxième expérience, les résultats montrent également un taux de classification plus élevé quand le modèle testé est entraîné sur un sous-ensemble intérieur à son domaine d'apprentissage pour les voyelles courtes (Table 7.8). Pour les voyelles

Apprentissage	Test	Position		
		les deux	initiale	finale
courte	courte	<b>66%</b>	<b>64%</b>	<b>70%</b>
courte	longue	<b>54%</b>	<b>54%</b>	<b>65%</b>
longue	courte	<b>65%</b>	<b>44%</b>	<b>67%</b>
longue	longue	<b>63%</b>	<b>44%</b>	<b>66%</b>

TABLE 7.8 – Taux de classification correcte des occlusives /b, d, g/ lorsque le modèle des LDA est entraîné sur les voyelles courtes puis longues et testé sur les occlusives produites avec les voyelles courtes puis longues. Les résultats sont divisés en fonction de la position syllabique (les deux positions, la position initiale et la position finale).

longues, le taux de classification est plus grand dans le contexte de la position initiale et quasiment identique dans celui de la position finale quand le modèle est testé sur un sous-ensemble extérieur à son domaine d'apprentissage.

En résumé, les résultats présentés ci-dessus mettent en exergue que les catégories temporelles intrinsèques impactent de façon importante la classification des occlusives voisées et non voisées. En effet, les voyelles longues portent une quantité d'informations consonantiques plus basse que celle existante dans les voyelles courtes. Par conséquent, l'entraînement sur les voyelles longues et le test sur les courtes conduisent à des taux de classification inférieurs à ceux provenant de la phase d'apprentissage appliqué sur le sous-ensemble à l'intérieur de son domaine (courte → courte).

### 7.2.2.2 Catégories extrinsèques

Pour la troisième expérience, le modèle a été entraîné sur les données des séquences prononcées dans les débits rapide, moyen puis lent de l'échantillon d'apprentissage pour le tester sur les données des séquences réalisées 1) en débit rapide, 2) en débit moyen puis 3) en débit lent de l'échantillon du test.

Les résultats des LDA mettent en avant que les catégories temporelles extrinsèques ont un impact réduit sur la classification des occlusives voisées et non voisées (Table 7.9). De ce fait, le taux de classification correcte obtenu lorsque le modèle est testé sur un sous-ensemble à l'intérieur de son domaine d'apprentissage (rapide → rapide, moyen → moyen ou lent → lent) est globalement similaire à celui issu du modèle testé sur un sous-ensemble à l'extérieur de son domaine d'apprentissage dans la majorité des cas (lent → rapide, moyen → rapide, lent → moyen, rapide → moyen, rapide → lent ou moyen → lent). Ces constatations laissent supposer

Apprentissage	Test	b d g			p t k		
		2 pos	initiale	finale	2 pos	initiale	finale
rapide	rapide	45%	47%	49%	48%	55%	53%
rapide	moyen	44%	46%	49%	48%	52%	53%
rapide	lent	48%	49%	48%	47%	49%	52%
moyen	rapide	46%	48%	54%	43%	56%	39%
moyen	moyen	45%	44%	54%	49%	55%	47%
moyen	lent	46%	50%	52%	44%	49%	44%
lent	rapide	43%	42%	54%	53%	58%	49%
lent	moyen	42%	46%	60%	54%	58%	52%
lent	lent	43%	49%	55%	50%	52%	48%

TABLE 7.9 – Taux de classification correcte des occlusives voisées et non voisées lorsque le modèle des LDA est entraîné sur les séquences produites dans les débits rapide, moyen puis lent et testé sur celles réalisées dans les débits rapide, moyen et lent. Les résultats sont divisés selon la position syllabique (les deux positions, la position initiale et la position finale).

que la quantité d'informations consonantiques contenues dans les voyelles est généralement semblable peu importe le débit de parole utilisé. Par conséquent, les modifications spectrales sur les voyelles et les consonnes provoquées par les changements des débits de parole ne semblent pas clairement perturber la classification des classes consonantiques.

En résumé, d'après les résultats des LDA sur les données du français, les catégories temporelles extrinsèques influencent de manière faible et aléatoire les taux de classification correcte des occlusives voisées et non voisées dans les deux positions syllabiques.

## 7.3 Conclusion

L'objectif principal de ce chapitre était d'évaluer l'impact des variations temporelles intrinsèque et extrinsèque dans ces deux formes - continue et catégorielle - sur la classification des occlusives. Les résultats révèlent que l'effet des deux types de variations temporelles sur la classification est distinct. En effet, l'ajout d'une information temporelle (durée vocalique ou vélocité de la transition du F2) n'améliore pas le taux de classification de manière importante lorsque l'origine des variations temporelles est intrinsèque aux voyelles (voyelles longues vs. courtes). Au contraire, une augmentation systématique du taux de classification a été remarquée quand l'information de la vélocité (pour les occlusives voisées) ou de la durée (pour les occlusives non voisées) est ajoutée aux informations spectrales ( $F2_{onset}$  et / ou  $F2_{offset} + F2_{mid}$ ). D'après ces résultats, les variations temporelles, dépendant de la longueur vocalique et du débit de parole, qui ont conduit à des modifications semblables de degrés de coarticulation (c'est-à-dire – de temps = + de coarticulation) ne participent pas aux mêmes niveaux dans la catégorisation des classes consonantiques. Autrement dit, il a été remarqué dans les chapitres précédents que l'accélération du débit de parole ainsi que la longueur vocalique courte (– de temps) amènent à une augmentation du degré de coarticulation (+ de pente des équations de locus et – de dispersion). Cette similarité au niveau du degré de coarticulation n'apparaît pas dans celui des catégorisations des classes consonantiques.

Ces effets différents sur la classification malgré une similarité sur le degré de coarticulation peuvent être expliqués par la nature et les conséquences des deux types de variations temporelles. Dans le cas des changements des débits de parole, la durée vocalique ainsi que la durée de l'ensemble de la séquence CVC sont réduites en passant d'un débit lent à un débit rapide conduisant à des chevauchements qui influencent les consonnes et la voyelle inter-consonantique. Autrement dit, le temps nécessaire pour prononcer les trois segments phonétiques dans la séquence CVC est diminué lorsque le débit de parole s'accélère. Concernant l'opposition de la longueur, la variation temporelle affecte principalement la durée vocalique. Par conséquent, l'augmentation du degré de chevauchement est présente dans les deux types de variations mais pour des raisons différentes. Pour le débit de parole, les consonnes et les voyelles s'influencent mutuellement dans les trois débits et la consonne transmet aux voyelles des informations spectrales de son identité globalement similaires dans les trois débits puisque les deux segments (la consonne et la voyelle) subissent des modifications spectrales à cause des variations temporelles. Quant aux variations temporelles liées à l'opposition de la longueur vocalique, la consonne transmet un niveau d'informations spectrales plus élevé aux voyelles courtes qu'aux voyelles longues. Dans le cas de ces dernières, la distance qui sépare la consonne et le milieu de la voyelle est

plus grande et donc l'impact est affaibli. Ces arguments sont soutenus par le taux de classification plus élevé pour les occlusives dans le contexte des voyelles courtes par rapport à celui des mêmes occlusives avec les voyelles longues. C'est pourquoi l'ajout de la dimension temporelle dans les modèles des LDA améliore la classification dans le contexte des variations extrinsèques mais pas pour celles intrinsèques. Pour les premières, cette troisième dimension offre des informations sur le contexte temporel de la réalisation alors que pour les secondes, ces informations sont déjà présentes dans la voyelle qui est peu influencée quand elle est longue et très influencée lorsqu'elle est courte.

L'impact des catégories temporelles sur la classification observé pour les variations intrinsèques sont également liées aux explications présentées ci-dessus. Le test d'un modèle sur un sous-ensemble à l'intérieur de son domaine d'apprentissage révèlent toujours des taux de classification plus élevés pour les voyelles courtes, au contraire de ceux des voyelles longues qui restent globalement similaires. Cette constatation signifie également que les voyelles longues ont un niveau d'informations consonantiques inférieur à celui contenu dans les voyelles courtes et, par conséquent, elles n'atteignent pas leurs taux de classification correcte. Ce phénomène n'est pas remarqué dans le cas des variations temporelles extrinsèques car le test sur un sous-ensemble extérieur ou intérieur à son domaine d'apprentissage ne change pas de manière importante le taux de classification des occlusives. En effet, les voyelles semblent porter une quantité d'informations similaire dans les trois débits. Ce postulat peut être soutenu par les taux de classification correcte dans les contextes des trois débits qui sont globalement similaires lorsque les deux prédicteurs spectraux sont utilisés dans les tests de la première approche des LDA.

# Conclusion générale

L'effet des variations temporelles sur la consonne, sur la voyelle mais également sur la relation spectrale entre les consonnes et les voyelles adjacentes a été décrit et analysé pour mieux appréhender la variabilité des caractéristiques acoustiques des segments phonétiques provoqués par ces phénomènes. Pour ce faire, il a été nécessaire, tout d'abord, de présenter les aspects théoriques contenant les définitions des propriétés acoustiques des segments consonantiques et vocaliques. Toutefois, ces derniers, caractérisés par un ensemble d'indices acoustiques qui leur apporte leur identité et permet leur identification, sont influencés par les différents contextes phonétiques adjacents. En effet, les consonnes et les voyelles connaissent des influences mutuelles quand elles sont prononcées successivement dans la même séquence ou dans des séquences voisines. Ce phénomène entraîne une coarticulation entre les segments successifs qui engendre des changements spectraux sur la qualité d'un ou des deux sons. Par conséquent, une grande partie des variations spectrales des segments phonétiques peut être expliquée par l'analyse de la coarticulation qui sous-entend qu'un son n'est pas réalisé de la même manière lorsqu'il est produit avec d'autres segments. C'est pourquoi un grand nombre de chercheurs se sont intéressés à la coarticulation et ont proposé plusieurs théories et modèles qui donnent la possibilité d'avoir une vision détaillée de son fonctionnement et, plus généralement, de la production de la parole.

Certains de ces modèles exposent que les variations temporelles influencent le degré de coarticulation, impactant les fréquences formantiques de la consonne et / ou de la voyelle ainsi que les transitions formantiques (Lindblom, 1963a; Krull, 1987). Parmi les modèles de coarticulation formulés dans la littérature, les équations de locus offrent un outil de mesure fiable des degrés de chevauchement coarticulatoire entre une consonne et les voyelles adjacentes, calculés à l'aide du début ( $F2_{onset}$ ) et de la fin ( $F2_{mid}$ ) des transitions du deuxième formant, car elles ont été testées dans plusieurs langues et par de nombreux chercheurs (Duez, 1992; Lindblom, 1963b; Sussman et al., 1991, 1993). Elles peuvent être utilisées pour calculer le degré de coarticulation entre les occlusives et les voyelles et pour mesurer leur degré de chevauchement spectral dans le cas des variations temporelles (Krull, 1989; Berry et Weismer, 2013). Le degré de coarticulation est désigné par les paramètres des équations de locus, notamment par leurs pentes. Une pente proche de zéro signifie que l'occlusive est moins coarticulée et plus résistante alors qu'une pente autour de 1 signale un degré de coarticulation élevé et un degré de résistance minimal. Ces variations de pente en fonction du degré de coarticulation de l'occlusive ont ins-

piré Sussman et ses collègues qui ont suggéré que les équations de locus peuvent également être employées comme des indices relationnels pour décrire le lieu d'articulation des occlusives (Sussman, 1989; Sussman et al., 1998b). Ce postulat n'a pas reçu l'unanimité des chercheurs provoquant un débat entre Sussman et ses collègues et de nombreux chercheurs – notamment Fowler (1994); Fowler et Brancazio (2000) (voir également l'article de Sussman et al. (1998b) sur ce débat). Toutefois, l'utilisation des équations de locus comme un outil de mesure du degré de coarticulation performant a été acceptée par la majorité des chercheurs (à l'exception, entre autres, de Löfqvist (1999)).

Les recherches menées sur les équations de locus ont révélé que les vélares sont les occlusives qui ont les pentes les moins élevées suivies par les bilabiales puis les vélares. Les pentes de ces dernières dépendent de l'antériorité *vs.* postériorité des voyelles à cause des deux variantes allophoniques (palatale et vélaire). Lorsque les équations de locus d'une vélaire sont calculées avec les voyelles postérieures, les valeurs des pentes sont supérieures à celles des bilabiales alors que quand elles sont estimées avec les voyelles antérieures, les pentes deviennent inférieures à celles des bilabiales. Ces valeurs de pentes des équations de locus impliquent l'existence d'une influence coarticulatoire plus importante de la voyelle suivante sur les fréquences de  $F2_{onset}$  des bilabiales comparée aux alvéolaires. Autrement dit, ces dernières résistent de façon importante aux effets coarticulatoires de la voyelle suivante. Les vélares, quant à elles, sont désignées par deux pentes distinctes. Les pentes sont basses lorsqu'elles sont produites avec les voyelles antérieures, indiquant que la coarticulation entre la consonne et les voyelles suivantes est minimale. Au contraire, quand elles sont prononcées avec les voyelles postérieures, les pentes sont très élevées (proche de 1) soulignant un impact maximal de ces voyelles sur la consonne. Le calcul des équations de locus avec les deux catégories vocaliques (voyelles antérieures et postérieures) conduit également à des pentes élevées dévoilant que ce lieu d'articulation est très impacté par les voyelles adjacentes.

Plusieurs études ont montré la capacité des paramètres des équations de locus à capter les changements du degré de chevauchement coarticulatoire entraînés par plusieurs phénomènes liés aux variations temporelles, comme le style de parole, l'accent emphatique et le débit de parole (Krull, 1989; Duez, 1992; Lindblom et al., 2007; Berry et Weismer, 2013). De ce fait, la réduction temporelle des segments crée un rapprochement physique entre eux qui amène à une augmentation du chevauchement coarticulatoire. Selon le modèle de *l'undershoot*, les valeurs formantiques de la voyelle et de la consonne sont modifiées en raison du chevauchement coarticulatoire élevé, impliquant la perte d'informations spectrales. Toutefois, Sussman et al. (1998b) minimisent l'effet des variations temporelles sur les paramètres des équations de locus, puisqu'ils séparent les catégories consonantiques, malgré leurs variations.

La présente étude a eu pour objectif d'éclaircir cette ambiguïté sur l'étendue de l'influence des variations temporelles sur les informations spectrales des sons et sur leurs relations coarticulatoires. Par ailleurs, elle a exposé l'efficacité des paramètres des équations de locus à désigner le degré de coarticulation et le lieu d'articulation des segments phonétiques qui su-

bissent différents changements temporels. Ces derniers peuvent être associés à deux types de variations temporelles, soit intrinsèques ou soit extrinsèques à la voyelle. Le premier type correspond à l'opposition phonologique de la longueur – voyelles longues *vs.* courtes – et le second aux changements des débits de parole. L'arabe jordanien et le français ont alors été choisis afin d'examiner l'impact des variations du premier et du second types sur les informations spectrales des occlusives et des voyelles. L'arabe jordanien représente un cas intéressant, car il possède un grand espace consonantique qui permet d'étudier de nombreux lieux d'articulation dont certains ont très peu été exploités dans la littérature. De plus, cette richesse consonantique est accompagnée par une pauvreté de son système vocalique, se caractérisant par l'existence de cinq timbres vocaliques, parmi lesquels trois s'opposent en durée phonologique. Quant au français, il offre l'opportunité de comparer les effets temporels sur les consonnes et les voyelles liés à l'opposition de la longueur vocalique de l'arabe jordanien (variations intrinsèques) à ceux engendrés par les changements des débits de parole (variations extrinsèques).

Trois protocoles expérimentaux ont été conçus de manière à évaluer l'impact temporel des occlusives et des voyelles adjacentes sur les informations spectrales. La première expérience a porté sur l'étude des cinq occlusives non voisées /t, t<sup>ʕ</sup>, k, q, ʔ/ associées aux trois voyelles courtes et aux cinq longues /i, a, u, i:, a:, u:, e:, o:/ de l'arabe jordanien. L'examen de l'espace vocalique a révélé que les voyelles longues et courtes ne se distinguent pas que par leur durée vocalique, mais elles s'opposent également sur le plan spectral. L'écart acoustique entre ces voyelles varie selon le timbre vocalique, qui est faible entre les /a, a:/, moyen entre les /u, u:/ et élevé entre les /i, i:/. En effet, l'opposition spectrale entre les voyelles longues et courtes apparaît dans le domaine F1-F2 par des positions plus périphériques pour les longues par rapport aux courtes. L'observation de cet espace laisse suggérer que les voyelles courtes /i, u/ sont plus proches spectralement des voyelles /e:, o:/ que de celles /i:, u:/ respectivement. Ces distinctions ont ainsi permis de former deux groupes de voyelles à partir des oppositions spectrale et temporelle. Le premier comprend des voyelles longues et courtes qui s'opposent au niveau spectral et temporel (/i, a, u, i:, a:, u:/, premier sous-ensemble), tandis que le second est composé des voyelles longues et courtes qui sont différentes au niveau temporel mais similaires au niveau spectral (/i, a, u, e:, a:, o:/, second sous-ensemble). Cette division a donné la possibilité de comparer l'influence de l'opposition temporelle sur la relation coarticulatoire entre la consonne et la voyelle dans les deux contextes.

Les équations de locus ont alors été employées de façon à estimer les degrés de coarticulation des occlusives dans les contextes des voyelles longues et courtes dans les deux sous-ensembles vocaliques. Parallèlement aux équations de locus, les analyses de dispersion - mesurées à l'aide de l'écart-type - ont également été utilisées pour calculer le degré de la résistance coarticulatoire des voyelles longues et courtes aux différentes influences consonantiques. De même, elles mesurent le degré de la résistance coarticulatoire des occlusives faces aux différents contextes vocaliques (temporels et spectraux). Une dispersion élevée des  $F2_{onset}$  et / ou des  $F2_{mid}$  dans le contexte d'une occlusive produite avec plusieurs voyelles signale une

grande dépendance de la consonne sur les voyelles, alors qu'une dispersion basse signifie que la consonne résiste au maximum aux effets vocaliques. D'un autre côté, une dispersion élevée des  $F2_{onset}$  et / ou des  $F2_{mid}$  pour une voyelle prononcée avec plusieurs consonnes montre que la voyelle ne résiste pas aux influences consonantiques, tandis qu'une dispersion réduite indique une résistance maximale des voyelles contre les impacts des consonnes adjacentes.

Les résultats des analyses de dispersion ainsi que ceux des équations de locus soulignent que le degré de coarticulation des occlusives est influencé de manière comparable dans les deux sous-ensembles. La production des occlusives (dans les deux positions syllabiques) avec les voyelles courtes amène toujours à des pentes plus élevées (+ de coarticulation) que celles des voyelles longues, peu importe la présence ou l'absence d'une opposition spectrale des deux catégories (cas du premier et du second sous-ensembles, respectivement). Autrement dit, ces résultats avancent que l'aspect temporel des voyelles qui s'opposent phonologiquement joue un rôle important sur la relation coarticulatoire entre les consonnes et les voyelles, quelque soit le degré de la différence spectrale entre les voyelles longues et courtes. Les paramètres des équations de locus et les analyses de dispersion ont classé l'alvéolaire pharyngalisée comme l'occlusive la moins coarticulée, suivie par l'alvéolaire non pharyngalisée, l'uvulaire, la glottale puis la vélaire. Ce classement reste identique pour ces occlusives au sein de chaque longueur vocalique (longue et courte) dans les deux sous-ensembles (1 et 2).

Les équations de locus ont à nouveau été générées en remplaçant les  $F2_{onset}$  par des valeurs de fréquences prédites au moment du relâchement de la consonne ( $F2_{@burst}$ ) en utilisant des méthodes d'extrapolation linéaire. Ces nouvelles équations ont pour but d'éviter que les résultats obtenus avec les  $F2_{onset}$  ne soient perturbés par l'effet de l'intervalle temporel relativement long des occlusives non voisées. De fait, ces valeurs prédites sont plus proches du moment de la réalisation de la consonne et, par conséquent, plus représentatives de son lieu d'articulation. Les résultats de ces équations sont en accord avec ceux des équations de locus estimées avec les  $F2_{onset}$ .

La deuxième expérience a été mise en place afin d'examiner l'impact des mêmes phénomènes sur les occlusives voisées et non voisées /b, d, g, t, k/ en arabe jordanien. Cette expérience permet également de tester la fiabilité des méthodes d'extrapolation des  $F2_{@burst}$  en comparant les occlusives voisées /d, g/ avec les occlusives non voisées /t, k/. Les résultats ont dévoilé les mêmes effets au niveau des analyses de dispersion et des équations de locus. Les occlusives produites avec les voyelles courtes sont plus coarticulées que quand elles sont réalisées avec les voyelles longues. De plus, l'ordre du degré de coarticulation est le même au sein de chaque catégorie de longueur vocalique, les alvéolaires étant les occlusives les moins coarticulées suivies par la bilabiale puis les vélaire. Comme attendu, les occlusives non voisées ont des pentes plus élevées que leurs contreparties voisées, mais l'estimation des équations de locus avec les  $F2_{@burst}$  a conduit à des valeurs de pentes similaires. En outre, l'impact de la longueur vocalique reste toujours présent dans ces calculs. Ces résultats offrent des arguments en faveur de l'efficacité des méthodes d'extrapolation utilisées et de l'impact des catégories temporelles

(courte vs. longue) sur la coarticulation.

Concernant la troisième expérience, elle a eu pour objectif d'évaluer l'impact des variations temporelles extrinsèques liées aux changements des débits de parole sur les voyelles, sur les consonnes et sur leurs degrés de coarticulation en français. Les occlusives /b, d, g, p, t, k/ produites avec les voyelles /i, a, u/ dans trois différents débits (lent, moyen et rapide) ont été étudiées pour pouvoir les comparer ensuite avec les données de l'arabe jordanien. Dans un premier temps, les analyses statistiques spectrales sur le début et sur la fin des transitions formantiques ont montré que les fréquences des voyelles sont significativement modifiées par les changements temporels, alors que celles des consonnes semblent résister aux effets de ce facteur. Dans un second temps, les résultats des analyses de dispersion et des équations de locus ont mis en avant que les variations des débits de parole impactent de façon systématique le degré de coarticulation. Les occlusives prononcées dans un débit rapide ont des degrés de coarticulation plus élevés que quand elles sont produites dans un débit plus lent. Par ailleurs, les alvéolaires sont les occlusives les moins coarticulées suivies par les bilabiales puis les vélares à l'intérieur de chaque catégorie du voisement (occlusives voisée et non voisée). Ce classement reste également identique au sein de chaque débit de parole. Il est à noter que les occlusives non voisées ont des valeurs de pentes élevées et proches les unes des autres. Ce phénomène peut être expliqué par l'effet de l'intervalle temporel, le VOT, relativement long – plus long que celui de l'arabe, qui sépare le début des transitions formantiques et le moment du relâchement de l'occlusive. Par conséquent, l'impact du lieu d'articulation est diminué en raison de cette distance temporelle entre les deux événements.

Les méthodes de prédictions des  $F2_{@burst}$  n'ont pas donné de résultats satisfaisants pour les données des occlusives non voisées de cette expérience, car les valeurs de pentes générées avec ces données prédites amènent à des valeurs erronées et inattendues. Ces valeurs peuvent être expliquées principalement par la sensibilité des deux méthodes d'extrapolation aux durées longues des VOT qui caractérisent les occlusives non voisées du français. En effet, l'estimation des valeurs fréquentielles au moment du relâchement des occlusives non voisées représente un problème et un défi pour les phonéticiens. L'une des méthodes proposées dans la littérature afin de répondre à cette problématique est la méthode manuelle décrite dans l'article de Sussman et Shore (1996). Toutefois, elle est subjective et suppose l'existence des résonances de F2 pendant l'intervalle du VOT, or celles-ci sont souvent absentes ou sont cachées par le bruit de friction de l'occlusive. C'est pourquoi, le choix a été fait d'utiliser la prédiction par extrapolation linéaire, exposée dans les deux premières expériences. Elle apporte une solution pour automatiser l'estimation de  $F2_{onset}$  au moment du relâchement dans les grands corpus, qui peut remplacer l'estimation visuelle utilisée auparavant. Néanmoins, les valeurs prédites avec ces méthodes ne sont pas seulement influencées par les longs VOT mais aussi par les changements brusques et rapides qui peuvent être présents dans le tracé formantique. En effet, ces méthodes se basent sur les informations formantiques de deux points de mesure uniquement, déterminés au début et au milieu (première méthode) ou aux deux premiers points (seconde méthode) du tracé forman-

tique du F2. Des méthodes d'extrapolation plus précises (polynomiale par exemple) ou d'autres moyens de prédiction, qui réussiraient à diminuer l'effet des VOT longs et des variations rapides du tracé, pourraient peut-être permettre d'obtenir des  $F2_{@burst}$  plus représentatifs.

Les résultats globaux de la coarticulation dans les trois expériences ont amené des preuves sur la relation entre la temporalité et la coarticulation. Ces expériences dévoilent que le chevauchement coarticulatoire (degré de coarticulation) augmente lorsque la durée vocalique est réduite, peu importe l'origine des variations temporelles (intrinsèque ou extrinsèque). Par ailleurs, le classement du degré de coarticulation (valeurs de pentes) des occlusives révèle toujours des ordres identiques au sein d'une même catégorie temporelle (longueur : longue et courte ou débit : lent, moyen et rapide). Cependant, ce classement connaît des chevauchements entre les valeurs de pentes des différentes occlusives quand elles sont comparées avec toutes les catégories temporelles confondues, notamment en position finale et dans la troisième expérience. Ces observations fournissent des arguments à l'encontre du postulat de Sussman et al. (1998b) cité ci-dessus concernant l'importance des variations temporelles sur les informations obtenues par les équations de locus sur le lieu d'articulation.

Les analyses discriminantes linéaires ont ainsi été utilisées de manière à examiner plus en détail l'étendue de l'impact des variations temporelles sur la classification des occlusives dans les trois expériences. Les résultats ont montré des effets distincts en fonction de l'origine intrinsèque ou extrinsèque des variations temporelles. Dans le premier cas, les deux formes continues des informations temporelles étudiées (durée vocalique en ms ou vitesse de la transition du F2) n'améliorent pas la classification des occlusives lorsqu'elles sont ajoutées aux prédicteurs spectraux traditionnels ( $F2_{onset}$  et  $F2_{mid}$ ) dans les deux premières expériences (variations temporelles intrinsèques). Au contraire, l'ajout de ces informations temporelles augmente le taux de classification correcte dans le cas des occlusives de la troisième expérience (variations temporelles extrinsèques). Ces constatations laissent supposer que, dans le premier cas, la durée ou la vitesse n'apporte pas d'informations supplémentaires sur les classes consonantiques, puisque les informations les plus importantes dans la classification existent déjà dans les prédicteurs spectraux. Autrement dit, les  $F2_{onset}$  et les  $F2_{mid}$  suffisent à eux seuls à fournir le meilleur taux de classification sans avoir recours à l'utilisation de l'information temporelle. Quant au second cas, les deux prédicteurs temporels semblent compenser une partie de la perte d'informations provoquées par les changements des débits de parole. En outre, l'évaluation de l'influence des catégories temporelles sur la classification a mis en avant que les occlusives réalisées avec les voyelles courtes sont mieux classifiées que celles produites avec leurs contreparties longues. Cette observation signifie que les voyelles courtes acceptent de recevoir une quantité d'informations consonantiques plus grande que celle des voyelles longues. Les voyelles articulées dans les trois catégories temporelles extrinsèques, quant à elles, semblent recevoir une quantité d'informations similaire sur le lieu d'articulation des occlusives. En effet, leur taux de classification n'est pas modifié de façon importante quand les sous-ensembles "test-apprentissage" font partie du même domaine ou non, pour les trois débits. Ces constatations peuvent être expliquées par

la perte mutuelle de l'information spectrale des consonnes et des voyelles, lors du changement des débits de parole.

Les ressemblances et les différences des résultats observés, concernant l'impact temporel des deux types dans le domaine de la coarticulation et dans celui de la classification, peuvent être dues à plusieurs raisons. Dans le cas de la coarticulation, une petite distance temporelle entre la consonne et la voyelle (représentée par la distance entre le  $F2_{onset}$  et le  $F2_{mid}$ ) entraîne un degré de chevauchement coarticulatoire élevé, parce que les deux segments sont physiquement proches permettant un maximum d'influences mutuelles entre eux. Au contraire, une grande distance temporelle séparant les deux sons implique un degré de chevauchement coarticulatoire moins important, car les deux événements sont éloignés. Ces arguments justifient alors les résultats similaires des analyses de coarticulation dans les deux types de variations temporelles. Par ailleurs, dans le cas des variations temporelles intrinsèques à la voyelle, les consonnes et les voyelles sont réalisées – idéalement – avec leur durée "naturelle", tandis que, dans le cas des variations extrinsèques, les durées de la réalisation "idéale" des voyelles et des consonnes sont modifiées compte tenu de l'accélération de l'articulation de chacun des deux événements. Le degré de transmission de l'information consonantique dans le premier cas est élevé parce que la distance temporelle est petite, mais également parce que la durée "naturelle" de chacun des deux segments reste généralement non modifiée (d'où une production "idéale" de la consonne et de la voyelle). Au contraire, le degré de transmission de l'information consonantique n'est pas impacté de manière importante par les changements temporels extrinsèques, car la consonne et la voyelle ne sont pas prononcées avec leurs durées "idéales". L'articulation n'est alors pas "idéale" et, par conséquent, une perte d'informations existe pour les deux événements.

Pour conclure, cette recherche a donné des arguments en faveur de l'existence de différents impacts temporels sur la coarticulation CV/VC et sur la classification des occlusives à des degrés différents. Des réponses ont pu être apportées aux questions posées sur ce sujet, grâce à l'analyse de nouvelles données recueillies au cours de cette étude portant sur deux langues distinctes, l'arabe jordanien et le français, qui contiennent deux systèmes vocaliques et consonantiques différents. La confrontation de ces données a enrichi les connaissances et donné des éclaircissements sur la question. Cette recherche s'inscrit dans une problématique plus large qu'il serait intéressant de poursuivre, notamment avec la conception d'expériences de perception qui viseraient à évaluer les résultats développés ici sur les données des LDA et des équations de locus. De même, la poursuite des recherches dans ce domaine pourraient conduire à l'acquisition de nouvelles informations permettant de mieux comprendre l'importance des variations temporelles chez les êtres humains.



# **Les annexes**



# Annexe A

## Les Tables

		F2 <sub>onset</sub>			F2 <sub>@burst1</sub>			F2 <sub>@burst2</sub>		
C	Longueur	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
t <sup>f</sup>	courte	0.13	158	0.04	0.14	201	0.13	0.17	234	0.15
	longue	0.10	102	0.05	0.13	139	0.19	0.16	180	0.09
t	courte	0.11	176	0.02	0.15	221	0.15	0.15	217	0.20
	longue	0.10	176	0.03	0.15	261	0.21	0.16	262	0.11
q	courte	0.06	65	0.03	0.11	113	0.13	0.11	133	0.16
	longue	0.14	169	0.11	0.16	208	0.13	0.15	177	0.13
ʔ	courte	0.06	90	0.04	0.07	119	0.09	0.08	139	0.08
	longue	0.06	104	0.05	0.08	144	0.05	0.06	116	0.08
k	courte	0.07	126	0.03	0.15	260	0.09	0.15	273	0.08
	longue	0.08	182	0.05	0.15	316	0.09	0.13	269	0.07

TABLE A.1 – Écart-types des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus (calculées avec les F2<sub>onset</sub> et les F2<sub>@burst</sub>) des cinq occlusives produites avec les voyelles /i, a, u/ longues vs. courtes (expérience I).

		F2 <sub>onset</sub>			F2 <sub>@burst1</sub>			F2 <sub>@burst2</sub>		
C	Longueur	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
t <sup>f</sup>	courte	0.12	158	0.04	0.14	201	0.13	0.17	234	0.15
	longue	0.05	139	0.23	0.09	171	0.33	0.13	217	0.36
t	courte	0.11	176	0.02	0.15	221	0.15	0.15	217	0.20
	longue	0.09	187	0.04	0.17	345	0.28	0.17	353	0.28
q	courte	0.06	65	0.03	0.11	113	0.13	0.11	133	0.16
	longue	0.18	261	0.23	0.17	234	0.07	0.17	219	0.07
ʔ	courte	0.07	90	0.04	0.07	119	0.09	0.08	139	0.08
	longue	0.09	166	0.01	0.10	202	0.02	0.09	172	0.03
k	courte	0.07	126	0.03	0.15	260	0.09	0.15	273	0.08
	longue	0.14	256	0.03	0.26	488	0.06	0.25	476	0.05

TABLE A.2 – Écart-types des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus calculées avec les voyelles /i, a, u/ vs. /eɪ, aɪ, oɪ/ (expérience I).

		F2 <sub>onset</sub>			F2 <sub>@burst1</sub>			F2 <sub>@burst2</sub>		
longueur		P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>	P	OO	R <sup>2</sup>
<b>b</b>	courte	0.06	74	0.04	0.07	92	0.05	0.07	97	0.05
	longue	0.10	120	0.03	0.11	132	0.05	0.07	88	0.06
<b>d</b>	courte	0.13	200	0.06	0.17	263	0.21	0.18	281	0.20
	longue	0.07	132	0.04	0.08	161	0.08	0.03	93	0.08
<b>g</b>	courte	0.11	144	0.02	0.15	194	0.04	0.15	202	0.02
	longue	0.08	148	0.03	0.09	154	0.03	0.07	141	0.03
<b>t</b>	courte	0.07	114	0.03	0.17	253	0.22	0.13	190	0.18
	longue	0.06	109	0.02	0.08	139	0.17	0.05	100	0.10
<b>k</b>	courte	0.14	217	0.03	0.31	487	0.16	0.25	427	0.13
	longue	0.03	53	0.02	0.04	65	0.05	0.06	84	0.07

TABLE A.3 – Valeurs des écart-types des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus des cinq occlusives de l'arabe jordanien en position initiale (expérience II).

Débit	C	P	OO	R <sup>2</sup>	C	P	OO	R <sup>2</sup>
lent		0.60 (0.17)	614 (235)	0.83 (0.11)		0.94 (0.34)	20 (588)	0.77 (0.09)
moyen	<b>b</b>	0.55 (0.07)	654 (102)	0.90 (0.06)	<b>p</b>	1.03 (0.16)	-212 (285)	0.85 (0.06)
rapide		0.61 (0.11)	560 (156)	0.88 (0.06)		1.01 (0.28)	-221 (547)	0.85 (0.07)
lent		0.32 (0.03)	1255 (95)	0.70 (0.13)		0.70 (0.35)	518 (542)	0.64 (0.29)
moyen	<b>d</b>	0.32 (0.20)	1216 (297)	0.47 (0.30)	<b>t</b>	0.70 (0.28)	542 (391)	0.67 (0.28)
rapide		0.40 (0.19)	1088 (314)	0.53 (0.32)		0.85 (0.22)	315 (363)	0.68 (0.08)
lent		1.19 (0.19)	-257 (338)	0.67 (0.11)		1.36 (0.63)	-492 (1147)	0.56 (0.18)
moyen	<b>g</b>	1.27 (0.24)	-422 (468)	0.75 (0.11)	<b>k</b>	1.46 (0.27)	-686 (563)	0.61 (0.14)
rapide		1.38 (0.21)	-1129 (379)	0.74 (0.17)		1.69 (0.42)	-1130 (833)	0.74 (0.10)

TABLE A.4 – Valeurs des moyennes et des écart-types (entre parenthèses) des pentes (P), des ordonnées à l'origine (OO) et des R<sup>2</sup> des équations de locus estimées avec les F2<sub>@burst2</sub> des occlusives finales produites par les cinq locuteurs dans les trois débits de parole.

## Annexe B

### Les Corpus

TABLE B.1 – Liste des mots CVC écrits en alphabet arabe et en transcription SAMPA utilisée pour l'expérience I

			Position initiale		Position finale		
C	V	CV	Mot	SAMPA	VC	Mot	SAMPA
t	a:	ta:	تَاب	ta:b	a:t	بَات	ba:t
t	a	ta	تَل	tal	at	فَت	fat
t	u:	tu:	تُوب	tu:b	u:t	فُوت	fu:t
t	u	tu	تُف	tuf	ut	كُت	kut
t	i:	ti:	تَيْن	ti:n	i:t	جِيَت	Zi:t
t	i	ti	تِف	tif	it	بِت	bit
t	o:	to:	تُوم	to:m	o:t	بُوت	bo:t
t	e:	te:	تَيْس	te:s	e:t	بَيْت	be:t
k	a:	ka:	كَاز	ka:z	a:k	جَاك	Za:k
k	a	ka	كَز	kaz	ak	فَاك	fak
k	u:	ku:	كُوع	ku:ʔ	u:k	جُوك	Zu:k
k	u	ku	كُم	kum	uk	حُك	X\uk
k	i:	ki:	كَيْس	ki:s	i:k	فِيك	fi:k
k	i	ki	كِت	kit	ik	فِيك	fik
k	o:	ko:	كُول	ko:l	o:k	شُوك	So:k
k	e:	ke:	كَيْف	ke:f	e:k	بِيك	be:k
q	a:	qa:	قَاف	qa:f	a:q	عَاق	ʔ'a:q
q	a	qa	قَد	qad	aq	عَق	ʔ'aq
q	u:	qu:	قُوم	qu:m	u:q	بُوق	bu:q
q	u	qu	قُم	qum	uq	طُق	t'uq

q	i:	qi:	قِيم	qi:m	i:q	سِيَق	si:q
q	i	qi	قِف	qif	iq	ثِق	Tiq
q	o:	qo:	قُوس	qo:s	o:q	ذُوق	Do:q
q	e:	qe:	قِيح	qe:X\	e:q	ضِيَق	d'e:q
?	a:	?a:	أَل	?a:l	a:?	بَاء	ba:?
?	a	?a	أَخ	?ax	a?	لَاء	la?
?	u:	?u:	أُوم	?u:m	u:?	دُوء	du:?
?	u	?u	أُم	?um	u?	نُء	nu?
?	i:	?i:	إِيَم	?i:m	i:?	دِيء	di:?
?	i	?i	إِم	?im	i?	نِء	ni?
?	o:	?o:	أُول	?o:l	o:?	فُوء	fo:?
?	e:	?e:	أَيْل	?e:l	e:?	شِيء	Se:?
t'	a:	t'a:	ظَال	t'a:l	a:t'	جَاط	Za:t'
t'	a	t'a	طَب	t'ab	at'	حَط	X\at'
t'	u:	t'u:	طُوب	t'u:b	u:t'	لُوط	lu:t'
t'	u	t'u	طَخ	t'ux	ut'	حُط	X\ut'
t'	i:	t'i:	طِير	t'i:r	i:t'	شِيَط	Si:t'
t'	i	t'i	طِب	t'ib	it'	غِطَاء	Git'
t'	o:	t'o:	طُوق	t'o:g	o:t'	غُوط	Go:t'
t'	e:	t'e:	طَيْر	t'e:r	e:t'	خِيَط	xe:t'

TABLE B.2 – Listes des consonnes (C), des voyelles (V), des portions CV/VC et CVC transcrites en SAMPA et le mot CVC en alphabet arabe constituant le corpus de la seconde expérience de l'arabe.

Position initiale				Position finale			
C	V	CV	Mot	SAMPA	VC	Mot	SAMPA
b	i:	bi:	بِير	bi:r	i:b	سِيْب	si:b
b	i	bi	بِس	bis	ib	سِب	sib
b	a:	ba:	بَاس	ba:s	a:b	شَاب	Sa:b
b	a	ba	بَس	bas	ab	شَب	Sab
b	u:	bu:	بُوس	bu:s	u:b	ذُوب	Du:b
b	u	bu	بُخ	bux	ub	حُب	X\ub
d	i:	di:	دِير	di:r	i:d	شِيد	Si:d
d	i	di	دِز	diz	id	شِد	Sid
d	a:	da:	دَاس	da:s	a:d	صَاد	s'a:d
d	a	da	دَس	das	ad	صَد	s'ad
d	u:	du:	دُوس	du:s	u:d	جُود	Zu:d
d	u	du	دُش	duS	ud	رُد	rud
g	i:	gi:	قِيس	gi:s	i:g	رِيق	ri:g
g	i	gi	قِش	giS	ig	شِق	Siq
g	a:	ga:	قَاس	ga:s	a:g	سَاق	sa:g
g	a	ga	قَش	gaS	ag	شَق	Sag
g	u:	gu:	قُول	gu:l	u:g	سُوق	su:g
g	u	gu	قُف	guf	ug	رُق	rug
t	i:	ti:	تِين	ti:n	i:t	حِيت	Zi:t
t	i	ti	تِف	tif	it	فِت	fit
t	a:	ta:	تَاه	ta:h	a:t	هَات	ha:t
t	a	ta	تَل	tal	at	زَت	zat
t	u:	tu:	تُوه	tu:h	u:t	شُوت	Su:t
t	u	tu	تُف	tuf	ut	هُت	hut
k	i:	ki:	كِيس	ki:s	i:k	فِيك	fi:k
k	i	ki	كِش	kiS	ik	فِك	fik
k	a:	ka:	كَاز	ka:z	a:k	جَاك	Za:k
k	a	ka	كَز	kaz	ak	فَاك	fak
k	u:	ku:	كُوخ	ku:x	u:k	جُوك	Zu:k
k	u	ku	كُر	kur	uk	حُك	X\uk

TABLE B.3 – Liste des consonnes (C), des voyelles (V), des portions CV/VC transcrites en SAMPA et des mots en alphabet latin pour le corpus de l’expérience du français.

		Position initiale		Position finale	
C	V	CV	Mot	VC	Mot
b	i	bi	bic	ib	rib
b	a	ba	bac	ab	rab
b	u	bu	bouc	ub	loub
d	i	di	dites	id	bide
d	a	da	date	ad	fade
d	u	du	doute	ud	coude
g	i	gi	guide	ig	figue
g	a	ga	gaffe	ag	bague
g	u	gu	goutte	ug	fougue
p	i	pi	pic	ip	type
p	a	pa	patte	ap	cape
p	u	pu	pouce	up	soupe
t	i	ti	type	it	dites
t	a	ta	tache	at	date
t	u	tu	touche	ut	douche
k	i	ki	quitte	ik	chique
k	a	ka	cape	ak	bac
k	u	ku	coude	uk	bouc

# Annexe C

## Les Figures

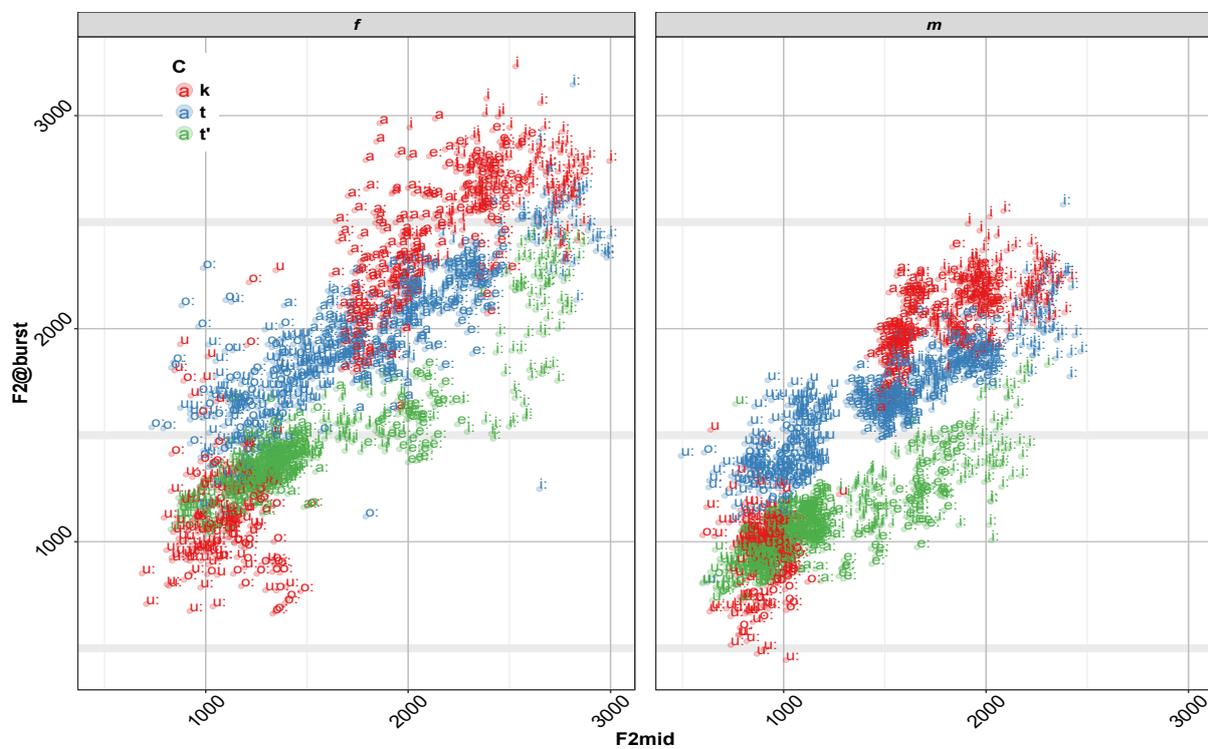


FIGURE C.1 – Illustration des zones du chevauchement spectral dans le domaine  $F2_{onset} - F2_{mid}$  des occlusives /k, t, t<sup>f</sup>/ prononcées par les locuteurs féminins (f) et masculins (m) (Expérience I).

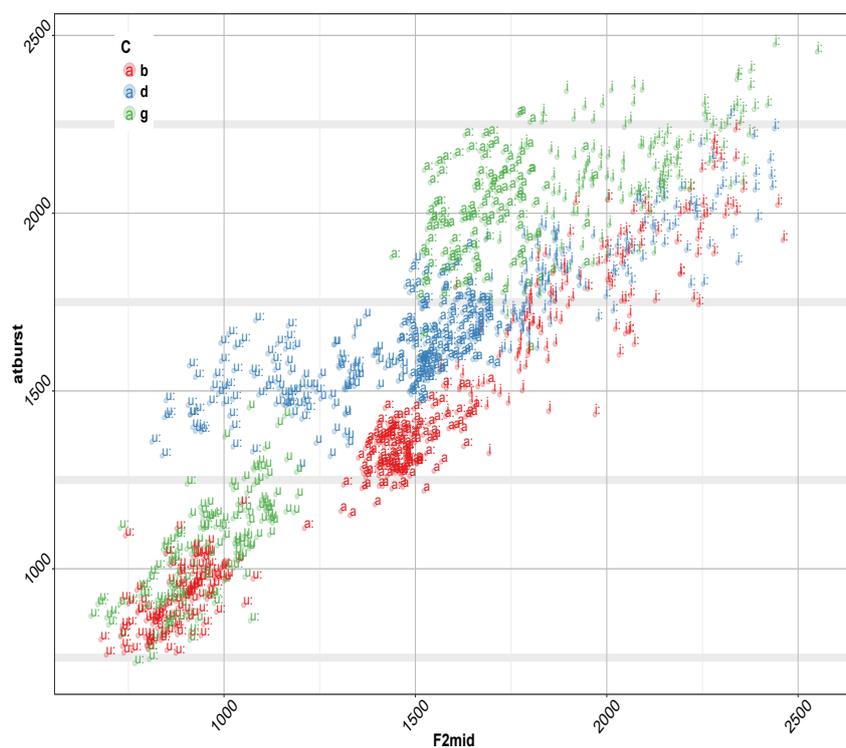


FIGURE C.2 – Illustration des zones du chevauchement spectral dans le domaine  $F2_{@burst}$  -  $F2_{mid}$  des occlusives /b, d, g/ prononcées par les spet locuteurs en position initiale (Expérience II).

# Bibliographie

- M. Abuoudeh and O. Crouzet. A study on locus equations of Jordanian Arabic stop consonants : Do vowel time-based variations impact consonant classification properties ? In *Consecutivity and Simultaneity in Linguistics, Language and Speech (C&S)*. Université de Stransbourg-France, 1-3 July 2015. URL [https://www.researchgate.net/publication/286453845\\_A\\_study\\_on\\_locus\\_equations\\_of\\_Jordanian\\_Arabic\\_stop\\_consonants\\_Do\\_vowel\\_time-based\\_variations\\_impact\\_consonant\\_classification\\_properties](https://www.researchgate.net/publication/286453845_A_study_on_locus_equations_of_Jordanian_Arabic_stop_consonants_Do_vowel_time-based_variations_impact_consonant_classification_properties). (page 249).
- A. Agwuele, H. M. Sussman, B. Lindblom, and A. Miller. Differentiating the effect of speech tempo on CV coarticulation. In *ICPhS XVI*, pages 617–620, 2007. (pages 142 et 233).
- A. Agwuele, H. M. Sussman, and B. Lindblom. The effect of speaking rate on consonant vowel coarticulation. *Phonitica*, 65(4) :194–209, 2008. (pages 142, 174, 233, 233, 234, 234, 235, 235, 235 et 238).
- W. A. Ainsworth. Duration as a cue in the recognition of synthetic vowels. *The Journal of Acoustical Society of America*, 51 :648–651, 1972. (pages 22 et 23).
- S. H. Al-Ani. *Arabic phonology : an acoustical and physiological investigation. (Janua Linguarum. Series Practica, 61.) 104 pp., 1970. Guilders 35.* The Hague, Paris : Mouton, 1970. (pages 19, 20, 224, 225, 228, 229, 229 et 229).
- A. Al-Bamerni and A. Bladon. One-stage and two-stage temporal patterns of velar coarticulation. *The Journal of the Acoustical Society of America*, 72(S1) :S104–S104, 1982. (pages 89 et 90).
- F. Al-Tamimi. *Phonetic and phonological variation in speech of rural migrants in a Jordanian city.* Phd thesis, The University of Leeds, 2001. (pages 224, 228, 228, 228 et 259).
- F. Al-Tamimi and B. Heselwood. Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic. In Z. M. Hassan and B. Heselwood, editors, *Instrumental Studies in Arabic Phonetics*, page 163–192. John Benjamins Publishing Company, 2011. (pages 225, 225, 226 et 226).

- F. Al-Tamimi, F. Alzoubi, and R. Tarawnah. A videofluoroscopic study of the emphatic consonants in Jordanian Arabic. *Folia Phoniatrica et Logopaedica*, 61 :247–253, 2009. (pages 225, 225, 225 et 226).
- J.-E. Al-Tamimi. L'équation de locus comme mesure de la coarticulation vc et cv : étude préliminaire en arabe dialectal jordanien. In *JEP*, 2004. (pages 224 et 227).
- J.-E. Al-Tamimi. *Indices dynamiques et perception des voyelles : Étude translinguistique en arabe dialectal et en français*. Thèse de doctorat, Université Louis Lumière - Lyon 2, 2007. (pages 120, 122, 221, 222, 222, 222, 223, 223, 223, 224, 233, 237, 252 et 254).
- E. Al-Wer and B. Herin. The lifecycle of Qaf in Jordan. *Langage et Société*, 132 :59–76, 2011. (pages 228 et 259).
- O. Allatif. Adaptabilité des paramètres temporels et spectraux dans l'opposition de quantité vocalique de l'arabe de mayadin (syrie). In *25e Journées d'Étude sur la Parole*. Fès, Maroc, 2004. (page 29).
- O. Allatif. Le poids perceptif de la durée de la voyelle face à sa qualité acoustique dans la discrimination des voyelles longues et brèves. In *7ème Rencontres des Jeunes Chercheurs en Parole 2007 (RJCP 2007)*. Paris, France, 2007. (page 29).
- O. Allatif. *Contrôle des corrélats temporels et spectraux de la quantité vocalique : de l'arabe syrien de l'Euphrate au français de Savoie*. PhD thesis, Université Stendhal - Grenoble III, 2008. (pages 29, 29, 30, 30, 31, 31, 32, 223, 231, 237 et 254).
- A. Amelot. Dispositif d'imagerie pour l'observation de l'activité vélo-pharyngée. In A. Marchal and C. Cavé, editors, *L'imagerie médicale pour l'étude de la parole*, pages 65–85. Lavoisier, 2009. (pages 105 et 116).
- P. F. Assman, T. M. Nearey, and J. T. Hogan. Vowel identification : Orthographic perceptual, and acoustic aspects. *The Journal of Acoustical Society of America*, 71 :975–989, 1982. (pages 37 et 38).
- M. Aylett and A. Turk. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119 (5) :3048–3058, 2006. (page 121).
- H. R. Baayen. *Analyzing Linguistic Data : A practical introduction to statistics*. Cambridge University Press, 2008. (page 330).
- M. Bahloul. Variation and attitudes : A sociolinguistic analysis of the Qaaf. In M. Embarki and M. Ennaji, editors, *Modern trends in Arabic dialectology*, chapter 4, pages 69–94. Red Sea Press, 2011. (page 228).

- P. Bailly and C. Carrère. *Statistique descriptive : L'économie et les chiffres*. PUG, 2015. (page 146).
- J. Bakran and V. Mildner. Effect of speech rate and coarticulation strategies on the locus equation determination. In *ICPhS, Stockholm*, volume 1, pages 26–29, 1995. (pages 187 et 188).
- M. Barkat. *Détermination d'indices acoustiques robustes pour l'identification automatique des parlars arabes*. Thèse de doctorat, Université Lumière - Lyon 2, 2000. (pages 221, 222 et 224).
- F. Bell-Berti. A spatial-temporal model of velopharyngeal function. In N. J. Lass, editor, *Speech and Language : Advances in Basic Research Practice*, volume IV, pages 291–313. Academic, New York, 1980. (pages 84 et 88).
- F. Bell-Berti and K. Harris. A temporal model of speech production. *Phonetica*, 65(5) :1268–1270, 1981. (pages 79, 82, 85, 85 et 88).
- F. Bell-Berti and K. S. Harris. Anticipatory coarticulation : Some implications from a study of lip rounding. *The Journal of Acoustical Society of America*, 90(1) :112–123, 1979. (page 82).
- F. Bell-Berti and K. S. Harris. Temporal patterns of coarticulation : Lip rounding. *The Journal of Acoustical Society of America*, 71(2) :449–454, 1982. (page 82).
- F. Bell-Berti and R. Krakow. Anticipatory velar lowering : A coproduction account. *The Journal of Acoustical Society of America*, 90(1) :112–123, 1991. (pages 79, 81, 83, 85, 88, 90 et 90).
- A. P. Benguerel and H. Cowan. Coarticulation of upper lip protrusion in French. *Phonetica*, 30 (1) :41–55, 1974. (page 88).
- A. P. Benguerel, H. Hirose, M. Sawashima, and T. Ushijima. Velar coarticulation in French : A fiberoptic study. *Journal of Phonetics*, 5 :149–158, 1977. (page 88).
- J. R. Benki. Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, 29 :1–22, 2001. (page 14).
- D. C. Bennett. Spectral Form and Duration as Cues in the Recognition of English and German Vowels. *Language and Speech*, 11(2) :65–85, 1968. (pages 22, 23 et 23).
- J. Berry and G. Weismer. Speaking rate effects on locus equation slope. *Journal of Phonetics*, 41 :468–478, 2013. (pages 142, 233, 235, 235, 236, 237, 238, 347 et 348).
- M. Bin-Muqbil. *Phonetic and phonological aspects of Arabic emphatics and gutturals*. Phd, University of Wisconsin-Madison, 2006. (page 229).

- J. W. Black. Natural frequency, duration, and intensity of vowels in reading. *Journal of Speech and Hearing Disorders*, 14(3) :216–221, 1949. (pages 20 et 21).
- R. A. W. Bladon and A. Al-Bamerni. Coarticulation resistance in English /l/. *Journal of Phonetics*, 4 :135–150, 1976. (pages 82, 82, 108, 108 et 109).
- S. E. Blumstein and K. N. Stevens. Acoustic invariance in speech production : Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of Acoustical Society of America*, 66(4) :1001–1017, 1979. (pages 47, 49, 50 et 50).
- S. E. Blumstein and K. N. Stevens. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of Acoustical Society of America*, 67(2) :648–662, 1980. (pages 47, 50, 51, 51, 52, 67, 70 et 71).
- P. Boersma and D. Weenink. Praat : doing phonetics by computer [computer program], 2017. URL <http://www.praat.org/>. (pages 212 et 243).
- G. J. Borden, K. S. Harris, and L. J. Raphael. *Speech Science Primer : Physiology Acoustics, and Perception of Speech*. Williams & Wilkins, third edition, 1994. (pages 12, 12, 17 et 109).
- O. Bouferroum and M. Boudraa. Cv coarticulation, locus and locus equation perspective on the invariance issue involving Algerian Arabic consonants. *Journal of Phonetics*, 50 :120–135, 2015. (page 229).
- J. Bourque, J.-G. Blais, and F. Larose. L'interprétation des tests d'hypothèses : p, la taille de l'effet et la puissance. *Revue des sciences de l'éducation*, 35(1) :211–226, 2009. (page 256).
- S. Boyce, R. Krakow, F. Bell-Berti, and C. E. Gelfer. Converging sources of evidence for dissection articulatory movements into core gestures. *Haskins Laboratories Status Report on Speech Research*, SR-103/104 :39–50, 1990. (page 90).
- L. Brancazio and C. A. Fowler. On the relevance of locus equations for production and perception of stop consonants. *Perception & Psychophysics*, 60(1) :24–50, 1998. (pages 124, 124, 179, 187, 188, 188, 188, 189, 189, 190 et 215).
- C. P. Browman and L. Goldstein. Dynamic modeling of speech structure. *Haskins Laboratories Status Report on Speech Research*, SR-79/80 :1–17, 1984. (page 131).
- C. P. Browman and L. Goldstein. Towards an articulatory phonology. *Phonology Yearbook*, 3 : 219–252, 1986. (pages 80, 85 et 125).
- C. P. Browman and L. Goldstein. Articulatory gestures as phonological units. *Phonology*, 6 : 201–251, 1989. (pages 85, 87, 124 et 125).

- C. P. Browman and L. Goldstein. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18 :299–320, 1990a. (page 125).
- C. P. Browman and L. Goldstein. Representation and reality : Physical systems and phonological structure. *Haskins Laboratories Status Report on Speech Research*, SR-105/106 :83–92, 1990b. (page 86).
- C. P. Browman and L. Goldstein. Dynamic and articulatory phonology. *Haskins Laboratories Status Report on Speech Research*, SR-113 :51–62, 1993. (page 86 et 86).
- R. Carré and S. Chennoukh. Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures. *Journal of Phonetics*, 23(1-2), 1995. (pages 106, 107, 107, 107 et 107).
- E. Celdran and X. Villalba. Locus equations as a metrics for place of articulation in automatic speech recognition. In *ICPhS, Stockholm*, volume 1, pages 30–33, 1995. (pages 156, 157, 157 et 182).
- M. Chafcouloff. Voir la parole. In *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, volume 23, pages 23–65. Laboratoire Parole et Langage, 2004. (page 76).
- S. Chennoukh, R. Carré, and B. Lindblom. Locus equations in the light of articulatory modeling. *The Journal of the Acoustical Society of America*, 71(2) :437–448, 1997. (pages 167, 167, 168, 169, 169, 169, 187, 193 et 195).
- T. Cho and P. Ladefoged. Variation and universals in VOT : Evidence from 18 languages. *Journal of Phonetics*, 27 :207–229, 1999. (pages 59, 60 et 61).
- J. Cohen. *Statistical power analysis for the behavioral sciences*. LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS, 1988. (page 256).
- F. S. Cooper, A. M. Liberman, and J. M. Borst. The interconversion of audible and visible patterns as basis for research in the perception of speech. *Poceedings of the National Academy of Sciences*, 37(5) :318–325, 1951. (pages 9 et 55).
- F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst, and L. J. Gerstman. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 24(6) :597–606, 1952. (pages 9, 13, 14, 54, 55, 56, 94 et 95).
- P.-A. Cornillon, F. Husson, N. Jégou, and E. Matzner-Lober. *Statistiques avec R*. PU Rennes, 2012. (pages 329 et 330).
- C. S. Crowther. Modeling coarticulation and place of articulation using locus equations. *UCLA Working Papers in Phonetics*, 88 :127–148, 1994. (pages 184, 185, 186, 187, 187 et 188).

- D. Crystal. *A dictionary of Linguistics and Phonetics*. Blackwell Publishing Ltd, vi edition, 2008. (page 83).
- T. H. Crystal and A. S. House. The duration of American-English vowels : An overview. *Journal of Phonetics*, 16 :263–284, 1988. (page 21).
- R. G. Daniloff and R. E. Hammarberg. On defining coarticulation. *Journal of Phonetics*, 1 : 239–248, 1973. (pages 78, 80, 81 et 81).
- R. G. Daniloff and K. L. Moll. Coarticulation of lip rounding. *Journal of Speech and Hearing Research*, 11 :707–721, 1968. (page 81 et 81).
- P. Delattre. La forece d’articulation consonantique en français. *The French Review*, 14(3) : 220–232, 1941. (page 61).
- P. Delattre. The physiological interpretation of sound spectrograms. *Publications of the Modern Language Association of America (PMLA)*, 66 :864–875, 1951. (pages 19, 19 et 201).
- P. Delattre. Le jeu des transitions de formants et la perception des consonnes. In *IVth International Congress of Phonetic Sciences, Helsinki*, volume 4, pages 407–418, 1961. (pages 12, 13, 13, 14, 14, 14 et 15).
- P. Delattre. Coarticulation and the locus theory. *Studia Linguistica*, 23(1) :1–26, 1969. (pages 95, 96, 97, 98 et 105).
- P. Delattre, A. M. Liberman, F. S. Cooper, and L. J. Gerstman. An experimental study of the acoustic determinants of vowel color ; observations on one- and two-formant vowels synthesized from spectrographic patterns. *WORD*, 8(3) :195–210, 1952. doi : 10.1080/00437956.1952.11659431. (page 19).
- P. Delattre, A. Liberman, and F. Cooper. Acoustical loci and transitional cues for consonants. *The Journal of the Acoustic Society of America*, 27(4) :769–773, July 1955. (pages 9, 14, 14, 14, 56, 56, 64, 66, 93, 94, 95, 95, 96, 105, 105, 122, 133, 143, 143, 205, 224, 225, 228, 232 et 232).
- Y. Dodge. *Statistique : Dictionnaire encyclopédique*. Springer, second edition, 1988. (page 247).
- M. F. Dorman, M. Studdert-Kennedy, and L. J. Raphael. Stop-consonant recognition : Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 22(2) :109–122, 1977. (pages 64, 64, 65 et 71).
- D. Duez. Second formant locus-nucleus patterns in spontaneous speech : Some preliminary results. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 10 :109–114, December 1989. (page 141).

- D. Duez. Some evidence on second formant locus-nucleus patterns in spontaneous speech in French. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 12 :109–126, May 1991. (page 141).
- D. Duez. Second formant locus-nucleus patterns : An investigation of spontaneous French speech. *Speech Communication*, 11(4-5) :471–427, October 1992. (pages 91, 91, 122, 141, 142, 142, 142, 142, 184, 232, 232, 233, 347 et 348).
- J. Durand. La phonétique classique : l'Association Phonétique Internationale et son alphabet. In *Phonologie et phonétique : forme et substance*, page 35. Hermès Lavoisier, 2005. (page 16).
- A. Eek and E. Meister. The perception of stop consonants : Locus equations and spectral integration. In *ICPhS, Stockholm*, volume 1, pages 18–21, 1995. (pages 175, 176, 176, 176, 178 et 179).
- M. Embarki. Les dialectes arabes modernes : état et nouvelles perspectives pour la classification géo-sociologique. *Arabica*, 55 :583–604, 2008. (page 221 et 221).
- M. Embarki. Acoustic cues for the classification of arabic dialects. In M. Embarki and M. En-naji, editors, *Modern trends in Arabic dialectology*, chapter 3, pages 47–68. Red Sea Press, 2011. (page 228).
- M. Embarki, C. Guilleminot, and M. Yeou. Equation de locus comme indice de distinction consonantique pharyngalisé vs non-pharyngalisé en arabe. In *XXV JEP*, 2006. (page 227).
- M. Embarki, M. Yeou, C. Guilleminot, and S. Al Masri. An acoustic study of coarticulation in Modern Standard Arabic and dialectal Arabic : Pharyngealized vs. non pharyngealized articulation. In *ICPhS XVI*, 2007. (page 227).
- M. Embarki, S. Ouni, M. Yeou, C. Guilleminot, and S. Al-Maqtari. Acoustic and electromagnetic articulographic study of pharyngealisation : Coarticulatory effects as an index of stylistic and regional variation in arabic. In Z. M. Hassan and B. Heselwood, editors, *Instrumental Studies in Arabic Phonetics*, pages 194–215. John Benjamins Publishing Company, 2011. (page 227).
- O. Engstrand and B. Lindblom. The locus line : does aspiration affect its steepness. In *Reports from the Department of Linguistics : Umea University (PHONUM)*, volume 4, pages 101–104, 1997. (page 208).
- C. Essner. Recherche sur la structure des voyelles orales. *Archives Néerlandaises de Phonétique Expérimentale*, 20 :40–77, 1947. (page 17).
- G. Fant. The acoustics of speech. In *The third International Congress on Acoustics, Stuttgart*, 1959. (page 12).

- G. Fant. *Acoustic Theory of Speech Production : With Calculations based on X-Ray Studies of Russian Articulations*. Description and Analysis of Contemporary Standard Russian. Mouton, 1960. (pages 9, 9, 13 et 16).
- G. Fant. *Speech sounds and features*. Cambridge, MA : MIT Press, 1973. (pages 42 et 208).
- E. Farnetani and D. Recasens. Coarticulation models in recent speech production theories. In W. J. Hardcastle and N. Hewlett, editors, *Coarticulation, Theory, Data and Techniques*, Cambridge Studies in Speech Science and Communication, pages 31–65. Cambridge University Press, 1st edition, February 1999. (pages 79, 80, 83, 85, 88 et 92).
- E. Farnetani and D. Recasens. Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, and F. E. Gibbon, editors, *The Handbook of Phonetic Sciences*, pages 316–352. Wiley-Blackwell, second edition, 2010. (pages 80, 80, 89 et 90).
- D. C. Fitzpatrick, J. S. Kanwal, J. A. Butman, and N. Suga. Combination-sensitive neurons in the primary auditory cortex of the moustached bat. *Journal of Neuroscience*, 13 :931–940, 1993. (page 183).
- E. Flemming. On line course : Linguistic Phonetics. *Massachusetts Institute of Technology : MIT OpenCourseWare*, License : Creative Commons BY-NC-SA, 2005. URL <http://mit.sustech.edu/OcwWeb/Linguistics-and-Philosophy/24-963Fall-2005/CourseHome/index.htm>. (page 10).
- C. Fougeron. *Variations articulatoires en début de constituants prosodiques de différents niveaux en français*. Thèse de doctorat, Université Paris III - Sorbonne Nouvelle, 1998. (pages 63 et 270).
- C. Fougeron. Phonologie articulatoire. In N. Nguyen, S. Wauquier-Gravelines, and J. Durand, editors, *Phonologie et phonétique : forme et substance*, pages 265–290. Hermès Lavoisier, 2005. (page 125).
- C. Fowler. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8 :113–133, 1980. (pages 79, 80, 82, 85, 85 et 85).
- C. Fowler. Invariants, specifiers, cues : An investigation of locus equations as information for place of articulation. *Perception and Psychophysics*, 55 :597–610, 1994. (pages 91, 124, 124, 135, 150, 157, 157, 157, 158, 158, 158, 159, 159, 159, 161, 161, 162, 164, 164, 164, 164, 169, 169, 189, 190, 215, 224, 225, 228, 228, 232 et 348).
- C. Fowler. *Speech production*, chapter 29, pages 489–501. Oxford : Oxford University Press, 2007. (pages 80 et 85).
- C. Fowler and Saltzman. Coordination and coarticulation in speech production. *Language and speech*, 5 :171–195, 1993. (pages 86, 86, 87, 87, 88, 124, 126, 130 et 130).

- C. A. Fowler and L. Brancazio. Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, 43(1) : 1–41, 2000. (pages 199, 199, 200, 200, 201, 215 et 348).
- D. Fruchter and H. M. Sussman. The perceptual relevance of locus equations. *The Journal of the Acoustical Society of America*, 102(5) :2997–3008, 1997. (pages 174, 174, 175, 177, 177, 178, 178 et 179).
- T. Gay. Articulatory movements in vcv sequences. *The Journal of the Acoustical Society of America*, 62(1) :183–193, 1977. (pages 106, 106, 106 et 107).
- T. Gay. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 63 :223–30, January 1978a. (pages 20, 91, 101, 101, 101, 102, 102, 132, 231, 231 et 305).
- T. Gay. Articulatory units : Segments or syllables? *Haskins Laboratories Status Report on Speech Research*, SR-54 :53–63, 1978b. (page 74).
- S. Ghazali. *Back consonants and backing coarticulation in Arabic*. Phd thesis, University of Texas, 1977. (page 228).
- K. Hadding-Koch and A. S. Abramson. Duration versus spectrum in swedish vowels : Some perceptual experiments. *Studia Linguistica*, 18(2) :94–107, December 1964. (pages 27, 28 et 231).
- M. Halle, G. W. Hughes, and J. A. Radley. Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, 29(1) :107–116, 1957. (pages 41, 44, 45, 57, 57 et 58).
- R. Hammarberg. The metaphysics of coarticulation. *Journal of Phonetics*, 4 :353–363, 1976. (pages 80 et 81).
- J. Harrington. Acoustic phonetics. In W. J. Hardcastle, J. Laver, and F. E. Gibbon, editors, *The Handbook of Phonetic Sciences*, pages 81–129. Wiley-Blackwell, second edition, 2010. (pages 12, 12, 12, 12, 18, 19, 20, 42, 42, 119 et 232).
- K. S. Harris. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1(1) :1–7, 1958. (page 14).
- K. S. Harris. Vowel duration change and its underlying physiological mechanisms. *Language and Speech*, 21(4), 1978. (page 102).
- K. S. Harris, G. F. Lysaught, and M. M. Schvey. Some aspects of the production of oral and nasal labial stops. *Language and Speech*, 8(3) :135–147, 1965. (page 62).

- S. Hawkins. An introduction to task dynamics. In G. J. Docherty and D. R. Ladd, editors, *Papers in Laboratory Phonology : Gesture, Segment, Prosody*, volume II, pages 9–25. Cambridge : Cambridge University Press, 1992. (pages 125 et 129).
- W. Henke. *Dynamic articulatory model of speech production using computer simulation*. PhD thesis, Massachusetts Institute of Technology, 1966. (pages 79 et 82).
- B. Heselwood and F. Al-Tamimi. A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectrography. In Z. M. Hassan and B. Heselwood, editors, *Instrumental Studies in Arabic Phonetics*, pages 101–127. John Benjamins Publishing Company, 2011. (page 229 et 229).
- T. Hill and P. Lewicki. *Statistics : Methods and Applications*. StatSoft, Inc., 2005. (page 117).
- J. M. Hillenbrand. Static and dynamic approaches to vowel perception. In G. S. Marrison and P. F. Assmann, editors, *Vowel Inherent Spectral Change*, pages 87–115. Springer, 2013. (pages 33, 39, 39, 39 et 40).
- J. M. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97 :3099–3111, 1995. (pages 20, 21, 21, 25, 39, 39 et 39).
- J. M. Hillenbrand, M. J. Clark, and R. A. Houde. Some effects of duration on vowel recognition. *The Journal of Acoustical Society of America*, 108(6) :3013–3022, 2000. (pages 20, 21, 24, 25, 26, 26, 27, 27 et 314).
- J. M. Hillenbrand, M. J. Clark, and T. M. Nearey. Effects of consonant environment on vowel formant patterns. *The Journal of Acoustical Society of America*, 109(2) :748–763, 2001. (page 22, 22 et 22).
- G. E. Hinton and K. J. Lang. The development of the time-delay neural network architecture for speech recognition. *Technical Report Carnegie Mellon University*, CMU-CS :88–152, 1988. (page 183).
- H. Hirose. Investigating the physiologie of laryngeal structures. In W. J. Hardcastle, J. Laver, and F. E. Gibbon, editors, *The Handbook of Phonetic Sciences*, pages 130–152. Wiley-Blackwell, second edition, 2010. (page 102).
- C. F. Hockett. *A manual of phonology*. Baltimore, : Waverly Press, 1955. (page 74).
- C. Huang. The effect of formant trajectory and spectral shape on the tense/lax distinction in american vowels. In *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 893–896, Apr 1986. (pages 22 et 23).

- M. Huckvale. Experimental phonetics, 2016. URL <http://www.phon.ucl.ac.uk/courses/spsci/expphon/week7.php>. (pages 84 et 86).
- K. Iskarous, C. a. Fowler, and D. H. Whalen. Locus equations are an acoustic expression of articulator synergy. *The Journal of the Acoustical Society of America*, 128(4) :2021–32, oct 2010. ISSN 1520-8524. (pages 122, 123, 124, 201, 202, 203, 204, 204, 204, 215 et 215).
- R. Jakobson and M. Halle. Tenseness and laxness. In R. Jakobson, editor, *Selected Writings*, volume I of *Phonological Studies*, pages 550–555. De Gruyter Mouton, 1962. (page 61).
- R. Jakobson, G. Fant, and M. Halle. *Preliminaries to speech analysis*. Cambridge, Mass. : MIT, 1963. (page 7).
- J. J. Jenkins, W. Strange, and T. R. Edman. Identification of vowels in "vowelless" syllables. *Perception & Psychophysics*, 34 :441–450, 1983. (pages 34, 34, 35, 36 et 36).
- K. Johnson. *Acoustic and Auditory Phonetics*. Blackwell Publishing, second edition, 2003. (pages 7 et 17).
- A. Jongman, W. Herd, M. Al-Masri, J. Sereno, and S. Combest. Acoustics and perception of emphasis in Urban Jordanian Arabic. *Journal of Phonetics*, 39 :85–95, 2011. (pages 226, 226 et 227).
- M. Joos. Acoustic phonetics. *Language*, 24(2) :5–136, Jun. 1948. (pages 17, 77, 77, 78, 78, 79 et 201).
- P. Keating. The window model of coarticulation : articulatory evidence. In Beckman and Kingston, editors, *Papers in Laboratory Phonology*, volume I, pages 451–470. Cambridge University Press, 1990. (pages 82, 83, 83, 83, 84, 84 et 86).
- P. Keating, R. Wright, and J. Zhang. Word-level asymmetries in consonant articulation. In *UCLA Working Papers in Phonetics*, 2001. (pages 63, 63 et 270).
- J. A. S. Kelso, B. Tuller, E. Vatikiotis-Bateston, and C. A. Fowler. Functionally specific articulatory cooperation following jaw perturbations during speech : Evidence for coordinative structures. *Journal of Experimental Psychology : Human Perception and Performance*, 10 (6) :812–832, 1984. (page 127).
- J. A. S. Kelso, E. L. Saltzman, and B. Tuller. The dynamical perspective on speech production : data and theory. *Journal of Phonetics*, 14 :29–59, 1986. (page 127).
- R. D. Kent and F. D. Minifie. Coarticulation in recent speech production models. *Journal of Phonetics*, 5 :115–133, 1977. (pages 75, 79 et 110).

- R. D. Kent and C. Read. *Acoustic analysis of speech*. Singular Publishing Group, Inc, 1996. (pages 12, 13, 14, 16, 18, 41, 42, 43, 48 et 57).
- D. Kewley-Port. Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America*, 72(2) :379–389, aug 1982. (pages 66, 66, 66, 67, 67, 69, 69, 71, 95, 143, 232 et 232).
- D. Kewley-Port. Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73(1) :322–335, Jan 1983. (pages 67, 67, 68, 68, 69, 69, 70 et 71).
- D. Kewley-Port, D. B. Pisoni, and M. Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 73(5) :1779–1793, May 1983. (page 69).
- D. H. Klatt. Linguistic uses of segmental duration in English : Acoustic and perceptual evidence. *The Journal of Acoustical Society of America*, 59(5) :1208–1221, 1976. (page 20).
- D. H. Klatt. Synthesis by rule of consonant-vowel syllables. *MIT Speech Communication Group Working Papers*, 3 :93–104, 1979. (pages 70, 135, 135, 136, 136, 146 et 146).
- D. H. Klatt. Review of text-to-speech conversion for english. *The Journal of the Acoustical Society of America*, 82(3) :737–793, 1987. (pages 136 et 146).
- D. H. Klatt and L. C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of Acoustical Society of America*, 87 :820–857, 1990. (page 25).
- W. Koenig, H. K. Dunn, and L. Y. Lacy. The sound spectrograph. *The Journal of the Acoustical Society of America*, 18(1) :19–49, 1946. doi : <http://dx.doi.org/10.1121/1.1916342>. URL <http://scitation.aip.org/content/asa/journal/jasa/18/1/10.1121/1.1916342>. (page 8 et 8).
- V. A. Kozhevnikov and Chistovich. *Speech : Articulation and Perception*. Washington, U.S. Joint Publications Research Service, 1965. (pages 79, 82, 82, 82, 108 et 108).
- D. Krull. Second formant locus patterns as a measure of consonant-vowel coarticulation. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 5 :57–75, 1987. (pages 91, 91, 91, 91, 135, 139, 139, 140, 140, 140, 141, 193, 196, 232 et 347).
- D. Krull. Acoustic properties as predictors of perceptual responses : A study of Swedish voiced stops. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 12 :66–71, May 1988. (pages 137, 138, 139, 139, 139, 168 et 170).

- D. Krull. Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 10 :87–108, December 1989. (pages 91, 91, 91, 91, 91, 122, 132, 133, 133, 139, 140, 141, 141, 141, 141, 142, 142, 155, 155, 157, 184, 188, 196, 232, 233, 233, 233, 347 et 348).
- B. Kühnert and F. Nolan. The origin of articulation. In W. J. Hardcastle and N. Hewlett, editors, *Coarticulation, Theory, Data and Techniques*, Cambridge Studies in Speech Science and Communication, pages 8–30. Cambridge University Press, 1st edition, February 1999. (pages 73, 74, 74 et 110).
- F. Lacerda. Descriptive acoustic studies for the synthesis of spoken Swedish. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 4 :51–65, 1985. (pages 135, 136, 137, 137, 137, 170 et 174).
- F. Laclotte. L'harmonie vocalique. *La Parole. Revue internationale de rhinologie, otologie et laryngologie*, 1 :177–188, 1899. (pages 75, 75 et 76).
- P. Ladefoged. *Linguistic phonetics*. UCLA Working Papers in Phonetics 6, 1967. (page 81).
- P. Ladefoged. *A Course in Phonetics*. NY : Harcourt Brace Jovanovich, 1975. (page 16).
- P. Ladefoged. *Vowels and consonants : An introduction to the sound of language*. Blackwell, second edition, 2005. (pages 220, 229 et 229).
- I. Lehiste and G. E. Peterson. Transitions, glides and diphthongs. *The Journal of the Acoustical Society of America*, 33 :268–277, 1961. (page 37 et 37).
- A. M. Liberman, P. Delattre, and F. S. Cooper. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*, 65(4) : 497–516, 1952. (pages 43, 44 et 44).
- A. M. Liberman, P. C. Delattre, F. S. Cooper, and L. J. Gerstman. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs : General and Applied*, 68(8), 1954. (pages 54, 56, 56, 94, 95, 174, 174, 175, 178 et 179).
- A. M. Liberman, P. Delattre, and F. Cooper. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1(3) :153–167, 1958. (pages 9, 14, 14, 14 et 58).
- A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. Perception of the speech code. *Psychological Review*, 74(6) :431–461, 1967. (pages 14 et 15).

- B. Lindblom. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35(11) :1773–1781, November 1963a. (pages 18, 20, 90, 99, 99, 100, 101, 116, 121, 122, 132, 133, 134, 231, 238, 305 et 347).
- B. Lindblom. On vowel reduction. Technical Report 29, Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, 1963b. (pages 91, 99, 122, 133, 133, 134, 134, 135, 135, 143, 155, 155, 170, 174, 224, 232 et 347).
- B. Lindblom. The interdisciplinary challenge of speech motor control. In S. Grillner, B. Lindblom, J. Lubker, and A. Persson, editors, *Speech Motor Control*, pages 3–18. Pergamon Press, Proceedings of an International Symposium on Speech Motor Control, 1982. (page 79, 79 et 79).
- B. Lindblom. Economy of speech gestures. In P. M. MacNeilage, editor, *The production of speech*, pages 217–245. Springer-Verlag, 1983. (pages 80, 90, 91, 92, 92 et 92).
- B. Lindblom. Economy of speech gestures. In B. A. G. Elsendoom and H. Bouma, editors, *Working models of human perception*, pages 139–173. London : Academic Press, 1989. (pages 80, 90 et 91).
- B. Lindblom. Explaining phonetic variation : A sketch of H&H theory. In W. Hardcastle and A. Marchal, editors, *Speech production and speech modelling*, pages 403–439. Kluwer Academic Publishers, 1990. (pages 74, 80, 90, 91, 119, 119, 119, 120 et 121).
- B. Lindblom. Role of articulation in speech perception : Clues from production. *The Journal of the Acoustical Society of America*, 99(3) :1683–1692, 1996. (pages 119 et 120).
- B. Lindblom. An articulatory perspective on the locus equation. *Behavioral and Brain Sciences*, 21(2 Open Peer Commentary) :274–275, 1998. (page 169 et 169).
- B. Lindblom. A numerical model of coarticulation based on a principal components analysis of tongue shapes. In *ICPhS*. Barcelona, 2003. (page 204).
- B. Lindblom and R. Lindgren. Speaker-listener interaction and phonetic variation. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm-PERILUS*, 4 : 77–85, 1985. (pages 90 et 91).
- B. Lindblom and J. Sunberg. Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, 50(4B) :1166–1179, 1971. (pages 12, 12 et 91).
- B. Lindblom and H. M. Sussman. Articulatory and acoustic bases of locus equations. In *Fonetik*. Dept. of Linguistics, Stockholm University, 2004. (page 204).

- B. Lindblom and H. M. Sussman. Dissecting coarticulation : How locus equations happen. *Journal of Phonetics*, 40(1) :1–19, January 2012. (pages 123, 123, 135, 138, 174, 204, 205, 205, 206, 207, 215 et 232).
- B. Lindblom, S. Pauli, and J. Sundberg. Modeling coarticulation in apical stops. In G. Fant, editor, *Speech Communication*, volume II, pages 87–94. Uppsala : Almqvist and Wiksell, 1975. (page 91).
- B. Lindblom, S. Brownlee, B. Davis, and S.-J. Moon. Speech transforms. *Speech Communication*, 11(4) :357–368, 1992. (page 91).
- B. Lindblom, A. Agwuele, and H. M. Sussman. The effect of emphatic stress on consonant vowel coarticulation. *The Journal of Acoustical Society of America*, 121(6) :3802–3813, June 2007. (pages 142, 173, 174, 235 et 348).
- B. Lindblom, D. Krull, L. Hartelius, and E. Schalling. Formant transitions in normal and disordered speech : An acoustic measure of articulatory dynamics. In *Fonetik*. Dept. of Linguistics, Stockholm University, 2009. (pages 174, 205 et 205).
- B. Lindblom, D. Krull, and H. M. Sussman. Coarticulation as incomplete interpolation. In *Fonetik*. Sweden : Department of Phonetics, Center for Language and Literature, Lund University, 2010. (page 174).
- L. Lisker. On explaining vowel duration variation. *Glossa*, 2 :233–246, 1974. (page 20).
- L. Lisker and A. S. Abramson. A cross language study of voicing in initial stops : Acoustical measurements. *Word*, 20(3) :527–565, 1964. (pages 58, 59, 59, 59, 59, 60, 60 et 61).
- L. Lisker and A. S. Abramson. Some effects of context on voice onset time in english stops. *Language and Speech*, 10(1) :1–28, 1967. (pages 58 et 59).
- A. Löfqvist. Interarticulator phasing, locus equations, and degree of coarticulation. *The Journal of Acoustical Society of America*, 106(4) :2022–2030, 1999. (pages 193, 193, 194, 195 et 348).
- J. Lubker. Temporal aspects of speech production : Anticipatory labial coarticulation. *Phonetica*, 38(1-3) :51–65, 1981. (page 88).
- J. Lubker and T. Gay. Anticipatory labial coarticulation : Experimental, biological, and linguistic variables. *The Journal of the Acoustical Society of America*, 71(2) :437–448, 1982. (page 88).
- A. MacLeod, M. Baillargeon, D. E. Metz, N. Schiavetti, and R. L. Whitehead. Locus equations as a source of relational invariance for stop place categorization : A direct replication of

- Sussman, McCaffrey, and Mathews. *Contemporary Issues in Communication Science and Disorders*, 28 :98–103, 2001. (pages 149 et 150).
- P. MacNeilage and P. Ladefoged. The production of speech and language. In E. C. Carterette and M. P. Friedman, editors, *Handbook of perception : Language and Speech*, chapter 3, pages 75–120. Academic Press, New York, San Francisco, London, 1976. (page 7).
- P. F. MacNeilage. Motor control of serial ordering of speech. *Psychological Review*, 77(3) : 182–196, 1970. (page 74).
- A. Malécot. The role of releases in the identification of released final stops : A series of tape-cutting experiments. *Language*, 34(3) :370–380, 1958. (pages 45, 46 et 47).
- A. Malécot. The lenis-fortis opposition : Its physiological parameters. *The Journal of the Acoustical Society of America*, 47(2) :1588–1592, 1970. (page 62, 62 et 62).
- A. Marchal and T. Legou. L'électropalatographie : questions de méthodologie, données et nouveaux développements. In A. Marchal and C. Cavé, editors, *L'imagerie médicale pour l'étude de la parole*, pages 87–123. Lavoisier, 2009. (page 115).
- R. McAulay and T. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(4) :744–754, 1986. (page 26).
- D. R. McCloy. Normalizing and plotting vowels with phonR 1.0.7, 2016. URL <http://drammock.github.io/phonR/>. (page 252 et 252).
- E. McDermott and S. Katagiri. Phoneme recognition using kohonen's learning vector quantization. *ATR Workshop on Neural Networks and Parallel Distributed Processing*, Japan, 1988. (page 183).
- P. Menzerath and A. de Lacerda. *Koartikulation, Steuerung und Lautabgrenzung : eine experimentelle Untersuchung*. Number 1 in *Phonetische Studien*. Berlin : F. Dümmler, 1933. (pages 75, 75, 76 et 77).
- C. Meunier. Invariants et Variabilité en Phonétique. In *Phonologie et phonétique : forme et substance*, pages 350–374. Hermès Lavoisier, 2005. (pages 74 et 120).
- Y. Meynadier. éléments de phonétique acoustique. In N. Nguyen and M. Adda-Decker, editors, *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, pages 25, 83. Hermès Lavoisier, 2013. (page 84).
- G. Modarresi, H. M. Sussman, B. Lindblom, and E. Burlingame. Stop place coding : an acoustic study of CV, VC#, and C#V sequences. *Phonetica*, 61 :2–20, 2004a. (pages 172, 173, 173, 173, 173, 214, 224 et 228).

- G. Modarresi, H. M. Sussman, B. Lindblom, and E. Burlingame. An acoustic analysis of the bidirectionality of coarticulation in VCV utterances. *Journal of Phonetics*, 32 :291–312, 2004b. (page 173 et 173).
- G. Modarresi, H. M. Sussman, B. Lindblom, and E. Burlingame. Locus equation encoding of stop place : revisiting the voicing/VOT issue. *Journal of Phonetics*, 33 :101–113, 2005. (pages 209, 209, 210 et 246).
- M. R. Molis, B. Lindblom, W. Castelman, and R. Carré. Cross-language analysis of VCV coarticulation. *The Journal of the Acoustical Society of America*, 95 :2925, 1994. (page 208).
- K. L. Moll and R. G. Daniloff. Investigation of the timing of velar movements during speech. *The Journal of the Acoustical Society of America*, 50(2B) :678–684, 1971. (pages 79, 81 et 82).
- A. Montgomery, P. E. Reed, K. Crass, I. Hubbard, and J. Stith. The effects of measurement error and vowel selection on the locus equation mesure of coarticulation. *The Journal of the Acoustical Society of America*, 136(5) :2747–2750, 2014. (pages 212, 213 et 214).
- J. Morais, L. Cary, J. Alegria, and P. Bertelson. Does awareness of speech as a sequence of phones arise spontaneously ? *Cognition*, 7 :323–331, 1979. (page 73 et 73).
- M. Mrayati, R. Carre, and B. Guerin. Distinctive regions and modes : a new theory of speech production. *Speech Communication*, 7(3) :257–286, 1988. (page 107).
- A. Nawafleh. *Difficultés de prononciation et de perception de voyelles du Français par des apprenants jordaniens*. Thèse de doctorat, Université Sorbonne Nouvelle - Paris 3, 2012. (pages 221 et 222).
- T. M. Nearey and S. E. Shammass. Formant transitions as partly distinctive invariant properties in the identification of voiced stops. *Canadian Acoustics*, 15(4), 1987. (pages 123, 133, 135, 143, 145, 156 et 157).
- T. M. Neary and P. F. Assman. Modeling the role of vowel inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, 80 :1297–1308, 1986. (pages 37, 37, 38, 38 et 39).
- A. Noiray, L. Ménard, and K. Iskarous. The development of motor synergies in children : Ultrasound and acoustic measurements. *The Journal of the Acoustical Society of America*, 133(1) :444–452, 2013. (pages 204, 213 et 215).
- J. J. Ohala. The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage, editor, *The Production of Speech*, pages 189–216. Springer Verlag, Heidelberg and Berlin, 1983. (page 41 et 41).

- R. N. Ohde and K. N. Stevens. Effect of burst amplitude on the perception of stop consonant place of articulation. *The Journal of the Acoustical Society of America*, 74(3) :706–714, 1983. (pages 52, 52, 53, 64 et 64).
- S. Öhman. Coarticulation in VCV utterances : Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1) :151–168, January 1966. (pages 14, 85, 95, 95, 103, 103, 104, 104, 105, 105, 105, 107, 107, 107, 107, 143, 167, 232 et 232).
- S. Öhman. title. *STL-QPSR*, 8(2-3) :20–54, 1967. (page 102).
- D. O’Shaughnessy. The effects of speaking rate on formant transitions in French synthesis-by-rule. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP ’86.*, volume 11, pages 2027–2030, 1986. (pages 20, 91, 102, 132, 231, 233 et 305).
- J. S. Perkell. Testing theories of speech production : Implications of some detailed analyses of variable articulatory data. In W. Hardcastle and A. Marchal, editors, *Speech Production and Speech Modelling*, pages 263–288. Dordrecht, 1990. (page 89 et 89).
- J. S. Perkell and C. Cohen. Preliminary support for a "hybrid model" of anticipatory coarticulation. In *Proceedings of the 12th International Congress of Acoustics*, pages A3–6, 1986. (page 89 et 89).
- J. S. Perkell and M. L. Matthies. Temporal measures of anticipatory labial coarticulation for the vowel /u/ : Within- and cross-subject variability. *The Journal of the Acoustical Society of America*, 91(5) :2911–2925, 1992. (page 90).
- J.-M. Peterfalvi. La perception de la parole d’après les expériences de synthèse acoustique. *L’année psychologique*, 66(2) :559–577, 1966. (page 93).
- G. E. Peterson and H. L. Barney. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24 :175–184, 1952. (pages 33, 33, 33 et 34).
- Population and Social Statistics Directorate of the Hashemite Kingdom of Jordan. The population of the kingdom by administrative divisions, according to the general census of population and housing, 2015. URL [http://www.dos.gov.jo/dos\\_home\\_a/main/population/census2015/No\\_of\\_pop\\_depand\\_on\\_GOV.pdf](http://www.dos.gov.jo/dos_home_a/main/population/census2015/No_of_pop_depand_on_GOV.pdf). (page 221).
- R. K. Potter, G. A. Kopp, and H. C. Green. *Visible speech*. New York : Van Nostrand, 1947. (page 54).
- R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0. (page 243).
- D. Recasens. V-to-V coarticulation in Catalan VCV sequences : an articulatory and acoustical study. *Journal of Phonetics*, 12 :61–73, 1984a. (pages 87, 87, 113 et 199).

- D. Recasens. V-to-V coarticulation in Catalan VCV sequences. *The Journal of the Acoustical Society of America*, 76(6) :1624–1635, December 1984b. (pages 87, 113, 114 et 199).
- D. Recasens. Coarticulatory patterns and degrees of coarticulatory resistance in catalan VCV sequences. *Language Speech*, 28 :97–114, 1985. (pages 87, 111, 111, 111, 113, 114, 115, 117, 189, 199, 199, 199, 200 et 259).
- D. Recasens. Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences. *Speech Communications*, 8 :299–312, 1989. (page 87).
- D. Recasens. The articulatory characteristics of palatal consonants. *Journal of Phnoetics*, 18 : 267–280, 1990. (page 113).
- D. Recasens. An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phnoetics*, 19 :179–192, 1991. (page 115).
- D. Recasens. Adaptation du lieu d’articulation dans les groupes consonantiques du catalan à la lumière du modèle DAC. In M. Embarki and C. Dodane, editors, *La coarticulation. Des indices à la représentation*, pages 19–35. L’Harmattan, 2011. (pages 113 et 118).
- D. Recasens. *Coarticulation and Sound Change in Romance*. John Benjamins, 2013. (page 117).
- D. Recasens and A. Espinosa. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, 125(4) :2288–2298, 2009. (pages 111, 116, 117, 203 et 261).
- D. Recasens, M. D. Pallarès, and J. Fontdevila. A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, 102 :544–561, 1997. (pages 87, 113, 113, 116, 116, 224, 225 et 261).
- B. H. Repp and H. Lin. Acoustic properties and perception of stop consonant release transients. *The Journal of the Acoustical Society of America*, 85(1) :379–396, 1989. (page 42).
- A. Rhone and A. Jongman. Modified locus equations categorize stop place in a perceptually realistic time frame. *The Journal of Acoustical Society of America*, 131(6) :EL487–EL491, May 2012. (pages 210, 211, 212 et 232).
- W. Richard. Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, and R. Temple, editors, *Phonetic Interpretation : Papers in Laboratory Phonology VI*, pages 75–87. Cambridge University Press, Cambridge, 2004. (page 121).
- B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, and D. Firth. Support Functions and Datasets for Venables and Ripley’s MASS. <https://cran.r-project.org/web/packages/MASS/MASS.pdf>, 2017. Version 7.3-48. (page 332).

- J.-P. Rousselot. *Principes de phonétique expérimentale*, volume 2. H. WELTER, 1901-1908. (page 75).
- E. Saltzman. Task dynamic coordination of the speech coarticulation : A preliminary model. *Haskins Laboratories Status Report on Speech Research*, SR-84 :1–18, 1985. (page 124 et 124).
- E. Saltzman. The task dynamic model in speech production. In H. Peters, W. Hulstijn, and W. Starkweather, editors, *Speech motor control and stuttering*, pages 37–52. Elsevier Science Publishers B. V., 1991. (pages 125, 127, 128 et 129).
- E. Saltzman and S. Kelso. Skilled actions : A task–dynamic approach. *Haskins Laboratories Status Report on Speech Research*, SR-76 :3–50, 1983. (pages 124, 124, 124 et 126).
- E. Saltzman and K. Munhall. A dynamical approach to gestural patterning in speech production. *Haskins Laboratories Status Report on Speech Research*, SR-99(100) :38–68, 1989. (pages 80, 87, 87, 124, 124 et 127).
- J. R. Sawusch. Effects of duration and formant movement on vowel perception. In *4th International Conference on Spoken Language Processing (ICSLP 96)*. Philadelphia, PA, USA, 1996. (pages 24, 24, 24, 25 et 25).
- J. R. Sawusch. Acoustic analysis and synthesis of speech. In D. B. Pisoni and R. E. Remez, editors, *Handbook of speech perception*, chapter 1, pages 7–27. Blackwell Publishing, 2005. (page 10).
- H. K. Sfford, H. S. Hoffman, A. M. Liberman, D. P. C., and F. S. Cooper. Effect of third-formant transitions on the perception of the voiced stop consonants. *The Journal of the Acoustical Society of America*, 30(2) :122–126, 1958. (page 56).
- S. Shammass. *Formant transitions, spectral shape, and vowel context in perception of voiced stops*. PhD thesis, University of Alberta, 1985. (pages 143, 143, 144, 144, 145, 145, 145, 174 et 207).
- Sharf and Hemeyer. Identification of place of consonant articulation from vowel formant transitions. *The Journal of the Acoustical Society of America*, 51(2) :652–658, 1972. (page 63).
- K. Sjolander and J. Beskow. WaveSurfer, 2013. URL <http://www.speech.kth.se/wavesurfer/>. (page 212).
- K. Stevens. Design features of speech sound systems. In P. M. MacNeilage, editor, *The production of speech*, pages 248–261. Springer-Verlag, 1983. (page 7).
- K. Stevens and H. Arthur. Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America*, 27(3) :484–493, 1956. (pages 9 et 16).

- K. Stevens and A. House. Studies of formant transitions using a vocal tract analog. *The Journal of the Acoustical Society of America*, 28(4) :578–585, 1956. (page 94).
- K. N. Stevens. Effect of duration upon vowel identification. *The Journal of the Acoustical Society of America*, 31(1) :109, 1959. (pages 22, 22 et 23).
- K. N. Stevens and S. Blumstein. The search for invariant acoustic correlates of phonetic features. In D. P. Eimas and J. L. Miller, editors, *Perspectives on the study of speech*, chapter 1, pages 1–38. Lawrence Erlbaum Associates, Inc. Publishers, Hillsdale, New Jersey, 1981. (pages 47, 143 et 143).
- K. N. Stevens and S. E. Blumstein. Invariant cues for place of articulation in stop consonants. *The Journal of Acoustical Society of America*, 64(5) :1358–1368, 1978. (pages 47, 47, 47, 48, 49, 49, 50, 64, 69 et 232).
- K. N. Stevens, S. Kasowski, and C. G. M. Fant. An electrical analog of the vocal tract. *The Journal of the Acoustical Society of America*, 25(4) :734–742, 1953. (page 95).
- N. K. Stevens. *Acoustic phonetics*. The MIT Press, first edition, 1998. (page 230).
- S. S. Stevens, J. Volkman, and E. B. Newman. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3) :185–190, 1937. (page 109).
- M. Stone. Laboratory techniques for investigating speech articulation. In W. J. Hardcastle, J. Laver, and F. E. Gibbon, editors, *The Handbook of Phonetic Sciences*, pages 9–38. Wiley-Blackwell, second edition, 2010. (pages 115 et 116).
- W. Strange. Information for vowels in formant transitions. *Journal of Memory and Language*, 26 :550–557, 1987. (page 34).
- W. Strange and J. J. Jenkins. Dynamic specification of coarticulated vowels : Research chronology, theory and hypotheses. In G. S. Marrison and P. F. Assmann, editors, *Vowel Inherent Spectral Change*, pages 9–30. Springer, 2013. (pages 35, 36 et 37).
- W. Strange, R. R. Verbrugge, D. P. Shankweiler, and T. R. Edman. Consonant environment specifies vowel identity. *The Journal of Acoustical Society of America*, 60(1) :213–224, 1976. (page 34, 34, 34 et 34).
- W. Strange, R. R. Verbrugge, D. P. Shankweiler, and T. R. Edman. Dynamic specification of coarticulated vowels. *The Journal of Acoustical Society of America*, 74(3) :695–705, 1983. (pages 34, 34, 36, 37 et 37).

- N. Suga, K. Kujirai, and T. Manabe. Specificity of combination-sensitive neurons for processing of complex biosonar signals in auditory cortex of the moustached bat. *Neurophysiology*, 49 : 1573–1627, 1983. (page 182).
- H. M. Sussman. Neural coding of relational invariance in speech : Human language analogs to the barn owl. *Psychological Review*, 96(4) :631–642, 1989. (pages 145, 146, 146, 146, 147, 179 et 348).
- H. M. Sussman. The phonological reality of locus equations across manner class distinctions : Preliminary observations. *Phonetica*, 51(1-3) :119–131, 1994. (pages 134, 135, 159, 160, 160, 161, 161, 169, 199 et 200).
- H. M. Sussman. Representation of phonological categories : A functional role for auditory columns. *Brain and Language*, 80 :1–13, 2002. (page 184).
- H. M. Sussman. Neuroethology in the service of neurophonetics. *Journal of Neurolinguistics*, 26 :511–525, 2013. (page 184 et 184).
- H. M. Sussman and J. Shore. Locus equation as phonetic descriptors of consonantal place of articulation. *Psychonomic Society, Inc.*, 58(6) :936–946, 1996. (pages 159, 161, 161, 162, 162, 162, 163, 163, 163, 164, 165, 165, 165, 165, 207, 208, 209, 209, 232, 246, 247, 247, 248 et 351).
- H. M. Sussman and J. Westbury. The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation. *Journal of Speech Language and Hearing Research*, 24(1) :16–24, 1981. (page 88 et 88).
- H. M. Sussman, H. A. McCaffrey, and S. A. Matthews. An investigation of locus equations as a source of relational invariance for stop consonant place categorization. *Journal of the Acoustical Society of America*, 90 :1309–1325, November 1991. (pages 91, 99, 122, 123, 123, 124, 133, 135, 143, 148, 148, 149, 149, 149, 149, 149, 150, 150, 150, 151, 153, 155, 156, 156, 157, 157, 158, 158, 158, 158, 158, 159, 160, 160, 163, 164, 164, 169, 174, 175, 177, 177, 177, 199, 200, 220, 224, 228, 228, 232, 330 et 347).
- H. M. Sussman, K. A. Hoemeke, and H. A. McCaffrey. Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech, Language, and Hearing Research*, 35 :769–781, 1992. (pages 150, 150, 151, 152 et 152).
- H. M. Sussman, K. A. Hoemeke, and F. S. Ahmed. A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *The Journal of the Acoustical Society of America*, 94(3) :1256–1268, 1993. (pages 143, 154, 154, 155, 155, 155, 156, 156, 157, 157, 161, 168, 169, 182, 227 et 347).

- H. M. Sussman, D. Fruchter, and A. Cable. Locus equations derived from compensatory articulation. *The Journal of the Acoustical Society of America*, 97(5) :3112–3112, 1995. (pages 159, 165, 166, 166 et 166).
- H. M. Sussman, F. D. Minifie, E. Buder, C. Stoel-Gammon, and J. Smith. Consonant vowel interdependencies in babbling and early words : Preliminary examination of a locus equation approach. *Journal of Speech and Hearing Research*, 39 :424–433, 1996. (pages 122, 123, 124, 152, 153 et 153).
- H. M. Sussman, N. Bessell, E. Dalston, and T. Majors. An investigation of stop place of articulation as a function of syllable position : A locus equation perspective. *The Journal of the Acoustical Society of America*, 101(5) :2826–2838, 1997. (pages 170, 171, 171 et 173).
- H. M. Sussman, E. Dalston, and S. Gumbert. The effect of speaking style on a locus equation characterization of stop place of articulation. *Phonetica*, 55 :204–225, 1998a. (pages 191, 192, 192 et 193).
- H. M. Sussman, D. Fruchter, J. Hilbert, and J. Sirosh. Linear correlates in the speech signal : The orderly output constraint. *Behavioral and Brain Sciences*, 21 :241–299, 1998b. (pages 99, 123, 124, 179, 180, 180, 182, 182, 182, 183, 183, 184, 215, 215, 232, 233, 348, 348, 348 et 352).
- M. Tabain. Coarticulation in CV syllables : a comparison of locus equation and EPG data. *Journal of Phonetics*, 28 :137–159, 2000. (pages 195, 196, 196, 197, 197, 197 et 198).
- M. Tabain. Voiceless consonant and locus equations : A comparison with electropalatographic data on coarticulation. *Phonetica*, 59 :20–37, 2002. (pages 197, 198 et 198).
- M. Tekieli and W. L. Cullinan. The perception of temporally segmented vowels and consonant-vowel syllables. *Journal of Speech, Language, and Hearing Research*, 22 :103–121, 1979. (page 210).
- S. Thornton and A. Rex. *Physique moderne*. De Boeck, 3e edition, 2009. (page 181).
- W. R. Tiffany. Vowel recognition as a function of duration, frequency modulation and phonetic context. *Journal of Speech and Hearing Disorders*, 18(3) :289–301, 1953. (page 22).
- H. Traunmüller. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, 88(1) :97–100, 1990. URL <https://doi.org/10.1121/1.399849>. (page 254).
- K. Tsukada. An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai : Durational and spectral. *International Journal on Asian Language Processing*, 19(4) :127–138, 2009. (pages 20 et 231).

- UNESCO. Message from Ms Irina Bokova, Director-General of UNESCO, on the occasion of the first World Arabic Language Day 18 December 2012. URL <http://www.unesco.org/new/en/unesco/events/prizes-and-celebrations/celebrations/international-days/world-arabic-language-day/>. (page 220).
- J. van Santen. Contextual effects on vowel duration. *Speech Communication*, 11(6) :513 – 546, 1992. (page 21).
- R. J. J. H. Van Son and L. C. W. Pols. Formant frequencies of dutch vowels in a text, read at normal and fast rate. *The Journal of the Acoustical Society of America*, 88(4) :1683–1693, 1990. (page 18).
- B. Vaxelaire, A. Marchal, F. Hirsch, and R. Sock. Apports des techniques radiologiques et de la radiocinématographie à l'étude de la production de la parole. In A. Marchal and C. Cavé, editors, *L'imagerie médicale pour l'étude de la parole*, pages 125–145. Lavoisier, 2009. (page 105).
- K. Versteegh. *The Arabic language*. Edinburgh University Press Ltd, second edition, 2014. (page 221, 221 et 221).
- A. Wagner, M. Ernestus, and A. Cutler. Formant transitions in fricative identification : The role of native fricative inventory. *The Journal of Acoustical Society of America*, 120(4) : 2267–2277, 2006. (page 14 et 14).
- H. Wagner, T. Takahashi, and M. Konishi. Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *Journal of Neuroscience*, 7 :3105–3116, 1987. (page 182).
- C. I. Watson and J. Harrington. Acoustic evidence for dynamic formant trajectories in australian english vowels. *The Journal of Acoustical Society of America*, 106(1) :458–468, 1999. (page 22).
- J. C. E. Watson. *The phonology and morphology of Arabic*. Oxford University Press, first edition, 2002. (page 224).
- G. Weismer and J. Berry. Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *The Journal of Acoustical Society of America*, 113(6) :3362–3378, June 2003. (page 102).
- J. Westbury. *Xray Microbeam Speech Production Database Users Handbook*. University of Wisconsin, Madison, WI, 1994. (page 201).
- G. Willett. Paradigme, théorie, modèle, schéma : qu'est-ce donc ? *Communication et organisation*, 10 :1–19, 1996. (pages 78 et 79).

- M. Yeou. Locus equations and the degree of coarticulation of Arabic consonant. *Phonetica*, 54 :187–202, 1997. (pages 164, 164, 164, 165 et 227).
- Zahorian and Jagharghi. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, 94(4) :1966–1982, 1993. (page 21 et 21).
- B. Zawaydeh. An acoustic analysis of uvularization spread in Ammani-Jordanian Arabic. *Studies in the Linguistic Sciences*, 27(1) :185–200, 1997. (pages 226, 229 et 229).
- B. Zawaydeh. *The Phonetics and Phonology of Gutturals in Arabic*. PhD thesis, Indiana University, 1999. (pages 225, 225, 225 et 226).
- B. Zawaydeh and K. de Jong. Uvularization spread in Arabic. *IULC Working Papers*, 2(2) : 93–107, 2002. (page 226).
- B. Zawaydeh and K. de Jong. The phonetics of localising uvularisation in Ammani-Jordanian Arabic. In Z. M. Hassan and B. Heselwood, editors, *Instrumental Studies in Arabic Phonetics*, page 257–276. John Benjamins Publishing Company, 2011. (page 226).

# Thèse de Doctorat

Mohammad ABUOUEH

De l'impact des variations temporelles sur les transitions formantiques

On the impact of temporal variations on formant transitions

## Résumé

Cette étude permet d'explorer les variations spectrales provoquées par deux types de changements temporels qui résultent respectivement de l'opposition de longueur vocalique (variations intrinsèques) et des variations du débit de parole (variations extrinsèques). Trois protocoles expérimentaux ont été conçus, deux en arabe jordanien et un en français, pour examiner ce phénomène. Un intérêt particulier a été porté aux occlusives produites dans des séquences CVC dans le but d'évaluer la coarticulation de la consonne en position initiale et finale. La durée des voyelles et la fréquence des trois premiers formants au début ou à la fin et au milieu de chaque séquence ont été mesurées dans chaque condition de longueur et de débit. Les équations de locus ont été utilisées afin de décrire la relation CV/VC quand elle subit ces deux types de variations. Selon les résultats, la qualité de la voyelle et de la consonne est influencée par l'opposition de durée et par le débit de parole de manière distincte. Le changement généré par les variations temporelles est évalué à l'aide des équations de locus et à travers des analyses de dispersion. Ces outils révèlent qu'il existe un chevauchement coarticulatoire plus important quand la durée de la voyelle est réduite, peu importe l'origine des variations temporelles dans les deux langues (intrinsèque et extrinsèque). Par contre, les variations temporelles impactent la classification des occlusives de manière distincte selon leur origine.

## Mots clés

Variations temporelles, débit de parole, longueur vocalique, coarticulation, équations de locus, arabe jordanien, français, phonétique, classification des occlusives.

## Abstract

The aim of this study is to examine the different spectral variations that are impacted by two types of temporal modifications: phonological vowel length contrast (intrinsic variations) and varying speaking rates (extrinsic variations). Three experiments, two on Jordanian Arabic and one on French, were conducted to investigate this phenomenon. We were interested in stop consonants produced in the initial and the final positions in CVC sequences. Vowel duration and the frequency of the first three formants were measured for each vowel length and speaking rate. In addition, locus equations were computed to measure the impact of time variations on CV / VC coarticulation. According to the results, the vowel and consonant qualities are influenced both by vowel duration contrasts and by speaking rate modifications in different ways. Such modifications are evidenced by locus equations and by dispersion analyses as well. These two quantification tools revealed that there is greater coarticulatory overlap when the duration of the vowel is reduced, regardless of the origin of temporal variations in the two languages (intrinsic and extrinsic). On the other hand, temporal variations affect the classification of the stop consonant differently according to their origin.

## Key Words

Temporal variations, speaking rate, vowel length, coarticulation, locus equations, Jordanian Arabic, French, phonetics, stop consonant classification.