

Étude phonétique des dialectes modernes de
l'anglais des Îles Britanniques :
vers l'identification automatique du dialecte

THÈSE

présentée et soutenue publiquement le 4 juin 2008

pour l'obtention du

Doctorat en Lexicologie et Terminologie Multilingues, Traduction
(Phonétique)

par

Emmanuel Ferragne

sous la direction de

M. le Professeur Claude Boisson

Composition du jury

Rapporteurs : M. Jacques Durand Professeur à l'Université de Toulouse-Le Mirail

M. Francis Nolan Professeur à l'Université de Cambridge

Examineurs : M. Claude Boisson Professeur à l'Université Lumière Lyon 2

M. Daniel Hirst D.R. CNRS - Laboratoire Parole et Langage

M. François Pellegrino C.R. CNRS - Laboratoire Dynamique Du Langage

Mis en page avec la classe thloria.

Remerciements

Je tenais à remercier en premier lieu M. Claude Boisson pour avoir assuré la direction de ce travail, et, plus généralement, pour son rôle déterminant dans mon cursus universitaire.

M. François Pellegrino a co-encadré cette thèse ; qu'il soit vivement remercié pour son investissement considérable au quotidien dans la réalisation de mon projet.

J'ai passé l'intégralité de ces années de thèse au Laboratoire Dynamique Du Langage, à Lyon. Je remercie tous les membres du laboratoire – qu'ils soient doctorants, ingénieurs, chercheurs, enseignants-chercheurs ou membres du personnel administratif – pour leur accueil et l'excellente atmosphère de travail dans laquelle nous avons évolué.

Je remercie, pour leur accueil et pour leur aide, les collègues et amis d'outre-Manche : à Londres, Volker Dellwo et Mark Huckvale, à Cambridge, Francis Nolan et tous les membres du Phonetics Laboratory.

Merci enfin à toutes les personnes ayant contribué, d'une manière ou d'une autre, à l'aboutissement de ce travail, et en particulier, ma famille et mes amis.

I wryte moued by good wil, and not to shewe my skill.

Gascoigne, 1575, Certayne notes. . .

Résumé

Parmi les phénomènes qui affectent la manière dont nous parlons, l’accent est une des composantes principales de la variation observée. La prononciation d’un locuteur peut en effet nous renseigner sur son origine, géographique et sociale. La description des caractéristiques phonétiques qui sous-tendent les différences d’accent perçues constitue donc un intérêt scientifique particulier. De plus, la recherche dans le domaine des accents contribue à l’amélioration d’applications technologiques telles que la reconnaissance de la parole ou l’indexation du locuteur.

Cette thèse propose une étude phonétique acoustique d’enregistrements de 13 accents (264 locuteurs) tirés du corpus Accents of the British Isles (ABI). Nous abordons dans un premier temps les caractéristiques essentielles à une meilleure compréhension de la variation induite par des facteurs tels que le dialecte, les spécificités physiologiques du locuteur et le contenu informationnel. Nous définissons la variation dialectale à travers les concepts d’ensembles lexicaux, de scissions et convergences phonémiques, de changements en chaîne, et mentionnons également, entre autres, les phénomènes d’hypercorrection et de « quasi convergences ». De plus, certaines études traitant de questions connexes, comme l’identification du dialecte par des auditeurs ou le nivellement dialectal, sont passées en revue, et les traits phonétophonologiques les plus saillants de chaque dialecte du corpus sont listés. Les aspects physiologiques sont abordés en référence à la question de la normalisation du locuteur ; une solution à ce problème est suggérée à la fin de la thèse. La variation induite par le contexte informationnel est analysée par le biais de la théorie H&H de Lindblom et de la *Smooth Signal Redundancy Hypothesis* de Aylett ; ce point permet de justifier l’utilisation de mots à structure /hVd/ dans notre étude. Nous posons en outre les bases d’un cadre théorique décrivant les relations entre la phonétique et les représentations phonologiques prototypiques, et l’interprétation de la distance phonétique entre locuteurs comme un degré d’appartenance à des catégories de prototypes dialectaux.

La section empirique de nos travaux peut-être divisée en trois parties :

1. Rythme de la parole ;
2. Description phonétique des systèmes vocaliques ;
3. Classification automatique et représentations multidimensionnelles.

Dans la première partie, des procédures classiques dans l’étude du rythme – basées sur des mesures de durée – sont appliquées au passage lu du corpus ABI. Nous introduisons en outre une nouvelle mesure qui prend en compte l’intensité. Une série d’analyses discriminantes montrent que ces paramètres ne permettent pas une bonne « séparabilité » entre les dialectes ; cependant, notre mesure d’intensité semble présenter un potentiel discriminant supérieur à celui des mesures courantes. Les résultats suggèrent également que, au vu de la grande variabilité du débit de parole, et malgré l’utilisation d’indices dits « normalisés », la faisabilité de telles études à partir de données non contrôlées est douteuse. Néanmoins, le potentiel discriminant du rythme, tel que nous le mesurons, semble varier d’un dialecte à l’autre.

Dans la deuxième partie, les systèmes vocaliques sont décrits à partir de F1, F2 et de la durée. Les données sont composées de 19 voyelles présentées dans un contexte consonantique /h_d/ ; une discussion des avantages et des inconvénients de ces mots-test est proposée au début de la thèse. Les formants ont été extraits automatiquement avec le logiciel Praat. Étant donné le peu de fiabilité de l’estimation de formants, les mesures ont été ensuite traitées par le biais de la régression linéaire et polynomiale afin de lisser les trajectoires formantiques et d’écarter les valeurs déviantes. Pour chaque dialecte, nous représentons graphiquement les monophthongues et les diphtongues, et nous nous concentrons sur les paires de voyelles qui sont particulièrement proches. La variation intra-dialectale est illustrée au moyen de spectrogrammes individuels.

La troisième partie aborde la classification automatique de locuteurs en fonction du dialecte, partant de l’idée selon laquelle les accents – tels qu’ils sont mesurés à travers le système vocalique des individus – sont mieux représentés, et mieux préservés de la variation individuelle d’ordre physiologique, lorsqu’on utilise les coordonnées relatives des voyelles dans l’espace acoustique plutôt que les valeurs absolues comparées à une norme. Notre méthode permet d’atteindre un peu plus de 90 % de classification correcte dans une tâche à 13 dialectes. Nous démontrons ensuite l’intérêt de techniques d’analyse multivariée pour la projection de données de l’espace des *Mel-Frequency Cepstral Coefficients* dans un espace favorisant l’interprétation phonétique. Les résultats offrent un nouvel éclairage sur les systèmes vocaliques et montrent que le regroupement des dialectes est en accord avec nos prédictions.

Mots-clés: Accents et dialectes, Îles Britanniques, phonétique acoustique, classification automatique.

Abstract

Among the many features affecting the way people speak, accents constitute a major component of variation. A speaker's pronunciation can indeed be very informative as to who s/he is, and where s/he comes from, both geographically and socially. A description of the phonetic phenomena underlying perceived differences in accent are therefore of particular scientific interest. Besides, accents research is crucial to the improvement of technological applications such as speech recognition and speaker indexing.

This Ph.D is an acoustic-phonetic investigation of speech samples of 13 accents (264 speakers) from the Accents of the British Isles (ABI) corpus. The dissertation starts by outlining some key features for a better understanding of the dialect-related, the physiological, and the informational factors involved in phonetic variation. Dialect-specific variation is examined mainly through the concepts of lexical sets, phonemic splits and mergers, and chain shifts, and phenomena such as hypercorrection and near-mergers are mentioned. We also review a number of studies covering connected issues ranging from accent identification by listeners to dialect levelling, and draw a list of the most salient phonetic or phonological features for each dialect in the corpus. The physiological aspects are discussed with reference to speaker normalization techniques; a solution to this problem is suggested at the end of the dissertation. The phonetic variation caused by the informational content of an utterance is analyzed thanks to Lindblom's H&H theory and Aylett's Smooth Signal Redundancy Hypothesis; this point is relevant to justifying the use of /hVd/ words. We also sketch out a theoretical framework describing the mapping between phonetics and prototype phonological representations, and the interpretation of phonetic distance between speakers as degrees of membership to prototype dialect categories.

The empirical section of this research can be divided into three parts :

1. Speech rhythm ;
2. Phonetic description of vowel systems ;
3. Automatic classification and multidimensional representations.

In the 1st part, standard procedures in the study of speech rhythm – based on duration measurements – are applied to a read passage from the ABI corpus. We also introduce a new metric that takes intensity into account. The results of a series of discriminant analyses show that separability between accents cannot be achieved on the basis of rhythm only ; however, intensity as a parameter performed better than duration. The findings also suggest that, on account of great variation in speech rate, and despite the use of so-called normalized metrics, the feasibility of such studies with uncontrolled data is highly questionable. However, the discriminatory power of the metrics we use depends on the dialect under study.

In the 2nd part, the vowel systems are described using F1, F2, and duration. The material consists of 19 vowels embedded in /h_d/ frames; the advantages and drawbacks of using such test words are thoroughly described at the beginning of the dissertation. The formants were extracted automatically using the Praat software. Given the lack of total reliability of formant extraction, the measurements were post-processed using linear and polynomial regression in order to smooth formant trajectories and rule out outliers. For each dialect, a plot of the monophthongs and diphthongs is displayed, and particular attention is paid to vowels that are conspicuously close to one another in the acoustic space. Variation within a dialect is illustrated through individual spectrograms.

The 3rd part tackles the automatic classification of speakers into dialects, building on the idea that accents – as measured through speakers' vowel systems – are better represented, and better preserved from individual physiology-related acoustic information, by the relative locations of their vowels, rather than the absolute acoustic values compared to a norm. The method yields more than 90 % correct classification in a 13-class task. We then demonstrate the relevance of multivariate techniques to map the high dimensional Mel-Frequency Cepstral Coefficient acoustic space into an easily interpretable phonetic space. The findings offer new insight into a speaker's vowel system and show that dialects cluster together in accordance with our predictions.

Keywords: Accents and dialects, British Isles, acoustic phonetics, automatic classification.

Table des matières

Partie I Cadre de l'étude

1

Introduction

2

Orientation théorique et méthodologique, et notions élémentaires

2.1	Cadre dialectologique	9
2.2	Variation dialectale : évolution des systèmes vocaliques et cas limites	15
2.2.1	Les changements en chaîne	15
2.2.2	Les convergences phonémiques	16
2.2.3	Les scissions phonémiques	19
2.2.4	Les quasi-convergences	21
2.3	Cadre phonético-phonologique : variation, information et prototypes	24
2.3.1	Variation phonético-acoustique	26
2.3.2	Caractéristiques phonétiques des mots-test et information	32
2.3.3	Théorie du prototype et théorie des ensembles flous	40
2.3.4	La notion de catégorie dialectale	52
2.4	Cadre statistique	58
2.4.1	Les mesures de distance	59

2.4.2	La classification hiérarchique	61
2.4.3	L'analyse en composantes principales	62
2.4.4	L'analyse linéaire discriminante	62
2.4.5	« K-means clustering » et qualité d'une partition	63
2.4.6	La régression	64

3

Situation géolinguistique et anglais standard

3.1	L'anglais standard	67
3.1.1	Système vocalique de l'anglais standard	70
3.1.2	Changements récents et perspectives	80
3.2	Caractéristiques phonético-phonologiques des dialectes des Îles Britanniques	83
3.2.1	Partition dialectale de l'Angleterre	85
3.2.1.1	Partition nord/sud : scission FOOT-STRUT et BATH-broadening	85
3.2.1.2	La rhoticité	90
3.2.1.3	L'occlusive vélaire dans <ng>	92
3.2.1.4	Le Yod-dropping	93
3.2.1.5	Le <i>happy</i> -tensing	95
3.2.1.6	La voyelle de FACE	96
3.2.1.7	La vocalisation du /l/ (L-vocalization)	98
3.2.2	Le Pays de Galles	99
3.2.3	L'Écosse	102
3.2.4	L'Irlande	104
3.2.4.1	Le sud	106
3.2.4.2	Le nord	108

Partie II Description phonético-phonologique des dialectes et classification automatique

4**Introduction à l'analyse empirique**

- 4.1 Description du corpus Accents of the British Isles (ABI) 114
- 4.2 Homogénéité et typicalité des dialectes de ABI : analyse auditive . . . 116

5**Le rythme**

- 5.1 Préambule 121
- 5.2 Genèse historique de la notion phonétique de rythme 122
- 5.3 Perspectives multilingues : « stress-timing » et « syllable-timing » . . . 124
- 5.4 Analyse du rythme à partir de ABI 136
- 5.5 Note sur le débit de parole 156
- 5.6 Synthèse des résultats et discussion 157

6**Les voyelles**

- 6.1 Méthodes de représentation 163
- 6.2 Analyse des voyelles 170
- 6.3 Description acoustique des voyelles 181
 - 6.3.1 L'anglais standard (*sse*) 182
 - 6.3.1.1 Les monophthongues 182
 - 6.3.1.2 Les diphtongues 185
 - 6.3.1.3 Analyse des voyelles d'un locuteur type de *sse* 188
 - 6.3.2 Birmingham (*brm*) 206
 - 6.3.2.1 Les monophthongues 206
 - 6.3.2.2 Les diphtongues 209
 - 6.3.3 Cornwall (*crn*) 215
 - 6.3.3.1 Les monophthongues 215
 - 6.3.3.2 Les diphtongues 218
 - 6.3.4 East Anglia (*ean*) 221
 - 6.3.4.1 Les monophthongues 221
 - 6.3.4.2 Les diphtongues 226
 - 6.3.5 East Yorkshire (*eyk*) 228
 - 6.3.5.1 Les monophthongues 228

6.3.5.2	Les diphtongues	231
6.3.6	Glasgow (<i>gla</i>)	235
6.3.6.1	Les monophtongues	235
6.3.6.2	Les diphtongues	238
6.3.7	Lancashire (<i>lan</i>)	242
6.3.7.1	Les monophtongues	242
6.3.7.2	Les diphtongues	244
6.3.8	Liverpool (<i>lvp</i>)	246
6.3.8.1	Les monophtongues	246
6.3.8.2	Les diphtongues	251
6.3.9	Newcastle (<i>ncl</i>)	254
6.3.9.1	Les monophtongues	254
6.3.9.2	Les diphtongues	259
6.3.10	North Wales (<i>nwa</i>)	263
6.3.10.1	Les monophtongues	263
6.3.10.2	Les diphtongues	266
6.3.11	Republic of Ireland (<i>roi</i>)	268
6.3.11.1	Les monophtongues	268
6.3.11.2	Les diphtongues	271
6.3.12	Scottish Highlands (<i>shl</i>)	273
6.3.12.1	Les monophtongues	273
6.3.12.2	Les diphtongues	277
6.3.13	Ulster (<i>uls</i>)	279
6.3.13.1	Les monophtongues	279
6.3.13.2	Les diphtongues	283
6.4	Synthèse des résultats et discussion	285
6.4.1	Monophtongues	285
6.4.2	Diphtongues	303

7

Classification automatique et représentations multidimensionnelles

7.1	Extraction des paramètres et segmentation	310
7.2	Méthode des matrices de distances (ACCDIST)	311
7.3	Portée descriptive et diagnostic	321
7.4	Représentations arborées	323

7.5	Détermination des voyelles discriminantes	328
7.6	Vers une interprétation dans le cadre de la théorie des ensembles flous .	338
7.7	Synthèse des résultats et perspectives	347

8

Conclusion générale

A

Annexe : passage lu du corpus ABI
--

B

Annexe : rythme et voisement

Index	377
Bibliographie	381

Première partie

Cadre de l'étude

Chapitre 1

Introduction

L'étude du langage est un domaine particulièrement enclin aux changements de paradigme, à la prolifération des théories et au cloisonnement – parfois arbitraire – d'une multitude de sous-disciplines. Deux axes d'évolution épistémologique nous semblent caractériser la linguistique, au sens large. D'une part, on a pu assister au basculement d'une quête de l'invariant vers une analyse raisonnée de la variation, ce qui a permis, entre autres, le développement des études dites « variationnistes ». D'autre part, la linguistique est, à notre sens, parvenue à un stade de maturation avancé qui autorise la prise en compte de l'articulation entre constructions théoriques et phénomènes empiriques. Ainsi, la phonologie et la phonétique, disciplines naguère distinctes, présentent aujourd'hui une intersection – parfois étiquetée « phonologie de laboratoire » – qui constitue un paradigme particulièrement fécond. C'est dans ce type d'approche que s'inscrit notre travail de thèse, qui examine la structure des systèmes de sons individuels par le biais de la substance acoustique, pour ensuite regrouper les individus sur des critères de proximité phonétophonologique. Cette approche est donc, par nature, pluridisciplinaire, puisqu'elle allie la phonétique acoustique, la phonologie, la dialectologie et d'autres domaines de la linguistique dans le but de mieux cerner la complexité de la notion de dialecte sur les Îles Britanniques.

L'étude des dialectes¹ présente des enjeux de société singulièrement forts dans le contexte des Îles Britanniques. Les particularismes phonético-phonologiques ruraux et urbains, fréquemment érigés en stéréotypes dans les médias et les arts, reflètent l'origine géographique ou sociale d'un locuteur. C'est donc sans surprise que la dialectologie a une incidence dans des domaines aussi variés que la sociologie, la psychologie, l'ingénierie de la langue et la publicité. Au-delà de l'image vieillissante du Professeur Higgins, ce phonéticien de fiction né sous la plume de G.B. Shaw, qui s'impose la lourde tâche de faire passer une marchande de fleurs pour une duchesse en corrigeant sa prononciation, nous livrons, en guise de préliminaires, quelques exemples de l'implication directe de la problématique des dialectes dans la société.

D'après un sondage récent² mené auprès de 5000 personnes, l'acteur Sean Connery aurait l'accent le plus plaisant du Royaume-Uni. La reine, quant à elle, apparaît aussi bien dans les 10 accents préférés que dans les 10 accents les moins plaisants. Les accents les moins appréciés sont ceux de Birmingham, Liverpool et Glasgow. Notons encore que 59% des personnes interrogées souhaiteraient changer d'accent. Enfin, concernant l'accent gallois, les personnes interrogées en dehors du Pays de Galles préfèrent entendre l'anglais parlé par un Australien ou même par un Français ! La stigmatisation de certains dialectes peut avoir de lourdes conséquences. Par exemple, un article du *Irish Post* du 17 mars 2001³ rapporte la mésaventure d'une jeune femme de Liverpool qui se voit refuser un emploi à Dublin en raison de son accent. Pire encore, dans une étude de Dixon et collègues ([DMC02]), un interrogatoire de police, dans lequel le suspect plaide son innocence, est présenté à des auditeurs. Cet interrogatoire, simulé, a été enregistré en deux versions ; une où le suspect adopte un accent de type RP⁴, une autre où il prend l'accent de Birmingham. Les auditeurs ont pour tâche d'attribuer un score de culpabilité à ce suspect.

¹Nous emploierons ce terme pour désigner l'aspect phonétique de la prononciation d'une communauté, qu'elle soit régionale ou sociale ; il s'agira donc sans équivoque d'accents. Nous reviendrons *infra* (Section 2.1) sur le terme dialectologie.

²<http://news.bbc.co.uk/1/hi/uk/4180373.stm> ; site consulté le 2 avril 2008.

³Disponible à l'adresse <http://www.phon.ucl.ac.uk/home/estuary/> ; site consulté le 10 avril 2008.

⁴« Received Pronunciation » ; il s'agit de l'accent britannique de prestige. Un exposé détaillé est présenté dans la Section 3.1.

Sans surprise, le locuteur « Brummie » est considéré comme plus coupable. Pour prendre un autre exemple, réel cette fois-ci, la police britannique, qui enquête sur une série de meurtres perpétrés par le Yorkshire Ripper à la fin des années 1970, reçoit l'enregistrement d'un homme se désignant comme l'assassin ([Ell94, Bil03]). Détail symptomatique de l'importance des dialectes sur le sol britannique, c'est à Stanley Ellis, un dialectologue renommé, que la cassette est confiée afin qu'il détermine l'origine géographique du locuteur. L'expert conclut que le locuteur est originaire de la région de Sunderland, une ville industrielle côtière du nord-est de l'Angleterre. L'enregistrement était en réalité un canular dont l'auteur n'a pas pu être identifié jusqu'en 2005 : l'ADN a en effet récemment permis de confondre celui que la presse surnomme Wearside Jack ⁵, un certain John Humble, originaire, comme l'avait justement envisagé Ellis, de Sunderland !

En 1963 sort le premier album des Beatles, intitulé *Please Please Me* : la proportion de /r/ non pré-vocaliques effectivement réalisés atteint les 47 % ([Tru83]). En 1969 paraît le dernier album des Liverpooliens, *Abbey Road* ⁶. Dans ce dernier, seulement 3 % de ces /r/ sont chantés. Dès l'album *Sergeant Pepper* en 1967, les mots <can't> et <half> ne sont plus exclusivement prononcés avec la voyelle /æ/ : un /a:/ plus britannique vient parfois se substituer au /æ/, rangé par Trudgill ([Tru83], 152), à l'instar de la réalisation de /r/ non pré-vocaliques, dans la catégorie des « quasi-American features ». Un début d'explication réside dans le fait qu'avant 1964, les États-Unis exerçaient une domination culturelle presque exclusive dans l'industrie du disque par le biais du rock'n'roll ; et quiconque produisait ce genre de musique se devait d'imiter le modèle américain, y compris l'accent. Après cette date, la « pop » anglaise acquiert ses lettres de noblesse et devient une culture à elle seule, ce qui permet à ses représentants de s'affranchir de la nécessité d'employer un accent américain pour vendre des disques. Quelques années plus tard apparaît le mouvement punk. La convention est alors d'adopter un accent britannique associé au bas de l'échelle sociale. Les coups de glotte se substituent à /t/ en position finale et intervocalique, les /l/ sont vocalisés, les diphtongues sont plus marquées : [æɪ] et [æʊ]

⁵http://news.bbc.co.uk/1/hi/england/west_yorkshire/4825820.stm ; site consulté le 2 avril 2008.

⁶L'album *Let It Be* est publié en 1970, mais il a été enregistré avant *Abbey Road*.

pour <face> et <go>, respectivement ([Tru83]). L'accent, le dialecte, comme vecteur de l'identité culturelle, constitue donc un procédé publicitaire de choix.

La thématique des dialectes revêt donc un rôle crucial dans les Îles Britanniques. Notre objectif, dans ce travail, est de décrire les aspects phonétiques de 13 dialectes et de mettre au point une procédure de classification automatique des locuteurs en fonction de leur dialecte. À travers l'analyse acoustique d'enregistrements, nous explorons la prononciation de 264 anglophones dans le but de déterminer les particularités les plus diagnostiques de chaque dialecte. Si l'aspect descriptif est justifié par les points que nous venons d'énumérer, quel pourrait être l'utilité d'un système de classification automatique d'un individu selon son dialecte à partir d'un enregistrement de sa voix ? Afin de répondre à cette question, nous présentons brièvement, entre autres applications potentielles, l'intérêt de la classification du dialecte pour la reconnaissance automatique de la parole et l'indexation des voix de comédiens.

Les systèmes de reconnaissance automatique de la parole sont moins performants lorsque le dialecte de l'échantillon sur lequel s'effectue la reconnaissance ne correspond pas au dialecte d'apprentissage (i.e. le modèle). Par exemple, Yan et Vaseghi ([YV02]) ont observé, à partir d'anglais américain et britannique, une augmentation d'environ 140 à 230 % (selon la condition) du taux d'erreur lorsque le dialecte de l'échantillon et celui du modèle étaient différents.

Le site Internet <http://www.sueterryvoices.com/>⁷ est une base de données regroupant des comédiens anglophones. Immédiatement après le sexe, le deuxième critère de recherche possible est l'accent. Certes, la base est de taille modeste, mais on peut néanmoins imaginer un système d'indexation automatique des locuteurs en fonction de leur accent qui permettrait de disposer d'une base d'enregistrements de voix de comédiens bien utile dans l'industrie du spectacle et de la publicité.

Les études actuelles qui abordent les dialectes des Îles Britanniques sous l'angle de la phonétique acoustique se bornent souvent à la variation induite par des facteurs socio-

⁷Site consulté le 2 avril 2008.

économiques dans un seul et même centre urbain. Rassemblées dans des ouvrages collectifs (e.g. [FD99, SBK⁺04], etc.), ces études offrent donc en apparence l'équivalent de ce que nous proposons dans notre thèse. En réalité, si les chapitres dans [FD99] présentent l'avantage d'une description fine du lien entre structure sociale et phonétique (description rendue de fait impossible par la nature de nos données, voir Section 4.1), les méthodes et les échantillons de parole employés par les différents auteurs varient. Notre objectif est, au contraire, d'analyser la lecture d'un texte identique pour tous les locuteurs par le biais d'une méthodologie unique et reproductible (car automatique).

Dans le reste de la Partie I, nous introduisons des concepts fondamentaux : certains ont une utilité directe pour nos travaux (Sections 2.3.2 et 2.4), d'autres sont mentionnés parce qu'ils reflètent, parfois de manière indirecte, la vision épistémologique qui a accompagné nos recherches (e.g. Section 2.3.3). La Section 2.1 examine la diversité des études regroupées sous l'appellation générique de « dialectologie » et traite certains aspects de l'évolution des systèmes vocaliques. Dans la Section 2.3, les caractéristiques phonétophonologiques des voyelles sont étudiées sous l'angle de la variation acoustique, qui est ensuite éclairée par le concept d'information. Cette discussion, mise en relation avec les mots-test de notre corpus, conduit à examiner la manière dont la variation acoustique des voyelles s'inscrit dans une théorie du prototype. Enfin, la Section 2.3 s'achève sur un bref examen de ce qu'est une catégorie dialectale. Dans la Section 2.4, nous décrivons brièvement les méthodes d'analyse de données utilisées dans nos travaux. Le Chapitre 3 passe en revue les traits phonétophonologiques des variétés de l'anglais britannique. On y aborde, dans un premier temps (Section 3.1), la notion d'anglais standard dans le contexte des Îles Britanniques, accompagnée d'une revue succincte des études récentes sur le système vocalique de cette variété et sur les changements phonétiques en cours. Dans un second temps (Section 3.2), nous établissons l'inventaire des particularités phonétophonologiques les plus marquantes des aires dialectales des Îles Britanniques. L'analyse acoustique de notre corpus est détaillée dans la Partie II, qui aborde dans un premier temps la notion de rythme (Chapitre 5) pour ensuite se concentrer sur la description des

systemes vocaliques (Chapitre 6). Enfin, le Chapitre 7 est consacré à la classification automatique des dialectes et aux représentations des voyelles et des locuteurs dans un espace acoustique multidimensionnel.

Chapitre 2

Orientation théorique et méthodologique, et notions élémentaires

2.1 Cadre dialectologique

La dialectologie, au sens large, rassemble des études dont les méthodes et les objectifs sont très variés. Pour notre propos, il est possible de distinguer trois paradigmes perméables les uns aux autres :

1. la dialectologie traditionnelle, dont l'objet d'étude est principalement la dimension aréale des variations et dont les méthodes se limitent souvent à l'analyse auditive ou au questionnaire. Ce domaine s'inscrit notamment dans une optique de conservation du patrimoine ;
2. la sociolinguistique, qui étudie la variation en fonction de facteurs socio-économiques. Cette discipline constitue le paradigme dominant de nos jours, et fait appel à des techniques d'instrumentation modernes comme le spectrogramme ;
3. la dialectométrie, terme sous lequel nous regroupons les approches comportant des techniques d'analyses de données élaborées et la classification automatique.

Nous avons choisi quelques exemples d'études qui nous paraissent représentatives pour illustrer ces trois paradigmes. Le *Linguistic Atlas of England* ([OSW78]) constitue l'aboutissement d'un gigantesque projet réalisé entre 1950 et 1961 : le *Survey of English Dialects*. Durant cette période, onze enquêteurs ont sillonné l'Angleterre avec un questionnaire visant à mettre en évidence les caractéristiques des parlers de pas moins de 313 localités. Le but avoué de ce projet n'était pas d'enregistrer les variations dialectales représentatives de l'époque mais bien de garder une trace des formes de parlers traditionnels les plus conservateurs car on avait entrevu que la mobilité géographique et sociale engendrée par la guerre, accompagnée d'un effet de nivellement dû à la radio et à la télévision, allait conduire à l'extinction des particularités archaisantes de certains dialectes. Les critères retenus pour la population cible traduisent tout à fait cette volonté d'effectuer des relevés sur des traits ou des variétés moribonds : les informateurs devaient être prioritairement des hommes puisque les femmes semblent être plus disposées à assimiler les innovations linguistiques, et affichent une propension à imiter les variétés de prestige (voir [OSW78], Introduction). Ces hommes devaient être « over 60 years of age, with good mouths, teeth and hearing, and of the class of agricultural workers » ([OSW78], Introduction). En outre, dans la phase d'interprétation précédant la conception des cartes de l'atlas, Orton, dans les cas où il était confronté à des données contradictoires pour une même aire, semble avoir souvent tranché arbitrairement en faveur des formes les plus anciennes ou de celles qui déviaient le plus de l'anglais standard, ce qui n'aura pas manqué de mettre en avant des particularités qui n'étaient en réalité qu'anecdotiques.

Les accents et dialectes, notamment traditionnels, se prêtent particulièrement bien à une approche populaire – « folk linguistics », nous entendons – si bien que la collecte des données se fait souvent dans un double objectif d'étude linguistique et de divertissement familial. Cette remarque s'applique tout à fait à « The VOICES survey », projet de BBC Radio 4, qui peut être considéré comme une prolongement du *Survey of English Dialects*, et dont est tiré l'ouvrage de Elmes ([Elm05]). L'aspect scientifique de cette forme de dialectologie est assuré par des universitaires de renom, tels Stanley Ellis, l'un des investigateurs

originaux du *Survey of English Dialects*, et Clive Upton, l'un des directeurs du *Handbook of English Varieties* ([SBK⁺04]). Malgré cela, il apparaît comme évident que la finalité de cette enquête ressortit davantage à la préservation du patrimoine qu'à l'étude scientifique *stricto sensu* : dans les enregistrements, on retrouve, par exemple, trois coiffeuses de Truro (Cornouailles) qui livrent leurs impressions sur l'accent local, des fans de rugby à Leeds (Yorkshire) disséquant certains aspects du lexique local ; même exercice pour le pêcheur de l'île de Shetland, mais, cette fois-ci, la compréhension est ardue pour l'oreille novice. D'ailleurs, Elmes ([Elm05]) ne propose pas de transcription phonétique mais bien plutôt une tentative de rendre les particularités dialectales en orthographe standard ; le terme « eye dialect » désigne parfois ce genre d'exercice. La description des sons est d'ailleurs très impressionniste, voire fantaisiste : « pinched flat vowels » *vs* « full-throated [vowels] » ([Elm05], 32). Les enregistrements, environ 250 interviews, ont été effectués entre fin 2004 et début 2005 par une cinquantaine d'enquêteurs, et ils couvrent tout le Royaume-Uni. Le point commun entre les interviews (qui durent entre 45 minutes et 2 heures) réside dans la présentation d'un questionnaire, sur le modèle des questionnaires utilisés pour le *Survey of English Dialects*, avec des questions du type : « What do you call the toilet ? »

Voilà, donc, le type d'étude typique que nous classons dans le paradigme de la dialectologie traditionnelle. L'ouvrage de Wells ([Wel82]), quoique en partie inspiré de cette forme de dialectologie, témoigne de l'influence du tournant sociolinguistique des années 1960-70. Kerswill ([Ker03]) voit ce tournant comme « a move from the rural to the urban ». Les études actuelles s'inscrivent, pour la plupart, dans ce cadre (voir [LP97] pour un aperçu historique de cette discipline). Nous en voulons pour preuve la nature des travaux réunis dans l'ouvrage de Foulkes et Docherty ([FD99]), qui ont pour dénominateur commun la stratification sociale systématique des participants, la quantification des réalisations de ces participants (souvent limitée à des pourcentages) et, parfois, l'utilisation d'outils modernes de mesure et de visualisation du signal.

Enfin, le troisième paradigme mentionné, la dialectométrie, regroupe les études impliquant un traitement de données multidimensionnelles généralement plus élaboré, sou-

vent dans une optique de classification automatique. Ce genre d'étude regroupe, entre autres, les travaux de Barry et collègues ([BHN89]) et Huckvale ([Huc04, Huc07a, Huc07b]) pour la classification automatique dans les Îles Britanniques, Hansen et collègues ([AH02, HYHI04, AH03]) et Berkling et collègues ([BZVC98]), pour la classification automatique de dialectes anglais et américains (incluant des accents étrangers, voir aussi [BVK93]) et les travaux de Heeringa et collègues (e.g. [HN01, HG03]), ten Bosch ([tB00]), Wilhelm et Sander ([WS98]), notamment sur l'idée de distance phonétique entre dialectes.

Notre étude, contrainte dans ses objectifs par les données dont nous disposons, s'inscrit dans la lignée des travaux de dialectologie traditionnelle britannique. Certes, nous avons consulté les références provenant de la sociolinguistique variationniste, et avons rapporté les résultats pertinents émanant de ce domaine, mais la nature de nos données nous a cantonné à l'étude de la variation géographique. Si nous avons précisé « traditionnelle britannique », c'est que des habitudes codifiées distinguent les spécialistes britanniques des autres. Et s'il est un ouvrage dans lequel (la plupart de) ces conventions sont enregistrées, c'est indéniablement le *Accents of English* de John Wells ([Wel82]). Nous donnerons comme exemples, pour ne relever que ces deux-là, la terminologie et l'utilisation de la notion d'ensemble lexical. Wells ([Wel82], *passim*), même s'il n'en est pas nécessairement l'inventeur, a contribué à normaliser la dénomination de nombre de concepts, et les références actuelles en dialectologie des Îles Britanniques ([FD99], [SBK⁺04]) font un usage quasi-exclusif de ces termes. Concernant les ensembles lexicaux, il s'agit, à notre avis, d'une avancée méthodologique majeure qui constitue l'unique système consensuel de mots-clés dans la tradition britannique. Le principe est détaillé dans [Wel82], *passim*. Nous allons définir ce concept et tenter de prouver son utilité.

Les mots-clés désignant les ensembles lexicaux sont un moyen commode de faire référence à un groupe de mots qui ont tendance à posséder la même voyelle quel que soit l'accent, son timbre variant bien évidemment d'un dialecte à l'autre. Ils constituent ainsi une norme de description panlectale particulièrement appropriée. En règle générale, on a recours à un plus grand nombre d'ensembles lexicaux que de phonèmes pour décrire

TAB. 2.1 – Les 24 ensembles lexicaux standards, d’après Wells ([Wel82], xviii)

Mot-clé	Exemples
KIT	ship, sick, bridge, milk, myth, busy...
DRESS	step, neck, edge, shelf, friend, ready...
TRAP	tap, back, badge, scalp, hand, cancel...
LOT	stop, sock, dodge, romp, possible, quality...
STRUT	cup, suck, budge, pulse, trunk, blood...
FOOT	put, bush, full, good, look, wolf...
BATH	staff, brass, ask, dance, sample, calf...
CLOTH	cough, broth, cross, long, Boston...
NURSE	hurt, lurk, urge, burst, jerk, term...
FLEECE	creep, speak, leave, feel, key, people...
FACE	tape, cake, raid, veil, steak, day...
PALM	psalm, father, bra, spa, lager...
THOUGHT	taught, sauce, hawk, jaw, broad...
GOAT	soap, joke, home, know, so, roll...
GOOSE	loop, shoot, tomb, mute, huge, view...
PRICE	ripe, write, arrive, high, try, buy...
CHOICE	adroit, noise, join, toy, royal...
MOUTH	out, house, loud, count, crowd, cow...
NEAR	beer, sincere, fear, beard, serum...
SQUARE	care, fair, pear, where, scarce, vary...
START	far, sharp, bark, carve, farm, heart...
NORTH	for, war, short, scorch, born, warm...
FORCE	four, wore, sport, porch, borne, story...
CURE	poor, tourist, pure, plural, jury...

un dialecte car outre la diversité phonologique, la variation allophonique et l’évolution diachronique des classes de mots constituent des objets d’étude que la dialectologie ne néglige pas⁸.

Les ensembles lexicaux standards donnés dans la Table 2.1 ont été enrichis depuis [Wel82] au gré des besoins de chaque dialecte. Cet enrichissement permet de mieux cerner le fait que ces ensembles lexicaux ne procèdent pas de la même démarche qu’une

⁸Chevillet ([Che91], 32) rapporte que Gilliéron, un des pionniers de la tradition des atlas linguistiques, avait fait sienne la phrase de Grimm : « jedes Wort hat seine Geschichte ».

analyse phonologique parcimonieuse. Bien qu'ils ne soient pas mentionnés dans l'énumération dont s'inspire la Table 2.1, Wells ([Wel82], 165 *sqq.*) utilise également les ensembles *happy*, *letter* et *comma*. Or un éventuel statut phonologique de la voyelle finale de *happy* reste à démontrer (voir néanmoins la discussion détaillée dans la Section 3.2.1.5). Cependant, l'utilité d'un tel ensemble est incontestable puisque ce dernier permet d'exprimer avec concision le comportement de la voyelle inaccentuée finale d'un bon nombre des mots en <-y>, <-i>, <-ie>, <-ey>, etc. Pour prendre un autre exemple, l'ouvrage de référence de Foulkes et Docherty ([FD99]) contient un mot-clé supplémentaire : *horses*⁹. Phonologiquement, il s'agit d'une voyelle inaccentuée qui selon le dialecte aura la voyelle de KIT ou un schwa. Mais l'utilisation d'un nouvel ensemble lexical permet, au-delà d'une description strictement phonologique, de faire intervenir explicitement les critères morphologiques qui régissent le choix de l'une des deux voyelles citées dans les dialectes n'ayant pas encore subi ce que Wells ([Wel82], 85) nomme « Weak Vowel Merger ». Ce phénomène traduit le fait que dans certains dialectes, l'opposition /ɪ/-/ə/ permettant de distinguer traditionnellement les mots <chatted> et <chattered>, et <offices> et <officers>¹⁰, a été neutralisée (ce thème sera abordé dans les Sections 3.1.1 et 3.2.1.4).

En ce qui concerne la méthodologie utilisée dans le traitement des données, notre approche s'apparente davantage aux méthodes employées dans le domaine que nous avons désigné par le terme générique de « dialectométrie ». En effet, les deux principaux objectifs de notre étude sont de classer automatiquement les locuteurs en fonction de leur dialecte et de décrire ces dialectes via des techniques de traitement de données multidimensionnelles (voir notamment le Chapitre 7). À notre connaissance, certaines méthodes utilisées ici l'ont rarement (voire jamais) été en dialectologie traditionnelle ou en sociolinguistique. Un survol sommaire de ces outils (voir *infra*, Section 2.4) permettra de mieux comprendre le reste de notre étude.

⁹Nous avons rétabli la convention typographique de Wells, qui consiste à noter la voyelle qui nous intéresse avec une petite majuscule et le reste en italiques.

¹⁰Ceci s'applique aux accents non rhotiques.

Une description des mécanismes d'évolution des systèmes vocaliques permettra d'éclairer certains aspects de la variation dialectale actuelle ; c'est l'objectif de la section suivante.

2.2 Variation dialectale : évolution des systèmes vocaliques et cas limites

Dans cette section, nous examinons certains facteurs expliquant les mécanismes responsables de la variation observée de nos jours dans les dialectes de l'anglais des Îles Britanniques avec une attention particulière à l'évolution des systèmes vocaliques. Nous passons en revue les trois phénomènes qui affectent l'organisation des systèmes vocaliques – changements en chaîne, convergences phonémiques et scissions phonémiques – avant d'aborder le cas limite des quasi-convergences (« near mergers »).

2.2.1 Les changements en chaîne

Il est nécessaire d'aborder ici la notion de changements en chaîne (« chain shifts ») car elle renforce l'idée que les inventaires vocaliques sont des systèmes (chaque voyelle n'est pas indépendante), et elle recouvre des principes qui nous permettront de mieux comprendre la variation actuelle dans les dialectes de l'anglais des Îles Britanniques. Nous nous appuyerons sur les ouvrages de Martinet ([Mar05]) et Labov ([Lab94]). L'idée de changements en chaîne se base sur les observations empiriques de la phonologie diachronique dans plusieurs langues. L'anglais a connu un tel phénomène au cours de son histoire, le Grand Changement Vocalique (GCV), et ce précédent éclaire les structures phonologiques actuelles. Labov ([Lab94], 116 *sqq.*) reconnaît trois principes généraux relatifs aux changements en chaîne :

- Principe I : les voyelles longues se ferment ;
- Principe II : les voyelles brèves s'ouvrent (non attesté dans le GCV puisque ce dernier ne concerne que les longues) ;

- Principe IIA : le noyau des diphtongues fermantes s’ouvre ;
- Principe III : les voyelles postérieures s’antériorisent (non attesté dans le GCV)¹¹.

Un changement en chaîne élémentaire se produit lorsqu’un phonème A quitte sa position initiale dans le trapèze vocalique, et qu’un phonème B vient occuper cette position. Ceci ne nous dit rien sur le sens de la causalité ; il se peut que ce soit A qui ait amorcé le mouvement, auquel cas on parlera de chaîne de traction ; si c’est B, on parlera de chaîne de propulsion. Lorsque choisir l’une de ces deux options a un sens, la décision sera guidée par les principes énoncés plus haut ainsi que par des critères sociolinguistiques.

Il apparaît également que l’évolution des systèmes vocaliques préserve une certaine symétrie entre les voyelles antérieures et postérieures. Ceci a été vrai diachroniquement : les voyelles de PRICE et MOUTH étaient les monophthongues longues les plus fermés du système moyen-anglais ; le processus de fermeture des longues (ici « remplacé » par une diphtongaison à cause de l’impossibilité d’atteindre un degré de fermeture supérieur) suivi de l’ouverture des noyaux des diphtongues fermantes a affecté ces voyelles de la même manière. En synchronie, il est intéressant de noter une symétrie sur le plan de la réalisation pour des voyelles d’aperture égale ; par exemple, dans ce que nous appelons plus loin « anglais standard », les deux voyelles actuelles les plus fermées (FLEECE et GOOSE) sont légèrement diphtonguées et pourraient, incidemment, emprunter le même chemin que PRICE et MOUTH cinq siècles plus tôt.

2.2.2 Les convergences phonémiques

Les convergences phonémiques (« phonemic mergers ») s’opposent au phénomène de changements en chaîne en cela que, schématiquement, le mouvement d’un phonème A vers le phonème B ne provoque pas un mouvement de ce dernier, ce qui conduit A et B à être confondus au final (le cas décrit est en réalité un des trois mécanismes reconnus dans les convergences). Labov précise que les convergences phonémiques sont bien plus courantes

¹¹Mais, nous le verrons plus loin, ce phénomène est attesté, notamment en anglais standard, pour les ensembles FOOT et GOOSE.

dans l'histoire des langues que les changements en chaîne ([Lab94], 310).

D'après Labov ([Lab94], 310-348), les convergences phonémiques obéissent à deux règles : le principe de Garde¹² et le principe de Herzog¹³. Le premier pose que les convergences phonémiques sont irréversibles ou, comme le formule Labov ([Lab94], 311) : « Mergers are irreversible by linguistic means ». Le second complète le précédent : l'expansion des convergences se fait au détriment des distinctions ([Lab94], 311 : « Mergers expand at the expense of distinctions »).

La justification du principe de Garde est très intuitive : une fois que deux classes de mots, deux ensembles lexicaux, ont fusionné dans une langue ou un dialecte, le locuteur n'a dans sa phonologie plus qu'une seule catégorie et ne peut normalement pas reconstituer les deux classes historiques. C'est par exemple le cas chez les locuteurs d'anglais d'Écosse qui ne présentent pas la distinction FOOT-GOOSE. Si l'un d'entre eux souhaite, mû par une quelconque pression sociolinguistique, réapprendre ces deux catégories qui continuent d'être distinctes dans la majorité des dialectes des Îles Britanniques, il devra réapprendre la prononciation de chaque item. L'orthographe n'étant pas un guide fiable pour ces deux classes, il courra en outre le risque de l'hypercorrection¹⁴. Labov [Lab94], 312, ajoute que ce principe bénéficie d'une validité empirique certaine :

at no known time in the history of languages has such a reversal been accomplished by enough individual speakers to restore two original word classes...

Il existe cependant des exemples de cas limites où les deux classes de mots apparemment impliquées dans une convergence ont été reconstituées. Labov ([Lab94], 306-309) et Wells ([Wel82], 208-210) expliquent qu'une convergence (au moins partielle) entre les ensembles PRICE et CHOICE¹⁵ est attestée dès le 15^e siècle. Ce phénomène semble avoir atteint son paroxysme au 17^e siècle. À la fin du 18^e, il s'agissait d'un trait hautement stigmatisé. De nos jours, la plupart des accents ont rétabli la distinction. Wells ([Wel82],

¹²Du nom d'un universitaire slavisant français, Paul Garde.

¹³Marvin Herzog est un collègue de Labov, spécialiste du yiddish.

¹⁴Par exemple, la forme graphique ne permet pas de prédire qu'en anglais standard <wood> ne rime pas avec <mood> ; statistiquement, /u:/ est plus courant pour le digraphe <oo> ([Jon03], 381).

¹⁵Notre formulation suppose un certain raccourci : ces ensembles lexicaux sont ceux de l'anglais actuel ; or la convergence que nous rapportons ne semble concerner qu'un sous-ensemble historique de CHOICE.

209-210) attribue ceci à l'influence de RP (sans plus de précision) et à la pression exercée par l'orthographe. Ces deux facteurs sont probablement de deux types bien différents ; si l'orthographe constitue très vraisemblablement une condition nécessaire pour le rétablissement d'une distinction perdue après convergence, l'hypothèse de pressions sociolinguistiques causant le processus paraît la moins extravagante. Et si des pressions sociales ont pu initier ce processus, c'est qu'il n'était pas achevé (voire pas initié) dans les variétés qui ont fini par imposer leur norme, d'où une certaine réserve quant à l'ampleur géographique et sociale de cette convergence, et son degré de complétude. En étudiant le phénomène de plus près et en suivant l'un de ses principes méthodologiques récurrents (étudier le présent pour comprendre le passé), Labov ([Lab94], 371-384) observe la convergence PRICE-CHOICE dans l'Essex. Il conclut que ces deux voyelles n'ont probablement pas connu une totale convergence au 18^e, mais une quasi-convergence (« near merger », phénomène déconcertant pour le linguiste structuraliste ; voir ci-dessous 2.2.4). En tout état de cause, pour Labov, s'il n'y a pas eu convergence, c'est donc qu'on ne peut pas véritablement parler de restauration de deux classes après convergence ; le principe de Garde reste donc intact.

D'après Labov ([Lab94], 321 *sqq.*), il existe trois mécanismes distincts pouvant aboutir à une convergence phonémique :

- la convergence par approximation ; les réalisations de deux phonèmes se rapprochent au point de finir par se confondre. Le phonème unique résultant de ce processus présente soit une qualité phonétique intermédiaire entre les deux timbres originaux, soit une réalisation correspondant à celle de l'un des deux phonèmes de départ. Ce genre de convergence peut se mettre en place en trois ou quatre générations ;
- la convergence par transfert ; les mots d'une classe rejoignent progressivement une autre classe. C'est le processus le plus long ;
- la convergence par expansion ; les réalisations de deux phonèmes se dilatent dans l'espace vocalique au point de finir par se chevaucher. L'espace phonétique occupé par les réalisations du nouveau phonème englobe les deux espaces précédemment

occupés par les réalisations des deux phonèmes. C'est le type de convergence le plus rapide ; il peut s'effectuer en une seule génération.

2.2.3 Les scissions phonémiques

Les scissions phonémiques (« phonemic splits ») sont l'inverse des convergences : la naissance d'une nouvelle distinction phonémique dans le système due à la scission d'une seule catégorie phonémique préexistante. Abordons dans un premier temps l'influence des emprunts pour ensuite détailler le cas où une classe de mots en vient à être scindée sans règle apparente.

Les emprunts à d'autres langues peuvent enrichir un inventaire phonémique de deux manières différentes : soit en apportant un nouveau phonème, comme cela semble être le cas pour les mots de l'ensemble CHOICE ([Wel82], 150), soit en permettant la phonologisation de ce qui n'était qu'une variante allophonique dans la langue qui emprunte. Ce dernier point est illustré par la phonologisation de l'opposition de voisement pour certaines fricatives en moyen-anglais. Hormis dans les régions du sud-ouest de l'Angleterre, le vieil-anglais ne possède pas de fricatives voisées à l'initiale d'un mot ; en revanche, celles-ci se sont voisées dans cette période lorsqu'elles apparaissaient entre deux segments voisés, d'où les doublets actuels : <wife>/<wives>¹⁶, <bath>/<bathe>, etc ([Sté97], 18 *sqq.*). Il s'agit donc ici de variation allophonique conditionnée par le contexte. En conjonction avec d'autres facteurs (dont le voisement dû à l'affaiblissement des fricatives à l'initiale de mots grammaticaux, e.g. <this>, <thy>, etc.), les emprunts au français ont permis la présence de fricatives voisées à l'initiale, et ont ainsi contribué à la formation de paires minimales (e.g. <fine>/<vine>).

Ce que Labov ([Lab94], 333 *sqq.*) nomme « lexical split » implique un processus de scission d'un ensemble lexical dont les raisons sont multiples et moins intuitives que l'influence des emprunts. L'exemple du « broad a », ce phénomène historique typiquement

¹⁶Le <-e> du singulier n'est qu'un diacritique conventionnel qui signale la longueur de la voyelle, et non une relique désinentielle qui aurait pu favoriser le voisement de la fricative ; en vieil-anglais : <wif> [wi:f]/<wifas> [wi:vas] ([Sté97], 18).

TAB. 2.2 – Illustration de la difficulté du choix de l’affectation d’un item à l’ensemble BATH.

TRAP : /æ/	BATH : /ɑː/
classic	class
gas	grass
ample	sample
land	demand
math	bath
plastic	plaster
cancer	answer

britannique qui consiste en l’allongement et la postériorisation de la voyelle dans les mots de l’ensemble BATH illustre la complexité des facteurs qui interviennent. Certains critères phonologiques (« broad a » notamment devant /f/, /θ/, /s/ et /n/ + obstruante sourde), ou morphémiques (<chancy> est un dérivé; le dérivant remplit une condition phonologique favorable au « broad a », alors que <fancy>, qui n’est pas un dérivé, a /æ/) permettent à qui voudrait apprendre cette distinction de connaître certaines tendances. Mais, comme le montre la Table 2.2 les irrégularités sont nombreuses, d’où, une fois de plus, le risque d’hypercorrection.

Il apparaît donc qu’acquérir ce type de distinction nécessite souvent l’apprentissage un à un des items susceptibles d’appartenir à la nouvelle classe. Pour prendre un autre exemple, les dialectes du nord de l’Angleterre, qui ont la même voyelle pour FOOT et STRUT jouissent de relativement peu de prestige (voir Section 3.2.1.1). Il est donc envisageable qu’un locuteur natif d’un de ces dialectes soit amené à apprendre cette distinction au contact prolongé de locuteurs méridionaux ou de locuteurs de classes supérieures. Or, si l’orthographe constitue une règle probabiliste assez fiable (la graphie <u> étant typique de STRUT), il risque cependant de produire certaines formes hypercorrectes s’il ne se fie qu’à celle-ci. Ainsi l’orthographe le conduira à inférer à raison que <but> appartient à STRUT, mais que <butcher> est dans le même cas à tort. S’il prononce <Hull> avec /ʌ/, il devra se garder de généraliser cette prononciation à <full>, <pull>, etc. Et s’il constate que les singularités orthographiques du genre de <love> ou <come>¹⁷ ont la voyelle de

¹⁷Historiquement, ces mots avaient comme voyelle un <u> dans la graphie; s’il est conservé dans la prononciation, il a en revanche été remplacé par un <o> dans l’écriture cursive pour faciliter sa lisibilité,

STRUT, il devra s'abstenir d'en déduire que les étrangetés orthographiques que sont la première voyelle de <woman> et de <bosom> appartiennent au même ensemble que <love> et <come>. À ce propos, Gerry Knowles, pourtant phonéticien (mais originaire du nord) écrivait en 1978 ([Kno78], 86) :

There are just a few words like *putty* and *nasty* which baffle me; I cannot recall whether RP has /pʊtɪ/ or /pʌtɪ/, /næstɪ/ or /nɑːstɪ/, and I have to look them up in a pronouncing dictionary.
(Soulignement ajouté)

2.2.4 Les quasi-convergences

Le phénomène des quasi-convergences (« near mergers ») est un cas délicat car il illustre les limites d'une phonologie structuraliste trop conservatrice.

Labov ([Lab94], 371-384) aborde plus en détail la convergence PRICE-CHOICE. Au début des années 1970, il se rend à Tillingham¹⁸ pour mener une étude sur trois informateurs âgés de 69 à 87 ans. Il parvient à éliciter les paires <voice>/<vice> et <loin>/<line>. À la question de savoir si les deux membres de chaque paire sont identiques ou différents, deux informateurs répondent qu'ils sont identiques, le troisième, différents. Sur le plan de la production, les trois informateurs ont des valeurs sur un plan F1/F2 qui, quoique assez semblables pour PRICE et CHOICE, font apparaître deux groupes plutôt distincts. En faisant écouter aux trois informateurs les hypothétiques paires minimales produites par l'un d'entre eux, aucun des trois ne parvient à distinguer les membres de chaque paire. La discussion qui suit entre l'investigateur et deux des sujets amène ces derniers à conclure qu'une différence existe, l'un d'entre eux tentant de l'exagérer (probablement inspiré par ce qu'il connaît des autres dialectes ou par la graphie), et l'autre en fournissant une imitation très approximative, le contraste étant surtout rendu par le contexte sémantique de ses commentaires. Les conclusions de cette étude sont appuyées par les résultats d'autres études résumées par Labov ([Lab94], 357-370).

et l'habitude a été conservé malgré l'avènement de l'imprimerie pour certains mots fréquents ([Sté97], 42).

¹⁸Petit village de l'Essex où les enquêteurs du *Survey of English Dialects* avaient diagnostiqué la persistance de la convergence PRICE-CHOICE.

La linguistique pose généralement une symétrie entre production et perception : quand il parle, le locuteur est aussi auditeur, et quand il écoute, l'auditeur est aussi locuteur ([Lab94], 352). Or le cas des quasi-convergences fait apparaître qu'il est possible pour un seul et même locuteur de produire des différences phonétiques de façon cohérente sans les percevoir. Se pourrait-il, donc, qu'une paire minimale continue d'exister sur le plan phonétique tout en ayant perdu sa valeur phonologique pour le locuteur qui la produit ? De plus, se pourrait-il que ce phénomène infraliminal ait une quelconque influence sur l'évolution d'un système ?

Le problème des quasi-convergences peut être mis en parallèle avec certains aspects de phonologie développementale, notamment les « covert contrasts ». On sait que le nourrisson est capable de distinguer très tôt des contrastes phonologiques dans une langue étrangère qui sont implémentés phonétiquement de manière différente de sa langue maternelle ([Jus96]). Il lui est en outre possible de distinguer des contrastes inexistant dans sa langue maternelle, comme c'est le cas de ces nourrissons canadiens (« anglophones ») âgés de 1 à 4 mois qui discriminent /pa/ et /pã/ en français dans une étude rapportée par [Jus96], 334. Le nourrisson est donc un incroyable phonéticien, et s'il est doué d'une perception avec une granularité aussi fine à ce stade immature de l'acquisition de sa phonologie, ceci lui confère probablement la capacité de produire de la variation phonétique (pertinente ou non) imperceptible pour un adulte. Cette remarque est corroborée par un certain nombre d'études, dont celle de Scobbie *et al.* ([SGHF00]). La plupart des études sur l'acquisition de la phonologie se basent sur des transcriptions phonétiques d'enregistrements. On sait par ailleurs que, dans ce domaine, la comparaison de transcriptions d'un même passage par plusieurs transcrip-teurs expérimentés montre des différences très importantes ([SGHF00], 194-196), d'où une fiabilité douteuse. Ces transcriptions font croire que beaucoup d'oppositions présentes chez l'adulte sont neutralisées chez l'enfant. Or les études citées par [SGHF00] prouvent que l'enfant dans la phase d'acquisition de la phonologie parvient à contraster en production de façon cohérente certains phonèmes de sa langue, le contraste étant imperceptible pour un adulte. Par exemple, Scobbie *et al.* ([SGHF00]) montrent

dans une étude sur l'acquisition du trait de voisement, que leur sujet produit des valeurs de VOT et de pentes spectrales différenciées selon que la consonne articulée est phonologiquement voisée ou non. Ces mesures parviennent, là où l'oreille adulte échoue, à prouver l'existence de catégories phonologiques bien établies, quoique opposées en surface par des indices acoustiques qu'on pourrait qualifier de subliminaux pour un auditeur adulte. Ces faits appellent un constat et soulèvent une question. D'abord, l'étude de l'acquisition de la phonologie sans l'appui de la phonétique expérimentale est une entreprise hasardeuse. Ensuite, le fait que l'enfant en phase d'acquisition de sa phonologie puisse produire et, peut-être, percevoir des contrastes que le système de l'adulte ne reconnaît pas, ou plus du tout, peut-il jouer un rôle dans l'évolution d'une langue ou d'un dialecte ?

Dans le même ordre d'idées, Scobbie ([Sco07]) étudie le phénomène de dérhoticisation chez deux adolescents écossais issus de la classe ouvrière. Ce phénomène fait référence à la disparition (sur le plan de la perception) du /r/ en position de coda. L'étude compare dans un premier temps un spectrogramme de <rain>, où le /r/ initial se traduit par un mouvement rapide vers le haut de F2 et F3 sur un peu plus de 100 ms, avec un spectrogramme de <car> où l'on observe, là où le /r/ est attendu, un mouvement extrêmement réduit des deux formants, avec une intensité bien moindre que celle de la voyelle, et sur une durée particulièrement faible ; en d'autres termes, rien sur le spectrogramme n'indique la présence d'une approximante alvéolaire ou post-alvéolaire. En utilisant une technique d'imagerie à ultrasons¹⁹ (« Ultrasound Tongue Imaging »), Scobbie met en évidence, à la fin de <car>, un mouvement de la lame de langue en direction de la région post-alvéolaire tout à fait semblable à celui que le locuteur produit à l'onset de <rain>. Son absence sur le plan de la perception semble provenir du fait que ce mouvement de la langue intervient plus tard que chez les locuteurs qui préservent une rhoticité perçue. Il apparaît donc que, chez ce locuteur, le /r/ imperceptible en coda soit néanmoins présent acoustiquement sous une forme vestigiale.

Ce rapide tour d'horizon des phénomènes « évolutionnistes » avait pour but de caracté-

¹⁹Proche, quant au principe et à l'image obtenue, d'une échographie classique.

tériser la variation systémique dans les dialectes. Dans la section suivante, nous abordons la question de la variation acoustique.

2.3 Cadre phonético-phonologique : variation, information et prototypes

L'orientation phonético-phonologique de nos travaux est en grande partie déterminée par la nature de nos données. En effet, la taille importante du corpus utilisé (voir Section 4.1) justifie une étude exclusivement acoustique ; les aspects de perception constituant un développement ultérieur potentiel. Le lien entre phonétique et phonologie ou, autrement dit, entre des paramètres physiques variant dans un espace continu et des catégories linguistiques, nécessite une approche qui, au moins dans un premier temps, s'éloigne de la représentation classique structuraliste qui prend la forme d'une matrice de traits binaires. Pour Chomsky et Halle, la projection de la substance phonétique (par nature gradiente) dans un espace phonologique binaire (ou inversement) n'a rien de contradictoire ([CH68], 297) :

In view of the fact that phonological features are classificatory devices, they are binary, [...] for the natural way of indicating whether or not an item belongs to a particular category is by means of binary features. This does not mean that the phonetic features into which the phonological features are mapped must also be binary²⁰.

Une revue des développements postérieurs à Chomsky et Halle sur l'idée de correspondance entre les représentations phonétique et phonologique est présentée dans les travaux de Chitoran et Cohn ([CCis]). Les auteurs abordent dans un premier temps la représentation modulaire classique, qui fait co-exister une phonologie strictement catégorielle avec une phonétique gradiente. Elles traitent ensuite l'approche modulaire alternative, qui plaide en faveur d'une phonétique discrète et d'une phonologie plus gradiente. Comme nous le verrons, notamment lorsque nous ferons allusion à la théorie quantique de Stevens (Section 2.3.3), il existe dans la substance phonétique – supposée varier de

²⁰Il nous apparaît néanmoins que, contrairement à ce que semblent penser les auteurs, une logique binaire – quoique autorisant des représentations parcimonieuses – n'est *a priori* pas plus « naturelle » qu'une logique plurivalente comme, par exemple, celle que nous abordons *infra* Section 2.3.3.

manière continu – des zones de stabilité qui imposent une certaine discontinuité à cette substance²¹. Un troisième paradigme propose une approche unifiée de la phonétique et de la phonologie²². En effet, la frontière entre les deux n'est pas toujours évidente, comme le montrent Scobbie et Stuart-Smith ([SSS06]) dans une étude sur les contrastes « quasi-phonémiques ». Les auteurs donnent l'exemple de la *Scottish Vowel Length Rule*, qui aboutit à des paires comme <crude> [k^hrʉd] vs <crewed> [k^hrʉ:d], <side> [sʌɪd] vs <sighed> [sʌɪd], etc²³. Chaque paire pourrait être considérée comme une paire minimale puisque une différence phonétique engendre une différence de sens. Or l'alternance de la voyelle – au moins pour les deux paires citées – est entièrement prévisible par la morphologie. Ces phénomènes sont en outre complexifiés par le fait que seules certaines voyelles sont soumises à cette règle d'allongement. Les auteurs concluent que ces phénomènes sont pertinents pour une analyse phonologique (même conservatrice), et qu'il est nécessaire, dans le cadre d'une phonétique/phonologie modulaire, de pouvoir les classer soit comme entités phonologiques, soit comme formes de surfaces dérivables d'une représentation sous-jacente. Mais, pour ce faire, il faut laisser à la phonologie la flexibilité de représenter plusieurs degrés de phonologisation afin d'intégrer dans l'analyse des cas limites comme la *Scottish Vowel Length Rule*.

Dans cette section, après avoir passé en revue les facteurs engendrant la variation des quantités physiques que nous étudions, l'approche des catégories que nous avons employée est détaillée. Cette discussion est indispensable compte tenu du fait que dans le Chapitre 6.2, nous inférons le système phonologique des dialectes à partir de mesures acoustiques : la correspondance entre la substance et la forme doit donc être rendue explicite pour justifier nos inférences.

²¹La représentation de l'intonation par le biais d'un « prosogramme » (voir par exemple Patel et collègues [PIR06]) opère une discrétisation de la courbe de F0 qui illustre un type de phonétique plus catégoriel.

²²Le succès de la « phonologie de laboratoire » illustre bien l'importance de ce nouveau paradigme (voir, à ce sujet, [DL02], 38-40).

²³Ce phénomène est en réalité beaucoup plus complexe.

2.3.1 Variation phonético-acoustique

Le thème de la variation, qui est récurrent en phonétique (notamment acoustique), semble exercer une certaine fascination. Il recouvre plusieurs concepts que nous allons tenter de démêler, tout en nous efforçant de démontrer que la variation phonétique n'est pas plus imprévisible (ou mystérieuse) que d'autres types de variations auxquelles l'être humain est confronté quotidiennement. Les facteurs de variabilité acoustique peuvent être regroupés en trois catégories :

1. la variation déterminée par des propriétés physiologiques ;
2. la variation déterminée par le contenu informationnel du signal de parole ;
3. la variation linguistique ; celle qui persiste une fois que les deux autres (ci-dessus) ont été écartés.

On sait depuis les premières études de grande ampleur en phonétique acoustique qu'une voyelle, telle qu'elle est mesurée par le biais des proéminences spectrales appelées formants, varie d'un locuteur à l'autre, entre hommes, femmes et enfants, et également chez un seul et même locuteur répétant plusieurs fois la même voyelle (e.g. [PB52]). Concernant les deux premiers types, cette variation est, toutes choses (dialecte, style, etc.) égales par ailleurs, le corollaire de différences physiologiques permanentes. En effet, les hommes ayant, par exemple, des conduits vocaux en moyenne plus longs que ceux des femmes, les formants de ces dernières sont plus élevés et plus espacés (voir e.g. [KDL96] ; d'autres variables physiologiques sont mentionnées dans [Cal89], 86-87). Malgré le fait que les valeurs acoustiques d'un /i/ en anglais américain soient en moyenne, pour les deux premiers formants, de 270 et 2300 Hz pour les hommes, et de 300 et 2800 Hz pour une femme ([KDL96], 205), c'est bien à une seule et même catégorie phonologique que l'auditeur associe ce son. Il n'y a aucune raison *a priori* de penser que cette variation ne suive pas une loi normale. En revanche, ceci n'implique pas que les différences de valeurs acoustiques observables entre les hommes et les femmes soient exclusivement une conséquence de la variation physiologique ; en comparant les valeurs formantiques des trois premiers formants vocaliques entre hommes et femmes dans 26 langues, Johnson ([Joh05], 379-381) démontre, au vu de la variation importante d'une langue à l'autre, que des aspects culturels ou sociophonétiques entrent en ligne de compte. Ce double constat nous poussera à dissocier les hommes et les femmes dans l'analyse de la Section 6.2. En ce qui concerne la

variation de valeurs formantiques pour un seul et même locuteur produisant n répétitions d'une voyelle, tous les autres facteurs (nous les détaillons *infra*) étant contrôlés, elle est imputable, en grande partie, aux limites inhérentes à la précision du contrôle moteur des articulateurs et à la tolérance du système auditif et de l'organisation mentale des catégories phonémiques vis-à-vis de cette imprécision et de la distance objective entre deux sons. À ce sujet, Broad ([Bro76]) réexamine les productions de plusieurs voyelles d'un même locuteur issues d'études antérieures et constate une variation intra-catégorielle (suivant la loi normale) très faible et une distance inter-catégorielle importante ([Bro76], 403) :

all the studies demonstrate a remarkably tight clustering of the formant frequencies for repeated items. In fact, the standard deviations are of the same order as the human auditory difference limen for formant frequencies. . .

Un argument censé prouver la variabilité atypique de la parole consiste à remarquer la grande dispersion et le chevauchement inter-catégoriel dans l'espace des deux premiers formants mesurés en un point de l'état stable d'une voyelle. Or d'autres paramètres moins variables ont été mis en évidence, par exemple, dans l'étude de Verbrugge et Rakerd ([VR86]). Les sujets ont pour tâche d'identifier des voyelles présentées dans un contexte consonantique /b_b/²⁴. Les stimuli originaux sont des monosyllabes produits en isolation par un homme et une femme. Les sujets sont soumis à 5 conditions :

- la syllabe entière, c'est la condition de contrôle ;
- les 20 % du début et de la fin de la syllabe, avec un silence (correspondant à la partie stable de la voyelle) entre les deux ;
- idem que la condition précédente, excepté que les premiers 20 % proviennent des réalisations du locuteur et les derniers, de la locutrice, ou vice-versa ; c'est la condition dite « hybride » ;
- uniquement les 20 % à l'initiale ;
- uniquement les 20 % en position finale.

Contrairement aux résultats d'études antérieures qu'ils citent, les auteurs notent une légère dégradation de l'identification entre les deux premières conditions. En revanche, les

²⁴Ce type d'expérience est regroupé sous l'étiquette « silent centers » ; voir Jenkins et collègues ([JST99]).

taux d'erreurs dans la deuxième et la troisième condition (23,1 et 27,4 % respectivement) ne sont pas significativement différents. Les deux dernières conditions ont des taux d'erreur de 56,4 et 73,8 %. Cette étude confirme ce qui avait déjà été établi : la conjonction des informations dynamiques présentes au début et à la fin d'une syllabe CVC permettent une bonne identification de la voyelle nucléaire lorsque la cible de celle-ci est remplacée par un silence. La présentation de la partie initiale ou finale seule de cette syllabe engendre des taux d'erreur élevés, ce qui suggère que ce sont bien les deux transitions à la fois qui sont pertinentes dans l'identification d'une voyelle. Le résultat le plus inattendu réside dans le fait que l'identification est aussi bonne dans la condition hybride que dans la deuxième condition. Ce dernier point permet d'émettre l'hypothèse que l'information dynamique aux marges d'une voyelle dans un contexte CVC varie peu par rapport à la cible vocalique entre un homme et une femme. Les auteurs réaffirment²⁵ le rôle des transitions comme indices acoustiques propres et rejettent l'idée selon laquelle ces transitions ne feraient que permettre à l'auditeur une extrapolation visant à prédire le spectre de la partie stable (la cible) de la voyelle (conclusion identique dans une étude de Jenkins et collègues ([JST99])). Dans le cadre de notre brève discussion sur l'apparent paradoxe de l'invariance, cette étude démontre que l'information acoustique classiquement mesurée au niveau de la cible vocalique, et dont l'extrême variabilité sert d'argument appuyant l'impossibilité de résoudre ce paradoxe, peut être contrebalancée par des indices dynamiques relativement insensibles aux variations physiologiques entre locuteurs²⁶.

Voyons maintenant la contribution des limites du système auditif au degré de liberté articulatoire (et donc acoustique). Le système auditif est décrit en détails dans [Gre96]. L'exemple de l'échelle des fréquences illustre comment celui-ci autorise la variation acoustique. Le fonctionnement de la cochlée engendre des limites de résolution fréquentielle variables. Autrement dit, la fréquence objective et la fréquence telle qu'elle est restituée

²⁵La complémentarité des valeurs d'état stable et de pentes formantiques dans l'identification des voyelles avait déjà été démontré par Lindblom et Studdert-Kennedy ([LSK67]) entre autres.

²⁶Al-Tamimi ([AT07]) aborde la question des pentes formantiques dans les dialectes de l'arabe et en français ; il ré-affirme l'importance de la conjonction des caractéristiques spectrales statiques et dynamiques dans la production et la perception de la parole.

au niveau de la cochlée n'entretiennent pas une relation linéaire. Plus précisément, plus la fréquence est élevée, plus l'écart de fréquence objective entre deux sons différents doit être élevé pour être interprété comme identique par le système auditif. Ceci a pour conséquence directe de favoriser la variation acoustique objective au fur et à mesure que l'on monte dans l'échelle des fréquences. Cette variation est donc une nouvelle fois imposée par des caractéristiques physiologiques ; elle ne présente donc que peu d'intérêt en tant qu'objet d'étude *per se* pour la description phonétique, d'autant plus que des échelles psycho-acoustiques ont été mises au point pour tenter de la maîtriser (voir *infra*). Corrélativement à ces phénomènes, et en rapport direct avec notre étude, la question des seuils différentiels de perception des formants vocaliques doit être abordée. Il s'agit ici de déterminer la déviation nécessaire dans le plan F1/F2 pour qu'un stimulus vocalique soit jugé comme différent d'un stimulus de référence. Les formants étant étudiés séparément, le seuil différentiel de perception est la différence entre la fréquence formantique du stimulus de référence et celle du stimulus le plus proche constituant néanmoins une différence perceptible.

Flanagan ([Fla55]) obtient des seuils variant de 12 à 27 Hz pour F1 (entre 300 et 700 Hz), et de 20 à 90 Hz pour F2 (1000 à 2000 Hz). Il observe en outre une asymétrie entre les valeurs de seuil inférieures et celles supérieures au stimulus de référence. Les études suivantes (rapportées dans [KPW94], 486) font apparaître des seuils allant, en moyenne, de 2,8 à 9,6 % de la valeur F1, et de 2 à 8,9 % de la valeur de F2. L'étude de Kewley-Port et Watson ([KPW94]) ne rapporte pas une asymétrie marquée ; leur seuil, peu variable d'un timbre à l'autre, est de 14,5 Hz (à peu près constant) sur F1 (<800 Hz), et augmente linéairement de 10 Hz tous les 1000 Hz sur F2. Ces études, menées dans des conditions de laboratoire, avaient pour but de détecter un changement perceptible de qualité du timbre vocalique et non un changement de catégorie phonémique. Il est donc évident que :

- la différence de fréquence opposant les réalisations types de deux phonèmes distincts est largement supérieure à ces seuils ; plus généralement, les études sur la perception

catégorielle montrent que la granularité de l'identification ne coïncide jamais avec celle de la discrimination (elle est plus grossière, [DH00]) : l'être humain est capable de discriminer au-delà de ce qu'il peut identifier, autrement dit, il est possible de discriminer à l'intérieur de catégories et pas simplement entre catégories²⁷ ;

- en situation de communication écologique²⁸, ces seuils seraient probablement insuffisants pour permettre de détecter une différence de timbre.

Ces estimations nous seront néanmoins utiles lorsqu'il s'agira de comparer deux voyelles ayant un timbre proche. La variation induite par le contexte informationnel sera traitée plus loin en référence directe aux mots-test à structure /hVd/ que nous utilisons dans notre étude (Section 2.3.2). Mais nous pouvons toutefois affirmer pour l'instant que, en l'absence de contexte informationnel, le contraste entre les voyelles des mots-test ne repose que sur la qualité phonétique des timbres ; il est donc hautement probable que si deux voyelles ont des valeurs formantiques inférieures ou à peine supérieures aux seuils que nous venons de présenter, elles font probablement partie du même phonème.

Lorsqu'il s'agit de représenter graphiquement des voyelles à partir de mesures acoustiques, il est légitime de se demander si l'on souhaite ou non que la représentation fasse sens sur le plan de la perception. Nous n'entrerons pas dans les détails de la genèse des différentes échelles utilisées de nos jours car certains manuels d'introduction à la phonétique acoustique traitent le sujet (voir, par exemple, [Hay00], 140-145). Des études expérimentales ont abouti à deux échelles très employées en phonétique : les Bark (du nom du physicien allemand, Barkhausen) et les ERB (Equivalent Rectangular Bandwidth). L'Équation 2.1 permet la transformation des Hz en Bark que nous avons utilisée dans cette étude (adaptée de [Hay00], voir [Tra90] pour les détails). Cette Équation est valable pour $x > 200$.

²⁷Pour être plus précis : Liberman et collègues ([LCSSK67]) soutiennent une version forte de la perception catégorielle pour les consonnes occlusives (la discrimination est à peine plus fine que l'identification) et une version plus faible – qui tend vers une perception continue – pour les voyelles (la variation intra-catégorielle est aisément perçue).

²⁸Par opposition aux conditions de laboratoire.

$$y = (26,81 \times x)/(1960 + x) - 0,53 \quad (2.1)$$

Intuitivement, si la taille des espaces vocaliques sur le plan F1/F2 varie d'un locuteur à l'autre en raison – entre autres – de différences d'ordre physiologique, on peut envisager qu'en procédant à une compression ou une expansion linéaire, les espaces individuels où seuls les aspects physiologiques sont facteurs de variation finissent par se ressembler. On sait d'ailleurs, d'après l'étude de Ladefoged et Broadbent ([Lad67]) que les valeurs acoustiques d'un son sont vraisemblablement jugées en référence à des caractéristiques acoustiques propres à chaque locuteur afin de déterminer l'appartenance de ce son à un phonème. En faisant varier les formants d'une phrase porteuse dans laquelle apparaît un monosyllabe test dont la voyelle reste constante, cette dernière est catégorisée différemment par les auditeurs. Par exemple, le stimulus synthétique <bit> a 375 et 1700 Hz pour F1 et F2, respectivement. Lorsqu'il est présenté après la phrase introductive « Please say what this word is » dans laquelle le premier formant varie de 275 à 500 Hz, environ 88 % de leurs 60 sujets identifient <bit>. Mais lorsqu'il est introduit par la même phrase avec un F1 variant de 200 à 380 Hz, 90 % des auditeurs entendent <bet>.

La question de la normalisation du locuteur a fait (et continue de faire) l'objet d'un grand nombre d'études (e.g. [Lob71], [Mil89], [RP94], [HR04], [ASvH04], etc.), et la plupart des travaux descriptifs basés sur des mesures acoustiques abordent des aspects de normalisation (e.g., [PB52], [Wel62], [Lad67], [Bla83], [HGCW95], [HM05], etc.).

Adank et collègues ([ASvH04]) évaluent 11 méthodes de normalisation en comparant leurs effets sur les 9 monophthongues du néerlandais produites par 160 locuteurs répartis en groupes sociolinguistiques. À partir de pourcentages de classification issus d'une série d'analyses discriminantes dont le but est de tester dans quelle mesure les procédures de normalisation préservent les contrastes phonémiques, réduisent la variation d'origine physiologique et maintiennent la variation sociolinguistique, les auteurs concluent que la méthode dite « de Lobanov » (centrer-réduire les fréquences des formants, indépendamment pour chaque formant), en référence à [Lob71], est la plus appropriée. C'est donc

cette méthode que nous utiliserons dans la suite de notre travail. ²⁹

2.3.2 Caractéristiques phonétiques des mots-test et information

La validité écologique des listes de mots du type *heed*, *hid*, *head* etc. est certes contestable dans une étude de phonétique, mais ces derniers offrent néanmoins la possibilité de comparer des voyelles en écartant les facteurs de variation importants que constituent le contexte consonantique et, plus généralement, le contexte informationnel (concept détaillé *infra*). Les variations de style ne sont pas abordées car elles ne sont pas spécifiques aux mots à structure /hVd/. En revanche, il est certain que le degré de formalité avec lequel les sujets ont abordé les enregistrements varie entre les individus, et il est, par conséquent, justifié de penser que ceci est un facteur supplémentaire de variation phonétique inter-individuelle. Une bonne compréhension de l'influence des contextes consonantique et informationnel permettra de mieux appréhender en quoi les formants mesurés en contexte /hVd/ peuvent se distinguer de ceux relevés en parole spontanée.

Pendant l'émission du [h], la langue n'étant pas requise comme articulateur dans la cavité supra-glottique, celle-ci peut donc anticiper la position de la voyelle subséquente. Par conséquent, l'effet de co-articulation de [h] sur la voyelle est quasi-nul. Le /d/ de la plupart des dialectes que nous étudions est une occlusive dont le lieu d'articulation correspond à la zone alvéolaire. Une étude de Delattre et collègues ([DLC55]) montre, à partir de stimuli synthétiques de type CV, que [d] possède un locus qui se situe aux alentours de 1800 Hz ; en d'autres termes, le deuxième formant de la voyelle suivant le [d]

²⁹Pour rappel, l'écart-type d'une variable aléatoire discrète est une mesure de sa variation donnée dans l'Équation 2.2, où \bar{x} est la moyenne. Le centrage d'une variable consiste à soustraire \bar{x} et la réduction, à diviser le résultat de cette soustraction par s pour chaque valeur prise par cette variable (Équation 2.3). Les légendes des axes des figures de la Partie II désignent le centrage et la réduction d'une variable par le terme « z-score ».

$$s = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (2.2)$$

$$z_i = \frac{(x_i - \bar{x})}{s} \quad (2.3)$$

tend vers une fréquence de 1800 Hz au fur et à mesure que l'on se rapproche de la barre d'explosion de la consonne. Une étude de Sussman et collègues ([SBDM97]) étudie l'effet du type de syllabe (CV, VC, CVC ou VCV) sur la valeur du locus. Par le biais d'une régression linéaire, les auteurs obtiennent une ordonnée à l'origine moyenne de 1232 Hz pour le cas qui nous occupe, i.e. le [d] en position de coda. Le type de syllabe semble ne pas avoir d'effet sur le locus de [d], contrairement à ce que les auteurs remarquent pour [b] et [g].

Le contexte informationnel regroupe, entre autres, les effets potentiels de la probabilité d'occurrence d'un mot³⁰, de sa probabilité d'occurrence étant donnés les n mots qui le précèdent ou qui le suivent, ou encore étant donné le contexte sémantique ou situationnel et la position dans la phrase, sur la forme phonétique de ce mot. On sait par exemple (voir l'étude de Fidelholz rapportée dans [Jur03], 45) que la première voyelle des mots fréquents <forget> ou <forgive> est un meilleur candidat à la réduction vocalique que celle des items rares que sont <forgo> et <forfend>³¹. La réduction vocalique est d'ailleurs souvent la norme pour les mots grammaticaux, qui ont la spécificité d'être particulièrement fréquents ([Byb01, BJFL⁺03, JBFL⁺98]). Lorsque van Bergem ([vB93]) montre l'effet de l'accent de phrase, l'accent de mot et la classe (lexicale *vs* grammaticale) sur la réduction vocalique phonétique en néerlandais, on peut considérer qu'il n'étudie en réalité qu'un seul facteur sous-jacent : l'information.

Les études qui ont examiné l'effet de la probabilité d'occurrence de plusieurs mots ([FLM99, BJFL⁺03]), qu'il s'agisse de $P(m_i|m_{i-1})$, la probabilité du mot m_i sachant que

³⁰Les probabilités sont généralement déduites de fréquences d'occurrence des mots (nous utilisons ce dernier terme pour simplifier le propos) dans des corpus. L'estimation de cette probabilité ne correspond pas au rapport de l'effectif d'un mot donné sur le nombre total de mots d'un corpus ; en effet, ceci conduit à attribuer à tort une probabilité nulle aux mots absents du corpus et donc, *a contrario*, à surestimer les probabilités des items présents dans le corpus. Ce biais peut être atténué par une estimation de Good-Turing ([GS95], voir aussi [Jur03], 206 *sqq.*).

³¹Ces résultats ne disent toutefois pas si le processus est productif ou bien si la répartition entre voyelles réduites et non réduites a été fixée une fois pour toutes en diachronie. Il est d'ailleurs vraisemblable – et ceci répond peut-être en partie à la question – que ces quatre items ne suivent pas exactement le même schéma en anglais britannique et américain. En effet, les dictionnaires de prononciations ([Wel90] et [Jon03]) notent une alternative entre voyelle pleine et voyelle réduite en anglais américain pour la première syllabe de <forget> et <forgive>, mais une seule valeur phonologique (le schwa) pour l'anglais britannique.

le mot m_{i-1} a été prononcé, ou $P(m_{i-1}m_i)$, la probabilité d'occurrence des mots m_i et m_{i-1} , montrent un effet de cette probabilité sur la substance phonétique : les phénomènes de réduction (chute d'une consonne en coda, réduction vocalique, durée plus brève d'un mot) sont associés à des items dont la probabilité d'occurrence (ou de co-occurrence) est élevée. Par exemple, la probabilité conditionnelle plus élevée des séquences <did you>, <told you> et <would you> donne lieu à un nombre de palatalisations plus important des coronales que les paires <at you>, <but you> et <good you>, pour lesquelles $P(m_i|m_{i-1})$ est moins élevé ([Jur03], 52). D'autres auteurs cités dans [Jur03], 53, ont démontré une corrélation entre la probabilité conditionnelle de bigrammes et la localisation de l'accent de phrase : plus la probabilité du bigramme est faible, plus la probabilité qu'il reçoive l'accent de phrase est importante.

En résumé, l'entropie linguistique, i.e. le degré d'incertitude associé à la réalisation d'un mot dans un certain contexte, affecte sa forme phonétique. Plus précisément, il apparaît qu'à une entropie faible correspond une réalisation phonétique plus déviante d'une réalisation prototypique (telle qu'elle est suggérée par la forme phonologique d'un mot), et inversement dans le cas d'une entropie élevée. On peut donc avancer que, lorsque l'occurrence d'un mot est totalement déterminée par le contexte, ou qu'elle est bien plus probable que celles d'autres mots dans ce même contexte (dans les deux cas, l'entropie est faible), sa forme phonétique a un degré de liberté accru par rapport à une forme canonique. La théorie H&H de production de la parole et le Neighborhood Activation Model (NAM) constituent un cadre particulièrement adapté à la description et la compréhension de ces phénomènes.

La théorie H&H, proposée par Lindblom ([Lin90]), met l'accent sur la relation entre la production du locuteur et ce que ce dernier anticipe des attentes de l'auditeur pour que l'accès lexical soit optimal. Optimal... et non maximal³², car la théorie H&H prévoit une

³²D'ailleurs, même dans des conditions d'entropie élevée, incitant le locuteur à hyper-articuler, sa production n'ira jamais jusqu'à satisfaire pleinement les attentes de l'auditeur, comme le prouve l'effet « hyperspace » ([JFW93, Joh00]) : lorsque des sujets sélectionnent des voyelles de synthèse prototypiques de leur système, l'espace acoustique obtenu est plus grand que celui qui est généré par des mesures acoustiques de voyelles prototypiques produites par les mêmes sujets. Si Whalen et collègues ([WMP⁺04])

interaction et la recherche constante d'un équilibre entre :

- le contexte informationnel, ou, dans la théorie H&H, les « signal-complementary processes » ;
- les contraintes « output-oriented » ; il s'agit d'optimiser l'information acoustique pour faciliter l'accès lexical. C'est l'« hyperspeech », auquel le locuteur a recours lorsque l'entropie est élevée³³ ;
- le contrôle « system-oriented », qu'on peut assimiler à la notion d'économie de l'articulation : « [u]nconstrained, a motor system tends to default to a low-cost form of behavior. » ([Lin90], 413).

La forme acoustique, qui dépend de l'équilibre entre ces trois paramètres, est donc censée constituer un contraste optimal. Reste à définir comment ce contraste est implémenté dans le domaine de la production, en particulier, sur le plan acoustique. Une très brève description du Neighborhood Activation Model (NAM) est instructive, et adaptée à notre propos en vertu du fait que ce dernier rend compte des effets de fréquences, et de l'entropie, qu'il est cité par Lindblom ([Lin90], 409-410), et que ses prédictions ont été utilisées dans une étude acoustique d'un grand intérêt pour la suite de notre propos.

Le modèle NAM ([GPL96], 311-313) prévoit que les mots fréquents ont une plus forte probabilité d'être reconnus, et de l'être rapidement, mais cette probabilité est pondérée à la baisse par :

- la densité du voisinage, i.e. le nombre de mots possédant des caractéristiques phonético-acoustiques proches du stimulus et leur degré de confusion possible avec le stimulus ;
- la fréquence du voisinage : la fréquence des candidats du voisinage par rapport à la fréquence du stimulus.

Si Goldinger et collègues ([GPL96], 311-313) rapportent que ces deux aspects affectent la reconnaissance des mots, il serait donc informatif, si l'on se replace dans le cadre de la théorie H&H, de connaître ce que le locuteur anticipe des difficultés de l'auditeur

répliquent en partie – mais de façon flagrante – cet effet, ils émettent quelques réserves quant à l'utilité du concept d'« hyperspace » pour expliquer l'apparent décalage entre production et perception.

³³Nous laissons de côté l'adaptation du locuteur au bruit environnant, parfois appelée « effet Lombard » car, bien que cet aspect fasse partie du contexte informationnel, nous postulons que les effets de ce type de bruit sont constants dans nos données.

à reconnaître les mots en contexte /hVd/. Malheureusement, tel qu'il est décrit dans [GPL96], le modèle NAM ne règle pas la question de la distance entre le stimulus et ses voisins. S'agit-il d'une simple distance de type Levenshtein ([DHS01], 418-420), où les unités primitives sont les phonèmes, ce qui aboutirait, par exemple, à considérer que la distance entre /hi:d/ et /hɪd/ est la même que celle qui sépare /hi:d/ et /hæd/? On peut encore imaginer une distance du même type, mais cette fois entre des traits phonologiques, ou peut-être une distance phonétique euclidienne dans l'espace des trois premiers formants et de la durée. Quelle que soit la forme de cette distance, on doit pouvoir considérer qu'elle contraint le degré de liberté acoustique des listes de mots en /hVd/. En effet, si l'on pose que la liste de 19 mots à structure /hVd/ constitue un lexique de taille finie où les items sont équiprobables³⁴, on est dans la situation où la densité du voisinage et la fréquence du voisinage sont extrêmement élevés. Par conséquent, du point de vue de l'auditeur, l'accès au lexique se trouve compliqué, et du point de vue du locuteur, celui-ci doit tout mettre en œuvre pour maximiser les contrastes³⁵. Il faut donc s'attendre à ce que les distances acoustiques entre deux mots-test soient magnifiées par rapport à de la parole spontanée. Notons en outre que le recours au décodage en phonèmes dans la reconnaissance lexicale est vraisemblablement cantonné à certains cas bien précis. D'après Sendlmeier ([Sen95], 140) :

a listener can switch to single sounds or even distinctive features when discriminating, for example, minimal pairs or difficult words such as proper names, words of a foreign language or pseudowords, and he/she can then switch back again to the syllable or even word level when progressing in the recognition process.

Le recours au phonème comme unité primitive pour l'accès au lexique semble donc s'appliquer aux cas où l'entropie est élevée (paires minimales hors contexte, noms propres, pseudo-mots, etc.), ce qui est le cas dans la liste des mots /hVd/ de notre corpus. À redondance linguistique élevée, la reconnaissance lexicale semble ne plus impliquer le phonème ; produire un prototype dans ce cas-là devient non seulement inutile, mais également dis-

³⁴On peut en effet concevoir que les locuteurs ont rapidement réalisé que leur tâche consistait à prononcer un faible nombre de mots ayant un contexte consonantique constant. Si l'on écarte le cas de *hured*, où la correspondance entre forme phonologique et graphique est loin d'avoir fait l'objet d'un consensus, nous doutons que les probabilités d'occurrence différentes de chacun des mots-test dans la parole naturelle affecte leur forme phonétique dans cette tâche artificielle.

³⁵Ceci est probablement contrebalancé par le fait qu'une liste de mots est un cas limite, peu écologique, d'acte de communication.

pendieux en termes de coût articulatoire, comme le prévoit la théorie H&H. Il s'ensuit que la distance que le locuteur cherche à maximiser dans le cas des listes de mots en /hVd/ concerne très probablement le phonème (voire une entité de taille inférieure)³⁶.

Aylett ([Ayl00]) reformule la théorie H&H dans un cadre plus proche de la théorie de l'information. Pour ce faire, il distingue trois types de redondances :

- la redondance linguistique, qui traduit la proportion des entités (mots, syllabes, phonèmes, etc.) dont la présence est statistiquement déterminée dans un message. C'est dans cette acception que le terme est utilisé par Shannon et Weaver ([SW63], 43-44) ; par exemple, <to> dans <I'm going to leave you> est très redondant ;
- la redondance acoustique : à contexte informationnel égal, un mot est plus facilement identifiable si sa redondance acoustique est élevée. Cette redondance se traduit par une articulation plus soignée, très vraisemblablement synonyme du « hyperspeech » de Lindblom ([Lin90]) ;
- la redondance du signal : il s'agit de la combinaison des deux types de redondance que nous venons de mentionner. Aylett formule l'hypothèse suivante : la redondance du signal tend à varier le moins possible au cours du temps. Autrement dit, il prévoit que la faible redondance d'une entité linguistique sera compensée par une redondance acoustique accrue, et, inversement, la redondance linguistique élevée d'une entité ne nécessitera qu'une faible redondance acoustique pour remplir le critère de variation minimale de la redondance du signal au cours du temps.

Le lien avec la théorie H&H est évident, surtout si l'on admet que la redondance acoustique est proportionnelle à l'effort d'articulation. Jusqu'ici, les études empiriques auxquelles nous avons fait allusion confirment ces deux formulations de la théorie H&H. Mais elles n'apportent rien à la compréhension de la manière dont les formants des voyelles de nos mots-test peuvent être affectés. En effet, les résultats décrits jusqu'ici sont binaires :

³⁶La question de l'unité primitive de perception de la parole n'est toujours pas résolue ([GA03]) ; par exemple [MWW94] soutient l'interprétation directe des traits phonétiques, sans passer par le niveau phonémique. Dans une autre perspective, Durand et Laks ([DL02], 30) désignent la syllabe comme unité fondamentale « which provides the link between a physically driven system and higher-level units ». Ce que nous venons de décrire est donc, certes fondé empiriquement et intuitivement, mais nous ne prétendons pas qu'il s'agit d'une réponse définitive.

soit la voyelle est pleine, soit elle est réduite ; soit la consonne est produite, soit elle est élidée. Or les mots-test à structure /hVd/ sont des formes isolées accentuées ; leur voyelle ne peut donc pas être candidate à la réduction phonologique.

Wright ([Wri03]) analyse dans quelle mesure la densité du voisinage et la fréquence relative d'un stimulus affectent les valeurs des formants. Il utilise deux types de stimuli :

- les mots « faciles », qui ont un voisinage peu dense et une fréquence élevée par rapport à leurs voisins ;
- les mots « difficiles », avec les caractéristiques opposées : un voisinage dense et une fréquence relative faible.

F1 et F2 sont ensuite mesurés au point temporel où ils atteignent une distance maximale par rapport au centre de l'espace vocalique, puis transformés en Bark. Conformément aux études précédentes (e.g. [Lin83]), la taille de l'espace vocalique délimité par les timbres moyens de la condition « mots difficiles » est supérieure à celle de l'espace des « mots faciles ». Mais le résultat le plus intéressant est le suivant : la dispersion autour de la moyenne des valeurs de chaque timbre est plus importante (pour la majorité des voyelles) dans la condition « difficile ». Les différences de dispersion les plus marquées entre les deux conditions concernent les voyelles situées aux extrémités de l'espace vocalique ; et ce sont celles-là même qui exhibent la différence moyenne la plus saillante entre les deux conditions. Autrement dit, l'augmentation moyenne de distance entre les timbres est accompagnée d'une augmentation de leur variance individuelle ; mais l'étude ne précise pas explicitement si les contrastes sont améliorés ou non, i.e. si les frontières entre les nuages de dispersion de timbres dans F1/F2 sont plus ou moins nettes.

Aylett et Turk ([AT06]) ont des résultats comparables, qui confirment dans l'ensemble l'hypothèse de la centralisation dans F1/F2 des voyelles associées à une redondance élevée. Ils ajoutent que la présence de proéminences prosodiques est associée à des syllabes de faible redondance linguistique ; la prosodie permet donc elle aussi de moduler la redondance acoustique en fonction de la redondance linguistique afin d'assurer la stabilité de la redondance du signal dans le temps.

À ce stade de la discussion, il apparaît que la variation acoustique est dans une large mesure déterminée par des différences physiologiques et par le contexte informationnel. Ces dernières seront en partie écartées par les méthodes dites « de normalisation » employées dans la Section 6.2. Le contexte informationnel est, quant à lui, de fait neutralisé par l'utilisation des mots à structure /hVd/, ce qui constitue un atout majeur pour une étude des propriétés acoustiques des voyelles. Nous pouvons donc affirmer que les mots à structure /hVd/ constituent un outil d'étude des dialectes particulièrement adapté (ceci sera d'ailleurs mis en évidence par les scores de classification obtenus dans la Section 7.2) car ils offrent un contrôle des facteurs qui est optimal pour une étude observationnelle. Ils comportent néanmoins certaines faiblesses qui ne sont pas dues au fait qu'il s'agisse de formes isolées, mais plutôt à l'exhaustivité de la liste de ces mots-test, au degré de consensus avec lequel les locuteurs vont attribuer une forme phonologique à certains mots rares ou pseudo-mots (nous mentionnons le cas de *hured* dans la Section 6.3.1.3) et à l'impossibilité de garantir *a priori* l'appartenance d'un mot-test à un ensemble lexical donné. En ce qui concerne l'exhaustivité de la liste, les 19 mots disponibles ne couvrent pas l'ensemble des contrastes possibles dans les 13 dialectes étudiés. Par exemple, le phénomène de convergence COT-CAUGHT³⁷, caractéristique, notamment, de certaines variétés de l'anglais d'Écosse ne peut pas être mis en évidence puisque aucun des mots de la liste de ne peut être assimilé à l'ensemble CAUGHT. En effet, le seul mot-test présentant *a priori* une voyelle postérieure mi-fermée est *hoard* ; or s'il peut, faute de mieux, faire l'affaire pour tester une éventuelle convergence COT-CAUGHT (non attestée) en anglais standard par exemple, ce n'est pas le cas en anglais d'Écosse car la rhoticité fait de *hoard* un mauvais candidat pour l'ensemble CAUGHT. Les mots <wait> et <weight> ne riment pas dans certains dialectes du nord de l'Angleterre ([ei] et [ɛi], respectivement) ; or il n'est pas possible de révéler ce contraste avec le mot-test *hade*, représentant de la supra-catégorie FACE. Il arrive également que la variation allophonique ait un intérêt dans la description des dialectes. En effet, nous verrons plus loin (Section 3.2.1) que l'ensemble PRICE

³⁷On rencontre souvent cette dénomination bien que les ensembles lexicaux standards soient LOT et THOUGHT.

peut-être subdivisé en deux sous-ensembles à Hull et Newcastle (la qualité de la voyelle ou son degré de diphtongaison étant affecté par le voisement de la consonne qui suit), et ici encore, le mot-test *hide* ne permet pas de révéler cette variation supplémentaire (par rapport à l'anglais standard).

Il convient maintenant d'examiner plus précisément la manière dont l'espace multidimensionnel des paramètres physiques est projeté dans un espace de catégories phonémiques.

2.3.3 Théorie du prototype et théorie des ensembles flous

Dans la pensée structuraliste ([Sau95, Jak63]), le système phonologique d'une langue contient des entités discrètes formant des catégories hermétiques les unes aux autres. Pour accéder au statut d'entité pertinente, de phonème, un son doit constituer un contraste fonctionnel qu'il est possible de révéler par le biais du test des paires minimales. Les catégories regroupent les formes abstraites de sons définis par des traits distinctifs. Pour appartenir à une catégorie, une entité doit en posséder tous les traits distinctifs et les posséder pleinement. Il n'est pas ici question de rejeter ce point de vue, mais il nous incombe de trouver une solution alternative nous permettant de résoudre l'apparent paradoxe selon lequel un seul et même phonème a pour équivalents tangibles des événements acoustiques très variables – mais, comme nous venons de le voir, cette variabilité est en grande partie prévisible. La théorie du prototype constitue un cadre tout à fait adéquat dont nous introduisons quelques rudiments avant d'aborder cette notion par le biais du cadre formel de la théorie des ensembles flous.

Un prototype désigne souvent le membre d'une catégorie qui est (quasi) universellement considéré comme un bon représentant de celle-ci. Par exemple, les études sur la catégorisation des couleurs par Berlin et Kay, Eleanor Rosch et d'autres (rapportées dans [Tay95], 1-20) font apparaître que certaines couleurs – 11 pour Berlin et Kay – forment un inventaire que l'on retrouve dans presque toutes les langues étudiées. Ces couleurs, qui ont donc un statut particulier, sont qualifiées de « focales » ; elles exhibent, en outre, les

caractéristiques suivantes (d'après Rosch) :

- elles sont stables quelle que soit la langue parlée : lorsqu'on demande à des sujets de cultures diverses de choisir un bon exemple de rouge, ils ont tendance à retenir la même nuance ; il existerait donc une nuance de rouge qui puisse probablement être universellement considérée comme le centre de la catégorie « rouge » ;
- une expérience impliquant des locuteurs natifs de 23 langues différentes, à qui on a demandé de nommer les couleurs présentées sur des échantillons, répondent plus rapidement lorsqu'il s'agit de couleurs focales ;
- dans une autre expérience, une couleur est présentée à des sujets pendant 5 secondes ; puis, après 30 secondes, leur tâche consiste à reconnaître la couleur qu'ils ont vu parmi d'autres ; les couleurs focales sont reconnues plus facilement. Un aspect particulièrement intéressant de cette expérience réside dans le fait que les sujets Dani (tribu papoue de Nouvelle-Guinée), ne disposant que de deux termes pour tout le spectre des couleurs, montrent également de meilleures performances pour les couleurs focales.

Si ces catégories de couleurs « focales » semblent faire l'objet d'un consensus entre les cultures, il est fort probable que leur existence soit une conséquence de caractéristiques inhérentes des systèmes perceptif et cognitif de l'être humain. Comme le formule Taylor ([Tay95], 52) : « colour categories pre-exist their linguistic encoding »³⁸. D'ailleurs, ce caractère « focal » ne se limite pas aux couleurs. Il inclut, entre autres, les formes géométriques, avec probablement une interaction entre facteurs purement biologiques et facteurs culturels³⁹. Il semble en outre que certaines voyelles obéissent à ce principe. En effet, la description de Stevens ([Ste89]) des relations entre les domaines articulaire, acoustique et auditif peut être interprétée comme une description des raisons naturelles permettant l'émergence de prototypes de type focal, i.e. universellement contraints. L'auteur démontre que ces relations ne sont pas monotones ; c'est ce qui est illustré dans la

³⁸Voir aussi Lakoff ([Lak87], 24-30).

³⁹Voir, par exemple, l'étude de Raghurir et Greenleaf ([RG06]) sur les dimensions préférées des rectangles, et la référence au fameux rectangle d'or.

Figure 2.1. Dans la zone 1, le mouvement articuloire n'a que peu d'impact sur le signal acoustique, puis, dans la zone 2, le mouvement articuloire (pourtant moindre que celui de la zone 1) engendre une modification importante de la valeur que prend la variable acoustique ; enfin, la zone 3 illustre le même schéma que la zone 1. Les zones où le paramètre acoustique reste relativement insensible aux perturbations affectant le paramètre articuloire ont, d'après Stevens, des conséquences sur la formation de l'inventaire phonémique des langues. Donc, de même que les cônes de la rétine humaine sont naturellement plus sensibles à certaines longueurs d'ondes (phénomène avancé pour l'existence des couleurs focales, [Tay95], 13-14), ces zones de stabilité acoustique constituent de véritables « niches » propices à l'apparition de prototypes de voyelles.

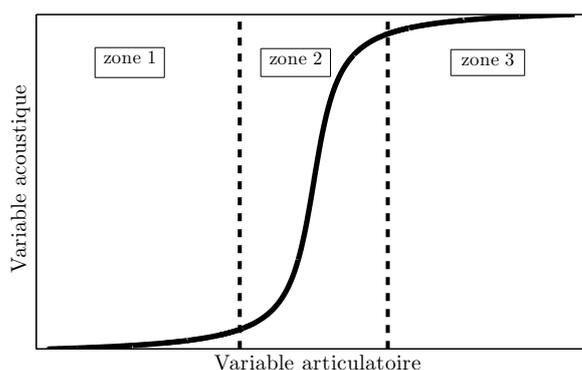


FIG. 2.1 – Illustration de la théorie quantique (adapté de [Ste89], 4).

Les prototypes auxquels nous faisons référence dans cette étude ne sont pas de type focal (tel que nous venons de définir ce terme), en cela qu'ils ne font pas référence à des représentants universellement jugés comme bons d'une catégorie. Ils sont dépendants du système vocalique de la langue, du dialecte, voire du locuteur. En d'autres termes, si la réalisation du /u/ de l'anglais britannique standard ([u]) n'est pas universellement considérée comme un exemple typique de [u], c'est néanmoins le prototype du /u/ de l'anglais britannique standard. La notion de prototype que nous utilisons ici est spécifique à une communauté linguistique précise, et est donc différente du prototype « focal » universel et biologiquement contraint. Puisque nos travaux s'appuient sur des mesures de grandeurs

physiques, le prototype dans notre étude n'est pas une représentation abstraite du centre d'une catégorie, mais bel et bien un « exemplaire »⁴⁰. En d'autres termes, dans notre étude, le prototype d'une catégorie phonémique est un point dans un espace multidimensionnel dont les dimensions sont des variables corrélées aux grandeurs physiques mesurables dans le signal. Pour prendre un exemple simplifié, si l'on considère que la valeur des deux premiers formants au milieu temporel de la voyelle décrit de façon exhaustive la manière dont les voyelles s'organisent en catégories, on pourrait considérer que la valeur médiane⁴¹ de chaque phonème représenté dans le plan F1/F2⁴², est le prototype de cette catégorie. Autrement dit, le degré d'appartenance d'une voyelle à cette catégorie peut-être directement estimé par une mesure appropriée de la distance entre les coordonnées de cette voyelle et celles de la voyelle possédant les valeurs médianes dans le plan F1/F2. Cette dernière voyelle n'est donc pas une simple « instanciation » d'un prototype abstrait ; elle est le prototype.

Le prototype « appris » (par opposition à focal) d'un phonème a dans une langue donnée présente certaines caractéristiques intéressantes. Comme le résumant Iverson et Kuhl ([IK95], 553) :

Stimuli judged as exceptionally good instances of phonetic categories (prototypes) make neighboring tokens in the vowel space seem more similar, exhibiting a *perceptual magnet effect*.

Dans l'étude de Kuhl ([Kuh91]), l'auteur a synthétisé plusieurs exemples de /i/, et les a soumis à des sujets américains adultes afin qu'ils les jugent. Puis, le timbre perçu comme le meilleur /i/ ainsi que le plus mauvais ont été choisis, et leurs formants manipulés afin de créer des variantes de ces deux stimuli de référence. Une tâche de discrimination impliquant des sujets adultes et des nourrissons a démontré ce qui est depuis connu sous le nom de « perceptual magnet effect » : les variantes du prototype sont plus difficiles à discriminer que ne le sont les variantes du timbre non prototypique. D'autres études (citées dans [IK95], 553) ont ensuite démontré que :

⁴⁰Traduction littérale de l'anglais <exemplar>.

⁴¹Il s'agit ici de la médiane comme estimateur de la tendance centrale des distributions de F1 et F2.

⁴²Voir les illustrations de la Partie II : Figure 6.12 et suivantes du même type.

- il est attesté chez le nourrisson dès l'âge de six mois ;
- le prototype, dans ce cas-là est une entité apprise, propre à un système phonologique, comme le démontrent des études translinguistiques ;
- il implique une distorsion de l'espace acoustique qui conduit à une réduction de la distance perceptive autour d'un prototype.

Ceci a des conséquences directes sur la manière dont les catégories sont représentées, et, plus particulièrement, sur la variation acoustique tolérée à l'intérieur d'une catégorie phonémique. En effet, ce phénomène peut être interprété comme un degré de liberté acoustique supplémentaire car la distance perceptive d'un stimulus au prototype est virtuellement nulle quand sa distance objective est très faible. Dans la discussion de leurs résultats, Iverson et Kuhl ([IK95], 560) insistent sur l'apport de cet effet attracteur au domaine de la perception catégorielle : il avait déjà été démontré que la discrimination de voyelles aux frontières de catégories était plus aisée que celle impliquant des timbres à l'intérieur de catégories. L'effet attracteur du prototype démontre en outre qu'il existe à l'intérieur d'une seule et même catégorie un degré de conformité au prototype qui influence la distance perçue entre deux exemplaires de voyelles.

Broad ([Bro76]) aborde la question des catégories par le biais de la production et non pas de la perception. Son approche est intéressante car, bien que la théorie des ensembles flous et la théorie des probabilités offrent des perspectives différentes ([Zad95]), sa description du chevauchement des catégories par le biais de gaussiennes dont il ajuste de manière heuristique les paramètres pour les besoins de sa démonstration, n'est pas très éloignée, conceptuellement, de la philosophie des ensembles flous. À partir de la dispersion des trois premiers formants de voyelles répétées par un même locuteur, il partitionne l'espace vocalique en plusieurs sous-espaces incluant 1,28 écart-type au-dessus et au-dessous de la moyenne (i.e. ces sous-espaces incluent 80 % de la distribution théorique). En première approximation, cela aboutit à des valeurs, à peu près constantes pour toutes les voyelles, de 40, 140 et 160 Hz pour F1, F2 et F3 (ces valeurs concernent des voyelles isolées). Ces résultats donnent une idée de la taille des catégories phonémiques dans l'espace des trois

premiers formants.

La catégorisation des voyelles dans un contexte multi-dialectal a été étudiée par Evans et Iverson ([EI04]). Dans cette étude, les sujets jugent des voyelles dans des monosyllabes présentés à l'intérieur d'une phrase porteuse, tantôt prononcée avec un accent du nord, tantôt avec un accent du sud-est. Dans une première expérience, deux groupes de sujets, vivant tous à Londres, dont l'accent d'origine est soit du nord, soit du sud-est, ont pour tâche de déterminer la meilleure voyelle dans des mots synthétisés tels que <bud> ou <bath> (ces stimuli comportent les traits déterminant la partition nord/sud de l'Angleterre). Lorsque la phrase porteuse est produite dans l'accent d'origine des sujets, ceux-ci choisissent les stimuli dont les formants sont en adéquation avec la production attendue d'un locuteur de cet accent. Par exemple, pour <bud> présenté dans une phrase porteuse dans leur accent d'origine, les sujets originaires du sud choisissent [ʌ] et ceux du nord, [ʊ]. Lorsque la phrase n'est pas produite dans leur accent d'origine, les sujets des deux groupes ne choisissent aucune des ces deux variantes, mais plutôt une voyelle intermédiaire de type [ə]. Dans un second temps, un troisième groupe de sujets, composé d'adolescents entre 16 et 17 ans originaires de Ashby de la Zouch (nord linguistique de l'Angleterre), est soumis au même test. Leur meilleur exemplaire, quelle que soit la phrase porteuse, est celui qui correspond à la production attendue d'un locuteur septentrional. Ceci démontre que :

- des locuteurs vivant à Londres, et étant de ce fait en contact avec des accents du nord et du sud, sont capables d'adapter leur représentation prototypique d'un timbre vocalique. Cette représentation est toutefois en partie erronée (cas du [ə]) lorsque l'accent de la phrase porteuse est différent de leur accent d'origine.
- des locuteurs n'ayant été que très peu en contact avec un accent différent du leur ne procèdent pas à cette adaptation au dialecte.

Le premier point démontre que des auditeurs en contact prolongé avec d'autres accents ajustent leur représentation prototypique d'un timbre en fonction du dialecte du locuteur. De plus, cet ajustement ne semble pas être conforme à la manière dont l'apprenant assimile les phonèmes d'une langue étrangère à ceux de sa langue maternelle. En effet, les valeurs

intermédiaires entre [ʌ] et [ʊ] obtenues dans le cas où l'accent d'origine des sujets était différent de celui de la phrase porteuse tendent à prouver, d'après Evans et Iverson, que la constitution éventuelle d'un système phonologique dialectal supplémentaire passe par la formation de prototypes à partir d'épisodes stockés dans la mémoire à long terme, et non par l'assimilation des voyelles du nouveau système à celles de celui du dialecte natif. L'absence de correspondance exacte entre ce prototype nouvellement formé et les productions typiques du dialecte pourrait être imputable à la distorsion induite par l'effet attracteur des prototypes du dialecte d'origine.

Dans leur étude suivante, Evans et Iverson ([EI07]) poussent davantage leur analyse. Ils inspectent l'évolution des productions de locuteurs originaires du nord sur leurs deux premières années à l'université, période pendant laquelle, selon les auteurs, ils sont confrontés à l'anglais standard (du sud). L'hypothèse de départ est que le système vocalique des sujets va tendre vers celui de l'anglais standard. Dans un premier temps, les sujets sont enregistrés à 4 points temporels sur les deux ans, et leur production est jugée par 6 phonéticiens sur un continuum nord/sud. Les résultats font apparaître que, malgré des différences individuelles importantes (certains locuteurs avaient au départ un degré de proximité au pôle « nord » très différents), la tendance générale montre un déplacement du nord vers le sud. Les auteurs se penchent ensuite sur la reconnaissance de phrases dans le bruit : une série de phrases sont présentées aux sujets, tantôt lues avec un accent du nord, tantôt avec en anglais standard, avec différents rapports signal/bruit. Les résultats montrent que les sujets originaires du nord, mais dont le système vocalique atteint les meilleurs scores de proximité au sud dans l'expérience décrite plus haut, identifient mieux les phrases prononcées en anglais standard.

Compte tenu des éléments passés en revue dans cette section, la formalisation de la théorie du prototype par le biais de la théorie des ensembles flous semble justifiée. Nous développons dans ce qui suit un cadre conceptuel qui, sans apparaître explicitement dans la suite de notre travail (sauf Sections 7.6 et 7.7), a sous-tendu notre vision de l'interface phonétique/phonologie.

Nous nous basons notamment sur Negnevitsky ([Neg02], 87-128) et Jang et Gulley ([JG99]); la description originale de la théorie est exposée dans Zadeh ([Zad65]). Notre exemple, simplifié, est le suivant : les voyelles peuvent être antérieures, centrales ou postérieures ; comment caractériser l'appartenance d'une voyelle à l'une de ces trois catégories à partir des valeurs de F2 ? Soit X , l'univers du discours (i.e. ici, les fréquences de F2 centrées-réduites⁴³), contenant quatre éléments (voyelles) : $X = \{x_1, x_2, x_3, x_4\}$. Le sous-ensemble A (les voyelles antérieures) de X contient deux éléments : $A = \{x_1, x_2\}$. En logique binaire classique, le sous-ensemble A peut-être décrit de la manière suivante : $A = \{(1/x_1), (1/x_2), (0/x_3), (0/x_4)\}$, où chaque élément x_i est caractérisé par sa fonction d'appartenance $\mu_A(x_i)$ au sous-ensemble A sous la forme $\{\mu_A(x_i)/x_i\}$. Dans ce cas, $\mu_A(x_i)$ ne peut prendre que deux valeurs :

1. $\mu_A(x_i) = 1$, si $x_i \in A$;
2. $\mu_A(x_i) = 0$, si $x_i \notin A$.

Ceci conduit à la situation illustrée dans la Figure 2.2⁴⁴, qui va à l'encontre de l'intuition et de ce que nous venons de voir concernant la catégorisation. On y constate que tout x ayant une fréquence $Freq$ comprise entre environ 0,833⁴⁵ et 2,5 écart-types appartiennent pleinement et exclusivement au sous-ensemble A , et subitement, à partir de $Freq < 0,833$, x ne fait plus du tout partie du sous-ensemble A .

La théorie des ensembles flous propose une *via media* évitant cette rupture abrupte, particulièrement aberrante lorsque l'univers du discours prend la forme d'une variable continue⁴⁶. Dans ce contexte, $\mu_A(x_i)$ peut prendre comme valeur tout nombre réel entre 0 et 1 compris. La Figure 2.3 représente une première tentative de modélisation des catégories antérieur (A), central (C) et postérieur (P) par le biais d'ensembles flous. Les

⁴³Pour l'exemple, les valeurs de F2 sont bornées entre -2,5 et 2,5 écart-types.

⁴⁴Toutes les figures dans ce travail ont le point comme séparateur de décimales ; la virgule est utilisée partout ailleurs.

⁴⁵Cette valeur délimite le premier tiers de l'univers du discours.

⁴⁶Sauf, peut-être, dans les cas avérés de perception catégorielle au sens fort, comme, par exemple, la perception de [b], [d] et [g] à partir de modifications progressives des transitions de F2 ([LCSSK67], 442 *sqq.*), qui correspondent, sur le plan de la perception, à des sauts quantiques d'une catégorie à l'autre.

triangles⁴⁷ représentent les fonctions d'appartenance d'une voyelle à l'une des trois catégories. Par exemple, on sait que $x_1 \in A$; en imaginant que $x_1 = 1,3$ dans l'univers du discours, la ligne discontinue schématise le degré d'appartenance de x_1 au sous-ensemble A , et la ligne en pointillés, le degré d'appartenance de x_1 à C . On peut donc noter que $\mu_A(x_1) = 0,4146$ et $\mu_C(x_1) = 0,1333$. La voyelle correspondant à x_1 est donc partiellement antérieure, et centrale dans une moindre mesure.

Un avantage supplémentaire de la théorie des ensembles flous réside dans le fait qu'il est possible de créer de nouveaux sous-ensembles, comme l'être humain le fait tous les jours, par le biais de variables linguistiques, ce que les anglophones nomment « hedges ». Par exemple, les adverbes d'intensité permettent une telle modification : les hommes peuvent être grands, très grands, extrêmement grands, plus ou moins grands, etc. De même, il est possible de considérer qu'une voyelle peut être perçue comme très centrale, plus ou moins centrale, etc. La théorie des ensembles flous met à disposition des expressions mathématiques pour tenter de capturer la manière dont ces sous-ensembles formés à partir de « hedges » peuvent être dérivés des fonctions caractérisant des sous-ensembles de départ. Dans la Figure 2.4, nous donnons l'exemple du sous-ensemble des voyelles très centrales et de celui des voyelles plus ou moins centrales, obtenus, en suivant [Neg02], 97 :

$$\mu_C^{tres}(x) = [\mu_C(x)]^2 \quad (2.4)$$

$$\mu_C^{+/-} = \sqrt{[\mu_C(x)]} \quad (2.5)$$

Si l'on pose que le prototype x_{proto} d'une catégorie (ou sous-ensemble) A est caractérisé par le degré d'appartenance $\mu_A(x_{proto}) = 1$, il s'ensuit que, en utilisant une fonction triangulaire, x_{proto} ne peut prendre qu'une seule valeur dans l'univers du discours. Autre-

⁴⁷Ces fonctions sont conventionnellement qualifiées de triangulaires, mais il ne s'agit là que d'une image facilitant la visualisation des sous-ensembles. Il est par exemple évident que, si le triangle qui matérialise le sous-ensemble C était un véritable triangle (i.e. si tous les sommets étaient reliés), la fonction de x_1 prendrait deux valeurs : 0 et 0,1333 (le point x_1 est illustré dans la Figure 2.3).

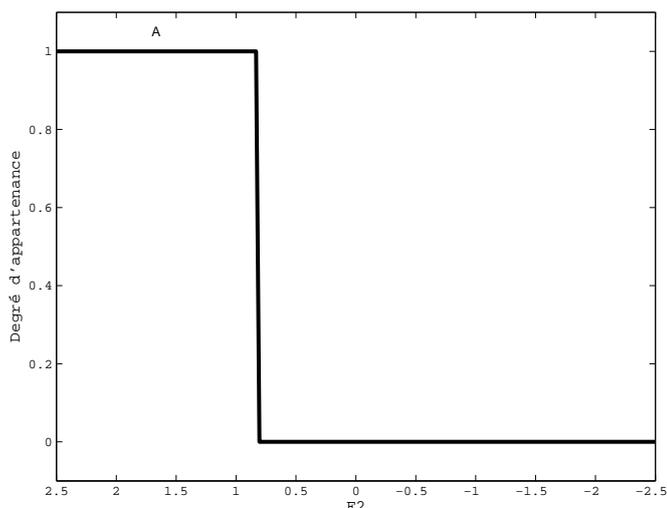


FIG. 2.2 – Illustration des limites de la logique binaire pour la catégorisation.

ment dit, cela impliquerait que la voyelle prototypique d'une des trois catégories de notre espace unidimensionnel est exclusivement représentée par une seule valeur de F2 et que le degré de ressemblance au prototype des autres x_n voyelles décroît de façon linéaire au fur et à mesure que $|Freq(x_i) - Freq(x_{proto})|$ croît. Ceci soulève deux problèmes : d'une part, étant donné le degré de liberté des valeurs formantiques, en partie imputable aux limites de résolution de l'appareil auditif (comme nous l'avons noté plus haut) on peut considérer intuitivement que, lorsque $|Freq(x_i) - Freq(x_{i+1})|$ est très faible, $\mu_A(x_i) \approx \mu_A(x_{i+1})$. D'autre part, étant donné l'effet attracteur des prototypes que nous avons mentionné plus haut, on peut concevoir qu'en deçà d'une certaine valeur de $|Freq(x_i) - Freq(x_{proto})|$, le degré d'appartenance à la catégorie décroît très lentement pour ensuite, passée cette valeur, diminuer de façon plus abrupte. Cette notion est illustrée dans la Figure 2.5 ; les fonctions triangulaires ont été remplacées par des sigmoïdes⁴⁸.

La théorie des ensembles flous permet de modéliser la correspondance entre les paramètres acoustiques et les catégories phonémiques prototypiques. Un intérêt supplémentaire réside dans le fait que le degré d'appartenance aux catégories reflète à bien des égards le jugement humain. Nous n'utilisons pas ce type de représentation dans ce qui suit, mais il

⁴⁸La courbe représentant le sous-ensemble C est obtenue grâce à la fonction `psigmf` de la Fuzzy Logic Toolbox du logiciel Matlab ; celles des sous-ensembles A et P utilisent la fonction `dsigmf` (voir le guide de l'utilisateur pour les détails).

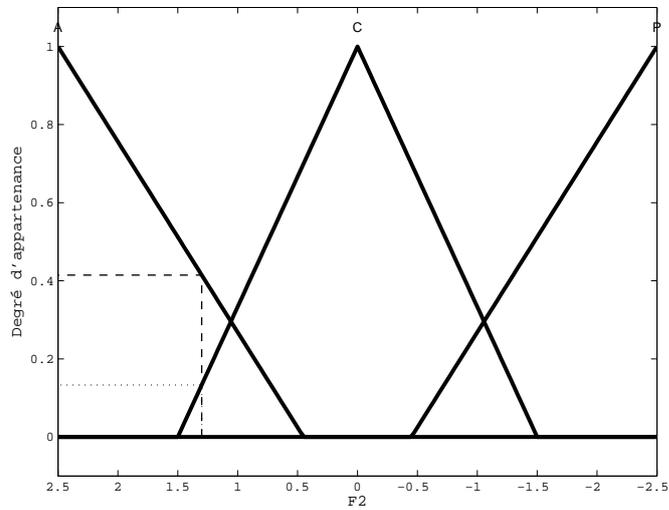


FIG. 2.3 – Illustration du degré d'appartenance d'un élément à un sous-ensemble dans le cadre de la théorie des ensembles flous.

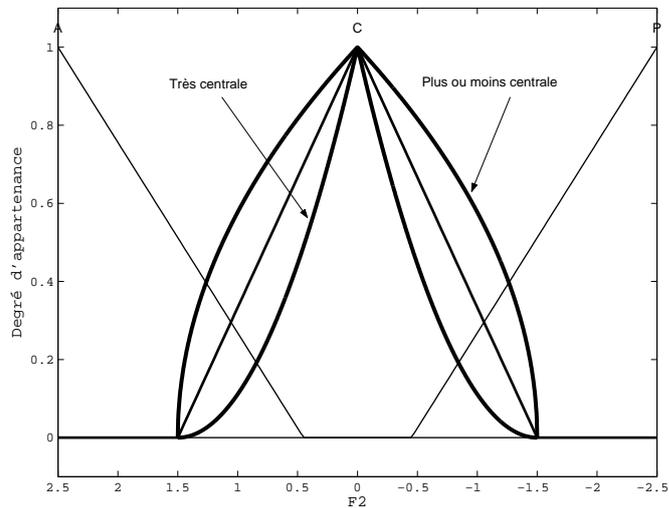


FIG. 2.4 – Illustration de la modification par les adverbes d'intensité de la forme des fonctions d'appartenances.

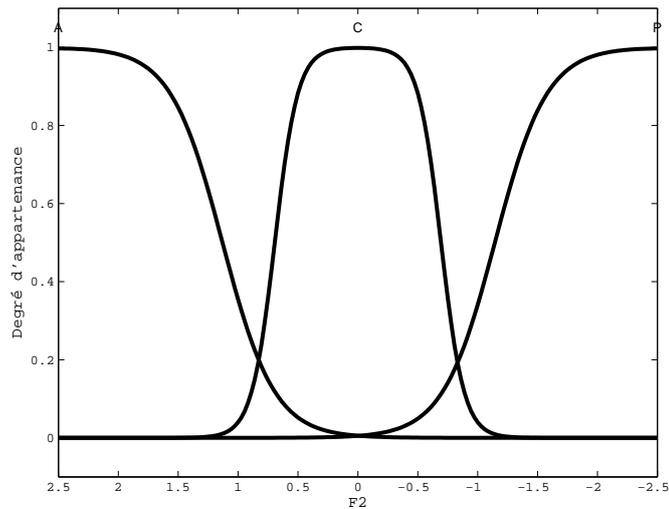


FIG. 2.5 – Fonctions d'appartenance conformes à la notion d'effet attracteur des prototypes.

sous-tend l'approche de la phonétique qui est développée dans nos travaux. Nous aurons cependant l'occasion d'y faire une brève allusion dans la Section 7.6 pour modéliser le degré d'appartenance du locuteur à un dialecte, ainsi que dans la Section 7.7, pour poser les bases d'un système expert flou susceptible d'améliorer les scores de classification et d'envisager cette classification sous un angle plus linguistique.

Il est en outre légitime de se demander si les dialectes constituent des sous-ensembles flous, et si, par exemple, la forme de ces sous-ensembles obéit au principe attracteur du prototype. Cela revient à se poser la question de savoir si les locuteurs possèdent des degrés d'appartenance à un dialecte, si les catégories les plus proches se chevauchent (au fur à mesure qu'un locuteur s'éloigne du prototype de la catégorie A, se rapproche-t-il de la catégorie B ?), etc. Nous fournissons une réponse très sommaire dans la section suivante, et nous opterons, sur des bases très hypothétiques, en faveur d'une représentation floue dans la Section 7.7. Une réponse plus empirique pourra être obtenue par le biais d'expériences de perception s'inscrivant dans le prolongement de nos travaux.

2.3.4 La notion de catégorie dialectale

Dans cette Section, nous abordons la question du nombre de classes dialectales et de leur forme à travers les études de perception, de contact linguistiques et de classification automatique.

Daniels ([Dan90], 27) rappelle l'aspect artificiel que comporte une étude d'identification des accents par des sujets humains :

Clearly, not all native speakers of English are interested in regional accents, so that the attempt to identify consciously a speaker's regional accent is not the everyday objective of a listener.

À partir de 40 échantillons d'accents anglais du monde entier d'une durée moyenne de 7 secondes, les 30 auditeurs anglophones natifs de l'étude de Daniels (majoritairement britanniques) sont capables, en moyenne, d'identifier avec précision environ 11 accents. Il était également demandé aux sujets de spécifier quels éléments leur avaient permis de reconnaître ces accents. Leurs réponses font apparaître deux types de critères : un trait linguistique ou phonétique particulier et la ressemblance avec l'accent d'un individu qu'ils connaissent, qu'il s'agisse d'une connaissance directe ou d'une célébrité. Parmi les traits phonétiques les plus mentionnés (qualité des voyelles et des consonnes, etc.) 15 % des réponses d'ordre linguistique concernent la rhoticité ou la réalisation du /r/. La Table 2.3 récapitule les accents des Îles Britanniques de l'étude de Daniels du mieux au plus difficilement identifié ; le score correspond au nombre d'identifications correctes. Dans ce tableau, les accents apparaissant sur la même ligne sont æquo ; pour des raisons d'homogénéité de la présentation, la terminologie anglaise a été conservée.

Malheureusement, la littérature qui traite de la perception des dialectes ne s'est, à notre connaissance, que rarement penchée sur la manière dont sont organisées les catégories dialectales dans l'esprit du locuteur natif. En 2004, Clopper et Pisoni ([CP04]) passent en revue les études marquantes dans le domaine de la perception du dialecte. Ces études abordent l'attitude des sujets vis-à-vis d'un échantillon de parole dialectale, leur capacité, parfois limitée, à identifier cet échantillon comme provenant d'une zone géographique

TAB. 2.3 – Accents des Îles Britanniques en fonction de leur rang d'identification.

Rang	Score sur 30	accent
1	28	RP, Cockney
2	23	(Adoptive) RP
3	19	Newcastle
4	14	South Wales
5	13	Leeds
6	11	Dublin
7	9	Glasgow
8	8	Belfast, Liverpool
9	7	North Yorkshire
10	6	Devon, Durham
11	5	North Wales
12	4	Norfolk, Lancashire
13	3	Cornwall, Birmingham, Black Country
14	2	Aberdeen, Ayrshire, Southern Ireland, Leicestershire, Bristol
15	1	Somerset
16	0	Hampshire, Oxford, Worcestershire

déterminée, ou encore les corrélats acoustiques des indices utilisés par les sujets dans leur tâche de classification. Aucune ne semble aborder la question de l'organisation des classes, qui est pourtant déterminante, ne serait-ce que pour pouvoir véritablement utiliser des termes tels que classes, classification ou catégories. L'étude de Clopper et Pisoni apporte un début de réponse concernant le nombre de classes que des sujets naïfs sont capables d'identifier : en anglais américain, il semblerait que trois régions dialectales, la Nouvelle Angleterre, le sud et l'ouest, puissent être identifiées de façon très fiable. Dans leur tâche de catégorisation avec un choix forcé à six classes, le taux de classification correcte est, certes, au-dessus du hasard, mais néanmoins très faible. Clopper et collègues ([CLP06], 567) rapportent, à partir d'une tâche de classification libre, que des sujets naïfs sont capables de distinguer 4 variétés régionales. Il est donc possible d'inférer que le nombre de catégories dialectales identifiables en anglais américain se situe entre 3 et 6. Cependant, les auteurs montrent ([CLP06], 569) que le taux de classification correcte pour un choix forcé avec 4 classes est d'environ 42 %, ce qui n'est pas très élevé.

Un concept intéressant des études de Clopper et collègues est celui de la saillance d'un

dialecte (notre traduction de « markedness » dans ce contexte). Clopper *et al.* ([CLP06], 567) le définissent ainsi :

markedness refers to the extent to which a given dialect has unique characteristics that differentiate it from [...] other dialects...

Les auteurs ajoutent que certains dialectes sont plus saillants que d'autres, et semblent définir cette saillance comme une fonction du nombre de traits phonétiques dont un dialecte a l'exclusivité. À partir de jugements de similitude entre locuteurs de 4 dialectes présentés par paires, [CLP06] concluent, par le biais du Multidimensional Scaling, que les deux dimensions les plus pertinentes utilisées dans les jugements de similitude sont la géographie (partition nord/sud) et la saillance. S'il était avéré – ce dont nous ne doutons pas – que ce concept puisse être transposé aux Îles Britanniques, ceci constituerait une piste pour la suite de nos travaux.

Les études sur les contacts entre dialectes offrent une ébauche de réponse à la question de la forme et du nombre de classes dialectales qu'il est possible de distinguer sur les Îles Britanniques. Le concept qui aborde indirectement ce thème est celui du nivellement dialectal (« dialect levelling »). Ce phénomène, souvent mentionné dans la littérature ([Bri02, Ker01, Ker03]), fait référence à l'érosion des formes dialectales marquées ; il peut conduire au regroupement de plusieurs variétés locales « minoritaires » en une seule. Certains considèrent d'ailleurs que « Estuary English » est une variété issue d'un nivellement ([FD99], 11, [Bri02], 63). Est-ce à dire qu'entre l'époque des données représentées dans le *Linguistic Atlas of England* ([OSW78]) et aujourd'hui, le nombre de classes qu'il est possible de mettre en évidence a diminué ? Kerswill apporte un début de réponse ([Ker01], 45) :

over one or more generations, families have abandoned these [traditional rural] dialects [and] there are fewer differences between ways of speaking in different parts of the country...

Ce commentaire peut-être interprété de deux manières : soit le nombre de classes reste le même mais la séparation entre classes repose sur des indices de plus en plus subtils ; soit le nombre de classes diminue effectivement. Le nivellement a été accompagné, dans un deuxième temps, par l'émergence de variétés urbaines très distinctes. Ces variétés urbaines subissent à leur tour un nivellement qui semble se manifester, pour Kerswill ([Ker01]),

par le biais d'une standardisation. On pourrait donc croire dans un premier temps que la bonne option consiste à admettre que le nombre de classes diminue. Cependant, la formation d'une nouvelle variété passe par l'apparition de nouveaux traits ([FD99]); il ne s'agit donc pas d'un regroupement de plusieurs variétés locales sur le critère du grand dénominateur de traits communs. Il ne s'agit pas non plus de l'adoption de la variété de prestige par toute la population ([Bri02]). Les études que nous avons consultées sur les contacts linguistiques, quoique informatives, ne permettent pas non plus de répondre à la question du nombre de catégories dialectales et des relations qu'elles entretiennent.

Pour ce qui est de la classification automatique, nous rapportons deux études qui concernent les dialectes de l'anglais. Barry et collègues ([BHN89]) proposent une approche phonétique de la question des accents dans la reconnaissance automatique de la parole, à partir de 4 accents régionaux : américain (A), écossais (E), anglais du nord (N) et anglais du sud (S). Leur méthode, dont l'un des objectifs est de maîtriser l'impact de la variation acoustique inter-locuteurs, consiste à obtenir une représentation de l'espace acoustique d'un locuteur en comparant les voyelles de son système entre elles. Ne plus comparer la production d'un locuteur à des valeurs acoustiques de référence (comme la moyenne d'un groupe) permet de s'affranchir en partie du problème de la normalisation. En outre, l'approche est intéressante car elle exploite explicitement la connaissance phonologique des systèmes vocaliques des dialectes. Par exemple, dans les quatre phrases test qu'utilisent les auteurs, une très faible distance entre <two> et <cooked> révèle la présence probable de la convergence FOOT-GOOSE, typique de E, et qu'on ne retrouve dans aucune des trois autres zones. Une proximité remarquable entre <cooked> et <butter> permettrait de mettre en évidence l'absence de scission FOOT-STRUT, diagnostique de N, et excluant les trois autres variétés. On pourrait même aller jusqu'à considérer – ce que les auteurs ne font pas – qu'une distance relativement importante entre la voyelle de <puddings> (qui serait prononcé [ʌ]) et celle de <cooked> trahirait un locuteur du nord pris en flagrant délit d'hypercorrection (voir Section 3.2.1.1). Les trois premiers formants vocaliques sont ensuite extraits, transformés en ERB, et une distance euclidienne est calculée entre chaque

TAB. 2.4 – Exemple de pondération pour la classification des dialectes.

Antécédent	Conséquent			
	S	N	E	A
$d(\text{father}, \text{after}) > d(\text{cat}, \text{after})$	-1	1	0	1
$d(\text{father}, \text{after}) < d(\text{cat}, \text{after})$	1	-1	0	-1
$d(\text{father}, \text{after}) = d(\text{cat}, \text{after})$	-1	0	1	-1

voyelle. Les comparaisons de distances subséquentes sont effectuées entre deux paires de voyelles ayant une voyelle en commun ; l'une des deux paires comporte deux phonèmes différents quel que soit le dialecte. Un système de pondération trivalent à partir de règles est ensuite mis en place : -1 indique que le trait phonético-phonologique en question joue en défaveur du dialecte concerné, 0 ne fait pencher la balance ni dans un sens ni dans l'autre, et 1 indique que le trait phonétique donne l'avantage au dialecte concerné. Nous avons réadapté un exemple de règle de la publication originale ([BHN89], 360) dans la Table 2.4. La première règle se lit de la façon suivante : si la distance entre la voyelle accentuée de <father> et celle de <after> est supérieure à la distance entre <cat> et <after>, alors, le poids en faveur de l'anglais du sud est de -1, etc. En effet, le système de l'anglais du sud n'a qu'un phonème pour PALM et BATH – i.e. <father> et <after> sont phonologiquement identiques – et deux phonèmes pour TRAP et BATH – <cat> et <after> sont phonologiquement différents. Or, si l'antécédent de cette première règle est vrai, alors le système vocalique concerné a vraisemblablement un seul phonème pour TRAP et BATH, ce qui n'est pas le cas de l'anglais du sud, mais correspond, par exemple, aux variétés du nord de l'Angleterre.

Sur les 58 locuteurs testés dans la tâche de classification (11 A, 6 E, 19 S et 20 N + une locutrice d'Irlande du Nord et un locuteur australien), 43 (73%) sont correctement classés, 9 ont été classés dans le mauvaise catégorie et 4 avaient des scores ex æquo. Parmi les caractéristiques de cette classification, on retiendra que certains locuteurs du nord de l'Angleterre avaient une distinction entre FOOT et STRUT, ce qui a influencé le taux de classification correcte pour N (13/20). Ce dernier point, que nous aborderons à la lumière

de nos données (Sections 7.5 et 7.7), est fondamental dans l'attribution d'une classe à un locuteur. En effet, faut-il considérer qu'un locuteur du nord de l'Angleterre qui a appris la distinction FOOT-STRUT se situe quelque part entre le pôle accent du nord typique/accent du sud typique ? Est-il possible, instructif ou souhaitable de lui attribuer un degré d'appartenance à chacun de ces pôles ? Comment quantifier d'ailleurs l'appartenance à ces pôles ? Une mesure telle que la différence entre le nombre de caractéristiques phonétiques que le locuteur partage avec chacun des pôles n'est pas adaptée puisque, en raison de degrés de stigmatisation variant selon le trait (l'absence de scission FOOT-STRUT est très marquée), toutes les caractéristiques n'ont pas le même impact.

Huckvale ([Huc04]) utilise la même idée que Barry et collègues ([BHN89]), selon laquelle les distances entre les voyelles d'un même locuteur constituent une représentation robuste de son espace acoustique, et surtout de son système. Il n'utilise cependant pas des règles explicites pour la pondération des distances entre voyelles comme celles rapportées dans la Table 2.4. Nous décrivons la méthode de calcul des distances ACCDIST dans la Section 7.2 car c'est celle que nous utilisons également. Dans la procédure, seuls les paramètres acoustiques diffèrent entre l'étude de Huckvale et la tâche de classification automatique que nous décrivons dans la Section 7.2. Ce dernier utilise les 20 phrases du corpus *Accents of the British Isles* ([DRBT04] ; nous employons le même corpus, mais une sous-partie différente). Les voyelles accentuées sont coupées en deux et des paramètres sont calculés sur chaque moitié. À partir des mesures des 4 premiers formants, Huckvale obtient des scores de classification correcte des locuteurs en dialectes entre environ 50 et 60 %, pour les conditions « tous sexes » et « même sexe », respectivement ; une dégradation importante (score de 35 %) intervenant dans la condition « autre sexe », où le locuteur testé et le modèle sont de sexes différents. L'utilisation de formants centrés-réduits (par locuteur) fait passer les scores à environ 72 et 73 %, la condition « autre sexe » ne donnant que 59 %. Il apparaît donc que la normalisation des formants, quoique importante dans l'amélioration des scores des conditions « tous sexes » et « même sexe », ne règle pas la question de la variation entre hommes et femmes. À partir de mesures de l'enveloppe

spectrale obtenues par le biais d'un banc de filtre auditif à 19 canaux⁴⁹, la méthode ACC-DIST (Section 7.2) est utilisée pour la classification. Les taux de classification correcte pour les conditions tous, même et autre sexe sont de 86,9 %, 87,2 % et 81,4 %. Cette étude démontre qu'il est possible d'obtenir une représentation spectrale de l'espace vocalique d'un locuteur qui préserve l'information dialectale tout en étant relativement insensible aux caractéristiques non phonétiques propres à ce locuteur (e.g. le sexe).

Définir un nombre de classes dialectales sur les Îles Britanniques *a priori* semble donc être une entreprise hasardeuse. L'étude de Daniels ([Dan90]) fait apparaître des taux d'identification faibles (Table 2.3); en effet, sur les 28 dialectes britanniques de l'étude, seuls 4 ont été identifiés avec précision par plus de la moitié des sujets. Il semble évident qu'une tâche d'identification de 40 accents du monde entier présente un haut niveau de difficulté; un paradigme de choix forcé révélerait très probablement une représentation des dialectes plus fine par les sujets. Les études ayant trait aux contacts linguistiques n'apportent pas de réponse tranchée quant au nombre de classes; elles mettent cependant en avant une instabilité des catégories dialectales qu'il convenait de mentionner. Enfin, les deux études de classification automatique ([BHN89, Huc04]) décrites prouvent que la machine peut classer de façon satisfaisante jusqu'à 14 dialectes; rappelons néanmoins que, comme c'est le cas dans nos travaux, la méthode présente certes l'avantage d'inclure la connaissance phonologique, mais elle reste dépendante du texte.

2.4 Cadre statistique

Nos travaux de recherche nous ont conduit à accorder une place prépondérante aux méthodes d'analyse quantitative. Au fil de notre recherche, le bien-fondé d'une dichotomie entre études expérimentales et études observationnelles s'est imposé de manière évidente (voir [vB02]). L'emploi de l'inférence statistique, domaine regroupant des tests

⁴⁹Chaque voyelle est définie par un vecteur de 40 paramètres : 19 paramètres pour chaque moitié d'une voyelle; ces données sont réduites par soustraction de la moyenne, et celle-ci est ajoutée comme paramètre supplémentaire.

d'hypothèse comme l'analyse de la variance (ANOVA), suppose l'élaboration préalable d'un plan d'expérience extrêmement minutieux visant à contrôler l'effet des facteurs que l'on ne souhaite pas prendre en compte. De plus, ces tests présupposent un caractère aléatoire, objectif qu'il n'est possible d'atteindre que par le biais de la mise en œuvre de techniques d'échantillonnage éprouvées, et ce critère n'est que très rarement satisfait en phonétique acoustique. En effet, la phonétique acoustique se place par nature dans le domaine de l'observation. Nous nous sommes donc tourné vers les méthodes descriptives, qui offrent bien plus de souplesse quant à leurs réquisits. Les diverses techniques utilisées sont brièvement présentées ci-après. Nous n'aurons recours au formalisme mathématique que lorsque celui-ci nous paraît indispensable.

2.4.1 Les mesures de distance

Calculer une distance dans un espace unidimensionnel, et *a fortiori* multidimensionnel, ne va pas de soi ; il convient que cette distance ait un sens. Ceci n'est pas anodin : par exemple, nous avons vu (Section 2.3.1) que calculer l'écart absolu en Hz entre deux fréquences ne correspond pas à la manière dont l'être humain perçoit la hauteur. Il est donc légitime de soulever la question d'une éventuelle transformation des variables et de la distorsion engendrée par la réduction de p paramètres à un score composite censé représenter la distance entre deux objets. Nous utiliserons principalement la distance dite « euclidienne », donnée dans l'Équation 2.6, où $x_{ik} - x_{jk}$ est la différence entre les points x_i et x_j mesurée sur le paramètre numéro k . Cette distance est un cas particulier de la distance de Minkowski (Équation 2.7). En pratique, nous nous sommes limité aux cas où $r = 2$ (distance euclidienne) et $r = 1$, distance de Manhattan (ou city block, ou encore taxicab). D'autres métriques existent ; un grand nombre d'entre elles est répertorié dans [ELL01] et [GL86].

$$d_{ij}^2 = \left(\sum_{k=1}^p (x_i^k - x_j^k)^2 \right)^{\frac{1}{2}} \quad (2.6)$$

$$d_{ij}^r = \left(\sum_{k=1}^p |x_i^k - x_j^k|^r \right)^{\frac{1}{r}} \quad (2.7)$$

Si le choix d'une mesure de distance est primordial, et est déterminé *in fine* par le type de données spécifique dont on dispose, le problème connexe de la pondération des variables (des dimensions ou paramètres) l'est tout autant. La distance de Minkowski attribue de fait une pondération équivalente à chacune des variables d'origine ; par conséquent, cette mesure est sensible aux différences d'échelles. En outre, plus le nombre r de l'Équation 2.7 est élevé, plus important est le poids affecté aux variables dont la différence absolue entre deux objets est grande. Le fait de centrer-réduire les variables originales revient à leur accorder un poids identique implicitement ; mais il est également possible de pondérer explicitement. À ce propos, [ELL01], 48-49, notent qu'on peut considérer que la pertinence d'une variable est inversement proportionnelle à l'amplitude de sa variation. En réalité, cette approche risque de masquer la variation inter-classe, et c'est précisément cette dernière qui devrait être pondérée à la hausse dans une tâche de classification. Une solution consiste, surtout pour une tâche supervisée, à donner plus de poids aux dimensions pour lesquelles le rapport de la variation inter-classe sur la variation intra-classe est maximal. Dans l'Équation 2.8, la pondération w est incluse dans le calcul de la distance de Minkowski.

$$d_{ij}^{rw} = \left(\sum_{k=1}^p w_k |x_i^k - x_j^k|^r \right)^{\frac{1}{r}} \quad (2.8)$$

Nous avons par exemple tenté de déterminer empiriquement les poids relatifs de la durée et des paramètres spectraux dans [FP06] : en utilisant la méthode décrite dans la Section 7.2, nous avons fait varier un facteur de pondération w de 0 à 1 pour la durée, et avons multiplié à chaque itération les paramètres spectraux par $1 - w$. La pondération optimale correspondait au taux de classification le plus élevé. Mentionnons également le programme Optimal Variable Weighting ([ML01]). Ce programme s'efforce de trouver des poids optimaux pour chaque variable afin que la distance entre classes soit maximale, et

celle à l'intérieur des classes, minimale. Ce programme a été utilisé sans succès sur nos données, nous nous en tiendrons donc à cette rapide allusion.

2.4.2 La classification hiérarchique

La classification hiérarchique est une technique qui vise à obtenir une représentation arborée, le dendrogramme, ou arbre hiérarchique, d'un ensemble d'individus. Nous avons utilisé cette technique dans la Section 7.3 ; les arbres représentés sont binaires en cela que les individus y sont regroupés deux à deux. Détaillons à présent la méthode employée. Dans un premier temps, la procédure consiste à obtenir les distances entre les individus à classer pris deux à deux ; ce point vient d'être abordé dans la Section 2.4.1. Puis, à partir de la partition de départ, où chaque individu constitue une classe, les deux individus les plus proches sont réunis. Les distances entre classes sont alors recalculées. C'est ici qu'intervient la notion de critère d'agrégation ; la distance entre la classe qui vient d'être formée et les autres classes peut être exprimée de différentes façons (un exemple très clair du fonctionnement des méthodes agglomératives est donné dans [ELL01], 57-58). Nous avons déterminé empiriquement que la méthode du lien moyen⁵⁰ impliquait la distorsion moyenne la plus faible entre les distances originales et les distances telles qu'elles sont représentées dans l'arbre. La méthode du lien moyen consiste à définir la distance entre deux classes comme la moyenne des distances entre chaque paire d'individus n'appartenant pas à la même classe. L'opération est ainsi répétée jusqu'à ce que toutes les classes soient réunies en une seule (pour d'autres méthodes, voir [ELL01] 55-89). Un diagnostic de la distorsion entre la matrice de distances originale et l'arbre final peut être obtenu au moyen du coefficient de corrélation cophénétique. Il consiste à calculer une corrélation linéaire classique entre la matrice de distances originales et la matrice cophénétique, i.e. la matrice des distances entre individus deux à deux, ces distances étant données par la hauteur à laquelle deux individus sont regroupés dans l'arbre.

⁵⁰ « Average linkage » ou encore « unweighted pair-group method using the average ».

2.4.3 L'analyse en composantes principales

S'il est impossible de représenter graphiquement des individus (voyelles, locuteurs, etc.) dans un espace à plus de deux ou trois dimensions, ceci ne doit en aucun cas constituer un frein à l'inclusion d'autres variables. D'ailleurs, il est toujours possible de représenter ces individus par leurs coordonnées sur toutes les dimensions prises deux à deux (ce que certains logiciels nomment « scatterplot matrix ». Voir, par exemple, la Figure 5.6 dans la Section 5.4.). Ceci constitue un excellent moyen de visualiser les variables corrélées ; mais en représentant autant de nuages de points sur des graphes bidimensionnels qu'il y a de combinaisons de variables prises deux à deux, la figure finale, quoique instructive, n'est pas synthétique. En effet, outre la taille importante de cette figure, les variables corrélées fournissent une information redondante (puisque en partie prévisible à partir des autres variables) car une multiplicité de variables n'obéissent souvent qu'à un nombre restreint de principes sous-jacents. L'analyse en composantes principales permet de réduire le nombre de variables initiales (corrélées) à de nouvelles variables orthogonales dans un espace réduit. On calcule les vecteurs propres et valeurs propres de la matrice de covariance des paramètres centrés-réduits. Ces vecteurs propres, orthogonaux, sont ensuite ordonnés en fonction de la valeur propre qui est associée à chacun. Cette valeur traduit la proportion de variation dans l'espace des paramètres originaux qui est conservée si on projette les données sur le vecteur propre correspondant. La première composante principale est définie par le vecteur propre associé à la valeur propre la plus élevée, et ainsi de suite. Pour plus de détails, voir [WFH86], 273-290, [Sap90], 159-186, [Lag98], *passim*.

2.4.4 L'analyse linéaire discriminante

L'analyse linéaire discriminante est une technique de classification avec apprentissage supervisé. Là où l'analyse en composantes principales vise à décrire au mieux la dispersion des données dans un espace à dimension réduite (sans tenir compte de leur éventuelle organisation en classes), l'analyse discriminante consiste à déterminer le meilleur hyper-

plan séparant les différentes classes d'individus. Pour ce faire, on cherche à maximiser le rapport de la variation inter-classe (e.g. la matrice de covariance calculée sur l'ensemble des données) sur la variation intra-classe (e.g. la moyenne des matrices de covariances de chaque classe). Les données originales sont alors projetées dans l'espace des vecteurs propres contenant le plus d'information discriminante. Pour un aperçu de la méthode et des détails plus techniques, voir [Lag98], 113-125, [Sap90], 403-428, [DHS01], 114-124 et 215-281, pour une approche plus exhaustive des fonctions discriminantes ; dans ce cadre plus général, voir également [DMS⁺02].

2.4.5 « K-means clustering » et qualité d'une partition

Le « k-means clustering » est une méthode de classification non supervisée ([DHS01], 526-528, [Nab02], 101-104) qui, dans sa forme élémentaire, consiste à localiser c vecteurs moyens (ou centres de classes) – c étant déterminé à l'avance – dans un espace multidimensionnel, pour lesquels la somme des distances de chaque individu au centre le plus proche est minimale. La fonction `kmeans` du logiciel Matlab prévoit une première phase pendant laquelle, à chaque itération, les points sont associés au centre le plus proche, qui est ensuite recalculé jusqu'à ce que les coordonnées des centres soient relativement stables. Dans un second temps, à chaque itération, chaque point est associé à un centre différent si cela permet une diminution de la somme des distances, et les centres sont mis à jour.

Parmi les nombreuses techniques permettant de juger la qualité d'une partition ([ELL01, CH06]), nous avons utilisé la valeur de silhouette, dont le calcul est donné dans l'Équation 2.9. s_j est la valeur de silhouette de l'objet j , a_{pj} représente la distance moyenne de l'objet j aux autres objets appartenant à la classe p . Si d_{qj} est la distance moyenne entre l'objet j et tous les objets appartenant à une classe q , $q \neq p$, b_{pj} est la valeur d_{qj} minimale calculée pour $q = 1 \dots c$, $q \neq p$ (c étant le nombre de classes). Autrement dit, b_{pj} mesure la dissimilarité de l'objet j par rapport à la classe (autre que la sienne) la plus proche. La qualité de la partition est définie par S , la moyenne des s_j pour $j = 1 \dots N$; dans le cas où la classe p est un singleton, $s_j = 0$. L'objectif consiste donc, si l'on cherche à déterminer

le nombre optimal de classes dans une tâche non supervisée, à trouver la partition pour laquelle S est maximal. C'est ce que nous avons fait, par exemple, dans la Section 5.6.

$$s_j = \frac{b_{pj} - a_{pj}}{\max\{a_{pj}, b_{pj}\}} \quad (2.9)$$

La valeur est bornée entre -1 et 1 ; plus la valeur est élevée, plus j est distant de la classe q la plus proche; une valeur de 0 caractérise les points qui ne peuvent pas être clairement attribués à une classe, et une valeur proche de -1 indique que le point en question n'appartient vraisemblablement pas à la bonne classe. Everitt *et al.* ([ELL01], 105) ajoutent qu'une valeur de silhouette moyenne inférieure à $0,2$ révèle une absence totale de structure en classes, et une valeur supérieure à $0,5$ traduit une bonne classification.

2.4.6 La régression

Nous avons employé la technique de la régression dans le but de décrire la trajectoire des formants des voyelles. La régression permet d'exprimer une relation entre plusieurs variables. Dans notre cas, on recherche les paramètres d'une fonction affine (Équation 2.10) passant, pour chaque voyelle, au plus près des valeurs formantiques dans le plan temps (variable indépendante)/fréquence (variable dépendante).

$$f(x) = a + bx \quad (2.10)$$

« Passer au plus près » se décline en deux versions : calcul des moindres carrés ordinaires et calcul des moindres carrés pondérés. Le premier attribue un poids équivalent à chaque point, le second pondère à la baisse les points apparemment déviants. Le calcul de la pente, b , par la méthode des moindres carrés ordinaires est donné dans l'Équation 2.11, où n , dans notre cas, est le nombre de valeurs formantiques, x est la valeur de temps et y , la valeur fréquentielle.

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.11)$$

La régression utilisant les moindres carrés pondérés est une procédure itérative donnant une estimation robuste à la présence de valeurs déviantes. On peut considérer que les moindres carrés ordinaires sont un cas particulier des moindres carrés pondérés où le poids de chaque point est équivalent. L'algorithme estime dans un premier temps la droite des moindres carrés (Équation 2.11), puis, l'écart entre chaque point observé et cette droite est mesuré, et chaque point observé reçoit un poids inversement proportionnel à sa distance avec la droite. Une nouvelle droite est calculée en tenant compte de ces poids, et ainsi de suite jusqu'à ce qu'il n'y ait plus de modification importante de la droite (voir [WW91]). La fonction de Matlab utilisée pour ce type de régression est `robustfit`; et pour la régression simple, `regress`.

Comme nous le verrons dans la Section 6.2, les formants peuvent décrire une trajectoire courbe, auquel cas ajuster une droite aux points dans l'espace temps/fréquence devient aberrant. Dans ces cas-là, l'estimation des paramètres d'une fonction polynôme sera réalisée avec la fonction `polyfit` du logiciel Matlab. L'Équation 2.12 donne la forme d'une fonction polynôme, où n est le degré du polynôme et $b_i \dots_n$, les coefficients. L'estimation de ces coefficients se fait, comme pour la régression ordinaire, avec le critère des moindres carrés. Ce qu'il est, en pratique, important de retenir pour nos besoins, c'est qu'un polynôme d'ordre n ne peut pas générer une courbe comportant plus de $n - 2$ points d'inflexion.

$$f(x) = a + b_1x + b_2x^2 + \dots + b_nx^n \quad (2.12)$$

Les autres techniques utilisées dans ce travail seront présentées au fur et à mesure que nous les rencontrerons.

Après avoir introduit dans ce chapitre les notions fondamentales de notre étude, nous passons en revue les caractéristiques phonético-phonologiques des dialectes de l'anglais des Îles Britanniques dans le Chapitre 3.

Chapitre 3

Situation géolinguistique et anglais standard

Ce chapitre a pour but de faire le point sur les connaissances actuelles concernant les aspects phonético-phonologiques des dialectes des Îles Britanniques. Nous nous concentrons sur le domaine segmental ; certaines caractéristiques suprasegmentales seront néanmoins abordées plus bas dans le Chapitre 5. Dans ce qui suit, les particularités des systèmes vocaliques constituent l'essentiel du propos. Cependant, bien que nous n'abordions pas le cas des consonnes dans notre étude empirique (Partie II), nous avons tenu à mentionner les traits consonantiques les plus saillants des zones géographiques passées en revue. Ce chapitre s'articule en deux temps : la Section 3.1 traite de la notion d'anglais standard et fournit une description détaillée du système vocalique de cette variété ; et la Section 3.2 dresse un inventaire des traits phonético-phonologiques diagnostiques des principales zones dialectales des Îles Britanniques.

3.1 L'anglais standard

Une description exhaustive de l'histoire de l'anglais standard et une rétrospective des points de vue des experts sur la question seraient pour le moins rebutantes. Nous nous

contenterons de définir ce que recouvre ce concept et de fournir une description phonétophonologique détaillée de cette variété. Ce dernier volet est capital puisque c'est à l'aune de cette description que seront examinés les dialectes de notre corpus.

Le terme « Received Pronunciation » (RP) a longtemps été employé, notamment dans le domaine de l'enseignement de l'anglais comme langue étrangère. Il désigne sans ambiguïté un type de prononciation (et ne fait donc aucune référence au lexique ou à la syntaxe), qui, dans sa forme la plus conservatrice, serait à peine parlée par 3 à 5 % de la population anglaise ([HTW05], 3). À notre sens, le terme est très lourdement connoté puisque, bien que Wells ([Wel90]) tente de l'associer à un effort descriptiviste, c'est bien une norme prescrite qui est évoquée par RP. Ce terme a été employé pour la première fois en 1869 par le philologue Alexander Ellis, puis remis au goût du jour en 1926 par Daniel Jones, figure emblématique de la phonétique moderne ([Upt04], 217). Ce concept a été désigné depuis le début du 20^e siècle par les termes « Public School Pronunciation », « BBC English » (c'est ce terme qu'utilisent Roach et collègues [Jon03]), ou encore « Queen's English », entre autres, ce qui traduit admirablement le caractère hautement élitiste souvent associé à cette prononciation. C'est la raison pour laquelle nous avons abandonné le terme. Wells ([Wel82], 279) parlait de « Varieties of RP » ; c'est précisément cette acception plus large (mais cependant limitée aux variétés « natives ») que recouvre notre conception de l'anglais standard.

Lorsque l'on parle d'anglais standard⁵¹ sur les Îles Britanniques, on s'empresse de signaler que cette variété n'a aucune connotation régionale, et qu'il s'agit donc d'un « sociolecte ». Certes. Mais il n'en demeure pas moins que les traits de prononciation de l'anglais standard sont typiques du sud-est de l'Angleterre. D'ailleurs, l'histoire de l'anglais standard commence dans le sud-est, comme le rappelle Honey ([Hon97], 93 *sqq.*), et Nolan ([Nol98]) considère que ce qu'il nomme « RP » ou « near-RP » est lié aux accents du sud-est⁵². Après avoir identifié les principaux facteurs qui ont conduit à l'émergence des

⁵¹Ce terme fera exclusivement référence à l'aspect phonétophonologique.

⁵²Rappelons aussi que, dans l'étude de Evans et Iverson ([EI07]) décrite dans la Section 2.3.3, plus les voyelles d'un locuteur originaire du nord se rapprochent de l'anglais standard, plus ce locuteur est jugé proche du sud (sur un continuum nord/sud) par des phonéticiens.

variétés standards en Europe occidentale dès le 15^e siècle⁵³, l'auteur rappelle que l'anglais standard est associé aux zones de Londres, Cambridge et Oxford. Sa forme écrite trouve comme vecteur de diffusion l'imprimerie naissante⁵⁴ et sa forme orale trouvera un moyen de propagation bien plus tard par l'intermédiaire du réseau des « public schools ». D'après Honey ([Hon97], 94), la radio dès 1922, puis la télévision, ont pris le relais.

Dans une tentative de définition de ce que nous appellerons « anglais standard », nous donnerons, en guise de préliminaire, l'avertissement de Wells ([Wel91]) :

it is important not to lose sight of the fact that accents, and more generally varieties of a language, are not objective entities so much as mental constructs. As such, their definition and circumscription are open to considerable debate.

Parmi les définitions possibles de ce que certains nomment RP, Wells ([Wel91]) se demande, sur un ton légèrement ironique, si l'on ne ferait pas mieux de considérer qu'il ne s'agit de rien d'autre qu'une norme codifiée à l'usage des apprenants étrangers. Si cette idée est effectivement étroitement liée au concept que nous tentons de définir, il n'en demeure pas moins qu'elle ne constitue pas une définition. En effet seuls les critères présidant à l'élaboration de cette norme sont susceptibles de posséder un caractère définitoire d'une quelconque utilité pour la suite de notre propos, l'utilisation qui est faite de cette norme n'étant que secondaire ici. En d'autres termes, s'il est vrai que la description d'une norme de prononciation de l'anglais britannique a été en grande partie motivée par l'enseignement de l'anglais comme langue étrangère⁵⁵, cela ne doit pas occulter le fait que le choix des traits de prononciation est motivé par une réalité linguistique déterminante dans la société britannique. Remarquons cependant que l'anglais standard auquel nous faisons référence n'est pas une norme de prestige pour toutes les Îles Britanniques ; il l'est en Angleterre, mais, concernant les autres pays, Foulkes et Docherty ([FD99], 11) observent que : « RP has little cachet for many speakers, particularly those outside of England [and is] often regarded with 'hostility' in Glasgow. »

⁵³Avènement de l'imprimerie, des états-nations dotés de systèmes éducatifs au niveau national ; mobilité géographique et sociale associée à l'industrialisation et à l'urbanisation.

⁵⁴Caxton installe ses presses à Westminster en 1476.

⁵⁵Nous en voulons pour preuve l'habitude prise dans les ouvrages semblables à [Jon67] et [Gim80] de consacrer une place importante aux conseils pour les apprenants.

Quel que soit le terme utilisé pour désigner cette norme de prononciation, il va de soi que cette dernière est profondément instable en diachronie ; l'anglais standard de Daniel Jones n'est pas celui de John Wells. Il apparaît dès lors capital de comprendre qu'une description phonétique n'est rien d'autre qu'un instantané ; un anglais standard ne peut être défini qu'à travers l'époque pendant laquelle il a cours. Il est la marque d'une communauté, et le fait que cette communauté disparaisse, ou que ses frontières deviennent floues (comme ce fut le cas de classes sociales en Angleterre ces quarante dernières années) est lourd de conséquences.

Nous allons certes définir l'anglais standard en énumérant les traits de prononciation qui le caractérisent, mais une caractérisation plus parcimonieuse consisterait à le définir négativement ; nous dirions alors que l'anglais standard est dénué de traits de prononciation qui trahiraient l'origine géographique ou sociale d'un locuteur.

3.1.1 Système vocalique de l'anglais standard

Nous nous appuyerons sur plusieurs ouvrages ([Jon67], [Wel82], [Wel90], [Jon03], [Upt04]) pour une description phonétique traditionnelle du système vocalique de l'anglais standard. Les commentaires ayant trait à l'évolution acoustique des voyelles sur les quarante dernières années se fondent sur [Wel62], [Hen83], [Det97] et [HM05].

La description est structurée autour des mots-clés représentant les ensembles lexicaux popularisés par Wells ([Wel82]). L'ordre d'apparition de ces ensembles est identique à celui de Upton ([Upt04]) ; les voyelles brèves précèdent les voyelles longues, et l'exposé s'achève avec les voyelles réduites. Dans les Figures 3.1, 3.2 et 3.3, nous avons adapté les trapèzes vocaliques de [Jon03], viii-ix. Il est à noter, au vu de la présentation phonétique détaillée qui suit, que ces trapèzes présentent une prononciation assez conservatrice.

KIT a un timbre mi-fermé, proche de [ɪ], quoique plus postérieur. En position inaccentuée, cette voyelle est encore utilisée dans les morphèmes <-ed> et <-es> (certains dialectes ont [ə], voir 3.2.1.4). En revanche, [ɪ] a largement cédé la place à [ə] dans les autres contextes non-accentués où la réduction vocalique opère. Wells ([Wel91]) retrace

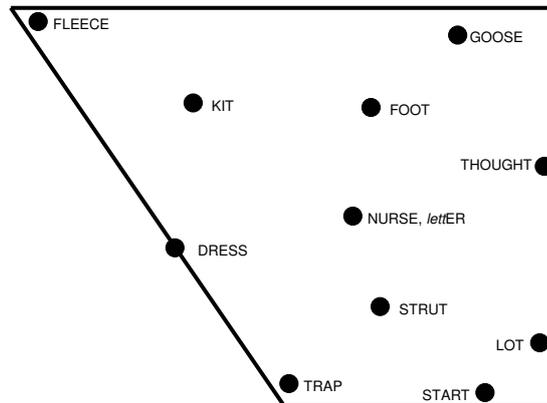


FIG. 3.1 – Monophthongues de l'anglais standard (adapté de [Jon03]).

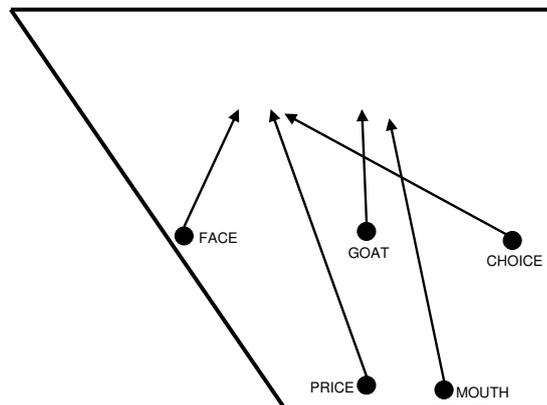


FIG. 3.2 – Diphtongues fermantes de l'anglais standard (adapté de [Jon03]).

brèvement l'histoire de l'acceptation de ce [ə] dans les descriptions de l'anglais standard : jusqu'en 1967, Jones continuait de penser que la prononciation [-ləsnəs] dans <carelessness> ne méritait pas d'être incluse dans son dictionnaire. Dix ans plus tard, dans le même dictionnaire, révisé par Gimson, ce dernier jugeait cette variante comme acceptable. Treize ans plus tard, Wells, dans son dictionnaire ([Wel90]), en fait la variante principale, reléguant [-ɪsnɪs] en seconde position. Ce choix est toujours d'actualité puisque c'est celui qui a été retenu dans la 16^e édition du dictionnaire de Jones ([Jon03]). Wells ([Wel99]) a étudié ce phénomène des voyelles faibles pré-consonantiques par le biais d'un questionnaire

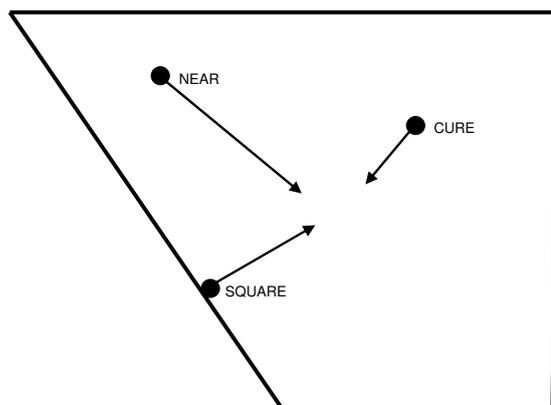


FIG. 3.3 – Diphtongues centripètes de l’anglais standard (adapté de [Jon03]).

auprès d’environ 2000 personnes. Les résultats font apparaître que 76 % des personnes interrogées préfèrent [ə]. En termes de système, on peut supposer (comme le fait [Gim80], 105) qu’une certaine pression existe encore contre une neutralisation complète de /ɪ/ et /ə/ en syllabe inaccentuée : <offices> et <officers> peuvent être distingués si les voyelles finales de ces deux mots sont prononcées respectivement /ɪ/ et /ə/. Nuancions toutefois : dans l’histoire de l’anglais, si ces risques d’homophonie ont peut-être retardé certaines neutralisations et convergences phonémiques, ils n’ont jamais constitué une entrave suffisante. Nolan ([Nol98], 80) considère que ce changement s’inscrit dans une tendance générale de réduction des voyelles non accentuées qui n’avaient pas encore schwa ; d’après l’auteur, le changement est plus avancé en anglais australien, et nous savons par ailleurs qu’une étude récente sur l’anglais américain ([FJ07]) prouve l’existence d’un contraste dans la voyelle finale des (rares) paires du type <Rosa’s> vs <roses>⁵⁶.

DRESS a une voyelle de type [ɛ]. Une forme plus conservatrice de l’accent présente une voyelle plus fermée. Upton ([Upt04], 222) note l’ouverture accrue relativement récente de cette voyelle, qui participe d’une tendance affectant les voyelles antérieures brèves. L’ouverture de DRESS est visible dans l’étude acoustique de Hawkins et Midgley ([HM05]) ; son évolution semble toutefois moins flagrante que celles de TRAP, FOOT et GOOSE (voir

⁵⁶Les paires du type <offices>/<officers> ne permettent pas de tester ce type d’oppositions puisque la plupart des accents américains sont rhotiques.

infra).

La voyelle de TRAP est traditionnellement représentée par le symbole æ dans les dictionnaires de prononciation. Or il semblerait que la tendance récente à l'ouverture des voyelles antérieures brèves mentionnée plus haut soit particulièrement saillante pour celle de TRAP. Il apparaîtrait donc que la voyelle de TRAP en anglais standard ressemble de nos jours à celle des dialectes septentrionaux de l'Angleterre : [a]. Cette observation est largement corroborée par l'étude acoustique transversale de Hawkins et Midgley ([HM05]) : F1 passe de 644 Hz, pour les locuteurs âgés de 65 ans ou plus, à 917 Hz pour les 20-25 ans. Cette ouverture s'accompagne d'une postériorisation du timbre qui n'est en grande partie que le résultat de contraintes articulatoires et de propriétés acoustiques des voyelles bien connues. En effet, et ceci est particulièrement flagrant pour les voyelles antérieures, le degré de fermeture est positivement corrélé au degré d'antériorité⁵⁷. Wells ([Wel82], 291-292) remarquait déjà le phénomène d'ouverture de TRAP en citant un article de journal de 1978 dans lequel un journaliste vilipendait cette tendance qu'il considérait comme une conséquence des années 1960, époque à laquelle, précise le journaliste, « working class became beautiful ». Wells poursuit son exposé en notant que cette ouverture de TRAP conduit à une convergence potentielle (et partielle chez certains locuteurs) avec la voyelle de STRUT qui, comme nous le verrons plus loin, est particulièrement ouverte en anglais standard (ceci est toujours d'actualité puisque [Upt04], 222, reprend le même exposé). Pour [Wel91], cette ouverture de la voyelle de TRAP en anglais standard peut s'expliquer comme une volonté de s'éloigner de la réalisation en Cockney, [ɛ]. Le fait que ce « flat A » s'apparente à celui des dialectes du nord n'est qu'une conséquence de ce phénomène pour [Wel91] alors que [Upt04], 222-223, semble penser que ce sont précisément les locuteurs originaires du nord qui ont fait pénétrer cette réalisation dans l'anglais standard. Fudge ([Fud77]) traite la question de la voyelle de TRAP en référence à sa propre prononciation. Il note que dans le sud de l'Angleterre et en RP, cette voyelle est parfois allongée, et tente

⁵⁷Le trapèze vocalique actuel, qu'il soit d'inspiration articulatoire ou acoustique, présente une partie antérieure oblique, illustrant l'impossibilité physiologique de conserver le même degré d'antériorité au fur et à mesure que l'aperture s'accroît.

de détecter les facteurs de cet allongement. Cette variante allongée semble ne pas obéir à des règles bien précises, mais Fudge envisage la possibilité de deux phonèmes distincts. Le degré de formalité constitue un facteur probable ; pour l'anecdote, l'auteur offre au lecteur un moyen mnémotechnique avec les mots br[æ]ndy et sh[æ:]ndy. La longueur de la voyelle est prévisible par le degré de formalité des situations associées à la consommation de ces deux boissons (cognac et panaché, respectivement) et si cela ne suffit pas, il faut se souvenir que le brandy est un « short drink », et le shandy, un « long drink » . . .

La voyelle de LOT est postérieure, de degré d'aperture maximum (ou très légèrement fermée), et arrondie. Cette voyelle est brève, contrairement à son équivalent dans le système de l'anglais américain standard, où <dog> est volontiers prononcé [dɔ:ɡ].

Upton ([Upt04], 222) décrit la voyelle de STRUT de la manière suivante : elle est centralisée et légèrement plus ouverte que [ʌ]. Il ajoute que pour un bon nombre de locuteurs, STRUT s'apparente davantage à un [a] légèrement plus fermé et centralisé que la voyelle désignée par ce symbole ([ɐ] serait peut-être tout à fait adéquat). Nolan ([Nol98], 81) inclut cette voyelle dans les changements récents⁵⁸ en RP.

FOOT a une voyelle mi-fermée, antériorisée et arrondie. L'antériorisation de ce timbre est attestée dans les données de [HM05], et Henton ([Hen83], 365) notait déjà en 1983 que cette centralisation de FOOT et GOOSE était le phénomène le plus fréquemment mentionné dans la littérature ayant trait à l'évolution du système de l'anglais standard.

Traditionnellement, la voyelle de BATH est ouverte, postérieure et longue ([Jon67], 73-77). Cependant, [Upt04], 223, constate une évolution imputable à deux phénomènes distincts. D'abord, la voyelle de BATH tend à être de plus en plus brève et de plus en plus centrale sur l'axe antérieur/postérieur ; la variante longue et postérieure étant progressivement associée à un accent affecté. Ensuite, le [a] caractéristique des accents du nord semble avoir été accepté dans le système de l'anglais standard, tant et si bien, qu'il n'est pas rare d'entendre des locuteurs du nord (i.e. dont l'accent d'origine n'a pas de distinction TRAP-BATH) présentant toutes les caractéristiques de l'anglais standard utiliser [a] pour

⁵⁸Ici, « récent » qualifie les tendances de prononciation apparues au 20^e siècle.

l'ensemble BATH, voire [a] et [ɑ:] de façon interchangeable. Ceci rejoint un commentaire de Wells dans une interview au *Guardian* de 1999 faisant suite aux sondages de préférences de prononciation pour la deuxième édition du *Longman Pronouncing Dictionary* ([Byr99]) : « There is a northern pronunciation increasing among the young which the word 'chance' clearly shows ».

L'étude acoustique de Hawkins et Midgley ([HM05]) ne permet pas de commenter le cas de BATH. Des trois ensembles lexicaux susceptibles de contenir un phonème /ɑ:/ (BATH, START et PALM), seul START est présent dans les mots-test de l'étude ; or c'est précisément celui qui tend à préserver un timbre postérieur non stigmatisé.

La voyelle de CLOTH est aujourd'hui identique à celle de LOT. Dans son fameux poème pour apprenants, Trénité⁵⁹ avait écrit :

Mind the o of off and often
Which may be pronounced as orphan

Cette variante ([ɔ:]) s'avère aujourd'hui totalement archaïque en anglais standard ; Upton ([Upt04]), 224 la juge même « risible ». Gimson ([Gim80], 115) explique l'abandon de la variante longue et plus fermée de ce phonème par sa similitude avec la variante utilisée en Cockney.

D'après [Upt04], 224, la voyelle de NURSE est très variable, de mi-ouverte à mi-fermée ; cette caractéristique est également mentionnée chez [Jon67], 88. La variante la plus ouverte était associée à U-RP⁶⁰ d'après Wells ([Wel82], 281).

FLEECE a une voyelle longue, antérieure, fermée. Elle est souvent légèrement diphtonguée : [ii]. Jones [Jon67], 65-66, conseillait aux apprenants de produire une monophthongue. Déjà en 1980, Gimson ([Gim80], 102-103) remarquait que parmi les locuteurs de RP, une légère diphtongue était plus courante qu'une véritable monophthongue. Il ajoutait que produire une monophthongue lorsque la voyelle de FLEECE était en position finale relevait d'une prononciation « over-cultivated ». Il recommandait aux apprenants de tenter de produire une diphtongue, mais seulement avec précaution.

⁵⁹<http://www.hep.wisc.edu/~jnb/charivarius.html> ; site consulté le 8 avril 2008.

⁶⁰Une abréviation fréquente pour « upper » ou « upper-crust » RP.

La voyelle de FACE en anglais standard est une diphtongue fermante dont les points de départ et d'arrivée sont respectivement [e] et [ɪ]. Le timbre de la première partie de la diphtongue varie entre [ɛ] et [e] ([Gim80], 129). La limite de fermeture est imposée sur le plan articulaire; la limite d'ouverture est, elle, sociolinguistiquement déterminée : un premier élément de type [æ] ou [a] est typique de certains accents populaires, notamment le Cockney. Un mouvement articulaire ample étant, comme nous venons de le voir, associé à des variétés non prestigieuses, certains locuteurs marquent leur distance avec ces variétés stigmatisées en utilisant un premier élément très fermé, occasionnant un mouvement articulaire restreint.

La voyelle de PALM est longue, ouverte et quelque peu centralisée. Plus elle est postérieure, plus elle évoque ce que Upton [Upt04], 224, nomme « Refined RP ». Cet ensemble, décrit par Wells ([Wel82], 142-144), est, aux dires de ce dernier, plutôt difficile à cerner avec précision. Il comprend peu de mots courants (e.g. *father*, *tomato*, etc.) et est essentiellement constitué d'emprunts récents à d'autres langues, le timbre postérieur soulignant le caractère exotique du mot en question (e.g. *sonata*, *rajah*, *façade*, etc.).

La voyelle de THOUGHT est généralement notée [ɔ:] ([Wel90, Jon03, Upt04, OS06]). Elle s'est fermée depuis le début du 20^e siècle; Nolan ([Nol98], 82) considère qu'il s'agit, dans une certaine mesure, d'un prolongement du Grand Changement Vocalique; il note [ɔ̟:].

GOAT a une diphtongue de type [əʊ]. La variante traditionnelle et rare, [oʊ], était naguère la norme. La variante [ɛʊ] appartient à une forme d'accent raffiné qui confine à la caricature. Nolan ([Nol98], 82) semble faire de cette dernière variante une question de génération :

'no joke' as spoken by some younger speakers seems to old speakers to sound more like 'nay
Jake'⁶¹.

La voyelle de GOOSE est souvent décrite comme voyelle postérieure longue et fermée. En réalité, un timbre antérieur et une diphtongue de type [ʊɪ] sont très répandus, le

⁶¹On remarque, en effet, que chez certains locuteurs le deuxième élément de la diphtongue est au moins partiellement délabialisé : [ɛɪ] tendant vers [ɛi].

degré d'antériorisation étant négativement corrélé à l'âge des locuteurs. En examinant les relevés de formants de l'étude de Hawkins et Midgley ([HM05]), où les locuteurs sont divisés en classes d'âge, on relève immédiatement que GOOSE et FOOT ont connu une antériorisation flagrante, notamment chez les locuteurs de 20 à 40 ans. En effet, les valeurs moyennes de F2 pour GOOSE passent de 994 Hz chez les 65 ans et plus à 1616 Hz chez les 20-25 ans. Cette dernière valeur correspond, si l'on accepte le parallèle entre trapèze vocale articulaire et plan F1/F2, à un degré d'antériorité supérieur à la voyelle de TRAP. Pour FOOT, la tendance est identique, à ceci près que, dans le cas de GOOSE, l'évolution de F2 en fonction de l'âge semble plus graduelle, alors que pour FOOT, l'antériorisation est soudaine et spécifique aux 20-25 ans. Gardons-nous cependant d'en tirer des conclusions trop hâtives, chaque classe d'âge dans l'étude citée n'étant constituée que de 5 locuteurs. Gimson ([Gim80], 122) identifie deux facteurs, l'un systémique, l'autre sociolinguistique, influant sur la position de la voyelle de GOOSE en anglais standard. L'absence de phonème dans la zone de [y] permet à GOOSE de migrer vers la partie antérieure du système. En revanche, une antériorisation trop marquée est associée aux dialectes de la région de Londres, notamment populaires ; ceci juggle dans une certaine mesure une potentielle antériorisation supplémentaire en anglais standard. Sur le plan de la perception, Harrington et collègues ([HKR07]) ont montré, à partir de stimuli sur un continuum /i/-/u/, que la frontière entre les catégories /i/ et /u/ dépendait de l'âge des auditeurs. Plus précisément, les sujets les plus jeunes catégorisent dans /u/ des stimuli avec un F2 élevé (i.e. proches de [i]) que les plus âgés classent comme /i/ ; il y a donc, compte tenu notamment des résultats de Hawkins et Midgley, une symétrie entre production et perception. Les auteurs démontrent en outre une absence de compensation perceptive des effets de coarticulation (e.g. l'antériorisation après [j] n'est plus perçue comme une conséquence de la coarticulation) dans le groupe des sujets jeunes, qui pourrait être une des causes de l'antériorisation de la voyelle de GOOSE en anglais standard.

L'accent le plus traditionnel a [aɪ] pour la voyelle de PRICE. Upton ([Upt04], 225) et Olausson et Sangster ([OS06]) trouvent la transcription suivante adéquate : [Λɪ].

La réalisation attendue de la voyelle de CHOICE en anglais standard est [ɔɪ].

La voyelle de MOUTH est une diphtongue qu'on peut symboliser [aʊ] ou [ɑʊ]. Le point de départ le plus postérieur est symptomatique d'un accent raffiné. Nous reviendrons *infra* (Section 6.3.1.2) plus en détail sur les caractéristiques acoustiques de cette voyelle et sur le choix d'un symbole phonétique adapté.

NEAR présente une diphtongue centripète de type [ɪə]. Les réalisations [ɪə:] et [ɪɑ:] traduisent une affection excessive.

Upton ([Upt04], 226) suggère que la forme la plus actuelle de SQUARE est une monophthongue ([ɛ:]) ou, tout au plus, une monophthongue suivie d'un léger glide vers le centre de l'espace vocalique. La diphtongue [ɛə] stigmatise les formes les plus conservatrices de cet accent. Il y a plus de quarante ans, [Jon66], 64, notait que la réalisation monophthongale n'était qu'occasionnelle.

La voyelle de START est réalisée comme [ɑ:]. Upton ([Upt04], 226) précise que même les locuteurs qui ont une voyelle courte pour BATH, ont une longue dans START. En revanche, ces derniers sont parmi ceux qui ont la réalisation la plus antérieure.

Les ensembles lexicaux NORTH et FORCE se confondent avec THOUGHT en anglais standard. Jusqu'à récemment, les items de FORCE avaient /ɔə/ chez les locuteurs les plus âgés. Voici ce qu'écrivait O'Connor à ce sujet en 1973 ([O'C73], 153) :

/ɔ:/ and /ɔə/ are not often separated : relatively few RP speakers make a contrast, for instance, between 'paw' with /ɔ:/ and 'pour' with /ɔə/ [...] but some still make the contrast and it must be reckoned with.

En ce qui concerne CURE, les formes les plus conservatrices de l'accent ont [ʊə]. Il est de plus en plus fréquent de nos jours d'entendre [ɔ:]. Jones ([Jon66], 66) décrivait un stade intermédiaire entre [ʊə] et [ɔ:], en l'occurrence : [ɔə]. Aux dires de Jones, nombre de locuteurs du sud de l'Angleterre utilisaient déjà [ɔ:] à son époque, surtout dans les mots fréquents. La substitution de [ɔ:] à [ʊə] dans l'ensemble CURE était considérée par Wells ([Wel82], 237 et 287-288) comme un changement en cours ayant pour conséquence, selon le point de vue, une migration d'un grand nombre d'items de CURE vers FORCE-NORTH-THOUGHT, ou même une convergence totale de CURE et du super-ensemble que

nous venons de mentionner, avec toutefois, dans cette seconde perspective, l'existence d'un allophone [ʊə] chez certains locuteurs dans le contexte /Cj_/. Le dictionnaire de prononciation de Jones ([Jon03]) donne les deux variantes ([ʊə] et [ɔː]) pour la prononciation de <poor>; la monophthongue y est considérée comme la plus usuelle. /ʊə/ est le phonème vocalique le moins fréquent en anglais ([Gim80], 149); ceci constitue probablement un facteur aggravant qui facilite sa disparition, ou la disparition de son allophone principal historique. Gimson en 1980 ([Gim80], 146) ne conseillait que la diphtongue [ʊə] pour les apprenants. En 1998, Nolan ([Nol98], 79) remarquait que lui-même utilisait /ʊə/, et considérait qu'il était peu probable que des mots tels que <sewer> ou <Ruhr> soient prononcés avec /ɔː/.

FIRE et POWER sont des triphthongues que [Upt04], 226, symbolise [ɪɪə] et [aʊə], respectivement. Elles peuvent être réduites à des diphtongues ou à des monophthongues (phénomène du « Smoothing ») chez tous les locuteurs à débit rapide. Curieusement, le Smoothing à débit lent semble traduire une élocution raffinée en RP, alors que c'est également un cockneyisme notoire. Le passage de la triphthongue à la monophthongue peut conduire à une homophonie entre <buyer> (FIRE), <bower> (POWER) et <bar> (START) (exemple tiré de [Gim80], 140).

Pour l'ensemble *happy*, l'anglais standard actuel présente une voyelle tendue, résultant du *happy-tensing*, phénomène détaillé dans la Section 3.2.1.5. On trouve [ɪ] dans une forme conservatrice d'anglais standard.

Les ensembles *letter* et *comma* ont des réalisations en [ə] et appartiennent au même phonème. L'anglais standard n'étant pas rhotique, le /r/ n'apparaît dans *letter* que lorsqu'il sert à faire la liaison avec le mot suivant si ce dernier commence par une voyelle. Traditionnellement, dans un registre très soigné, ce type de liaison était évité au profit d'un coup de glotte : <far away> [fɑː?əweɪ] (exemple repris de [Upt04], 228)⁶². Le r intrusif (une explication plus détaillée est donnée dans la Section 3.2.1.2) semble être la norme, même à l'intérieur d'un mot : <drawing> [dɹɔːɪŋ].

⁶²Le symbole utilisé dans la référence correspond en API à une occlusive épiglottale; nous ignorons s'il s'agit d'une coquille, mais avons tout de même noté un coup de glotte.

Cette présentation du système vocalique de l'anglais standard sera confrontée aux données empiriques de notre corpus dans la Section 6.3.1. Avant de clore notre discussion sur cette variété, nous abordons, dans la section suivante, les changements phonéto-phonologiques en cours.

3.1.2 Changements récents et perspectives

Dans cette section, nous tirons un bilan des évolutions récentes de l'anglais standard et commentons les changements en cours. Les changements récents dont il est question sont ceux qui sont intervenus entre les premières descriptions de Daniel Jones dans les années 1910 et aujourd'hui.

Comme le rappellent très justement Hawkins et Midgley ([HM05]), une étude phonétique avec des sujets classés par tranches d'âge (i.e. transversale) est quelque peu différente d'une véritable étude longitudinale⁶³. En effet, la prononciation d'un locuteur donné évolue tout au long de sa vie. L'exemple le plus flagrant (mais peut-être pas le plus représentatif⁶⁴) nous est donné par la reine d'Angleterre. Dans une série de publications, Harrington a examiné les allocutions (annuelles) à l'occasion de Noël de la reine Élisabeth II depuis 1952. Par exemple, dans une étude préliminaire et succincte, Harrington et collègues [HPW00] analysent les deux premiers formants des 11 monophthongues de l'anglais standard produites par la reine dans les années 1950, les années 1980, et par un panel de présentatrices de la BBC enregistrées dans les années 1980. Concernant la prononciation de la reine, l'étude montre qu'au moins un des deux formants varie de façon significative pour 10 des 11 voyelles, et que les deux formants varient pour 5 des 11 voyelles entre 1950 et 1980. La comparaison avec le corpus de présentatrices de la BBC fait apparaître que les valeurs formantiques des enregistrements les plus récents de la reine sont à mi-chemin entre les valeurs de 1950 et celles du corpus de présentatrices. [HPW00] concluent que la

⁶³Les auteurs anglophones utilisent les adjectifs « apparent-time » et « real-time » pour qualifier les études transversales et longitudinales, respectivement ; voir [Lab94], 43 *sqq.*

⁶⁴Labov commente les résultats de Harrington ([Har06]) et estime pourtant que la reine devrait être particulièrement résistante aux changements phonétiques de la communauté ([Lab06], 501-502).

reine Élisabeth II a inconsciemment modifié sa prononciation pour suivre les tendances inspirées par les locuteurs les plus jeunes d'une forme de prononciation standard du sud de l'Angleterre, moins élitiste que RP. Tout ceci a été rendu possible, entre autres, par l'effondrement d'une stratification sociale rigide, notamment entre les années 1960 et 1980. L'évolution de la prononciation de la reine est particulièrement visible pour la voyelle de TRAP. Guère après la publication de [HPW00], un article paru dans *The Independent* ([Hen00]) dénonçait une étude « disrespectful to the point of *lèse-majesté* ». Le journaliste continuait en remarquant qu'il était bien évident que chacun d'entre nous modifie son accent tout au long de sa vie, et confiait sa vision de l'évolution de la norme de prestige : pour lui, l'accent aristocrate d'il y a cinquante ans a disparu et l'accent des classes moyennes d'hier est devenu la norme de prestige d'aujourd'hui.

Le cas de la voyelle de TRAP présente un intérêt particulier dans une discussion sur l'évolution des systèmes vocaliques. En effet, [Par98], 22, rappelle qu'en 1874, Sweet notait la diffusion d'une réalisation plus fermée dans les classes supérieures. Or depuis quelques décennies, la voyelle de TRAP est de plus en plus ouverte; elle est donc montée puis redescendue dans le triangle vocalique en l'espace d'un siècle environ⁶⁵. Cet exemple appelle plusieurs commentaires. D'abord, si nous nous référons aux principes qui régissent les changements en chaîne (voir Section 2.2.1), on remarque que la fermeture de TRAP est contraire au fait que, dans ce type de changement, on constate généralement que les voyelles brèves s'ouvrent⁶⁶. Ensuite, puisque la « case » laissée vide par TRAP n'a pas été occupée par un autre phonème, et au vu du fait que la pression exercée par TRAP n'a pas été suffisante pour contraindre FLEECE à quitter sa position de voyelle antérieure de degré de fermeture maximale, on peut conclure, avec Labov ([Lab94], 138), qu'il s'agit d'une simple compression des voyelles antérieures. Nous ajouterons qu'il nous apparaît que cette fermeture, puis ouverture, de la voyelle de TRAP en l'espace d'un siècle consti-

⁶⁵On compare parfois la voyelle de TRAP au Grand Old Duke of York (image tirée de [Nol98], 81), dans la comptine éponyme : « The Grand Old Duke of York, He had 10,000 men, He marched them up to the top of the hill, And he marched them down again . . . ».

⁶⁶Les observations empiriques sur lesquelles Labov ([Lab94]) base ces principes sont cependant loin d'être exhaustives. Parallèlement, le statut phonologique bref de la voyelle de TRAP peut être remis en cause pour certains mots d'après [Fud77].

tue certainement un exemple de conflit entre une innovation motivée par des phénomènes sociolinguistiques et les tendances universelles qui semblent régir la « directionnalité » des changements vocaliques, ces dernières ayant fini par l'emporter.

L'antériorisation des voyelles de GOOSE et de FOOT constitue l'une des évolutions les plus marquantes du système vocalique de l'anglais standard ([Wel62, Hen83, HM05, dJMHN07]). Si l'on ajoute à cela l'ouverture plus importante des voyelles antérieures, qui est très marquée pour TRAP, il semble justifié d'émettre l'hypothèse selon laquelle un changement en chaîne, se traduisant par un mouvement des voyelles dans le sens inverse des aiguilles d'une montre, affecte le système des monophthongues de l'anglais standard ([HM05]).

Bien que nous ne traitons pas la question des consonnes, mentionnons brièvement certains traits consonantiques qui pourraient peut-être un jour avoir leur place dans une description de l'anglais standard. Le TH-fronting – la transformation de /θ/ et /ð/ en /f/ et /v/ respectivement – est attesté dans beaucoup de centres urbains en Angleterre (voir [Bri02], 58 *sqq.*). Il en va de même pour le /r/ labiodental ([Nol98, FD00, Mar06]). On sait que ces traits sont perçus comme non-standards⁶⁷, mais on ne peut pas exclure que leur diffusion entraîne une réinterprétation de leur connotation sociale. En effet, ce genre de phénomène est déjà attesté; par exemple, le coup de glotte remplaçant le /t/ intervocalique est stigmatisé dans le sud-est de l'Angleterre comme une forme typique des hommes et de la classe ouvrière. Or il a été démontré, sur la base de l'utilisation croissante de cette variante à Newcastle (remarquable notamment chez les jeunes femmes de classe moyenne), que la réalisation du /t/ intervocalique en coup de glotte correspondait plutôt à une forme de « chic » urbain ([Bri02]).

D'autres innovations concernent le schéma intonatif; par exemple, Nolan ([Nol98], 84) relève chez les locuteurs les plus jeunes l'adoption du « High Rising Terminal » (voir aussi [SH03]). Enfin, le schéma accentuel d'un mot, qui détermine l'alternation voyelles pleines/voyelles réduites, a lui aussi connu quelques changements. Par exemple, d'après

⁶⁷À noter néanmoins, concernant le statut du /r/ labiodental, que Nolan ([Nol98], 83-84) l'inclut dans les innovations de RP.

[Nol98], <formidable> est passé de /'fɔ:mɪdəbl/ à /fə'mɪdəbl/. Le mot <armchair> semble également avoir connu un déplacement de l'accent primaire de la seconde vers la première syllabe. Ceci contribue probablement à ce qu'il tende à être interprété comme un tout et non comme deux entités séparables ; de là à imaginer qu'il subisse le même sort que <cupboard> au point où il en viendrait à être prononcé /'ɑ:mtʃə/, cela paraît très improbable étant donné l'omniprésence de l'orthographe dans nos cultures.

3.2 Caractéristiques phonético-phonologiques des dialectes des Îles Britanniques

La description géolinguistique qui suit s'appuie en particulier sur les ouvrages de Orton et collègues ([OSW78]), Wells ([Wel82]), Trudgill ([Tru90]), Foulkes et Docherty ([FD99]), et Schneider *et al.* ([SBK⁺04]). L'énumération des traits diagnostiques de chaque variété dans cette section ne présume en rien de leur présence dans nos données ; nous nous contentons de rapporter les propos des spécialistes. Ces éléments sont mis en relation avec les dialectes de notre corpus dans la Partie II. Trudgill ([Tru90]) mentionne le fait que la plupart des Britanniques sont des locuteurs de dialectes modernes, et que si l'on souhaite localiser géographiquement l'un d'entre eux (et c'est précisément notre dessein), il convient d'avoir recours à des critères de prononciation différents de ceux utilisés pour les dialectes traditionnels. Notre analyse des dialectes de l'Angleterre s'appuie sur la classification établie par Trudgill ([Tru90], 50-78). Pour l'intégralité des Îles Britanniques, nous suivrons la typologie de Hughes et collègues ([HTW05]) ; elle est illustrée dans la Table 3.1 et sert de fondement à la description qui suit.

La Table 3.1 reprend la terminologie de Hughes et collègues ([HTW05]) ; or certains des termes employés font référence à des concepts que nous désignerons autrement. En effet, /ʌ/ dans *mud* est un trait qui caractérise les dialectes dans lesquels la scission FOOT-STRUT a eu lieu (voir Section 3.2.1.1). /ɑ:/ dans *path* s'applique aux dialectes qui ont le BATH-broadening, cet allongement accompagné d'une postériorisation de la voyelle

TAB. 3.1 – Classification des dialectes des Îles Britanniques (d’après [HTW05]). Les zones non représentées dans notre corpus sont signalées par ***.

Aire dialectales	Dialectes de ABI	/ʌ/ dans <i>mud</i>	/ɑ:/ dans <i>path</i>	/ɑ:/ dans <i>palm</i>	/g/ dans <i>sing</i>	/j/ dans <i>few</i>	/i:/ dans <i>hazy</i>	/ɪ/ dans <i>bar</i>	/ʊ/ dans <i>pull</i>	/h/ dans <i>harm</i>	[ei] dans <i>gate</i>
Scot. & North. Ire.	<i>gla/shl/uls</i>	+	-	-	-	+	-	+	-	+	-
S. Ireland	<i>roi</i>	+	+	+	+	+	+	+	-	+	-
Northeast	<i>ncl</i>	-	-	+	+	-	+	+	-	+	-
Central north	***	-	-	+	-	-	+	-	-	+	-
Central Lancashire	<i>lan</i>	-	-	+	-	+	+	-	+	+	-
Merseyside	<i>lvp</i>	-	-	+	+	-	+	-	+	+	+
Humberside	<i>eyk</i>	-	-	+	+	-	+	-	-	+	-
N.W. Midlands	***	-	-	+	-	-	+	-	+	+	+
E. Midlands	***	-	-	+	-	-	+	-	-	-	+
W. Midlands	<i>brm</i>	-	-	+	+	-	+	-	+	+	+
S. Midlands	***	+	+	+	+	-	+	-	-	-	+
E. south-west	***	+	-	-	+	+	+	-	-	+	+
W. south-west	<i>crn</i>	+	-	-	+	+	+	-	-	+	-
South-east	***	+	+	+	+	-	+	-	-	+	+
East Anglia	<i>ean</i>	+	+	+	+	-	+	+	-	-	+
Wales	<i>nwa</i>	+	-	+	+	-	+	-	-	+	-

historique /a/ devant fricative sourde. /ɑ:/ dans *palm* désigne le phénomène qui consiste à avoir une voyelle postérieure longue dans les mots de l’ensemble lexical PALM. /i:/ dans *hazy* fait référence au *happy*-tensing, phénomène décrit plus bas (Section 3.2.1.5). /ɪ/ dans *bar* est plus connu sous le nom de rhoticité. /ʊ/ dans *pull* renvoie à la convergence FOOT-GOOSE, caractéristique des dialectes de l’Écosse et de l’Irlande du Nord, que nous regrouperons sous l’appellation générique « Scots ». Pour /h/ dans *harm*, c’est l’absence de ce trait qui porte un nom : le H-dropping. /g/ dans *sing* dénote la réalisation du digraphe <ng> par [ŋg], différente de la réalisation la plus fréquente : [ŋ]. Concernant /j/ dans *few*, c’est son absence qui est appelée Yod-dropping. Enfin, [ei] dans *gate* fait référence à l’emploi d’une diphtongue fermante pour les mots de l’ensemble lexical FACE. Nous avons inclus dans la Table 3.1 les abréviations des dialectes de notre corpus ; ce dernier est décrit dans la Section 4.1.

L’aspect le plus saillant de cette typologie réside dans le fait que 9 des 10 traits retenus sont d’ordre systématique, le dixième se situant sur le plan de la réalisation. Il semblerait donc que la plupart des caractéristiques essentielles des dialectes des Îles Britanniques

soient d'ordre phonologique et non réalisationnel. Nuançons toutefois : une logique bivalente ne peut guère être utilisée que pour une classification employant comme traits distinctifs des phénomènes phonologiques (car ils sont binaires par nature : présence ou absence du phonème, voyelle arrondie ou non, etc.) ; une valence plus élevée serait nécessaire pour caractériser les aspects phonétiques, et le tableau perdrait en lisibilité.

3.2.1 Partition dialectale de l'Angleterre

Notre analyse géolinguistique de l'Angleterre est nettement plus détaillée que celle des autres pays des Îles Britanniques. Il y a deux raisons à cela : il apparaît, d'une part, que la situation est beaucoup mieux documentée, et que les documents sont plus accessibles ; d'autre part, et c'est bien là la raison principale, la majorité des dialectes de notre corpus (huit sur douze, en excluant la variété étiquetée « anglais standard » du calcul) concernent l'Angleterre.

3.2.1.1 Partition nord/sud : scission FOOT-STRUT et BATH-broadening

Nous dirons, en première analyse, que l'Angleterre est traversée par une frontière – dont la localisation précise est malaisée – qui a un statut particulier puisqu'elle est généralement connue de tous les Anglais : « [it] has an acknowledged folk-linguistic status » ([AW04], 178). Cette frontière traduit une réalité qui peut se résumer à l'expression évocatrice « North-South divide » ([Wal00]). Du point de vue de la dialectologie, cette démarcation est matérialisée par deux isoglosses qui, au reste, ne coïncident pas exactement. Le nord est caractérisé par l'absence de scission FOOT-STRUT ; en d'autres termes, la voyelle de ces deux mots appartient au même phonème, dont la réalisation s'approche de [ʊ]. Dans le nord, donc, <luck> et <look> sont homophones. La délabialisation de la voyelle de STRUT (dans certains contextes), qui a conduit à une phonémisation de l'opposition FOOT-STRUT, semble s'être amorcée dans le sud de l'Angleterre vers le milieu du 17^e siècle ([Wel82], 197). Non seulement cette absence d'opposition FOOT-STRUT est elle-même un trait hautement diagnostique, mais aussi, le phénomène d'hypercorrection qui consiste, pour le locuteur

du nord, à se risquer à une prononciation méridionale en utilisant la voyelle de STRUT pour un mot qui appartient à l'ensemble lexical de FOOT dans le sud, et qui aboutit à des prononciations dont la forme phonologique serait /ʃʌgə/ pour <sugar>, permet de trahir l'origine dialectale de ce locuteur. Knowles ([Kno78], 86), dans une analyse informelle de sa propre prononciation, discerne un second type d'hypercorrection agissant entre les voyelles de deux mots différents, et qui revient à modifier la mauvaise voyelle. Ainsi, dans ce cas, <good luck> et <black castle> en viennent à être prononcés /gʌd lʊk/ et /blɑ:k kæsl/, respectivement.

L'absence de scission FOOT-STRUT est tellement caractéristique du nord que Trudgill ([Tru90], 51) va jusqu'à affirmer qu'elle est connue de « everybody who has spent any time in England » (Soulignement ajouté). La seconde isoglosse qui scinde l'Angleterre en deux est située légèrement au nord de la précédente. Elle sépare le nord linguistique, où la voyelle de BATH est, d'un point de vue systémique, identique à celle de TRAP (avec une qualité proche de [a]), du sud, où a eu lieu le phénomène de BATH-broadening. Ce dernier s'est traduit dans un premier temps par l'allongement de la voyelle de BATH vers la fin du 17^e siècle, notamment devant une fricative sourde, et a abouti, dans un second temps, à une qualité postérieure de la voyelle au 19^e siècle. Le nord linguistique, comme le précise Wells ([Wel82], 349) ne correspond pas exactement au nord géographique puisqu'il inclut, en plus de ce dernier, les Midlands.

À l'instar d'un certain nombre d'autres pays, la partition nord/sud dialectologique en Angleterre est également perçue comme une division socio-économique. Il semblerait donc que quand bien même l'on voudrait se concentrer exclusivement sur le facteur géographique, on ne pourrait pas s'affranchir de la dimension sociolinguistique. Par exemple en Italie, le sud véhicule, dans la conscience collective, l'idée de pauvreté et de ruralité. Les traits de prononciation du *mezzogiorno* sont considérés comme diagnostiques de l'appartenance du locuteur à cette région, avec toutes les connotations péjoratives qui en découlent. Le nord, quant à lui, incarne la réussite industrielle et la richesse. En Angleterre, la polarité est inversée : le nord est perçu – et ce cliché est largement entretenu par

les média - comme le royaume des villes crasseuses où règnent les taux de chômage les plus élevés (voir, par exemple, certains films de Ken Loach).

Concernant les deux variables permettant de distinguer le nord du sud, notons d'emblée qu'à l'intérieur du nord linguistique, l'absence de scission FOOT-STRUT et de BATH-broadening ne se manifestent pas de manière homogène. La scission FOOT-STRUT étant tributaire de facteurs sociolinguistiques, Wells ([Wel82], 352) remarque que plus on monte vers le nord, plus le passage d'un système de 5 voyelles brèves à un système à 6 (opposition FOOT-STRUT rétablie dans les classes supérieures) est situé haut dans l'échelle sociale. Ajoutons à cela qu'il existe un stade intermédiaire entre l'absence totale d'opposition FOOT-STRUT et son existence : sur le plan de la réalisation, la forme la plus marquée de l'accent du nord a [ʊ] pour FOOT-STRUT, alors que certains locuteurs tendant vers la variété de prestige (méridionale), mais ne possédant pas le contraste phonologique, vont produire ce qu'on pourrait considérer comme un timbre intermédiaire entre les réalisations de /ʌ/ et /ʊ/ en anglais standard, à savoir, [ə].

Les deux variables marquant la distinction nord-sud n'ont par ailleurs pas exactement le même statut ; Wells ([Wel82], 354) indique :

There are many educated northerners who would not be caught dead doing something so vulgar as to pronounce STRUT words with [ʊ], but who would feel it to be a denial of their identity as northerners to say BATH words with anything other than short [a].

Trudgill ([Tru90], 50-78) propose une classification des dialectes modernes de l'Angleterre qui a l'avantage d'être exposée de façon succincte. Nous suivrons son usage, et conserverons par souci d'homogénéité, sauf pour désigner les quatre points cardinaux, la terminologie anglaise. La Table 3.2 est une matrice de traits binaires offrant une représentation synoptique de la classification opérée par Trudgill ([Tru90]). À partir de 7 traits de prononciation, l'Angleterre peut être divisée en 16 aires dialectales. Ceci n'est d'ailleurs pas tout à fait vrai, puisqu'une inspection sommaire de la Table 3.2 fait apparaître que les traits utilisés ne sont pas suffisants pour distinguer les dialectes étiquetés Northeast et Humberside, et Merseyside et West Midlands, ce qui n'est pas explicitement mentionné

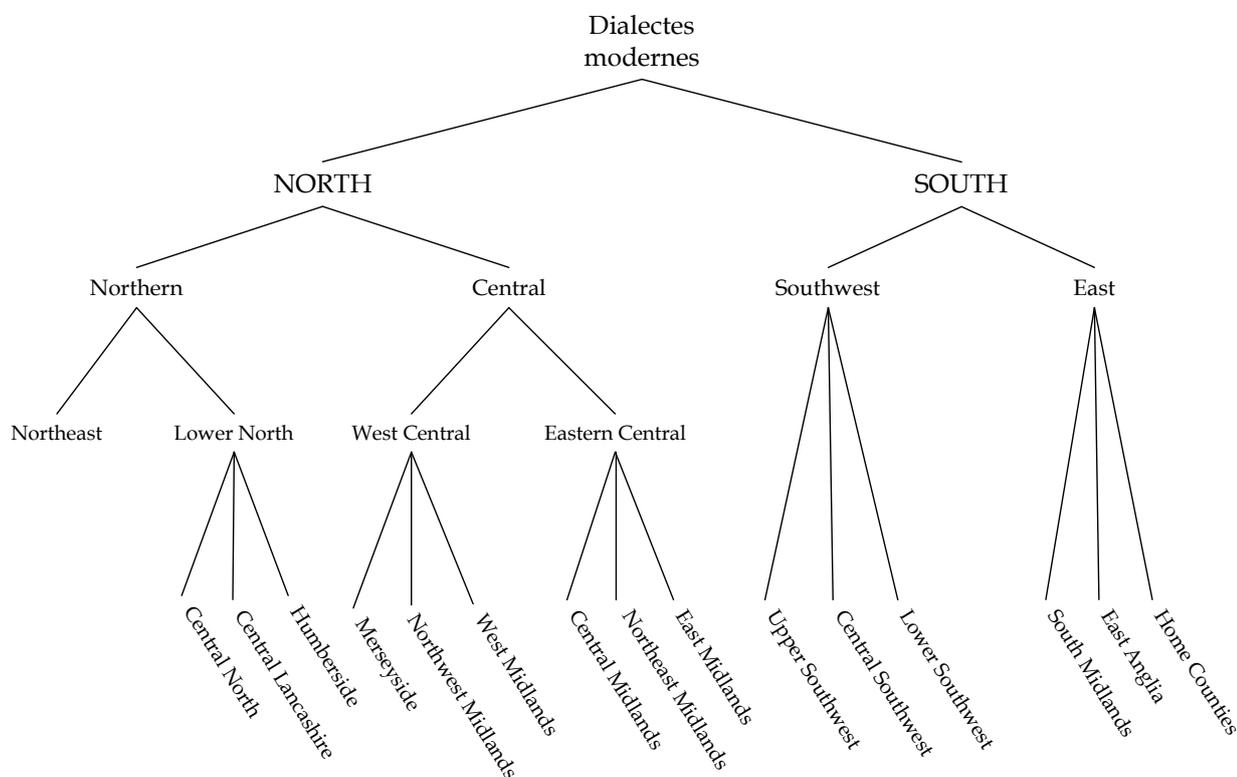


FIG. 3.4 – Regroupement des zones dialectales en supra-régions.

par Trudgill.

Trudgill regroupe ces 16 aires en supra-régions ; nous avons reproduit cette classification arborée dans la Figure 3.4.

Abordons à présent tour à tour chacun des traits employés dans la Table 3.2 : la scission FOOT-STRUT, la rhoticité, la prononciation de [g] dans le digraphe <ng>, le Yod-dropping, le *happy*-tensing, la réalisation de FACE et la vocalisation du /l/. À l'occasion, les paragraphes qui suivent nous donnerons l'opportunité de détailler les caractéristiques

TAB. 3.2 – Classification des dialectes de l'Angleterre (d'après [Tru90]).

Dialectes de ABI	Scission FOOT-STRUT	Rhoticité [g] dans <ng>	Yod-dropping	happy-tensing	monophthongue dans FACE	L Vocalization
Northeast	-	-	-	+	+	-
Central North	-	-	-	-	+	-
Central Lancashire	-	+	-	-	+	-
Humberside	-	-	-	+	+	-
Merseyside	-	+	-	+	-	-
Northwest Midlands	-	+	-	-	-	-
West Midlands	-	+	-	+	-	-
Central Midlands	-	-	-	-	-	-
Northeast Midlands	-	-	-	+	-	-
East Midlands	-	-	+	+	-	-
Upper Southwest	+	+	-	+	-	-
Central Southwest	+	+	-	+	-	+
Lower Southwest	+	-	-	+	+	-
South Midlands	+	-	+	+	-	+
East Anglia	+	-	+	+	-	-
Home Counties	+	-	-	+	-	+

phonétiques d'un ou plusieurs dialectes, notamment lorsque le lien avec les dialectes de notre corpus est évident.

Nous avons déjà traité la question de la scission FOOT-STRUT ; nous ne reviendrons donc pas dessus. Précisons toutefois, comme le fait Trudgill ([Tru90], 66) que le BATH-broadening aurait tout aussi bien pu être utilisé dans cette table en lieu et place de la scission FOOT-STRUT. Voyons le cas de l'accent de Hull, puisqu'il figure dans nos données. Son système vocalique est typique du nord de l'Angleterre car les deux phénomènes que nous venons de mentionner n'ont pas eu lieu. La convergence NURSE-SQUARE est fréquente ([WK99], 146). La voyelle de PRICE a deux variantes : une monophthongue devant consonne voisée (type [a:], apparemment propre à la classe ouvrière [WK99]) et une diphtongue partout ailleurs. Traditionnellement, dans certaines zones du Yorkshire, la voyelle de GOAT est une monophthongue postérieure à mi-chemin entre [ɔ] et [o] ([O'C73], 166-167).

Sur le plan consonantique ([WK99], 147-148), le H-dropping est particulièrement répandu. Le TH-fronting (réalisation de /θ, ð/ en [f, v]) est fréquent dans la jeune génération. Le /r/ labiodental ([ɹ]) existe chez les enfants et les jeunes adultes. /l/ est légèrement vélarisé dans toutes les positions. Le remplacement du [t] (sauf à l'initiale) par un coup de glotte est fréquent chez les locuteurs les plus jeunes.

3.2.1.2 La rhoticité

Un accent est dit rhotique lorsque tous les <r> graphiques – et donc souvent historiques – sont prononcés. Ceci s'oppose aux accents non-rhotiques dans lesquels seuls les <r> en position pré-vocalique sont produits. Le <r> pré-consonantique ou en finale de mot a commencé à disparaître dans la prononciation des dialectes du sud-est de l'Angleterre aux alentours du 18^e siècle⁶⁸ ([Tru90], 51, [Wel82], 218 *sqq.*). La perte de la rhoticité

⁶⁸Précisons qu'il réapparaît dans les dialectes non-rhotiques en finale de mot lorsque le mot suivant commence par une voyelle ; autrement dit, il sert à l'euphonie qui caractérise les phénomènes de liaison. Ce peut être un <r> orthographique, auquel cas on parle généralement de *linking r*, ou absent de la graphie, et dans ce cas, il s'agit de *intrusive r*. Lorsque le <r> est absent de la graphie et qu'il est néanmoins prononcé, mais pas dans le but de faire la liaison avec un mot commençant par une voyelle, Wells ([Wel82], 343) parle d'*hyper-rhoticité*. Une illustration très connue du r intrusif est *Australia and Asia* : [ɔstreiliə r ən eɪzə]. Autre exemple, apocryphe : *Are you xeroxing it ? No, I'm minoltaing it* ([mɪnɔltæɪŋ]).

continue de gagner du terrain en Angleterre : « Every year the r-pronouncing area gets smaller » ([Tru90], 52). Ceci implique que dans certains dialectes, les locuteurs les plus âgés la conservent, alors que les plus jeunes l'ont perdue. La rhoticité est également marquée socialement. Trudgill ([Tru90], 53) explique que l'introduction précoce de la non-rhoticité dans l'accent de la BBC confère à ce trait un statut social relativement élevé. Ceci revient à dire – et c'est un leitmotiv dans nos travaux – que la dialectologie aréale ne peut pas s'affranchir de considérations ressortissant à la sociolinguistique. En Angleterre, la rhoticité caractérise les dialectes du sud-ouest, ainsi que le Central Lancashire, dans lequel se trouve une des villes où ont eu lieu des enregistrements de notre corpus : Burnley. Si nous nous attardons un instant sur le sud-ouest, nous noterons que les dialectes traditionnels de cette région ont une tendance au voisement des fricatives, d'où l'orthographe humoristique <Zomerzet> ou <Mummerzet>. Par exemple, détail très connu, le féminin de <fox>, en l'occurrence, <vixen>, provient des dialectes du sud-ouest de l'Angleterre, d'où la fricative voisée à l'initiale. Cette prononciation, quoique largement attestée dans le projet VOICES de la BBC⁶⁹, reste l'apanage des locuteurs les plus âgés habitant les endroits les plus reculés ([Elm05], 27). Wells [Wel82] classe la Cornouailles (une partie de notre corpus a été enregistré à Truro) comme faisant partie du sud-ouest. Il s'empresse cependant de préciser qu'elle se singularise par une tradition différente des autres comtés de ce sous-groupe, et par le fait qu'on y parlait le cornique⁷⁰.

La rhoticité, présente donc en Cornouailles, se traduit, sur le plan phonétique, par une approximante rétroflexe, [ɻ]. En plus des traits de prononciation rapportés dans la Table 3.2, on notera que les voyelles de GOAT et FACE semblent être de plus en plus souvent réalisées comme des diphtongues ([Wel82], 347). Wells souligne également que dans certaines parties de la Cornouailles, les voyelles de FOOT et GOOSE ont des qualités antérieures.

⁶⁹<http://www.bbc.co.uk/voices/> ; site consulté le 8 avril 2008.

⁷⁰Les derniers locuteurs natifs du cornique (*kernevek* en langue originale), langue celtique du groupe brittonique, ont vécu à la fin du 18^e siècle. Une certaine Dolly Pentreath, décédée en 1777 ([Elm05], 5) semble être la toute dernière locutrice native du cornique, et Chesten Marchant, mort un siècle plus tôt, le dernier locuteur monolingue.

3.2.1.3 L’occlusive vélaire dans <ng>

La prononciation systématique de [g] dans <ng> est diagnostique des dialectes libellés Central Lancashire, Merseyside, Northwest Midlands et West Midlands. Ces dialectes n’ont pas subi le phénomène que Wells ([Wel82], 188-189) nomme NG-coalescence. Aux alentours de 1600, dans la séquence /ng/ en fin de mot (qui était probablement homorganique : [ŋg]), le [g] a cessé d’être prononcé dans l’accent des classes instruites à Londres. Là où ce phénomène n’a pas eu lieu, la variable [g] dans <ng> est très stable sur toute l’échelle sociale; seuls quelques locuteurs de RP conservateurs ne prononcent pas [ŋg] ([Wel82], 365-366).

Nous allons décrire les variétés de Liverpool et Birmingham, qui appartiennent respectivement aux zones Merseyside et West Midlands. L’accent de Liverpool, le Scouse, se distingue de celui des régions environnantes à cause, dit-on, de l’afflux massif d’immigrés irlandais au 19^e siècle ([Wel82], 371). On constate, par exemple, qu’en 1851, plus de 22 % de la population de Liverpool étaient composée d’Irlandais ([Hon07]); 25 % en 1861 ([Kno78]). Cette vision est nuancée par Honeybone ([Hon07]) : il note que la plus grande communauté d’immigrants à cette époque reste tout de même celle en provenance de l’Angleterre, et ajoute que d’autres villes, comme Glasgow ou Newcastle, ont elles aussi connu une importante immigration d’Irlandais. Il n’en demeure pas moins que l’accent de Liverpool est ([Kno78], 80) :

an interesting hybrid : on the phonological level, it remains similar to the dialects of neighbouring Northern towns, but phonetically it has been heavily influenced by Anglo-Irish.

En ce qui concerne le système vocalique, le trait le plus typique est la convergence NURSE-SQUARE, plutôt caractéristique de la classe ouvrière. La réalisation de cette unique voyelle est variable : [ɜ:] ou [ɛ:] ([Bea04], 125). La scission FOOT-STRUT n’a pas eu lieu et le BATH-broadening est attesté à partir des classes moyennes ([Wat06], 57).

Sur la plan réalisationnel, il est intéressant de noter que, contrairement à la plupart des dialectes du nord de l’Angleterre, les voyelles de FACE et GOAT sont des diphtongues fermantes; ces diphtongues ont des mouvements formantiques restreints. Les mots terminant en <ook> ont conservé la voyelle longue du moyen-anglais dans une forme dite « very broad » d’anglais liverpuldien, si bien qu’ils appartiennent à GOOSE, et non à FOOT

([Wel82], 373).

Au niveau des consonnes, l'influence de l'anglais d'Irlande se manifeste à travers la réalisation de /θ/ et /ð/ en occlusives dentales ou alvéolaires ([Wel82], 371), phénomène parfois appelé TH-stopping. Paradoxalement, au moins en apparence, les occlusives sourdes (les sonores, plus rarement) en fin de syllabe peuvent être parfois légèrement spirantisées, ou affriquées (et ceci dans toutes les positions).

Passons à présent en revue les traits caractéristiques de l'anglais de Birmingham, souvent appelé « Brummie »⁷¹. Dans la dichotomie nord-sud, Birmingham est probablement une zone intermédiaire : Wells ([Wel82], 364) mentionne explicitement le fait que FOOT et STRUT puissent être distincts. PRICE et CHOICE peuvent être homophones. La convergence NURSE-SQUARE se rencontre parfois.

Sur le plan réalisationnel, la voyelle de KIT est particulièrement fermée ([O'C73], 155). FACE et GOAT sont des diphtongues à mouvement formantiques amples ([HTW05], 87-88). O'Connor ([O'C73], 154) note que FLEECE a une diphtongue relativement marquée.

3.2.1.4 Le Yod-dropping

Le Yod-dropping caractérise les dialectes East Midlands, South Midlands et East Anglia. La chute du [j] dans les mots <tune>, <dude> et <news>, par exemple, est typique d'un grand nombre de dialectes d'anglais américain ; et le profane considère parfois, à tort, que ce trait distingue ces derniers des dialectes britanniques. En Angleterre, certains contextes consonantiques ont favorisé la disparition de [j] devant [u:] dans une grande majorité de dialectes : lorsque la voyelle était précédée d'une palatale (e.g. <chew>, <juice>), précédée de /r/ (e.g. <crew>, <shrew>) et après une consonne suivie d'un /l/ (e.g. <blue>, <flue>). Ceci correspond à ce que Wells nomme « Early Yod Dropping » ([Wel82], 206). En *General American* s'applique le « Later Yod Dropping » ([Wel82], 247-248), phénomène qui a étendu la chute du [j] devant [u:] aux consonnes coronales. Certains dialectes d'Angleterre, peut-être sous l'influence du prestige que peut dégager la norme

⁷¹Souvenons-nous qu'il s'agit, selon un sondage de la BBC conduit auprès de 5000 personnes, de l'accent que les Anglais affectionnent le moins.

américaine, connaissent une situation proche.

Au-delà, il existe un Yod-dropping total très typique des dialectes mentionnées au début de ce paragraphe, et notamment East Anglia. Cette énumération des traits utilisés dans la taxinomie de Trudgill nous ayant conduit à mentionner l'East Anglia, nous nous arrêtons un instant sur cette variété car elle est présente dans nos données (enregistrements effectués à Lowestoft), et profitons de ce qu'elle a été minutieusement étudiée pour en fournir une description circonstanciée. Le dialecte d'East Anglia est la spécialité du variationniste Peter Trudgill, qui a conduit notamment des études sur la ville de Norwich (située dans le Norfolk) dans les années 1970. L'accent de cette région a joué un rôle capital dans la formation de l'anglais standard. Par ailleurs, si l'on accepte le point de vue selon lequel la langue anglaise est née quand les Anglo-Saxons ont commencé à s'installer en Angleterre, alors on peut considérer que c'est très probablement en East Anglia que l'anglais a été parlé pour la toute première fois dans le monde ([Tru04], 163). Ce dialecte est non rhotique. Entre autres particularités, nous retiendrons (d'après [Tru04]) :

- le Yod-dropping systématique, au point que <dew>/<do> et <Hugh>/<who> sont homophones ;
- l'emploi de /ə/ en position inaccentuée là où on attendrait plutôt /ɪ/ en anglais standard, dans <wanted>, <naked>, etc. Incidemment, ce phénomène est typique de l'accent de Tony Blair (qui n'est pourtant pas originaire de cette région) ;
- l'utilisation, dans les formes les plus archaïsantes, de /ɛ/ pour certains items qui, en anglais standard, appartiennent à TRAP : <catch>, <have>, etc. ;
- certains mots possédant le /ɔ:/ du moyen-anglais peuvent rimer avec *good* : e.g. <road>, <home>, <stone> (/ʊ/), mais pas <coal> (/u:/). L'incidence lexicale de GOOSE et FOOT est différente de celle de l'anglais standard ;
- les diphtongues fermantes – parmi lesquelles Trudgill ([Tru04], 169-172) compte FLEECE et GOOSE – ont la particularité de comporter un second élément plus fermé que dans les autres variétés méridionales ;
- le *happy*-tensing ;

- dans une forme conservatrice du dialecte, les Long Mid Mergers n’ont pas eu lieu, si bien que <days>/<daze> et <maid>/<made> ne constituent pas des paires d’homophones (respectivement pour chaque paire : /æ/ et /e:/);
- pour faire écho au point précédent, et toujours en raison de l’absence de Long Mid Mergers, les voyelles postérieures correspondant au degré d’aperture des deux voyelles que nous venons de mentionner, <nose>/<knows> et <sole>/<soul> ne sont pas homophones (respectivement pour chaque paire : /ʊu/ et /u:/). Trudgill ([Tru04], 170) précise que cette distinction, à la différence de la précédente, est encore bien attestée dans le nord de l’East Anglia;
- la convergence NEAR-SQUARE est attestée dans le nord de l’aire dialectale avec une réalisation proche de [e] ([Tru04], 172);
- certains items de l’ensemble lexical CURE de l’anglais standard ont la voyelle de NURSE. Ainsi <sure> rime avec <her>, et si l’on ajoute à cela le Yod-dropping, <pure> et <purr> sont des homophones dans le nord de la province;
- le Smoothing est largement attesté : <going> est prononcé [gɔ:n], <allow it>, [əlaɪt], etc.

Pour ce qui est des consonnes, la glottalisation des occlusives sourdes en position intervocalique et finale de mot est très répandue; il semblerait d’ailleurs que l’East Anglia soit un des centres à partir desquels ce phénomène s’est généralisé à tout l’anglais britannique. La vocalisation du /l/ est de plus en plus commune dans le sud de l’aire concernée, mais pas au nord. Trudgill ([Tru04], 175) note d’ailleurs que chez les locuteurs du nord de l’East Anglia nés avant 1920, on peut entendre un /l/ clair dans toutes les positions⁷², la distribution sombre *vs* clair actuelle étant la même que celle de l’anglais standard.

3.2.1.5 Le *happy-tensing*

Le *happy-tensing* se manifeste de la manière suivante : comme l’explique Wells ([Wel82], 257), la voyelle finale de mots comme <happy>, <coffee>, etc., a été décrite jusqu’aux

⁷²D’où une très faible probabilité de vocalisation puisqu’il semble que – toute pression sociolinguistique mise à part – seul un /l/ vélarisé puisse aboutir « naturellement » à une voyelle.

années 1950 comme appartenant à KIT. Or sa qualité actuelle est plus proche de [i], ce qui, intuitivement, pousserait à la classer avec FLEECE. D'après la description de Trudgill ([Tru90], 59-60), la version brève de la voyelle est confinée, dans les dialectes modernes, au Central North, Central Lancashire, Northwest Midlands et Central Midlands.

La variante tendue de la voyelle de *happy* soulève le problème du choix du symbole à utiliser dans les dictionnaires de prononciation. En effet, RP avait traditionnellement /ɪ/, et ce n'est vraisemblablement plus le cas aujourd'hui. Les éditeurs du dictionnaire de Jones ([Jon03]), ainsi que Wells dans son dictionnaire ([Wel90]), ont opté pour le symbole /i/. Les premiers considèrent que dans les mots du type de *happy*, il y a neutralisation entre /ɪ/ et /i:/; en disant cela, ils reconnaissent qu'il ne s'agit pas d'un véritable phonème. Si on adopte le principe selon lequel la transcription d'un dictionnaire de prononciation est un équilibre entre un alphabet phonologique (et donc minimaliste) et un guide de prononciation pour les apprenants (cette dernière motivation est au moins implicite à en juger par l'usage de diacritiques), le symbole /i/ peut être justifié. En revanche, dans une optique strictement phonologique, il conviendrait de trancher entre les symboles /ɪ/ et /i:/, et au vu de la proximité phonétique entre la voyelle finale de *happy* et les allophones les plus courants de FLEECE en anglais standard, /i:/ semble plus approprié. Un dernier argument consisterait à dire que la voyelle de *happy* est un phonème à part entière, si on considère que, pour certains locuteurs, <booty> et <bootee> forment une paire minimale⁷³.

3.2.1.6 La voyelle de FACE

La monophthongue longue dans FACE, d'après la typologie de Trudgill ([Tru90], 60-61), caractérise les dialectes étiquetés Northern et Lower SouthWest. Dans la partie nommée Northeast, le Geordie (accent de Newcastle-upon-Tyne, l'aire dialectale est souvent appelée « Tyneside ») traditionnel et/ou du bas de l'échelle sociale se singularise par la présence d'une diphtongue centripète dans FACE – phonétiquement [iə] – très proche de la

⁷³Francis Nolan (communication personnelle) a fait mention de la possibilité d'une telle paire.

prononciation de <fierce> en anglais standard, quoique légèrement plus brève. Ceci nous donne l'occasion d'aborder les traits caractéristiques de l'accent de Newcastle. Cette variété n'est pas rhotique. Le /r/ de liaison et le /r/ intrusif sont plus rares que dans d'autres variétés et sont influencés par des facteurs sociolinguistiques (voir [WM99], 31-32). Le /r/ est généralement réalisé comme une approximante alvéolaire, parfois une battue alvéolaire (surtout en position intervocalique). Le « Northumbrian burr », la fricative ou approximante uvulaire, est extrêmement rare et limité aux locuteurs les plus âgés dans les parlers ruraux (elle n'est pas attestée à Newcastle même d'après [Wel82], 368) ; en revanche, le /r/ labiodental ([ʋ]) est de plus en plus répandu, comme dans d'autres variétés britanniques⁷⁴.

Concernant les particularités du système, le Geordie est typiquement septentrional en cela que la scission FOOT-STRUT et le BATH-broadening n'ont pas eu lieu, ce qui revient à dire que nous sommes en présence d'un système à 5 voyelles brèves, sauf, variation sociolinguistique oblige, si un locuteur donné adopte la norme de prestige, en l'occurrence, l'anglais standard. Le dialecte traditionnel avait semble-t-il une convergence NURSE-NORTH, avec NURSE postériorisé jusqu'à [ɔ:] sous l'influence du [ɸ] ([Wel82], 374-375, [Bea04], 126). L'étude instrumentale de Watt et Milroy ([WM99], 38-40) fait apparaître que la réalisation [ɔ:] (donnant potentiellement lieu à une convergence NURSE-NORTH) est utilisée presque exclusivement par les locuteurs les plus âgés de la classe ouvrière (38 % de leurs réalisations), et que ce trait est clairement récessif : les jeunes de la même classe ne l'utilisent que dans 4 % des cas. On note également un potentiel phonème supplémentaire dans des mots tels que <talk>, <walk>, <call>, etc., dont la réalisation est [a:] ([O'C73], 161).

Sur le plan de la réalisation, les voyelles de FACE et GOAT ont un comportement similaire ; il semblerait, en effet, qu'une tendance à la symétrie avant-arrière dans les

⁷⁴D'après [FD00], le /r/ labiodental est connu depuis le milieu du 19^e siècle comme défaut de prononciation chez les enfants. Gimson, en 1980, y fait encore référence en qualifiant ce trait de « defective » ([FD00], 31). Paradoxalement, ce trait de prononciation semble également avoir été naguère symptomatique d'un style affecté typique des classes supérieures. La variante a désormais sa place dans les médias, puisque on peut voir dans une publicité pour de la lessive, un présentateur lancer son fameux « all right » avec un /r/ labiodental, jouant ainsi sur l'ambiguïté possible avec « all white ». Même Margaret Thatcher aurait adopté consciemment cette variante afin d'adoucir son image ([FD00], 33). L'innovation semble provenir du sud-est de l'Angleterre, et elle s'est ensuite diffusée au reste de l'Angleterre, mais pas encore, semble-t-il, aux autres pays des Îles Britanniques.

systèmes vocaliques soit universelle ([WM99], 32). Les variantes les plus attendues, et donc non marquées, sont de longues monophthongues du type [e:] et [o:]. Dans la classe ouvrière, en particulier chez les locuteurs masculins les plus âgés, on trouve les réalisations les plus conservatrices (en cela qu'elles correspondent au dialecte traditionnel de la zone concernée), i.e. les diphtongues centripètes [ɪə] et [ʊə] ([WM99], 34-38). La voyelle de NURSE possède trois variantes dans les descriptions traditionnelles :

- une voyelle postérieure de type [ɔ:] qui pourrait, dans certains cas, traduire une convergence partielle ou totale avec les ensembles THOUGHT, NORTH et FORCE. Watt et Milroy ([WM99], 33) regardent cette éventualité avec circonspection. Cette variante est l'apanage des locuteurs masculins les plus âgés de la classe ouvrière ;
- une variante proche de l'anglais standard, [ɜ:] ;
- une réalisation antérieure, arrondie à des degrés variables, [ø:], variante de prédilection des femmes jeunes, quelle que soit leur classe sociale.

La réalisation de la voyelle de PRICE est, comme à Hull (Section 3.2.1.1), conditionnée par l'environnement consonantique. Le premier élément de la diphtongue est plus fermé devant une occlusive ou une fricative sourde. Par exemple, <knife> est réalisé [nɛɪf], et <knives>, [naɪvz] ([WM99], 28-29).

3.2.1.7 La vocalisation du /l/ (L-vocalization)

Chez Trudgill ([Tru90], 60-62), la vocalisation du /l/ est diagnostique des Home Counties. Ce trait est typique d'un accent connu sous le nom de Estuary English⁷⁵. Le /l/ en position pré-consonantique est vélarisé dans nombre de variétés d'anglais : au contact de l'apex avec les alvéoles s'ajoute un renflement du dos de la langue en direction du voile du palais, ce qui confère à [ɫ] une qualité proche de [o] ou [u]. Wells ([Wel91]) reconnaissait à demi-mot que ce phénomène pouvait désormais être inclus dans les descriptions de l'anglais standard. On se souviendra que la vocalisation du /l/ s'est produite de ce côté-ci

⁷⁵C'est un terme qui désigne depuis une trentaine d'années l'accent caractéristique de l'estuaire de la Tamise qui, depuis, s'est diffusé au-delà.

de la Manche il y a plus de mille ans⁷⁶.

3.2.2 Le Pays de Galles

Nous commencerons cet exposé par un bref aperçu de la situation, non pas de l'anglais du Pays de Galles, mais du gallois. Le gallois est le descendant direct de la langue celtique parlée sur la plus grande partie du sol britannique avant les invasions anglo-saxonnes. Il comptait 21 583 locuteurs monolingues en 1981, soit environ 0,8 % de la population totale du Pays de Galles ([DO97], 584). 90 ans plus tôt, cette fraction atteignait environ les 30 %. Fennell ([Fen01], 197) constate qu'il ne reste plus un seul locuteur de gallois monolingue aujourd'hui⁷⁷. En ce qui concerne les locuteurs bilingues anglais/gallois, ils représentaient 50 % de la population du pays au début du 20^e siècle et moins de 20 % en 1981 ([DO97], 585). Toutefois, comme c'est le cas pour les autres langues celtiques encore parlées sur les Îles Britanniques, le gallois connaît un renouveau grâce à l'appui des pouvoirs publics et des médias. Sans entrer dans une analyse profonde de ce phénomène, nous livrerons les résultats éloquentes d'une étude de Drinkwater et O'Leary ([DO97], *passim*). À partir d'un examen du *Sample of Anonymised Records* extrait du *1991 Population Census*, les auteurs ont constaté, méthodes quantitatives à l'appui, que les locuteurs du gallois au Pays de Galles sont moins touchés par le chômage que leurs compatriotes qui ne possèdent aucune compétence dans cet idiome. Il se pourrait que ceci soit partiellement imputable à un phénomène de discrimination positive à l'embauche procédant du renforcement d'un sentiment patriotique. Cependant, Penhallurick ([Pen04], 99) soutient que, malgré les efforts récents, le gallois tend vers une « latinisation », processus par lequel une langue vient à n'être utilisée que dans un nombre très restreint de domaines.

En première approximation, notons que l'anglais du Pays de Galles peut être divisé en

⁷⁶Il semblerait que le /l/ latin devant consonne fût vélarisé. La perte du contact apical s'est amorcée, d'après Joly ([Jol95],100) aux alentours du 7^e siècle. [ɫ] a semble-t-il terminé de se vocaliser en [u] au 10^e siècle, formant ainsi une diphtongue avec la voyelle qui le précédait.

⁷⁷Il semblerait néanmoins que le gallois existe encore en tant que langue maternelle : Mark Jones (communication personnelle) rapporte que, alors qu'il effectuait un travail de terrain au Pays de Galles, il a constaté que certains jeunes privilégiaient le gallois jusqu'à ce qu'ils entrent dans la vie active.

deux variétés, l'une typique du nord-ouest, l'autre, du sud. Ces deux variétés sont, d'après Penhallurick ([Pen04], 99), le reflet des deux dialectes du gallois. Formulé autrement, cela revient à dire que le substrat gallois a constitué un facteur important dans la partition dialectale de l'anglais au Pays de Galles. Les variétés parlées en Angleterre dans les zones frontalières ont également influencé l'anglais du Pays de Galles. La description de Penhallurick ([Pen04]) s'appuie sur une variété étiquetée « traditional, rural Welsh English ». Nous ne saurions donc être trop prudent quant à la possibilité de généraliser ce qui suit. L'anglais du Pays de Galles n'est pas rhotique, mais l'influence du gallois dans les zones où ce dernier est encore très présent (i.e. la moitié ouest), la proximité géographique avec des accents d'Angleterre rhotiques (i.e. au sud-ouest) font que des /r/ peuvent être réalisés par certains locuteurs dans toutes les positions. Voici quelques traits caractéristiques :

- certains mots appartenant à LOT en anglais standard, mais orthographiés avec un <a> (<quarry>, <wash>, <wasp>, etc.), ont la voyelle de TRAP ;
- Wells ([Wel82], 380) mentionne la convergence STRUT-Schwa comme caractéristique incontournable. Penhallurick ([Pen04], 103) nuance le propos en parlant de « the blurring or even erasing of distinction » (Soulignement ajouté) ;
- au nord-est, certains mots de l'ensemble STRUT de l'anglais standard ont la voyelle de FOOT, typique des accents du nord de l'Angleterre ; cela semble dû à la proximité avec le Cheshire ;
- certains mots de l'ensemble ONE (cet ensemble lexical peut être considéré comme un sous-ensemble de STRUT regroupant les voyelles orthographiées <o>) sont prononcés avec la voyelle de LOT ; c'est très vraisemblablement le résultat d'une prononciation orthographique ;
- la voyelle de BATH est généralement réalisée en [a], mais il existe une variante allongée [a:], voire allongée et postérieure [ɑ:]. La réalisation précise semble être corrélée au registre de langue et à la classe sociale ([Pen04], 104) ;
- la situation de FACE-STAY et GOAT-SNOW est complexe car elle est loin d'être homo-

gène sur tout le territoire. Sans entrer dans les détails⁷⁸, retenons que le gallois n'a pas de diphtongue de type /ei/ et /ou/. Par conséquent, la phonétique du substrat influençant ici celle du superstrat, les voyelles de FACE et GOAT sont réalisées comme de longues monophthongues. Ce phénomène est, une fois de plus, renforcé au nord du Pays de Galles par la proximité avec les accents de l'Angleterre septentrionale, qui ont en général également une monophthongue. L'orthographe influence les deux voyelles qui nous intéressent dans l'autre direction : les graphies <ai>, <ay>, <ei> et <ey> ont tendance à engendrer la prononciation [ei], et <ou> et <ow> tendent à aboutir à [ou], d'où la nécessité d'identifier deux ensembles lexicaux supplémentaires : STAY et SNOW ;

- les diphtongues de PRICE, CHOICE et MOUTH⁷⁹ ont un second élément particulièrement fermé. Il semble y avoir deux réalisations de ces voyelles : [ai] et [əi] pour PRICE, et [au] et [əu] pour MOUTH. Les variantes avec un schwa semblent confinées au sud et sud-est, ce qui peut laisser penser, comme le note Penhallurick ([Pen04]) que les deux variantes de chaque ensemble sont déterminées par l'époque à laquelle l'anglais s'est implanté dans l'aire concernée ; le sud-est étant la première zone à avoir subi l'influence de l'anglais. Ceci est tout à fait en accord avec la chronologie du Grand Changement Vocalique ;
- CURE, POWER et FIRE sont très clairement dissyllabiques, avec un glide marqué intervenant entre les deux éléments de la diphtongue. Ce trait est hautement caractéristique, et c'est en cela que l'anglais gallois s'oppose à d'autres dialectes, comme East Anglia, dans lesquels se produit le phénomène de Smoothing ;
- l'anglais du Pays de Galles a tendance à avoir, en termes phonétiques, plus de monophthongues et moins de diphtongues que la plupart des autres dialectes ([Wel82], 379) ; par exemple, contrairement à un certain nombre d'autres dialectes, FLEECE

⁷⁸Nous reprenons l'argumentaire de [Pen04], 105-106, ainsi que les symboles qu'il utilise.

⁷⁹[Pen04], 106, utilise MOUSE et non MOUTH (étiquette conventionnelle) pour désigner cet ensemble lexical. Or ces deux mots contiennent visiblement, d'après l'*OED*, la même voyelle historique et le contexte consonantique est tellement similaire qu'il est fort peu probable que l'auteur ait souhaité désigner autre chose que MOUTH. Étant donné qu'à la page 101 dans [Pen04], c'est bien MOUTH qui est employé, nous avons rétabli ici l'appellation conventionnelle de cet ensemble lexical.

et GOOSE sont des monophthongues ;

- la voyelle de NEAR peut être prononcée [jɜ:] d’après [Wel82], 380, auquel cas <fear> rime avec <fur>.

Sur le plan consonantique, notons que les occlusives sourdes sont fortement aspirées, surtout au nord. La plupart des consonnes sont allongées, sous l’influence du gallois, en position médiane de mot ; ceci est également particulièrement singulier. L’articulation dentale (plutôt qu’alvéolaire comme en anglais standard) pour les apicales /t, d, n/ semble être la norme au nord. Les fricatives initiales peuvent être voisées comme c’est le cas dans certains accents du sud-ouest de l’Angleterre. /r/ est généralement réalisé en approximante (comme en anglais standard). Les roulées et les battues ne sont cependant pas rares. On note aussi sporadiquement des réalisations uvulaires dans le nord. Wells rapporte que le [l] clair est très répandu, surtout au sud ([Wel82], 379).

3.2.3 L’Écosse

D’après Wells ([Wel82], 394), l’anglais écossais est un anglais d’Angleterre standard prononcé avec un accent écossais, qui coexiste avec le scots, un peu comme un dialecte moderne peut coexister avec un dialecte traditionnel dans le nord de l’Angleterre. Stuart-Smith ([SS04], 47) parle de continuum corrélé à la classe sociale allant de « broad Scots » à « Scottish Standard English » (ScStE), le scots tendant à être restreint à certaines situations de communication informelle.

Il est généralement admis que l’anglais d’Écosse n’a pas de durée vocalique distinctive dans son système ; c’est d’ailleurs une caractéristique qui sera illustrée dans notre étude sur le rythme (Chapitre 5). En revanche, certains contextes consonantiques ou morphémiques ont une influence particulière sur la durée des voyelles ; c’est la *Scottish Vowel Length Rule* (SVLR), également connue sous le nom de loi d’Aitken (voir par exemple [McC77, SHT99]). Lorsqu’une voyelle précède une fricative voisée, un /r/, ou une frontière morphémique, elle est phonétiquement longue. D’après Scobbie *et al.* ([SHT99]), la SVLR ne s’applique qu’aux voyelles /i/, /u/ et /ai/. Par exemple, les voyelles de *breathe*,

beer, *bee* et *agreed* sont plus longues que celles de *brief*, *bead* et *greed* ([SS04], 57).

Pour rester à un niveau systémique, notons encore que les voyelles de FOOT et GOOSE ne sont pas distinctes ; ce phénomène existe également en Ulster et dans l'extrême nord du Northumberland ([Wel82], 402). Le BATH-broadening n'a pas eu lieu. La plupart des locuteurs de ScStE ont un seul phonème pour la voyelle de LOT, CLOTH et THOUGHT, faisant de <cot> et <caught> des homophones. Incidemment, cette convergence caractérise les variétés de l'ouest des États-Unis ([Gor04], 342). La convergence NORTH-FORCE n'a pas eu lieu pour l'échantillon de ScStE de Glasgow décrit dans [SS04], 53. Il existerait également, de façon sporadique, un phonème intermédiaire entre KIT et DRESS, auquel [SS04], 57 fait référence avec le mot-clé NEVER.

La voyelle de FACE et son homologue postérieur, GOAT, sont généralement réalisés comme des monophthongues. La voyelle de OUT semble être un indicateur assez fiable de la classe sociale : l'usage d'une monophthongue de type [ɥ] est typique de la classe ouvrière, une diphtongue étant de plus en plus fréquente au fur et à mesure qu'on remonte l'échelle sociale. La voyelle de *happy* n'est non seulement pas tendue, mais sa réalisation est plus ouverte, proche de [e].

Les variétés d'anglais d'Écosse sont généralement rhotiques ; Stuart-Smith ([SS04], 62) rapporte qu'environ 90 % des /r/ post-vocaliques de l'une de ses études sont articulés chez les locuteurs de classe moyenne⁸⁰. Les réalisations de /r/ sont très variables : une approximante post-alvéolaire (comme en anglais standard), une approximante rétroflexe, une battue, ou, plus rarement, une roulée. Il arrive parfois que /r/ soit réalisé en [ə]. Wells ([Wel82], 411) note la possibilité d'une fricative uvulaire comme variante idiosyncrasique. Le choix de ces variantes est conditionné par la position à l'intérieur du mot, le contexte phonétique, le sexe, ou encore la classe sociale (pour plus de détails, se référer à [SS04], 62-63, [SS99], 210, [Chi99], 228, [Wel82], 410-411).

En ce qui concerne les consonnes, on retiendra que les apicales peuvent être soit alvéolaires, soit dentales. ScStE possède deux phonèmes consonantiques en plus de la plupart

⁸⁰Une étude récente de Stuart-Smith ([SS07]) montre une dérhoticisation chez les adolescents à Glasgow.

des variétés d'anglais, /ʌ/ et /x/ (pour une discussion détaillée, voir [Chi99], 227 et [SSS06]). Le /l/ est généralement vélarisé dans toutes les positions du mot ; mais un [l] clair est possible, notamment dans les Hautes-Terres, avec une distribution clair *vs* sombre identique à l'anglais standard.

Pour ce qui concerne la dialectologie aréale, Wells ([Wel82], 397) propose 3 zones :

- les dialectes centraux des Basses Terres, incluant principalement la région d'Édimbourg (Lothians) et celle de Glasgow (Strathclyde) ;
- les dialectes méridionaux, avec la région de Dumfries-and-Galloway à l'ouest, et les Borders à l'est ;
- les dialectes septentrionaux, au nord de Stonehaven.

Stuart-Smith ([SS04], 47), quant à elle, remarque que le ScStE (anglais standard parlé avec un accent écossais) ne varie que dans une très faible mesure sur tout le territoire, alors qu'on peut distinguer quatre variétés de scots : « Mid or Central Scots, Southern or Border Scots, Northern Scots, and Insular Scots ».

Il nous est particulièrement malaisé à ce stade de notre étude d'entrevoir quels pourraient être les traits (notamment vocaliques) qui permettraient de distinguer les deux dialectes dont nous disposons dans notre base de données : celui des Hautes-Terres et celui de Glasgow.

3.2.4 L'Irlande

Bien que nous traitions les deux Irlandes dans la même section, souvenons-nous qu'il s'agit de deux zones dialectales très différentes.

Wells ([Wel82], 417) rappelle les 3 principales sources d'influence de l'anglais irlandais actuel :

- l'anglais venu de l'ouest de l'Angleterre ;
- le scots et l'accent qui lui est attaché ;
- le erse (gaélique irlandais), langue indigène, dont l'influence en tant que substrat semble incontestable.

Wells ([Wel82], 418) insiste sur le fait que l'anglais irlandais est demeuré relativement conservateur en cela qu'il est resté imperméable à nombre d'innovations tant britanniques qu'américaines. Nous calquerons la partition dialectale sur la division politique et démographique, qui oppose l'Ulster (ou Irlande du Nord) à la République d'Irlande. Ces régions sont représentées dans notre corpus par les villes de Belfast et Dublin, respectivement. Certes ces supra-régions dialectales peuvent être subdivisées en plusieurs zones, mais une granularité plus fine dépasserait les limites imposées par notre corpus.

À l'indépendance politique de la République d'Irlande (vis-à-vis du Royaume-Uni) semble correspondre une indépendance (socio-) linguistique. Ceci se manifeste à travers le fait que, pour un Irlandais, l'anglais standard du sud de l'Angleterre n'est pas vu comme une norme de prestige, mais plutôt comme une variété que Hickey ([Hic99], 265) qualifie de « extra-national » et de « definitely un-Irish »⁸¹. Ce rejet de l'anglais standard n'a, par exemple, pas été aussi ferme en Écosse.

À des fins typologiques, il peut parfois être intéressant de regarder l'Irlande comme une seule entité linguistique. Dans ce cas, Hickey, dans *A Sound Atlas of Irish English* ([Hic04a]), établit une liste de traits de prononciation partagés par la plupart des variétés vernaculaires le l'île :

- la lénition des occlusives alvéolaires, e.g. <city> [sɪrɪ];
- le [l] clair dans toutes les positions (ceci se limite aux variétés les plus conservatrices);
- la rhoticité : dans le sud, la réalisation traditionnelle est une approximante alvéolaire vélarisée. Au nord, on rencontre une approximante rétroflexe, comme en Écosse;
- la distinction entre les voyelles brèves devant /r/ dans les variétés conservatrices : <term> [tɛ.ɪm] vs <turn> [tʌ.m];
- la distinction entre /w/ et /ʌ/ (qui oppose, par exemple, <witch> et <which>);
- l'absence fréquente de distinction entre les voyelles de TRAP et BATH.

Si, comme le rappelle Hickey ([Hic04a], 41), certains de ces traits ont un potentiel

⁸¹Lors d'une conférence, nous avons, afin d'être concis, remplacé la formulation « dialectes des Îles Britanniques » par « dialectes britanniques ». Une participante nous avait vivement reproché ce raccourci qu'elle jugeait maladroit.

typologique limitée (puisqu'ils existent également dans d'autres variétés), la non rhoticité ou encore le H-dropping permettent de déduire sans hésitation que le locuteur n'est pas irlandais. Comme nous l'avons remarqué plus haut, l'Irlande du Nord et la République d'Irlande sont deux régions linguistiques bien distinctes ; nous conserverons donc la dichotomie dans notre exposé. Nous garderons également l'opposition entre variétés urbaines et rurales.

3.2.4.1 Le sud

Sur le plan systémique, Wells ([Wel82], 421) notait un phénomène particulièrement intéressant à rapprocher des ensembles lexicaux NURSE et SQUARE de l'anglais standard. Les mots-clés <pair>, <per> et <purr> possèdent trois, deux phonèmes différents ou un seul dans les variétés étiquetées « typical southern Irish provincial », « typical Dublin » et « smart Dublin », respectivement. L'auteur constate que la convergence NURSE-SQUARE, trait stigmatisant à Liverpool, est, au contraire, perçue favorablement par certains Dublinois. Hickey ([Hic04b], 91) considère la convergence NURSE-SQUARE comme possible dans ce qu'il intitule « fashionable Dublin » et habituel en « rural northern ». FORCE et NORTH sont distincts dans les 5 variétés présentées par Hickey ([Hic04b], 91), sauf potentiellement en « fashionable Dublin ».

À Dublin, traditionnellement, les voyelles longues fermées en syllabe fermée sont réalisées comme deux syllabes séparées par un hiatus : <clean> [kliːən], <fool> [fuːwəl] ([Hic04a], 45). Le /ʊ/ antérieur à la scission FOOT-STRUT est conservé, e.g <Dublin> [dʊblən]. La voyelle de LOT est allongée lorsqu'elle est suivie d'une fricative sourde ; c'est l'ensemble CLOTH (Hickey l'appelle SOFT) qui se distingue de LOT ici⁸².

Hickey ([Hic04a], 46-50) note la formation d'un nouveau dialecte urbain à Dublin ces vingt dernières années. Le grand nombre d'immigrants, dû à une prospérité sans précédent de la ville sur cette période, constitue un groupe d'individus socialement mobiles qui cherchent à se différencier de ce qu'ils perçoivent comme le dialecte populaire. Ceci a eu

⁸²Souvenons-nous que les mots de l'ensemble CLOTH avaient naguère une voyelle longue et relativement fermée dans une forme très conservatrice de RP (voir Section 3.1.1).

TAB. 3.3 – Changement en chaîne récent à Dublin. D’après ([Hic04a], 47)

<cot>	[kɒt] → [kɔt]
<caught>	[kɒ:t] → [kɔ:t]
<time>	[tʌɪm] → [tɔɪm]
<toy>	[tɔɪ] → [tɔɪ], [tɔɪ]

pour conséquence un véritable changement en chaîne dans le système vocalique, qui (avec d’autres traits de prononciation) tend à se diffuser au-delà de la métropole puisque c’est cette forme de prononciation que vont adopter de fait les locuteurs de l’Irlande du Sud à la recherche d’un accent non régional. Ce changement en chaîne se traduit par un mouvement des timbres dans le sens inverse des aiguilles d’une montre, ce mouvement étant également répercuté sur le premier élément des diphtongues. Les effets de ce changement sont récapitulés dans la Table 3.3.

Ce changement est encore plus marqué lorsque la voyelle précède /r/ ; à Dublin, c’est une approximante rétroflexe : [ɻ]. Cette nouvelle forme de prononciation possède en outre, parmi les traits vocaliques, un premier élément antérieur et relativement fermé dans MOUTH : [æʊ] ou [ɛʊ].

Le TH-stopping (également attesté à Liverpool) est cantonné à l’Irlande du Sud ([Hic04a], 38). Il est attesté dans le dialecte urbain traditionnel de Dublin au moins depuis le 17^e siècle ([Hic04a], 45). La lénition du /t/ en position intervocalique après une voyelle accentuée est également typique de cet accent. Des réalisations diverses montrent que ce processus peut atteindre différents niveaux de complétude selon le locuteur : <water> [wɑ:t̪ə] → [wɑ:ɹə] → [wɑ:hə] → [wɑ:ə]. Le /l/, traditionnellement clair dans toutes les positions, semble avoir une variante sombre dont la distribution est celle de l’anglais standard ([Hic04a], 49).

3.2.4.2 Le nord

Sur le plan du système vocalique, on remarque une possible homophonie entre <fir>, <fair> et <fur>. Si l'on s'en tient à la comparaison avec l'anglais standard, cela implique une convergence NURSE-SQUARE ; <fir> et <fur>, quant à eux, appartiennent déjà à l'ensemble NURSE en anglais standard. Contrairement à l'Irlande du Sud, on trouve au nord la convergence FOOT-GOOSE, caractéristique de l'anglais d'Écosse ([Wel82], 437). Milroy et Milroy ([MM78], 25) distinguent un ensemble lexical composé de 22 items (<pull>, <took>, <look>, etc.) comportant une voyelle tantôt réalisée [ʌ], tantôt [ɯ]. La durée vocalique semble ne pas être pertinente dans le système, mais plutôt influencée, comme c'est le cas en Écosse, par le contexte phonétique (pour plus de détails, voir [Wel82], 439-440). Les ensembles TRAP, BATH et PALM ont le même phonème. Les ensembles LOT et THOUGHT ont parfois deux phonèmes différents.

Au niveau de la réalisation phonétique, la voyelle de DRESS tend à être produite comme une diphtongue centripète : [eə]. La voyelle de *happy* s'approche de [e]. La voyelle de l'ensemble FOOT-GOOSE présente un timbre antérieur.

Concernant les consonnes, le TH-stopping est restreint aux cas où /θ/ et /ð/ sont suivis de /r/ (cf. 3.2.4.1, 3.2.1.3).

À Belfast, d'après Hickey ([Hic04a], 51), on a sur le plan de la réalisation :

- le son [ɑ] pour LOT ;
- la voyelle [ɛ:] dans <beat> ;
- une ouverture de DRESS ;
- /aʊ/ avant /l/ dans les monosyllabes, e.g. <old> [aʊl] ;
- TRAP se ferme devant les vélares (<back> [bɛk], <bag> [bɛg]) et après /k/ (e.g. cap [kɛp]) ;
- FLEECE et GOOSE ont des réalisations brèves ;
- KIT est plus ouvert, voire centralisé.

Notre revue des particularités phonético-phonologiques des dialectes des Îles Britanniques est loin d'être exhaustive. La multiplicité des facteurs de variation – âge, profil

sociologique, dialecte urbain *vs* rural – nous a conduit à simplifier une réalité hautement complexe. Nous nous sommes cependant efforcé de lister les principaux phénomènes phonologiques et réalisationnels susceptibles d'intervenir dans l'interprétation de l'analyse acoustique de la Partie II.

Deuxième partie

Description phonético-phonologique des dialectes et classification automatique

Chapitre 4

Introduction à l'analyse empirique

Nous avons introduit jusqu'ici le cadre théorique et descriptif de notre étude. Dans le Chapitre 2, les concepts nécessaires à la compréhension de notre étude empirique ont été abordés. Nous avons notamment présenté le thème de la variation dans une perspective dialectologique, puis à travers le point de vue de la phonétique acoustique. Dans le Chapitre 3, nous avons examiné la notion d'anglais standard ; les particularités phonétophonologiques des accents des Îles Britanniques ont ensuite été passées en revue. La Partie II est construite autour de l'étude phonétique du corpus *Accents of the British Isles*. Après avoir décrit le corpus (Section 4.1) et rapporté une expérience de perception visant à évaluer l'homogénéité des dialectes représentés (Section 4.2), nous traitons l'analyse du rythme dans le Chapitre 5. Nous abordons ensuite l'étude des systèmes vocaliques dans un cadre de phonétique acoustique traditionnelle (i.e. sur la base des formants et de la durée) dans le Chapitre 6. Enfin, nous présentons au Chapitre 7 des expériences de classification automatique et évaluons l'apport de représentations non conventionnelles issues d'une caractérisation multidimensionnelle des voyelles.

TAB. 4.1 – Dialectes du corpus ABI.

Abréviation	Dialecte	Lieu	Locuteurs (H/F)
<i>brm</i>	Birmingham	Birmingham	10/10
<i>crn</i>	Cornwall	Truro	11/9
<i>ean</i>	East Anglia	Lowestoft	9/10
<i>eyk</i>	East Yorkshire	Hull	13/12
<i>gla</i>	Glasgow	Glasgow	10/10
<i>ilo</i>	Inner London	Londres	10/11
<i>lan</i>	Lancashire	Burnley	11/10
<i>lvp</i>	Liverpool	Liverpool	10/10
<i>ncl</i>	Newcastle	Newcastle	10/9
<i>nwa</i>	North Wales	Denbigh	10/11
<i>roi</i>	Republic of Ireland	Dublin	10/10
<i>shl</i>	Scottish Highlands	Elgin	11/11
<i>sse</i>	Standard Southern English	Londres	10/6
<i>uls</i>	Ulster	Belfast	10/10

4.1 Description du corpus Accents of the British Isles (ABI)

Le corpus Accents of the British Isles a été enregistré dans les premiers mois de l'année 2003 par une équipe d'ingénieurs de l'Université de Birmingham ([DRBT04]). Il s'agit d'une base de données payante contenant des enregistrements censés représenter 14 dialectes des Îles Britanniques. Les enregistrements ont eu lieu dans des salles calmes (souvent dans des bibliothèques publiques). Le signal a été capté par le biais d'un micro-casque relié à une carte-son externe, et a été numérisé directement au format PCM Windows, mono, avec un taux d'échantillonnage de 22 050 Hz et une quantification de 16 bits. Ces dialectes, les abréviations que nous utiliserons pour les désigner, le lieu précis de l'enregistrement correspondant et le nombre de locuteurs sont détaillés dans la Table 4.1. Les lieux d'enregistrements sont localisés sur la carte de la Figure 4.1.

4.1. Description du corpus Accents of the British Isles (ABI)



FIG. 4.1 – Dialectes du corpus ABI.

La version de ABI que nous avons utilisée comporte trois composantes :

- une série de 19 mots à structure /hVd/ répétés cinq fois par chaque locuteur⁸³ ;
- une série de 20 phrases brèves et phonétiquement riches (que nous n'avons pas utilisées) ;
- un passage lu d'environ 300 mots (subdivisé en 3 fichiers son), reproduit dans l'Annexe A.

Idéalement, les locuteurs devaient avoir entre 18 et 50 ans ; les limites réelles sont de 16 et 79 ans. Ils ont été recrutés par le biais de publicités dans la presse et les radios locales. Les sujets percevaient 15 livres sterling pour leur participation. Nombre d'entre eux furent trouvés sur place à la dernière minute. Tous savaient que les enregistrements avaient pour but de mettre en évidence leur accent.

Nous ne saurions trop insister sur le fait qu'aucune donnée individuelle sur l'âge, la catégorie socio-professionnelle et l'histoire linguistique des locuteurs n'est disponible, ce qui nous contraint à limiter notre étude aux deux facteurs explicites utilisables, savoir le sexe et l'origine géographique des participants. Fort de ce constat, nous avons soumis des échantillons du corpus au Professeur Francis Nolan du Phonetics Laboratory de l'Université de Cambridge afin qu'il juge de la qualité des données.

4.2 Homogénéité et typicalité des dialectes de ABI : analyse auditive

Bien que notre méthode s'inspire d'expériences classiques dans le domaine de la perception, nous ne prétendons pas avoir élaboré un protocole des plus rigoureux, notamment en raison de contraintes de temps. L'analyse a été restreinte aux locuteurs masculins de la base ABI. Notre expert a pu écouter la première partie du texte lu produite par tous les locuteurs du corpus. L'analyse était divisée en 14 parties ; une par dialecte.

⁸³Voici la liste : *heed, hid, head, had, hard, hod, hoard, hood, who'd, Hudd, heard, hade, hide, hoid, hoed, howd, heered, hared, hured.*

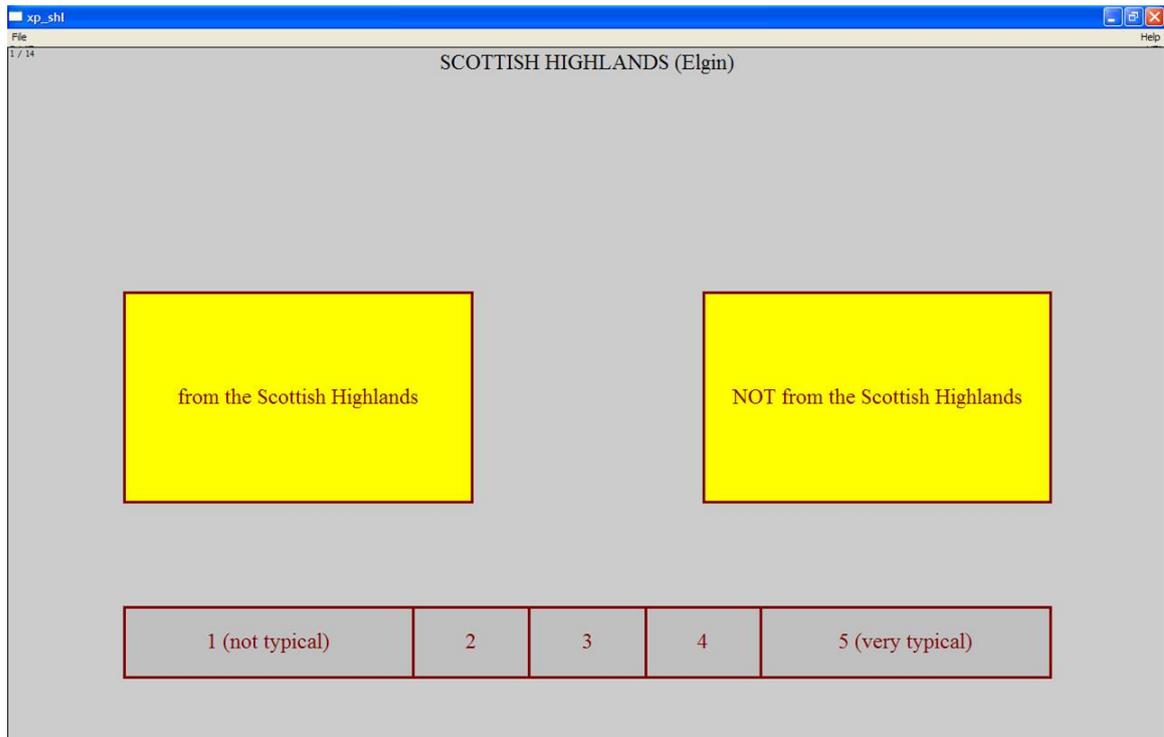


FIG. 4.2 – Interface pour l’analyse perceptive du corpus ABI.

Nous avons mis au point pour la circonstance une interface avec le logiciel Praat (voir Figure 4.2).

Installé dans la chambre anéchoïque du laboratoire de phonétique de l’Université de Cambridge, notre expert entendait via un casque audio la première partie du passage lu (Annexe A, Fichier 1) et devait décider s’il s’agissait d’un locuteur du dialecte en cours de test ou non. Dans un second temps, il devait choisir un chiffre entre 1 et 5 indiquant si les caractéristiques phonétiques du passage en faisait un extrait typique du dialecte identifié. Il avait le loisir d’interrompre à tout moment, mais de manière définitive, le son en cours de lecture afin de prendre une décision. Nous avons inclus pour chaque dialecte, en plus des fichiers son du dialecte en cours, 3 fichiers provenant d’autres dialectes ; un taux élevé de fausses acceptations pouvant remettre en cause la fiabilité de l’analyse. À la fin de chacune des 14 sessions, un questionnaire visant à fournir une appréciation plus générale sur le dialecte était soumis. L’expert devait :

- y indiquer sur une échelle de 1 à 5 le degré de conformité de tous les locuteurs du

- dialecte à son idée d'une prononciation typique de la zone concernée ;
- estimer, sur une échelle de 1 à 5, l'homogénéité des locuteurs du dialecte ;
- entourer, dans le passage reproduit sur le questionnaire, les voyelles et les consonnes lui ayant permis d'arrêter sa décision ;
- dire si des indices articulatoires permanents non linguistiques (« articulatory settings »), tels qu'un degré important de nasalité, de raucité, etc., l'avaient aidé dans sa tâche⁸⁴ ;
- relever l'éventuelle présence de traits d'intonation discriminants ;
- ajouter toute remarque qu'il jugeait nécessaire.

En considérant que les deux premiers points de l'énumération constituent un indice de fiabilité de nos données, nous les avons récapitulés dans les Tables 4.2 et 4.3. Ces deux tableaux contiennent les mêmes données ; la différence apparaît au niveau du tri. Dans la Table 4.2, les dialectes sont d'abord triés en fonction de leur conformité au prototype (du moins typique au plus typique), puis en fonction de l'homogénéité des locuteurs qui les composent (du moins homogène au plus homogène), attendu que le degré de confiance à allouer à chaque dialecte est proportionnel à sa conformité au prototype et à son homogénéité. Dans la Table 4.3, l'ordre du tri est inversé, donnant ainsi une prépondérance à la valeur de l'homogénéité. Intuitivement, ces deux indices méritent une pondération différenciée, la conformité au prototype nous paraissant plus déterminante dès lors qu'il s'agit de cerner ce que nous résumons avec le concept de fiabilité ; or nous avons renoncé à proposer un score composite, d'où les deux tableaux.

Ces jugements font apparaître que le dialecte *ilo* présente les plus mauvais scores. Une discussion postérieure à l'analyse auditive avec le Professeur Nolan nous a convaincu d'abandonner ce sous-ensemble du corpus. En effet, *ilo* est constitué de locuteurs d'origines ethniques trop diverses pour former un tout cohérent ; ce dialecte ne pouvait, aux dires de notre expert, en aucun cas être comparé aux 13 autres.

⁸⁴Ces « articulatory settings » ont une fonction sociolinguistique avérée. En 1999, Stuart-Smith ([SS99]) proposait une des rares analyses de la qualité de la voix dans une optique sociolinguistique. C'est ce point qui nous a conduit à intégrer dans le questionnaire la notion de « articulatory settings ».

TAB. 4.2 – Jugements de conformité au prototype.

Dialecte	Conformité	Homogénéité
ilo	1	5
nwa	1	2
crn	2	4
brm	3	5
ean	3	5
eyk	3	5
lan	3	3
shl	3	3
sse	3	3
uls	3	2
lvp	4	3
gla	4	2
ncl	5	4
roi	5	3

TAB. 4.3 – Jugements d’homogénéité du dialecte.

Dialecte	Conformité	Homogénéité
ilo	1	5
brm	3	5
ean	3	5
eyk	3	5
crn	2	4
ncl	5	4
lan	3	3
shl	3	3
sse	3	3
lvp	4	3
roi	5	3
nwa	1	2
uls	3	2
gla	4	2

Le dialecte *nwa* arrive en deuxième position des accents les moins typiques. Il est possible de relier ce jugement à ce que les auteurs du corpus disent de *nwa* dans la documentation : « Probably the worst set of data recorded, in terms of the typicality of the accents. »

Nous mettrons encore une fois en parallèle le score de conformité au prototype de *crn* et les propos des auteurs du corpus : « Accents appear very much diluted especially among younger people. Some people seemed to have Cornish accents when speaking but when reading they lost it very quickly... »

Sur le plan de l'homogénéité, les plus mauvais scores, ex æquo, reviennent à *ilo*, *brm*, *ean* et *eyk*. Nous ne reviendrons pas sur *ilo*. Dans ses remarques sur *brm*, l'expert avoue qu'il a été réduit à utiliser son intuition pour certains locuteurs qui semblaient tout juste posséder un substrat d'accent de Birmingham. Les auteurs du corpus ajoutent : « the accents are generally not all that strong. » Concernant *ean*, un début d'explication concernant le faible degré estimé d'homogénéité réside peut-être dans le fait que deux locuteurs étaient bien au-delà de la fourchette d'âges de 18-50 ans qui avait été définie *a priori*. Quant à *eyk*, nous relèverons cette phrase dans la documentation : « The level of literacy was lower for a number of the subjects in Hull than in other areas. »

Chapitre 5

Le rythme

5.1 Préambule

À l’instar de l’identification automatique du dialecte, l’analyse du rythme dans les variétés d’une même langue s’est inspirée de l’étude du rythme dans un contexte multilingue. Le sujet est vaste, et mériterait un exposé plus conséquent, qui inclurait par exemple des aspects perceptifs et cognitifs. Nous nous contenterons néanmoins, vu la nature de nos données, de tenter de répondre à la question suivante : quels peuvent être les corrélats physiques de l’impression de rythme dans la parole et, une fois ces corrélats identifiés, nous sont-ils utiles pour distinguer les dialectes de notre corpus ? Nous avons déjà proposé une ébauche de réponse dans [FP04c, FP04a].

La question de l’intonation (nous entendons plus spécifiquement : l’étude de la fréquence fondamentale), n’est pas abordée dans nos travaux faute de temps, mais aussi parce qu’elle a été traitée récemment dans un contexte multi-dialectal britannique, notamment via le projet IViE (e.g. [GPNF00, GPN01, GP02, KGCR05]).

5.2 Genèse historique de la notion phonétique de rythme

L'étude du rythme s'est longtemps appuyée sur ce qu'il était convenu d'appeler la métrique, autrement dit, l'étude des conventions de rythme dans la poésie. C'est vraisemblablement encore la norme en 1939 lorsque André Classe précise dans le titre de sa monographie qu'il s'agit de « English Prose » (par opposition à « verse » ; c'est nous qui soulignons). Un auteur du 18^e, Steele ([Ste69]), est parfois cité comme l'un des pères de l'étude objective du rythme de la parole. Ce dernier adopte un système extrêmement riche de transcription du suprasegmental ; on retiendra par exemple que ses portées musicales notent des intervalles de l'ordre du quart de ton. Phénomène assez rare à l'époque, Steele ([Ste69], 18) affiche d'emblée sa distance par rapport aux tant vénérés modèles de l'Antiquité : « the Greek feet [would not], under all their various names, answer in any suitable degree to the rhythmus [sic] of our language . . . ». Si Wells ([Wel82]) ne se montre guère loquace concernant la notion de rythme dans les dialectes (nous y reviendrons plus loin), Steele ([Ste69], 34), l'est à peine moins :

Take three common men ; one a native of Aberdeenshire, another of Tipperary, and the third of Somersetshire ; and let them converse together in the English language, in the presence of any gentleman of the courtly tone of the metropolis ; his ears will soon inform him, that every one of them talks in a tune very different from his own, and from each other ; and that their difference of tone is not owing merely to *loud* and *soft*, but to a variety both of melody and of measure, by a different application of *accents*, acute and grave ; and of *quantity*, short and long ; and of *cadence*, light and heavy. (Soulignement ajouté, italiques originales)

Steele est souvent cité comme l'auteur qui a imposé la notion d'isochronie⁸⁵ ; à notre connaissance, ceci n'est pas usurpé. Il affirme en effet ([Ste69], 68) : « the rhythmical pulsation is regularly periodical » et il insiste plus loin ([Ste69], 72) « rhythmus is to keep, by its pulsation, all the cadences of an equal length. ». On lui doit également une distinction nette entre rythme et débit de parole ; c'est au débit qu'il semble faire référence lorsqu'il affirme ([Ste69], 69) : « *Time* is measured by *pulsation*, quicker or slower . . . ». Steele peut

⁸⁵Cette notion a tout au plus le statut d'hypothèse, et certainement pas de théorie, comme on peut parfois le lire.

être crédité d'avoir émis l'hypothèse que le rythme de la parole serait un phénomène subjectif ([Ste69], 78) : « I consider our sense of *rhythmus* to be much more instinctive than rational. . . ». Cette idée est d'ailleurs reprise par le poète du 19^e, Coventry Patmore (cité dans [Boi80], 66-67) : « Yet, all-important as this time-beat is, I think it demonstrable that, for the most part, it has no material and external existence at all, but has its place in the mind, which craves measure in everything ». Enfin, nous concluons avec Abercrombie ([Abe65], 26) que Steele est certainement à l'origine de l'idée de « stress-timing » en anglais, notion largement employé au 20^e siècle dans les comparaisons multilingues.

L'idée d'isochronie est, quoi qu'il en soit, sous-jacente dans des travaux antérieurs ; nous relèverons pour l'anecdote ce qu'écrivait Gascoigne en 1575 dans son traité de versification à propos de quiconque lit la poésie de Chaucer ([Gas68], 34) :

he shall finde that although his lines are not alwayes of one selfe same number of Syllables, yet beyng redde by one that hath vnderstanding, the longest verse and that which hath most Syllables in it, will fall (to the eare) correspondent vnto that whiche hath fewest sillables in it : and like wise that whiche hath in it fewest syllables, shalbe founde yet to consist of woordes that haue suche naturall sounde, as may seeme equall in length to a verse which hath many moe sillables of lighter accenttes.

Si nous avons pris la peine de détailler la pensée de Steele, c'est que trois points qu'il mentionne font encore l'objet d'études de nos jours (l'isochronie, la dimension subjective et la notion de « stress-timing »), et que la notion de débit de parole reçoit une attention toute particulière dans les études phonétiques récentes ([BR03, DW03, DSA⁺04, PFR04, Del06, DFP06, JNQ03, Jan04]). Dans cette section, nous utiliserons « accentuel » et « syllabique » pour traduire « stress-timed » et « syllable-timed », respectivement.

En suivant une progression chronologique, on peut noter au passage que Daniel Jones écrivait en 1918, dans la première édition de son *Outline of English Phonetics* : « There is a strong tendency in connected speech to make stressed syllables follow each other as far as possible at equal distances » (cité dans [CM99], 235). L'une des premières analyses empiriques du rythme en anglais est celle de Classe en 1939 ([Cla39]). L'auteur définit le rythme en ces termes ([Cla39], 50) : « rhythm, generally speaking, is known to be due to the recurrence at more or less regular intervals of one given phenomenon ». On relève

ici l'emploi de « more or less », qui indique que certaines déviations par rapport à une cadence strictement régulière sont envisagées. Il précise que la taille de l'intervalle entre les accents (« stresses ») varie beaucoup, mais uniquement dans certaines proportions. L'auteur glisse ensuite petit à petit vers la notion de rythme subjectif, et identifie deux facteurs, sur le plan de la production et de la perception, qui concourent à l'isochronie de la parole ([Cla39], 51) :

It is highly probable, of course, that we do tend to equalize the groups we perceive and to minimize the differences. On the other hand, it is not less likely that we tend to place the stresses so as to facilitate the perception of groups as equal groups.

À partir de mesures de durées effectuées sur plusieurs locuteurs produisant des phrases comparables, Classe ([Cla39], 63-64) affirme : « none of them [the speakers] is very far from nearly perfect objective isochronism. » La conclusion de ses travaux fait apparaître qu'en anglais, il y a une tendance à l'isochronie qui est contrariée par certains facteurs comme ce qu'il nomme, sans vraiment le définir, « phonetic necessity » ([Cla39], 87). Cette tendance à l'isochronie se traduit par le fait qu'à nombre d'accents égal, le syntagme qui contient le plus de syllabes est produit à un débit syllabique supérieur. Enfin, chez Classe, la rupture entre l'étude phonétique de la poésie et celle de la parole est consommée ([Cla39], 135) : « It appears, therefore, that the rhythm of artistic prose must be the result of a more or less complete stylization of the rhythm of speech. »

5.3 Perspectives multilingues : « stress-timing » et « syllable-timing »

Abercrombie ([Abe65], 26) rapporte qu'en 1940, Arthur Lloyd James⁸⁶ avait identifié deux types de rythme dans les langues du monde : « machine-gun rhythm » *vs* « morse-code rhythm ». Ces termes impressionnistes – qui ont le mérite de ne faire appel à aucune construction phonologique – ont été abandonnés au profit de « syllable-timed » et « stress-

⁸⁶Lloyd James, phonéticien d'origine galloise, a rejoint Daniel Jones en 1920. En 1933, il devient le deuxième « Professor of Phonetics » de Grande Bretagne (chronologiquement ; le premier ayant été Jones). Il a largement contribué au développement de la notion de « BBC English » ([CM99], *passim*).

timed », respectivement. En 1945, Pike ([Pik45], 34) introduit ces nouveaux termes ; son intérêt pour la question du rythme lui vient de ce qu'en enseignant l'anglais (américain) à des Latino-américains, il constate que sensibiliser ses étudiants aux différences de rythme entre l'anglais et l'espagnol est hautement bénéfique ([Pik45], 35) :

Many non-English languages (Spanish, for instance) tend to use a rhythm which is more closely related to the syllable than the regular stress-timed type of English ; in this case, it is the syllables, instead of the stresses, which tend to come at more-or-less evenly recurrent intervals - so that, as a result, phrases with extra syllables take proportionately more time, and syllables or vowels are less likely to be shortened and modified.

Il continue ([Pik45], 35) : « For Latin-Americans, the shift from their tendency toward a syllable-timing rhythm to the English normal stress-timing rhythm is highly necessary, but comprises one of their greatest problems. » Ce transfert du rythme de L1 vers L2 pour les apprenants de l'anglais est également attesté dans l'étude de Bond et Fokes ([BF85]) à partir de l'examen du phénomène de compression temporelle de la voyelle d'un radical sous l'effet de la suffixation. Schaeffler ([Sch01]) démontre le transfert de rythme des Vénézuéliens apprenant l'allemand à partir des mesures utilisées dans les publications de référence de Ramus et collègues ([RNM99]) et Grabe et Low ([GL02]).

D'après Pike, il existe donc dans le monde des langues dont le rythme est caractérisé par la récurrence des syllabes à intervalles quasi-réguliers, et d'autres langues où ce sont les accents (« stresses ») et non les syllabes qui importent. Pour Pike, une seule et même langue peut avoir les deux types de rythme ; nous n'avons pas retrouvé cette remarque ailleurs. Voici ce qu'il dit à propos de ce qu'il nomme « spoken chant » en anglais ([Pik45], 35) : « English also has a rhythmic type which depends to a considerable extent upon the number of its syllables, rather than the presence of a strong stress, for some of its characteristics of timing ; in English, however, the type is used only rarely ».

On peut considérer qu'entre l'étude de Classe de 1939 et celle de Dauer de 1983 ([Cla39] et [Dau83]), les recherches empiriques sur le rythme de la parole se sont essentiellement concentrées sur la notion d'intervalles entre accents et la recherche de l'isochronie. Dauer s'éloigne quelque peu de cette notion en affirmant que la différence entre stress-timing

et syllable-timing est imputable aux différentes structures syllabiques des langues, à la réduction vocalique et à la réalisation phonétique de l'accent ([Dau83], 51). Ces caractéristiques sont connues au moins depuis l'étude de Delattre et Olsen ([DO69]), qui avait déjà mis en avant la contribution des différentes structures syllabiques, de la réduction vocalique et de l'accent à ce qu'ils nomment « phonic impression », à notre avis, une notion qui englobe entre autres ce que l'on range sous l'étiquette « rythme » de nos jours. Les auteurs avaient en effet comparé quatre langues et leur conclusion mettait en évidence les trois paramètres que nous venons de citer pour expliquer les différences d'impression phonique entre deux langues romanes et deux langues germaniques. Concernant les syllabes, deux facteurs, certes liés mais néanmoins distincts, sont à prendre en compte : dans les langues germaniques, les structures syllabiques sont non seulement plus complexes, mais également plus variées ; ou encore, comme le formulent Delattre et Olsen ([DO69], 167) : « the syllables of the two Germanic languages show both greater structural variety and greater complexity than those of the two Romance languages. » (Soulignement ajouté). Nous insistons délibérément sur ce point car il possède de fortes implications dès lors qu'il s'agit de concevoir des mesures objectives du rythme.

La notion d'isochronie, en tant que récurrence à intervalles réguliers des accents ou des syllabes, pour les langues accentuelles ou syllabiques, respectivement, est parfois rejetée en raison de l'absence d'isochronie stricte, telle qu'on peut la mesurer très précisément dans le signal avec les moyens technologiques actuels. Or démontrer l'absence d'isochronie de cette manière revient à inférer que l'appareil auditif et cognitif humain possède le même pouvoir de résolution temporelle qu'un logiciel de visualisation du signal, ce qui est aberrant. C'est en cela que nous ne rejoignons pas les conclusions fortes de l'étude de Pamies Bertrán ([PB99]). L'auteur, mesures de durée dans plusieurs langues à l'appui, rejette entièrement la notion d'isochronie et conclut sans nuance que l'idée de rythme dans les langues n'est peut-être qu'une simple métaphore musicale, héritée d'une esthétique normative de l'Antiquité, et qu'il faudrait aller jusqu'à envisager « the absolute lack of any kind of rhythm » ([PB99], 127). Cette dernière remarque est en contradiction totale avec ce que nombre de travaux sérieux rapportent : par exemple, [Kla76], 1218, rappelle qu'il est facile pour un locuteur anglophone de battre la mesure avec un doigt, et que chaque battement correspond à peu près au début d'une voyelle accentuée. L'étude de

Pamies-Bertrán tombe dans un travers que Benguerel et D’Arcy résument ainsi ([BD86], 244) :

It seems surprising that anyone interested in finding evidence for *perceptual* regularity would expect measurements [...] of the *acoustic* signal to provide such evidence.

Afin de tester véritablement l’isochronie, la première étape consisterait à obtenir une connaissance précise des seuils différentiels de perception de la durée dans la parole⁸⁷. L’étude de Benguerel et D’Arcy ([BD86]) démontre d’ailleurs que, la plupart du temps, une séquence de syllabes perçues comme isochrones ne le sont pas sur le plan acoustique⁸⁸. Mais ceci ne serait pas suffisant : comment savoir où mesurer exactement cette isochronie ? Si l’on prend l’exemple de la récurrence des syllabes, il semble intuitivement justifié de mesurer les intervalles entre les attaques, ou encore entre les milieux temporels des syllabes successives. Or ceci ne correspond pas à la manière dont l’être humain produit et perçoit l’isochronie ; en effet, Patel et collègues ([PLN99]) rappellent que des expériences ont démontré que lorsque des sujets avaient pour consigne de produire une série de syllabes (e.g. « bad - sad - bad - sad, etc. ») avec le plus de régularité possible, les locuteurs introduisaient une déviation systématique par rapport à une parfaite isochronie dans la succession des attaques. De plus, cette anisochronie était perçue comme isochronique. De ce type d’expérience est née la notion de « P-centers » (<P> pour « perceptual »), idée selon laquelle l’organisation temporelle d’une suite de syllabes ne peut pas être caractérisée, sur le plan de la perception, par la localisation d’un point fixe (comme l’attaque) sur chacune de ces syllabes. En d’autres termes, la pulsation caractérisant l’occurrence d’une syllabe sur le plan de la perception n’a à ce jour pas été localisée avec précision, et c’est probablement cette pulsation qui permettrait de tester le caractère isochrone d’une

⁸⁷Dans son étude de 1972, Rossi ([Ros72]) établit, à partir de la voyelle [a] du français et de sujets locuteurs du français méridional, un seuil constant d’environ 30 ms pour des stimuli de référence allant de 60 à 140 ms ; puis, entre 130 et 290 ms, c’est un seuil relatif de 22,5 % en moyenne. Pour la même tâche, des sujets tchécoslovaques (locuteurs d’une langue où la quantité phonologique existe ; l’auteur ne précise pas s’il s’agit du tchèque ou du slovaque) n’obtiennent pas des seuils plus fins. L’étude de Klatt et Cooper ([KC75]) fait apparaître des seuils différentiels variables selon la position dans la phrase de la séquence [il] dans <dealer> ; ces seuils vont de 22 à 59 ms.

⁸⁸Dans cette étude, les syllabes [ta] et [na] sont enregistrées par un locuteur francophone et organisées en séquences isochrones. Puis, ces séquences sont soumises à des sujets anglophones, francophones et nipponophones.

séquence de syllabes. Nous n'abordons pas ce débat à travers nos données ; mais il nous a paru important de rappeler qu'aucune véritable preuve en défaveur de l'isochronie n'a été apportée jusqu'ici en phonétique. D'ailleurs, récemment, Bouzon et Hirst ([BH04]) ont démontré la tendance à l'isochronie de l'anglais à partir du modèle de Jassem. Ce dernier considère qu'il y a deux types d'unités rythmiques en anglais : les « Narrow Rhythm Units » (NRU), qui comprennent une syllabe accentuée et toutes les syllabes non accentuées suivantes qui appartiennent au même mot, et les anacrouses (ANA), constituées de toutes les syllabes non accentuées ne faisant pas partie des NRU. Les auteurs démontrent que la compression temporelle des phonèmes est bien plus importante dans les NRU que dans les ANA, ce qui fait des NRU de bons candidats pour la recherche de l'isochronie en anglais. D'ailleurs l'étude de [BH04] n'est pas la seule à continuer d'envisager une éventuelle isochronie : par exemple, [Koh91] plaide en faveur d'une isochronie faible, [WD04] recommandent aux études futures de continuer de considérer le rythme comme une succession d'événements plus ou moins isochrones, Williams et Hiller ([WH94]) considèrent l'isochronie comme un concept pertinent pour l'anglais.

Deux études de ces dernières années ont particulièrement marqué le domaine du rythme de la parole : il s'agit de celle de Ramus et collègues ([RNM99]) et de celle de Grabe et Low ([GL02]). Ces études abordent essentiellement la question des classes de rythme (e.g. « stress-timed » *vs* « syllable-timed ») dans les langues du monde. L'aspect le plus directement lié à nos travaux concerne les mesures utilisées pour caractériser le rythme.

Ramus et collègues ([RNM99], [RM99]) rappellent que les psycholinguistes se sont penchés sur la question du rythme de la parole dans une optique développementale avec l'intuition que selon le type de langue, accentuel ou syllabique, les nourrissons segmentent le flux de parole différemment. En outre, les études citées par [RNM99] ont montré que le nourrisson est capable de distinguer sa langue maternelle d'une langue appartenant au type de rythme opposé sur la base du rythme, et, qui plus est, qu'il est à même de discriminer deux langues – autres que sa langue maternelle – à condition qu'elles appartiennent à des classes rythmiques différentes ([RNM99], 266-267). La conception des linguistes se-

lon laquelle les langues du monde appartiendraient à des classes rythmiques jouit donc d'un soutien incontestable provenant de la psycholinguistique. Il convient cependant de nuancer cette affirmation : dans une étude de Miller ([Mil84]), des échantillons de sept langues sont présentés à des phonéticiens anglais et français et à des non-phonéticiens de ces deux mêmes langues. Les sujets ont pour tâche de décider si la langue qu'ils entendent est accentuelle ou syllabique. L'arabe est considéré par tous les groupes de sujets comme très accentuel, l'espagnol est très accentuel pour les phonéticiens anglais et les Français non-phonéticiens alors qu'il n'affiche qu'une tendance vers le pôle accentuel pour les phonéticiens français. L'indonésien est très syllabique pour les phonéticiens ; aucune catégorisation n'émerge en revanche pour les non-phonéticiens. Les résultats pour le polonais sont contradictoires : les phonéticiens anglais le classent comme très accentuel, alors que les Anglais non-phonéticiens jugent qu'il est très syllabique. Le finnois n'est ni accentuel, ni syllabique pour tous les groupes. Le japonais ne semble pas plus facile à classer. Enfin, le yoruba est identifié comme syllabique par les deux groupes de phonéticiens alors que les deux autres groupes ne parviennent pas à lui assigner une classe précise. Cette étude laisse envisager l'existence d'un nombre de classes supérieur à 2 ou encore la possibilité que ces langues présentent en partie les propriétés d'une classe et en partie celles de l'autre. Les expériences de Ramus et Mehler ([RM99]), où des sujets francophones accomplissent une tâche de discrimination à partir de la resynthèse de phrases en anglais et en japonais avec différents types de dégradation du signal, tentent de mettre en évidence les caractéristiques pertinentes dans le signal pour une telle tâche. Il apparaît que les sujets sont capables de séparer ces deux langues lorsque le contour intonatif est aplati, que les voyelles sont remplacées par un [a], et les consonnes, par un [s]. Lorsque seul le contour intonatif original est préservé, la discrimination n'est pas possible. Ceci confirme le bien-fondé (voir paragraphe suivant) des mesures utilisées dans les travaux de [RNM99] et [GL02].

La véritable avancée que proposent les études s'inscrivant dans la lignée de [RNM99] et [GL02] réside dans le fait que leurs auteurs ont su remettre en question des prémisses

probablement mal fondées qui avaient été jusque là considérées comme acquises. En effet, les études antérieures se sont attachées à utiliser des constructions phonologiques telles que la syllabe ou l'accent pour décrire le rythme. Or, si ces concepts phonologiques ont des corrélats phonétiques certains, il n'est en revanche pas aisé de les définir, surtout dans une perspective multilingue. De plus, bien qu'il semble que le nouveau-né soit capable de compter des syllabes ([RNM99], 270), il s'appuie très vraisemblablement sur des événements sonores élémentaires (comme la saillance acoustique des segments vocaliques) et certainement pas sur des constructions phonologiques complexes. En conséquence, les mesures proposées dans ce type de littérature s'appuient sur des entités physiques affranchies de la moindre considération théorique : les intervalles vocaliques et consonantiques⁸⁹. Certaines conventions doivent cependant être appliquées concernant les segments qui, phonétiquement, peuvent appartenir à l'une ou à l'autre de ces catégories. Par exemple, pour Ramus *et al.* ([RNM99], 271) les glides pré-vocaliques sont considérés comme des consonnes, et les glides post-vocaliques (souvent le second élément d'une diphtongue) tombent dans la catégorie des voyelles. Un intervalle vocalique est borné par le début et la fin d'une ou plusieurs voyelles (sans consonne intermédiaire), et un intervalle consonantique est constitué d'une ou plusieurs consonnes successives (sans voyelle intermédiaire) ; les pauses sont exclues du calcul.

Les auteurs caractérisent ce qu'ils considèrent être les corrélats physiques du rythme de la parole à l'aide des trois indices suivants calculés à l'échelle de la phrase :

- le pourcentage de durée vocalique ; il s'agit plus précisément de la durée des intervalles vocaliques divisée par la durée totale de la phrase, que l'on note %V ;
- l'écart-type de la durée des intervalles vocaliques, noté ΔV ⁹⁰ ;
- l'écart-type de la durée des intervalles consonantiques : ΔC .

Bien que ces mesures n'aient jamais été utilisées auparavant, les auteurs reconnaissent volontiers qu'elles découlent de phénomènes phonologiques connus, et ajoutent : « this

⁸⁹Ces derniers sont souvent appelés « intervalles ».

⁹⁰L'idée d'utiliser un écart-type pour calculer des variations de durée n'est pas nouvelle : en 1982, Roach ([Roa82], 74) avait calculé l'écart-type des durées de syllabes dans six langues. Ses résultats ne faisaient apparaître aucune discrimination possible entre classes de rythmes sur la base de cette mesure.

study is meant to be an implementation of the phonological account of rhythm perception. » Lorsque les langues étudiées dans [RNM99] sont représentées dans un plan $\Delta C / \%V$, on voit apparaître deux classes : d'une part, le catalan, l'espagnol, le français et l'italien, et, d'autre part, l'anglais, le néerlandais et le polonais. Le japonais, quant à lui, semble ne pouvoir être intégré dans aucune catégorie. Les deux classes mentionnées correspondent, respectivement, aux catégories syllabique et accentuelle⁹¹. L'interprétation des mesures en des termes phonologiques est aisée : un $\%V$ faible suppose que la durée des intervalles vocaliques par rapport à celle des intervalles consonantiques est relativement faible dans la langue étudiée. En d'autres termes, cela permet d'inférer que la langue en question a des syllabes complexes (i.e. avec des groupes consonantiques de grande taille) et/ou qu'elle possède des voyelles réduites (phonétiquement au moins) sur le plan de la durée⁹². Le ΔC quant à lui traduit le fait que la complexité des syllabes varie dans la langue. Ces deux indices ne constituent au final qu'un moyen commode de mesurer ce que Delattre et Olsen ([DO69]) avaient décrit 30 ans plus tôt⁹³.

Ramus et collègues ([RNM99], 275) concluent que, au vu du lien évident entre les classes rythmiques définies traditionnellement par les linguistes et les classes obtenues à partir de leurs trois indices, on est autorisé à penser que « at least some rhythmic properties of languages can be extracted by phonetic measurements on the signal ».

Parallèlement aux travaux de Ramus et collègues, Grabe et Low, notamment dans ([GL02]), ont proposé d'autres types de mesures regroupées sous l'étiquette « Pairwise Variability Index » (PVI). Cet indice, dans sa forme élémentaire, a été utilisé pour la première fois par le second auteur, qui travaillait sur l'anglais de Singapour ; Francis Nolan est à l'origine de cette mesure ([LGN00], 382). Par exemple, dans [LG95], les auteurs

⁹¹Le japonais est classé aujourd'hui dans une catégorie dont le nom est « mora-timed » ; d'après Grabe et Low ([GL02]), Bloch est le premier linguiste à avoir utilisé cette notion en 1942. À propos du japonais, Bloch écrit ([Blo70], 121) : « The number of syllables in a phrase is therefore not found by counting peaks of sonority or chest pulses, but only by counting the temporally equal fractions contained in it. . . ». Wiik ([Wii91]) mentionne également le « foot timing » pour des langues comme le finnois et l'estonien.

⁹²À condition la variation de durée intrinsèque des segments soit comparable d'une langue à l'autre.

⁹³D'autres aspects liés aux éventuelles causes phonétiques de l'impression de rythme sont abordés dans [FKN91] pour le suédois, le français et l'anglais.

pensent pouvoir expliquer l'effet « staccato », i.e. la tendance syllabique, de l'anglais de Singapour par le fait que les voyelles phonologiquement réduites en anglais britannique le sont moins, phonétiquement parlant, à Singapour, notamment sur le plan de la durée⁹⁴. Il s'ensuit que la différence de durée entre voyelles non réduites et voyelles réduites est moindre qu'en anglais britannique. D'où la formule de base du PVI (5.1) :

$$PVI = \frac{\sum_{i=1}^{N-1} |D_i - D_{i+1}|}{N - 1} \quad (5.1)$$

où N représente le nombre d'intervalles d'un type donné (consonantique ou vocalique) à l'intérieur d'une portion de parole dont la taille reste à définir par l'investigateur (groupe entre pauses, phrase, passage intégral, etc.) et D_i est la durée de l'intervalle numéro i . Reformulé de manière moins formelle, cela revient à dire que le PVI mesure la différence moyenne, en valeur absolue, de durée entre deux intervalles (vocaliques ou consonantiques) consécutifs sur un passage de longueur donnée. Par conséquent, un PVI vocalique élevé traduit une variation importante de la durée des voyelles dans la langue étudiée (ce qui peut être dû au phénomène de réduction vocalique, ou découler du fait que le système étudié possède une durée contrastive phonologique), et un PVI consonantique élevé est symptomatique de structures syllabiques à complexité variable. En d'autres termes, plus les PVI, consonantique et vocalique, sont élevés, plus on s'approche des caractéristiques des langues accentuelles.

Un avantage incontestable du PVI, au moins sur le plan théorique, réside dans le fait que l'information chronologique séquentielle, est prise en compte. L'utilisation de l'écart-type, comme pour ΔV et ΔC , conduit à ignorer cette information, tant et si bien qu'on peut théoriquement objecter, comme le font Grabe et Low ([GL02], 521) que ΔV et ΔC mesurent certes la variation de durée des intervalles vocaliques et consonantiques, mais également les variations de débit.

Dans l'article de Grabe et Low ([GL02]), en plus de l'indice formalisé dans l'Équation 5.1, les auteurs proposent un PVI « normalisé » :

⁹⁴Voir également l'étude de Deterding ([Det01]) sur ces deux variétés d'anglais.

$$nPVI = \frac{\sum_{i=1}^{N-1} |(D_i - D_{i+1}) / ((D_i + D_{i+1}) / 2)|}{N - 1} \quad (5.2)$$

L'Équation 5.2⁹⁵ est justifiée par le fait que, au moins intuitivement, à débit variable au cours de la production des N intervalles, les valeurs du PVI non normalisé devraient varier, toutes choses égales par ailleurs. La normalisation consiste à diviser, à chaque itération, la différence de durée entre les deux segments de la paire par la moyenne de leurs durées. Un autre type de normalisation du débit a été proposé par Dellwo ([Del06]) : le coefficient de variation. Partant du constat que les écart-types proposés par Ramus ([RNM99]), ΔV et ΔC , sont très sensibles aux variations de débit (voir [DW03, DSA⁺04]), et donc corrélés à la durée moyenne des intervalles vocaliques et consonantiques, il paraît justifié de calculer le rapport de l'écart-type sur la moyenne en vue d'effacer au moins en partie le facteur débit. Cette approche est reprise par White et Mattys ([WM07]) ; les auteurs suggèrent que les versions normalisés d'intervalles vocaliques (*npviv*⁹⁶, i.e. le PVI vocalique normalisé, le coefficient de variation des intervalles vocaliques *VarcoV*, et %*V*) permettent une meilleure séparation entre les langues étudiées, alors que, pour les intervalles consonantiques, une telle normalisation aboutit à la perte d'une partie de l'information linguistique. Les auteurs recommandent donc l'utilisation des trois dimensions que sont %*V*, *VarcoV* et *npviv*.

La majeure partie des exemples que nous avons cités dans cette section concernent des études multilingues ; il est donc indispensable de donner explicitement les raisons qui nous ont poussé à estimer que ce type d'études pouvait être pertinent dans un contexte dialectal. Si l'on se réfère à l'étude de Low ([LGN00]), qui utilise avec succès les indices que nous venons de décrire pour distinguer non pas deux langues, mais bien deux variétés d'une même langue (anglais de Singapour *vs* anglais britannique), si éloignées soient-elles géographiquement et perceptivement, répliquer cette expérience à partir d'autres variétés ne semble pas incohérent. En 1982, Wells ([Wel82], 86) notait que le rythme avait une fonction de discrimination évidente entre les accents de l'anglais, mais qu'il restait

⁹⁵L'article original fait apparaître un facteur 100 devant l'Équation 5.2 qui n'a d'autre but que de faciliter la lecture de l'indice ainsi calculé.

⁹⁶Dans ce sigle et le suivants, la lettre à gauche de <pvi> indique s'il est normalisé (<n>) ou brut (<r> pour <raw>), et la lettre à droite précise s'il s'agit d'intervalles consonantiques ou vocaliques.

beaucoup à faire pour qu'il soit décrit de manière satisfaisante :

Regrettably, we are far from being able to give a coherent account of how these [rhythm, stressing, etc.] vary from accent to accent, even though it is clear that they too have an important indexical rôle [sic] helping the hearer to recognize particular accents...

Il poursuit avec quelques exemples ayant trait à des différences de syllabation, de durée, de débit (le parler citadin est plus rapide que le parler rural, [Wel82], 87) et de place de l'accent. Abercrombie ([Abe64], 222) fait allusion aux différences de quantité syllabique dans trois accents de l'anglais. À partir d'un système à 3 quantités différentes (longue, médium et brève), le mot <Peter> :

- possède le schéma long-bref dans le Yorkshire ;
- a la séquence bref-long en « Lowland Scots » ;
- peut être décrit par le schéma médium-médium en RP.

Dans une étude de 1967 ([Bus67]), Bush étudie les durées absolues et les rapports de durée entre consonnes et voyelles en anglais à partir de 3 groupes de sujets : des Américains, des Britanniques et des Indiens pour qui l'anglais est une langue seconde. Si l'on classe ces trois accents en fonction de la durée absolue des syllabes accentuées, on obtient, dans l'ordre décroissant : anglais américain, anglais britannique, anglais indien. L'étude montre ensuite que c'est le rapport entre la durée des consonnes et celle des voyelles qui offre l'indice le plus discriminant entre les 3 variétés d'anglais ; l'auteur explique, par exemple, que le rapport de la durée de [p] (occlusion et explosion) sur la durée de la voyelle qui suit, est de 1,4 pour l'anglais américain, 1,2 pour l'anglais britannique et 0,9 pour l'anglais indien. Nous remarquons que ceci justifie l'utilisation de la quantité %V dans des études inter-dialectales.

La dialectologie traditionnelle ne traite donc pas la question du rythme de façon détaillée. Ceci nous a conduit à suivre le raisonnement suivant : si l'on admet que les indices proposés par [RNM99] et [GL02] capturent, au moins dans une certaine mesure, les différences de rythmes entre les langues, et si ces indices varient en fonction du dialecte, on peut en déduire que cette variation est imputable à des différences de rythme. Il ne reste

plus qu'à déterminer *a priori* si la littérature nous laisse présager des résultats satisfaisants dans les dialectes des Îles Britanniques.

Dans son ouvrage de vulgarisation de la dialectologie sur les Îles Britanniques, Elmes met souvent en avant les particularités rythmiques (et plus généralement, suprasegmentales) des variétés décrites :

the countrymen spoke markedly more slowly than city folk, with words coming in clumps rather than in the city-dweller's tumble of syllables that often run into one another ([Elm05], 61, à propos du sud-est).

The core of this musicality lies in the way stressed syllables are given extra emphasis, combined with many elisions... ([Elm05], 92, au sujet du Pays de Galles).

So what's the key to the Geordie magic? Actually I think a lot of it lies in the rhythm ([Elm05], 252).

Wells ([Wel82], 362-363), quant à lui, rappelle une tendance à la non réduction de certaines voyelles dans le nord :

Northern speech tends to retain strong vowels in certain environments where RP and other accents show weakening. Notable among these are Latin prefixes such as *ad-*, *con-*, *ex-* when pretonic.

Ceci est confirmé pour l'anglais de Newcastle ([WA03], 269-270). Wells mentionne également le fait qu'en anglais du Pays de Galles, en syllabe finale de mot fermée, la réduction vocalique a tendance à être évitée ([Wel82], 387). Il note encore qu'en Irlande du Nord et en Écosse, la quantité, i.e. les différences phonologiques de durée, ont presque totalement disparu ⁹⁷ ([Wel82], 438-439 ; voir également [Jon63], 138-139). Si l'on utilise une mesure de la variation de la durée des voyelles, par exemple le PVI vocalique ou encore le ΔV , on peut donc s'attendre à ce que, toutes choses étant égales par ailleurs (et notamment le débit), les dialectes des régions que nous venons de citer aient des valeurs plus faibles, se rapprochant ainsi (toute proportion gardée) des langues syllabiques.

⁹⁷Exception faite, bien entendu, de certaines variations de durée dues à des facteurs morphologiques, comme c'est le cas dans ce qu'il est convenu d'appeler la *Scottish Vowel Length Rule* [McC77], [STH99], [SHT99] ; ce phénomène est également attesté chez des locuteurs âgés de l'extrême nord de l'Angleterre ([WI00], [WA03], 269).

Tout n'est cependant pas si simple puisque d'autres phénomènes phonétiques ont, sur le rythme, une influence contraire à celle qu'exercent ceux que nous venons d'énumérer. Par exemple, à Newcastle, les syllabes accentuées ont souvent tendance à être plus courtes que les non-accentuées ([WA03], 269-270). Wells ([Wel82], 362-363) note que dans le nord (en excluant les Midlands), les voyelles phonologiquement longues le sont toujours phonétiquement, et que, à titre d'exemple, la voyelle de <sheep> est plus longue dans ces dialectes qu'en RP. Interprété en référence à nos mesures, ceci ferait tendre ces dialectes vers le pôle accentuel, ce qui contredit ce que nous avons conclu plus haut. Pour complexifier encore davantage la situation, Jones relève ([Jon63], 138) une tendance des accents du sud de l'Angleterre à allonger les voyelles brèves, notamment lorsqu'elles sont en fin de phrase et qu'elles portent l'accent. De plus, la dichotomie longues/brèves semble ne pas toujours être observable à partir de mesures de durées sur de la parole spontanée (au moins en anglais américain [PL60], [LA95]).

White et collègues ([WMSG07]) ont très récemment mis au point une tâche de discrimination impliquant des dialectes de l'anglais à partir de stimuli de resynthèse du même type que ceux utilisés par [RM99]. Plus précisément, les auteurs utilisent le *sasasa* monotone, i.e. chaque voyelle est remplacé par [a], chaque consonne, par [s], et le contour intonatif est aplati. Les variétés d'anglais testées sont l'anglais britannique standard (SSE), l'anglais de Welsh Valleys (WV) et l'anglais de Orkney (OR). La tâche de discrimination des dialectes deux à deux montre que les sujets sont capables, mais avec des scores de classification correcte à peine supérieurs à 50 %, de séparer SSE de WV, mais pas WV de OR (la paire SSE/OR n'a pas été testée). Les auteurs soulignent le parallèle entre les scores de classification et les mesures *VarcoV* et *%V*.

5.4 Analyse du rythme à partir de ABI

Nous avons ensuite transposé les méthodes éprouvées en contexte multilingue (Section 5.3) à notre problématique inter-dialectale. Le passage lu du corpus ABI a été utilisé

à ces fins. Tous les dialectes de la base de données (excepté *ilo*, pour les raisons invoquées dans la Section 4.2), ont été pris en compte⁹⁸.

Le rythme a été mesuré sur 3 types de données distincts :

1. un extrait du passage lu de ABI segmenté manuellement en consonnes et voyelles⁹⁹ ;
2. le même extrait que précédemment, segmenté automatiquement en consonnes et voyelles (voir *infra* pour la méthode) ;
3. la totalité du passage lu (Annexe A) segmenté automatiquement en consonnes et voyelles.

Voyons le détail de la méthode de segmentation automatique du signal. Dans un premier temps, l'amplitude de chaque fichier a été normalisée par le biais d'un programme réalisé avec le logiciel Praat. Puis les pauses, les segments vocaliques et les consonnes sont détectés grâce à un algorithme implémenté dans le langage Tcl/Tk (les algorithmes sont décrits dans [PAO00]). Cette segmentation s'appuyant sur les propriétés acoustiques du signal (i.e. l'expert humain n'intervient pas), et particulier sur la détection de ruptures statistiques dans le signal, il s'ensuit que les frontières segmentales résultantes ne correspondent pas exactement à des entités phonologiques, mais plutôt infra-phonémiques. Par exemple, une certaine stabilité spectrale est nécessaire à une portion de signal pour que cette dernière accède au statut de voyelle selon les critères de l'algorithme¹⁰⁰ ; les transitions entre une consonne et une voyelle, caractérisées par des mouvements rapides des formants, et que le phonéticien aura tendance à inclure dans la voyelle, sont souvent considérés comme des consonnes par l'algorithme. De plus, les voyelles très brèves ou non-voisées sont traitées comme des consonnes. Les Figures 5.1 à 5.3 illustrent quelques spécificités de la segmentation automatique en consonnes et voyelles. La Figure 5.1 regroupe un spectrogramme, un oscillogramme et le résultat de la segmentation automatique en

⁹⁸Nous rapportons une expérience complémentaire dans l'Annexe B.

⁹⁹Voici l'extrait concerné : Boats are stronger and more stable, protecting against undue exposure. Tools and instruments are more accurate and more reliable, helping in all weather and conditions. Food and drink are better researched and easier to cook than ever before.

¹⁰⁰Comme le soulignent les auteurs ([PAO00], il s'agit plus d'une détection de noyaux vocaliques que de voyelles à part entière.

consonnes et voyelles pour la portion de signal correspondant à <(strong)er and more (stable)> ([əãmɔː])¹⁰¹ produit par une locutrice de *brm*. La première voyelle, qui correspond à <-er>, est correctement détectée et segmentée. Puis, parmi toutes les stratégies envisageables pour glisser de la voyelle de <-er> à celle de <and> (dont le /r/ de liaison), la locutrice passe en voix craquée, comme l'indiquent les stries espacées et irrégulières sur la partie du spectrogramme qui coïncide avec le premier segment identifié comme consonne. La voyelle phonologique de <and> est donc, probablement faute de pouvoir détecter une fréquence fondamentale, considérée par l'algorithme comme une consonne. Le deuxième C délimite le /n/ de <and>; on constate qu'il n'est suivi par aucun indice acoustique laissant présager la présence d'une occlusive; on peut donc avancer qu'aucun événement acoustique ne correspond au /d/ phonologique de <and>. Le deuxième segment noté V correspond au /m/ dans <more>. Le /m/ est réalisé en une consonne voisée à structure formantique stable clairement identifiable, accompagnée d'aucune vibration aperiodique visible, ce qui rend le son assimilable à une voyelle. L'amplitude moindre de ce segment aurait pu permettre de l'identifier comme consonne; cela n'a pas été le cas. Le dernier C de la Figure 5.1 est causé par la présence d'une barre d'explosion dans le spectrogramme; ceci n'a rien de surprenant car [m] est souvent accompagné d'une faible explosion bilabiale. La Figure 5.2 illustre la sur-segmentation produite par l'algorithme : le mot <vast> comprend trois consonnes phonologiques, et le programme en détecte neuf. Cette représentation n'a en aucun cas biaisé nos calculs puisque ce sont les durées des intervalles vocaliques et consonantiques qui nous occupent, quel que soit le nombre de segments qui les composent. Enfin, la Figure 5.3 représente la séquence <(un)due exposure>. Le bruit de friction identifié comme la première consonne de la figure correspond à la réalisation de /(d)j/ en /(d)ʒ/ (Yod-coalescence). La structure formantique instable qui suit (avant le premier V) correspond à la réalisation de /uː/, et est, à tort, identifiée comme consonne. Le reste de la figure n'est pas commenté car, hormis une sur-segmentation évidente, les étiquettes attribuées par l'algorithme aux bribes de signal restantes sont phonétiquement

¹⁰¹Les éléments entre parenthèses donnent le contexte; seuls les éléments dépourvus de parenthèses sont représentés sur les Figure 5.1 à 5.3.

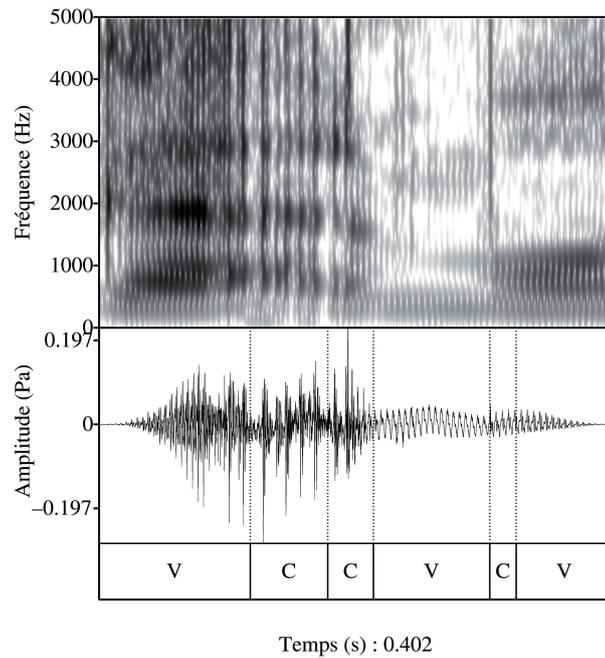


FIG. 5.1 – Segmentation automatique du signal en consonnes et voyelles : <(strong)er and more (stable)>.

justes. En résumé, la segmentation automatique a induit les biais suivants :

- une sous-estimation des durées vocaliques ;
- une mauvaise détection des voyelles très brèves et/ou particulièrement instables sur le plan spectral ;
- une fausse détection des consonnes à formants.

Une fois les frontières obtenues, elles sont importées sous Praat, puis les segments adjacents de même nature (voyelles ou consonnes) sont regroupés en une seule et même entité : un intervalle vocalique ou consonantique.

Détaillons à présent les mesures employées dans notre analyse du rythme. Rouas *et al.* ([RFP03]) avaient utilisé avec succès la durée brute des intervalles vocaliques et intervocaliques (i.e. consonantiques) pour l'identification automatique des langues ; nous avons fait de même dans Ferragne et Pellegrino ([FP04c]). Un test de Kruskal-Wallis¹⁰² ayant pour but de comparer les durées vocaliques et consonantiques moyennes entre les dialectes

¹⁰²L'équivalent d'une ANOVA simple, mais dans le domaine des statistiques dites « non-paramétriques », qui ont l'avantage d'être beaucoup moins contraignantes concernant leurs postulats de départ, notamment la forme des distributions des variables et la taille des échantillons.

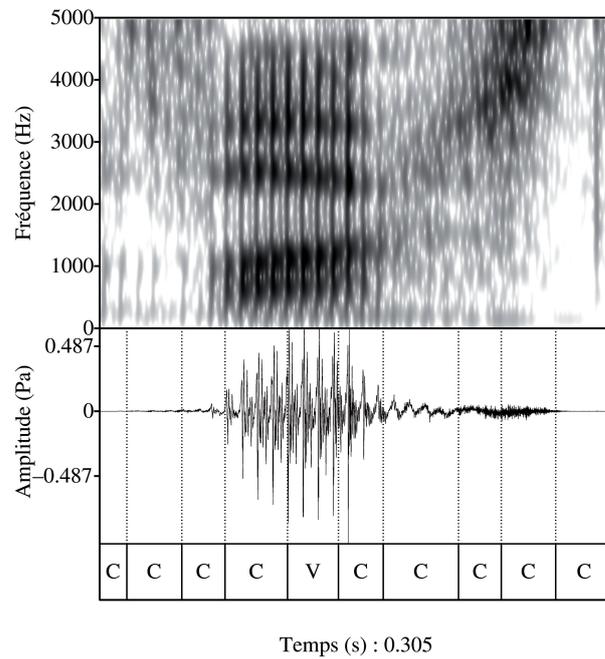


FIG. 5.2 – Segmentation automatique du signal en consonnes et voyelles : <vast>.

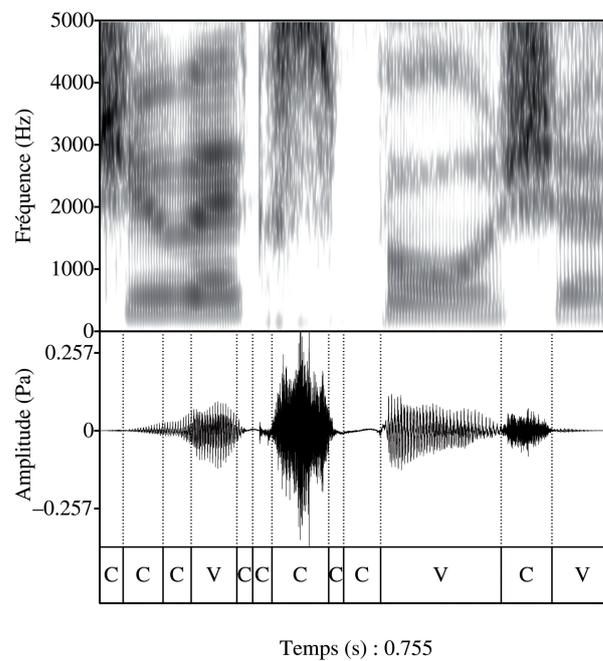


FIG. 5.3 – Segmentation automatique du signal en consonnes et voyelles : <(un)due exposure>.

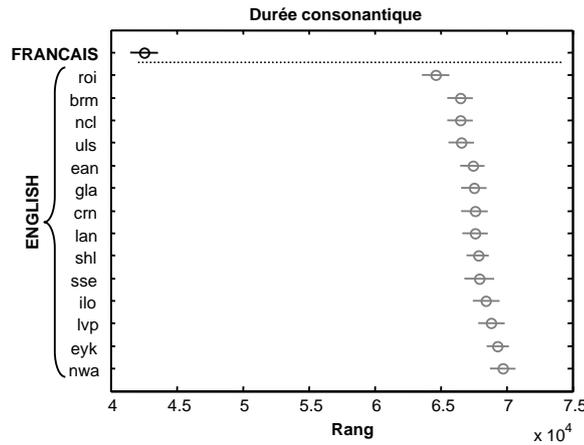


FIG. 5.4 – Durée consonantique moyenne et écart-type par dialecte.

s'avère significatif au niveau $p < 0,001$. La durée vocalique moyenne la plus faible revient à *sse* et l'on passe ainsi, presque insensiblement, de dialecte en dialecte, à des valeurs de plus en plus élevées jusqu'au groupe, constitué de *lan*, *roi* et *lvp*, qui présente les durées les plus élevées. La durée des voyelles est particulièrement sensible aux variations de débit, avec un coefficient de corrélation de 0,876 entre les deux dimensions. En ce qui concerne les intervalles consonantiques, *roi* a les durées les plus faibles et *nwa* les plus élevées. Pour ce dernier, on pourrait citer parmi les facteurs potentiels le fait que, comme nous l'avons noté dans le paragraphe traitant du Pays de Galles dans le Chapitre 3, beaucoup de consonnes peuvent être allongées dans cette variété en raison de l'influence du gallois. Mais ici encore, la corrélation avec le débit est très importante ($r = 0,890$). Ces résultats sont résumés dans les Figures 5.4 et 5.5¹⁰³.

Pour la partie segmentée manuellement, les paramètres suivants ont été calculés : $\%V$, ΔV , $VarcoV$, ΔC , $VarcoC$, les PVI bruts et normalisés des intervalles vocaliques et consonantiques, et le débit (voyelles par seconde, pauses exclues); soit, au total, 10 paramètres. Les PVI brut et normalisé (voir les Équations 5.1 et 5.2) ont été calculés pour les intervalles vocaliques et consonantiques. À noter que lorsque deux intervalles de même type étaient séparés par une pause (souvent due à une fin de phrase), ils étaient exclus du

¹⁰³Ces figures sont adaptées de [FP04c]; un échantillon de français avait été inclus ainsi que le dialecte *ilo*.

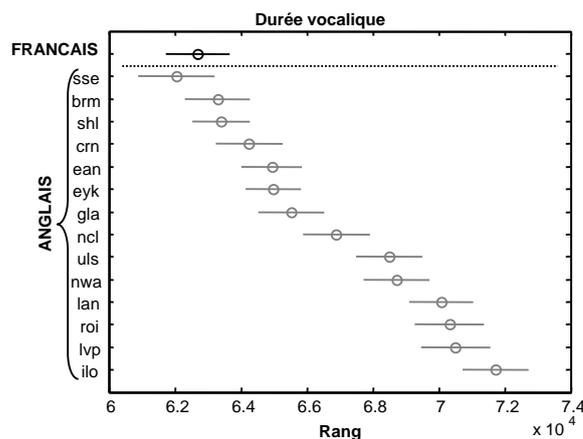


FIG. 5.5 – Durée vocalique moyenne et écart-type par dialecte.

calcul pour éviter le biais induit par l’allongement typique de ce genre de contexte¹⁰⁴.

La première étape consiste à étudier les corrélations entre les 10 variables incluses dans l’analyse. Une des faiblesses de la plupart des études multilingues ou multi-dialectales du rythme provient du fait que les auteurs s’astreignent souvent à représenter leurs objets dans un espace à deux dimensions seulement. Or, si cela peut sembler cohérent en vertu du fait que ce sont deux types d’événements phonétiques qui sont mesurés (voyelles et consonnes), rien n’indique *a priori* que deux dimensions soient suffisantes (nous verrons cela plus loin avec la classification automatique), et rien ne permet non plus d’affirmer *a priori* que ces deux dimensions ne soient pas redondantes. L’inclusion du débit permet également d’apprécier l’effet de la normalisation dans les paramètres de type nPVI, *VarcoV* et *VarcoC*. Les coefficients de corrélation de Pearson entre les 10 variables prises deux à deux sont représentés dans la Table 5.1. L’astérisque indique que la corrélation est significative au moins au niveau $p < 0,01$.

La Figure 5.6 fait écho à la Table 5.1 : elle représente les nuages de dispersion des 263 moyennes individuelles sur les 10 paramètres étudiés pris deux à deux. Toutes les variables sont corrélées au débit excepté *VarcoV* et *npviv* ; ce qui prouve que le but de ces deux variables – i.e. minimiser, voire faire disparaître, l’influence du débit sur ΔV et

¹⁰⁴On sait par ailleurs que la durée des voyelles finales de phrase varie en fonction du débit, mais pas de façon linéaire ([BGB95]).

TAB. 5.1 – Corrélations entre les 10 variables prises deux à deux, segmentation manuelle (r de Pearson).

	%V	ΔV	VarcoV	ΔC	VarcoC	rpviv	rpvic	npviv	npvic	débit
%V										
ΔV	0,492 *									
VarcoV	0,008	0,550 *								
ΔC	-0,120	0,539 *	0,081							
VarcoC	-0,062	0,256 *	0,189 *	0,709 *						
rpviv	0,486 *	0,932 *	0,447 *	0,516 *	0,212 *					
rpvic	-0,101	0,519 *	0,045	0,923 *	0,562 *	0,497 *				
npviv	0,003	0,333 *	0,659 *	0,033	0,108	0,479 *	-0,008			
npvic	-0,010	0,221 *	0,133	0,465 *	0,569 *	0,193 *	0,648 *	0,083		
débit	-0,187 *	-0,723 *	0,021	-0,784 *	-0,261 *	-0,716 *	-0,796 *	-0,261 *		

rpviv, respectivement – est atteint. En revanche, leurs pendants consonantiques (*VarcoC* et *npvic*) engendrent des valeurs corrélées au débit ; néanmoins, la corrélation est assez faible ($-0,261$ dans les deux cas). On peut également remarquer que les variables utilisées par Ramus et collègues ([RNM99]) et celles proposées par Grabe et Low ([GL02]) (e.g. ΔV et *rpviv*) sont extrêmement corrélées. On constate encore que les deux dimensions retenues dans les études que nous venons de citer – ΔC et $\%V$ pour [RNM99], et *rpvic* et *npviv* pour [GL02] – présentent, en plus de leur qualité de bons paramètres discriminants entre classes de rythmes mentionnée dans les études citées, l’avantage de ne pas être corrélées dans nos données, d’où une représentation au moins parcimonieuse si elle n’est pas exhaustive. Les mesures retenues dans l’étude de White et Mattys ([WM07]) – *VarcoV*, *npviv* et $\%V$ – ne sont pas, elles non plus, corrélées entre elles.

Nous nous devons d’insister sur le fait que les corrélations de la Table 5.1 ne sont valables que pour nos données ; en effet, il n’est pas à exclure qu’en dehors de l’empan des valeurs prises par ces variables dans les dialectes de l’anglais, les corrélations changent, disparaissent, ou s’inversent ; il convient donc d’éviter toute extrapolation. Par exemple, dans l’étude de Ramus et collègues ([RNM99], 273), on note l’évidente corrélation négative entre ΔC et $\%V$ dans ce contexte multilingue. En outre, le coefficient de Pearson ne mesure que le caractère linéaire d’une liaison entre deux variables.

Nous allons à présent examiner la dispersion des dialectes dans les trois espaces bi-dimensionnels employés dans les études de référence de Ramus et collègues ([RNM99]), Grabe et Low ([GL02]), et White et Mattys ([WM07]). La Figure 5.7 représente les 263 locuteurs et locutrices du corpus ABI dans le plan $\%V/\Delta C$: chaque dialecte est représenté par sa valeur moyenne et les barres d’erreurs s’étendent sur un écart-type. On constate d’emblée la très grande variation intra-dialectale, qui laisse présager que ces paramètres ne sont vraisemblablement pas robustes pour une tâche de classification. Les Figures 5.8 et 5.9 nous amènent au même constat. Si, en revanche, on examine les dialectes dans une optique purement descriptive, certains faits intéressants émergent. Puisque nous n’avons aucune hypothèse linguistique quant aux comportements des intervalles consonantiques,

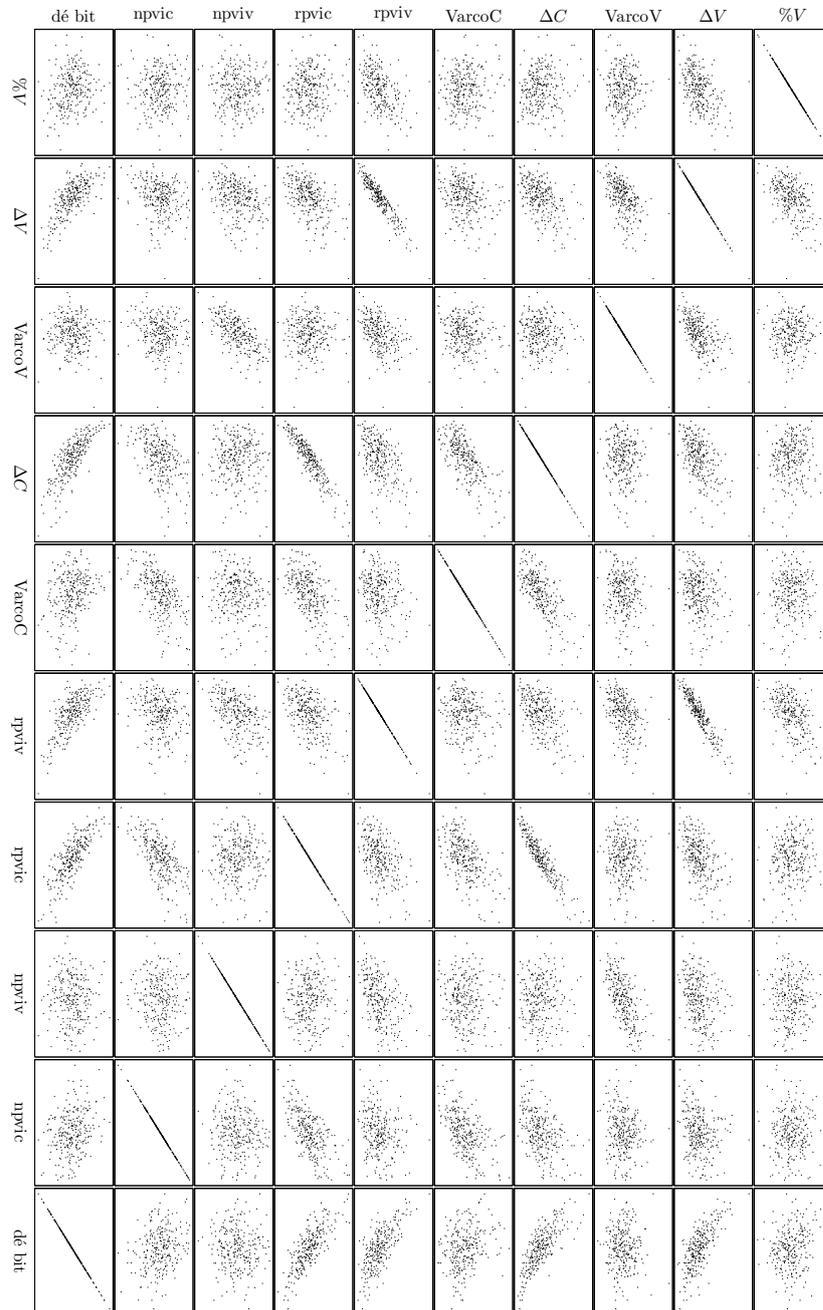


FIG. 5.6 – Nuages de dispersion des points individuels sur 10 dimensions prises deux à deux.

nous nous contenterons de commenter les mesures des intervalles vocaliques. La quantité %V (Figure 5.7) est difficilement interprétable dans notre contexte inter-dialectal : si elle permet, dans une étude multilingue, de distinguer les langues à structures syllabiques complexes des autres, il n'existe à notre connaissance pas de raison phonologique qui plaiderait en faveur d'intervalles consonantiques plus long dans l'un ou l'autre de nos dialectes, hormis la présence ou l'absence de rhoticité¹⁰⁵, le phénomène d'insertion d'un /t/ dans les mots du type de <chance>, qui en vient à rimer avec <chants> ([Wel99]), ou encore l'allongement typique d'un accent marqué de l'anglais du Pays de Galles des consonnes en position intervocalique ([Wel82], 387). L'indice %V peut certes être également affecté par la variation inter-dialectale sur le plan de la réduction vocalique, mais sa corrélation avec le débit nous conduit à nous tourner plutôt vers *npviv* et *VarcoV*. On note dans la Figure 5.8 que *uls* et *crn* ont des valeurs moyennes peu élevées sur l'axe *npviv*. En ce qui concerne *uls*, l'explication vient très probablement du fait qu'il ne semble pas exister de différence de quantité phonologique dans le système de ce dialecte. Pour ce qui est de *crn*, aucune explication phonologique *a priori* ne semble justifier cette valeur basse ; mais étant donné que le *npviv* mesure la différence de durée moyenne entre deux voyelles successives, on peut envisager :

- soit une absence de durée pertinente sur le plan phonologique, ce qui semble peu probable car non signalé dans la littérature ;
- soit une matérialisation phonétique de ce contraste moins marquée que dans d'autres dialectes ;
- soit, enfin, une résistance à la réduction vocalique.

À l'inverse, on remarque que *sse*, *brm* et *nwa* ont des valeurs relativement élevées, ce qui peut s'expliquer par la contrepartie d'une (ou de plusieurs) des trois propositions de l'énumération précédente. La dimension *VarcoV*, représentée dans la Figure 5.9, corrobore ce qui vient d'être dit.

Afin de tester le potentiel discriminant de ces 3 représentations bidimensionnelles, nous

¹⁰⁵Les valeurs équivalentes de %V pour *roi* (rhotique) et *eyk* (non rhotique), visibles dans la Figure 5.7, prouvent que la rhoticité seule ne permet pas d'expliquer la variation des durées consonantiques.

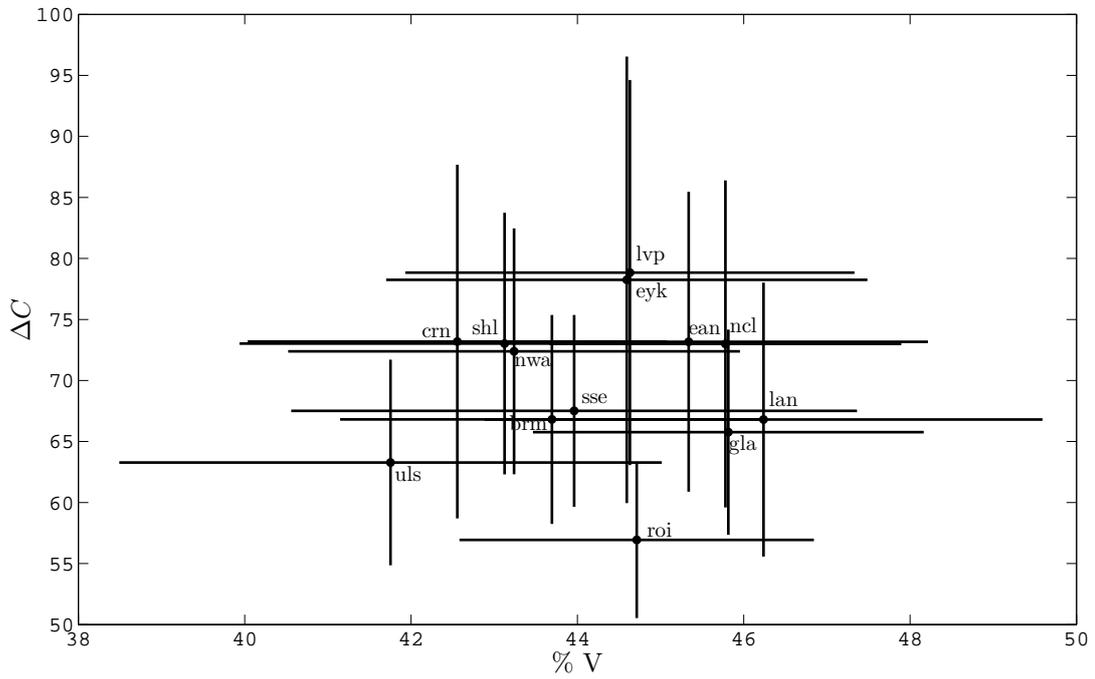


FIG. 5.7 – Dispersion des dialectes dans le plan $\%V/\Delta C$.

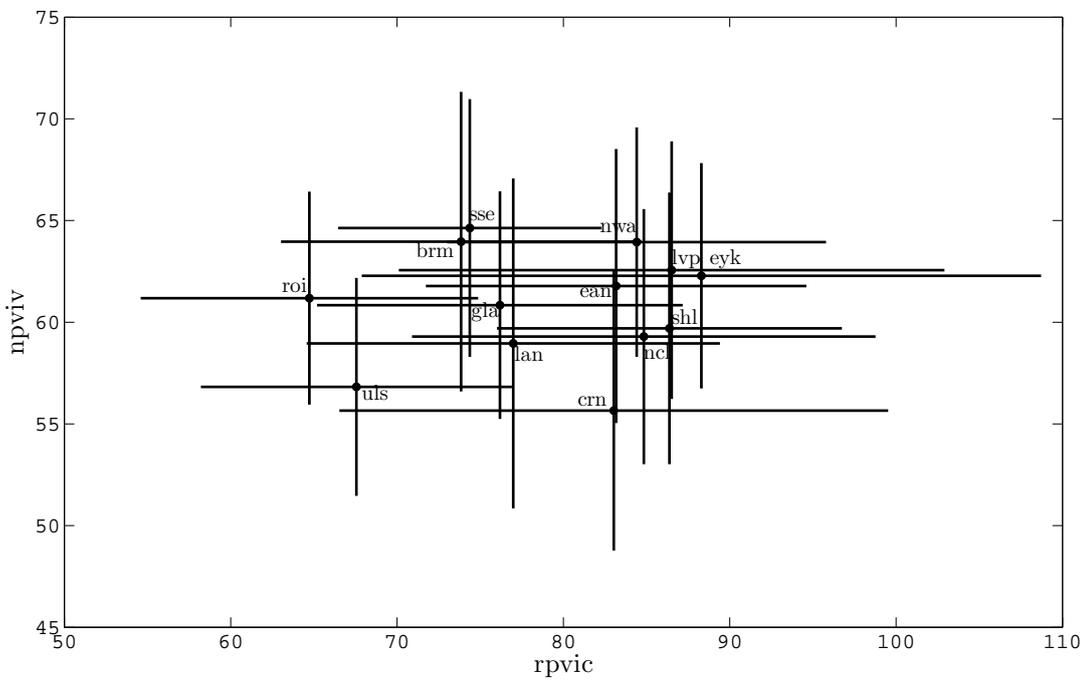


FIG. 5.8 – Dispersion des dialectes dans le plan $rpvic/npviv$.

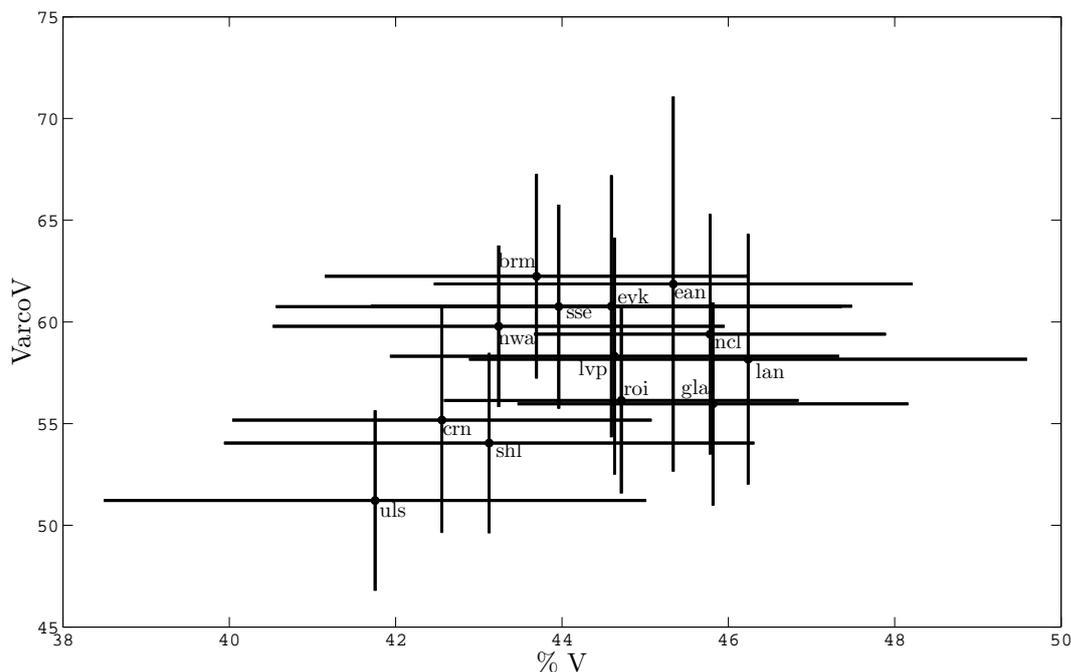


FIG. 5.9 – Dispersion des dialectes dans le plan $\%V/VarcoV$.

avons eu recours à une analyse discriminante¹⁰⁶. Les taux de classification correcte moyens pour chaque espace bidimensionnel sont les suivants :

- $\%V/\Delta C$: 16,73 % ;
- $rpvic/npviv$: 17,49 % ;
- $\%V/VarcoV$: 15,21 %.

Un test binomial montre que ces trois taux de classification sont supérieurs au hasard ($p < 0,0001$). Si l'on inclut tous les paramètres simultanément (à l'exception du débit), le taux de classification correcte passe à 30,80 %.

Le passage segmenté manuellement l'a ensuite été automatiquement en consonnes et voyelles, et les mêmes paramètres ont été calculés sur cette nouvelle segmentation. La Figure 5.10 représente les nuages de dispersion des 261¹⁰⁷ individus dans 100 espaces bi-

¹⁰⁶La fonction `classify` du logiciel Matlab a été utilisée. Le principe de validation est basé sur la méthode du « leave-one-out » : les 13 classes sont apprises sur tous les locuteurs sauf celui en cours de test, puis une classe est attribuée à ce dernier.

¹⁰⁷Deux individus ont dû être écartés en raison de problèmes avec la segmentation automatique.

dimensionnels ayant pour abscisses les mesures de rythme calculées via la segmentation automatique, et pour ordonnées, ces mêmes mesures calculées à partir de la segmentation manuelle. L'intérêt principal de cette représentation réside dans l'examen de la diagonale de cette matrice de graphes, qui permet de détecter les éventuelles corrélations d'un paramètre obtenu après segmentation manuelle avec son pendant automatique. Le calcul des coefficients de corrélation révèle, pour la diagonale, une corrélation significative au niveau $p < 0,01$ pour ΔC , $VarcoV$, $rpvic$ et le débit. Pour ces quatre variables, les coefficients eux-mêmes sont faibles : 0,408, 0,190, 0,397 et 0,560, respectivement. Ceci démontre qu'en règle générale, notre segmentation manuelle et la segmentation automatique ne mesurent pas exactement les mêmes propriétés ; la première est phonétique ; l'autre, strictement acoustique. Il serait faux de penser *a priori* que l'un des deux types de segmentation présente un avantage sur l'autre. En conséquence, les paramètres issus de la segmentation automatique, malgré leur faible (ou absence de) corrélation avec ceux obtenus après segmentation manuelle, seront également pris en compte.

La Figure 5.11 représente les nuages de dispersions des 261 moyennes individuelles sur les 10 paramètres étudiés, issus de la segmentation automatique, pris deux à deux. La Table 5.2 récapitule les coefficients de corrélation, et comporte un astérisque à chaque fois que la corrélation est significative au niveau $p < 0,01$. On remarque d'emblée qu'ici, contrairement à la Table 5.1, tous les paramètres sont corrélés au débit. Les indices normalisés (type $npvi$ et $Varco$) présentent une corrélation avec le débit inférieure à celle des indices non normalisés, mais le gain (en termes d'absence de covariation avec le débit) est bien inférieur à celui que nous avons pu observer sur la segmentation manuelle (Table 5.1). Si l'on se tourne à présent vers les trois espaces bidimensionnels de référence, on constate que les deux dimensions proposées par Ramus et collègues ([RNM99]), $\%V$ et ΔC sont négativement corrélées. Ceci revient à dire que la représentation bidimensionnelle dans cet espace n'est plus parcimonieuse comme c'était le cas avec la segmentation manuelle. Les espaces bidimensionnels ne sont pas représentés ici : ils comportent, à l'instar des Figures 5.7, 5.8 et 5.9, une très importante variation intra-classe associée à une faible

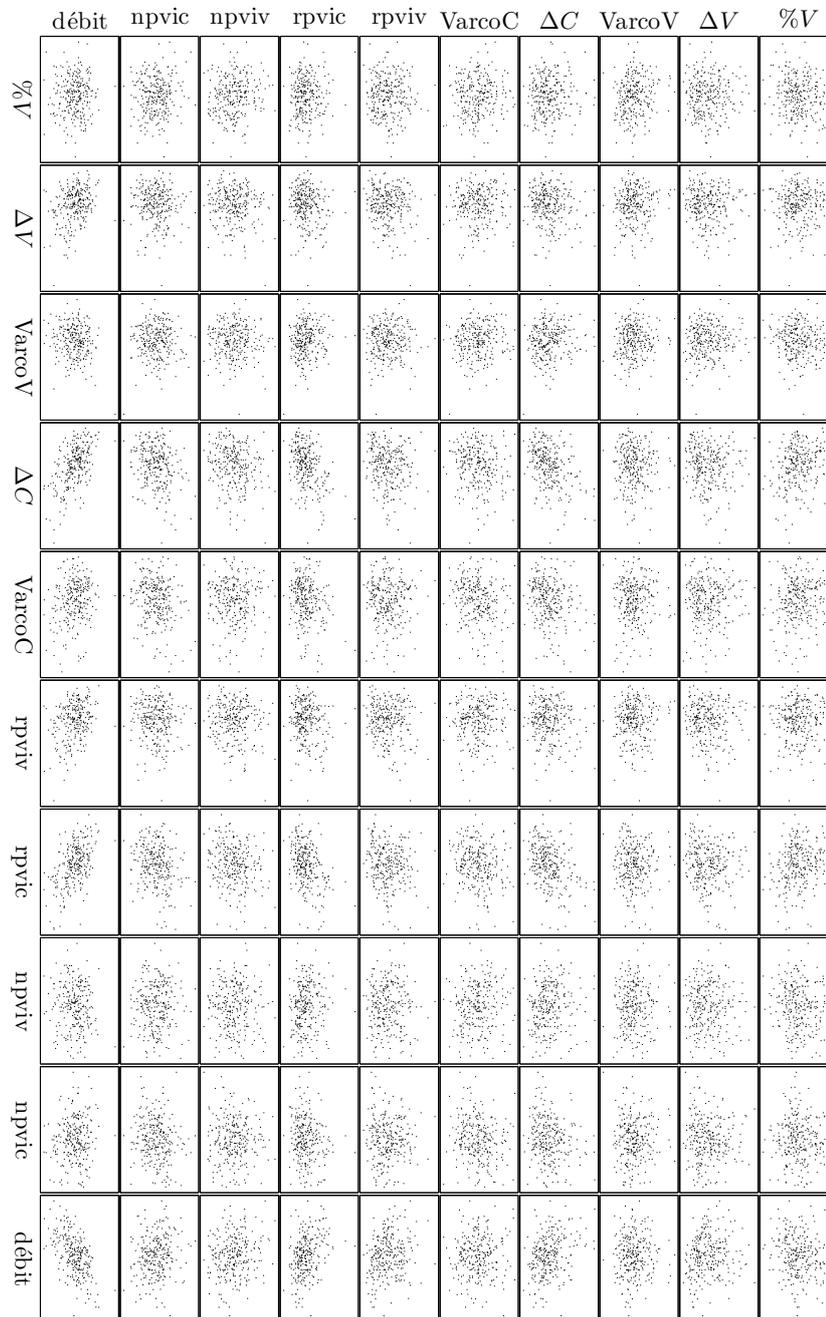


FIG. 5.10 – Nuages de dispersion des points individuels dans les espaces des dix paramètres manuels (longueur de la page) et automatiques (largeur de la page).

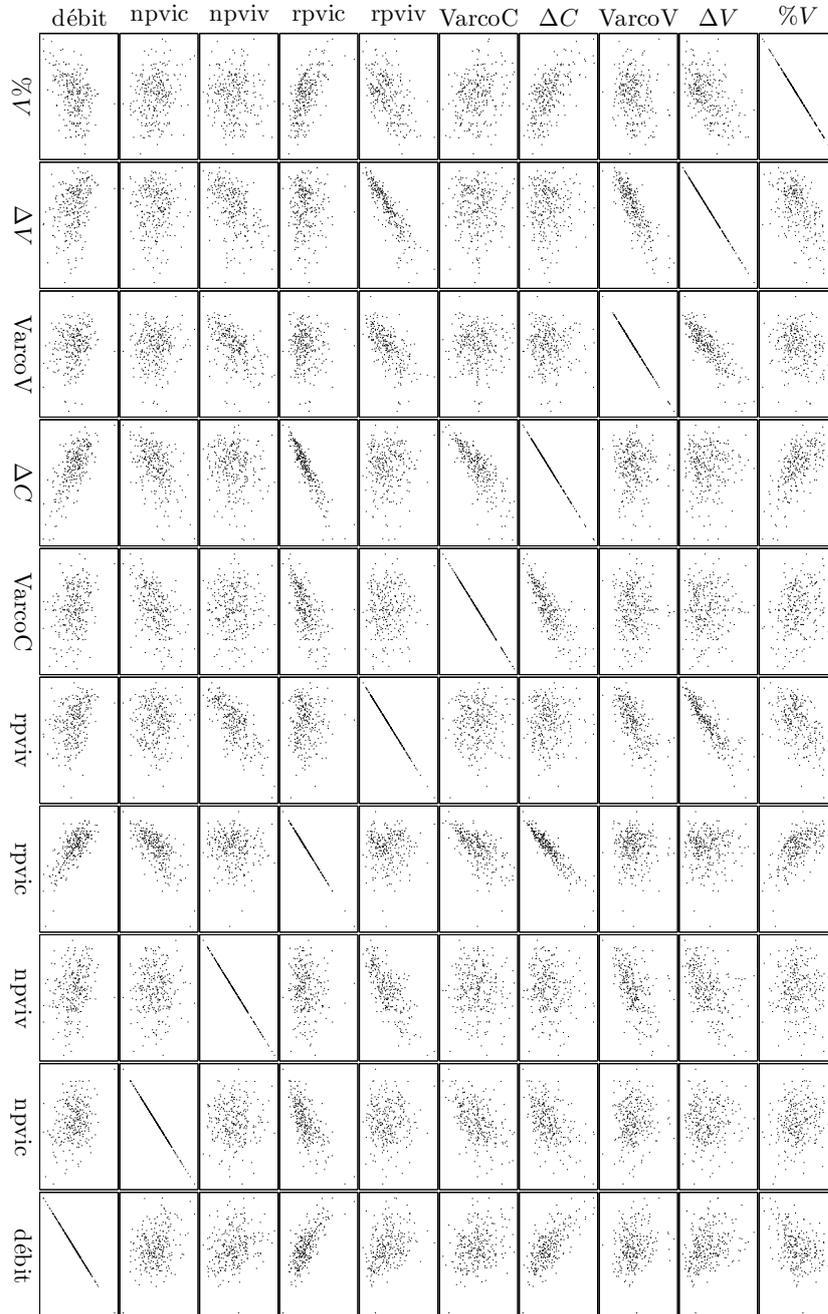


FIG. 5.11 – Nuages de dispersion des points individuels sur 10 dimensions prises deux à deux, à partir de la segmentation automatique.

TAB. 5.2 – Corrélations entre les 10 variables prises deux à deux, segmentation automatique (r de Pearson).

	%V	ΔV	VarcoV	ΔC	VarcoC	rpviv	rpvic	npviv	npvic	débit
%V		0,507 *	0,147	-0,611 *	-0,254 *	0,506 *	-0,605 *	0,040	-0,191 *	0,399 *
ΔV	0,507 *		0,833 *	-0,030	-0,028	0,879 *	-0,014	0,470 *	-0,041	-0,319 *
VarcoV	0,147	0,833 *		0,052	0,015	0,645 *	0,037	0,522 *	-0,061	-0,196 *
ΔC	-0,611 *	-0,030	0,052		0,792 *	-0,031	0,917 *	0,126	0,491 *	-0,693 *
VarcoC	-0,254 *	-0,028	0,015	0,792 *		-0,040	0,616 *	0,047	0,488 *	-0,266 *
rpviv	0,506 *	0,879 *	0,645 *	-0,031	-0,040		-0,023	0,725 *	-0,030	-0,352 *
rpvic	-0,605 *	-0,014	0,037	0,917 *	0,616 *	-0,023		0,088	0,669 *	-0,734 *
npviv	0,040	0,470 *	0,522 *	0,126	0,047	0,725 *	0,088		-0,009	-0,260 *
npvic	-0,191 *	-0,041	-0,061	0,491 *	0,488 *	-0,030	0,669 *	-0,009		-0,275 *
débit	0,399 *	-0,319 *	-0,196 *	-0,693 *	-0,266 *	-0,352 *	-0,734 *	-0,260 *	-0,275 *	

distance inter-classe.

Les taux de classification obtenus après analyse discriminante sont les suivants :

- $\%V/\Delta C$: 10,73 % ($p < 0,05$) ;
- $rpvic/npviv$: 11,88 % ($p < 0,05$) ;
- $\%V/VarcoV$: 14,56 % ($p < 0,0001$).

Les probabilités (test binomial) que ces taux soient dus au hasard sont très faibles. Cependant, pour la segmentation automatique comme pour la segmentation manuelle, les taux de classifications moyens dans une tâche à 13 classes sont nettement insuffisants pour être d'une quelconque utilité. En d'autres termes, les paramètres rythmiques calculés jusqu'ici à partir de la durée ne permettent pas de distinguer les 13 dialectes du corpus ABI.

Nous avons démontré, à ce stade de notre analyse, que les indices normalisés ($npvi$, $VarcoV$ et $VarcoC$) ont une corrélation plus faible (voire inexistante) avec le débit. Les mesures du PVI semblent mieux fondées puisqu'elles prennent en compte l'ordre chronologique des événements phonétiques et sont ainsi mieux adaptées aux variations locales de débit. Les mesures d'écart-type sont, à l'inverse, extrêmement sensibles à ces variations parce qu'elles ignorent précisément l'ordre des segments. Supposons que dans une langue syllabique, où, à débit constant, les voyelles successives ont tendance à posséder une durée équivalente, on dispose d'une phrase de 6 voyelles dont les trois premières sont produites à un débit extrêmement rapide, et les 3 dernières, à un débit extrêmement lent. Si l'on calcule un indice tel que le ΔV sur toute la phrase, la valeur obtenue sera élevée, et indiquera donc, à tort, qu'on est en présence d'une langue accentuelle. Pour éviter ce travers, il aurait fallu détecter le changement de débit, et calculer ΔV d'abord sur la partie rapide, puis sur la partie lente.

Après avoir pris note de l'avantage des mesures de type PVI, l'intégralité du passage lu du corpus ABI a été segmentée automatiquement, et les PVI consonantiques et vocaliques ont été calculés. Une analyse discriminante – avec pour paramètres $rpviv$, $npviv$, $rpvic$ et $npvic$ – atteint un taux moyen de 12,93 % de classification correcte (test binomial :

$p = 0,0021$).

Il est intéressant de noter que, à notre connaissance, toutes les études s'inscrivant dans la lignée de celles de Ramus et collègues ([RNM99]) et de Grabe et Low ([GL02]) se contentent du paramètre physique de la durée. Or, intuitivement, la notion de rythme n'est pas très éloignée de celle d'accent de mot et d'accent de phrase (voir, par exemple, Dauer [Dau83]). En effet, toutes les études qui utilisent le concept de pied, ou encore celui d'intervalle entre accents, impliquent de fait que l'accent est un aspect primordial de l'impression de rythme en anglais. Et puisqu'il est avéré que l'accent de mot en anglais et la proéminence se réalisent non seulement à travers la durée, mais également l'intensité ([Fry55] pour l'accent de mot et [KGCR05] pour la proéminence), mesurer l'intensité sous la forme d'un PVI ne semble pas tout à fait incohérent pour évaluer la pertinence de ces informations (voir ce que suggèrent Ramus et collègues [RDM03]¹⁰⁸). Les PVI vocaliques et consonantiques portant sur l'intensité ont été calculés à partir de la segmentation automatique de l'ensemble du passage lu de ABI. Le calcul est identique à celui décrit dans les Équations 5.1 et 5.2 si l'on remplace la durée de l'intervalle par l'intensité cet intervalle. L'intensité moyenne en dB est mesurée pour chaque intervalle avec le logiciel Praat. Les analyses discriminantes donnent les taux de classification correcte moyens suivants :

- *npviv-I* et *rpvic-I* : 22,05 % ;
- *npvic-I*, *npviv-I*, *rpvic-I* et *rpviv-I* : 33,84 % ;
- PVI d'intensité et de durée confondus : 30,80 %.

Un test binomial montre que ces taux de classification sont supérieurs au hasard ($p < 10^{-12}$). On remarque que lorsque les PVI d'intensité seuls sont inclus dans l'analyse, le taux de classification correcte est plus élevé que lorsque les paramètres de durée seuls sont employés. Il nous paraît donc justifié d'inclure cette mesure de la différence d'intensité moyenne entre deux intervalles vocaliques dans les études sur le rythme de l'anglais, et il

¹⁰⁸[RDM03], 341 : « languages differ in the way they use duration and intensity to signal phonological properties such as stress or quantity. It can therefore not be excluded that a similar quantitative, cross-linguistic study of intensity variations might provide yet another dimension for the study of rhythm classes. »

TAB. 5.3 – Matrice de confusion issue de la classification des dialectes à partir des paramètres d’intensité.

	brm	crn	ean	eyk	gla	lan	lvp	ncl	nwa	roi	shl	sse	uls
brm	6	2	2	-	-	2	-	-	1	2	-	4	1
crn	-	8	2	5	-	-	-	2	-	-	-	3	-
ean	1	2	11	3	-	-	-	-	-	-	1	1	-
eyk	-	1	7	6	-	-	-	-	2	-	9	-	-
gla	-	1	-	-	6	2	5	1	4	-	-	-	1
lan	1	1	1	-	3	-	3	2	4	4	1	1	-
lvp	-	-	-	-	3	-	11	-	-	3	-	1	2
ncl	3	3	-	-	2	-	-	2	4	2	-	2	1
nwa	1	5	1	1	4	1	-	1	3	-	-	1	3
roi	1	2	1	-	1	1	1	-	-	1	-	1	2
shl	-	1	-	6	-	-	-	-	-	-	15	-	-
sse	2	-	3	1	-	-	-	2	1	2	-	3	2
uls	1	-	-	-	3	-	3	3	-	-	-	2	8

serait également très intéressant de la tester dans le cadre de la classification automatique des langues. Il restera néanmoins à déterminer si d’autres mesures que l’intensité moyenne d’un intervalle (e.g. l’intensité maximale) ne seraient pas mieux adaptées.

La Table 5.3 donne la matrice de confusion qui rend compte des résultats de l’analyse discriminante à partir des quatre paramètres d’intensité. Les taux d’identification varient d’un dialecte à l’autre ; trois dialectes ont des taux supérieurs à 50 % : *shl* (15/22), *ean* (11/19) et *lvp* (11/20). S’il n’est pas aisé d’entrevoir les raisons pour lesquelles la classification fonctionnerait mieux pour ces dialectes, on peut néanmoins affirmer que leurs caractéristiques rythmiques (telles qu’elles sont mesurées à travers l’intensité) sont nettement distinctes des autres. À l’inverse, *lan*, avec 0% de classification correcte semble ne pas constituer un tout suffisamment cohérent et distinct des autres dialectes dans la dimension de l’intensité.

5.5 Note sur le débit de parole

L'expérience de tous les jours nous apprend que les variations de débit de la parole peuvent engendrer des variations sur le plan de l'intelligibilité. Il est en outre évident que le débit est utilisé à des fins communicatives, par exemple pour insister. Le fait de prendre en compte la variabilité du débit dans les systèmes de reconnaissance de la parole s'est avéré une avancée décisive ([WFS01]), et un certain nombre de publications ont pour unique objet la détection automatique du débit (par exemple, [VM96], [PR98], [PFR00], [RFP04], [WS05]). Son étude dans le domaine de la synthèse de la parole a également contribué au naturel de la parole générée artificiellement ([Zel98]).

Le débit varie sans cesse, et cette variation influence le rythme à la fois en termes de production et de perception. Koreman ([Kor06]) note que la parole à débit rapide est divisée en un plus grand nombre d'unités prosodiques qu'à débit normal ou lent, et introduit une dimension subjective : « both the intended and realized rate affect the perceived speech rate... » ([Kor06], 592). De la même manière, nous avons démontré ([DFP06]) que des sujets francophones à qui il est demandé d'estimer le débit de parole à partir de stimuli en allemand, anglais et français, tendent à retrouver le débit intentionnel¹⁰⁹, quel que soit le débit objectif.

Contrairement à ce qu'on pourrait croire de prime abord, parler vite ou lentement ne signifie pas qu'on multiplie par un facteur identique la durée de tous les types de segments produits à débit moyen. En effet, Fourakis ([Fou91]) a montré que le rapport de la durée des syllabes non accentuées sur la durée des syllabes accentuées (l'auteur entend « accent de phrase ») est plus haut à débit rapide qu'à débit lent ; autrement dit, la différence de durée entre voyelle accentuée et voyelle non accentuée est proportionnellement moins marquée à débit rapide¹¹⁰. Janse *et al.* ([JNQ03]) observent qu'à tempo élevé, la durée des voyelles est plus réduite que celle des consonnes¹¹¹, ce qui est confirmé dans la

¹⁰⁹Il avait été demandé aux sujets de lire un passage à vitesse moyenne, puis rapide et très rapide, et, enfin, lente et très lente.

¹¹⁰Le même type d'effet non-linéaire est constaté pour les voyelles tendues *vs* relâchées en anglais américain ([Gop90]).

¹¹¹Incidentement, Janse ([Jan04]) constate que l'augmentation artificielle du débit en utilisant la com-

Table 5.1 par la corrélation négative entre le débit et % V . La compression différentielle des segments de la parole affecte de façon évidente les indices que nous calculons. Si, par exemple, comme nous venons de le noter, le rapport de durée entre voyelle accentuée et voyelle non accentuée tend vers 1 à débit rapide (par rapport à un débit moyen), il s'ensuit que le calcul du PVI vocalique donnera des valeurs plus faibles et fera s'éloigner l'extrait de parole analysé du pôle accentuel pour tendre vers le pôle syllabique. Ce constat est illustré dans la Table 5.1 : ΔV et $rpviv$ sont négativement corrélés au débit, ce qui signifie que plus le débit augmente, plus leur valeur est faible ; autrement dit, plus le débit est rapide, plus la valeur des indices ΔV et $rpviv$ s'éloigne des valeurs caractéristiques du pôle accentuel. Si nos mesures de rythme sont influencées par le débit, il convient donc d'écarter ce facteur. À notre avis, contrôler le débit *a posteriori* comporte une double difficulté. Premièrement, le fait qu'il puisse varier considérablement, par exemple pour une seule et même phrase, pousse à se poser la question de la taille optimale de la fenêtre d'analyse¹¹². Ensuite, le traitement différencié des segments selon leur type, ce qu'on pourrait nommer la compression ou l'expansion non-linéaire de la parole en fonction du débit, constitue une deuxième difficulté. Il s'ensuit que les mesures utilisées jusqu'ici, y compris celles censées écarter l'impact du débit, demeurent imparfaites.

5.6 Synthèse des résultats et discussion

Si la plupart des résultats des analyses discriminantes sont très vraisemblablement meilleurs qu'un classement aléatoire, ils ne permettent en aucun cas une discrimination fiable entre les 13 dialectes du corpus ABI. Il est possible d'invoquer plusieurs raisons justifiant ces faibles performances. Premièrement, si les indices utilisés conviennent aux

pression linéaire (méthode avec laquelle la durée de tous les segments est réduite par un facteur identique) rend la parole plus intelligible que la parole naturelle au même tempo. Compte tenu du fait qu'en parole naturelle le facteur de réduction de durée varie en fonction du type de segment (voyelle *vs* consonne, accentué *vs* non accentué, etc.), les auteurs concluent qu'en situation naturelle, ce sont les contraintes physiologiques du système phonatoire qui prévalent sur celles des systèmes auditif et perceptif.

¹¹²C'est-à-dire : quelle étendue de parole (groupe intonatif, phrase, groupe entre pauses, etc.) permet de neutraliser l'impact du débit ?

tâches de discrimination entre plusieurs langues, rien n'indique qu'ils soient bien adaptés à une discrimination entre 13 dialectes de l'anglais des Îles Britanniques. Ensuite, rien ne permet non plus de penser que ces 13 dialectes soient « séparables » sur la base de leur rythme. S'il existe véritablement des classes de rythme dans les dialectes du corpus ABI, leur nombre est très certainement bien inférieur à 13. Une étude de perception permettrait d'obtenir une ébauche de réponse. Enfin, l'analyse phonétique *a posteriori* de corpus de parole qui n'ont pas été recueillis spécifiquement pour ce type d'études comporte malheureusement tous les inconvénients d'une expérience scientifique non contrôlée. En d'autres termes, les facteurs de variation induits par la lecture d'un texte sans consigne supplémentaire sont multiples, et ils peuvent en venir à perturber, voire occulter, l'effet étudié.

Revenons au second point abordé : le nombre de classes. À défaut d'une expérience de perception, une autre option consiste à utiliser une technique de classification avec apprentissage non-supervisé. Nous avons donc employé la méthode du *k-means clustering* pour tenter de déterminer l'existence d'un nombre de classes rythmiques optimal. Cette technique consiste à partitionner les données en un nombre déterminé de classes en minimisant la somme des sommes de distances d'un point au barycentre de sa classe¹¹³. Pour un nombre de classes allant de 2 à 13, la valeur moyenne des silhouettes est calculée. Les valeurs de PVI bruts et normalisés, vocaliques et consonantiques, ont été utilisées. Le nombre de classes optimal a été déterminé indépendamment pour la durée et l'intensité. Les valeurs de silhouette en fonction du nombre de classes sont données dans les Figures 5.12 et 5.13, respectivement pour la durée et l'intensité. On constate dans les deux cas que, d'après le critère de la valeur de silhouette moyenne, le nombre optimal de classe semble être 2. Afin d'apprécier la cohérence de chaque classe dans le cas à deux classes, A et B, le pourcentage d'individus appartenant à la classe A (pour le modèle de l'intensité) est représenté pour chaque dialecte dans la Figure 5.14. Si l'on observe les deux groupes de dialectes proches des deux pôles, *shl* et *ean vs lvp*, *gla* et *uls*, il est difficile de trouver

¹¹³La fonction `kmeans` du logiciel Matlab a été utilisée ; l'algorithme ainsi que les divers aspects de notre méthodologie sont détaillés dans la documentation du logiciel et dans la Section 2.4.5.

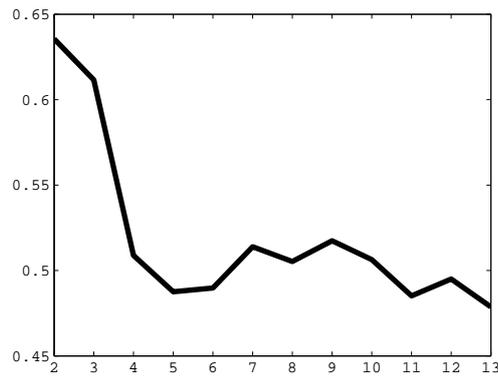


FIG. 5.12 – Moyenne des silhouettes en fonction du nombre de classes (durée).

une explication phonologique satisfaisante, notamment compte tenu du fait que les deux dialectes écossais sont aux antipodes l'un de l'autre. Une étude systématique des corrélats acoustiques de l'accent de mot dans les dialectes serait éventuellement informative. Nous avons également représenté les pourcentages de classification pour la solution à trois classes (A', B' et C') par le biais d'un diagramme ternaire (Figure 5.15). On y relève tout d'abord l'existence du groupe *brm*, *ean*, *crn* dont la majorité des locuteurs (tous pour *brm*) appartiennent à la classe A'. Ces trois dialectes forment une entité géographique cohérente : le sud de l'Angleterre. La proximité des dialectes *roi* et *lan* au barycentre du triangle démontre qu'ils ont chacun une proportion approximativement identique de locuteurs dans chacune des classes. En ce qui concerne *lan*, ce résultat n'est pas surprenant puisque l'analyse discriminante (voir la Table 5.3) donnait 0% de classification correcte pour ce dialecte. Les locuteurs des dialectes *ncl*, *lvp*, *gla* et *nwa* ont tendance à être regroupés dans la classe B'. La cohérence géographique ou linguistique ne peut pas véritablement justifier ce regroupement. Enfin, on voit émerger un troisième groupe rassemblant *sse* et *uls*. Pour ce dernier, s'il est certain que la durée et l'intonation systémiques diffèrent entre les deux dialectes, on peut néanmoins supposer que leurs schémas d'intensité sont proches.

Nous avons abordé la possibilité de mesurer le rythme dans un espace à de nombreuses dimensions (durée, intensité, etc.), mais certains auteurs proposent au contraire une seule dimension. Dans leur étude, O'Dell et Nieminen ([ON99]) rapportent une seconde analyse

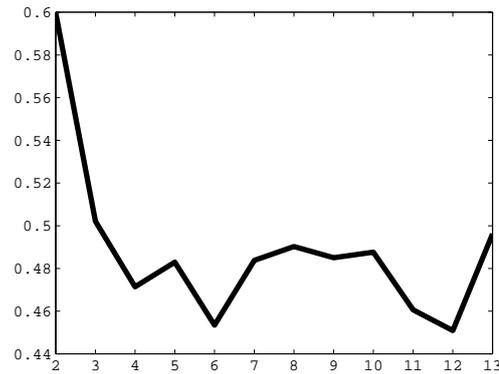


FIG. 5.13 – Moyenne des silhouettes en fonction du nombre de classes (intensité).

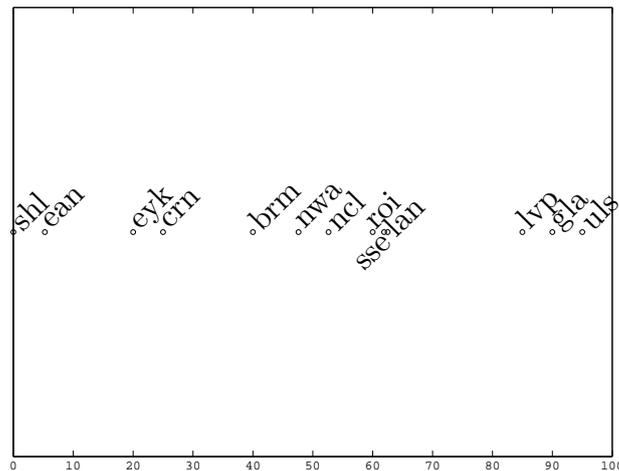


FIG. 5.14 – Pourcentage d’individus appartenant à la classe A (intensité).

des données de Dauer par Eriksson. Partant du constat que l’intervalle entre accents semble être une simple fonction linéaire du nombre de segments qu’il contient, l’auteur utilise la régression linéaire et obtient les résultats présentés dans la Table 5.4.

On peut noter tout d’abord les coefficients de corrélations remarquablement élevés, et ensuite, le terme constant, qui semble permettre la séparation des langues en deux groupes suivant la dichotomie traditionnelle : environ 100 ms pour les langues syllabiques et environ 200 ms pour les langues accentuelles. O’Dell et Nieminen ([ON99]) reconnaissent que cette méthode offre un moyen commode de classer les langues mais qu’elle n’a aucun

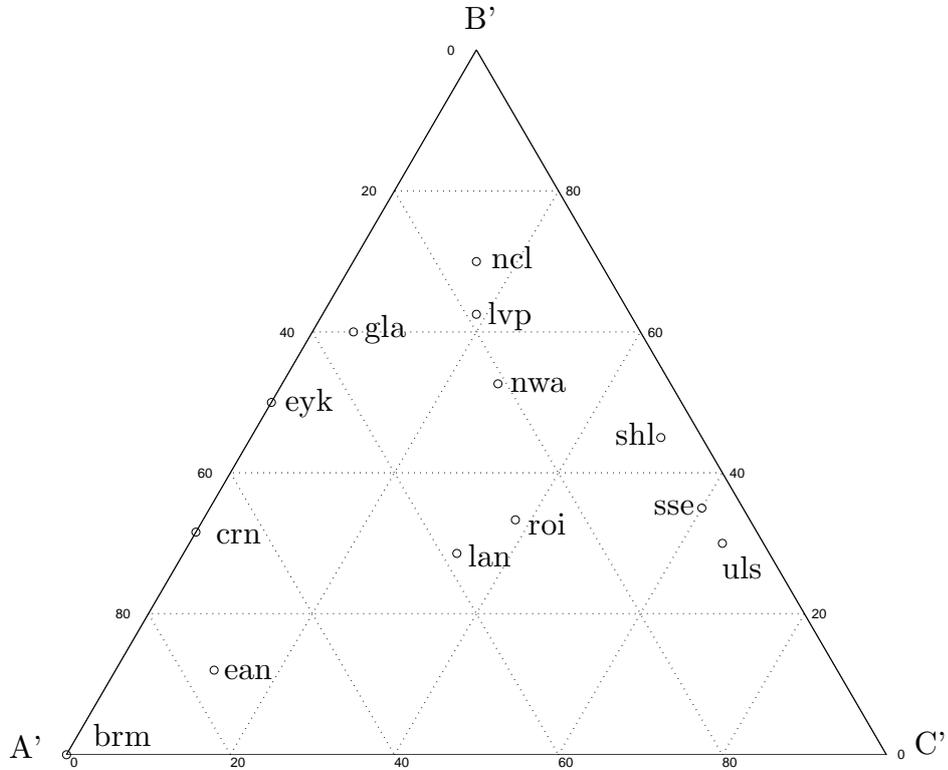


FIG. 5.15 – Diagramme ternaire (intensité).

TAB. 5.4 – Analyse en régression linéaire des données de Dauer par Eriksson ([ON99]).

Langue	Équation	Coefficient de corrélation
anglais	$ISI = 102n + 201$	$r = 0,996$
thai	$ISI = 97n + 220$	$r = 0,973$
espagnol	$ISI = 119n + 76$	$r = 0,997$
grec	$ISI = 104n + 107$	$r = 1,000$
italien	$ISI = 105n + 110$	$r = 1,000$

pouvoir explicatif. En effet, la différence entre les termes constants des équations de droite suggère que la syllabe accentuée dans les langues accentuelles est plus longue que dans les langues syllabiques. En réalité cette différence ne dit rien concernant la manière dont cette durée supérieure est distribuée sur toutes les syllabes des intervalles entre accents. O'Dell et Nieminen détaillent ensuite l'application du modèle mécanique des oscillateurs couplés à la description du rythme. Nous n'abordons pas cet aspect ici mais nous l'envisageons toutefois pour la suite de nos travaux car les oscillateurs couplés ont déjà été utilisés pour modéliser les comportements rythmiques biologiques et, en particulier, les aspects temporels de la parole (e.g. [GCS07]).

Avant de nous tourner vers l'étude des caractéristiques spectrales des voyelles, récapitulons ce que nous a appris notre étude du rythme :

- la transposition à la problématique des dialectes de méthodes issues d'études multilingues a permis de représenter la variation de rythme des dialectes du corpus ABI, à défaut de réellement les classer ;
- notre segmentation automatique, quoique non corrélée (ou très peu) aux valeurs obtenues par segmentation manuelle, offre un moyen potentiellement intéressant d'étudier le rythme sur des corpus de grande taille ;
- l'introduction du PVI d'intensité constitue une véritable nouveauté ; son pouvoir discriminant s'est révélé supérieur à celui des PVI de durée ;
- la faiblesse des résultats de classification, prévisible à partir des représentations bidimensionnelles typiques des études sur le rythme, s'explique en partie par la nature non-contrôlée du corpus, et en partie par le fait qu'il est malaisé de connaître *a priori* le nombre de classes de rythme qui caractérisent les dialectes de l'anglais britannique.

Chapitre 6

Les voyelles

Notre analyse de l'aspect segmental de la parole exclut, faute de temps et de place, les consonnes. Nous nous concentrerons donc sur les voyelles (accentuées) des 13 dialectes du corpus ABI, et tenterons de fournir une description actuelle aussi complète que possible. L'inventaire des spécificités dialectales décrites dans la littérature traitant des voyelles a été dressé plus haut ; ce qui suit fait donc écho au Chapitre 3.

6.1 Méthodes de représentation

Nous abordons ici les méthodes de représentation – notamment graphique – des voyelles. Afin d'être complet, nous ne saurions faire l'économie de quelques rudiments de phonétique. Nous inclurons en outre quelques aspects de l'histoire de la phonétique ; ces questions nous paraissent indispensables pour expliquer notre compréhension actuelle de ce qu'est une voyelle. Nous nous attarderons sur les développements des soixante dernières années.

Sur le plan articulatoire, les voyelles sont traditionnellement décrites en référence au point le plus haut de la langue dans la cavité orale. L'une des premières descriptions cohérentes dans cette perspective semble être celle de Robert Robinson en 1617 ([Lad67], 62-64). Chomsky et Halle ([CH68], 260 *sqq.*) mentionnent Hart, qui proposait déjà une des-

cription précise des voyelles dans les années 1570. Ladefoged cite ensuite John Wallis¹¹⁴, qui publie en 1653 son *Grammatica Linguae Anglicanae*, et qui aura une influence certaine sur les descriptions subséquentes. On attribue généralement l'invention du triangle vocalique (articulatoire) à l'Allemand Hellwag dans un traité sur les voyelles de 1781 (voir [Mal71], 29). C'est au 19^e siècle que la manière de décrire les voyelles sur des critères articulatoires que nous connaissons de nos jours se fixe avec des auteurs comme Alexander Ellis, Alexander Melville Bell (le père de l'inventeur du téléphone) avec son *Visible Speech*, ou encore Henry Sweet, qui propose jusqu'à 72 voyelles de référence. L'Alphabet Phonétique International liste 28 voyelles sous la forme d'un trapèze vocalique ; ce sont, à quelques détails près, les voyelles cardinales, dont la version actuelle est directement héritée de Daniel Jones et Paul Passy. Comme le notent Collins et Mees ([CM99]), Passy est probablement le premier phonéticien à avoir représenté les voyelles sous la forme d'un trapèze ; c'était en 1888. D'ailleurs, il n'est peut-être pas anodin que certaines voyelles cardinales soient extrêmement proches des voyelles du français, compte tenu de l'influence de Passy sur Jones, et du fait que ce dernier était résolument francophile. En 1967, Ladefoged fait un point sur la description phonétique de la qualité vocalique. Il estime que la rigueur scientifique fait défaut dans nombre de travaux de ses prédécesseurs et contemporains. On retiendra notamment que jusqu'à Jones exclus, les phonéticiens utilisent des étiquettes articulatoires pour exprimer ce qui s'apparente davantage à des critères auditifs. Notons toutefois que dès 1917, Daniel Jones avait publié des radiographies illustrant la position de la langue pour les voyelles [i a u] ([CM99], 188).

Sur le plan acoustique, ce sont les formants, des prééminences spectrales correspondant aux résonances des cavités supra-glottiques, qui sont encore majoritairement utilisés de nos jours pour décrire les voyelles. Les fondements de cette approche remontent, d'après Ladefoged, à Robert Willis qui, dans les années 1830, avait couplé une anche libre (la source) à des tubes de taille variable¹¹⁵ (le filtre) qui, en fonction de leur fréquence caractéristique, amplifiaient tel ou tel harmonique ([Lad67], 72). Incidemment, l'idée de

¹¹⁴L'inventeur du signe de l'infini en mathématiques, et des intégrales éponymes.

¹¹⁵C'est le principe de fonctionnement du « reed organ », cousin de l'harmonium.

résonance caractéristique d'une voyelle donnée était déjà bien comprise par Isaac Newton lui-même quand il explique dans ses notes aux alentours de 1665 que le fait de remplir une bouteille avec un flux constant de bière fait entendre, dans l'ordre [w u o ɔ a e i y]¹¹⁶. On peut inférer, en tenant compte de nos connaissances actuelles, que Newton entendait le deuxième formant vocalique. Il est particulièrement instructif de relever que Newton avait également entrevu la corrélation entre mouvements de la langue sur l'axe antérieur/postérieur et résonances (en l'occurrence, du deuxième formant). Toutefois, il semble faire erreur concernant l'aspect articulatoire quand il note que, par exemple, avec la langue qui s'éloigne du larynx, on passe de [i] à [u], alors que, nous le savons, c'est précisément l'inverse (le passage sur lequel nous basons nos propos est tiré de Ladefoged [Lad67], 65). L'approche de Willis a été reprise par Hermann von Helmholtz dans les années 1860 pour aboutir aux fondements de la théorie des résonances, qui préfigure la théorie acoustique utilisée de nos jours dans la description des voyelles. Helmholtz reprend en réalité l'approche Chord-Tone Theory de Charles Wheatstone, selon laquelle les fréquences amplifiées par les résonateurs sont des multiples entiers du fondamental, i.e. des harmoniques ([Mat99]). Mais il faudra encore attendre les années 1940 pour que se fixe la représentation actuelle. En effet, passées les machines singulières dont Rousselot fait l'inventaire [Rou01], on constate que la compréhension de la structure acoustique des voyelles s'améliore en fonction de l'évolution technologique des méthodes de visualisation. Illustrons ceci par l'exemple de deux articles publiés à un an d'intervalle dans le *Journal of the Acoustical Society of America*. En 1935, Scripture observe des tracés de voyelles obtenues de la même manière que l'était le son sur un film parlant : la bordure de la bande enregistre les variations de courant électrique provenant d'un microphone. Les tracés résultants, proches, conceptuellement, d'un oscillogramme, ne laissent pas apparaître, d'après Scripture ([Scr35], 171), « a trace of either the fundamental or a harmonic of an enforcing vibration ». Un an plus tôt, Steinberg ([Ste34]), au contraire, fait le lien entre le spectre des voyelles et les harmoniques du fondamental : les spectres (fréquence \times in-

¹¹⁶Ces symboles correspondent à notre interprétation des signes utilisés dans le texte original.

tensité) de chaque période d'une même voyelle sont disposés les uns à la suite des autres. Certains de ces tracés font apparaître plus de 3 formants. L'auteur reconnaît volontiers à propos de ces résonances au-delà des 3 kHz que « it is felt that the indicated amplitudes are beyond the resolving power of the equipment » ([Ste34], 18).

D'après l'historique des Bell Laboratories¹¹⁷, c'est en 1942 qu'est inventé le spectrographe¹¹⁸ qui permettra 13 ans plus tard à Potter et Peterson ([PP48]) d'affirmer, quoique avec une certaine circonspection : « Differences between vowel sounds can apparently be traced to differences in the frequency position of the resonances » (Soulignement ajouté). Ces derniers considèrent que le deuxième formant est le plus important, viennent ensuite le premier et le troisième, les formants supérieurs étant secondaires. En revanche, la représentation graphique n'est pas encore fixée : certes, les auteurs soutiennent que le graphe en deux dimensions (F1/F2) donne une définition suffisante des voyelles (ce point de vue est toujours d'actualité), mais l'axe des abscisses correspond à F1, et les ordonnées à F2, les axes n'étant pas inversés comme aujourd'hui. Autrement dit, ils ne recherchent pas encore de correspondance avec les triangles et quadrilatères articulatoires connus. Les auteurs constatent que, malgré des variations importantes de valeurs formantiques absolues d'un locuteur à l'autre, les rapports entre formants pour une même voyelle sont assez constants. En outre, ils proposent un système numérique servant à désigner une voyelle à partir de ses coordonnées sur F1/F2 : les deux premiers chiffres désignent la valeur des abscisses de toute voyelle située, sur cet axe, à 20 Hz (ou cycles) de la valeur précise donnée. Les deux derniers chiffres donnent le point de coordonnées de la deuxième dimension, et représentent toute voyelle située, sur cet axe, à 40 Hz. Autrement dit, les auteurs suggèrent ici la nature logarithmique de l'échelle subjective de perception de la fréquence. Ils vont encore plus loin en préconisant l'utilisation d'une échelle linéaire au-dessous de

¹¹⁷http://www.bell-labs.com/about/history/innovations_in_sound.html; site consulté le 7 avril 2008.

¹¹⁸[KDL46] détaillent le principe de fonctionnement de l'appareil en donnant des exemples de représentations graphiques antérieures du signal, ainsi que des illustrations de sons de la parole à l'aide du modèle le plus récent de l'appareil. L'accent est mis sur « visible speech », i.e. l'expression de Bell qui traduisait une volonté d'aider les sourds. Pour l'anecdote, ils mentionnent également une application potentielle pour l'apprentissage du chant. Le spectrographe est connu du grand public dès 1945 par le biais d'un article de R.K. Potter dans *Science* intitulé « Visible Patterns of Sound ».

1 kHz et logarithmique au-delà (possibilité également évoquée chez [KDL46] – le premier auteur donnera d’ailleurs son nom à une échelle psycho-acoustique de fréquence). La même année, le linguiste Martin Joos, qui avait servi comme oreille d’or pendant la guerre, met en lumière la correspondance, au prix de quelques raccourcis, entre la représentation acoustique des voyelles dans le plan F1/F2 et l’espace vocalique articulatoire ([Joo48], 52-53). Sa figure (page 52) semble être la première représentation en deux formants avec F1 sur les abscisses et F2 sur les ordonnées avec des axes inversés¹¹⁹. Il utilise une échelle logarithmique, l’échelle musicale.

À cette époque, la question – toujours d’actualité ! – se pose de l’utilité des innovations technologiques pour le linguiste. Deux points de vues opposés coexistent. Comme le note Mattingly ([Mat99], 2), Bloch et Trager soutiennent que les termes acoustiques ne sont d’aucune utilité pour la quasi-totalité des linguistes. En revanche, la terminologie des traits distinctifs de la phonologie de Roman Jakobson (par exemple, [Jak63], 128-130) s’inspire directement de la lecture de spectrogrammes (compact/diffus, grave/aigu, diésé/non-diésé, etc.)¹²⁰. En 1968, [CH68] fait déjà de la phonologie de laboratoire : le propos est hautement linguistique, en cela qu’on s’intéresse à la langue, au système, mais on y trouve néanmoins des références ([CH68], *passim*) à des données de phonétique instrumentale (données radiographiques, pression de l’air, etc.). Notons toutefois que les étiquettes phonologiques inspirées de la lecture de spectrogrammes sont explicitement abandonnées ([CH68], 306).

L’étude de référence sur la variation des formants vocaliques, encore souvent citée de nos jours, reste celle de Peterson et Barney ([PB52]). Les études précédentes réalisées par l’équipe des Bell Telephone Laboratories mettaient l’accent sur l’aspect technologique ; celle-ci est une véritable étude de phonétique en cela qu’elle rapporte des mesures effectuées sur un panel important de locuteurs (76), que femmes, hommes et enfants sont inclus, et que des différences d’accent sont mentionnées, quoique non exploitées. En outre

¹¹⁹Ladefoged (1967, 73) rapporte cependant que Essner avait déjà remarqué la correspondance entre les quadrilatères vocaliques articulatoire et acoustique en 1947.

¹²⁰La version anglaise de ce chapitre se trouve dans *Fundamentals of Language*, 13-66.

l'étude allie la production à la perception : 70 juges sont chargés d'identifier les voyelles enregistrées. Une de leurs figures ([PB52], 177) montre un trapèze vocalique (curieusement appelé « vowel loop ») sur un plan F1/F2 avec axes inversés ; les valeurs formantiques sont exprimées sur l'échelle psycho-acoustique des mels. Ce schéma résume l'expérience de classification des voyelles : on note d'importants écarts entre les différents timbres. Par exemple, [i ɜ a] et [u] sont souvent correctement identifiés alors que [a] ne l'est que 9 fois sur 152. Les stimuli utilisés pour la production et la perception sont des mots en /hVd/ (heed, hid, head, etc.). Les résultats des mesures acoustiques sont présentés en 2 dimensions avec F1 en abscisse et F2 en ordonnée. Les valeurs sont en Hertz¹²¹, mais représentées graphiquement en suivant l'échelle auditive de Koenig. Aucune innovation marquante n'a depuis remplacé la représentation en 2 dimensions (F1/F2) des voyelles. La correspondance du trapèze acoustique avec le trapèze articulaire a fini par imposer, dans les années 60, l'orientation de F1/F2 que nous connaissons aujourd'hui (par exemple, [Del65], *passim*). Nombre d'études phonétiques actuelles, et depuis les années 1960, dans des domaines variés tels que l'apprentissage d'une langue étrangère, la dialectologie et la sociolinguistique, se fondent sur [PB52]. Le cadre théorique acoustique demeure celui exposé chez Fant ([Fan60]).

En ce qui concerne le nombre de paramètres nécessaires à une représentation convenable des timbres vocaliques, il est généralement admis que F1 traduit l'ouverture, et F2, l'antériorité. Johnson ([Joh05], 365) cite sept études mettant en relation l'espace perceptif avec l'espace acoustique : en utilisant le *multidimensional scaling* (MDS)¹²² ces travaux démontrent que les deux premières dimensions de la perception des voyelles correspondent à F1 et F2. Les formants supérieurs ont un effet moindre, mais qui n'est pas négligeable. Kewley-Port et Atal ([KPA89]) examinent la perception de stimuli de synthèse en manipulant F1 et F2. Leur expérience s'appuie sur trois régions de l'espace vocalique : /i - ɪ/,

¹²¹A l'époque, on parlait encore de « cycles per second ».

¹²²Il s'agit d'une méthode proche, dans ses objectifs, de l'analyse en composantes principales, visant à réduire la dimensionnalité d'un espace caractérisant des objets (ici, les voyelles) afin de mettre en évidence de nouvelles dimensions plus pertinentes. L'un des intérêts de la méthode réside dans le fait qu'il est possible, après calcul, d'estimer la contribution des paramètres initiaux aux nouvelles dimensions pertinentes.

/u - υ/ et /ε - æ/. Pour chacune de ces zones, 2 stimuli prototypiques (e.g. /i/ et /I/), puis 9 autres ayant une différence (soit sur F1, soit sur F2) tout juste supérieure au seuil différentiel de perception des formants (d'après Flanagan [Fla55]) sont générés. Les sujets écoutent chaque paire de voyelles (pour chacune des trois zones) et jugent la proximité des deux timbres sur une échelle à dix points. Les voyelles sont tantôt présentées dans l'ordre A, puis B, tantôt dans l'ordre inverse¹²³. La matrice de distances est ensuite utilisée pour le MDS ; les voyelles sont projetées dans un espace à deux dimensions. Ce nouvel espace perceptif s'apparente à celui défini par F1 et F2. Une étude de corrélation montre que 88 % de la distance subjective après MDS entre deux voyelles est déterminée par la distance euclidienne calculée à partir de F1 et F2 convertis en Bark.

Les efforts ultérieurs se sont concentrés sur le problème de la normalisation du locuteur (ainsi que d'autres aspects de la variation, par exemple, intra-locuteur) : étant donnée une seule et même voyelle perçue, comment expliquer que les mesures objectives réalisées sur cette voyelle – produite par un ou plusieurs locuteurs – varient autant ? Il convient donc d'écarter les caractéristiques propres à chaque locuteur pour ne représenter que l'invariant. On peut affirmer, bien que ceci ne constitue qu'une bipartition grossière, que les deux moyens de limiter la variation afin de ne conserver que les caractéristiques pertinentes consistent à :

- transformer les valeurs originales brutes – notamment de fréquence – afin qu'elles reflètent la manière dont l'oreille perçoit les sons,
- exprimer ces valeurs en référence à un point qui définit les particularités d'un locuteur.

Ces deux aspects ont été brièvement introduits dans la Section 2.3.1. La section qui suit décrit la méthode employée pour l'analyse des voyelles.

¹²³Il est instructif de noter que certains sujets ont été éliminés *post hoc* à cause d'une asymétrie entre leurs scores pour une même paire de voyelles entre les conditions AB et BA.

6.2 Analyse des voyelles

L'extraction des valeurs formantiques de chaque occurrence a été réalisée avec le logiciel Praat. L'algorithme *Burg* a été employé, conformément aux recommandations du manuel du logiciel. Il est important de noter que la détection de formants reste du domaine de l'estimation, et il arrive parfois que les formants estimés soient trop éloignés des valeurs attendues pour être d'une quelconque utilité. D'aucuns pourraient relever que rejeter des estimations parce qu'elles s'éloignent de nos attentes implique une certaine circularité dans la méthode. Ceci est partiellement exact, mais il convient tout d'abord d'énumérer les facteurs responsables des erreurs d'estimation pour mieux cerner les limites de notre méthodologie.

Si, comme nous l'avons vu (Section 6.1), les voyelles peuvent être partiellement décrites au moyen de deux résonances caractéristiques, les deux premiers formants vocaliques sont parfois tellement proches pour certaines voyelles postérieures produites par des sujets masculins, que l'algorithme, et souvent l'oeil humain, ne peuvent distinguer plus d'une prééminence spectrale dans la région concernée¹²⁴. Si, en outre, l'un des deux premiers formants possède une intensité trop faible, l'algorithme va ignorer ce dernier. Ces phénomènes ne se manifestent pas nécessairement sur l'intégralité de la durée d'une voyelle ; il arrive parfois qu'un « saut de formants » n'intervienne que sur une partie de la voyelle, voire même sur un seul point d'analyse.

Afin de contourner ces inconvénients, Gendrot et Adda-Decker [GAD05] ont fixé *a priori* des seuils de rejet par type de voyelles dans une étude des voyelles orales du français et de l'allemand. On remarque que dans cette étude le taux de rejet semble particulièrement faible (4%), et, conformément au point que nous venons de mentionner, les taux de rejet sont relativement importants pour des voyelles ayant deux formants très proches

¹²⁴On sait d'ailleurs que l'oreille, dans ces cas-là, ne perçoit pas deux formants distincts ; qu'il s'agisse de F1 et F2 pour certaines voyelles postérieures, ou de F2 et F3 pour quelques voyelles antérieures, une seule prééminence spectrale est perçue. En effet, lorsque deux formants sont distants de moins de 3,5 Bark, ils sont « intégrés » sur la plan de la perception. En d'autres termes, l'oreille perçoit un formant « effectif » correspondant à une moyenne des fréquences des prééminences impliquées, pondérée par leur amplitude respective (voir par exemple, [Pic99], 154-155).

(voyelles dites « compactes »), en particulier pour /y/ et /u/ en français. Nous avons préféré ne pas utiliser cette méthode car, outre le fait qu'elle peut conduire à rejeter des voyelles dont les valeurs formantiques, quoique déviantes, ne sont pas le résultat d'une erreur de mesure (par exemple lorsqu'un locuteur interprète phonologiquement une forme graphique de façon inattendue), elle paraît peu applicable sur des diphtongues à mouvements formantiques amples (e.g. CHOICE). La méthode que nous avons retenue consiste à comparer le spectrogramme de chaque voyelle avec le tracé estimé des formants. Lorsque le tracé semble totalement incohérent, la voyelle est rejetée. Décrivons à présent pas à pas la façon dont notre méthode a été implémentée.

Les mots en /hVd/, rappelons-le, sont segmentés dans le corpus ABI. Afin d'extraire la partie vocalique de chaque mot, nous avons, dans un premier temps, procédé à une détection automatique de la fréquence fondamentale à l'aide d'un script en langage Tcl/Tk utilisant les fonctions du Snack Sound Toolkit¹²⁵. Les trames voisées recouvrent en général une zone plus grande que celle que l'expert humain considèrerait comme strictement vocalique : en effet, le voisement s'établit parfois alors que le segment phonétique n'est plus un [h] mais pas encore tout à fait une voyelle, et il persiste après la voyelle pendant la fermeture du [d]. Malgré cela, cette façon de déterminer les frontières des voyelles a été retenue car elle était réalisable automatiquement.

Une fois les frontières des voyelles ainsi déterminées, l'estimation de formants proprement dite peut commencer. Si nous nous étions contenté d'utiliser Praat et son langage de programmation, nous n'aurions pas pu implémenter notre méthode. En effet, et ceci est un choix délibéré du concepteur, l'utilisateur ne peut pas, à l'heure où nous écrivons ces lignes, interrompre l'exécution d'un script pour lui fournir un argument supplémentaire. Formulé de manière plus pragmatique, cela revient à dire qu'il est à l'heure actuelle impossible d'écrire un programme qui s'interrompt pour demander à l'utilisateur si la voyelle représentée sur l'écran doit être rejetée ou conservée. La solution consistait donc à faire usage d'un autre langage de programmation capable d'appeler Praat, de récupérer

¹²⁵<http://www.speech.kth.se/snack/>; site consulté le 7 avril 2008

les valeurs de formants calculées par ce dernier, et surtout, de s'interrompre pour proposer à l'utilisateur une boîte de dialogue. Nous avons opté pour le langage de programmation du logiciel Matlab ; bien qu'il ne soit pas plus adapté à la situation qu'un autre, le fait que nous soyons familier avec ce langage a constitué un avantage décisif. Ainsi, chacune des 26408 voyelles en contexte /hVd/ du corpus ABI a été visualisée par le biais d'un spectrogramme sur lequel étaient surimposées les valeurs de formants prédites par Praat.

Comme nous l'avons vu dans la Section 6.1, les deux premiers formants vocaliques permettent une représentation parcimonieuse et commode de l'espace vocalique, tant visuellement (deux dimensions corrélées aux axes du trapèze articulatoire) que perceptivement (pertinence de F1 et F2 dans les études de perception en MDS, e.g. [KPA89]). De plus, les valeurs des formants supérieurs se sont révélées difficiles à obtenir dès le début de l'analyse. Il s'ensuit que nous nous sommes contenté des valeurs de F1 et F2.

Mais les valeurs brutes n'ont pas été directement utilisées. En effet, il arrive parfois, comme nous l'avons déjà mentionné, que l'estimation de formants soit erronée sur une faible portion de la voyelle ; des valeurs incohérentes sont également présentes aux frontières des voyelles. Or ignorer toutes les voyelles concernées par ces deux cas de figure nous aurait conduit à un taux de rejet considérable. Un traitement *a posteriori* visant à minimiser l'impact des valeurs trop déviantes a été trouvé dans l'utilisation de la régression linéaire dite « robuste », telle qu'elle est implémentée dans le logiciel Matlab. Par rapport à la régression linéaire basée sur la technique des moindres carrés ordinaires, la régression robuste identifie les valeurs déviantes et les inclut dans le modèle avec une pondération moindre. Ce n'est donc plus le critère des moindres carrés que l'on cherche à minimiser, mais celui des moindres carrés pondérés (voir Section 2.4.6).

La méthode retenue implique cependant une simplification non négligeable : elle présume que les valeurs formantiques évoluent sur une droite en fonction du temps. Or ceci n'est pas exact, en particulier pour certaines diphtongues. Les Figures 6.1 à 6.3 permettent d'apprécier dans quelle mesure l'hypothèse de la linéarité constitue un biais certain. Dans la Figure 6.1, on constate l'adéquation de la régression linéaire classique pour modéliser

la course du premier formant d'une monophthongue, en l'occurrence, la voyelle de *hard*. La Figure 6.2, quant à elle, illustre le biais induit par l'ajustement d'une droite à un ensemble de points de F1 décrivant une trajectoire curvilinéaire pour une diphtongue : *hide*. La Figure 6.3 permet de visualiser la relative insensibilité aux valeurs déviantes de la droite de régression robuste (ligne continue) par rapport à la régression classique (ligne discontinue). Revenons à la Figure 6.2 avec une attention particulière aux valeurs de début et de fin, car ce sont elles qui sont utilisées dans la représentation graphique des diphtongues via un plan F1/F2. Si l'utilisation d'une droite renvoie une valeur estimée finale proche de la valeur réelle de F1, elle aboutit à une surestimation d'un peu plus de 100 Hz de la valeur initiale. Si l'on pose que la valeur initiale réelle dans la Figure 6.2 est tout à fait cohérente avec ce que nous avons pu observer sur l'ensemble des diphtongues du même type, alors l'expédient que nous avons choisi pour écarter les valeurs extrêmes (dues à des erreurs de mesure) se transforme en inconvénient. Afin de contourner ce biais, on peut supposer qu'ajuster une courbe polynomiale d'ordre supérieur plutôt qu'une droite fournirait une estimation plus proche de la réalité. Nous avons donc eu recours à la régression polynomiale. La Figure 6.4 permet de comparer l'ajustement d'une droite de régression robuste aux valeurs de F1 de l'occurrence de *hide* représentée dans la Figure 6.2 à celui d'une courbe obtenue par le biais d'une régression polynomiale de degré 3. On réalise aisément que le modèle polynomial est plus adapté à la course des formants de cette diphtongue car, outre le fait que la courbe passe plus près des valeurs de F1, la valeur initiale ainsi estimée ne s'écarte plus de la valeur réelle comme c'était le cas avec une droite. Mais en contrepartie, la sensibilité aux valeurs marginales extrêmes est accrue. Les Figures 6.5 et 6.6 sont un moyen alternatif de visualiser la Figure 6.4. Elles représentent l'écart entre les valeurs réelles de F1 (ligne interrompue) et les valeurs estimées par le modèle de régression (droite d'ordonnée 0), classique et polynomial, pour la Figure 6.5 et la Figure 6.6, respectivement. En plus de faire apparaître l'écart relativement important entre valeurs estimées et réelles pour la régression classique, le critère visuel de la dispersion aléatoire autour de zéro sur les ordonnées (synonyme d'adéquation du modèle) semble appuyer la

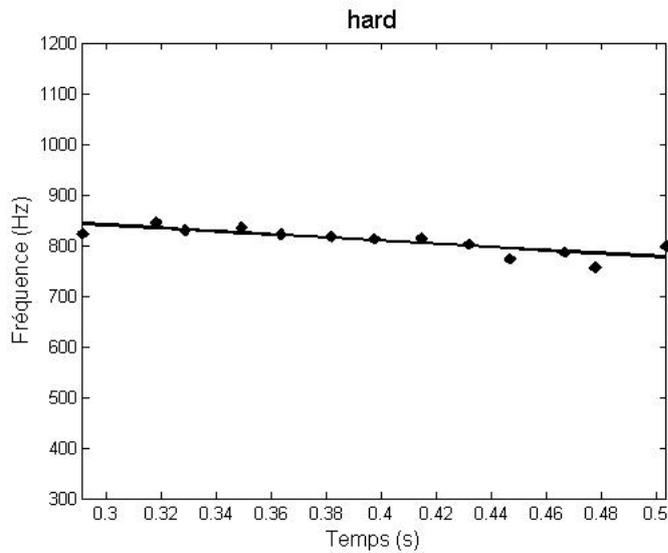


FIG. 6.1 – Valeurs de F1 pour la voyelle de *hard* (points) et droite de régression.

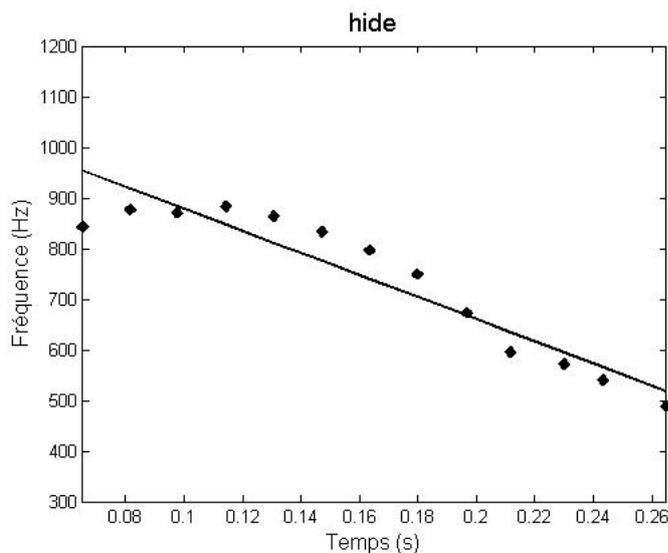


FIG. 6.2 – Valeurs de F1 pour la voyelle de *hide* (points) et droite de régression.

congruence du modèle polynomial. Mais la comparaison s'arrête ici car étant donné que les deux modèles ne minimisent pas le même critère (moindres carrés ordinaires pour le polynôme et moindres carrés pondérés pour la régression robuste), il serait aberrant de juger le modèle robuste par le biais de la somme des carrés de l'erreur.

Les techniques de régression linéaire robuste et régression polynomiale vont donc être utilisées pour modéliser les contours formantiques des monophthongues et des diphtongues,

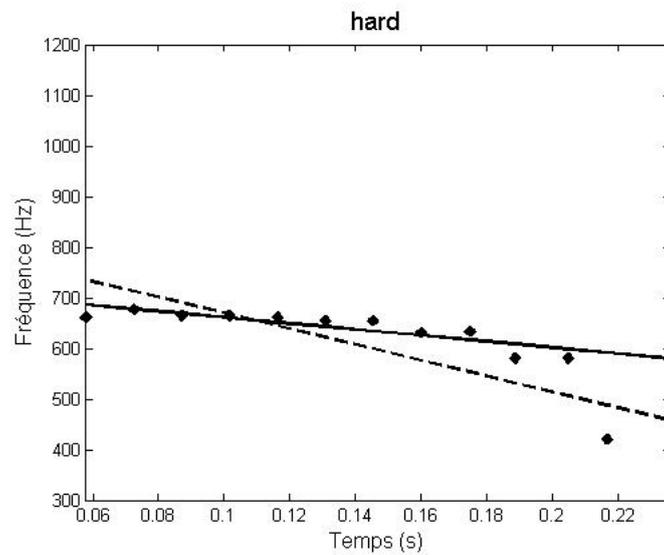


FIG. 6.3 – Valeurs de F1 pour la voyelle de *hard* (points), régression linéaire robuste (ligne continue) et régression linéaire classique (ligne discontinue).

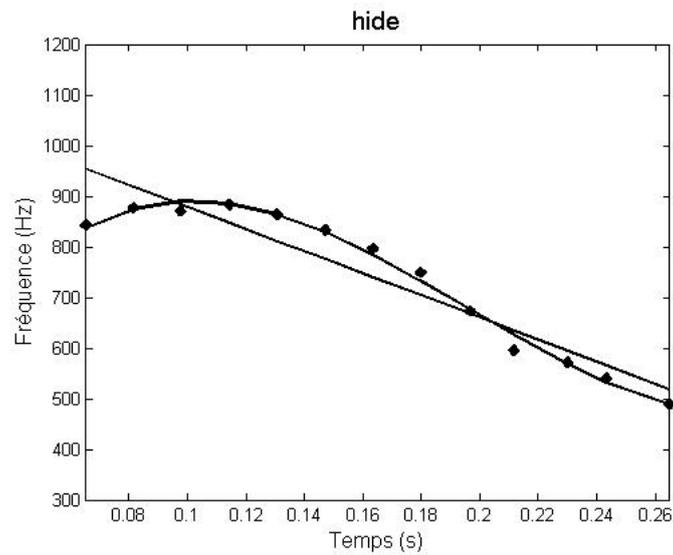


FIG. 6.4 – Valeurs de F1 pour la voyelle de *hide* avec droite de régression robuste et courbe issue d'une régression polynomiale cubique.

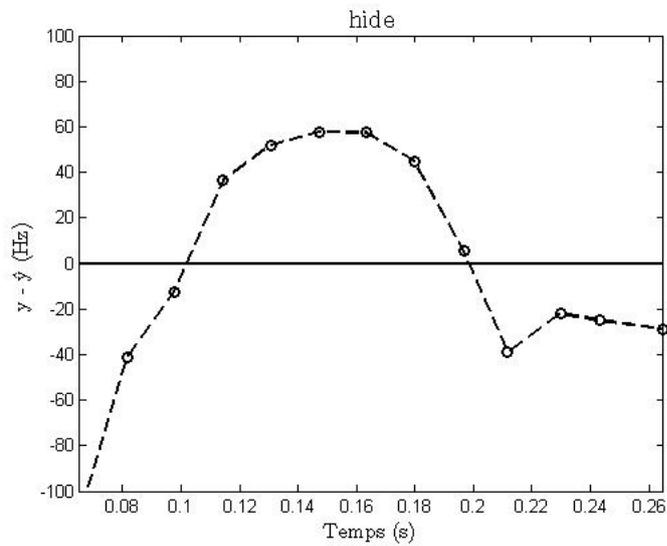


FIG. 6.5 – Graphe des résidus : écart entre les valeurs de F1 prédites (\hat{y}) par la régression linéaire robuste et les valeurs réelles (y , matérialisées par la droite d'ordonnée 0).

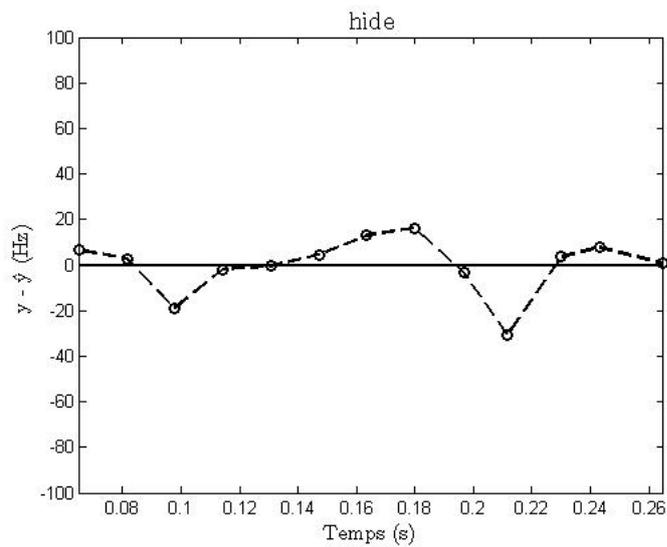


FIG. 6.6 – Graphe des résidus : écart entre les valeurs de F1 prédites (\hat{y}) par la régression polynomiale cubique et les valeurs réelles (y , matérialisées par la droite d'ordonnée 0).

respectivement. Reste à déterminer les points suivants :

- les diphtongues peuvent-elles être toutes caractérisées par le même modèle, tant sur la plan de la fonction algébrique utilisée que du nombre de paramètres ?
- comment choisir un modèle possédant des qualités de généralisation suffisantes ?

Si l'on accepte que la régression polynomiale constitue un modèle adéquat, il nous incombe de déterminer quel degré de polynôme caractérise le mieux nos données. Le choix d'un modèle repose sur un équilibre entre la parcimonie et l'ajustement aux données : on sait que plus un modèle possède de paramètres, plus il va être ajustable aux données réelles à partir desquelles il est estimé. Or maximiser l'ajustement n'est pas souhaitable puisqu'un modèle n'a d'intérêt que s'il permet une généralisation, une représentation stylisée, voire abstraite des données. Il est en outre souhaitable que le nombre de paramètres du modèle (dans notre cas, le degré du polynôme) soit aussi petit que possible. Le critère d'Akaike permet de juger l'équilibre que nous venons de mentionner. Le calcul de la quantité AIC (Akaike's Information Criterion) est donné dans l'Équation 6.1, où N est le nombre de données, SS , la somme des carrés de l'erreur et k , le nombre de paramètres plus 1. Il s'agit donc de minimiser ce critère sachant que, toutes choses égales par ailleurs, sa valeur est inversement corrélée à la taille des données, et croît avec le nombre de paramètres ou la somme des carrés de l'erreur.

$$AIC = N \times \ln \left(\frac{SS}{N} \right) + 2k \quad (6.1)$$

La version que nous utilisons comporte une correction qui pénalise davantage le nombre de paramètres lorsque la taille de l'échantillon est faible (rappelons qu'ici, $N = 13$) :

$$AIC_c = AIC + \frac{2k(k+1)}{N-k-1} \quad (6.2)$$

Le point faible de cette méthode vient du fait que, avec un nombre croissant de paramètres (i.e. degrés du polynôme), l'ajustement aux données (au sens des moindres carrés ordinaires) s'améliore inéluctablement et, par conséquent, la quantité SS diminue. Or

notre utilisation de la régression consiste avant tout à lisser les trajectoires des formants dans le but d'écarter les valeurs dues aux erreurs d'estimation. Malheureusement, si un point est particulièrement déviant, l'augmentation du nombre de paramètres conduit la courbe de régression à s'approcher de ce point et SS diminue (faisant diminuer AIC_c , *ceteris paribus*), alors que c'est précisément cette forme de surajustement que nous souhaitons éviter. Certes, la pénalisation du nombre de paramètres résout en partie ce problème, mais il est difficile de savoir dans quelle proportion. Nous ne sommes pas parvenu à implémenter une méthode qui pénaliserait les valeurs déviantes dans le calcul des moindres carrés (comme pour la régression robuste) dans le cas des polynômes ; nous utiliserons donc AIC_c avec circonspection. Les Figures 6.7 à 6.11 représentent la valeur de AIC_c en fonction du degré de polynôme pour le premier formant des 5 diphtongues fermantes produites par tous les locuteurs *sse*. Le critère a été calculé dans un premier temps sur la régression¹²⁶ de chaque occurrence ; les graphes présentés ici montrent la valeur médiane par type de voyelle. On y constate que le degré optimal (noté sur le graphe), d'après AIC_c , varie de 2 à 4 selon la diphtongue¹²⁷. Cela semble exagéré quand on sait que McDougall ([McD06]) et McDougall et Nolan ([MN07]) parviennent à capturer l'essentiel des caractéristiques individuelles de leurs 5 et 20 locuteurs (selon l'étude, respectivement) en utilisant des régressions polynomiales de degré 2 (quadratique) sur les trois premiers formants de /aɪ/ et /uɪ/.

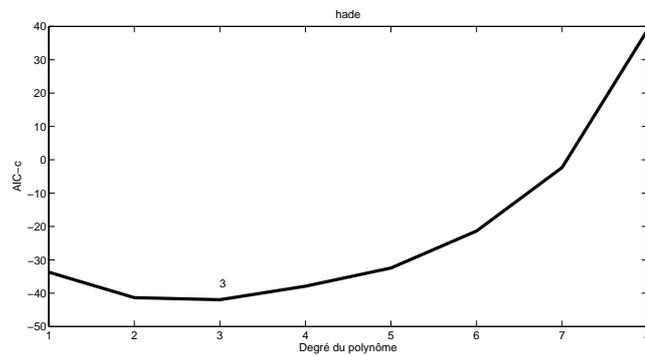
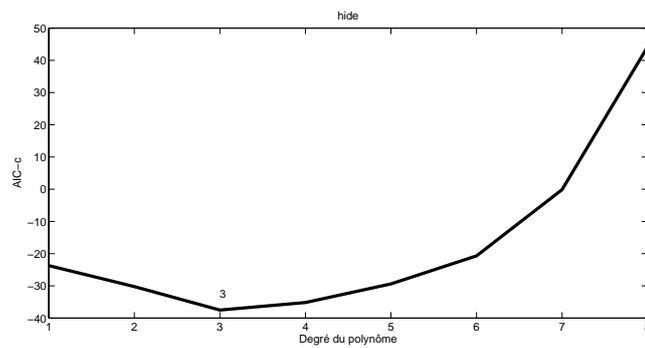
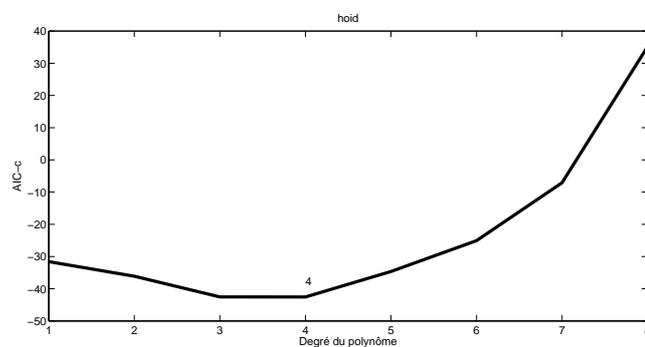
La solution que nous avons retenue nous est inspirée par l'observation empirique des trajectoires de formants, et en particulier celle de F2. Comme nous le verrons dans les spectrogrammes présentés dès la Section 6.3.1, on peut visuellement distinguer trois types de trajectoires formantiques pour les diphtongues :

- un long glide s'apparentant à une droite ;
- un état stable suivi d'un glide, ou l'inverse ;
- un état stable suivi d'un glide puis d'un nouvel état stable.

On sait qu'un polynôme de degré n ne peut pas produire une courbe comportant plus

¹²⁶Variable indépendante : temps ; variable dépendante : fréquence du formant en Bark.

¹²⁷Le même procédé a été appliqué à F2 avec des résultats similaires.

FIG. 6.7 – Critère d'Akaike en fonction du degré du polynôme pour le F1 de *hade* en *sse*.FIG. 6.8 – Critère d'Akaike en fonction du degré du polynôme pour le F1 de *hide* en *sse*.FIG. 6.9 – Critère d'Akaike en fonction du degré du polynôme pour le F1 de *hoid* en *sse*.

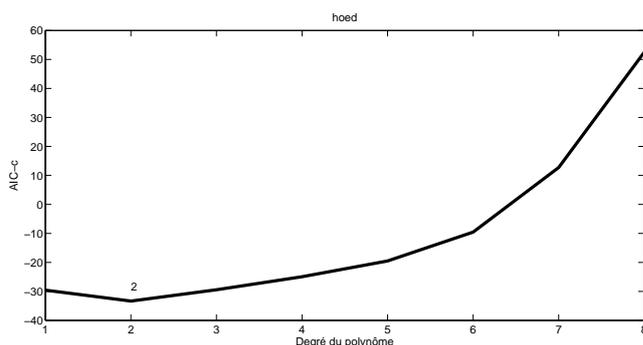


FIG. 6.10 – Critère d’Akaike en fonction du degré du polynôme pour le F1 de *hoed* en *sse*.

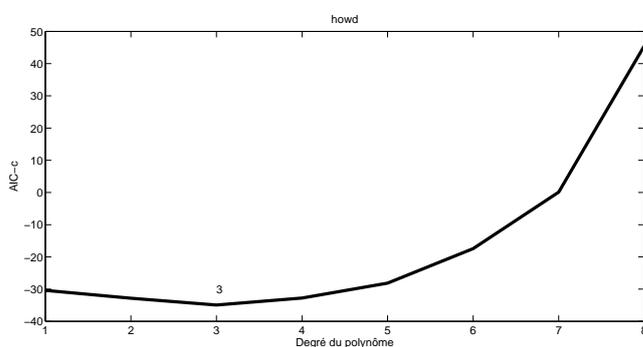


FIG. 6.11 – Critère d’Akaike en fonction du degré du polynôme pour le F1 de *howd* en *sse*.

de $n-2$ points d’inflexion. Dans le deuxième cas de l’énumération précédente, un polynôme quadratique paraît donc tout à fait indiqué. Un polynôme cubique est plus adapté dans le troisième cas puisque qu’il autorise un point d’inflexion dans la courbe, ce qui permet de capturer la dynamique sigmoïde des formants de ces voyelles. Le modèle à utiliser sera donc déterminé *ad hoc* pour chaque type de voyelle, et en fonction du dialecte, à partir de l’observation des tracés formantiques.

Pour chacune des 22331 occurrences de voyelles disponibles après inspection des spectrogrammes et tracés formantiques estimés, les vecteurs de temps et de fréquence des formants ont subi un ré-échantillonnage temporel afin que chaque voyelle comporte le même nombre de points¹²⁸, en l’occurrence, treize ; le nombre impair étant motivé par le

¹²⁸Ceci a été fait pour des raisons de commodité, mais nous nous devons d’insister sur le fait cette simplification n’est pas fondée phonétiquement. En effet, deux voyelles possédant un timbre perceptivement similaire, l’une étant phonétiquement longue, l’autre relativement brève, ne sont pas des versions obtenues par compression ou expansion linéaire l’une de l’autre.

fait que la valeur du milieu temporel nous intéressait pour la caractérisation des monophthongues. Dans un premier temps, les vecteurs de temps originaux de taille variable sont convertis en vecteurs de taille 13 grâce à la fonction `linspace` de Matlab. Puis, les valeurs des formants correspondant aux 13 nouveaux points temporels sont estimées par le biais d'une interpolation linéaire des valeurs formantiques originales via la fonction `interp1q`.

Ensuite, nous avons procédé à deux régressions « robustes » (une pour chaque formant) par le biais de la fonction `robustfit` de Matlab, avec pour variable indépendante, le temps, et la valeur du formant concerné comme variable dépendante. Ce sont les valeurs de F1 et F2 estimées par la régression qui sont ensuite utilisées dans le reste du calcul.

Les mots-clés contenant un <r> graphique ont pu ajouter un biais à l'analyse. En effet, si l'accent est rhotique, et que le <r> est réalisé comme une approximante alvéolaire ou rétroflexe, alors ce /r/ a été inclus dans la partie vocalique pendant la phase de segmentation automatique. Cette segmentation étant basée sur le voisement, et le [ɹ] étant une consonne voisée, ce dernier n'a donc pas pu être distingué de la voyelle le précédant. [ɹ] et [ɹ̥] sont des consonnes à formants; les valeurs formantiques des voyelles relevées dans les dialectes rhotiques sont donc influencées par la présence de ces approximantes. On sait qu'en moyenne, les trois premières proéminences spectrales du [ɹ] se situent aux alentours de 360, 1300 et 1600 Hz ([KDL96], 212; les valeurs données concernent l'anglais américain). Selon la nature du timbre qu'il suit, le [ɹ] pourra donc faire augmenter ou baisser les deux premiers formants de la voyelle, ou encore agir différemment sur l'un et l'autre.

6.3 Description acoustique des voyelles

La description qui suit est organisée par dialecte. Pour chacun d'entre eux, les voyelles ont été subdivisées en monophthongues et diphtongues. L'appartenance à l'une ou l'autre de ces catégories, calquée sur l'anglais standard, a été fixée *a priori* pour des raisons de commodité de l'exposé; ceci ne présume en rien de la qualité phonétique des voyelles dans

les différents dialectes. Pour les mots-clés de forme /hVd/ :

- les monophtongues sont : *heed, hid, head, had, hard, hod, hoard, hood, who'd, Hudd* et *heard* ;
- les diphtongues sont : *hade, hide, hoid, hoed, howd, heered, hared* et *hured*.

Nous avons séparé en outre les diphtongues fermantes des diphtongues centripètes.

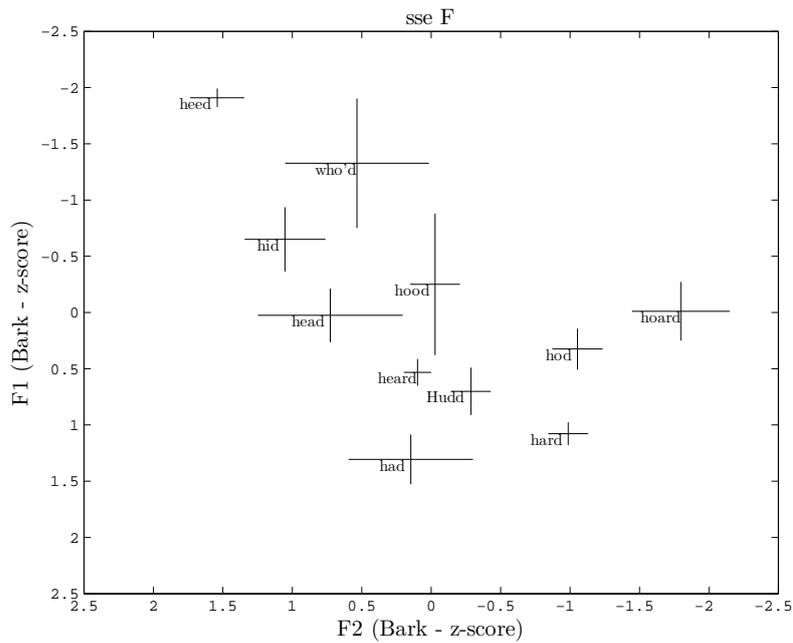
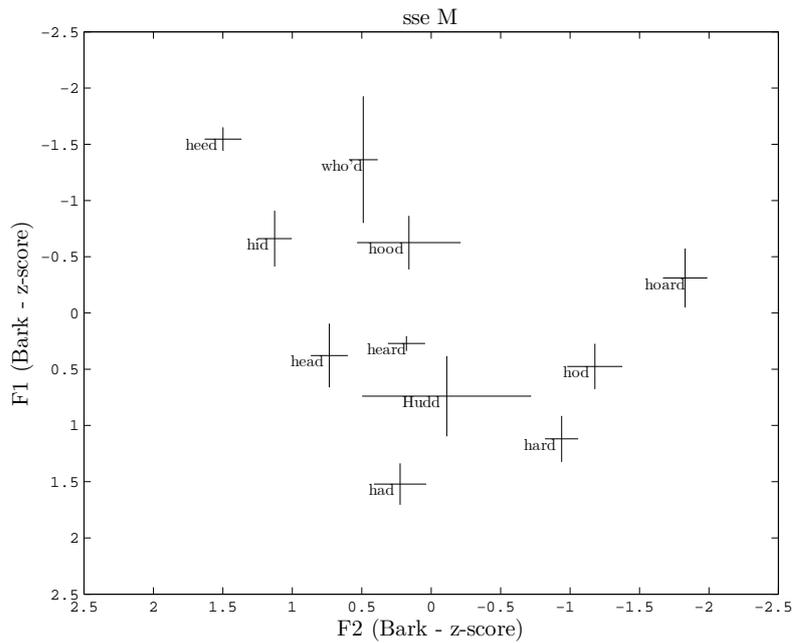
Pour les monophtongues, la valeur formantique en Bark du milieu temporel de la voyelle, telle qu'elle est donnée par la régression, a été extraite. Puis, les valeurs de l'ensemble des monophtongues ont été centrées-réduites par locuteur, indépendamment pour chaque formant, conformément aux résultats de l'étude comparative des méthodes de normalisation de Adank et collègues ([ASvH04])¹²⁹. Seuls les locuteurs ayant réalisé au moins une occurrence utilisable de chacune de ces 11 voyelles a été inclus dans l'analyse (centrer-réduire les valeurs formantiques pour un système lacunaire aurait induit un biais). Enfin, pour chaque locuteur, c'est la valeur médiane de chaque type de voyelle qui a été retenue comme estimateur robuste de la tendance centrale. Les résultats sont représentés dans un plan F1/F2 comme, par exemple, dans les Figures 6.12 à 6.13. Dans ces figures, chaque type de voyelle est matérialisé par sa valeur médiane et son écart interquartile (iqr). Dans la discussion qui suit, chaque dialecte est comparé (au moins implicitement) aux Figures 6.12 et 6.13, respectivement pour les locuteurs féminins et masculins. La médiane et l'écart interquartile de la durée de chacune des voyelles pour tous les locuteurs (hommes et femmes) sont listés sous forme de tableaux, comme, par exemple, dans la Table 6.1.

6.3.1 L'anglais standard (*sse*)

6.3.1.1 Les monophtongues

Dans la Figure 6.12, 4 locutrices ont produit l'ensemble des 11 voyelles. Ce sont 6 locuteurs qui ont produit l'ensemble des 11 voyelles dans la Figure 6.13. L'analyse d'un locuteur typique sera traitée dans un exposé à part (Section 6.3.1.3). Les Figures 6.12

¹²⁹Voir la Section 2.3.1 pour la conversion de Hertz en Bark et le calcul de valeurs centrées-réduites.

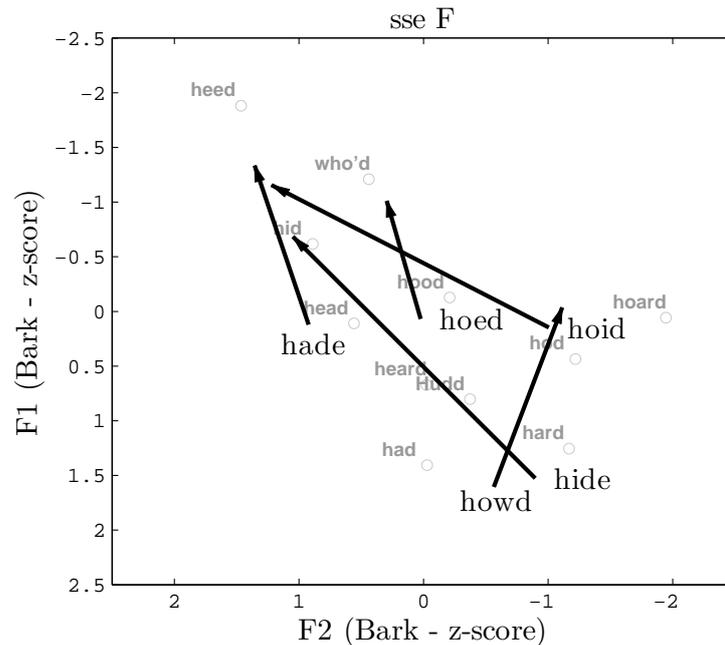
FIG. 6.12 – Monophthongs de *sse* pour les sujets féminins.FIG. 6.13 – Monophthongs de *sse* pour les sujets masculins.

et 6.13 constituent donc les systèmes de monophthongs de référence auxquels les autres systèmes de notre corpus sont comparés. Pour faire écho à ce qui a été dit dans la Section 3.1, on note l'antériorisation très marquée de *who'd* et *hood*. Ces deux timbres oc-

TAB. 6.1 – Durée des monophthongues (médiane et écart interquartile, *Iqr*) de *sse* comparée à la durée moyenne (3^e colonne) des monophthongues de l'étude de Wells ([Wel62]) en ms.

Voyelle	Médiane	<i>Iqr</i>	Moyenne ([Wel62])
heed	263	58	293
hid	178	56	139
head	194	64	170
had	213	63	210
hard	313	50	335
hod	207	50	178
hoard	313	60	330
hood	181	63	142
who'd	281	66	294
Hudd	194	50	148
heard	303	53	309

cupent d'ailleurs une aire relativement grande de l'espace vocalique chez les femmes. On remarque également que la voyelle de *had* a une position très ouverte, ceci étant tout à fait en adéquation avec les descriptions actuelles de l'anglais standard. On peut encore noter que *head* semble légèrement plus ouvert chez les hommes ; le processus de mouvement dans le sens inverse des aiguilles d'une montre, dont nous avons déjà fait mention (Section 3.1.2), serait-il plus avancé chez eux ? Les faibles écarts interquartiles pour *heard* ne nous permettent pas de confirmer la variation importante supposée affecter NURSE (voir Section 3.1.1). La distance séparant *Hudd* et *had* ne nous laisse pas non plus entrevoir une possible convergence entre ces deux ensembles. La durée médiane et l'écart interquartile des monophthongues de notre étude ainsi que la durée moyenne de chaque timbre de l'étude de Wells ([Wel62]) sont donnés dans la Table 6.1. Ces valeurs sont en accord avec l'analyse classique de ces timbres en voyelles phonologiquement longues et brèves ; la différence de durée moyenne entre ces deux catégories est de 95 ms.

FIG. 6.14 – Diphtongues de *sse* pour les sujets féminins.

6.3.1.2 Les diphtongues

Les données de 4 locutrices et 6 locuteurs sont représentées dans les Figures 6.14 et 6.15, respectivement. Les flèches représentent le déplacement des diphtongues dans l'espace F1/F2 ; l'origine de chaque flèche, accompagnée de l'étiquette du mot-test qui lui est associé, marque la valeur de départ de la diphtongue, et la pointe représente la valeur d'arrivée. Cet espace des diphtongues est superposé à celui des monophthongues. Les valeurs originales sont les coordonnées médianes (en Bark) de chaque monophthongue (les mêmes que dans les Figures 6.12 et 6.13), et les coordonnées médianes (en Bark) des points de départ et d'arrivée des diphtongues. Ces valeurs ont ensuite été centrées-réduites, indépendamment pour chaque formant. Toutes les diphtongues fermantes représentées dans ces figures ne feront pas l'objet de commentaires détaillés car les graphes sont suffisamment instructifs. Nous préférons nous concentrer sur quelques cas dont l'interprétation remet en cause certaines descriptions de ces voyelles.

Si l'on se fie à la description des diphtongues fournies dans la Section 3.1.1, et en particulier, à la Figure 3.2, on constate certaines divergences avec les Figures 6.14 et 6.15. Les

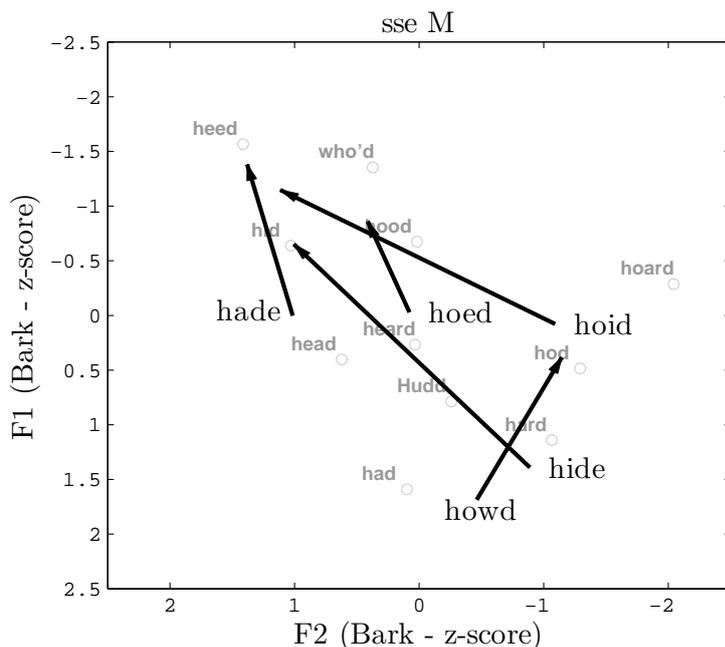


FIG. 6.15 – Diphtongues de *sse* pour les sujets masculins.

voyelles de *hoed* et *howd* sont habituellement représentées comme ayant le même point d'arrivée. Or nos données indiquent que, si le point d'arrivée de *hoed* correspond approximativement à la zone de la voyelle dont le symbole API est utilisé dans les dictionnaires de prononciation ([ʊ]), celui de *howd* est plus proche de la voyelle de *hod*. Nous ne saurions nous prononcer sur la pertinence de cette différence de timbre d'arrivée sur le plan de la perception ; lors de l'analyse auditive, ce point ne nous est pas apparu comme particulièrement saillant. La faible saillance perceptive du second élément de la diphtongue provient, très probablement, en partie du fait que les diphtongues fermantes présentent une amplitude relativement élevée sur le premier tiers de leur durée, et cette amplitude va décroissant par la suite.

La comparaison de *hide* et *howd* constitue un cas distinct du précédent : une différence de timbre sur le premier élément a été révélée dès l'écoute de ces mots-test. Rappelons que pour *hide* et *howd*, respectivement :

- Roach et collègues ([Jon03]) et Wells ([Wel90]) notent /aɪ/ et /aʊ/ ;
- Gimson ([Gim80]) et son contemporain, O'Connor ([O'C73]), transcrivent /aɪ/ et

/aʊ/;

- Olausson et Sangster ([OS06]) ont /ʌɪ/ et /aʊ/.

L'impression auditive qui se dégage de ces stimuli, doublée de l'analyse des formants nous conduit à la conclusion suivante : en moyenne, la voyelle de *hide* présente un timbre de départ plus postérieur que celle de *howd*. Ce qui nous amène à remarquer que, quelle que fût la réalisation prototypique réelle de ces deux voyelles à l'époque de Gimson¹³⁰, sa transcription ne reflète pas l'anglais standard actuel. Ce choix de Gimson en faveur d'une *via media* entre transcription phonologique minimaliste et guide de prononciation, s'il devait être fait aujourd'hui, nous conduirait vraisemblablement à utiliser comme premier élément de chaque diphtongue les symboles /ɑ/ et /a/, ou encore /ʌ/ et /a/ comme le proposent [OS06], pour *hide* et *howd*, respectivement.

À l'écoute, la différence de timbre initial entre *hide* et *howd* paraît plus marquée que ne le laissent supposer les Figures 6.14 et 6.15. C'est la raison pour laquelle nous avons souhaité examiner ces voyelles de façon plus approfondie. Pour tous les locuteurs masculins, la partie ayant la structure formantique la plus stable à l'initiale de ces deux voyelles a été segmentée manuellement. Il est possible de fournir une mesure de la stabilité formantique des parties segmentées par le biais de la variation des valeurs de formants sur la durée sélectionnée. La moyenne des écarts interquartiles sur toutes les occurrences de *hide* est de 43 et 68 Hz pour F1 et F2 respectivement. Pour *howd*, ces valeurs sont de 31 et 39 Hz. Pour chaque occurrence, le spectre moyen a été calculé sur toute la zone ainsi délimitée. Le spectre moyen englobant tous les locuteurs est représenté dans la Figure 6.16. En suivant la ligne continue, qui représente la voyelle de *hide*, le premier mode (aux alentours de 100 Hz) correspond à la fréquence fondamentale, le deuxième (vers 700 Hz), à F1, le troisième (vers 1000 Hz), à F2, et le quatrième (vers 2750 Hz), à F3. Les pics correspondant à F1 et F2 interviennent plus haut dans l'échelle des fréquences pour *howd* (ligne discontinue). F3 pour ce timbre se situe aux alentours de 2400 Hz. Plus généralement, l'aspect le plus saillant de la différence entre les spectres moyens de *hide*

¹³⁰Il est né en 1917 et décédé en 1985 (<http://www.phon.ucl.ac.uk/home/wells/gimson-ELL.pdf>; site consulté le 7 avril 2008).

et *howd* réside dans la différence d'amplitude sur une bande de fréquences comprise entre 1200 et 2800 Hz environ. Ceci rappelle que F1 et F2 ne permettent qu'une caractérisation économique (et donc partielle) des voyelles. En revanche, la représentation spectrale (fréquence \times intensité) autorise une vue plus exhaustive des indices acoustiques impliqués dans la différence de timbre perçue.

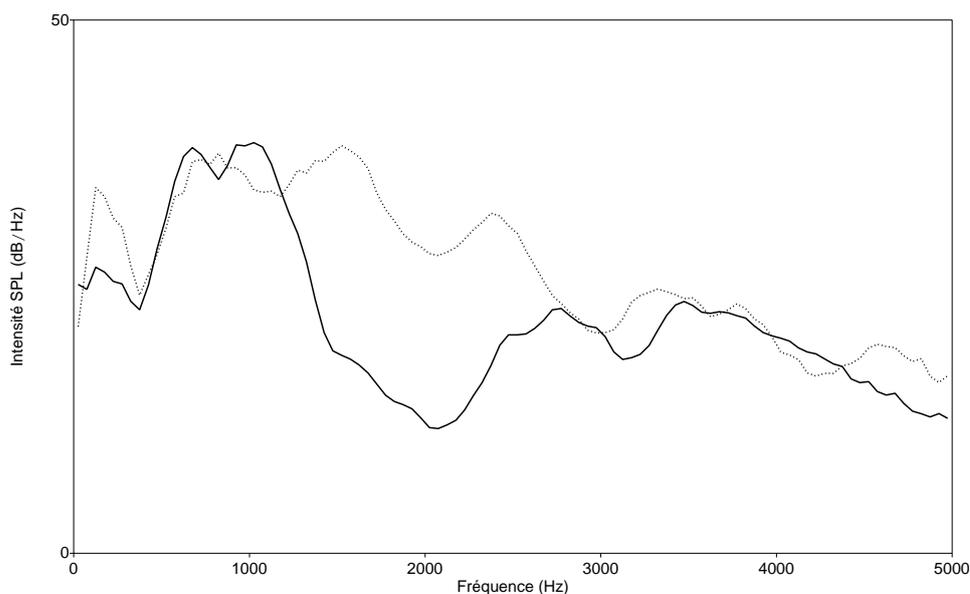


FIG. 6.16 – Spectre moyen du premier état stable de *hide* (ligne continue) et *howd* (ligne discontinue) des locuteurs masculins de *sse*.

L'analyse des diphtongues centripètes¹³¹ de *sse* est traitée dans la prochaine Section (6.3.1.3).

6.3.1.3 Analyse des voyelles d'un locuteur type de *sse*

Puisque *sse* constitue notre dialecte de référence, nous allons en fournir une description détaillée. L'expérience montre que si le calcul des valeurs acoustiques moyennes pour plusieurs locuteurs est nécessaire dans le traitement de corpus de grande taille (comme ABI), cela se révèle souvent dommageable, surtout lorsque ces valeurs sont utilisées pour

¹³¹Ces termes, rappelons-le, ne sont utilisés ici que pour la clarté de l'exposé ; en effet, une analyse en trois temps (monophtongues, diphtongues fermantes, diphtongues centripètes) permet de ne pas surcharger les diagrammes présentés. En revanche, ils ne font aucunement référence à la véritable qualité phonétique des voyelles.

une description fine. En effet, les caractéristiques centrales et la dispersion d'une variable aléatoire ne sont vraiment utiles que lorsqu'il s'agit véritablement d'une seule et même variable phonético-phonologique. Si nous prenons le raccourci de résumer les données de plusieurs locuteurs dans des représentations graphiques comme la Figure 6.12, c'est qu'il nous est matériellement impossible de présenter les données individuelles de tous les locuteurs de ABI; mais nous sommes bien conscient de la simplification qui sous-tend ces représentations. L'utilisation de l'échelle des Bark et le fait de centrer-réduire les valeurs formantiques permet, certes, d'effacer en grande partie la variation déterminée physiologiquement, mais quid de la variation linguistique à l'intérieur d'un dialecte? Partir du postulat que tous les locuteurs d'un dialecte possèdent le même système et les mêmes réalisations phonétiques, et que seules les différences physiologiques (et la variation aléatoire intrinsèque) engendrent la variation acoustique (ce que laissent croire la Figure 6.12 et les suivantes du même type) est tout à fait erroné. Par exemple, comme nous le verrons plus loin (Section 6.3.13.1), contrairement à la plupart des locuteurs de *uls*, le locuteur *fod* n'a pas d'homophonie entre *hood* et *who'd*, ce qui est matérialisé, notamment, par un F2 inférieur dans la voyelle *hood*. Si on mesure les valeurs de F2 pour *hood* produit par *fod* et par un autre locuteur ne présentant pas la distinction *hood-who'd*¹³², et qu'on calcule une moyenne sur ces deux locuteurs, la valeur obtenue (i.e. la moyenne d'une distribution bimodale) donne un F2 intermédiaire qui ne nous apprend rien car nous avons affaire à la forme de surface de deux phénomènes phonologiques différents. Ces raisons font que nous avons complété la description de tous les locuteurs de *sse* par celle d'un « locuteur type ». Notre choix a été guidé par les résultats de l'expérience de perception que nous avons détaillée dans la Section 4.2; de tous les locuteurs de *sse* (rappelons cependant que les locutrices n'ont pas été soumises à cette analyse), *rkm* est le seul à avoir obtenu la note maximale de 5 censée refléter son degré de conformité au prototype de l'anglais standard tel que le concevait l'expert.

Outre la potentielle perte d'information que nous venons de mentionner, et qui est

¹³²Dans ce cas, F2 est plus élevé.

occasionnée par la réduction des valeurs formantiques à une valeur centrale (moyenne ou médiane) sur plusieurs locuteurs, la nécessité d'analyser les voyelles d'un locuteur type s'explique par plusieurs autres raisons. D'abord, les graphiques (comme la Figure 6.12) qui illustrent nos descriptions ne donnent que la position relative moyenne des voyelles dans le système de tous les locuteurs d'un dialecte sans qu'il soit possible, du fait de la transformation en données centrées-réduites, de se référer à des valeurs absolues sur une échelle couramment utilisée par les phonéticiens (e.g. les Hertz). Nous allons donc décrire le système de *rk*m en nous appuyant sur des valeurs de formants en Hertz. Certes, cette échelle n'est pas conforme à la manière dont l'oreille perçoit les fréquences, mais nous la conservons car lorsqu'il s'agit de décrire des voyelles, ce sont les Hertz qui sont employés le plus souvent. Ensuite, puisque de nombreux spectrogrammes figurent dans les sections qui suivent, il est nécessaire de pouvoir se reporter à un ensemble de spectrogrammes de référence. Enfin, se pencher sur des réalisations individuelles permet d'aborder une simplification à laquelle nous avons eu recours dans le traitement de la dimension temporelle de nos données. En effet, la durée variable des voyelles, et donc le nombre variable de points d'analyse des formants¹³³, conduit à la situation où, pour un timbre donné, nous disposons d'un nombre différent de valeurs formantiques selon les occurrences. Or, par exemple, comment calculer le F2 moyen d'un ensemble de réalisations d'un timbre au temps 250 ms si une seule occurrence atteint cette durée ? La fonction `linspace` du logiciel Matlab nous a permis de « ré-échantillonner » le temps afin d'obtenir 13 valeurs pour toutes les voyelles. Or les voyelles se décomposent schématiquement en plusieurs phases ; par exemple, pour une monophthongue en contexte CVC, on observe une transition formantique, puis une structure relativement stable, et enfin une autre transition. Le ré-échantillonnage suppose que chacune de ces phases est une proportion constante de la durée de la voyelle, ce qui est inexact. Il semblerait en effet que la durée des transitions formantiques soit assez stable quel que soit le type de voyelle et que sa réduction sous l'effet d'un débit supérieur soit inférieure à celle que subit l'état stable ([Gay78]).

¹³³Étant donné un pas d'analyse constant.

L'analyse des voyelles de *rkm* a été conduite selon le schéma suivant : nous avons, dans un premier temps, visualisé le tracé des valeurs formantiques estimées superposé au spectrogramme de chaque voyelle à l'aide du logiciel Praat. Dans le cas où l'estimation de formants était incohérente avec le spectrogramme, nous avons ajusté empiriquement les paramètres de l'algorithme *Burg* jusqu'à obtenir une estimation satisfaisante. En pratique, cela consiste à faire varier la fréquence maximale au-dessous de laquelle sont attendus un nombre déterminé de formants. Nous avons ensuite délimité une zone temporelle correspondant à la voyelle : le début correspondait à l'apparition d'une structure formantique d'intensité relativement élevée et dépourvue du signal aperiodique du [h] précédant la voyelle ; la fin était marquée par une baisse d'amplitude soudaine dans le signal. En plus de cela, les valeurs des formants aux frontières de voyelles étant souvent mal estimées, nous avons parfois légèrement amputé les voyelles, le biais ne dépassant pas les 10 ms¹³⁴. Les durées des voyelles de *rkm* sont résumées dans la Table 6.2.

Les tracés des deux premiers formants ont été ensuite visualisés par type de voyelle afin de déterminer si les occurrences d'un même type étaient suffisamment homogènes pour n'être représentées que par l'une d'entre elles dans cette section. Enfin, nous avons délimité, pour les monophthongues, un état stable sur lequel ont pu être calculées des valeurs moyennes. Les moyennes et les écart-types de celles-ci sont présentés dans la Table 6.3. La Table 6.4 liste les valeurs de formants obtenues par Hawkins et Midgley dans leur étude de 2005 ([HM05]). Cette étude est une analyse des mesures de formants pour les 11 monophthongues de l'anglais standard produites par 20 locuteurs (masculins) divisés équitablement en 4 tranches d'âge. Les mots-test sont identiques aux nôtres, hormis *Hudd* et *heard*, qui sont remplacés par <hud> et <herd>. La Figure 6.28 représente les monophthongues du locuteur *rkm* dans le plan F1/F2¹³⁵.

La voyelle de *heed* est une monophthongue longue antérieure fermée. Au vu de ce que

¹³⁴On note que, contrairement aux autres mesures acoustiques obtenues dans nos travaux, celles-ci ont requis une intervention humaine importante.

¹³⁵Les axes ne sont pas à la même échelle ; ceci reflète en partie, mais peut-être de manière exagérée, le fait que, lorsque la fréquence est en Hz, un mouvement sur F1 est (en moyenne) perceptivement plus important que le même mouvement sur F2.

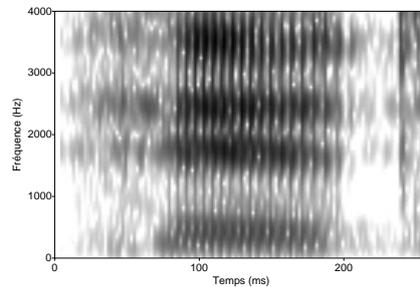
TAB. 6.2 – Durées des voyelles de *rkm* (ms).

Voyelle	Effectif	Moyenne	Écart-type
heed	5	199	12
hid	5	104	7
head	5	128	13
had	5	150	9
hard	4	253	4
hod	5	135	15
hoard	4	259	26
hood	4	97	8
who'd	3	215	19
Hudd	7	92	13
heard	5	249	18
hade	5	231	26
hide	5	255	28
hoid	7	221	22
hoed	5	220	24
howd	4	247	23
heered	5	219	9
hared	5	249	16
hured	7	246	13

nous savons de la voyelle de FLEECE (voir Section 3.1.1), il est assez surprenant de constater cette stabilité de la structure formantique qui correspond, à l'écoute, à une monophongue phonétique. Les locuteurs sont partagés de façon approximativement équitable entre la production d'une parfaite monophongue et d'une très légère diphtongue. Notons néanmoins que le /h/ qui précède cette voyelle dans les mots-test n'est pas un contexte favorable à la diphtongaison puisque les articulateurs sont déjà en position pour le [i] lors de l'émission du /h/¹³⁶. Dans la Figure 6.17, la stabilité de la structure formantique de *heed* par *rkm* (6.17(a)) peut être comparée à une réalisation de cette même voyelle par le locuteur *dme* (Figure 6.17(b))¹³⁷, qui a une diphtongue légère, certes, mais perceptible.

¹³⁶C'est sciemment que nous utilisons des barres obliques pour /h/ et des crochets pour [i] : si la qualité phonétique de la voyelle ne fait aucun doute, nous ignorons *a priori*, si, pour la fricative, un lieu d'articulation glottal – [h] – correspond à la réalité.

¹³⁷Ce locuteur a obtenu la note de 4 sur 5 dans l'évaluation de sa conformité au prototype (voir Sec-

FIG. 6.17 – *heed* produit par les locuteurs *rkm* et *dme* de *sse*.FIG. 6.18 – *hid* produit par le locuteur *rkm* de *sse*.

La voyelle de *hid* est une monophthongue brève mi-fermée légèrement rétractée : [ɨ]. Nous avons pu constater que le caractère rétracté de ce timbre était particulièrement flagrant chez le locuteur *rkm* (Figure 6.18) alors que les autres locuteurs de *sse* ont une réalisation plutôt antérieure.

La voyelle de *head* est une monophthongue brève mi-ouverte antérieure (Figure 6.19). Nous n'avons observé que très peu de variation entre les locuteurs de *sse*, hormis quelques rares variantes un peu plus fermées que [ɛ], par exemple pour le locuteur *prz*.

La voyelle de *had* est une monophthongue brève antérieure, ou parfois très légèrement rétractée, de degré d'aperture maximum (Figure 6.20(a)). Seul un locuteur présente une réalisation un peu plus fermée que les autres : *prz* (Figure 6.20(b))¹³⁸. On remarque en effet chez ce dernier que F2 est comparativement élevé.

tion 4.2).

¹³⁸C'est ce même locuteur qui produit la voyelle de *head* avec un timbre plus antérieur que les autres locuteurs; ce détail illustre bien le fait que les voyelles fonctionnent comme un système.

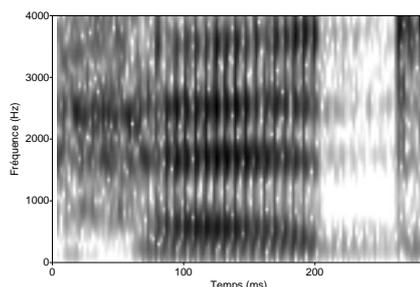


FIG. 6.19 – *head* produit par le locuteur *rkm* de *sse*.



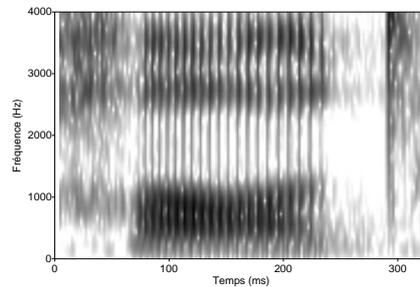
FIG. 6.20 – *had* produit par les locuteurs *rkm* et *prz* de *sse*.

La voyelle de *hard* est une monophtongue longue postérieure ouverte (Figure 6.21(a)). L’anglais standard, tel que nous l’avons déjà décrit dans la Section 3.1, n’est pas rhotique. Cependant, un des locuteurs de *sse* (*kmp*) produit systématiquement un [ɹ] dans *hard*. Et, sur ses cinq occurrences de *hard*, notre locuteur type a une réalisation rhotique (Figure 6.21(b)), ce qui illustre peut-être l’aspect artificiel des listes de mots à structure /hVd/¹³⁹.

La voyelle de *hod* est une monophtongue brève postérieure arrondie ouverte ou mi-ouverte (Figure 6.22). Notre analyse auditive ne fait apparaître que très peu de variation entre les locuteurs.

La voyelle de *hoard* est une monophtongue longue postérieure arrondie mi-fermée (Figure 6.23(a)). À l’instar de *hod*, notre impression auditive nous conduit à penser que

¹³⁹Il n’est pas non plus à exclure que cette réalisation rhotique illustre le fait que l’anglais standard est, pour certains locuteurs, un accent d’adoption.

FIG. 6.21 – *hard* par *rkm* ; variantes non rhotique et rhotique.FIG. 6.22 – *hod* produit par le locuteur *rkm* de *sse*.

cette voyelle varie peu. Comme ce fut le cas pour *hard*, le locuteur *kmp* produit un [ɪ] systématique, et le locuteur *rkm* a une seule réalisation rhotique (Figure 6.23(b)).

FIG. 6.23 – *hoard* par *rkm* ; variantes non rhotique et rhotique.

La voyelle de *hood*, telle qu'elle est produite par *rkm*, est une monophthongue brève postérieure mi-fermée : [ʊ] (Figure 6.24(a)). Cette voyelle varie d'un locuteur à l'autre.

Beaucoup de locuteurs ont une variante plus antérieure, de type [ø] (Figure 6.24(b)).



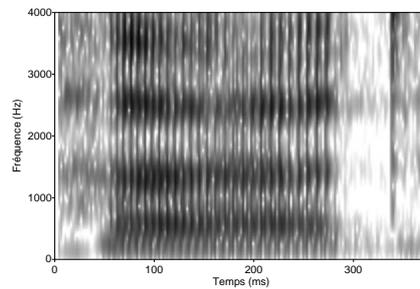
FIG. 6.24 – *hood* produit par les locuteurs *rkm* et *jph* de *sse*.

Conformément à ce que la description de l’anglais standard (Section 3.1) nous laissait espérer, les réalisations de *who’d* sont très variables entre les locuteurs. Le timbre est parfois très légèrement diphtongué et le degré d’antériorité varie considérablement. Le locuteur *rkm* a une réalisation plutôt conservatrice (voir Figure 6.25(a)), en cela qu’elle est relativement postérieure (cf. Figure 6.25(b), pour une variante antérieure).



FIG. 6.25 – *who’d* produit par les locuteurs *rkm* et *mjc* de *sse*.

La voyelle de *Hudd* est une monophthongue brève dont le timbre varie entre [ʌ], [ɜ] et [ɐ]. Le locuteur *rkm* n’a pas une variante très ouverte (Figure 6.26(a)). La Figure 6.26(b) illustre une prononciation plus ouverte ; les valeurs moyennes obtenues sur la partie stable de cette voyelle sont de 704 et 1200 Hz (cf. Table 6.3).

FIG. 6.26 – *Hudd* produit par les locuteurs *rkm* et *rws* de *sse*.FIG. 6.27 – *heard* produit par le locuteur *rkm* de *sse*.

La voyelle de *heard* est une longue monophthongue de type [ɜ] (Figure 6.27). Elle ne semble varier que faiblement entre locuteurs.

L'analyse des diphtongues de *rkm* s'est déroulée de la manière suivante. Comme les diphtongues fermantes tendent à faire diverger F1 et F2, et les diphtongues ouvrantes ont une tendance à faire converger ces formants, il paraît cohérent de mesurer :

- pour les diphtongues fermantes,
 - le point de départ en calculant le F1 maximum et le F2 minimum sur toute la durée de la voyelle ;
 - le point d'arrivée en calculant le F1 minimum et le F2 maximum ;
- pour les diphtongues ouvrantes,
 - le point de départ en calculant le F1 minimum et le F2 maximum ;
 - le point d'arrivée en calculant le F1 maximum et le F2 minimum.

À partir de notre segmentation manuelle des voyelles de *rkm*, nous avons dans un pre-

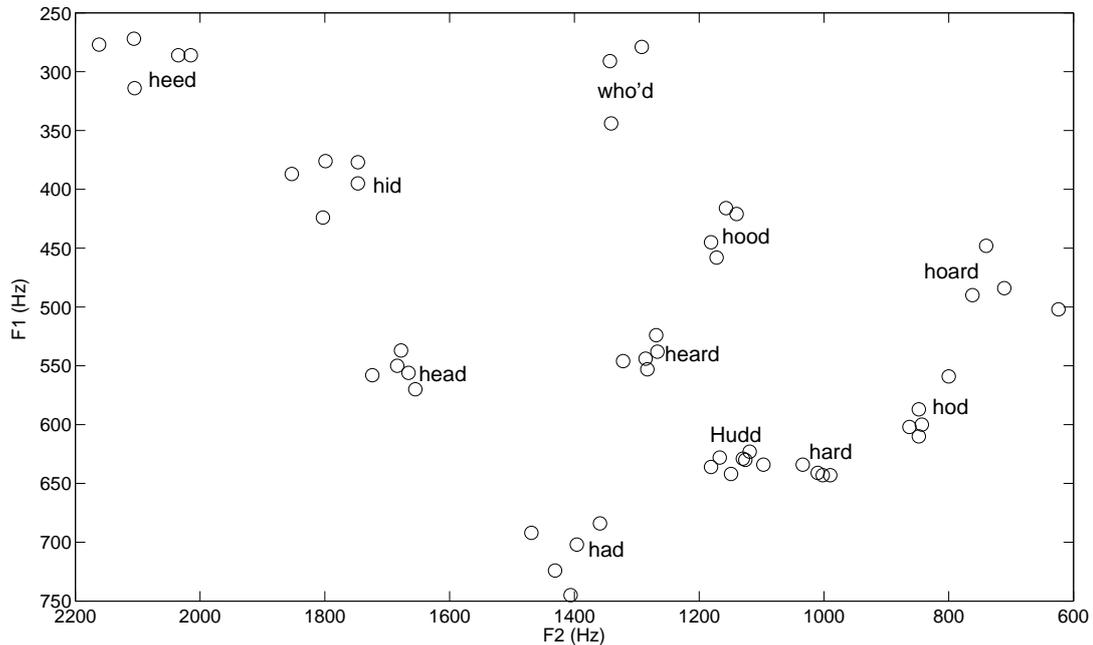
TAB. 6.3 – Formants des monophthongues de *rk*m (Hz); μ et σ représentent la moyenne et l'écart-type, respectivement.

Voyelle	Effectif	F1 μ	F2 μ	F1 σ	F2 σ
heed	5	287	2084	16	59
hid	5	392	1790	20	44
head	5	554	1681	12	26
had	5	710	1412	25	41
hard	4	640	1009	4	18
hod	5	592	840	20	24
hoard	4	481	709	23	60
hood	4	435	1163	20	18
who'd	3	305	1325	35	28
Hudd	7	632	1138	6	29
heard	5	541	1286	11	22

TAB. 6.4 – Formants des monophthongues de l'étude de Hawkins et Midgley ([HM05], moyenne par groupe d'âge en Hz).

Voyelle	65+		50-55		35-40		20-25	
	F1	F2	F1	F2	F1	F2	F1	F2
heed	285	2283	269	2355	269	2312	276	2338
hid	382	2024	341	2074	374	2115	393	2174
head	454	1962	489	1920	512	1888	600	1914
had	644	1678	693	1579	696	1574	917	1473
hard	665	1085	639	1041	608	1062	604	1040
hod	518	875	522	865	496	833	484	865
hoard	391	619	360	604	382	626	392	630
hood	376	990	371	975	381	984	413	1285
who'd	301	994	283	1112	288	1336	289	1616
hud	630	1213	643	1215	629	1160	658	1208
herd	475	1321	511	1340	497	1419	494	1373

mier temps inspecté tous les tracés des deux premiers formants afin de déterminer si les règles que nous venons d'énoncer pouvaient être appliquées. Constatant que c'était le cas dans une grande majorité des réalisations, nous avons calculé les points de départ et d'arrivée des diphtongues et avons retenu la médiane par type de voyelle dans la Figure 6.29 et

FIG. 6.28 – Monophthongues du locuteur *rk*m.

dans la Table 6.5. En réalité, cette stratégie n'est adéquate que pour les diphtongues dont le point d'arrivée se situe dans la zone de [i] ou [ɪ], puisque le locus du [d] final, approximativement égal à 1800 Hz, ne peut causer, dans le cas où le timbre est très antérieur, qu'une baisse de F2, ce qui n'affecte pas la valeur du maximum. En ce qui concerne les diphtongues dont le point d'arrivée est à proximité de [ʊ], [u] ou [ɯ], le [d] final occasionne une augmentation de F2 et, par conséquent, le calcul de la valeur maximale de ce formant est influencé à la hausse. Pour ces diphtongues-ci, nous avons déterminé empiriquement que la valeur de F2 du timbre d'arrivée se situait à environ 50 ms de la fin de la voyelle. Les diphtongues fermantes sont représentées dans l'espace vocalique conventionnel dans la Figure 6.29.

La voyelle de *hade* est une diphtongue fermante à mouvement formantique relativement restreint dont le point de départ se situe dans la zone de *head*, et le point d'arrivée est proche de *heed* (voir les valeurs formantiques des monophthongues dans la Table 6.3). Comme le montre la réalisation typique dans la Figure 6.30(a), la dynamique de cette

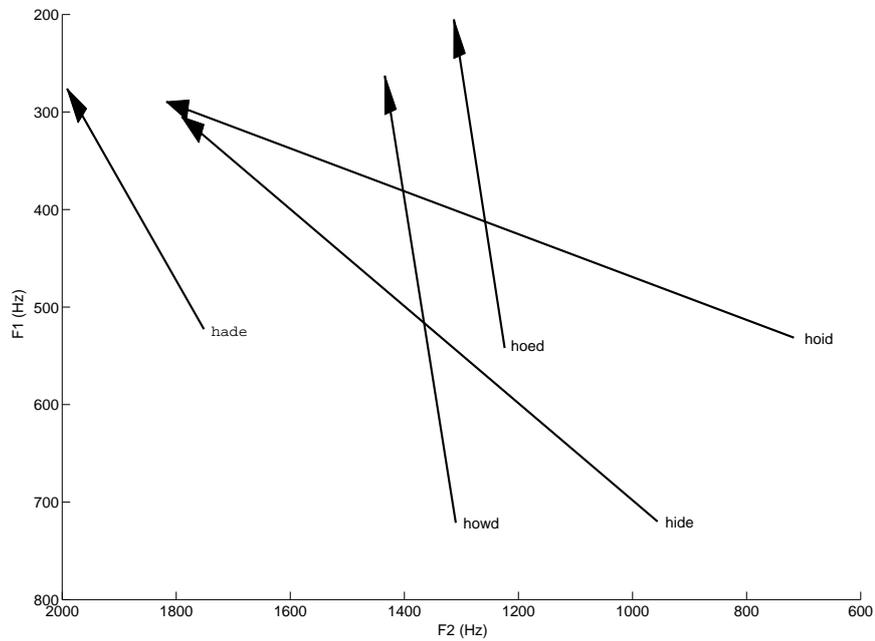


FIG. 6.29 – Diphtongues fermantes du locuteur *rk*m.

TAB. 6.5 – Médiane et écart interquartile (*iqr*) des diphtongues fermantes de *rk*m.

	Timbre initial				Timbre final			
	F1	F1 <i>iqr</i>	F2	F2 <i>iqr</i>	F1	F1 <i>iqr</i>	F2	F2 <i>iqr</i>
<i>hade</i>	523	42	1751	40	276	73	1992	23
<i>hide</i>	720	23	956	25	305	45	1790	67
<i>hoid</i>	531	21	717	74	290	37	1817	59
<i>hoed</i>	542	62	1224	32	205	80	1313	41
<i>howd</i>	721	27	1310	30	263	90	1434	162

voielle est caractérisée par un mouvement de formants continu qui semble se stabiliser dans la seconde partie de l'émission de la voyelle. On ne peut donc pas parler de deux voyelles liées par un glide, ni d'une voyelle stable suivie d'un glide, mais plutôt d'un glide qui tend vers une voyelle stable. Cette diphtongue ne varie que très peu parmi les locuteurs de *sse*. Un locuteur, *kmp*, produit une voyelle dont le caractère diphtongué est moins marqué que chez les autres sujets (Figure 6.30(b)).



FIG. 6.30 – *hade* produit par les locuteurs *rkm* et *kmp* de *sse*.

La voyelle de *hide* est une diphtongue fermante à mouvement formantique ample dont le timbre de départ s'approche de *hard*, et le timbre d'arrivée se situe aux alentours de *hid*. Dans la Figure 6.31, on constate que, contrairement à *hade*, cette diphtongue est probablement mieux décrite comme une voyelle suivie d'un glide, suivi d'une autre voyelle. En effet, on remarque sur ce spectrogramme que, schématiquement, les deux premiers formants sont relativement stables entre environ 100 et 200 ms ; puis on relève un mouvement divergent entre environ 200 et 300 ms (avec une intensité moindre), auquel succède une structure formantique plutôt stable. La variation entre locuteurs se limite à quelques différences de qualité du timbre de départ, qui reste néanmoins postérieur ou central (voir *supra* Section 6.3.1.2).

La voyelle de *hoid* est une diphtongue fermante à mouvement formantique ample dont le timbre de départ se situe aux environs de *hod* et *hoard*, et dont le timbre d'arrivée approche *heed*. Cette voyelle peut être décrite comme la succession d'un état stable, d'un

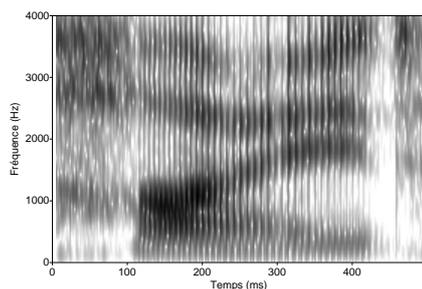
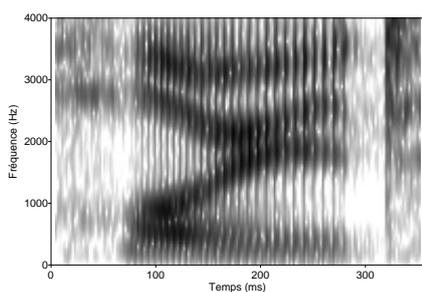
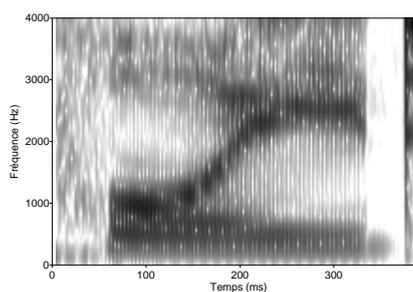


FIG. 6.31 – *hide* produit par le locuteur *rkm* de *sse*.

glide, et d'un autre état stable. La durée relative de ce dernier état stable, et l'impression auditive qui en découle, varie d'un locuteur à l'autre. Le locuteur *rkm* a un état stable final comparativement bref (Figure 6.32(a)). Dans la Figure 6.32(b), on constate que la locutrice *nta* produit un état stable final relativement long, qui renvoie l'impression auditive que nous avons affaire à deux syllabes. Ceci est peut-être imputable à un allongement anormal propre à ce type de mots-test. Comparativement à l'impression moyenne que dégagent les productions des locuteurs de *sse*, la voyelle de départ de *rkm* est légèrement fermée. On note en effet que le timbre de départ varie quelque peu entre les locuteurs. Ceci constitue, avec la taille relative variable du second état stable, le deuxième axe de variation entre locuteurs que nous avons pu observer.



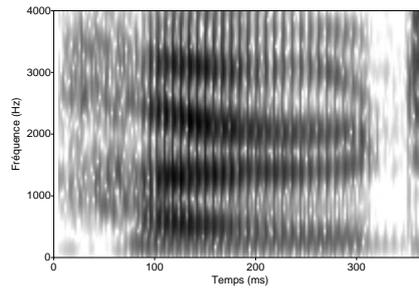
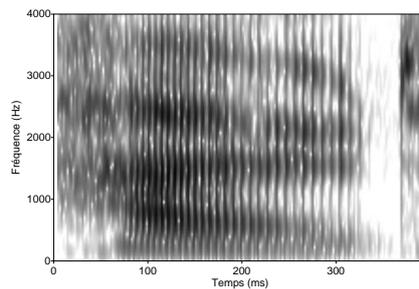
(a) *rkm*



(b) *nta*

FIG. 6.32 – *hoed* produit par le locuteur *rkm* et la locutrice *nta* de *sse*.

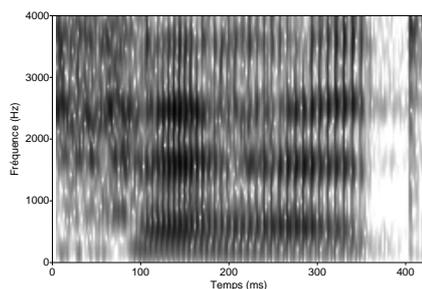
La voyelle de *hoed* est une diphtongue fermante à mouvement formantique relativement restreint dont le point de départ est localisé aux alentours de *heard*, et le point d'arrivée est

FIG. 6.33 – *hoed* produit par le locuteur *rkm* de *sse*.FIG. 6.34 – *howd* produit par le locuteur *rkm* de *sse*.

dans la zone de *who'd*, quoique nettement plus fermé (Figure 6.33). La variation que nous observons est due aux différences de timbre final, qui fait écho à la variation constatée pour les monophthongues *hood* et *who'd*. L'analyse auditive ne nous a pas permis de constater des différences flagrantes du timbre du premier élément, que ce soit dans le sens de [ɛ] (variante hypercorrecte ou d'un raffinement caricatural) ou dans celui de [o] (variante très conservatrice, voire archaïque). La prééminence du premier élément se traduit par une intensité plus élevée dans le premier tiers temporel de la voyelle.

La voyelle de *howd* est une diphtongue fermante dont l'ampleur des mouvements formantiques est intermédiaire. Son premier élément est proche de *had*, et le second se situe dans la zone de *who'd*. Comme c'était déjà le cas pour *hoed*, la variation est particulièrement visible au niveau du second élément. Le premier élément est antérieur ou légèrement centralisé (comme pour le locuteur *rkm* sur la Figure 6.34).

Contrairement au mode opératoire employé dans la description des monophthongues et des diphtongues fermantes, nous ne fournirons pas de trapèze vocalique représentant

FIG. 6.35 – *hared* produit par le locuteur *rkm* de *sse*.

les trajectoires de formants pour les voyelles que l'on range conventionnellement dans la catégorie des diphtongues ouvrantes, ou, plus précisément, centripètes. Et pour cause : une écoute attentive, doublée d'une analyse acoustique des voyelles de *hared*, *heered* et *hured*, nous fait douter de l'existence phonétique de diphtongues centripètes en anglais standard contemporain (au moins pour *hared* et *hured*).

Concernant *hared*, notre analyse auditive vient corroborer la description que donne Upton (Section 3.1.1) de la voyelle de SQUARE : la variante la plus fréquente est une monophthongue longue dans la zone de [e] ou [ɛ], et, plus rarement une monophthongue suivie d'un très léger glide vers le centre de l'espace vocalique. Trois des cinq occurrences du locuteur type *rkm* sont rhotiques (ainsi que toutes les réalisations du locuteur *kmp*). Dans la Figure 6.35, le spectrogramme d'une réalisation monophthongale de *hared* prononcé par *rkm* fait apparaître une structure formantique stable sur toute l'émission de la voyelle. Les fréquences moyennes de F1 et F2 relevées sur la première partie d'intensité élevée (ca. 100 à 150 ms), puis sur la seconde (ca. 160 à 240 ms) sont de 579 et 1591, et 565 et 1558 Hz ; il s'agit donc virtuellement de la même voyelle (donc, d'une monophthongue).

La voyelle de *heered* a des réalisations très variables sur le plan de la stabilité du timbre au cours de son émission. Ici, la variante diphtonguée, parfois très légèrement, constitue la norme (Figure 6.36(a)) ; mais les monophthongues sont très courantes (Figure 6.36(b)). Le locuteur type *rkm* produit une variante (sur les cinq qu'il prononce) avec un [ɪ]. Les valeurs médianes des deux premiers formants du premier élément et du deuxième élément des 4 occurrences non rhotiques de *heered* par *rkm* sont de 338 et 1916 Hz, et 426 et

1498 Hz, respectivement. Le premier élément a donc le degré de fermeture de *hid* et un degré d'antériorité proche de *heed*. Le second élément a un degré de fermeture proche de *hood* et un degré d'antériorité avoisinant celui de *had*.



FIG. 6.36 – *heered* produit par les locuteurs *rkm* et *ais* de *sse*.

Nous supposons que la voyelle de *hured* était censée illustrer l'ensemble CURE. Or il est évident après une analyse auditive des occurrences de ce mot-test, et après avoir constaté la très importante variation entre les sujets, et souvent pour un seul et même locuteur, que tous n'ont pas associé *hured* à l'ensemble CURE. Le locuteur *rkm* produit [jʊə] (voir Figure 6.37), ce qui correspond à la variante la plus conservatrice de CURE (Section 3.1.1). Mais nous doutons beaucoup du naturel de cette prononciation, eu égard au fait que d'après l'*OED* [OED07], <hure> (qu'il s'agisse de l'adverbe signifiant <au moins>, du nom désignant une casquette ou ayant le même sens que son homographe français) est obsolète. Pour corroborer l'aspect archaïque de ce terme, remarquons que le dictionnaire de prononciation de Jones ([Jon03]) ne le mentionne pas. Si l'on considère que *hured* fait effectivement partie de l'ensemble CURE, et que CURE (voir Section 3.1.1) a subi un changement phonétique presque achevé de nos jours¹⁴⁰, alors les hésitations des locuteurs proviennent peut-être de la résistance attestée des mots rares aux changements phonétiques réguliers ([Byb01], 11) : « phonetic change often progresses more quickly in items with high token frequency. » En outre, la variation entre les différentes prononciations (parfois d'un même locuteur) nous porte à croire que ce mot-test a engendré une

¹⁴⁰[ʊə] → [ɔ:].

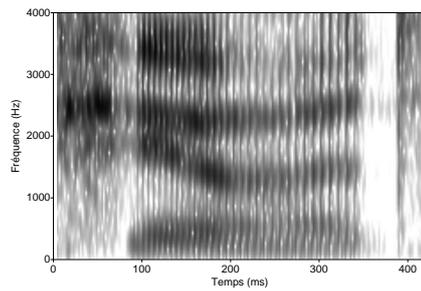


FIG. 6.37 – *hured* produit par le locuteur *rkm* de *sse*.

prononciation tout à fait artificielle. Notons, incidemment, l’amplitude relativement élevée d’une zone qui correspond au F3 de la voyelle lors de l’émission du /h/ : il s’agit très vraisemblablement de la friction créée par une constriction dans la zone palatale, phénomène courant en anglais lorsque [h] est suivi de [i] ou [j] (une description technique est donnée dans [Ste98], 441-445).

6.3.2 Birmingham (*brm*)

6.3.2.1 Les monophthongues

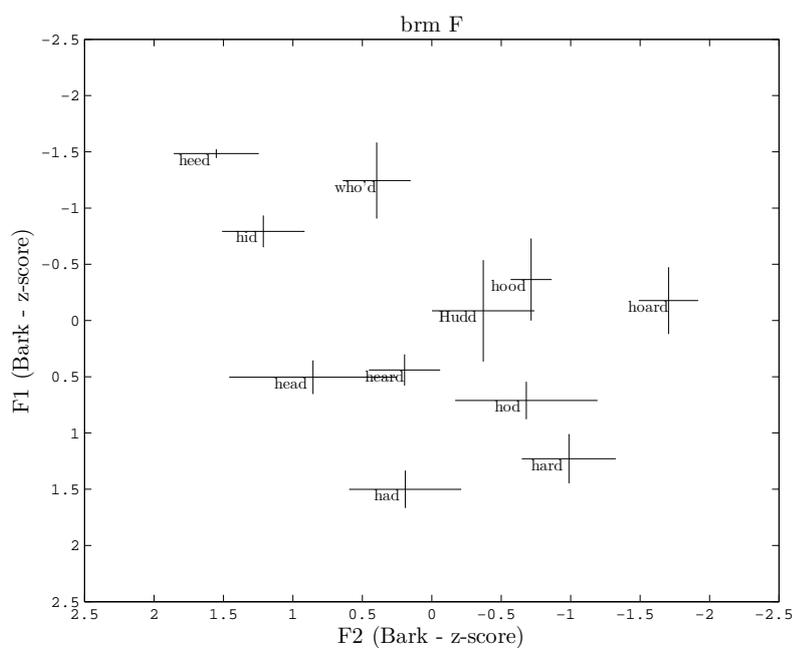


FIG. 6.38 – Monophthongues de *brm* pour les sujets féminins.

Dans la Figure 6.38, 8 locutrices ont produit l'ensemble des 11 voyelles. Leur système semble se distinguer de celui de *sse* sur les points suivants :

- la relative proximité de *head* et *heard* sur l'axe F1 ;
- la proximité, sur F1, de *hard* et *hod*, qui semble être due à une réalisation plus ouverte de *hod* ;
- la réalisation plutôt postérieure de *hood* ;
- la proximité de *Hudd* et *hood* causée par une réalisation avec un degré de fermeture élevé pour *Hudd*. Cette proximité des deux timbres peut nous laisser supposer une absence de scission FOOT-STRUT, au moins chez certains locuteurs ;
- la proximité de *heed* et *hid* ;

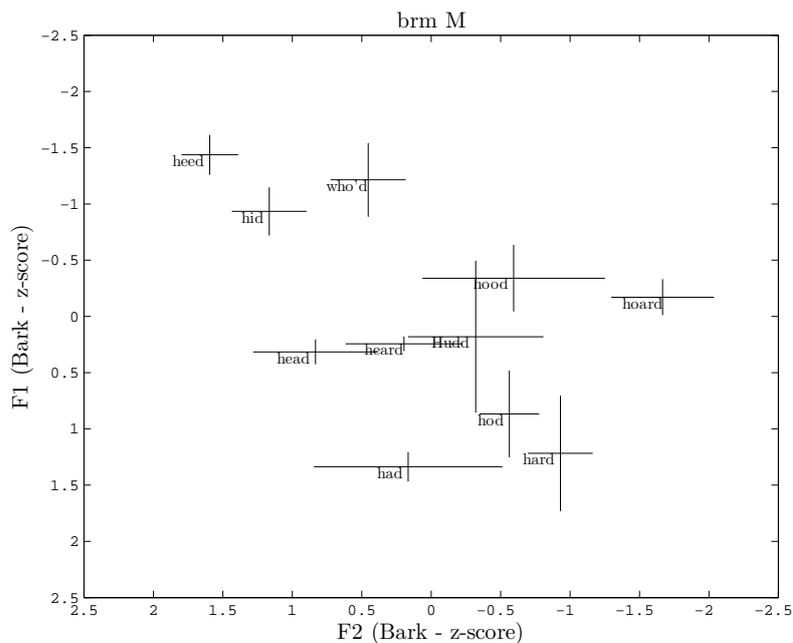


FIG. 6.39 – Monophthongues de *brm* pour les sujets masculins.

Dans la Figure 6.39, ce sont 7 locuteurs qui ont produit l'ensemble des 11 voyelles. Les remarques sont les mêmes que pour la Figure 6.38, auxquelles on peut ajouter un chevauchement accru entre *hard* et *hod* et un télescopage de *heard* et *Hudd*. Si les timbres de ces deux paires sont très proches, on peut néanmoins présumer, à ce stade descriptif de l'étude, que, au vu de la différence de durée entre chacun des membres d'une paire (voir

Table 6.6), l'opposition de quantité permet de les dissocier. Une analyse auditive ne nous laisse percevoir aucune rhoticité.

À l'écoute, les timbres de *hard* et *hod* apparaissent très proches, voire strictement identiques. Si l'on se réfère à la Table 6.6, il semble très vraisemblable que la durée soit le paramètre le plus fiable permettant de séparer les réalisations prototypiques des deux phonèmes. La Figure 6.40(a) représente le spectrogramme d'une occurrence typique de *hard* produite par le locuteur *ajh*. La valeur moyenne de F1 et F2 calculée sur la partie la plus stable de la voyelle est de 612 et 1102 Hz. Le mot-clé *hod*, produit par le même locuteur, est visible dans la Figure 6.40(b) ; les valeurs des deux premiers formants sont de 657 et 1109 Hz. Seule la différence sur F1 est légèrement supérieure au seuil différentiel de perception ; mais elle n'est très probablement pas suffisante pour coder une différence de catégorie phonémique. Il est donc vraisemblable que seule la durée permette de distinguer ces deux sons chez ce locuteur.

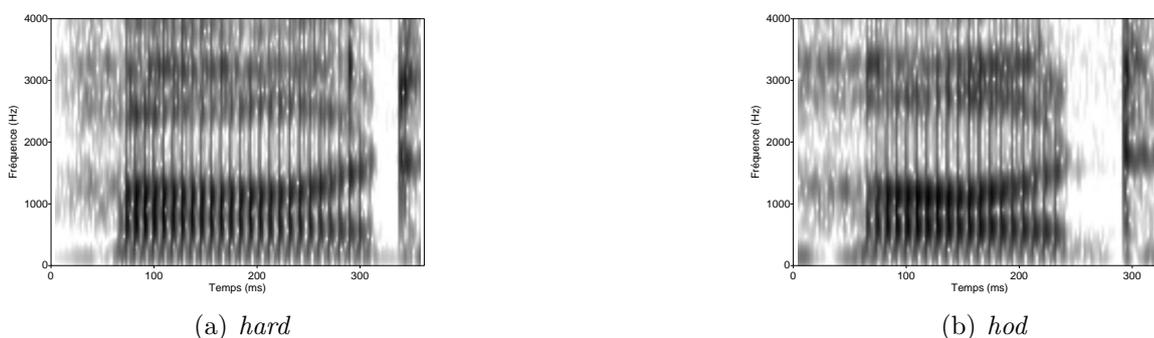
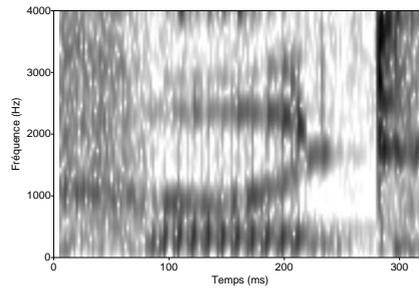


FIG. 6.40 – *hard* et *hod* produits par le locuteur *ajh* de *brm*.

La proximité de *hood* et *Hudd* n'est pas aussi marquée que, par exemple, pour *eyk* (Figures 6.67 et 6.68). Il est donc probable que certains systèmes individuels aient la scission FOOT-STRUT, et d'autres non. Cette variation systémique est révélée par notre analyse auditive : sur les 20 locuteurs et locutrices qui produisent au moins une occurrence de chacun des deux items qui nous concernent, la moitié présente une parfaite homophonie¹⁴¹.

¹⁴¹On note au passage que ceci illustre bien le biais induit par le calcul de valeurs moyennes sur tous les locuteurs.

FIG. 6.41 – *hood* produit par le locuteur *ikp* de *brm*.

L'autre moitié produit généralement une variante plutôt fermée pour *Hudd* (par rapport à *sse*), la réalisation de *hood* étant relativement postérieure. On peut, par exemple, comparer le *hood* du locuteur *ikp* (Figure 6.41, fréquence moyenne sur partie stable de F1 et F2 : 305 et 845 Hz) avec la réalisation de *hood* par le locuteur *rkm* de *sse* (Figure 6.24(a) et Table 6.3).

Le chevauchement partiel de *heard* et *Hudd* s'explique en partie par la scission FOOT-STRUT chez certains locuteurs. Si, comme nous venons de le voir, les locuteurs ayant une homophonie entre *Hudd* et *hood* produisent la première voyelle avec un timbre postérieur et semi-fermé, ceux qui ont cette distinction dans le système possèdent une réalisation de *Hudd* proche de *heard* au point où, chez certains, les deux timbres donnent l'impression d'être confondus. Les Figures 6.42(a) et 6.42(b) représentent les spectrogrammes d'une occurrence de *heard* et de *Hudd* chez le locuteur *axm*. Si la différence de durée est très nette, on remarque que les timbres sont proches (F1 et F2 : 593 et 1554 Hz pour *heard* et 548 et 1473 Hz pour *Hudd*).

6.3.2.2 Les diphtongues

Les réalisations de 8 locutrices et de 7 locuteurs sont représentées dans les Figures 6.43 et 6.44. Visuellement, l'ampleur des mouvements formantiques des hommes en référence à l'espace de leurs monophthongues paraît plus restreinte que chez les femmes. Une autre distinction entre les sexes réside dans le timbre de départ de *hoed* ; il est plus postérieur en moyenne chez les hommes. Seule une locutrice (*rkk*) présente une réalisation clairement



FIG. 6.42 – *heard* et *Hudd* produits par le locuteur *axm* de *brm*.

TAB. 6.6 – Durée des monophthongues de *brm* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	263	68
hid	163	56
head	175	62
had	193	52
hard	294	60
hod	181	44
hoard	294	67
hood	163	50
who'd	281	60
Hudd	163	50
heard	281	69

postérieure, les autres ayant une valeur centrale proche de celle de *sse*. Les hommes varient davantage pour ce timbre ; on peut, par exemple, comparer la réalisation du premier élément de *hoed* pour deux locuteurs : dans la Figure 6.45(a), la réalisation est relativement postérieure (F1=576 Hz, F2=1168 Hz) ; elle est nettement plus antérieure (et donc centrale) dans la Figure 6.45(b) (F1=590 Hz, F2=1512 Hz).

On constate également que *hoed* et *howd*, contrairement à leurs réalisations en *sse*, ont, pour les hommes comme pour les femmes, un élément final approximativement identique. La qualité du timbre initial de *hide* et *howd* est caractérisée par une différence plus importante qu'en *sse*. La qualité postérieure et quelque peu fermée du premier élément de

hide est remarquable à l'écoute des mots-test prononcés par la plupart des locuteurs. À titre d'exemple, les voyelles de *hide* et *howd* produites par le locuteur *apd* sont représentées dans les Figures 6.46(a) et 6.46(b) ; les F1 et F2 moyens calculés sur la partie stable du début de chaque voyelle sont de 612 et 1129 Hz, et 580 et 1695 Hz, respectivement.

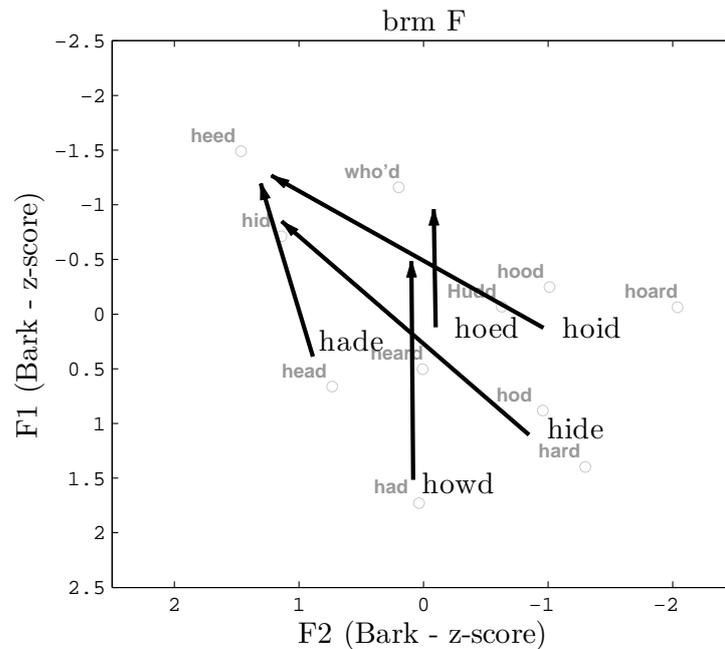


FIG. 6.43 – Diphtongues de *brm* pour les sujets féminins.

Nous avons mentionné (3.2.1.3) une possible convergence PRICE-CHOICE en *brm*. Les Figures 6.43 et 6.44 ne laissent aucunement apparaître un tel phénomène, mais force est de reconnaître que la distance perçue entre les deux timbres est parfois très faible. Compte tenu du fait que les mots à structure /hVd/ de notre étude conduisent probablement les sujets à maximiser les contrastes entre les réalisations de deux phonèmes adjacents dans le système, il est envisageable que la manifestation acoustique de cette opposition devienne moins robuste dans une situation de communication ordinaire. La différence subtile entre ces réalisations pourrait alors être perçue comme le reflet d'une convergence PRICE-CHOICE par des locuteurs d'autres variétés (qui, eux, ont un codage acoustique plus robuste de ce contraste).

La voyelle de *hared* est exclusivement produite comme une longue monophthongue.

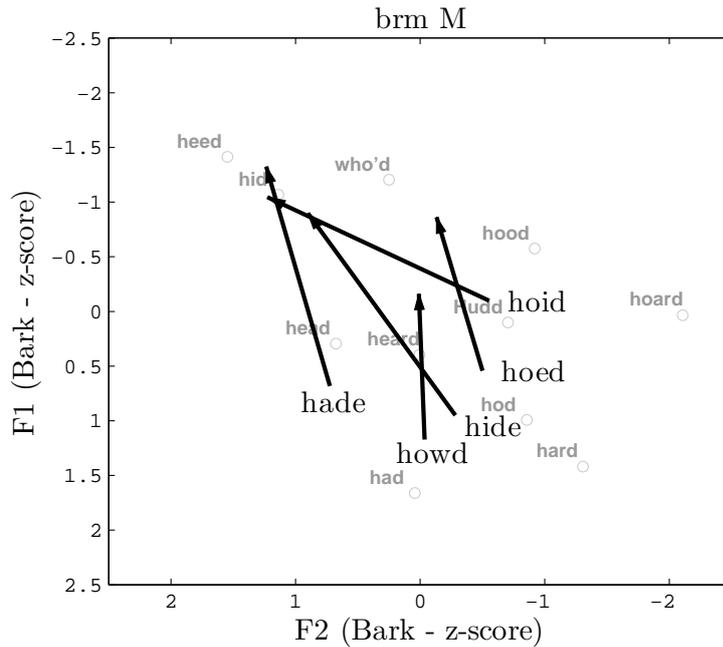


FIG. 6.44 – Diphtongues de *brm* pour les sujets masculins.



FIG. 6.45 – *hoed* produit par les locuteurs *apd* et *axm* de *brm*.

Notre impression auditive nous incite à penser que cette voyelle est une version longue du timbre de *head*. La Figure 6.47 tend à confirmer cette hypothèse. Chacun des trois graphes représente l'estimation de la densité de probabilité¹⁴² associée à *hared* (ligne continue) et

¹⁴²Sans entrer dans les détails, ce type de représentation peut être interprété comme un histogramme lissé. Les valeurs sur l'axe des ordonnées ne sont pas comparables d'un graphe à l'autre en raison de leur dépendance vis à vis des ordres de grandeur variables d'un paramètre à l'autre. Seules la forme des densités (aplatissement, asymétrie, nombre de modes) et l'ampleur du chevauchement entre les deux courbes de chaque graphe sont analysées ici. Ces estimations ont été réalisées avec la fonction `ksdensity` du logiciel Matlab. Une introduction très abordable sur les estimations de la densité par la méthode du noyau est consultable dans [ELL01], 16-23.

FIG. 6.46 – *hide* et *howd* produits par le locuteurs *apd* de *brm*.

head (ligne discontinue) en fonction des trois paramètres mesurés (F1, F2 et durée) chez les hommes. En réalité, afin d'écartier les valeurs déviantes dues aux erreurs de mesure, les densités de probabilité ont été estimées à partir de l'empan borné par le 10^e et le 90^e centile de chaque variable (paramètre)¹⁴³.

On constate le très important chevauchement entre les valeurs formantiques des deux voyelles. La valeur de F1 pour *head* est nettement plus variable que celle de *hared* ; l'origine de ce phénomène nous est inconnue. Après examen des courbes représentant F1 et F2, il est possible de conclure que, dans la majorité des cas, *hared* et *head* ont une timbre identique. Les estimations de densité de probabilité pour la durée font apparaître que ce paramètre semble particulièrement robuste pour différencier les deux voyelles. Les valeurs moyennes de F1, F2 et de durée sont de 573 Hz, 1941 Hz et 175 ms pour *hared*, et 533 Hz, 1994 Hz et 114 ms pour *head*.

La voyelle de *heered* est, dans la majorité des cas, une diphtongue centripète dont le timbre de départ se situe dans la zone de [i]. Certaines réalisations sont monophtongales. Si l'évolution fait tendre les diphtongues centripètes vers une monophtongaison, phénomène attesté récemment en *sse* pour la voyelle de SQUARE, alors la voyelle de NEAR pourrait bien être analysée comme le pendant phonologiquement long de KIT. En effet, certaines réalisations de *heered* et *hid* par le locuteur *jxs* font apparaître que les deux timbres ne peuvent être distingués que par la durée (Figure 6.48). Les valeurs moyennes de F1 et F2

¹⁴³Autrement dit, l'écart interdécile.

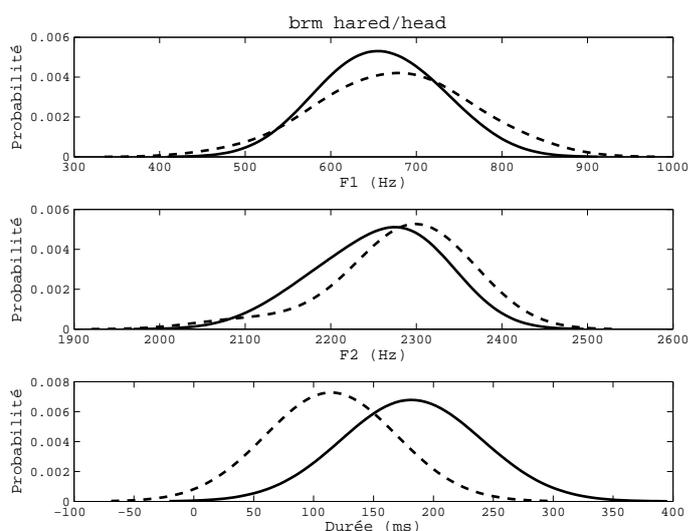


FIG. 6.47 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *hared* (ligne continue) et *head* (ligne discontinue) produites par les locuteurs masculins de *brm*.

pour la voyelle *heered* sont de 467 et 1888 Hz, et de 447 et 1927 Hz pour *hid*.



FIG. 6.48 – Spectrogrammes de *heered* et *hid* produits par le locuteur *jxs*.

La voyelle de *hured* a une nouvelle fois un nombre de variantes trop important, qui rend inutile le calcul d'une moyenne englobant tous les locuteurs. Les réalisations sont les suivantes : [hjøɾd], [hɜɾd], [hjød] et [hjuød]. En raison de cette variation considérable, et compte tenu du fait que les facteurs susceptibles de l'expliquer ne nous sont pas connus (i.e. aucune information individuelle disponible), nous ne pousserons pas plus avant l'analyse de ce timbre.

6.3.3 Cornwall (*crn*)

6.3.3.1 Les monophthongues

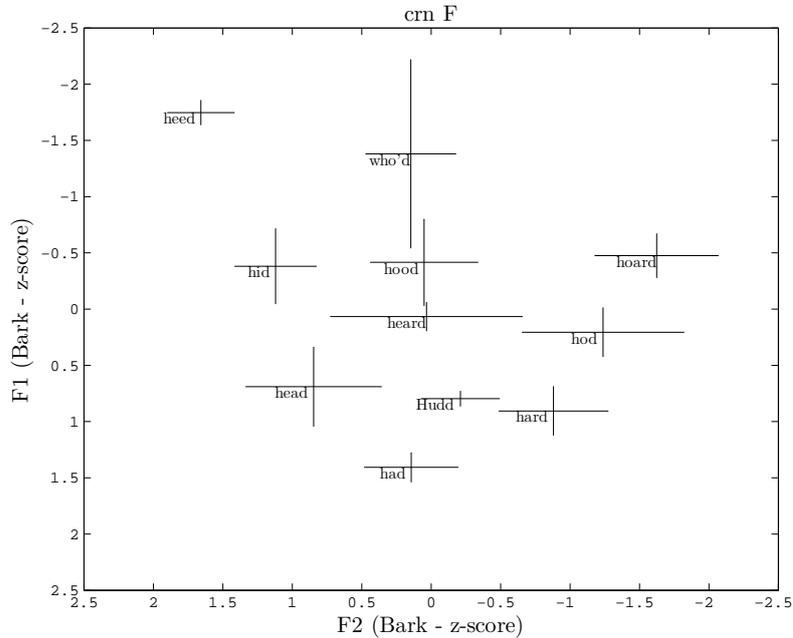


FIG. 6.49 – Monophthongues de *crn* pour les sujets féminins.

Dans la Figure 6.49, 6 locutrices ont produit l'ensemble des 11 voyelles. En ce qui concerne le système, nous sommes, semble-t-il, une nouvelle fois en présence de 11 phonèmes. Ce diagramme se distingue de celui de *sse* principalement sur le point suivant : *who'd* a une réalisation plutôt postérieure, et l'étendue de l'écart interquartile sur F1 rend nécessaire une analyse individuelle.

Dans la Figure 6.50, ce sont 2 locuteurs qui ont produit l'ensemble des 11 voyelles. La dispersion des valeurs de *hoard*, notamment sur F2, appelle la même remarque que pour *who'd* dans la Figure 6.49. Contrairement à *brm*, *crn* est indiscutablement rhotique, ce qui signifie que les <r> graphiques de nos mots-test ont une équivalent acoustique. Tâchons de déterminer jusqu'à quel point la rhoticité peut affecter nos mesures. La Figure 6.51(a) représente un spectrogramme de *hard* produit par un locuteur de *crn*. On constate que la fréquence de F3 baisse considérablement entre le début et la fin de la voyelle. On note également que F1 et F2 s'écartent légèrement à partir d'environ 400 ms.

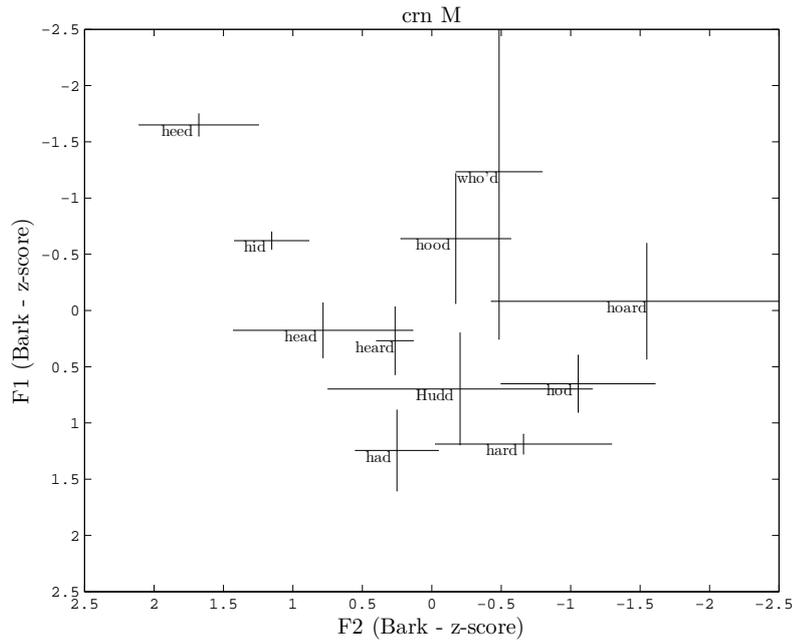


FIG. 6.50 – Monophthongues de *crn* pour les sujets masculins.

Dans la Figure 6.51(b), le même mot-test est produit par un locuteur de *sse*. On relève le même type de divergence de F1 et F2 vers la fin de la voyelle ; en revanche, F3 reste stable. Nous concluons donc que le mouvement des deux premiers formants est dû à la coarticulation avec le [d], et que la rhoticité en *crn* affecte principalement F3. La comparaison des Figures 6.52(a) et 6.52(b) appelle les mêmes conclusions : en *crn*, le premier tiers (voire la première moitié) de la voyelle est caractérisé par une structure formantique relativement stable ; la portion restante se singularise par un abaissement très marqué de F3. Il apparaît cependant que, pour certaines voyelles, la réalisation de /ɜr/ ne soit pas séquentielle, mais bien simultanée. En effet, dans la Figure 6.53(a), le même locuteur de *crn* produit *heard*. Or, si à l'écoute le caractère « rhotique » de ce mot est évident, sa réalisation diffère sensiblement de ce que nous avons pu constater dans les figures précédentes : on note une structure formantique stable sur toute la durée de la voyelle. En comparant ce tracé avec celui de la Figure 6.53(b), où le mot *heard* est produit par un locuteur de *sse*, on remarque, certes, que F1 et F2 n'ont pas le même écartement, mais surtout que, dans la Figure 6.53(a), F2 et F3 sont particulièrement proches. Dans la Figure 6.53(c), la

prononciation de *Hudd* par le locuteur de *crn* fait apparaître, pour un timbre (défini par F1 et F2) proche de *heard*, un F3 quasiment identique à celui du locuteur de *sse* (Figure 6.53(b)). Nous concluons donc que, comme nous l’avons dit plus haut, le corrélat acoustique principal de la rhoticité en *crn* est l’abaissement de F3 et, étant donnée la stabilité de ce formant dans la Figure 6.53(a), la voyelle et le /r/ de *heard* sont réalisés simultanément, on transcrirait donc [hɜ̃d] plutôt que [hɜ̃ɹd]. Une analyse alternative consisterait à considérer le /r/ comme syllabique : [hɹ̩d]. Incidemment, les Figures 6.51(a) à 6.52(b) illustrent le cas de voyelles ayant des valeurs de F1 et F2 très proches (voyelles compactes) ; c’est typiquement dans ce contexte que les algorithmes d’estimation de formants commettent des erreurs.

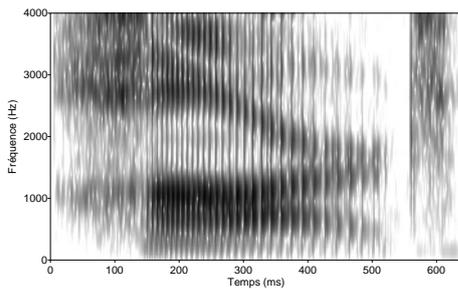
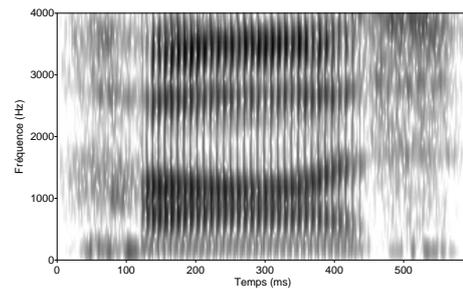
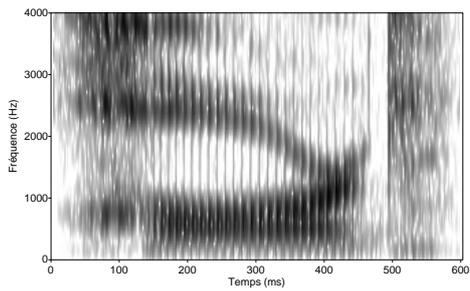
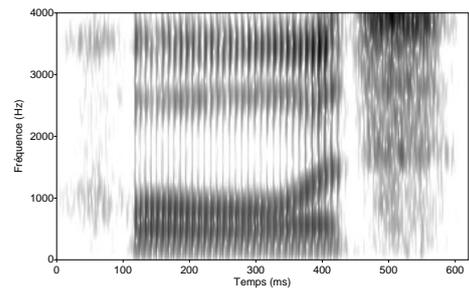
(a) locuteur *gmc* de *crn*(b) locuteur *jph* de *sse*

FIG. 6.51 – Spectrogrammes de *hard* produit par un locuteur de *crn* et un locuteur de *sse*.

La variabilité de *who'd* est flagrante à l’écoute. Selon le locuteur, la réalisation est plus ou moins fermée, plus ou moins antérieure, et parfois diphtonguée. Dans la Figure 6.54(a) une réalisation postérieure est attestée par la proximité de F1 et F2 dans les fréquences basses du spectrogramme. Au contraire, dans la Figure 6.54(b), la proximité de F2 et F3 confirme que ce locuteur-ci produit un timbre antérieur. Enfin, les mouvements de formants dans la Figure 6.54(c) illustrent que ce troisième locuteur a une variante légèrement diphtonguée.

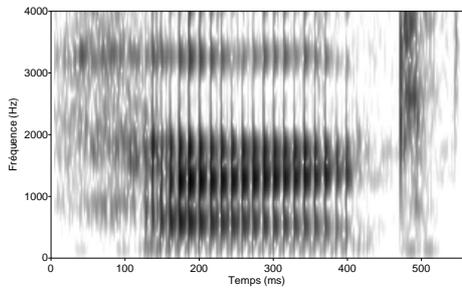


(a) variante rhotique, locuteur *gmc* de *crn*

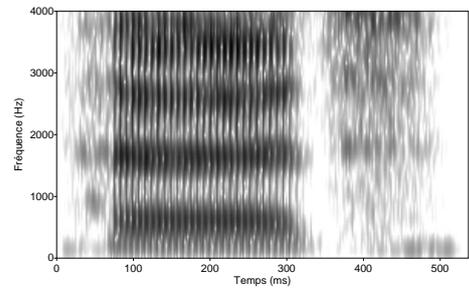


(b) variante non rhotique, locuteur *jph* de *sse*

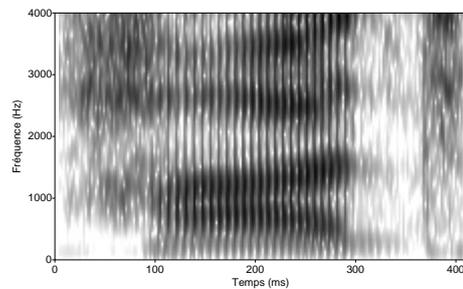
FIG. 6.52 – Spectrogrammes de *hoard* produit par un locuteur de *crn* et un locuteur de *sse*.



(a) *heard*, locuteur *gmc* de *crn*



(b) *heard*, locuteur *jph* de *sse*

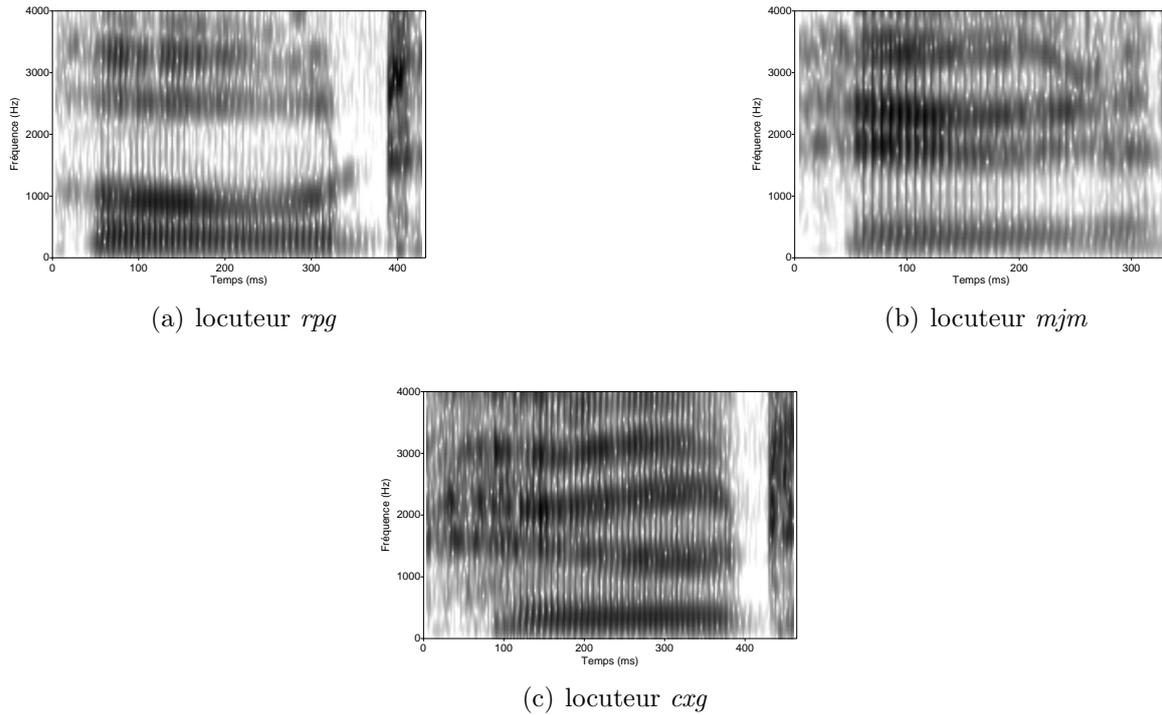


(c) *Hudd*, locuteur *gmc* de *crn*

FIG. 6.53 – Spectrogrammes de *heard* et *Hudd* produits par un locuteur de *crn* et un locuteur de *sse*.

6.3.3.2 Les diphtongues

Les Figures 6.55 et 6.56 montrent les diphtongues de *crn* produites par 6 locutrices et 2 locuteurs respectivement. L'ampleur du mouvement formatique sur *hade* est moins

FIG. 6.54 – Spectrogrammes de trois variantes de *who'd*.TAB. 6.7 – Durée des monophthongues de *crn* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	256	98
hid	175	71
head	203	95
had	213	106
hard	278	100
hod	194	92
hoard	263	65
hood	175	57
who'd	263	95
Hudd	181	75
heard	275	89

importante chez les femmes : son origine se situe entre *head* et *hid* chez ces dernières, alors qu'il correspond à *head* pour les hommes (pour un point d'arrivée identique relativement à

l'espace des monophthongues par sexe). Si *hide* et *howd* présentent une différence identique à celle de *sse* sur l'axe de F2, l'écart de degré d'aperture est plus important. À l'écoute, le premier élément de *howd* est effectivement plus fermé que son équivalent en *sse* ; le timbre de départ varie cependant beaucoup.

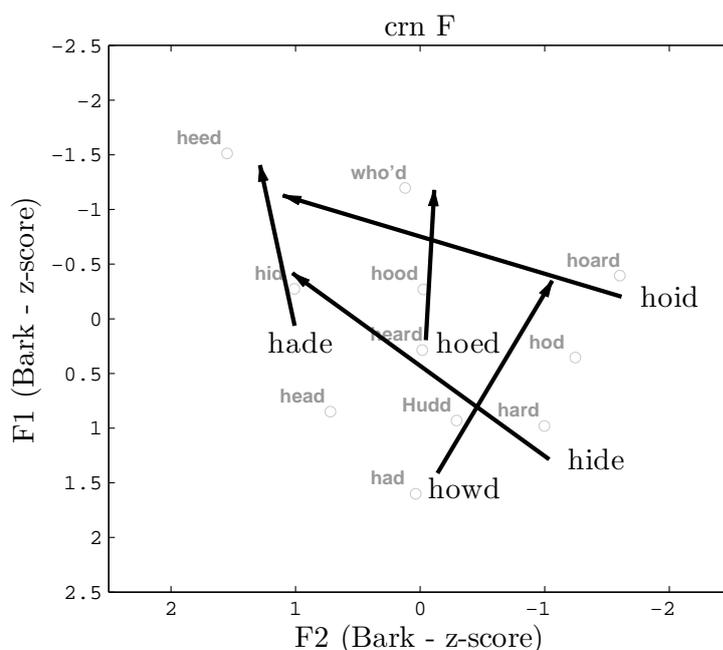
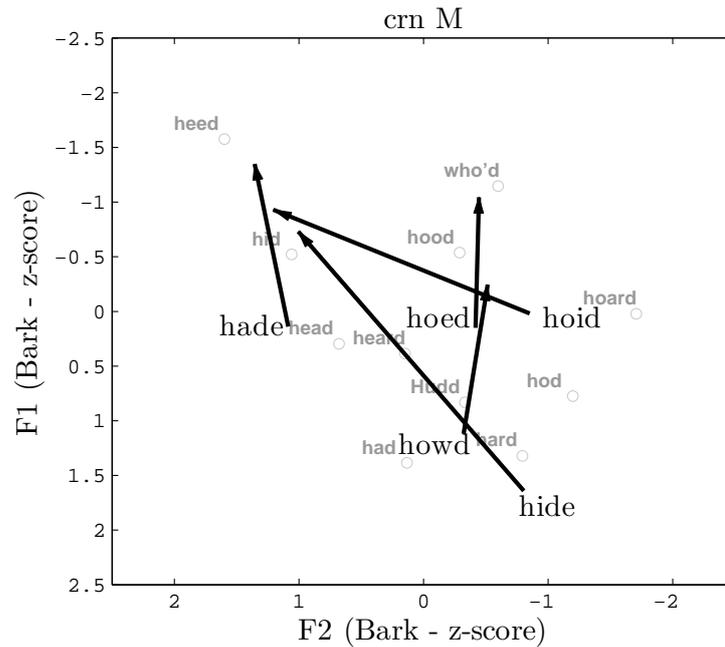
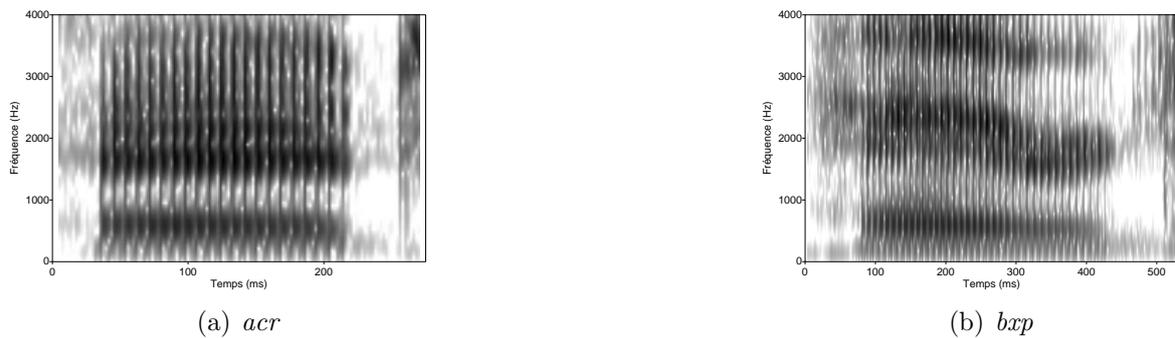


FIG. 6.55 – Diphtongues de *crn* pour les sujets féminins.

La voyelle de *hared* varie d'un locuteur à l'autre entre une réalisation monophthongale de type [ɛ] parfois colorée par un [ɛ^ɻ], et une monophthongue suivie d'une approximante : [ɛɪ]. Ces deux variantes sont illustrées dans la Figure 6.57.

Les réalisations de *heered* sont homogènes : il s'agit, dans la grande majorité des cas, de [iɪ]. La rhoticité a des degrés de saillance variables ; par exemple, le locuteur *acr* produit des réalisations qui, sur le plan auditif, s'approchent d'une diphtongue centripète [iə].

Le mot-test *hured* a deux variantes : [hɜ^ɻd], qui est la plus fréquente, et [hjuɪd]. Pour la première, il est raisonnable de considérer qu'il y a eu coalescence entre la voyelle et le /r/. Comme le montre la Figure 6.58, le F3 relativement bas et la stabilité du schéma formantique sur toute la durée d'émission confirment que cette voyelle est stable sur le plan articulaire, il s'agit donc de [ɜ^ɻ] plutôt que [ɜɪ] (cf. Figures 6.51(a) et 6.52(a) pour

FIG. 6.56 – Diphtongues de *crn* pour les sujets masculins.FIG. 6.57 – Spectrogrammes de *hared* produits par les locuteurs *acr* et *bxp* de *crn*.

des spectrogrammes montrant la succession d'une monophongue et d'un [ɪ]).

6.3.4 East Anglia (*ean*)

6.3.4.1 Les monophongues

Dans la Figure 6.59, 8 locutrices ont produit l'ensemble des 11 voyelles. Le système de *ean* n'est pas très éloigné de celui de *sse*. Dans la Figure 6.60, ce sont 7 locuteurs qui ont

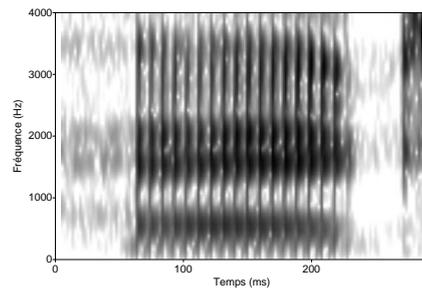


FIG. 6.58 – *hured* produit par le locuteur *acr* de *crn*.

produit l'ensemble des 11 voyelles. Ici encore, la dispersion importante de certains timbres rend nécessaire une analyse individuelle des locuteurs.

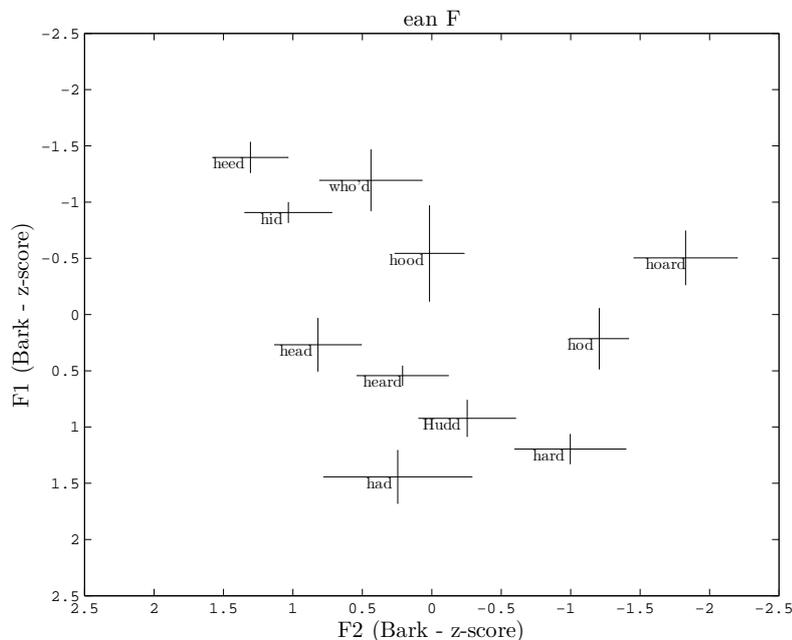


FIG. 6.59 – Monophtongues de *ean* pour les sujets féminins.

L'élément le plus flagrant qui distingue les femmes des hommes pour *ean* se situe incontestablement au niveau de la dispersion de chaque timbre : les écarts interquartiles de la Figure 6.59 sont à peu près équivalents entre les différentes voyelles, et surtout, ils ne font apparaître aucun chevauchement. À l'inverse, la dispersion autour de chaque voyelle dans la Figure 6.60 varie considérablement d'un timbre à l'autre, et certains chevauchements apparaissent. La voyelle de *heed* a parfois un timbre diphtongué comme l'illustre la

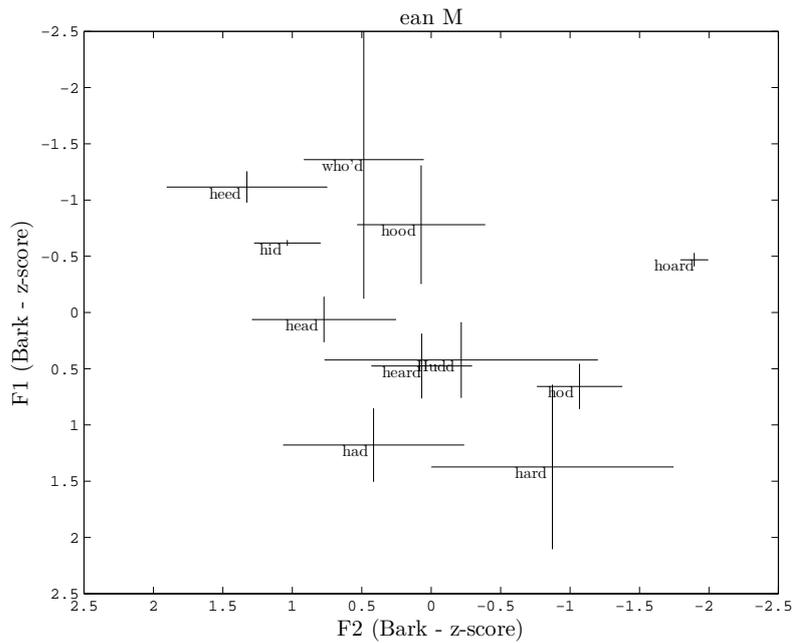
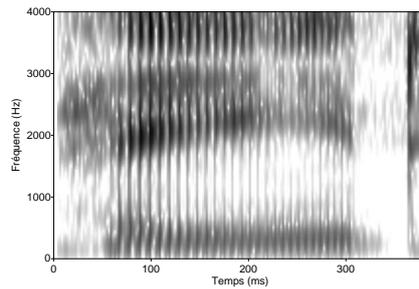
FIG. 6.60 – Monophthongues de *ean* pour les sujets masculins.FIG. 6.61 – *heed* produit par le locuteur *ifp* de *ean*.

Figure 6.61.

La Figure 6.60 montre un chevauchement important entre *Hudd* et *heard*. Ceci est confirmé dans la Figure 6.62 (*Hudd* : ligne continue ; *heard* : ligne discontinue). On y voit virtuellement une superposition des valeurs de F1 et F2 pour les deux voyelles étudiées. La densité de probabilité de F2 pour la voyelle de *Hudd* fait apparaître une bimodalité qui signale la présence probable de deux types de réalisation de *Hudd*. Les courbes de durée se chevauchent dans une grande proportion, mais on note une tendance de la voyelle de *heard* à être plus longue. Les valeurs moyennes de F1, F2 et de durée pour *Hudd* et *heard* sont, respectivement, 561 Hz, 1517 Hz et 108 ms, et 557 Hz, 1562 Hz et 185 ms. On comparera à

la figure précédente la Figure 6.63 illustrant les représentations des sujets féminins de *ean*. Sur cette figure, la différence de compacité¹⁴⁴ entre les deux timbres est flagrante. Les valeurs moyennes des paramètres sont de 750 Hz, 1601 Hz et 115 ms pour *Hudd*, et de 619 Hz, 1834 Hz et 209 ms pour *heard*. À ce stade, nous pouvons émettre deux hypothèses concernant les locuteurs masculins :

- soit les locuteurs n'utilisent pas les indices de durée et de compacité dans les mêmes proportions, i.e. certains utilisent davantage la compacité au détriment de la durée et vice-versa ; il se pourrait que, dans l'implémentation phonétique du contraste, ces indices acoustiques soient mutuellement exclusifs : les locuteurs choisissent soit la durée, soit la compacité ;
- soit une partie des locuteurs distinguent ces deux timbres (quel que soit le paramètre) et l'autre, non.

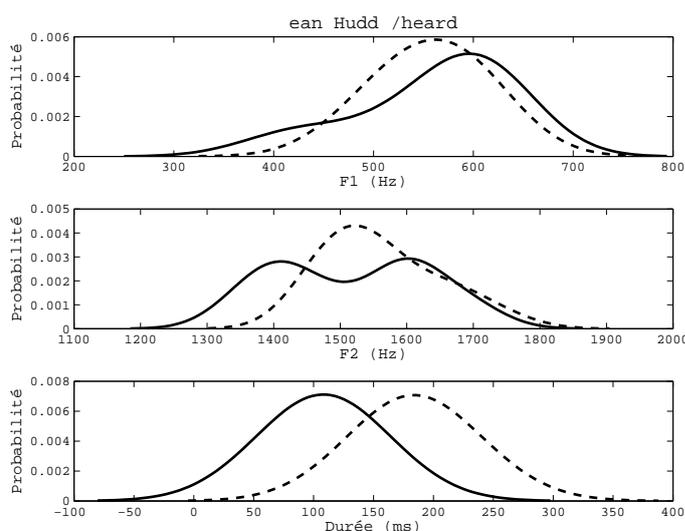


FIG. 6.62 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *Hudd* (ligne continue) et *heard* (ligne discontinue) produites par les locuteurs de *ean*.

À l'écoute, pour chaque locuteur et locutrice, il est toujours possible de distinguer *Hudd* de *heard*. Si le timbre des deux voyelles est parfois très proches, une différence de durée est toujours perceptible. Nous retiendrons donc la première hypothèse formulée plus haut : aucun locuteur n'a une véritable homophonie ; mais alors que les femmes codent

¹⁴⁴Dans ce chapitre, ce terme fait référence à la proximité de F1 et F2.

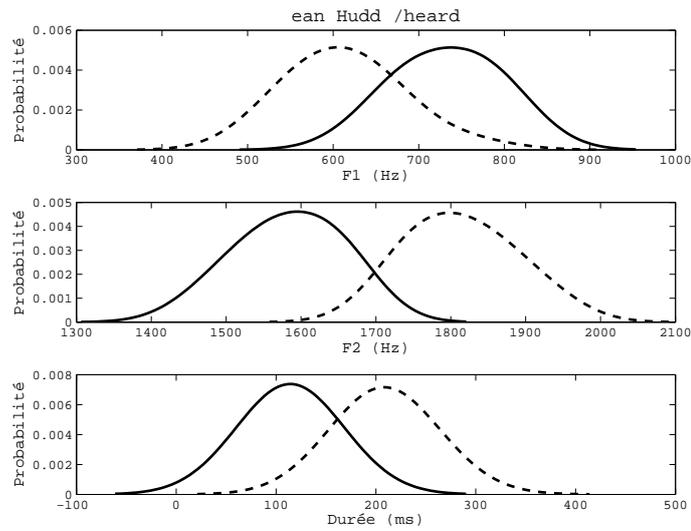


FIG. 6.63 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *Hudd* (ligne continue) et *heard* (ligne discontinue) produites par les locutrices de *ean*.

TAB. 6.8 – Durée des monophthongues de *ean* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	300	75
hid	175	50
head	194	56
had	222	85
hard	325	97
hod	213	81
hoard	325	105
hood	181	58
who'd	325	66
Hudd	181	53
heard	313	82

le contraste via et la durée et la compacité, les hommes ont probablement des stratégies individuelles favorisant tantôt l'un, tantôt l'autre paramètre. En tout état de cause, la redondance acoustique dans la distinction entre *Hudd* et *heard* chez ces derniers est plus faible.

6.3.4.2 Les diphtongues

4 locutrices et 5 locuteurs ont produit les diphtongues représentées dans les Figures 6.64 et 6.65. Conformément à nos attentes, la voyelle de *hade* a une trajectoire plus étendue que dans les autres dialectes présentant une diphtongue fermante. La voyelle de *hide* a un premier élément plus fermé qu'en *sse* ; ceci est particulièrement visible chez les hommes. Quelques rares locuteurs ont cependant un point de départ plus ouvert et légèrement plus antérieur (*nxp* et *asp*). Comme c'est le cas dans d'autres dialectes, les hommes ont un premier élément plus postérieur que les femmes dans *hoed*.

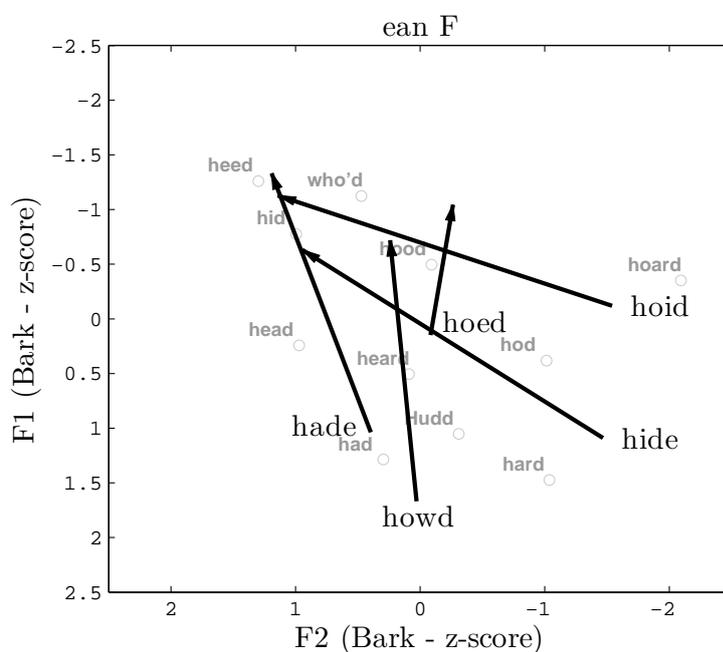
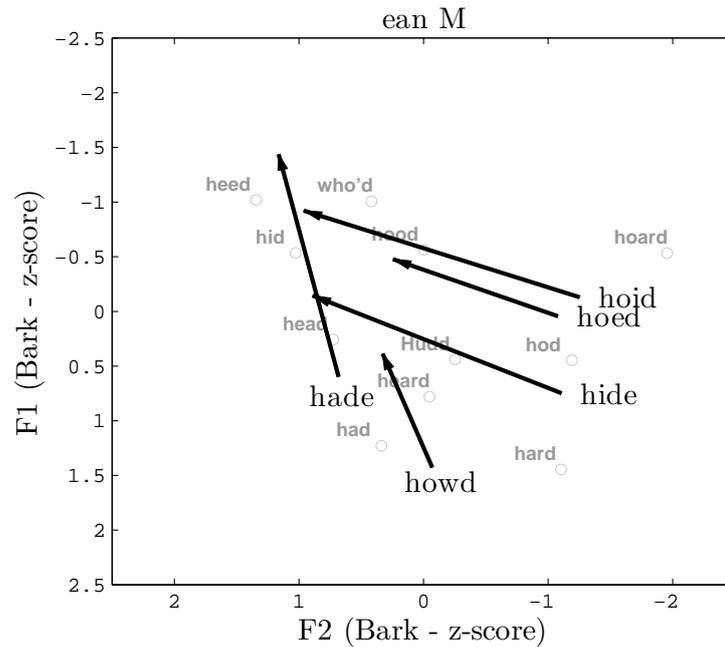


FIG. 6.64 – Diphtongues de *ean* pour les sujets féminins.

Si les Figures 6.64 et 6.65 font apparaître que, en moyenne, les points de départ de *hide* et *hoid* se situent dans la zone de [ɔ] et [o], respectivement, celles-ci masquent des faits individuels d'un grand intérêt. La locutrice *ald* a, semble-t-il, mais pas systématiquement, une homophonie entre ces deux mots-test¹⁴⁵.

La voyelle de *hared* suit le même schéma que son équivalent dans le dialecte *brm* ; il s'agit d'une réalisation longue de la voyelle de *head*. Les valeurs moyennes de F1, F2 et de

¹⁴⁵Voir notre allusion à la convergence historique PRICE-CHOICE dans la Section 2.2.2.

FIG. 6.65 – Diphtongues de *ean* pour les sujets masculins.

durée sont de 562 Hz, 1947 Hz et 169 ms pour *hared*, et 526 Hz, 1988 Hz et 105 ms pour *head*.

La quasi-totalité des réalisations de *heered* sont monophthongales ; par conséquent, ce timbre ne se distingue de *heed* que par son degré d'aperture (parfois tout juste) supérieur. Ceci est amplement confirmé dans la Figure 6.66. On y constate en effet que les densités de probabilité associées aux formants divergent entre *heered* (ligne continue) et *heed* (ligne discontinue). La voyelle de *heered* est plus compacte, avec un F1 et un F2 moyens de 444 Hz et 2053 Hz (contre 356 et 2277 Hz pour *heed*). Les durées, quant à elles, sont virtuellement équivalentes (174 et 180 ms pour *heered* et *heed*, respectivement).

La grande majorité des réalisations de *hured* présentent une longue monophthongue identique, sur le plan auditif, à la voyelle de *heard*.

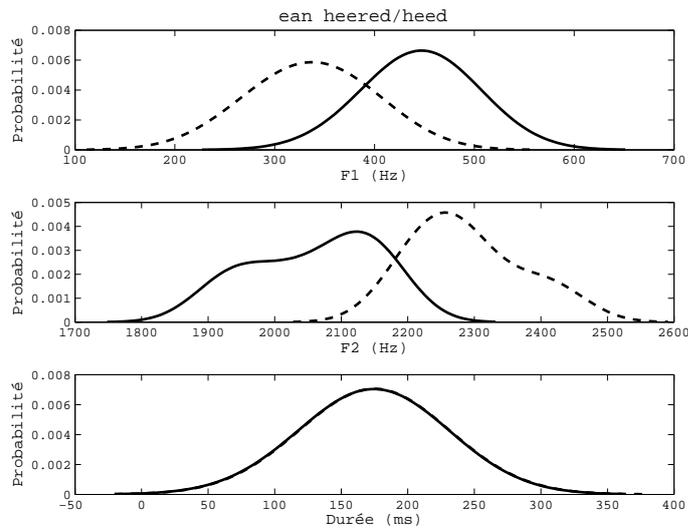


FIG. 6.66 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *heered* (ligne continue) et *heed* (ligne discontinue) produites par les locuteurs masculins de *ean*.

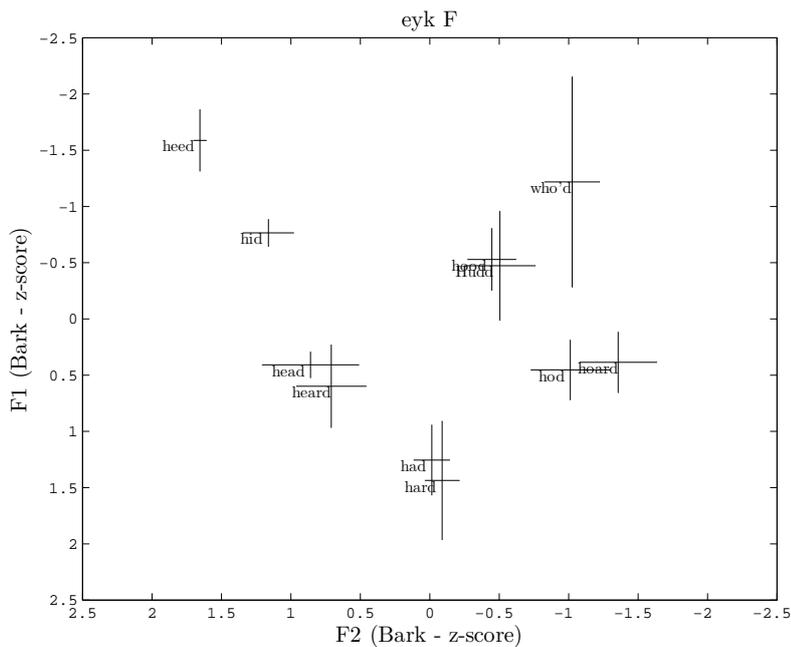
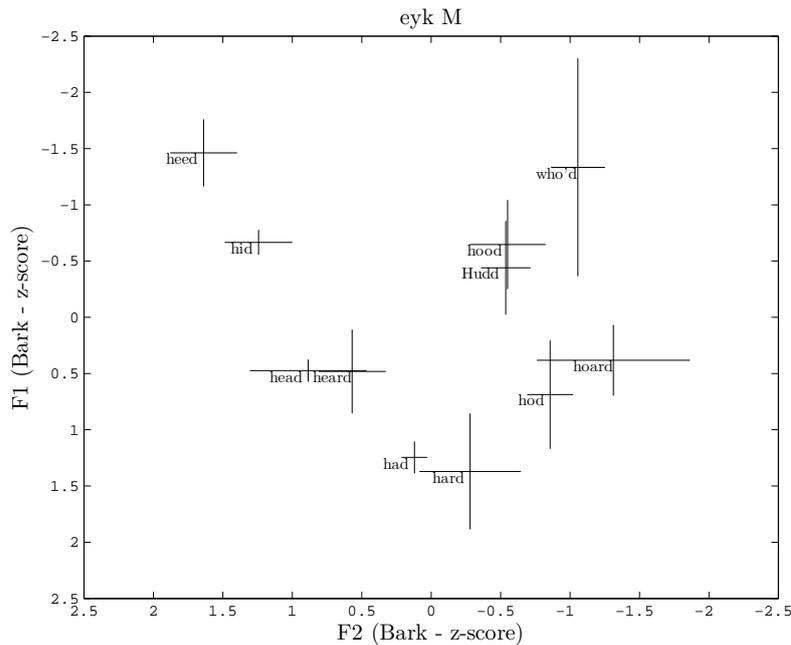


FIG. 6.67 – Monophtongues de *eyk* pour les sujets féminins.

6.3.5 East Yorkshire (*eyk*)

6.3.5.1 Les monophtongues

Dans la Figure 6.67, 8 locutrices ont produit l'ensemble des 11 voyelles. Nous avons ici affaire au premier système sensiblement éloigné de celui de *sse*. On y remarque :

FIG. 6.68 – Monophthongs de *eyk* pour les sujets masculins.

- la proximité relative de *head* et *heard* ;
- la proximité relative de *had* et *hard* ;
- le chevauchement presque parfait de *Hudd* et *hood* ;
- les qualités postérieures de *Hudd*, *hood* et *who'd*.

Dans la Figure 6.68, ce sont 8 locuteurs qui ont produit l'ensemble des 11 voyelles. Les remarques ayant trait à la Figure 6.67 s'appliquent également ici.

La proximité de *heard* et *head* peut s'expliquer comme suit. Des deux timbres, c'est *heard* qui a une réalisation très différente de *sse*. Une homophonie entre *head* et *heard* (i.e. une convergence entre DRESS et NURSE) est très improbable du fait de la différence de quantité qui oppose ces voyelles. En revanche, le timbre antérieur de *heard* pourrait bien traduire la convergence NURSE-SQUARE, que nous avons mentionnée en dans le Chapitre 3, avec une réalisation proche de [ɛ].

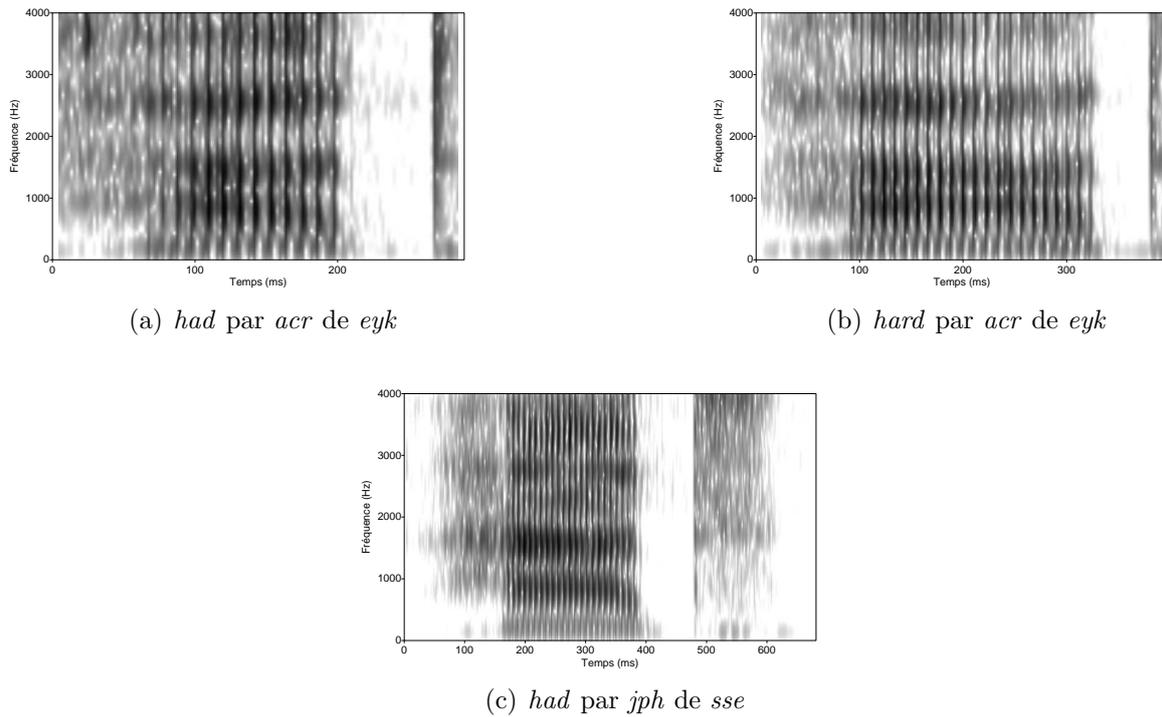
La proximité de *had* et *hard* est imputable, à notre avis, au fait que la réalisation typique de la voyelle de TRAP dans les accent septentrionaux est plus postérieure qu'en anglais standard (une question connexe a été abordée dans la Section 3.1.1). Une analyse

TAB. 6.9 – Durée des monophthongues de *eyk* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	256	75
hid	147	62
head	166	72
had	175	70
hard	281	100
hod	175	69
hoard	294	84
hood	163	55
who'd	275	73
Hudd	144	44
heard	281	100

auditive montre que chez certains locuteurs la voyelle de *hard* ne semble être qu'une version longue de celle de *had* ; chez d'autres, la distinction sur l'axe antérieur/postérieur est bien présente, mais beaucoup moins marquée que dans d'autres dialectes. Les Figures 6.69(a) et 6.69(b) représentent les mots *had* et *hard* produits par le locuteur *acr* de *eyk*. Les valeurs des deux premiers formants sont à peu près identiques, et les deux voyelles représentées sont très semblables (sinon identiques) sur le plan de la perception. Elles se distinguent en revanche par une différence de durée d'environ 100 ms ; il est donc vraisemblable que, chez ce locuteur, l'opposition TRAP-START ne soit visible en surface qu'à travers le paramètre de durée. Dans la Figure 6.69(c), le locuteur *jph* de *sse* produit la voyelle de *had* ; on remarque que F2 est comparativement plus élevé que pour le locuteur de *eyk* (Figure 6.69(a)). Et en mettant en parallèle *had* et *hard* produits par le locuteur de *sse* (respectivement Figure 6.69(c) et Figure 6.51(b)) on remarque la distinction spectrale qui fait défaut au locuteur de *eyk* : F1 et F2 sont plus espacés pour *had*.

Le fait que *Hudd* et *hood* soient proches reflète l'absence de scission FOOT-STRUT caractéristique des accents du nord de l'Angleterre.

FIG. 6.69 – Spectrogrammes de *had* et *hard* par un locuteur de *eyk* et un locuteur de *sse*.

6.3.5.2 Les diphtongues

Dans les Figures 6.70 et 6.71, ce sont 7 locutrices et 6 locuteurs qui ont produit l'ensemble des voyelles. Les voyelles de *hade* et *hoed* ont été considérées comme des monophtongues suite à l'analyse auditive. Pour la première, la réalisation alternative est une diphtongue fermante du type de celle de *sse* ; on la rencontre chez 5 des 25 sujets (avant analyse des formants). La voyelle de *hoed* est, quant à elle, beaucoup plus variable. Il peut s'agir d'une diphtongue fermante (Figure 6.72(a)), d'une monophtongue centrale (Figure 6.72(b)), ou encore d'une monophtongue postérieure semi-fermée (Figure 6.72(c)), voire complètement fermée. On peut par exemple, pour les deux monophtongues, comparer les valeurs moyennes de F2 sur la partie maximale stable¹⁴⁶ : 1379 Hz pour la voyelle représentée dans la Figure 6.72(b), contre 923 pour la voyelle de la Figure 6.72(c).

La voyelle de *hide* donne une impression auditive de trajectoire formantique plus res-

¹⁴⁶F1 étant virtuellement identique (524 et 538 Hz) pour ces deux voyelles.

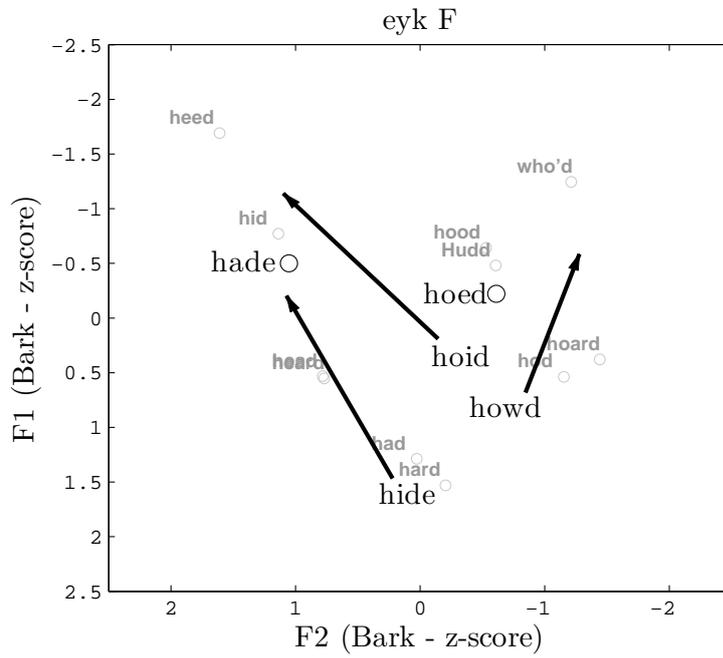


FIG. 6.70 – Diphtongues de *eyk* pour les sujets féminins.

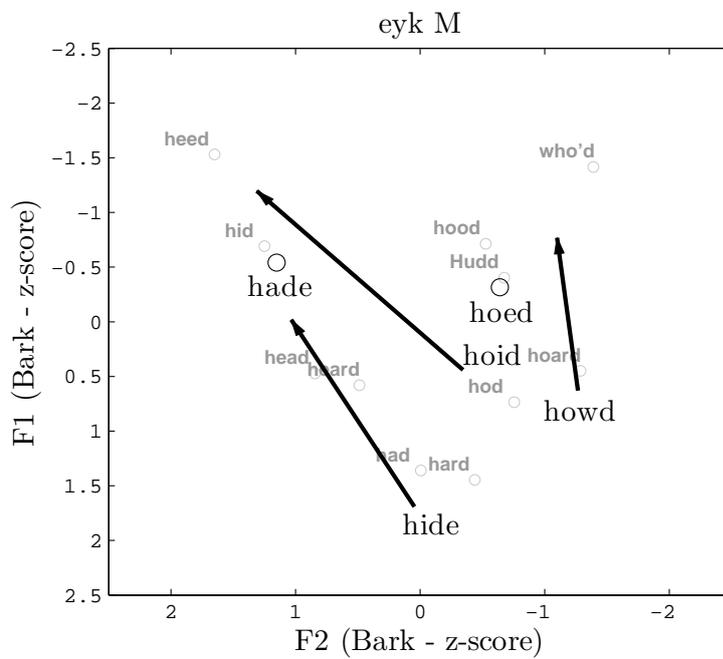


FIG. 6.71 – Diphtongues de *eyk* pour les sujets masculins.

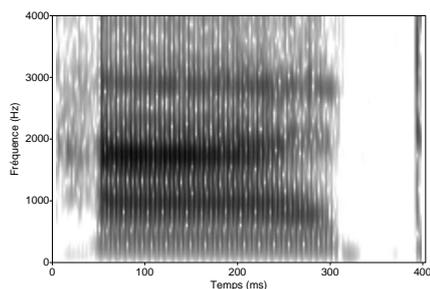


FIG. 6.73 – *hide* produit par la locutrice *cet* de *eyk*.

rhotiques (e.g. locutrice *wxr*) ou encore un long état stable suivi d'un très léger offglide. La Figure 6.74 représente l'estimation de la densité de probabilité de F1, F2 et la durée pour *hared* (ligne continue) et *head* (ligne discontinue) produits par les hommes. Les valeurs moyennes de F1, F2 et de durée sont de 562 Hz, 1737 Hz et 158 ms pour *hared*, et 537 Hz, 1868 Hz et 95 ms pour *head*. Par rapport à la Figure 6.47, le schéma est différent ; en effet, les courbes de F2 font apparaître plusieurs modes ; ceci reflète en partie la variation de timbre inter-individuelle évidente à l'écoute.

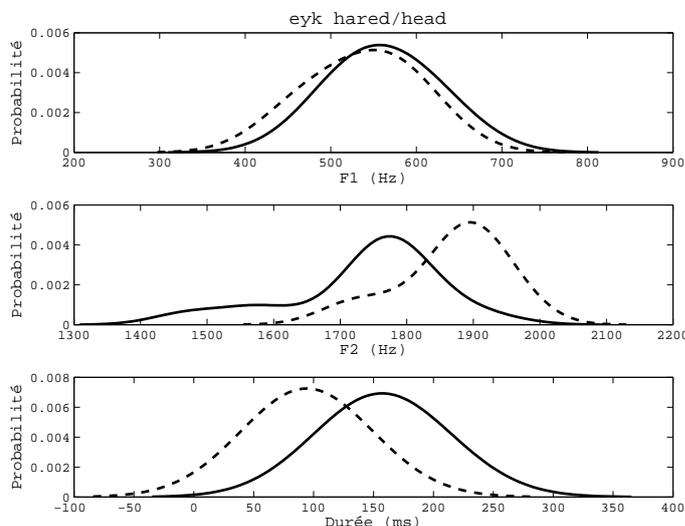
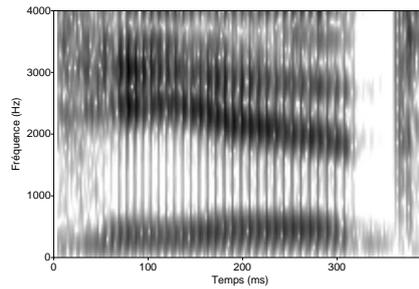
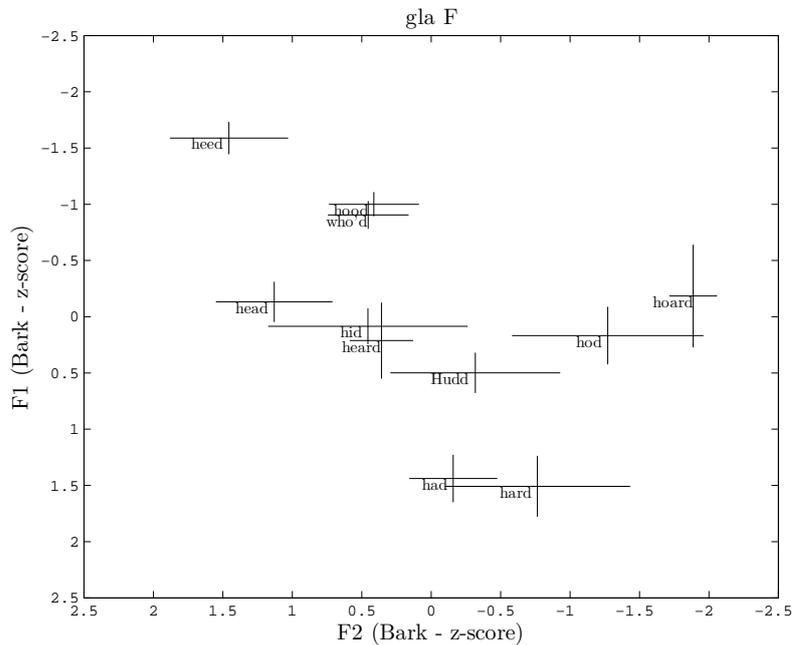


FIG. 6.74 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *hared* (ligne continue) et *head* (ligne discontinue) produites par les locuteurs masculins de *eyk*.

La voyelle de *heered* est une diphtongue centripète dont certaines réalisations tendent vers une monophtongaison. La Figure 6.75 illustre la variante la plus fréquente.

La voyelle de *hured* est trop variable ([jʊə], [ɜ], [jo], etc.) pour être décrite par le biais

FIG. 6.75 – *heered* produit par le locuteur *ccw* de *eyk*.FIG. 6.76 – Monophtongues de *gla* pour les sujets féminins.

d'une seule figure, ou même d'une seule paire de valeurs formantiques. Cette variation témoigne peut-être encore une fois de la difficulté des sujets à interpréter phonologiquement ce mot très rare.

6.3.6 Glasgow (*gla*)

6.3.6.1 Les monophtongues

Dans la Figure 6.76, 10 locutrices ont produit l'ensemble des 11 voyelles. Les principales différences avec le système de *sse* sont :

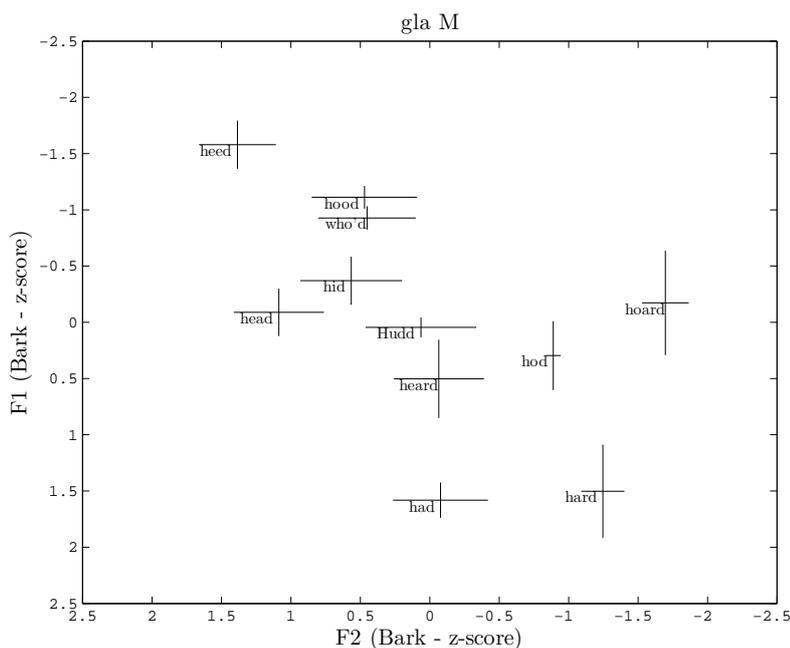


FIG. 6.77 – Monophthongues de *gla* pour les sujets masculins.

- la qualité nettement plus postérieure et ouverte de *hid* ;
- la proximité de *hid* et *heard* ;
- la dispersion importante de *hard* sur F2 ;
- le chevauchement de *hood* et *who'd* ;

Dans la Figure 6.77, ce sont 7 locuteurs qui ont produit l'ensemble des 11 voyelles. Les mêmes constatations que dans la Figure 6.76 s'appliquent hormis le fait que *hard* n'est pas particulièrement dispersé. On note également que *Hudd* est plus antérieur et plus fermé que dans la Figure 6.76. Ensuite, malgré leur proximité, *hood* et *who'd* ne se chevauchent pas dans la Figure 6.77. Notons encore que les positions relatives de *Hudd* et *heard* sont inversées dans les deux figures (sur l'axe F1). Enfin, si les dispersions de *had* et *hard* se chevauchent partiellement chez les femmes, elles sont en revanche entièrement distinctes chez les hommes, et cette remarque s'applique également à *hod* et *hoard*.

À l'écoute des locuteurs de *gla*, le trait le plus saillant provient d'une apparente absence de distinction de durée. Les informations de la Table 6.10 confirment en partie cette impression ; par exemple, *heed* et *hid* ont respectivement 144 ms et 150 ms, ce qui,

en pratique, correspond à une durée perçue égale. Le constat peut s'étendre aux paires *had/hard*, *hod/hoard*, mais pas à *hood/who'd*, ni à *Hudd/heard*.

Le timbre de *hid*, particulièrement ouvert et postérieur chez les femmes, nous a d'abord fait croire à des mesures erronées. Une analyse auditive confirme cependant que chez certains locuteurs, *hid* est plus ouvert et postérieur que *head*. La Figure 6.78(a) représente un spectrogramme de *hid* produit par le locuteur *gty* de *gla*. La Figure 6.78(b) montre un spectrogramme de *head* prononcé par ce même locuteur. La voyelle de *head* est incontestablement plus fermée et plus antérieure que celle de *hid*. Les voyelles représentées dans les Figures 6.78(a) et 6.78(b) ont une durée d'environ 90 et 200 ms, respectivement.

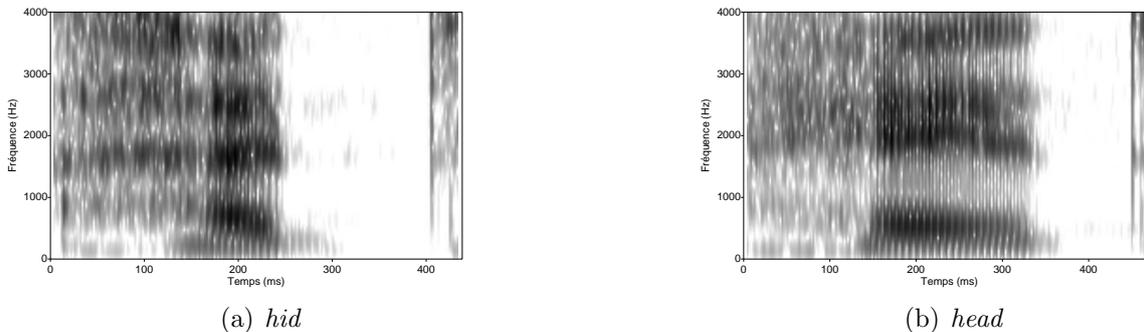


FIG. 6.78 – Spectrogrammes de *hid* et *head* par le locuteur *gty*.

Nous allons analyser la rhoticité de *gla* par le biais du mot-test *heard*. Ce mot est illustré par le spectrogramme de la Figure 6.79(a). On y relève une structure formantique plutôt stable sur le premier tiers de la voyelle, puis F1 et F2 amorcent un mouvement vers le bas jusqu'à la fin ; F2 passe d'environ 2000 Hz à presque 1400 Hz. Il va de soi qu'un tel mouvement a induit un biais dans nos mesures puisque nous n'avons retenu que la valeur médiane dans les Figures 6.76 et 6.77. L'écoute de ce mot-test chez le locuteur *iwd* signale clairement la présence d'une voyelle de type [ɛ] suivie d'une approximante. Autrement dit, si nous avions pu isoler la voyelle précédant l'approximante, nous aurions très vraisemblablement obtenu des valeurs de F2 plus élevées, et la voyelle serait alors apparue comme plus antérieure sur les Figures 6.76 et 6.77. Dans la Figure 6.79(b), où

est représenté le mot-test *heard* prononcé par une locutrice de *gla*, le schéma est différent. Certes, l’abaissement de F2 est toujours visible, et nos mesures souffrent donc du même biais, mais la fréquence de F3 reste d’abord stable pour augmenter légèrement à la fin de la voyelle. L’analyse auditive ne nous permet pas de détecter une quelconque approximante, ni quoi que ce soit qui puisse être assimilé à une consonne. Dans la Figure 6.80, on constate que *hard*, produit par un locuteur de *gla*, ne fait apparaître (ni sur le tracé, ni à l’oreille) aucun indice qui laisserait supposer une quelconque rhoticité (noter également que ce locuteur ne prononce pas les /h/ en début de mot). En règle générale, nous avons pu constater que la rhoticité n’était pas systématique en *gla* et qu’elle se manifestait sous différentes formes allant d’une probable approximante alvéolaire à une voyelle centrale. Il semblerait que nous soyons en présence d’un changement phonétique en cours semblable à celui qui a affecté les ensembles NORTH-FORCE au 18^e siècle en anglais standard¹⁴⁷.

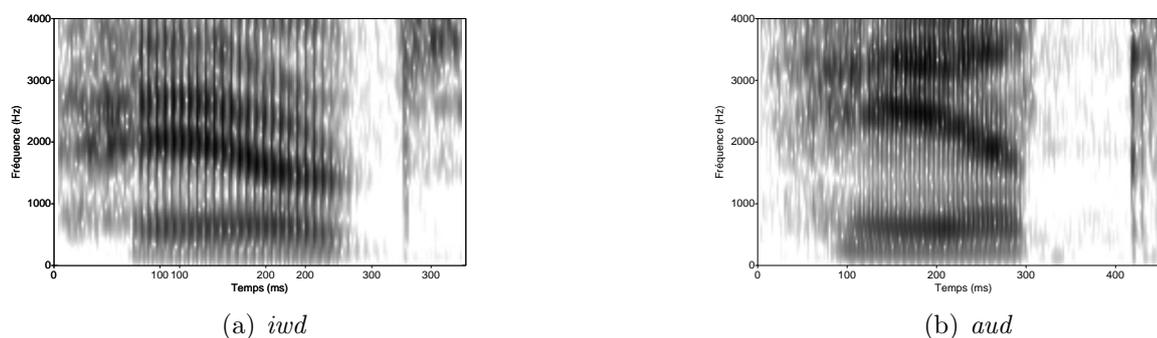


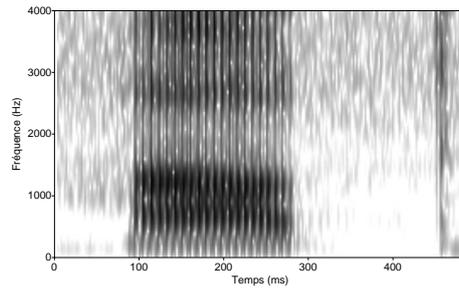
FIG. 6.79 – Spectrogrammes de *heard* par le locuteur *iwd* et la locutrice *aud*.

6.3.6.2 Les diphtongues

10 locutrices et 7 locuteurs sont représentés dans les Figures 6.81 et 6.82, respectivement. Sur le plan auditif, les traits les plus saillants sont :

- la qualité de *hide*, qui ressemble aux productions typiques de *hade* en *sse* ; cette réalisation fait l’unanimité ;

¹⁴⁷Schématiquement : [ɔr] → [ɔə] → [ɔ:], voir [Gim80], 118.

FIG. 6.80 – *hard* produit par le locuteur *wnh* de *gla*.TAB. 6.10 – Durée des monophthongues de *gla*(ms).

Voyelle	Durée médiane	Écart interquartile
heed	144	56
hid	144	50
head	225	81
had	237	61
hard	231	52
hod	231	58
hoard	243	50
hood	163	51
who'd	244	63
Hudd	150	50
heard	231	68

– les monophthongues, très fermées, dans *hade* et *hoed*.

Le timbre de *hide* est probablement le trait le plus typique des diphtongues de *gla* (et de *shl* également, voir Section 6.3.12.2). La monophthongue pour *hade* est produite par tous les locuteurs. En revanche, quelques rares locuteurs ont une diphtongue dans *hoed*.

Le timbre central du premier élément de *howd* est particulièrement saillant.

Les réalisations de *hared* semblent, à l'écoute, particulièrement homogènes pour le dialecte *gla*. En réalité, si tous les locuteurs ont une voyelle de type [e], voire [ɪ] avant le /r/, la réalisation de ce dernier est variable. Ceci se traduit soit par un mouvement conjoint de F2 et F3 vers le bas, c'est la marque d'une approximante (alvéolaire, post-alvéolaire ou rétroflexe), soit par un mouvement de F2 vers le bas, alors que F3 reste

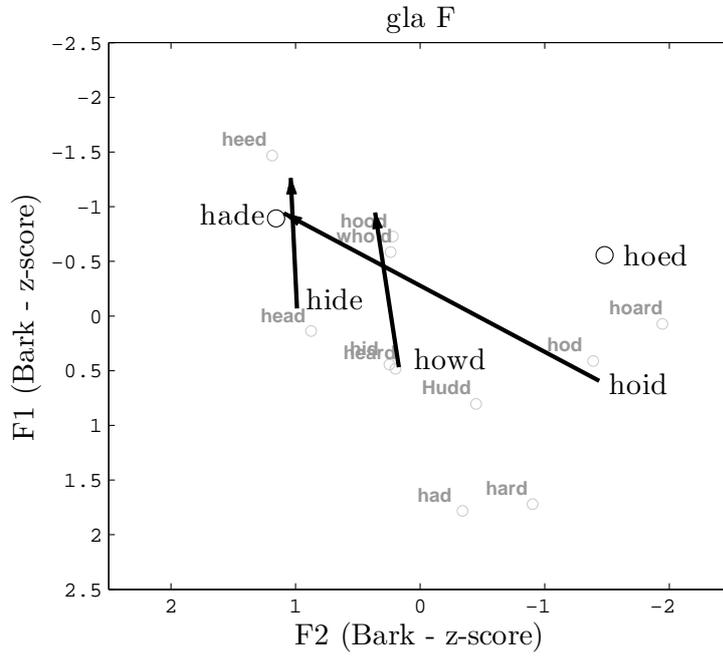


FIG. 6.81 – Diphtongues de *gla* pour les sujets féminins.

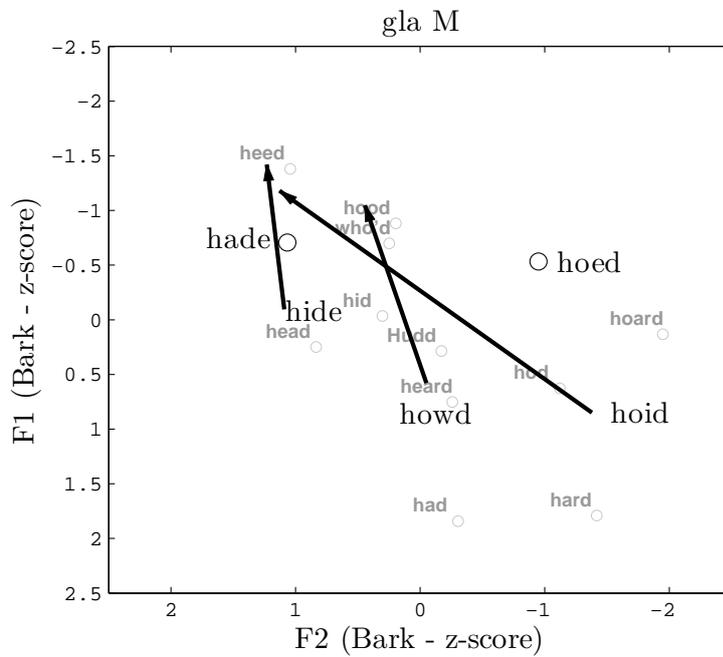


FIG. 6.82 – Diphtongues de *gla* pour les sujets masculins.

à peu près constant. Ce dernier cas semble lui-même recouvrir plusieurs variantes qui correspondent, au moins perceptivement, à une battue, voire à un mouvement vers un schwa. Une étude récente de Stuart-Smith ([SS07]) rapporte que cela fait environ 30 ans que les locuteurs de classe ouvrière ont tendance à ne pas produire le /r/ en position post-vocalique. L'auteur identifie à l'oreille pas moins de 8 types de réalisations pour le /r/ après [a] : [r], [r̥], [ɹ], [a^ɹ], [aħ] et [a]. Sur les 12 locuteurs analysés dans l'étude de Stuart-Smith, seuls deux sont systématiquement rhotiques. Toutes ces variantes sont probablement présentes dans nos données ; nous donnons, en guise d'illustration, deux réalisations différentes dans la Figure 6.83. La perte de la rhoticité est un changement en cours, et seule une étude de phonétique instrumentale articulatoire permettrait une analyse suffisamment fine du processus. Le détail des réalisations individuelles n'a que peu d'intérêt puisque les potentiels facteurs sociolinguistiques qui génèrent cette variation nous sont inaccessibles.



FIG. 6.83 – Spectrogrammes de *hared* produits par le locuteur *gty* et la locutrice *kam*.

La réalisation du locuteur *gty* traduit vraisemblablement une centralisation de la voyelle accompagnée d'une articulation secondaire uvulaire ou pharyngale, comme en témoigne le mouvement final de F3 vers le haut, accompagné d'une baisse d'amplitude. La variante de la locutrice *kam* est typique d'une approximante post-alvéolaire ou rétroflexe.

La voyelle de *heered* a un timbre de départ identique à celui de *heed*. Les remarques que nous avons émises concernant l'implémentation phonétique de la rhoticité dans *hared*

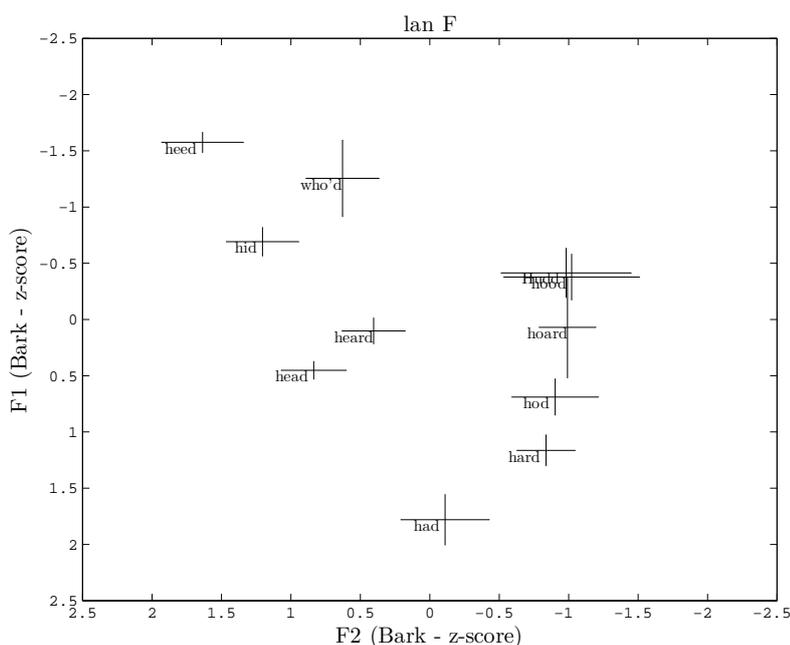


FIG. 6.84 – Monophtongues de *lan* pour les sujets féminins.

s'appliquent également ici.

Il est difficile de mettre en évidence un schéma de variation pour *hured*. On trouve principalement une voyelle de type [jʊ] ou [ʌ] suivie d'un /r/.

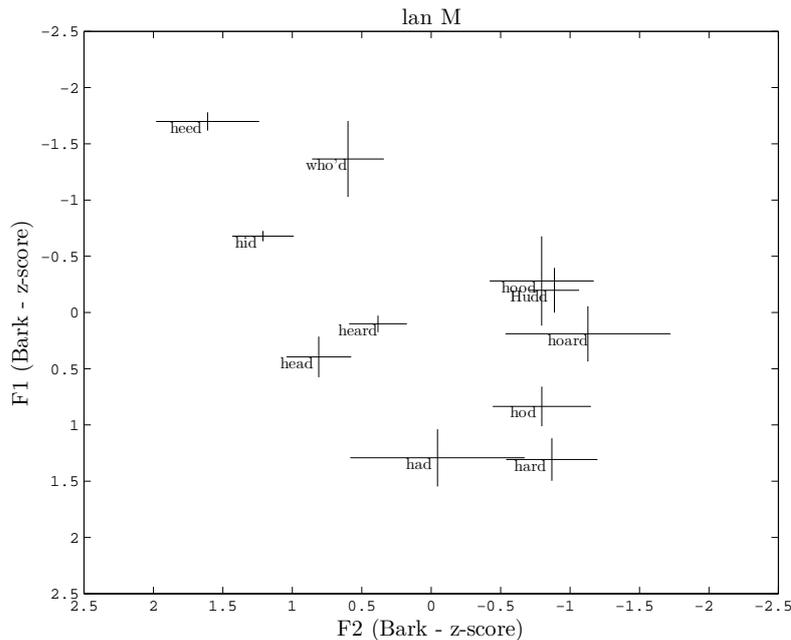
6.3.7 Lancashire (*lan*)

6.3.7.1 Les monophtongues

Dans la Figure 6.84, 8 locutrices ont produit l'ensemble des 11 voyelles. On y constate :

- une réalisation postérieure de *Hudd* et *hood* ;
- un chevauchement important entre *Hudd* et *hood* ;
- plus généralement, une zone postérieure très dense puisque 5 timbres – 4 en fait (phonologiquement), puisque *Hudd* et *hood* ne forment probablement qu'une seule catégorie – ont un degré de postériorité approximativement similaire, et ces mêmes timbres n'occupent qu'une partie réduite sur l'axe de l'aperture.

Dans la Figure 6.85, ce sont 10 locuteurs qui ont produit l'ensemble des 11 voyelles. Les mêmes remarques décrivant la Figure 6.84 sont valables ici.

FIG. 6.85 – Monophthongs de *lan* pour les sujets masculins.

L'écoute des stimuli ne fait aucun doute quant à l'absence de scission FOOT-STRUT. En ce qui concerne la densité particulière de la dimension postérieure, nous mettrons en parallèle nos résultats avec ce que rapporte Labov ([Lab94], 328), en se basant sur Martinet, d'une part, et Haudricourt et Juilland, d'autre part, à propos du nombre de distinctions possibles sur une seule dimension phonétique : « three levels of height are the most that can be maintained in the back ». Une éventuelle « surcharge » est avancée comme un facteur important de changement en chaîne ou de convergence.

Ce qui singularise *lan* d'après les Figures 6.84 et 6.85, c'est le fait que quatre timbres distincts, quatre phonèmes, se partagent la dimension de l'aperture avec un degré de postériorité équivalent. Or, jusqu'ici, nous n'avons rencontré que des systèmes vocaliques avec seulement 2 ou 3 timbres qui soient phonétiquement postérieurs. Pour *brm* (Figures 6.38 et 6.39) GOOSE et FOOT-STRUT présentent un degré d'antériorité qui conduit à ne pas considérer ces timbres comme strictement postérieurs. La même remarque s'applique à *crn*, *ean* et *gla*¹⁴⁸. Pour *eyk*, c'est l'ensemble FOOT-STRUT qui a une qualité nettement

¹⁴⁸Excepté le fait que, dans ces trois dialectes, FOOT et STRUT sont deux ensembles distincts.

TAB. 6.11 – Durée des monophthongues de *lan* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	244	94
hid	150	88
head	163	69
had	163	89
hard	275	69
hod	175	81
hoard	275	100
hood	157	63
who'd	250	104
Hudd	150	59
heard	263	75

postérieure. En ce qui concerne *lan*, la présence de quatre voyelles postérieures prouve qu'il est possible de maintenir quatre contrastes sur cet axe ; et il n'y a aucune raison de croire que les locuteurs de *lan* éprouvent plus de difficultés à communiquer que les autres. En revanche, il sera instructif d'observer l'évolution de ce système dans les années à venir ; si l'hypothèse d'un mouvement dans le sens inverse des aiguilles d'une montre est applicable à tous nos dialectes¹⁴⁹, et si la « surcharge » de la dimension postérieure est, à défaut d'être une cause suffisante, un facteur précipitant ce mouvement, on peut supposer que, d'une certaine manière, *lan* affiche un certain retard dans le cadre de ce changement en chaîne. S'il suit le schéma évolutif de *sse*, on peut anticiper une antériorisation de FOOT-STRUT et une fermeture de NORTH-FORCE.

6.3.7.2 Les diphtongues

Dans les Figures 6.86 et 6.87, ce sont 7 locutrices et 9 locuteurs qui sont représentés. Ici encore, les réalisations de *hade* et *hoed* sont majoritairement monophthongales. Une analyse auditive de cette dernière en parallèle avec *hoard* révèle que certains locuteurs semblent avoir une parfaite homophonie (Figure 6.88). Les F1 et F2 moyens sur la partie

¹⁴⁹Pour l'instant, à notre connaissance, cette hypothèse ne concerne que *sse*.

maximalement stable de ces voyelles sont de 629 et 949 Hz pour *hoard*, et 609 et 1011 Hz pour *hoed*. D'autres ont une différence de surface : *hoed* est une monophthongue et *hoard*, une diphtongue centripète (Figure 6.89). Les F1 et F2 moyens sur la partie maximalement stable de ces voyelles sont de 569 et 978 Hz pour *hoard*, et 560 et 1051 Hz pour *hoed* ; les deux états stables sont donc très proches phonétiquement, et leur faible distance acoustique n'est probablement pas suffisante pour coder un contraste. Ce qui a été dit plus haut (Section 6.3.5.2) sur la réalisation de *hide* est aussi valable pour *lan*, à ceci près que la réalisation monophthongale est beaucoup plus fréquente dans le cas présent.

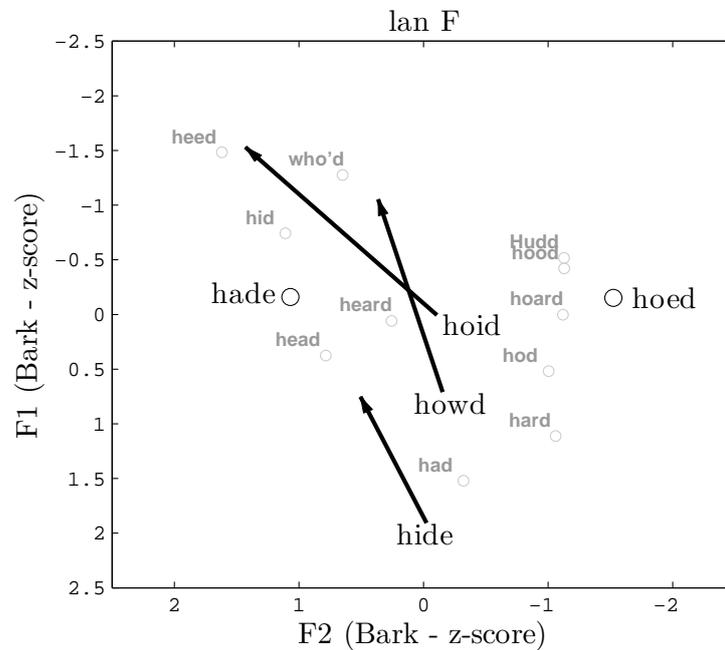


FIG. 6.86 – Diphtongues de *lan* pour les sujets féminins.

La voyelle de *hared* est très majoritairement réalisée comme une monophthongue longue, qui pourrait être analysée, une fois de plus, comme le pendant phonologiquement long de *head*. La voyelle de *heered* est une diphtongue centripète ; de rares réalisations rhotiques sont également à signaler. Dans le dialecte *lan*, la voyelle de *hured* ne fait pas exception à la règle : comme nous avons pu le constater pour la plupart des dialectes, cette voyelle est réalisée par de multiples variantes, ce qui révèle la difficulté qu'ont éprouvé les locuteurs à lui faire correspondre une forme phonologique.

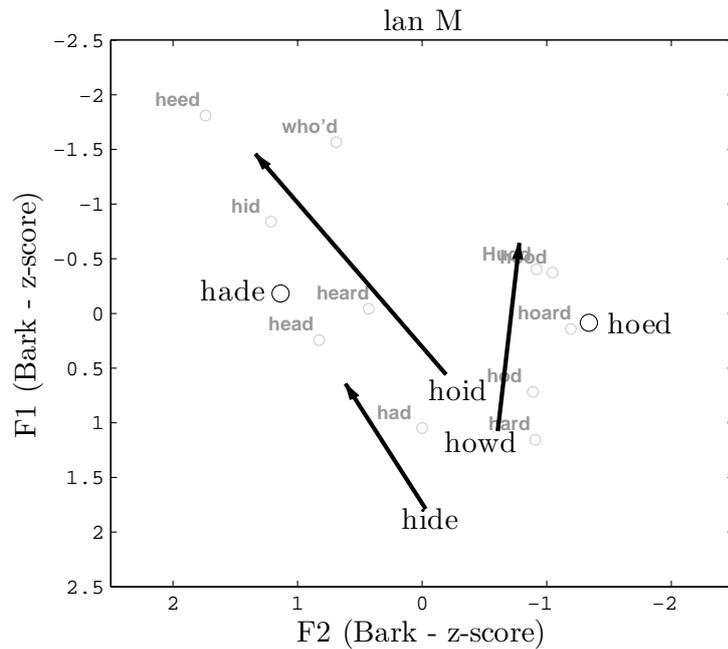


FIG. 6.87 – Diphtongues de *lan* pour les sujets masculins.



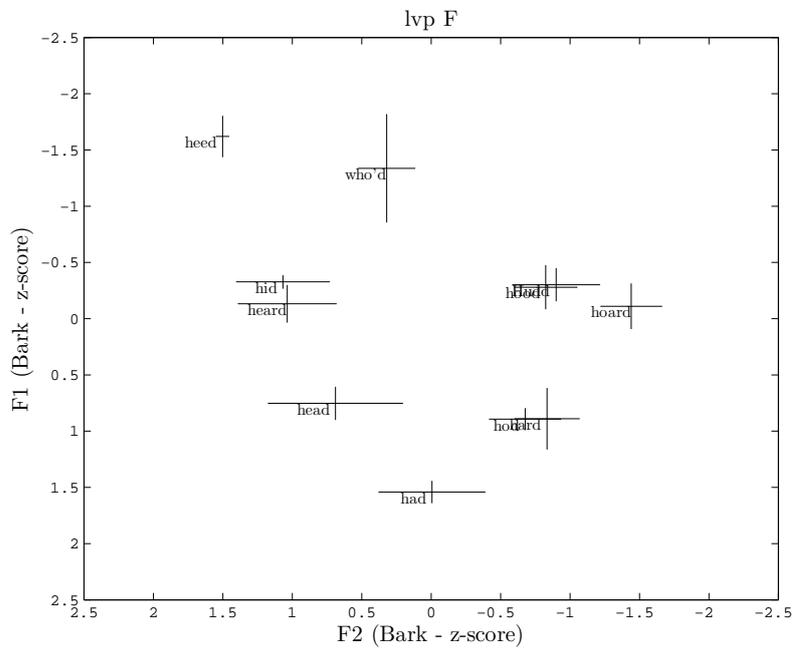
FIG. 6.88 – Spectrogrammes de *hoard* et *hoed* produits par la locutrice *mjh*.

6.3.8 Liverpool (*lvp*)

6.3.8.1 Les monophthongues

Dans la Figure 6.90, 10 locutrices ont produit l'ensemble des 11 voyelles. Ce diagramme fait apparaître :

- la proximité de *hid* et *heard* ;
- la proximité de *hod* et *hard* ;

FIG. 6.89 – Spectrogrammes de *hoard* et *hoed* produits par la locutrice *acf*.FIG. 6.90 – Monophthongues de *lvp* pour les sujets féminins.

– le chevauchement de *Hudd* et *hood*, avec une réalisation postérieure.

Dans la Figure 6.91, ce sont 8 locuteurs qui ont produit l'ensemble des 11 voyelles. Le schéma est à peu près identique à celui de la Figure 6.90. On note une plus grande dispersion de *hood* et *Hudd* conduisant à un chevauchement non seulement entre ces deux timbres, mais également avec *hoard*.

Les Figures 6.92(a) et 6.92(b) illustrent la proximité phonétique entre *hid* et *heard*. À l'écoute, le caractère centralisé de *hid* ne fait aucun doute. Mais c'est le degré d'antériorité

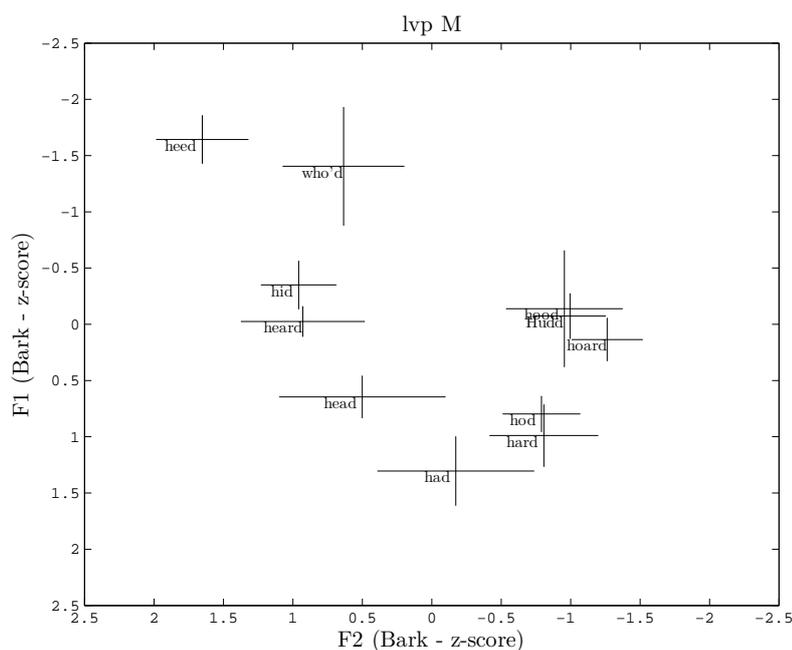
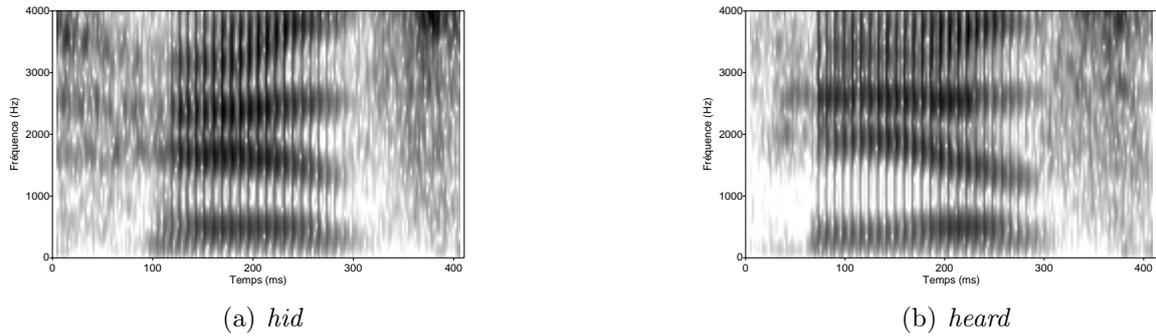


FIG. 6.91 – Monophthongs de *lvp* pour les sujets masculins.

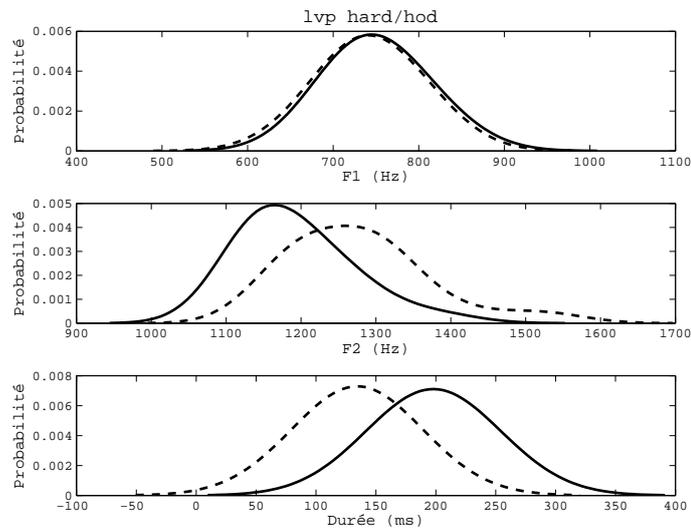
et de fermeture de *heard* qui constitue le trait le plus remarquable (comparé à *sse*). Nous nous trouvons en réalité dans la même situation que pour *heard* dans *gla* (voir, e.g. la Figure 6.79(b)) : l'abaissement de F2 sur la deuxième moitié de la voyelle dans la Figure 6.92(b) traduit une diphtongaison centripète, et l'utilisation de la médiane dans nos calculs vient masquer la dynamique des formants. En réalité, on constate que le timbre initial de *heard* (Figure 6.92(b)) est légèrement plus fermé que celui de *hid*. On remarque, incidemment, que chez ce locuteur, /d/ n'est pas réalisé comme une occlusive mais comme une fricative, probablement dentale, mais il semble qu'il s'agisse d'une caractéristique idiosyncrasique.

La proximité de timbre entre *hard* et *hood* est confirmée par l'analyse auditive ; il nous est difficile de les distinguer. Au cas où une analyse plus détaillée aboutirait à une équivalence de timbre, on peut d'emblée écarter l'hypothèse d'une totale homophonie en raison de la différence de durée médiane de presque 100 ms qui sépare ces deux voyelles. Cette proximité est illustrée dans la Figure 6.93. La superposition des courbes de F1, et le chevauchement important des courbes de F2 et de la durée témoignent du peu de

FIG. 6.92 – Spectrogrammes de *hid* et *heard* par le locuteur *aaaj*.

robustesse du codage acoustique de ce potentiel contraste. Les moyennes de F1, F2 et de durée sont de :

- 751 Hz, 1193 Hz et 198 ms pour *hard* ;
- 743 Hz, 1277 Hz et 134 ms pour *hod* ;

FIG. 6.93 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *hard* (ligne continue) et *hod* (ligne discontinue) produites par les locutrices de *lvp*.

Les Figures 6.94(a) et 6.94(b) montrent respectivement des spectrogrammes de *hard* et *hod*. La valeur moyenne de F1 et F2 calculée sur la partie stable de ces deux voyelles est de :

- 684 et 1128 Hz pour *hard* ;

– 726 et 1112 Hz pour *hod*.

Il y a donc une différence de 42 Hz et 16 Hz sur F1 et F2, respectivement. La comparaison de deux voyelles ne constitue pas une preuve mais suffit néanmoins à nous interpeler : les études sur le seuil différentiel de perception des formants ([Fla55], [KPW94]) montrent que si la différence de 42 Hz est suffisante sur F1 pour distinguer deux timbres dans des conditions de laboratoire (le seuil étant d'environ 14,5 Hz en dessous de 800 Hz, voir [KPW94], 492-493), il est difficile de savoir si la différence de 42 Hz sur F1 permet de distinguer *hard* et *hod*, sachant que les 16 Hz qui séparent F2 pour ces deux timbres sont en deçà du seuil, et que, par conséquent, F2 n'est pas discriminant ici. Comme le note Flanagan ([Fla55], 616-617), un seuil différentiel de perception constitue une mesure de la plus petite déviation d'un formant par rapport à un autre formant de référence nécessaire pour être perçue ; le passage d'un phonème à un autre requiert des valeurs plus élevées. Nous concluons donc qu'une identité entre *hard* et *hod* en termes de qualité vocalique est très vraisemblable. Ce point mériterait d'être testé par le biais d'une expérience de perception. La seule opposition qui nous paraisse pertinente est celle de la quantité : mesurées manuellement, les deux voyelles représentées affichent environ 280 et 190 ms (*hard* et *hod*, respectivement).

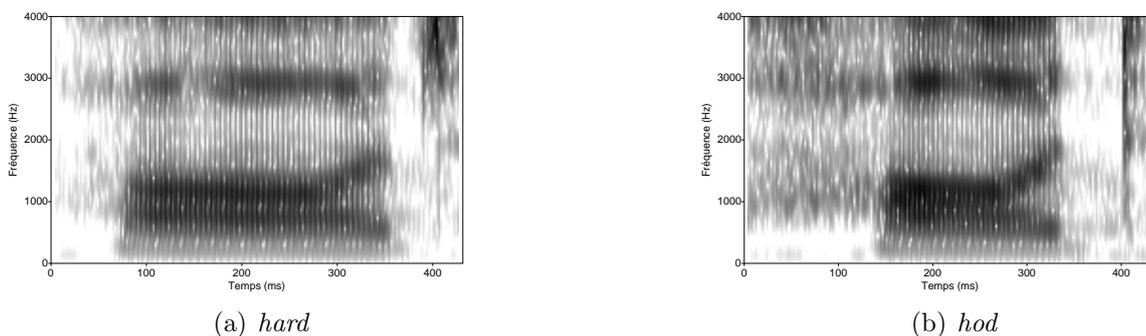


FIG. 6.94 – Spectrogrammes de *hard* et *hod* par la locutrice *lfm*.

TAB. 6.12 – Durée des monophthongues de *lvp* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	275	57
hid	181	49
head	193	52
had	194	68
hard	306	68
hod	209	56
hoard	294	79
hood	175	50
who'd	281	63
Hudd	175	48
heard	294	69

6.3.8.2 Les diphtongues

9 locutrices et 5 locuteurs ont produit les voyelles des Figures 6.95 et 6.96. Le mouvement de faible ampleur de *hade* n'est pas particulièrement évident sur le plan auditif. Concernant *hoed*, le timbre plus antérieur du premier élément chez les hommes est aisément perceptible à l'écoute des mots-test.

La voyelle de *hared* est une monophthongue longue dont le timbre est proche de [e]. Au contraire de ce que nous avons pu constater pour *brm* et *eyk*, *hared* présente une différence systématique avec *head* chez la majorité des locuteurs, ce dernier timbre étant sensiblement plus ouvert. Ceci est illustré dans la Figure 6.97 ; les graphes représentant F1 et F2 montrent qu'en moyenne, la voyelle de *head* est plus compacte. Les valeurs moyennes de F1, F2 et de durée sont de 501 Hz, 1796 Hz et 169 ms pour *hared*, et 585 Hz, 1652 Hz et 102 ms pour *head*. La Figure 6.98 démontre que, au vu du chevauchement entre les valeurs formantiques des deux voyelles, on peut considérer que *hared* peut être analysé comme le pendant phonétiquement (et probablement phonologiquement) long de *hid* pour la plupart des locuteurs. Enfin, la comparaison des mots-test *heard* (ligne continue) et *hared* (pointillés) dans la Figure 6.99 semble confirmer la convergence NURSE-SQUARE dans notre échantillon de *lvp*.

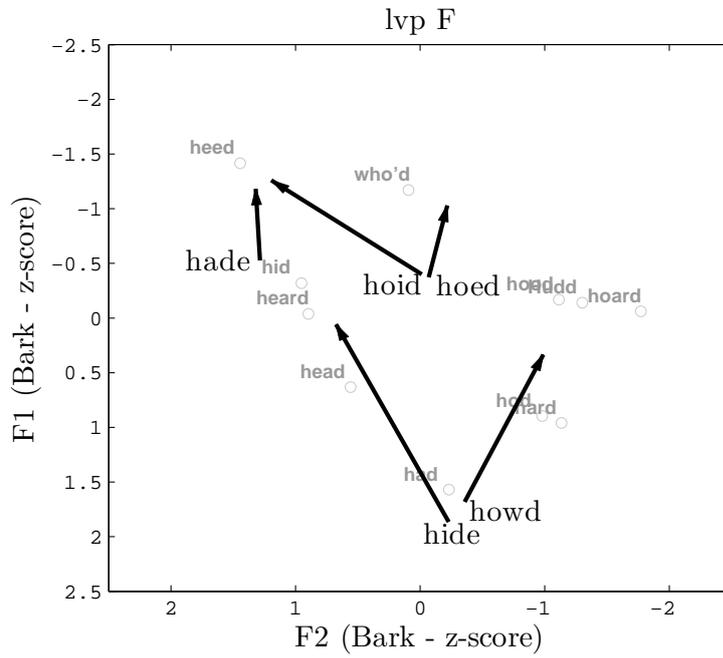


FIG. 6.95 – Diphtongues de *lvp* pour les sujets féminins.

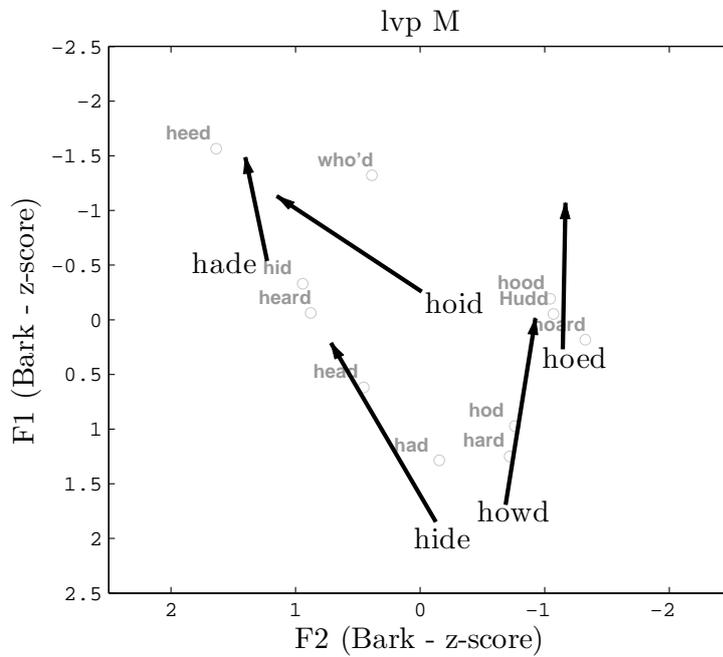


FIG. 6.96 – Diphtongues de *lvp* pour les sujets masculins.

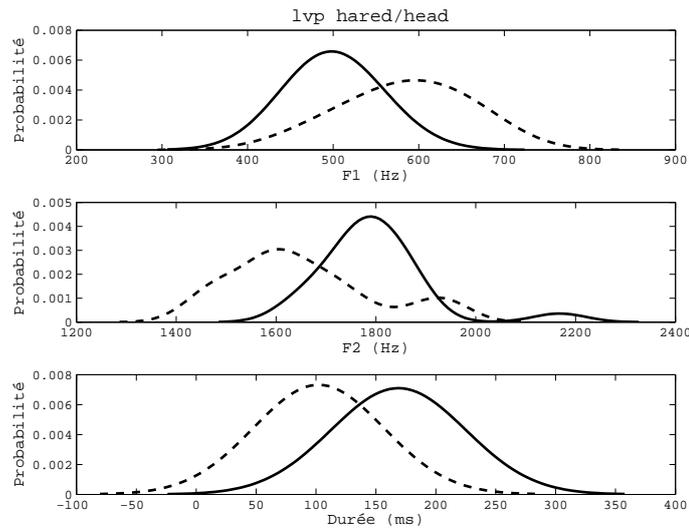


FIG. 6.97 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *hared* (ligne continue) et *head* (ligne discontinue) produites par les locuteurs masculins de *lvp*.

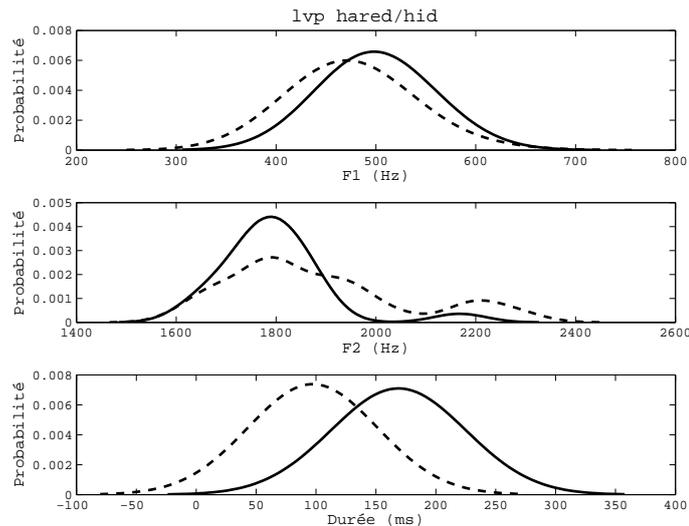


FIG. 6.98 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *hared* (ligne continue) et *hid* (ligne discontinue) produites par les locuteurs masculins de *lvp*.

La voyelle de *heered* est une diphtongue centripète dont le timbre initial est équivalent à *heed*. Quant à la voyelle de *hured*, elle varie moins que dans d'autres dialectes rencontrés jusqu'ici : une grande majorité des réalisations sont soit de type [jo], soit [o] ; les locuteurs emploient parfois les deux. Il s'avère que chez les locuteurs qui produisent cette dernière variante, *hured* et *hoard* sont homophones. Cette homophonie est illustrée dans la Figure 6.100 : les valeurs moyennes de F1 et F2 relevées sur la partie maximale

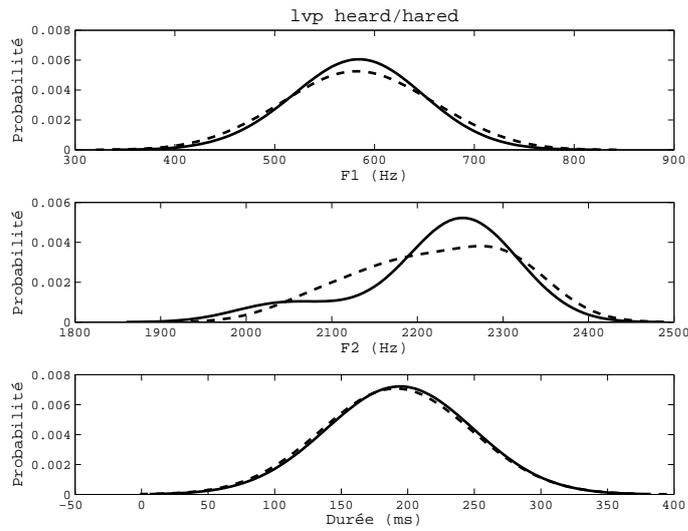


FIG. 6.99 – Densité de probabilité de F1, F2 et la durée pour les voyelles de *heard* (ligne continue) et *hared* (ligne discontinue) produites par les locutrices de *lvp*.

stable des deux voyelles sont de 531 et 881 Hz pour *hured*, et 530 et 841 Hz pour *hoard*. On note cependant sur cette figure une différence d'intensité importante au niveau de F3.

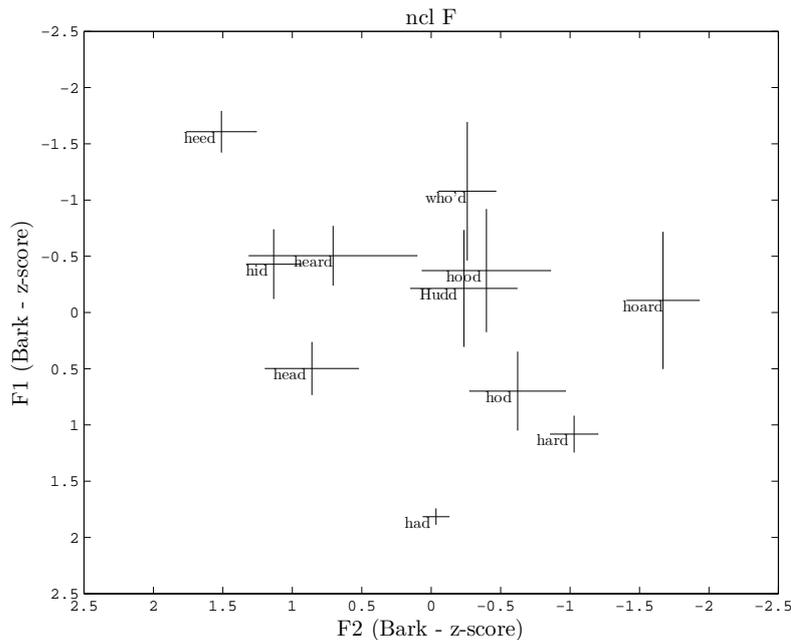


FIG. 6.100 – Spectrogrammes de *hured* et *hoard* pour le locuteur *djf* de *lvp*.

6.3.9 Newcastle (*ncl*)

6.3.9.1 Les monophongues

Dans la Figure 6.101, 5 locutrices ont produit l'ensemble des 11 voyelles. On remarque que :

FIG. 6.101 – Monophtongues de *ncl* pour les sujets féminins.

- *hid* et *heard* présentent un niveau d’aperture équivalent et ont une partie de l’axe de F2 en commun ;
- la partie postérieure du système est particulièrement dense, avec un chevauchement possible entre *Hudd*, *hood* et *who’d* (ce dernier dans une moindre mesure).

Dans la Figure 6.102, ce sont 2 locuteurs qui ont produit l’ensemble des 11 voyelles. Il s’ensuit que la représentation donnée ici du système de *ncl* est moins fiable que pour les autres systèmes (sauf *crn*). Nous reviendrons plus en détail sur ce système pour les analyses individuelles. On relève, en plus de ce qui a été dit à propos de la Figure 6.101, une certaine proximité entre *hard* et *hod*.

Le système de *ncl* est, à l’instar de celui de *lan*, particulièrement dense dans sa partie postérieure. À la différence de ce dernier, *ncl* ne semble pourtant pas présenter, à première vue, une antériorisation de GOOSE, caractéristique de tous les dialectes étudiés jusqu’ici, hormis *eyk* (Figures 6.67 et 6.68) et les locuteurs masculins de *crn* (Figure 6.50).

Une écoute attentive révèle que la voyelle de *heard* est variable : un locuteur (*ggc*) a une voyelle proche de celle de l’anglais standard ; les autres ont, pour la plupart, des

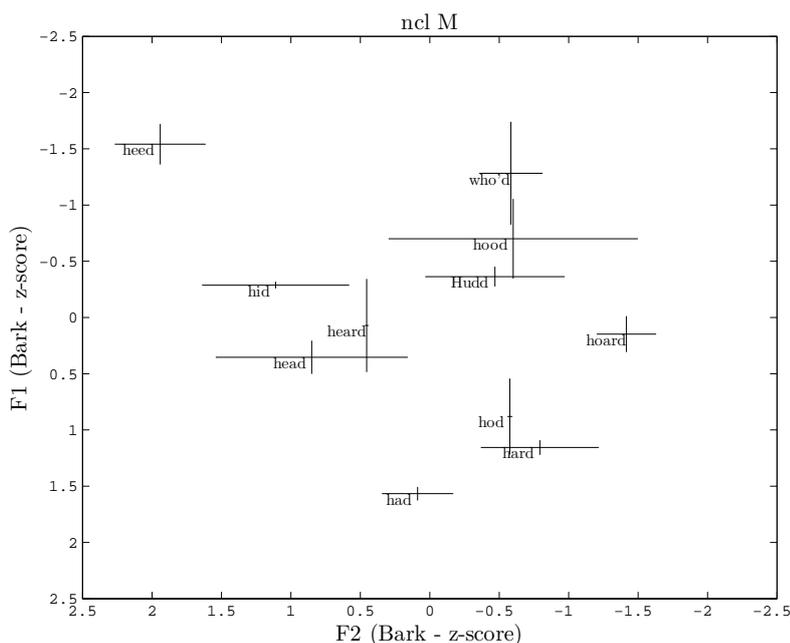


FIG. 6.102 – Monophtongues de *ncl* pour les sujets masculins.

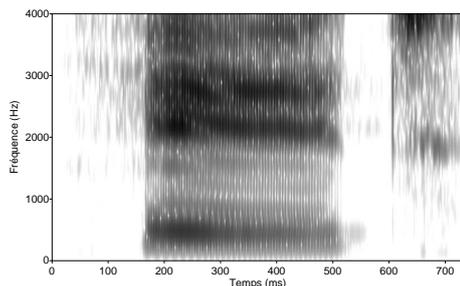


FIG. 6.103 – *heard* produit par la locutrice *jxh* de *ncl*.

productions très antérieures et plutôt fermées (allant de [œ] à [ø]; dans leur description, Watt et Allen ([WA03], 268) notent [ø]). La locutrice *jxh* se démarque des autres par une réalisation extrêmement fermée (Figure 6.103) de type [hɹ:d].

Les réalisations de *who'd* sont également très variables : certains locuteurs ont une monophtongue très fermée et postérieure de type [u] (Figures 6.104(a) et 6.104(b)), d'autres ont une monophtongue fermée antériorisée ([ɯ], Figure 6.104(c)), d'autres encore ont une diphtongue (type [əu] ou [əʊ], Figure 6.104(d)). Notre impression – mais ce commentaire reste très spéculatif – est que les locuteurs du corpus *ncl* ne sont pas tous au même stade d'un possible changement en chaîne causant l'antériorisation des voyelles postérieures les

plus fermées. La réalisation de l'opposition *hood-who'd* est très variable; nous avons pu relever les cas de figure suivants :

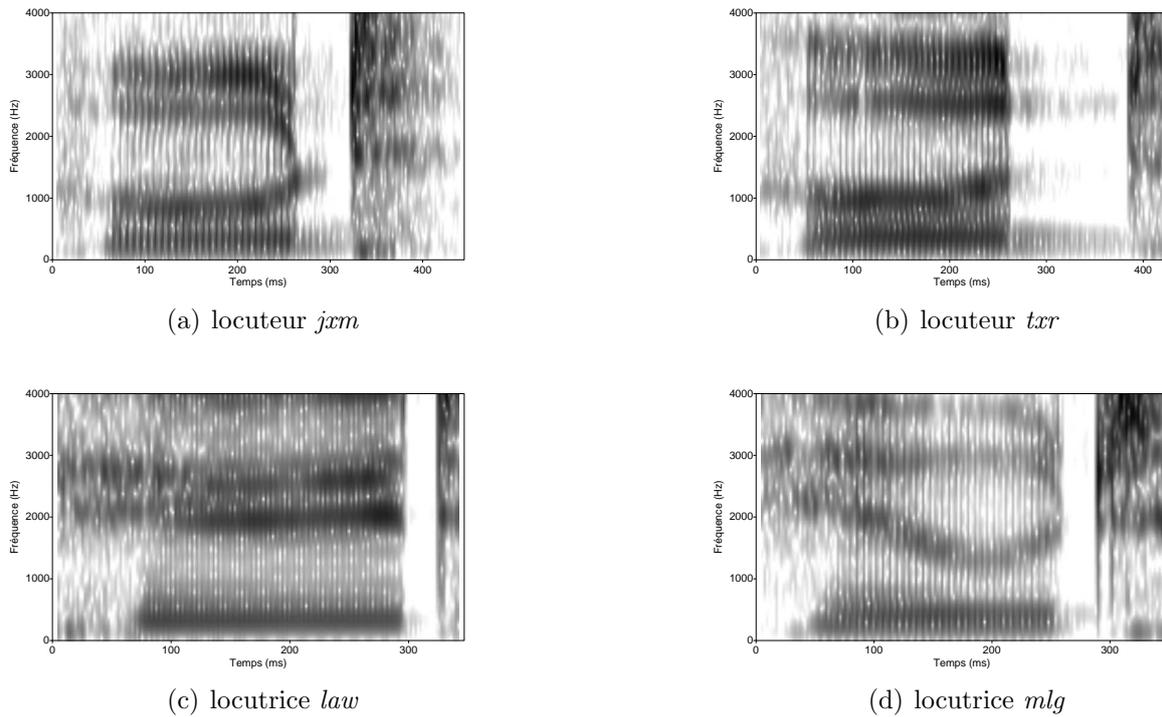
- cas A : *hood* et *who'd* sont de parfaits homophones (e.g. la locutrice *sxp* et le locuteur *krt*);
- cas B, lorsque la voyelle de *hood* est [u:], *who'd* est une diphtongue; mais les deux locutrices qui ont ce type de réalisation (*jxh* et *kjb*) ne l'emploient pas systématiquement, si bien que les deux mots sont parfois homophones;
- cas C1, *hood* a une réalisation brève de type [ʊ]; *who'd* est réalisé avec [ʊ] ou [əʊ]. Ce schéma correspond à celui de l'anglais standard; nous l'avons relevé chez les locutrices qui nous paraissaient les plus jeunes (e.g. *law*, *scn*);
- cas C2, *hood* a une réalisation brève qui varie entre [ʌ, ɔ, ɜ] et [o]; *who'd* peut être réalisé avec [u:] (locuteur *txp*), ou une légère diphtongue dont la voyelle d'arrivée est postérieure (locuteur *jxb*).

Le cas A va nous permettre de mettre en lumière un point qui, s'il n'est pas traité avec circonspection, peut conduire à des conclusions erronées. Si *hood* et *who'd* sont de parfaits homophones, on serait tenté d'inférer que les locuteurs concernés ont une convergence FOOT-GOOSE, phénomène typique des dialectes écossais. Or aucune des références consultées ([Wel82, WM99, WA03, Bea04]) ne mentionne cette possibilité pour *ncl*. Nous avons pu observer chez les locuteurs de *ncl* qui ont une distinction nette entre *who'd* et *hood* que *Hudd* et *hood* étaient homophones; l'absence de scission FOOT-STRUT est donc bien attestée. Chez les quelques locuteurs présentant l'homophonie entre *hood* et *who'd*, *Hudd* a une réalisation septentrionale typique : [ʊ]; ils n'ont donc pas, en surface, les caractéristiques d'un système écossais (où *Hudd* a une voyelle proche de celle de l'anglais standard). Notre hypothèse est que la variation apparente de système chez nos locuteurs de *ncl* n'est en réalité qu'un « accident » d'incidence lexicale. Nous supposons donc, à ce stade, que *hood* appartient à GOOSE pour les locuteurs chez qui l'homophonie a été constatée. Pour tester l'existence d'un ensemble FOOT confondu avec STRUT et non avec GOOSE chez les locuteurs présentant une homophonie *hood-who'd*, nous avons écouté les

mots du passage lu du corpus susceptibles d'appartenir à FOOT : *took*, *cook*, *looked* et *foot* (dans l'ordre d'apparition) chez *sxp* et *krt*. Notre première impression pour *sxp* indique que hormis *cook*, qui évoque une voyelle fermée faisant de cet item un membre possible de GOOSE, les autres mots-test ont une voyelle brève dans la région de [ʊ] ; notre conclusion penche, par conséquent, pour une incidence lexicale très différente de celle de l'anglais standard. Pour le locuteur *krt*, les voyelles des quatre mots-test sont plutôt fermées, et seule une analyse des formants pourra nous fournir une esquisse de réponse. Les valeurs moyennes de F1 et F2 sur l'état stable de la voyelle sont de 264 et 935 Hz pour *foot* et de 305 et 896 Hz pour *food*. En guise de comparaison, les fréquences des deux premiers formants relevées sur une occurrence de *but* dans un contexte accentué, sont de 486 et 922 Hz, ce qui suffit à distinguer la voyelle de *but* de celles de *foot* et *food*. Il est évident que le contexte sémantique, contrairement aux mots en /hVd/ qui en sont dépourvus, permet une plus grande latitude dans la réalisation des voyelles, et il est tout aussi certain que le contexte intonatif et consonantique (on sait que les occlusives voisées engendrent un allongement de la voyelle les précédant en anglais) ont une influence sur ces valeurs. On concède donc qu'il serait nécessaire d'obtenir un grand nombre d'occurrences de ces deux mots afin d'avoir une idée claire du chevauchement éventuel des distributions des valeurs formantiques avant de tirer des conclusions définitives. Néanmoins, le fait que la voyelle de *foot* (ainsi que les autres voyelles potentielles de FOOT chez *krt*) soit ici plus fermée et plus postérieure que celle de *food* nous met face à l'alternative suivante :

- soit le texte du corpus ABI ne contient pas de mot appartenant à FOOT dans l'idiolecte de *krt*, auquel cas on peut envisager que tous les items étudiés appartiennent à GOOSE¹⁵⁰ ;
- soit ce locuteur a dans son système la convergence FOOT-GOOSE.

¹⁵⁰Nous avons laissé *pulling* de côté car la brièveté de sa voyelle associée à une influence coarticulatoire importante du /l/ ne permet pas une analyse fiable.

FIG. 6.104 – Spectrogrammes de *who'd* par 4 locuteurs et locutrices de *ncl*.TAB. 6.13 – Durée des monophongues de *ncl* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	225	69
hid	181	73
head	194	75
had	222	94
hard	263	60
hod	213	74
hoard	278	53
hood	194	63
who'd	250	87
Hudd	181	63
heard	263	83

6.3.9.2 Les diphtongues

Les diphtongues de *ncl* ont été produites par 5 locutrices (Figure 6.105) et 2 locuteurs (Figure 6.106). Ces figures illustrent parfaitement le biais induit par l'utilisation de valeurs

centrales telles que la médiane ou la moyenne arithmétique. En effet, notre analyse auditive révèle que, conformément aux descriptions antérieures de *ncl* (e.g. [Wat00]), la voyelle de *hade* a trois réalisations possibles :

- une longue monophthongue de type [e:] ;
- une diphtongue fermante semblable à celle de *sse* : [ei] ;
- une diphtongue centripète, [iə].

À l'écoute, les locuteurs produisent une majorité de diphtongues centripètes, ce qui nous a conduit à considérer cette voyelle comme une diphtongue phonétique. Or les deux locuteurs ayant produit au moins une occurrence exploitable des 11 monophthongues et des 4 voyelles que nous avons classées dans la catégorie des diphtongues fermantes ont des productions diamétralement opposées :

- le locuteur *ggc* a une diphtongue fermante (Figure 6.107(a)) ;
- le locuteur *txr* a une diphtongue centripète (Figure 6.107(b)).

Or à quoi correspond la moyenne entre les valeurs formantiques d'une diphtongue fermante et d'une diphtongue centripète ? À une monophthongue...

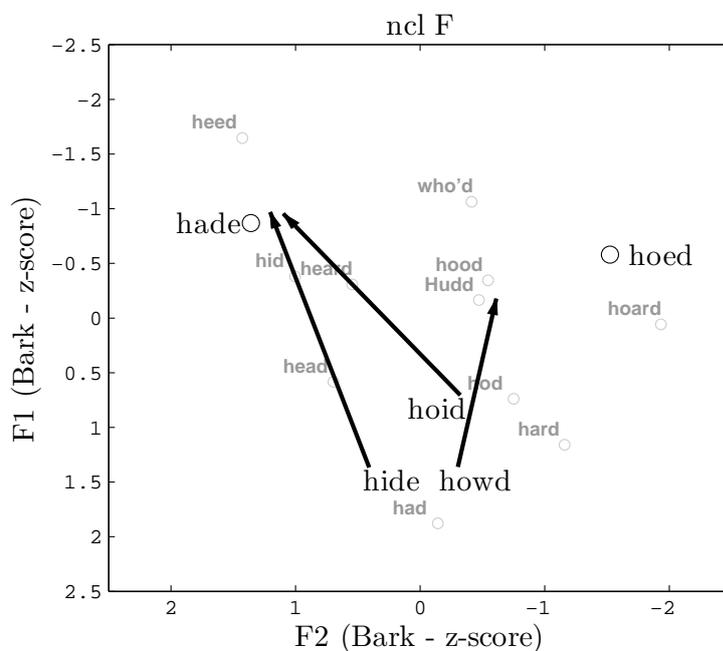


FIG. 6.105 – Diphtongues de *ncl* pour les sujets féminins.

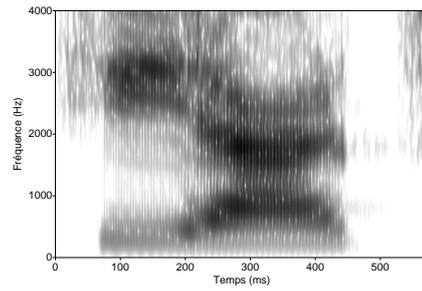


FIG. 6.108 – *heered* produit par la locutrice *jxh* de *ncl*.

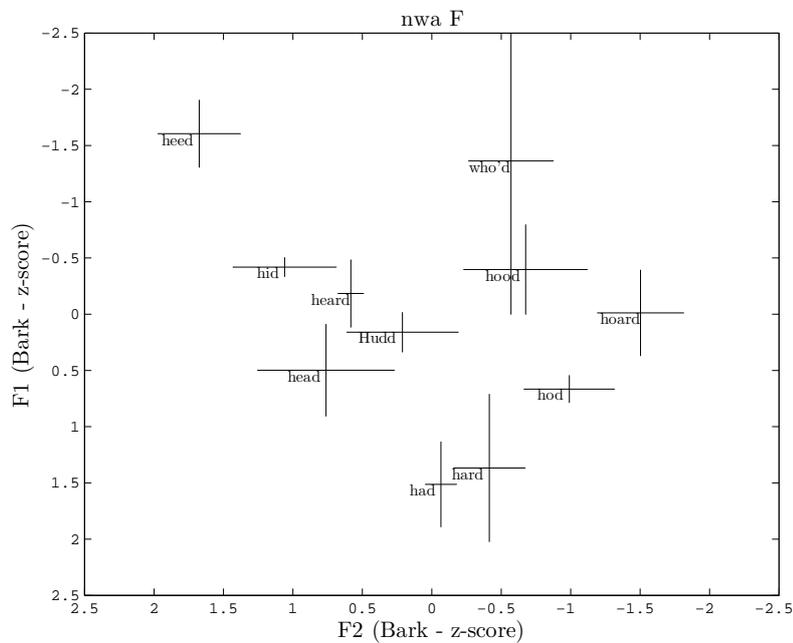
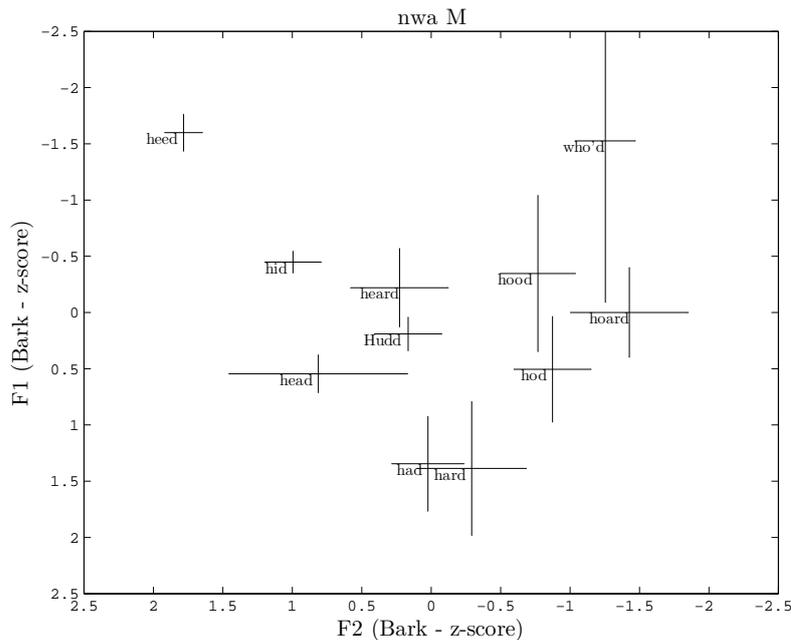


FIG. 6.109 – Monophthongues de *nwa* pour les sujets féminins.

phonétiquement au moins, nous avons affaire à une diphtongue montante¹⁵¹.

La voyelle de *hured* présente une variation inter- et intra-individuelle trop importante pour que sa description détaillée soit d'une quelconque utilité.

FIG. 6.110 – Monophtongues de *nwa* pour les sujets masculins.

6.3.10 North Wales (*nwa*)

6.3.10.1 Les monophtongues

Dans la Figure 6.109, 8 locutrices ont produit l'ensemble des 11 voyelles. On note :

- un léger chevauchement entre *hood* et *who'd*, et la qualité postérieure de ces timbres par rapport à l'anglais standard ;
- la proximité entre *had* et *hard* ;
- la proximité de *heard* et *Hudd*, et leur caractère antérieur.

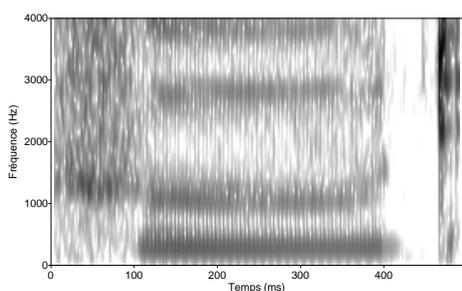
Dans la Figure 6.110, ce sont 7 locuteurs qui ont produit l'ensemble des 11 voyelles. La description fournie plus haut pour la Figure 6.110 s'applique ici.

En se référant à la Table 6.15, on constate que si les trois paires de timbres mentionnées plus haut (*hood-who'd*, *had-hard* et *heard-Hudd*) ont chacune des voyelles très proches, les différences de durée entre les deux membres de chaque paire laissent supposer une opposition de quantité. L'étude auditive de la paire *hood-who'd* fait apparaître les points

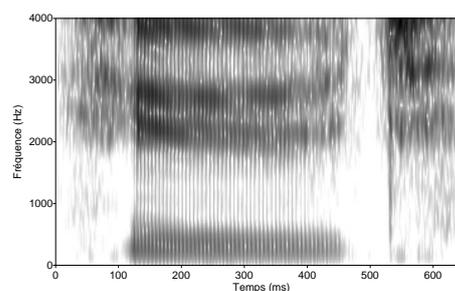
¹⁵¹Ce qualificatif s'applique aux diphtongues dont le second élément est plus saillant (en anglais : « rising diphthongs »). La plupart des diphtongues rencontrées jusqu'ici sont incontestablement du type opposé, i.e. descendantes (« falling diphthongs »).

suivants :

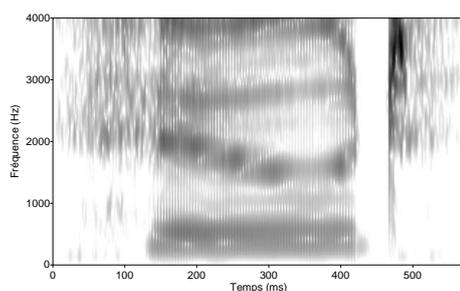
- la faible distance phonétique entre les deux timbres (visible dans les Figures 6.109 et 6.110) se traduit à l’écoute par deux timbres pourtant bien distincts, et cette distinction est appuyée par une différence de durée (voir Table 6.15) ;
- une grande majorité des occurrences de *who’d* ont une monophongue postérieure (Figure 6.111(a)), souvent très fermée ; quelques timbres antérieurs (Figure 6.111(b)) ou très légèrement diphtongués (Figure 6.111(c)) sont néanmoins attestés ;
- le peu de variation intra-individuelle, à l’exception de la locutrice *bea*, qui produit tantôt [u:] (Figure 6.111(a)), tantôt [ɯ:] (Figure 6.111(b)) ;
- *hood* a souvent une qualité plus postérieure qu’en *sse* et souvent plus ouverte.



(a) locutrice *bea*, variante postérieure



(b) locutrice *bea*, variante antérieure



(c) locutrice *sxt*, variante diphtonguée

FIG. 6.111 – Spectrogrammes de *who’d* par 2 locutrices de *nwa*.

Deux particularités ont attiré notre attention : deux des cinq occurrences de *hood* sont confondus avec *who’d* chez le locuteur *wxt*, et il existe une apparente homophonie entre *hood* et *hod* chez la locutrice *aaaj*, les deux voyelles ayant une qualité proche de

TAB. 6.14 – Valeurs de F1 et F2 pour *hod* et *hood* chez la locutrice *aa**j* de *nwa* (Hz).

Voyelle	F1 moyen	écart-type F1	F2 moyen	écart-type F2
<i>hod</i>	509	57	894	54
<i>hood</i>	484	58	863	67

[o]. En ce qui concerne *wxt* nous supposons qu'il s'agit d'un exemple d'incidence lexicale fluctuante ; en revanche, le cas de *aa**j* mérite un examen complémentaire. Pour appuyer notre analyse auditive, nous avons manuellement déterminé les frontières de l'état stable des 6 occurrences de *hod* et des 7 occurrences de *hood*, et avons calculé les valeurs moyennes des deux premiers formants. Les valeurs moyennes et les écart-types par timbre sont donnés dans la Table 6.14 ; ils corroborent l'impression auditive d'une parfaite homophonie entre *hod* et *hood*. Une éventuelle appartenance de *hod* à FOOT ou de *hood* à LOT nous semble tout à fait invraisemblable car ces deux ensembles, contrairement à FOOT et GOOSE, sont tout à fait distincts sur le plan orthographique. Il pourrait s'agir d'une mauvaise lecture des items de la part du sujet. Afin d'éclaircir ce point, nous avons analysé les mots *foot* et *gods* du passage lu du corpus, ce qui fut informatif à plus d'un titre. Les valeurs moyennes de F1 et F2 sur tout l'état stable sont de 518 et 920 Hz pour *foot* et 505 et 983 Hz pour *gods* ; malgré la variation induite par les facteurs non contrôlés habituels (contexte sémantique, segmental et suprasegmental), nous émettrons l'hypothèse qu'il n'est pas possible de distinguer ces deux timbres. Une preuve supplémentaire vient appuyer notre hypothèse d'une convergence au moins partielle entre LOT et FOOT : la locutrice *aa**j* lit la phrase suivante : « reduced their gods to one ». La personne en charge de l'enregistrement reprend *aa**j* en répétant cette phrase avec une emphase sur *gods*, prononcé avec une voyelle ouverte, comme pour signifier à la locutrice que cette dernière a lu <goods> au lieu de <gods>. La locutrice reprend avec une voyelle ouverte (F1=767 Hz, F2=1104 Hz), très éloignée de sa production initiale.

En ce qui concerne la proximité de timbre entre *had* et *hard*, nous avons constaté que le contraste était implémenté de deux manières différentes : sur les 20 locuteurs de

nwa, 13 ont une voyelle ne différant que par la durée, les 7 autres utilisent et la durée et l'opposition antérieur/postérieur pour démarquer les deux voyelles.

La plupart des locuteurs ont une différence qualitative évidente entre *Hudd* et *heard* : la voyelle de ce dernier étant plus fermée. Son degré de fermeture est sensiblement supérieur à celui qu'on peut rencontrer en anglais standard. Quelques réalisations de la voyelle de *heard* sont très antérieures et fermées. Par exemple, dans la Figure 6.112(a) où *heard* est prononcé par le locuteur *rwl*, on constate l'éloignement marqué de F1 et F2, 471 et 1833 Hz en moyenne, respectivement. À titre de comparaison, la voyelle de *hid*, produite par le même locuteur, est représentée dans la Figure 6.112(b) (F1=446 Hz, F2=1959 Hz) : la voyelle de départ de *heard*¹⁵² est acoustiquement très proche.

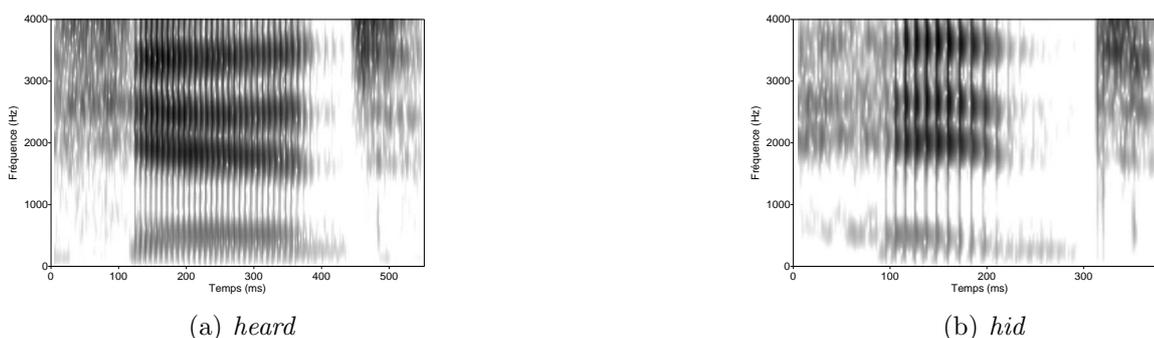


FIG. 6.112 – *heard* et *hid* produits par le locuteur *rwl* de *nwa*.

6.3.10.2 Les diphtongues

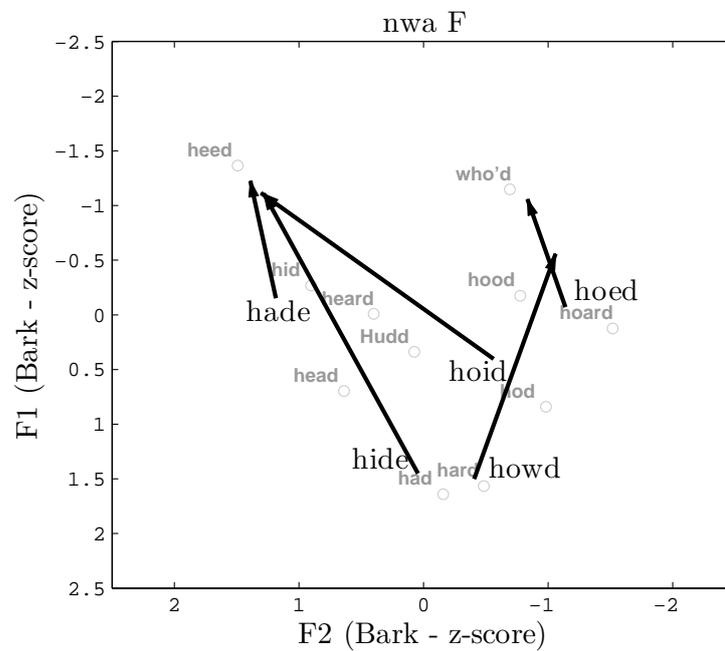
8 locutrices et 7 locuteurs ont produit les voyelles représentées dans les Figures 6.113 et 6.114. Comparativement à d'autres dialectes, on note une homogénéité remarquable entre les diphtongues des hommes et des femmes, excepté pour *hoed*. Pour ce dernier timbre, un certain nombre de réalisations monophthongales sont à noter ; c'est également le cas pour *hade*.

La voyelle de *hared* est une longue monophthongue dont le timbre est, la plupart du

¹⁵²Un faible mouvement des formants dans la Figure 6.112(a) confirme le caractère légèrement diphtongué de *heard* produit par *rwl*.

TAB. 6.15 – Durée des monophthongues de *nwa* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	231	63
hid	157	62
head	163	59
had	187	75
hard	294	87
hod	181	77
hoard	294	66
hood	163	50
who'd	263	78
Hudd	150	50
heard	278	82

FIG. 6.113 – Diphtongues de *nwa* pour les sujets féminins.

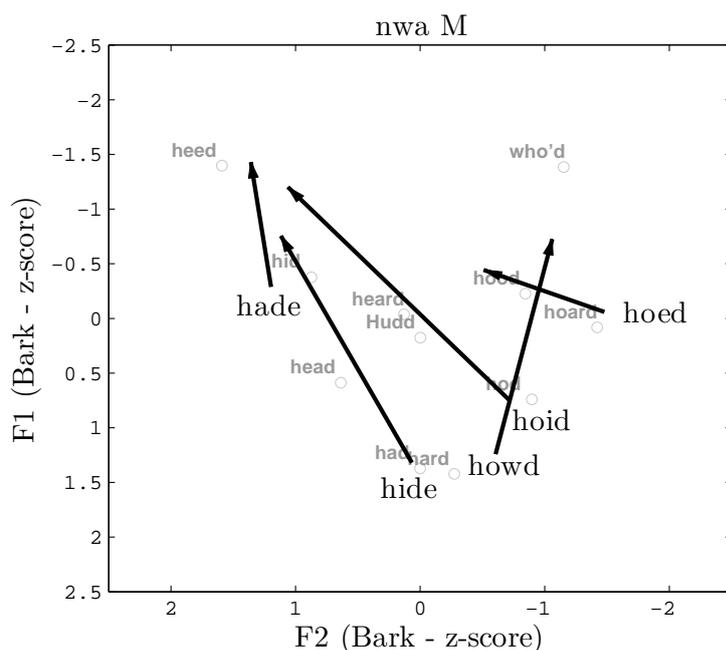


FIG. 6.114 – Diphtongues de *nwa* pour les sujets masculins.

temps, identique à celui de *head*. Les estimations des densités de probabilité (que nous n'avons pas reproduites ici) s'apparentent à celles que nous avons pu observer pour *brm* dans la Figure 6.47.

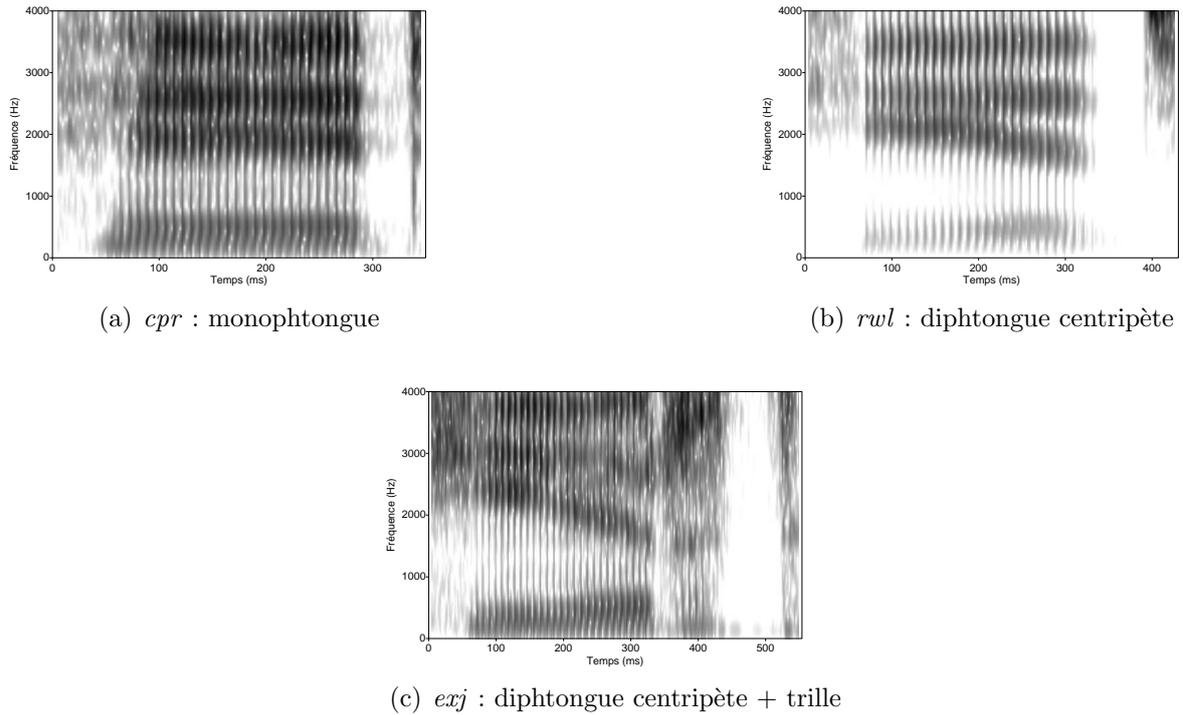
La voyelle de *heered* est, dans la majorité des cas, une diphtongue centripète. On dénombre quelques cas de monophthongues ainsi que de diphtongues suivies d'un /r/. La Figure 6.115 illustre ces trois variantes. La voyelle de *hured*, quant à elle, varie trop pour être décrite ici.

6.3.11 Republic of Ireland (*roi*)

6.3.11.1 Les monophthongues

Dans la Figure 6.116, 10 locutrices ont produit l'ensemble des 11 voyelles. Cette figure fait apparaître ;

- un chevauchement entre *hard* et *hod* ;
- un chevauchement entre *hood* et *Hudd*, qui tend à inclure *hoard* sur l'axe F2 ;

FIG. 6.115 – Spectrogrammes de *heered* produit par 3 locuteurs de *nwa*.

- le caractère relativement postérieur de *hood*, *Hudd* et *who'd*.

Dans la Figure 6.117, ce sont 7 locuteurs qui ont produit l'ensemble des 11 voyelles.

Par rapport à la Figure 6.116 :

- *Hudd*, *hood* et *hoard* sont plus proches ;
- *hard* est plus distinct de *hod*.

La faible distance entre *hard* et *hod* chez les femmes vient de ce que la voyelle de *hod* est très ouverte, ce qui indique que que celles-ci n'ont vraisemblablement pas été affectées par la fermeture des voyelles ouvertes postérieures caractéristique d'un changement récent à Dublin (voir Section 3.2.4.1). Les deux voyelles sont virtuellement superposées sur l'axe F1 alors que *hod* est plus postérieur. L'explication de cette différence acoustique sur F2 réside dans le fait que l'anglais de *roi* est rhotique. Les spectrogrammes des Figures 6.118(a) et 6.118(b) exhibent des valeurs formantiques initiales similaires ; puis on note sur la première l'abaissement de F3 caractéristique de la rhoticité, accompagné – et c'est ce

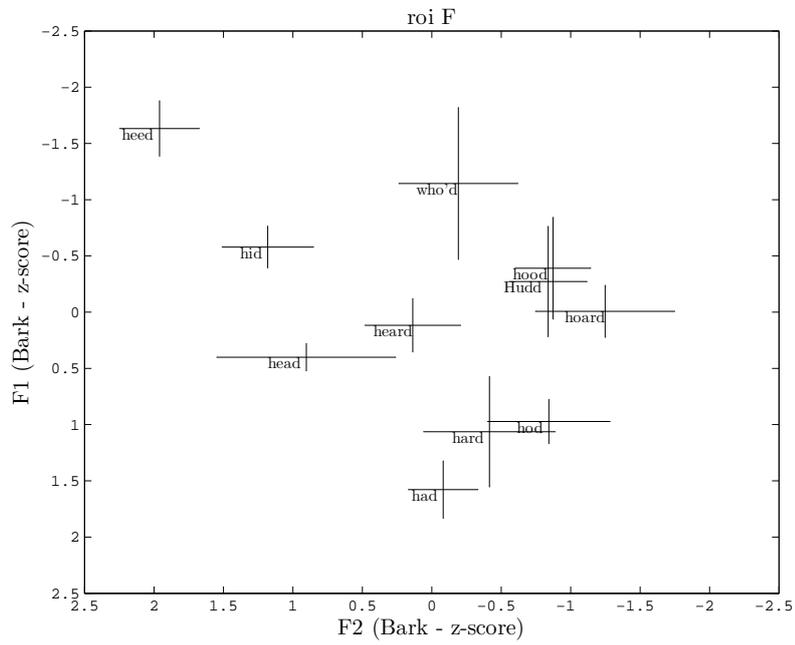


FIG. 6.116 – Monophthongues de *roi* pour les sujets féminins.

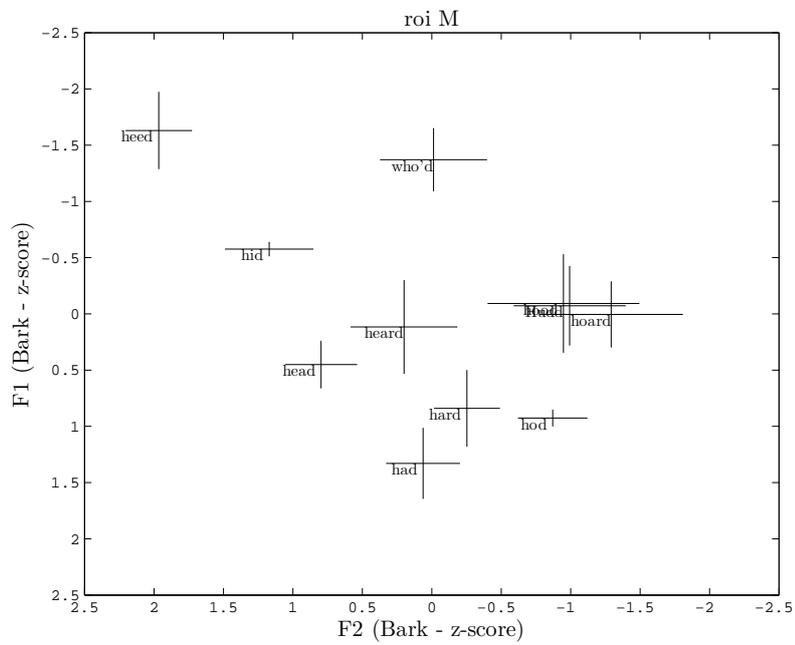


FIG. 6.117 – Monophthongues de *roi* pour les sujets masculins.

point qui introduit un biais – par une augmentation typique de F2 (voir e.g. Hagiwara [Hag95], 72). Quelques réalisations présentent une rhoticité qui colore la voyelle dès le début de son émission, à en juger par la valeur relativement basse de F3 (comparer les Figures 6.118(c) et 6.118(d)).

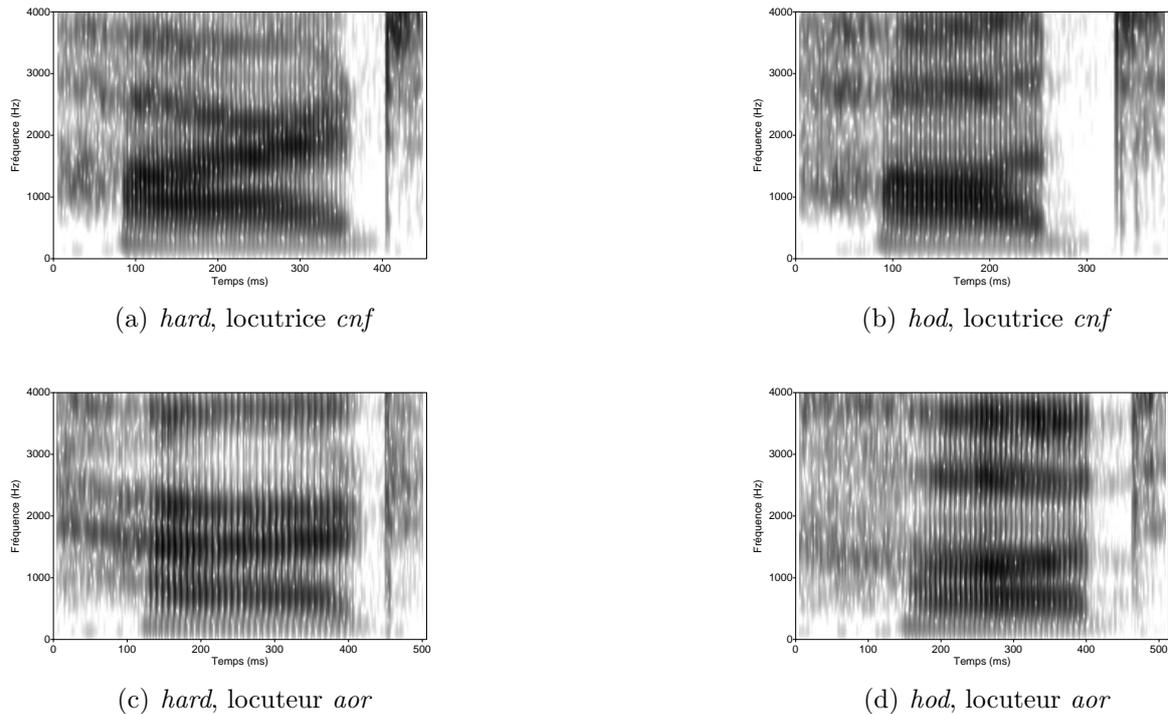


FIG. 6.118 – *hard* et *hod* produits par deux locuteurs de *roi*.

Concernant la proximité entre *Hudd* et *hood*, nous pouvons affirmer que l'absence de scission FOOT-STRUT constitue la norme dans notre échantillon de *roi*. D'après Hickey ([Hic04b], 91) ce trait est typique de « Popular Dublin », par opposition à « Fashionable Dublin », où la scission a eu lieu.

6.3.11.2 Les diphtongues

9 locutrices et 6 locuteurs ont produit les voyelles des Figures 6.119 et 6.120.

La mot-test *hared* a une réalisation rhotique. À l'écoute, le timbre de la partie strictement vocalique de *hared* (lorsqu'il n'est pas déjà coloré par l'approximante) et celui

TAB. 6.16 – Durée des monophthongues de *roi* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	225	53
hid	181	56
head	206	56
had	213	51
hard	275	61
hod	213	38
hoard	281	56
hood	181	54
who'd	248	68
Hudd	181	44
heard	250	56

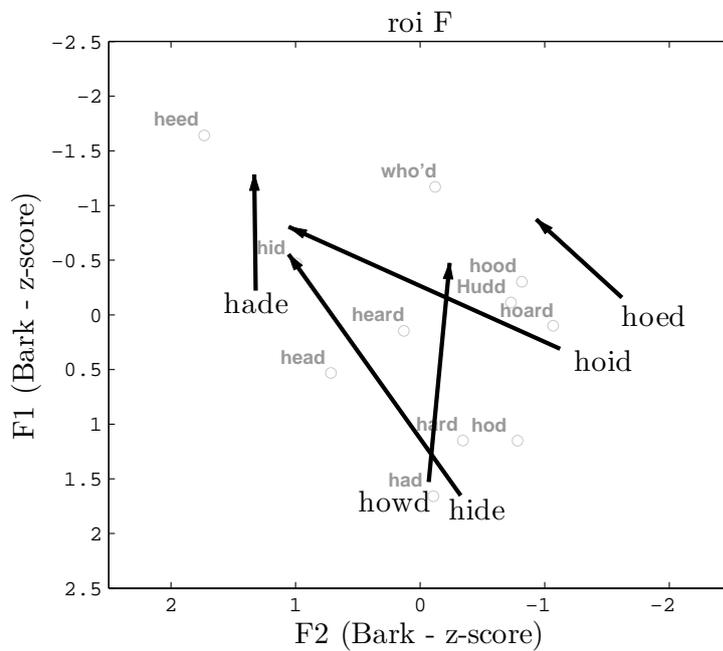
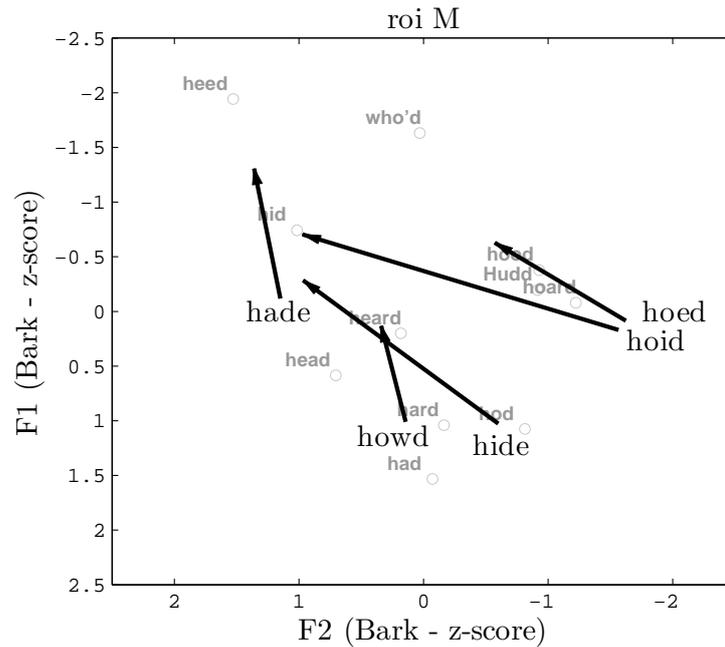


FIG. 6.119 – Diphtongues de *roi* pour les sujets féminins.

FIG. 6.120 – Diphtongues de *roi* pour les sujets masculins.

de *head* sont identiques. La voyelle de *heered* est systématiquement suivie d'un /r/. La voyelle de *hured*, suivie d'un /r/ également, varie entre les timbres [ju] et [u].

6.3.12 Scottish Highlands (*shl*)

6.3.12.1 Les monophthongues

Dans la Figure 6.121, 7 locutrices ont produit l'ensemble des 11 voyelles. On y relève :

- une qualité postériorisée pour *hid* ;
- le chevauchement de *had* et *hard* ;
- le degré d'aperture similaire de *hood* et *who'd* ;
- la dispersion importante de *Hudd* sur F2 et son degré d'aperture inférieur à celui de *heard* (contrairement à *sse*).

Dans la Figure 6.122, ce sont 11 locuteurs qui ont produit l'ensemble des 11 voyelles.

On relève la proximité entre :

- *Hudd* et *heard* ;

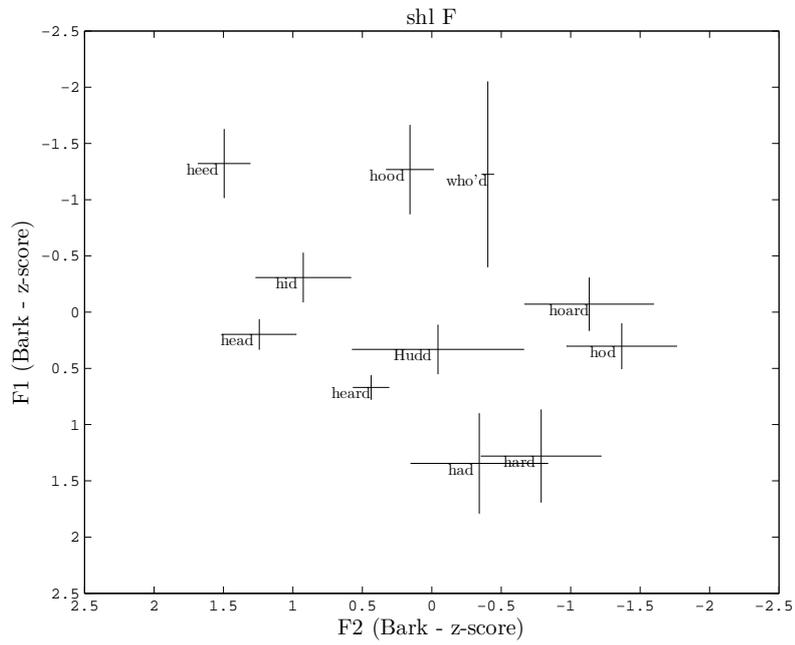


FIG. 6.121 – Monophthongues de *shl* pour les sujets féminins.

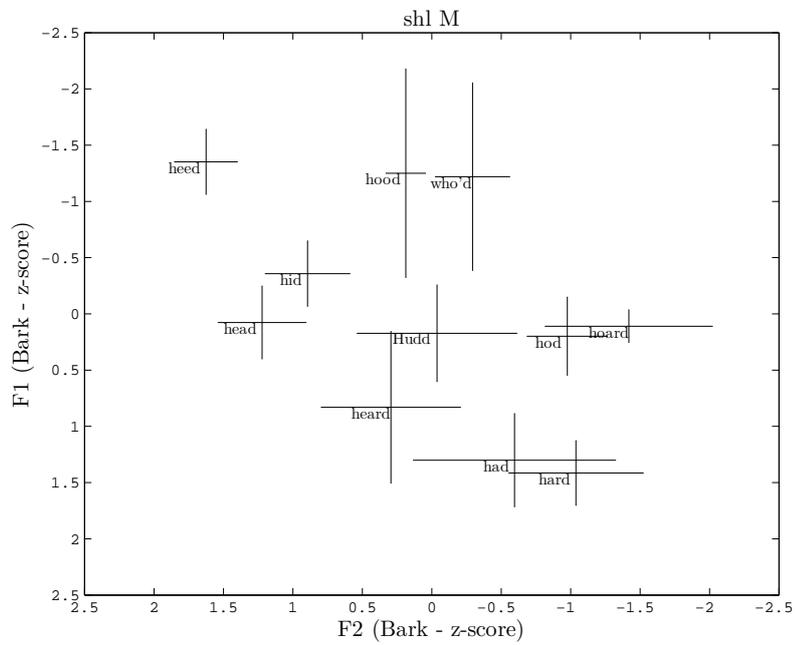


FIG. 6.122 – Monophthongues de *shl* pour les sujets masculins.

- *had* et *hard* ;
- *hod* et *hoard* ;
- *hood* et *who'd*.

Le degré de postériorisation de *hid* n'atteint pas celui que nous avons pu observer pour *gla*. Cette voyelle varie beaucoup d'un locuteur à l'autre, mais elle est néanmoins plus ouverte qu'en *sse* chez la plupart. La variante la plus courante est donc de type [e] ou [ɛ] (voir la Figure 6.123(a)). On trouve également, mais de façon sporadique, des réalisations proches de *sse* ([ɪ], voir Figure 6.123(b)) et des variantes semblables à la voyelle centralisée de *gla* (cf. Figure 6.78(a)) comme l'illustre la Figure 6.123(c).

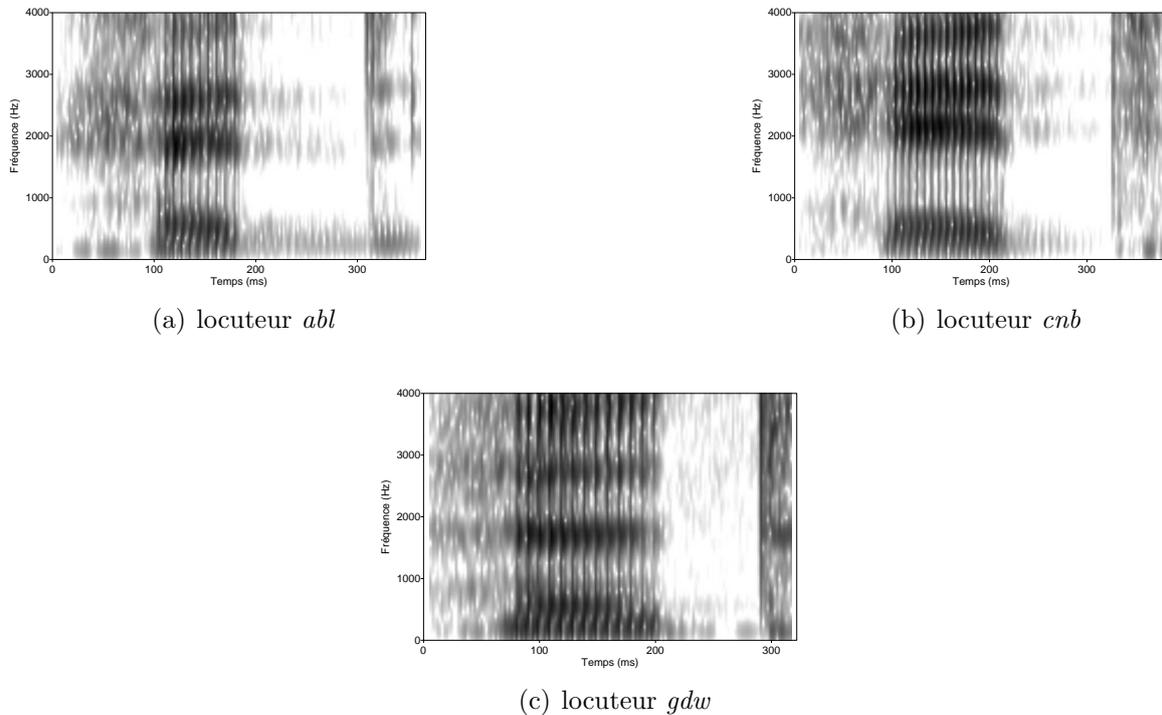
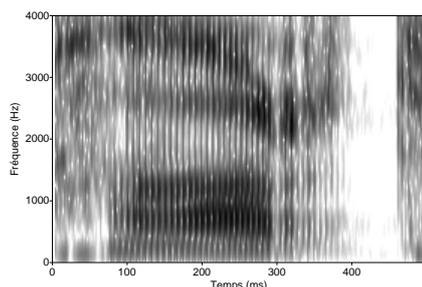


FIG. 6.123 – *hid* produit par 3 locuteurs de *shl*.

Voyons maintenant le cas de *had* et *hard*. La voyelle de *had* est en général très centrale : [ɛ], d'où une valeur de F2 relativement basse. Comme nous l'avons vu plus haut, le /r/ dans *hard* engendre une augmentation de F2 ; il s'ensuit donc que, même dans les cas où la voyelle a un timbre bien postérieur, l'utilisation de la médiane sur toute la durée

FIG. 6.124 – *hard* produit par le locuteur *jxw* de *shl*.

de la voyelle a été influencée par la rhoticité, d'où une proximité accrue entre *had* et *hard*. Incidemment, l'un des locuteurs produit une trille pour toutes ses réalisations de /r/ dans *hard* (voir Figure 6.124, succession de stries verticales marquées entre environ 300 et 400 ms).

Les timbres de *hod* et *hoard* sont tout à fait remarquables : ils sont très fermés, comme c'était déjà le cas pour *gla*. Si le /r/ permet de distinguer *hod* de *hoard*, il est vraisemblable que les deux timbres vocaliques soient identiques. Par exemple, la locutrice *ccm* de *shl* ne fait aucune différence de timbre dans la plupart de ses productions. En comparant les Figures 6.125(a) et 6.125(b) représentant des voyelles dont les valeurs moyennes sur la partie stable initiale sont de 465 et 876 Hz pour *hod* et 458 et 943 Hz pour *hoard*, il n'est pas absurde de supposer que nous avons peut-être affaire à la même voyelle. Le degré de fermeture de *hod* nous porte à croire que cette locutrice n'a pas d'opposition LOT-THOUGHT, trait typique mais très variable d'un locuteur à l'autre, de l'anglais d'Écosse (voir Section 3.2.3). La liste de mots-test dont nous disposons ne permet cependant pas de tester l'existence de cette convergence.

Les voyelles de *hood* et *who'd* sont, d'après notre analyse auditive, strictement identiques chez tous les locuteurs. En revanche, il se peut que certains appliquent une différence de longueur. La Table 6.17, dans laquelle sont regroupées les durées médianes des timbres, montre une faible différence entre *hood* et *who'd* (32 ms). Rappelons que l'anglais d'Écosse ne possède pas de quantité phonologique, contrairement à la plupart des autres dialectes, mais il existe une durée « quasi-phonémique » (voir [SHT99]) résultant de l'allongement

FIG. 6.125 – *hod* et *hoard* produits par la locutrice *ccm* de *shl*.TAB. 6.17 – Durée des monophthongues de *shl* (ms).

Voyelle	Durée médiane	Écart interquartile
heed	207	69
hid	203	62
head	272	75
had	263	87
hard	325	81
hod	269	82
hoard	325	69
hood	231	69
who'd	263	73
Hudd	204	68
heard	313	62

des voyelles dans certains contextes phonologiques ou morphémiques (voir Section 3.2.3).

6.3.12.2 Les diphtongues

Dans les Figures 6.126 et 6.127, 7 locutrices et 11 locuteurs sont représentés. Les diphtongues de *shl* sont très semblables à celles de l'autre dialecte écossais du corpus ABI, *gla*.

La voyelle de *hade* est unanimement monophthongale. La position de *hade* vis-à-vis de *hid* est inversée par rapport aux dialectes du nord de l'Angleterre (voir, e.g. *eyk*, Figures 6.71 et 6.70 ; ou *lan*, Figures 6.87 et 6.86). Il s'ensuit que le timbre de *hade* en *shl*

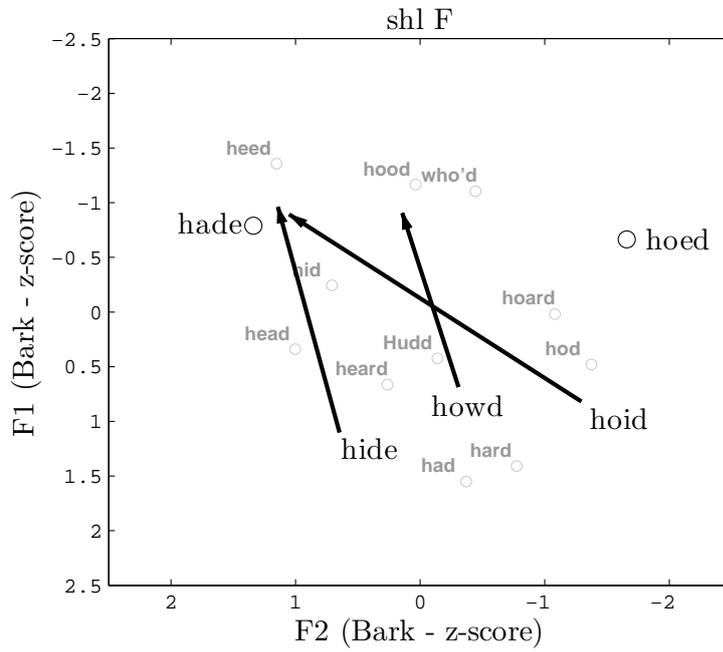


FIG. 6.126 – Diphtongues de *shl* pour les sujets féminins.

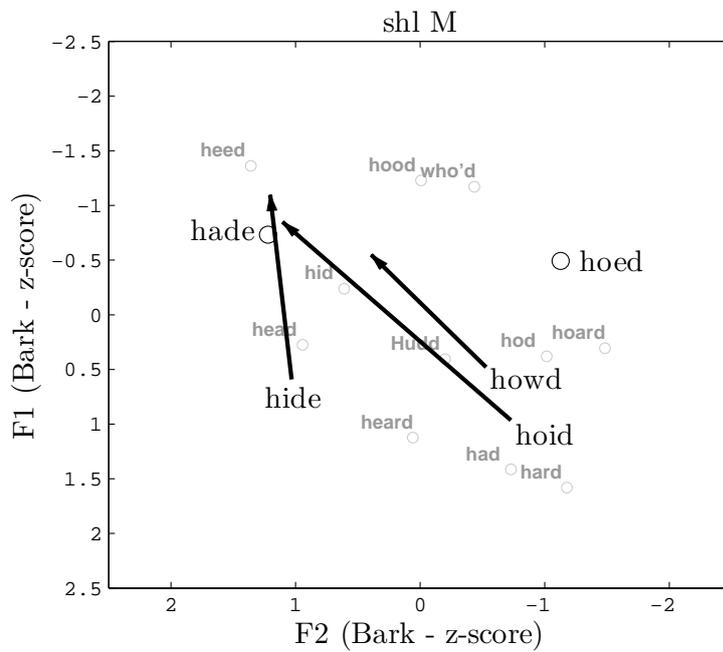
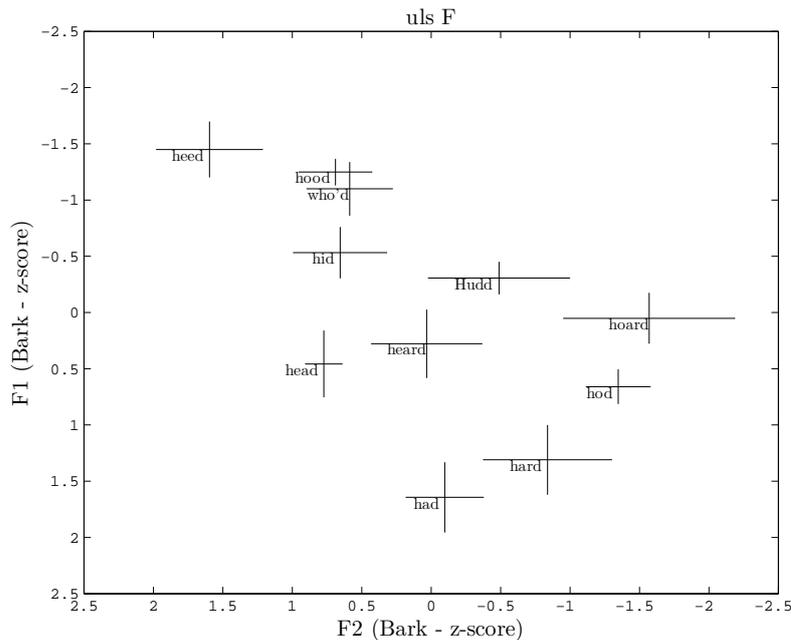


FIG. 6.127 – Diphtongues de *shl* pour les sujets masculins.

FIG. 6.128 – Monophtongues de *uls* pour les sujets féminins.

est très fermé.

Le dialecte *shl* étant rhotique, *hared* est donc produit avec un voyelle suivie d'une approximante. Le timbre de départ (i.e. avant le /r/) de *hared* est nettement plus fermé que celui de *head*, ce dernier donnant, à l'écoute, l'impression d'être plus fermé qu'en *sse*. On pourrait donc transcrire [hed] et [hɪɹd] pour *head* et *hared*, respectivement. La voyelle de *heered* suit le même schéma qu'en *gla*. Les variantes principales de *hured* sont une voyelle fermée, précédée ou non de [j], et toujours suivies d'un /r/ ; quelques réalisations de type [ɜ] apparaissent sporadiquement.

6.3.13 Ulster (*uls*)

6.3.13.1 Les monophtongues

Dans la Figure 6.128, 9 locutrices ont produit l'ensemble des 11 voyelles. Les points importants sont les suivants :

- comme pour *shl*, *hid* est plutôt rétracté ;
- *Hudd* est très fermé ;

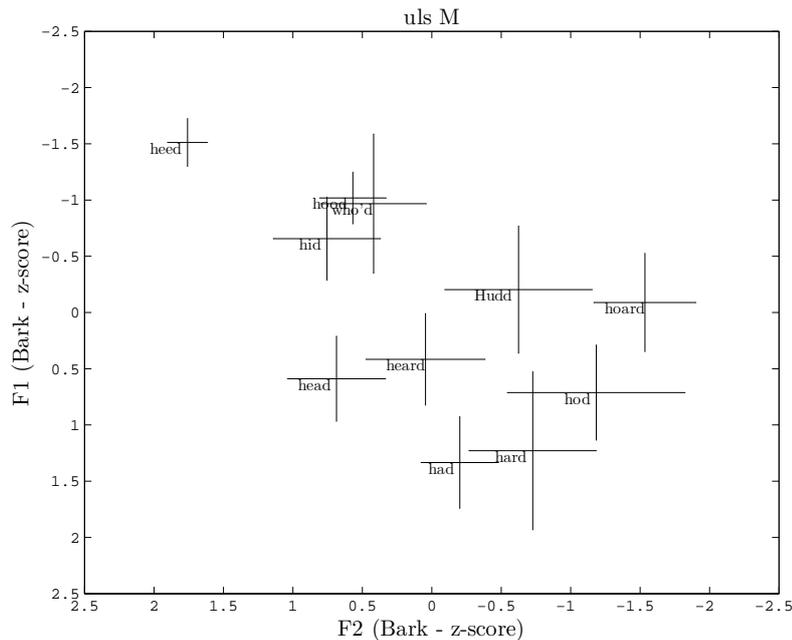


FIG. 6.129 – Monophthongues de *uls* pour les sujets masculins.

- *hood* et *who'd*, très antérieurs, se confondent.

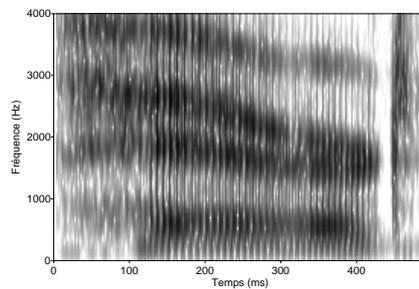
Dans la Figure 6.129, ce sont 6 locuteurs qui ont produit l'ensemble des 11 voyelles.

Par rapport à la Figure 6.128 ;

- globalement, le chevauchement entre timbres est plus important ;
- dans la partie supérieure de l'espace, *heed* est clairement isolé, alors que *hid*, *hood* et *who'd* sont regroupés ;
- *had* et *hard* sont très proches.

L'impression qui se dégage de l'écoute des occurrences de *hid* nous laisse penser que, malgré ce que suggère la Figure 6.128, cette voyelle, bien que plus ouverte qu'en *sse*, ne nous paraît pas particulièrement centralisée (sur l'axe F2). Chez les locuteurs masculins, les réalisations sont très variables, allant du timbre rétracté typique de *gla* et *shl* (Figure 6.130(a)) à une qualité plus antérieure et fermée (Figure 6.130(b)).

La voyelle de *heard* est également très variable : comme nous avons pu le rencontrer dans d'autres dialectes rhotiques, l'effet du [ɹ] peut soit affecter de manière uniforme la voyelle sur toute sa durée, soit intervenir en cours d'émission, ce qui se traduit notam-

FIG. 6.130 – *hid* produit par un locuteur et une locutrice de *uls*.FIG. 6.131 – *heard* produit par le locuteur *fod* de *uls*.

ment par un mouvement soudain de F3 vers le bas. Dans ce dernier cas, un bon nombre de locuteurs de *uls* ont un timbre de départ relativement antérieur, et parfois fermé (Figure 6.131).

En ce qui concerne *hood* et *who'd*, une homophonie totale chez une large majorité des locuteurs est incontestable. Le timbre de cette voyelle unique est très antérieur. La locutrice *mcw* n'a pas d'homophonie entre ces deux items. Le timbre plus postérieur et plus ouvert de sa réalisation de *hood* nous a poussé à comparer ce mot-test avec *Hudd* ; il ressort de cette comparaison que *mcw* a une homophonie entre *hood* et *Hudd*. Ceci signifie que soit *hood* fait partie de l'ensemble STRUT chez cette locutrice, soit son système est dépourvu de scission FOOT-STRUT. Notre corpus ne permet pas de trancher de façon fiable, mais les deux premières valeurs formantiques moyennes relevées sur l'état stable des voyelles de *looked* et *comes* fait apparaître un F2 différent (1530 et 1328 Hz, respectivement) pour un F1 virtuellement identique (695 et 702 Hz, respectivement). Deux locuteurs, *fod* et *mar*,

distinguent *Hudd*, *hood* et *who'd*. Des productions typiques de *fod* sont données dans la Figure 6.132. Notons que l'échantillon de *uls* semble n'être constitué que d'étudiants dans un « teacher training college » ; il n'est donc pas à exclure que *fod* et *mar* aient voulu adopter une norme de prononciation d'Angleterre.

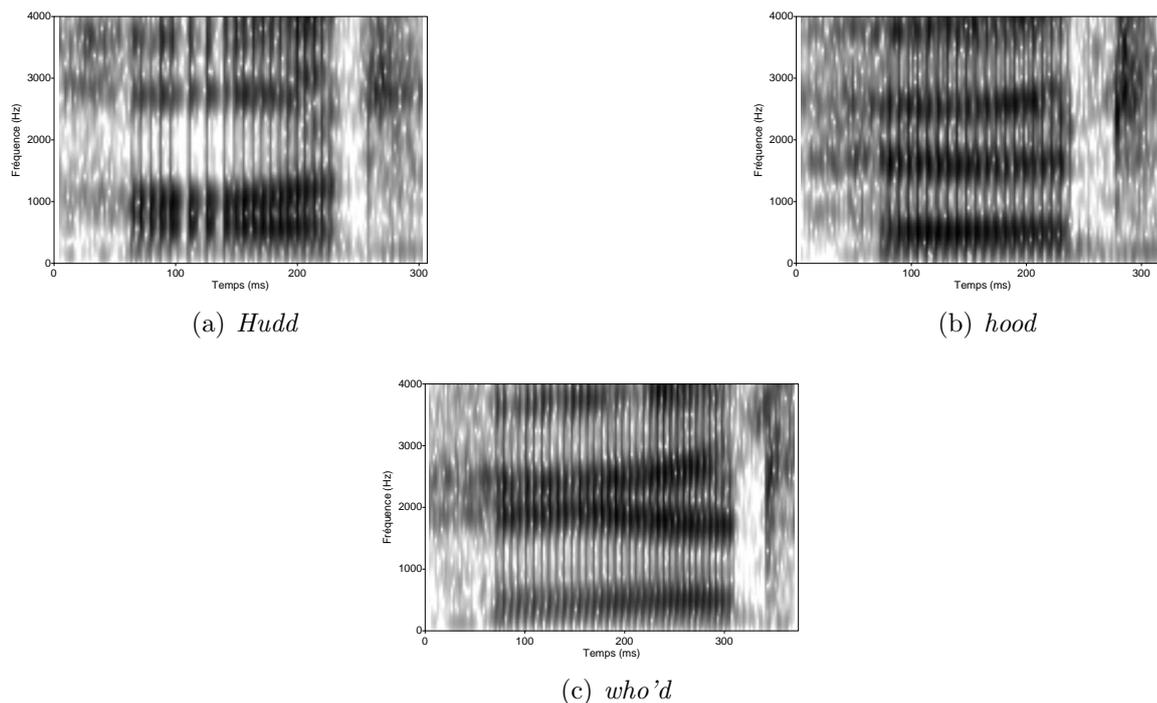
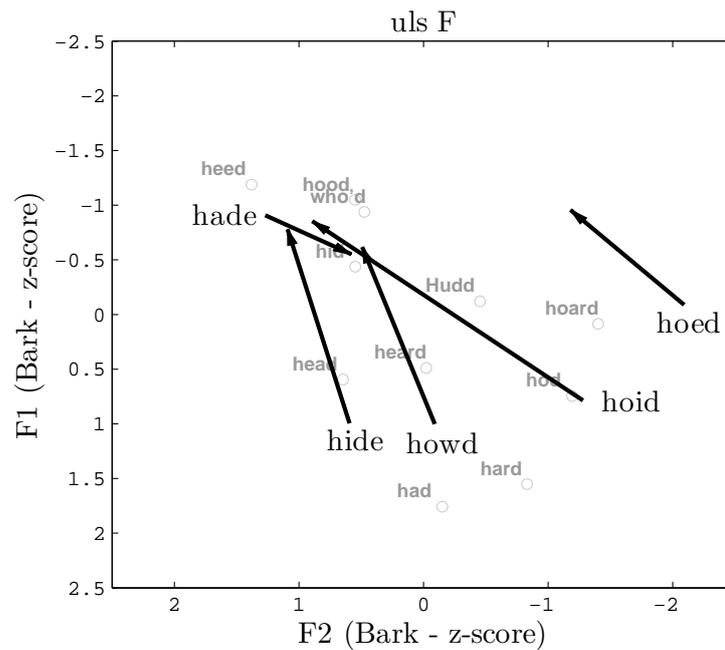


FIG. 6.132 – *Hudd*, *hood* et *who'd* produits par le locuteur *fod*.

La faible distance entre *had* et *hard* s'explique en partie par l'effet de [ɹ] sur le deuxième formant de la voyelle, comme nous avons déjà pu le constater dans d'autres dialectes. En règle générale, la voyelle de *hard* (avant qu'elle ne soit colorée par le [ɹ]) et celle de *had* ont des timbres différents. La majorité des réalisations de ce dernier item sont très clairement antérieures, contrairement à ce que nous avons pu noter pour *shl*. Quelques réalisations centralisées de cette voyelle ont été relevées, surtout chez les hommes. Seule la locutrice *mcw* semble avoir une identité de qualité vocalique entre ces deux timbres, identité qui serait totale (comme en *eyk* ou *nwa*) si *uls* n'était pas rhotique.

TAB. 6.18 – Durée des monophthongues de *uls*

Voyelle	Durée médiane	Écart interquartile
heed	231	82
hid	200	56
head	269	66
had	263	44
hard	312	67
hod	281	78
hoard	296	68
hood	231	85
who'd	263	69
Hudd	206	63
heard	300	62

FIG. 6.133 – Diphtongues de *uls* pour les sujets féminins.

6.3.13.2 Les diphtongues

9 locutrices et 5 locuteurs ont produit les voyelles des Figures 6.133 et 6.134. La majorité des occurrences de *hade* sont réalisées comme des diphtongues centripètes.

La voyelle de *hared* a deux réalisations majeures et une prononciation moins fréquente :

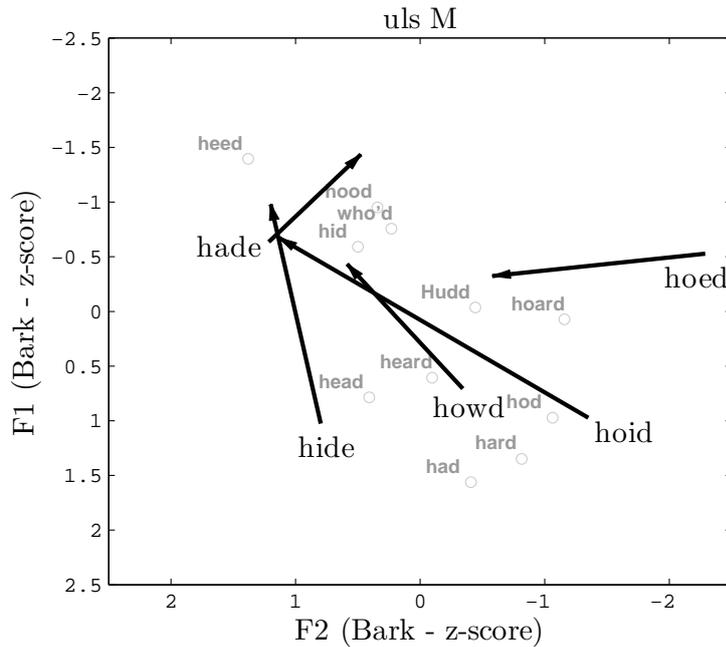


FIG. 6.134 – Diphtongues de *uls* pour les sujets masculins.

- une réalisation de type [ɛɪ] qui a le même timbre (avant rhoticité) que *head* chez les locuteurs qui la produisent ;
- une réalisation proche de [ɜɪ], typique des productions des femmes ;
- une production de type [iɪ] (e.g. locuteur *fmv*, locutrices *bmm*, *rho* et *yeq*).

Ces variantes peuvent être mises en relation avec la voyelle de *heard*. En écoutant attentivement les réalisations de ces deux timbres chez les 20 locuteurs et locutrices de *uls*, on constate qu'ils sont de parfaits homophones chez 14 sujets. Cette homophonie est cependant réalisée de manière différente selon le sujet :

- 9 d'entre eux, 6 hommes et 3 femmes, ont une réalisation de type [ɛɪ] ;
- 5 d'entre eux, 1 homme et 4 femmes, produisent un son de type [ɜɪ].

Parmi les 6 locuteurs restants, deux hommes ont un contraste cohérent [ɛɪ] vs [ɜɪ], pour *hared* et *heard*, respectivement. Un homme et deux femmes ont une homophonie, réalisée en [ɜɪ], sauf pour une occurrence de *heard*, qu'ils prononcent [iɪ]. Enfin, une locutrice, *bmm*, produit [ɜɪ] et [iɪ], tantôt pour *hared*, tantôt pour *heard*. Si l'on se fie aux réalisations de la majorité des locuteurs, il est possible d'inférer une convergence NURSE-SQUARE.

Les réalisations de *heered* sont du type de celles rencontrées pour *shl*. En ce qui concerne *hured*, le nombre important de variantes rend une analyse détaillée inutile.

6.4 Synthèse des résultats et discussion

6.4.1 Monophthongues

L'une des fins de notre étude est d'identifier les indices phonétiques qui permettraient de classer les locuteurs en fonction du dialecte. La synthèse des résultats que nous présentons ici est donc guidée par cet objectif. Il s'agit plus particulièrement de déterminer quelles sont les voyelles à fort potentiel discriminant. Nous discuterons également, plus généralement, les points qui ont émergé de la description des systèmes vocaliques.

Quantifier la variation dans un espace multidimensionnel implique le choix d'une mesure de distance. Les distances utilisées dans cette section sont toutes des distances euclidiennes (voir Équation 2.6) calculées tantôt à partir des coordonnées dans l'espace F1/F2 en Bark (valeurs centrées-réduites pour chaque locuteur, indépendamment pour F1 et F2), tantôt dans l'espace à trois dimensions incluant les deux précédentes et la durée. Dans ce cas, les durées ont également été centrées-réduites pour chaque locuteur, ceci étant motivé par la sensibilité de la distance euclidienne aux différences d'échelles entre les variables.

Afin d'apprécier visuellement le potentiel discriminant des monophthongues, celles-ci sont représentées dans la Figure 6.136 et les suivantes du même type. Pour chaque timbre, en séparant les hommes et les femmes, chaque dialecte est représenté par ses coordonnées médianes dans le plan F1/F2. Les axes de ces figures ont les mêmes dimensions, ce qui rend la dispersion des dialectes directement comparable d'un timbre à l'autre. Les limites des axes ont été déterminées en ajoutant et retranchant 1 écart-type aux coordonnées du barycentre du nuage de points visible dans ces figures. Ces limites ont ensuite été arrondies au dixième près afin de garantir une homogénéité dans la présentation. Seule la voyelle de *who'd* a nécessité un empan plus important, en l'occurrence, 1,3 écart-type a été ajouté et retranché aux coordonnées du barycentre pour permettre d'inclure la dispersion de tous

les points.

La Figure 6.135 représente la somme des distances entre les coordonnées médianes des 13 dialectes dans l'espace F1/F2. On y constate que, pour les hommes comme pour les femmes, les voyelles de *hood*, *Hudd* et *who'd* ont les valeurs les plus élevées, ce qui revient à dire que, en l'absence d'information complémentaire (e.g. la variation intra-dialectale), ces timbres jouissent d'un fort potentiel discriminant. La corrélation entre les sommes de distances par timbre relevées chez les femmes et chez les hommes est élevée : que l'information de la différence entre timbres soit préservée (coefficient de Pearson) ou non (coefficient de Spearman, i.e. seul l'ordre est pris en compte), les coefficients de corrélation arrondis au dixième près sont de $r = 0,9$. Pour plus de détails, abordons à présent chaque voyelle individuellement.

La densité des nuages de points dans la Figure 6.136 confirme le faible potentiel discriminant (si l'on s'en tient à F1 et F2, et à la somme des distances entre les coordonnées médianes de tous les dialectes) de *heed*. Dans la Figure 6.137, on retrouve le caractère centralisé de *hid*, qui est particulièrement marqué pour *gla*, et un peu moins pour *uls* et *shl*. À l'opposé, le degré de fermeture élevé pour *brm*, déjà noté dans la Section 3.2.1.3, apparaît clairement. La voyelle de *head*, représentée dans la Figure 6.138, illustre l'aspect antérieur et fermé de cette voyelle pour *gla* et *shl*, qui est corrélé à la qualité centrale de *hid* : en effet, sur le plan systémique, la zone du [ɪ] n'étant pas occupée par KIT, DRESS peut donc s'y fixer. La voyelle de *had* (Figure 6.139), dont le potentiel discriminant, tel qu'il est mesuré dans la Figure 6.135, est faible, montre une densité importante du nuage de points avec, surtout chez les hommes, un extremum postérieur pour *shl*. En ce qui concerne *hard*, il semble possible de considérer, au moins visuellement, qu'un groupe constitué de *eyk*, *nwa* et *roi* se distingue par des valeurs relativement antérieures, aussi bien pour les hommes que pour les femmes, dans la Figure 6.140. Nous avons noté dans les Sections 6.3.5.1 et 6.3.10.1 que les timbres de *had* et *hard* étaient, en excluant la durée, strictement identiques pour beaucoup de locuteurs de *eyk* et *nwa*. Quant à la présence de *roi* dans cette zone antérieure, on peut probablement invoquer un artefact dû

à la rhoticité très marquée, engendrant un F2 plus élevé. La voyelle de *hod*, représentée dans la Figure 6.141, a un potentiel discriminant intermédiaire, voire faible, d'après la Figure 6.135. On remarque par exemple le caractère fermé de ce timbre pour les deux dialectes d'Écosse, et son timbre ouvert pour *lvp* et *roi*. La voyelle de *hoard* (Figure 6.142) possède un potentiel discriminant du même ordre que celle de *hod*. La Figure 6.142(b) fait apparaître la position fermée et postérieure de *ean* et *sse*. Si l'on replace ce commentaire dans le cadre du mouvement des systèmes vocaliques dans le sens inverse des aiguilles d'une montre, on peut affirmer que *ean* et *sse* sont plus avancés que les autres dialectes. En ce qui concerne *hood*, la Figure 6.143 confirme le score élevé de cette voyelle constaté dans la Figure 6.135. Si l'on s'arrête plus particulièrement sur la Figure 6.143(b), on peut distinguer :

- un groupe composé de *gla*, *shl* et *uls* présentant un timbre très antérieur et fermé. Ces trois dialectes, qu'il est possible de désigner, pour des raisons géographiques et linguistiques historiques, par « Scots », doivent en partie la qualité de leur voyelle dans *hood* au fait que FOOT et GOOSE ne sont pas distincts ;
- un groupe composé de *crn*, *ean* et *sse* formant une entité géographique cohérente : le sud de l'Angleterre ;
- un groupe formé par *eyk* et *ncl* ; ici encore, une certaine cohérence géographique émerge : ce sont tous les deux des dialectes du nord-est de l'Angleterre ;
- un groupe rassemblant *brm*, *lan* et *nwa*. Si la distance linguistique connue entre ces dialectes est plus grande que pour les trois groupes que nous venons de commenter, il n'en demeure pas moins qu'ils forment un tout représentant l'ouest de la Grande-Bretagne ;
- un dernier groupe composé de *lvp* et *roi* ; la proximité géographique et surtout l'histoire des migrations viennent justifier ce regroupement.

La voyelle de *who'd* a nécessité un empan de valeurs plus grand que les autres timbres pour être représentée dans la Figure 6.144. À la différence de la voyelle de *hood*, il est malaisé de regrouper les points de la Figure 6.144 de façon cohérente ; on note en revanche

que la variabilité se manifeste principalement sur l'axe F2. La voyelle de *Hudd*, quant à elle, représente un intérêt particulier puisqu'elle fait écho à l'isoglosse traditionnelle qui démarque le sud du nord linguistiques en Angleterre. Sur un plan systémique, le timbre de *Hudd* devrait permettre de distinguer les dialectes ayant subi la scission FOOT-STRUT des autres ; et c'est précisément ce qui se produit dans la Figure 6.145. En effet, on peut considérer que la diagonale qui relie l'angle supérieur gauche à l'angle inférieur droit de l'espace vocalique représenté constitue une démarcation entre ces deux types de dialectes. On remarque l'incohérence du comportement de *brm* entre les hommes et les femmes, ce qui reflète bien qu'il s'agit d'une zone intermédiaire. D'ailleurs, si l'on se fie à ces valeurs, la Figure 6.145 va à l'encontre du principe selon lequel les femmes tendent vers la variété standard d'une langue¹⁵³ ; en effet, leurs productions s'approchent davantage des dialectes ayant un système qui ne distingue pas FOOT et STRUT ; les productions des hommes tendant vers un timbre plus « méridional », linguistiquement parlant. Enfin, *heard*, qui occupe la quatrième position dans le classement des voyelles à fort potentiel discriminant, est représenté dans la Figure 6.146. La qualité fermée et antérieure de ce timbre pour *lvp* et *ncl* confirme l'analyse faite dans les Sections 6.3.8.1 et 6.3.9.1. Le caractère fermé de *heard* en *nwa* a également été constaté dans la Section 6.3.10.1.

L'estimation du potentiel discriminant d'une voyelle par le biais de la somme des distances entre les coordonnées médianes de chaque dialecte dans F1/F2 est imparfaite dans la mesure où elle ne fournit aucune indication concernant la variation des distances entre dialectes. On peut en effet imaginer deux situations schématiquement très différentes :

1. les 13 dialectes sont approximativement équidistants ; la voyelle concernée a donc un potentiel discriminant pertinent dans une tâche de classification à 13 classes ;
2. 12 dialectes sont très proches et le treizième est particulièrement éloigné de tous les autres ; le potentiel discriminant de la voyelle est très faible pour une tâche à 13

¹⁵³En réalité, Foulkes et Docherty ([FD99], 16) notent que dans le domaine des attitudes spécifiques au genre (« gender ») vis-à-vis des diverses normes de prononciation, le point de vue traditionnel selon lequel les femmes adoptent plus volontiers une prononciation standard n'est plus vraiment valide. En effet, la dichotomie standard *vs* non standard s'est transformée en local *vs* non local : les femmes utilisent moins de variantes locales que les hommes.

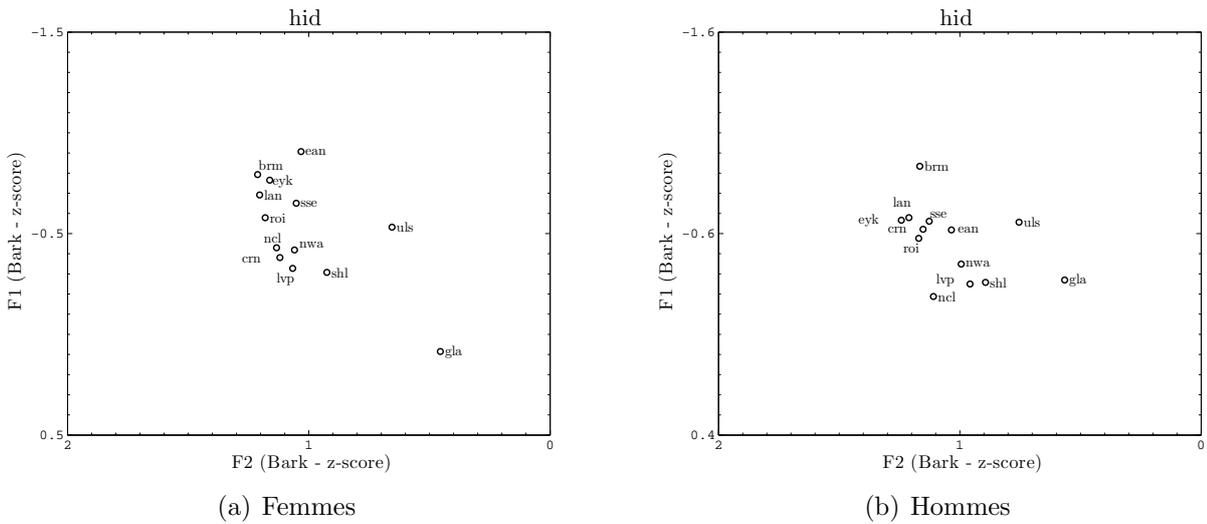


FIG. 6.137 – Synthèse des valeurs de F1/F2 pour la voyelle de *hid*.

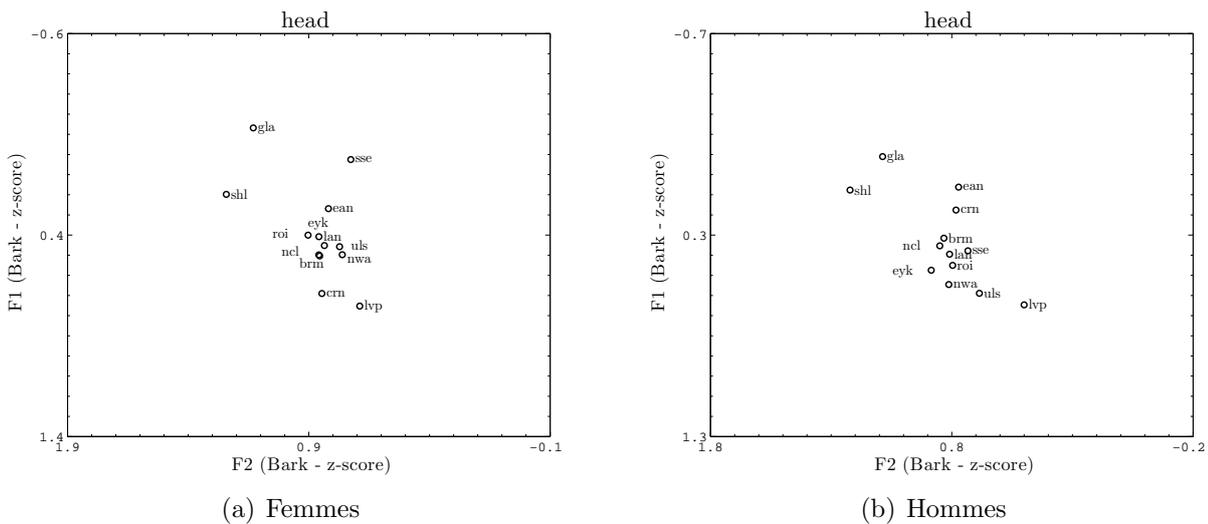
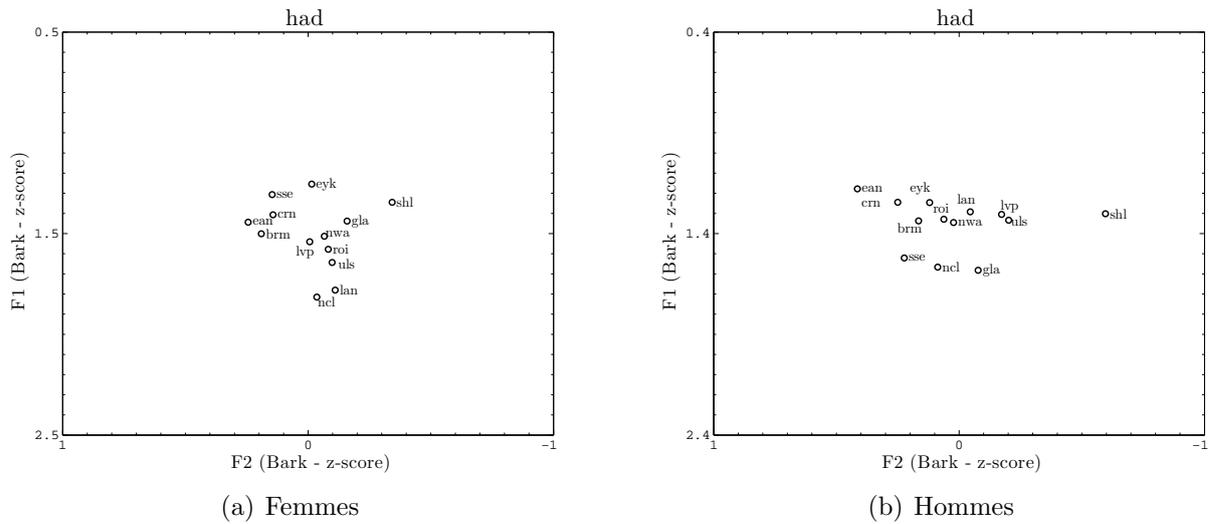
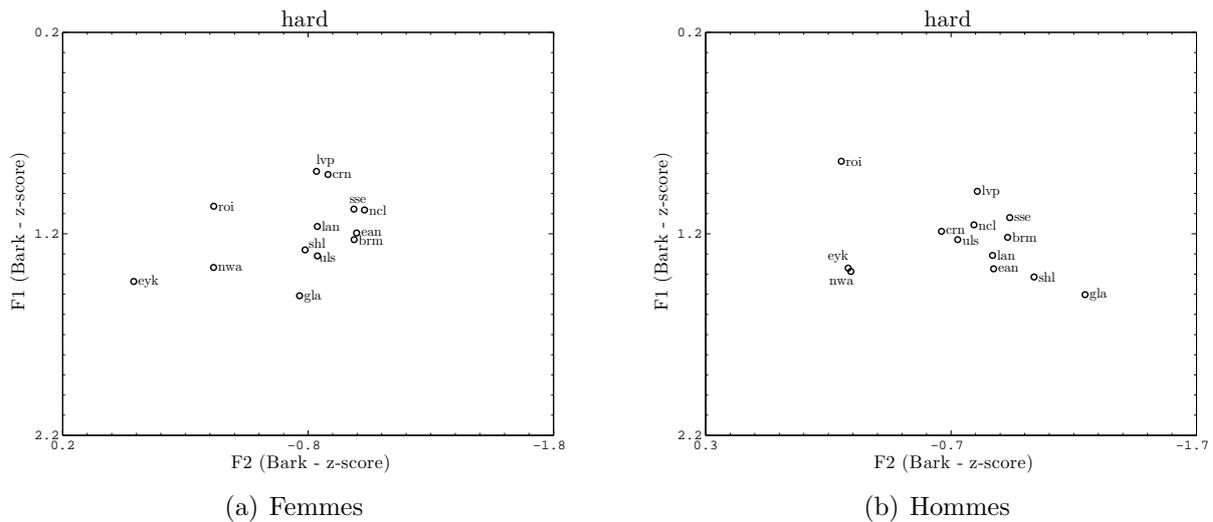


FIG. 6.138 – Synthèse des valeurs de F1/F2 pour la voyelle de *head*.

deux variables, ce qui laisse supposer que le deuxième cas exposé dans l'énumération qui précède ne semble pas réellement se produire (il aurait été matérialisé par une voyelle dans le coin supérieur gauche).

Avant d'inclure la durée comme variable supplémentaire dans le calcul des distances, il est instructif de comparer indépendamment l'apport de cette variable au potentiel discriminant. La Figure 6.148 est une représentation des 11 monophthongues dans l'espace

FIG. 6.139 – Synthèse des valeurs de F1/F2 pour la voyelle de *had*.FIG. 6.140 – Synthèse des valeurs de F1/F2 pour la voyelle de *hard*.

des différences de durée et des distances dans F1/F2. Il est intéressant de constater que, loin de concourir à une bonne discrimination entre les dialectes, ces deux variables, dans leur contribution au potentiel discriminant, semblent mutuellement exclusives, notamment pour les hommes. En effet, si on examine la Figure 6.148(b), on constate que la variation mesurée par la somme des distances s'exprime, selon le timbre, plutôt dans la dimension de la durée, ou plutôt dans la dimension F1/F2, ce qui est traduit par l'absence de voyelles

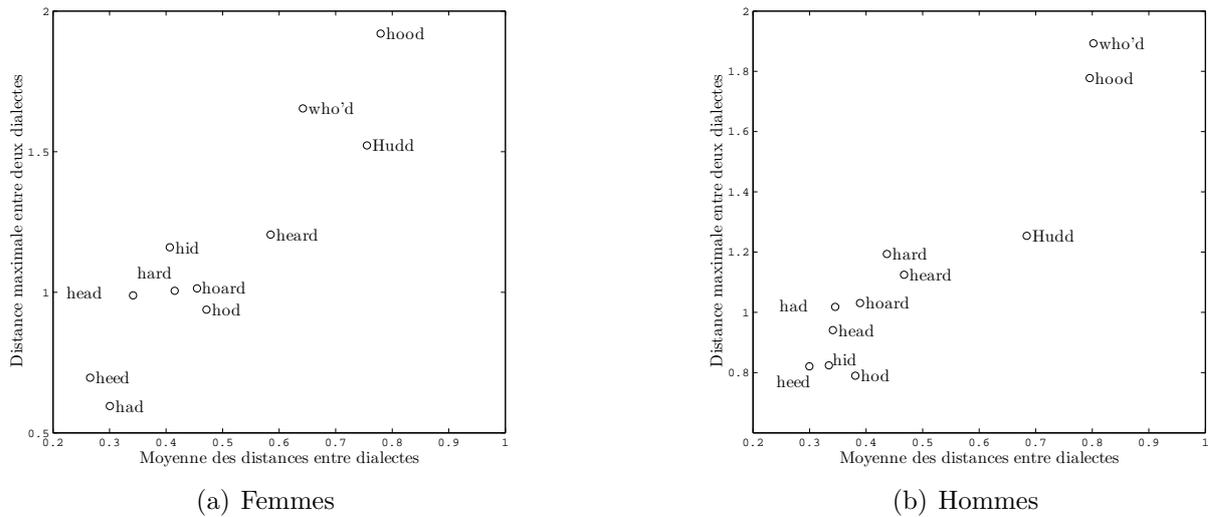


FIG. 6.147 – Moyennes et maxima des distances entre dialectes dans F1/F2.

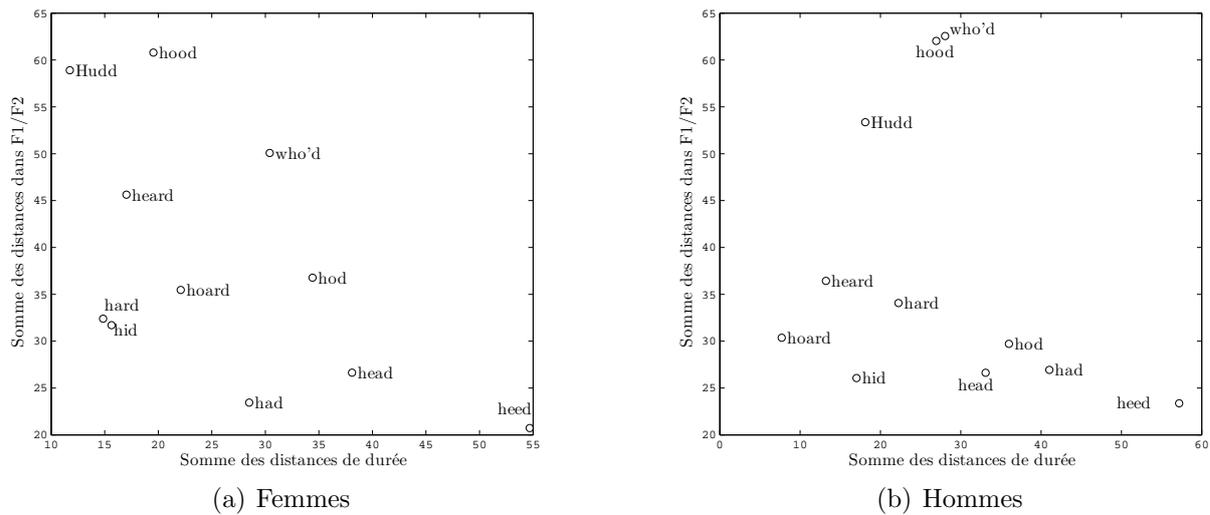


FIG. 6.148 – Somme des différences de durée et des distances dans F1/F2.

de chaque locuteur, indépendamment pour chaque formant ; en réalité, ceci revient à calculer une ANOVA univariée à un facteur (le dialecte). Afin de conserver une certaine cohérence avec la représentation canonique de l'espace vocalique, F2 est en abscisses, et F1 en ordonnées. Chez les hommes comme chez les femmes, on remarque que :

- la voyelle de *hood* possède le potentiel discriminant le plus élevé, qui se manifeste par une valeur élevée de F aussi bien sur F1 que sur F2 ; la contribution des deux

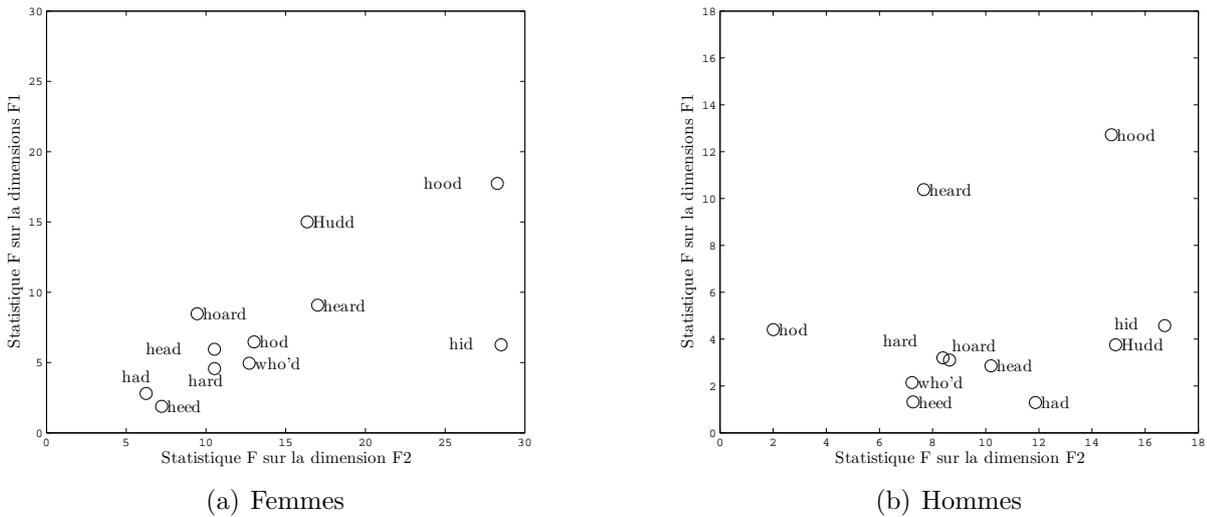


FIG. 6.149 – Statistique F de l'ANOVA calculée sur F1 et F2 indépendamment.

dimensions au potentiel discriminant de *hood* est apparente dans la Figure 6.143 ;

- la voyelle de *hid* a le potentiel discriminant le plus élevé sur F2, ceci permet d'affirmer, en référence à la dispersion approximativement équivalente sur les deux dimensions de la Figure 6.137, que la variabilité intra-catégorielle est plus importante sur F1 que sur F2 ; on peut donc conclure que c'est le degré d'antériorisation/postériorisation de *hid* qui est pertinent pour la classification des dialectes ;

Il est intéressant de noter la différence de magnitude des valeurs de F entre les hommes – $F_{max} \approx 13$ pour F1 et $F_{max} \approx 17$ pour F2 – et les femmes – $F_{max} \approx 18$ pour F1 et $F_{max} \approx 29$ pour F2. Cette différence n'est pas la conséquence de magnitudes différentes des fréquences de formants puisque ces dernières ont été centrées-réduites. Il semblerait donc que, à partir des deux premiers formants vocaliques, il soit plus facile de distinguer les dialectes des femmes de ceux des hommes.

Plus spécifiquement, dans la Figure 6.149(a), on constate que la statistique F est généralement plus élevée sur la dimension F2. En revanche, elle est à peu près équivalente sur les deux dimensions pour *Hudd*, ce qui s'accorde parfaitement avec le fait que, dans la Figure 6.145(a), les dialectes ayant la scission FOOT-STRUT soient séparés des autres par une diagonale. Curieusement, nous ne retrouvons pas ce schéma dans la Figure 6.149(b) :

la voyelle de *Hudd* possède un potentiel discriminant approximativement trois fois plus élevé sur F2 que sur F1. En observant la Figure 6.145(b), qui ne fait apparaître que la variation inter-dialectale, la dispersion semble approximativement identique dans les deux dimensions. On peut donc conclure que l'écart important entre les valeurs de F pour F1 et F2 (Figure 6.149(b)) est imputable à une variabilité intra-catégorielle sur F1 plus importante chez les hommes. Nous supposons que cette variabilité est en partie liée à du bruit (variation aléatoire, et peut-être erreurs de mesure), mais également à de véritables différences systémiques intra-dialectales (i.e. dans un même dialecte, certains sujets ont la scission FOOT-STRUT, d'autres non).

Si l'on s'en tient à la mesure du potentiel discriminant tel qu'il peut être calculé via la statistique F (Figure 6.149), on peut conclure que de F1 et F2, c'est cette dernière dimension qui est en moyenne la plus pertinente pour la classification des dialectes. Si l'on trace la diagonale descendante dans les Figures 6.149(a) et 6.149(b), on remarque, chez les hommes comme chez les femmes, que les voyelles de *hid*, *hood* et *Hudd* sont dans le triangle supérieur ; il est donc possible de conclure que ces trois voyelles (en ajoutant peut-être *heard*) sont les timbres qui possèdent le plus fort potentiel discriminant.

Everitt *et al.* ([ELL01], 46-47) recommandent l'utilisation de la distance de Mahalanobis D^2 entre deux groupes, qui est donnée par l'Équation 6.5, où \bar{x}_A et \bar{x}_B sont les barycentres des groupes A et B , respectivement, et W est la somme des matrices de covariance intra-groupes. La valeur de D^2 croît avec une distance croissante entre \bar{x}_A et \bar{x}_B et une dispersion décroissante autour de \bar{x}_A et de \bar{x}_B . Everitt et collègues ([ELL01], 46-47) signalent cependant que cette mesure n'est valable que lorsque les matrices de covariance des deux groupes sont approximativement identiques. Pour s'affranchir de cette contrainte, les auteurs suggèrent la formule de l'Équation 6.6, qui permet de calculer le « normal information radius » (NIR). Le NIR peut-être considéré, d'après [ELL01], 47, comme une généralisation de D^2 pour les cas où les matrices de covariance (W_A et W_B) sont hétérogènes ; la relation entre D^2 et NIR dans le cas de matrices de covariances équivalentes est explicitée dans l'Équation 6.7.

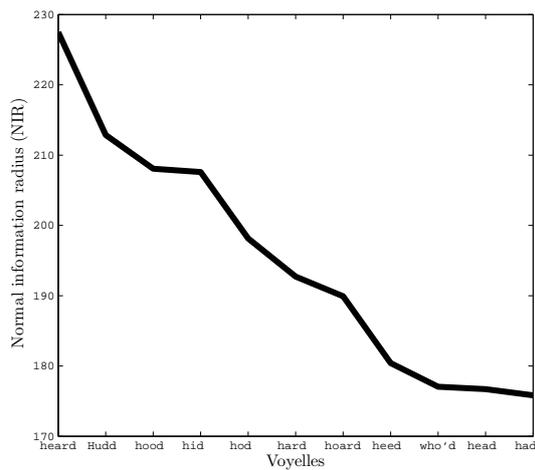
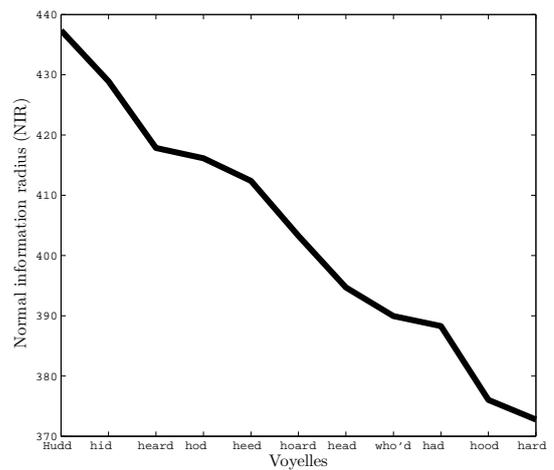
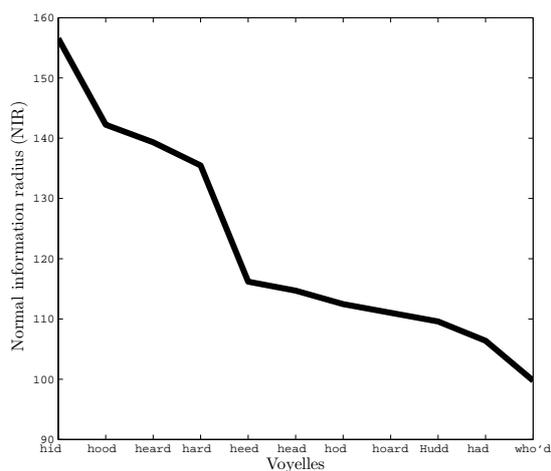
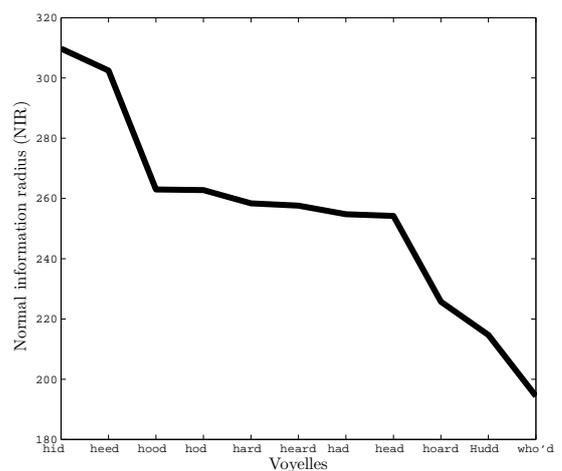
$$D^2 = (\bar{x}_A - \bar{x}_B)'W^{-1}(\bar{x}_A - \bar{x}_B) \quad (6.5)$$

$$NIR = \frac{1}{2} \log_2 \left(\frac{\det\{\frac{1}{2}(W_A + W_B)\} + \frac{1}{4}(\bar{x}_A - \bar{x}_B)'(\bar{x}_A - \bar{x}_B)}{\det(W_A)^{1/2}\det(W_B)^{1/2}} \right) \quad (6.6)$$

$$NIR = \frac{1}{2} \log_2 \left(1 + \frac{1}{4} D^2 \right) \quad (6.7)$$

Dans la Figure 6.150(a), le *NIR* a été calculé à partir de F1 et F2 pour chaque voyelle chez les femmes. La courbe confirme le potentiel discriminant supérieur de *heard*, *Hudd*, *hood* et *hid*, que nous avons déjà noté dans les représentations précédentes. En revanche, contrairement à ce que pouvaient laisser croire la Figure 6.144, mais conformément à la Figure 6.149, le potentiel discriminant de *who'd* est plutôt faible. Cependant, l'absence de « coude » dans la courbe signale qu'il n'existe pas véritablement de seuil délimitant les voyelles à fort potentiel discriminant des autres. L'ajout de la durée au calcul du *NIR* (Figure 6.150(b)) confirme le potentiel discriminant de *Hudd*, *hid* et *heard*. Par contre, on assiste à une chute notable de la valeur de *NIR* pour *hood*. Ceci soulève la question de la pondération des variables : un déplacement dans la dimension F1 ou F2 est-il perceptivement équivalent à un déplacement de même amplitude dans la dimension de la durée ? De plus, on peut considérer que la pertinence de la durée varie d'un timbre à l'autre. Par exemple, pour certaines voyelles dans les systèmes pour lesquels la durée est phonologiquement pertinente, on peut inférer que cette variable a une distribution linguistiquement conditionnée ; produire une durée trop déviante du prototype accroît les risques de confusion. Mais dans le cas où la durée n'a qu'un rôle marginal, sa variation possède un degré de liberté supérieur : sa déviance par rapport à une durée prototypique ne peut engendrer aucune confusion. La Figure 6.151 représente la valeur du *NIR* pour les hommes, avec et sans la durée, dans les Figures 6.151(b) et 6.151(a), respectivement. Les dialectes *crn* et *ncl* n'ont pas été inclus pour les hommes car le faible nombre de locuteurs disponibles dans ces deux dialectes ne permet pas de calculer la valeur de *NIR* convenablement ; les

Figures 6.150 et 6.151 ne sont donc pas directement comparables. Cette dernière représentation semble confirmer le potentiel discriminant de *hid*, *hood* et *heard* ; en revanche, *Hudd* a une valeur de *NIR* dans F1/F2 comparativement faibles. L'inclusion de la durée dans la Figure 6.151(b) fait émerger *heed* comme voyelle discriminante. De manière générale,, ces graphiques confirment le faible potentiel discriminant de *who'd*.

(a) *NIR* F1/F2(b) *NIR* F1/F2/duréeFIG. 6.150 – *NIR* calculé sur F1/F2 et F1/F2/durée chez les femmes.(a) *NIR* F1/F2(b) *NIR* F1/F2/duréeFIG. 6.151 – *NIR* calculé sur F1/F2 et F1/F2/durée chez les hommes.

Il nous est difficile de statuer sur la réalité phonologique des différences physiques observées entre deux timbres. Le premier obstacle réside dans l'impossibilité de déterminer *a priori*, bien qu'il existe de fortes tendances, à quel ensemble lexical un mot-test appartient. Par exemple, comme nous avons pu le constater pour *ncl*, l'homophonie entre *hood* et *who'd* nous laissait devant l'alternative suivante :

- soit *hood* et *who'd* sont de bons représentants (prototypes) des ensembles lexicaux FOOT et GOOSE, auquel cas, leur homophonie permet d'inférer une convergence des deux ensembles ;
- soit, contrairement à ce qu'on observe dans la majorité des autres dialectes, les mots-test *hood* et *who'd* font partie du même ensemble lexical, quel qu'il soit, alors qu'en réalité des paires minimales supplémentaires auraient permis d'attester la distinction entre les ensembles FOOT et GOOSE.

Un deuxième problème consiste à déterminer à partir de quel degré de dissimilitude (telle qu'elle est mesurée sur les paramètres physiques) on peut considérer que deux mots-test constituent une paire minimale, en l'absence d'étude de phonétique perceptive. Si l'on pose, comme nous pensons l'avoir démontré (Section 2.3.2), que les listes de mots en /hVd/ engendrent des productions qui reposent, du fait de la faible redondance linguistique, sur un contraste acoustique maximal et, par conséquent, des réalisations prototypiques, nous pouvons affirmer que le faible (ou l'absence de) chevauchement dans l'espace F1/F2/durée entre les réalisations de deux timbres en fait de bons candidats à une opposition phonologique.

Cette approche phonétique de la phonologie, que nous avons soutenue tout au long de ce travail, reçoit un appui supplémentaire illustré dans la Figure 6.152, qui représente le système des monophthongues de *ean*. Cette figure représente un arbre de classification généré par la fonction `classregtree` de Matlab. Les valeurs de F1 et F2 sont centrées-réduites (indépendamment pour chaque formant) à partir de valeurs en Bark ; la durée est exprimée en secondes. La méthode est la suivante : un premier arbre de classification est généré à partir des paramètres de F1, F2 et la durée. La fonction `test` utilise ensuite le

principe de la validation croisée pour calculer l'erreur (fonction du nombre de voyelles mal classées) associée aux arbres obtenus par élagage (« pruning ») du premier. On estime ainsi l'arbre d'erreur minimum. Puisque le but est d'atteindre un équilibre entre l'ajustement aux données et la complexité de l'arbre, une pénalité proportionnelle à cette dernière sera incluse dans la décision finale pour retenir le meilleur arbre. Le meilleur arbre (celui de la Figure 6.152) est le plus petit arbre dont l'erreur est inférieure à 1 erreur-type de l'arbre minimum. Dans le cas présent, il permet de classer correctement environ 76 % des monophthongues produites par les locuteurs masculins de *ean*.

On peut considérer que cet arbre de décision constitue un scénario possible dans le cadre d'un modèle à exemplaire d'accès au lexique. Le scénario proposé suppose, implicitement, une équiprobabilité de chacun des mots-test ; rendre compte d'un contexte informationnel plus complexe impliquerait une pondération de chaque branche. Son intérêt descriptif, pédagogique, voire théorique, est incontestable. Certes, il ne s'agit que d'une reformulation du trapèze vocalique traditionnel incluant également la durée, mais cet arbre de décision permet une classification parcimonieuse des timbres du système de *ean*. En allant de la racine vers les feuilles, une première partition sépare les voyelles fermées des voyelles ouvertes. Parmi les voyelles fermées, un critère de postériorité (relative) (premier noeud à gauche) sépare *hoard* des autres. Puis, un second impose une partition entre, d'un côté, *hood* et *who'd*, que l'on pourrait qualifier de voyelles centrales (sur l'axe antérieur/postérieur), et de l'autre, *heed* et *hid*, que l'on appellerait antérieures. On peut ensuite considérer que *hood* est une version brève de *who'd*, et que *hid* est une version plus ouverte de *heed*. Si nous revenons à la racine et empruntons le chemin des voyelles ouvertes, un premier critère de postériorité sépare *had* et *head* de *hod*, *Hudd*, *heard* et *hard*. À ce stade, le critère prévalant (qui pourrait par la suite être érigé en trait phonologique) qui permet de distinguer *had* et *head* est l'antériorité. Parmi les voyelles restantes (comparativement postérieures), *hard* se distingue par un degré d'aperture supérieur. Puis, un nouveau critère d'antériorité sépare *hod*, généralement considérée comme voyelle postérieure, de *Hudd* et *heard*, qui sont habituellement décrites comme centrales. Enfin,

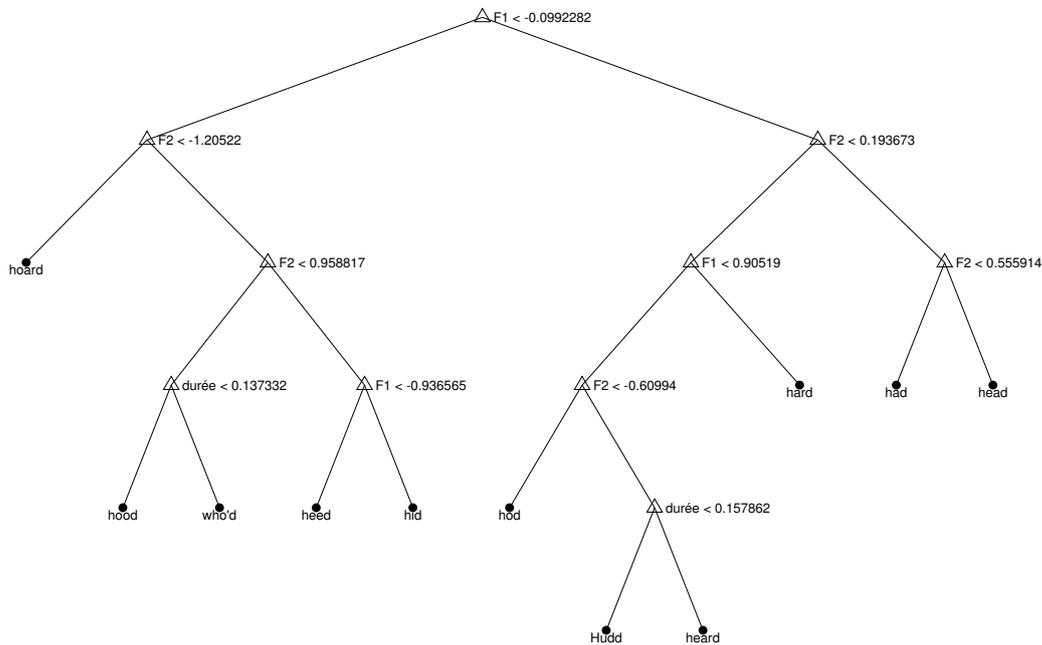


FIG. 6.152 – Arbre de classification des monophthongues de *ean* produites par les hommes.

l'arbre fait apparaître que *Hudd* est le pendant bref de *heard*. L'étape suivante consisterait donc à faire du paramètre déterminant chaque partition un trait phonologique. Cette méthode présente deux avantages intimement liés. Le premier vient de ce qu'elle produit une phonologie qui s'appuie sur la substance phonétique et qui reçoit donc – pour peu qu'elle veuille bien le réclamer – un appui empirique. Le second avantage réside dans le fait qu'un arbre binaire permet de réconcilier les paramètres physiques, par nature continus, et donc plurivalents, avec une phonologie structuraliste binaire, à condition de permettre un codage hiérarchisé des traits. Par exemple, le trait de fermeture à la racine de l'arbre n'est pas équivalent à celui qui fait contraster *heed* et *hid* plus loin dans l'arborescence. Cette hiérarchisation implique donc qu'un même trait puisse apparaître à différents niveaux de l'arborescence ; on peut parler de modèle « multi-échelle » puisque le critère (ou trait) de fermeture intervient à la fois aux niveaux macro-systémique – séparation des voyelles fermées du reste – et micro-systémique – distinction entre *heed* et *hid*.

6.4.2 Diphtongues

L'étiquette « diphtongues » a été utilisée par convention pour désigner les voyelles habituellement décrites comme des diphtongues en anglais standard (voir, e.g. [Wel90, Jon03]); or, nous le savions, tous les dialectes n'ont pas des diphtongues phonétiques pour ces timbres. L'impression générale qui se dégage de notre caractérisation des diphtongues par le biais des valeurs de F1 et F2 au début et à la fin de la voyelle est que les erreurs de mesures et la variation de réalisation entre locuteurs d'un même dialecte (e.g. monophtongue ou diphtongue à *ncl*) conduisent à certaines approximations.

De plus, l'approche traditionnelle qui consiste à caractériser une diphtongue par le biais des coordonnées de deux points, l'un au début temporel $M_1(x_1, y_1)$, l'autre à la fin $M_n(x_n, y_n)$, dans le plan F1/F2 implique une simplification :

- la distance euclidienne entre ces deux points dans F1/F2 caractérise, au sens strict, le déplacement, mais ne donne aucune information sur la distance (au sens cinématique) effectivement parcourue, i.e. sur le chemin emprunté entre M_1 et M_n . Dans le cas où tous les points successifs $M_{2\dots n-1}$ sont situés sur la droite qui passe par M_1 et M_n , la distance et le déplacement sont équivalents; mais une inspection sommaire de la Figure 6.153¹⁵⁴ suffit à démontrer que ce n'est pas le cas;
- la description utilisée jusqu'ici pourrait en outre laisser croire que les distances entre paires de points successifs $dist(M_i, M_{i+1})$ sont égales; autrement dit, que la vitesse instantanée¹⁵⁵ est constante; or la Figure 6.153 fait apparaître, pour une seule et même voyelle, et d'une voyelle à l'autre, une distance variable entre chaque paire de points successifs.

Les spectrogrammes du Chapitre 6 ont déjà permis de visualiser le fait que la distance parcourue dans F1/F2 entre $M_1(x_1, y_1)$ et $M_n(x_n, y_n)$, pour la plupart des diphtongues, n'est pas équivalente au déplacement, donné par :

¹⁵⁴Sur cette figure, les coordonnées du point de départ de toutes les occurrences sont (0,0).

¹⁵⁵Ici, les vitesses instantanées correspondent approximativement, dans notre espace temporel normalisé avec 13 valeurs discrètes équidistantes, aux 12 distances euclidiennes entre chaque paire de points successifs.

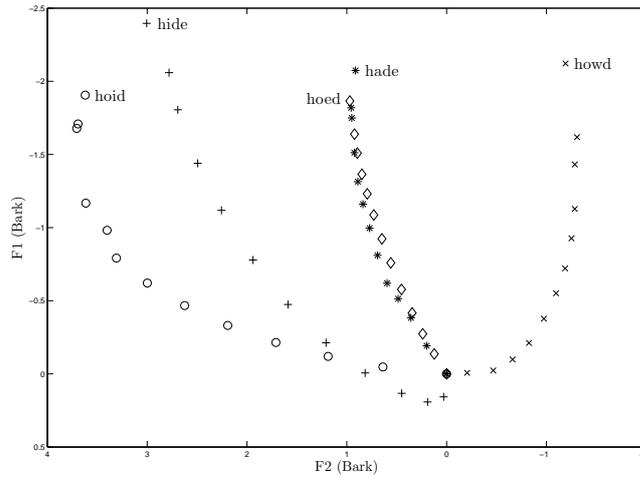


FIG. 6.153 – Trajectoires des diphtongues fermantes de la locutrice *wjh* de *sse*.

$$\Delta r = \sqrt{(M_{1x} - M_{nx})^2 + (M_{1y} - M_{ny})^2} \quad (6.8)$$

Théoriquement, calculer la distance consiste à mesurer la longueur de la courbe (estimée par le biais de la régression polynomiale) entre $M_1(x_1, y_1)$ et $M_n(x_n, y_n)$. Une approximation suffisante consiste à additionner les distances euclidiennes entre chacune des 12 paires successives de points d'analyse :

$$\Delta s \approx \sum_{i=1}^{n-1} \sqrt{(M_{i+1x} - M_{ix})^2 + (M_{i+1y} - M_{iy})^2} \quad (6.9)$$

Or c'est à ce stade qu'on réalise que la stratégie visant à écarter les erreurs de formants – l'utilisation de la régression (voir Section 6.2) – se transforme en inconvénient majeur. Le bénéfice que pourrait apporter la notion de distance par rapport à celle de déplacement est en partie perdu en raison du fait que la distance est très largement tributaire de la fonction (affine ou polynôme) utilisée pour approcher le tracé formantique sous-jacent. Il est donc probable qu'il faille, au vu du peu de fiabilité de l'extraction automatique des formants, avoir recours à des mesures de l'évolution de la forme globale du spectre (centre de gravité, aplatissement, asymétrie) sur la durée de la voyelle.

Pour finir, nous rapportons une expérience de classification automatique de la diph-tongue de FACE publiée dans Ferragne et Pellegrino ([FP04b]). La voyelle de FACE est l'un des 7 traits de la typologie de Trudgill (Section 3.2.1.6); tester son potentiel discriminant peut donc s'avérer instructif.

La voyelle du mot *faces* du passage lu du corpus ABI a été segmentée manuellement. Les formants ont été estimés automatiquement sans avoir recours à une inspection visuelle des tracés comme ce fut le cas dans la Section 6.2. Puis, une régression linéaire robuste a été calculée sur les deux premiers formants vocaliques dont les valeurs ont été au préalable converties en Bark (le temps étant la variable indépendante). La classification a été effectuée au moyen d'un perceptron à une couche¹⁵⁶. Nous n'avons inclus en entrée du classifieur que les deux valeurs de pentes de l'équation de régression¹⁵⁷ et non les ordonnées à l'origine puisque nous souhaitons nous concentrer sur l'aspect dynamique du spectre. Les typologies de Trudgill et de Hughes et collègues (voir Section 3, Tables 3.2 et 3.1) sont par nature bivalentes et nous laissent donc espérer l'existence potentielle de deux classes de dialectes au maximum sur la base de la voyelle de FACE. Notre étude des voyelles dans la Section 6.2 nous laisse cependant penser que trois types de schémas de F1 et F2 existent pour cette voyelle :

- le schéma parallèle, pour la monoph-tongue ;
- le schéma divergent, pour la diphtongue ouvrante ;
- le schéma convergent, pour la diphtongue centripète.

Puisqu'une tâche à 13 dialectes ne serait pas fondée, nous avons opéré une classification des dialectes deux à deux¹⁵⁸. Les taux de classification correcte varient d'environ 43 % à presque 98 %.

Dans la Figure 6.154 les dialectes sont représentés par leur valeur de pente (médianes et écarts interquartiles) pour F1 (abscisse) et F2 (ordonnée) mesurées sur la voyelle de

¹⁵⁶Nous avons utilisé la « toolbox » Netlab ([Nab02]); le « leave-one-out » est utilisé comme méthode de validation.

¹⁵⁷C'est-à-dire, le terme b de l'Équation 2.10.

¹⁵⁸Soit, au total, $14 \times (14 - 1) / 2 = 91$ paires de dialectes; *ilo* est également inclus puisque l'étude de [FP04b] était antérieure à l'étude perceptive qui a abouti au rejet de ce dialecte.

faces. On remarque la corrélation attendue entre les deux variables. La dispersion des dialectes fait apparaître des regroupements allant dans le sens de nos prédictions :

- un premier groupe (*ean, crn, brm, sse, lvp, roi* et *nwa*) présente des valeurs absolues élevées, ce qui indique que nous avons affaire à des diphtongues ouvranes. Celles-ci sont particulièrement diphtonguées pour *ean* et *crn* ; elles le sont nettement moins en moyenne pour *roi* et *nwa*, qui semblent former un sous-groupe à part ;
- un deuxième groupe, composé de *shl, lan, gla, eyk* et *ncl*, rassemble les dialectes d'Écosse et du nord de l'Angleterre, caractérisés par une monophthongue. Comme nous avons pu le constater dans la Section 6.3.9.2, la norme pour *ncl* semble être la monophthongue et non la variante historique stigmatisée : [iə]. Néanmoins, l'amplitude de la variation pour *ncl* permet de constater que de multiples variantes sont réalisées par les locuteurs de ce dialecte. Concernant *shl, lan, gla* et *eyk*, leur variation est relativement faible, ce qui laisse supposer une homogénéité comparativement élevée dans les réalisations des locuteurs de ces quatre dialectes. On aurait pu également émettre l'hypothèse d'une corrélation entre la variation et la moyenne de chaque dialecte, mais cela ne semble pas être le cas. En effet, la corrélation entre les médianes et les écarts interquartiles de F1 est de $r = -0,251$ ($p = 0,407$) ; pour F2, cette corrélation est de $r = 0,243$ ($p = 0,423$) ;
- un dernier groupe où *uls* se distingue avec des valeurs de pente légèrement positives sur F1 et très légèrement négatives sur F2 en moyenne. Il va de soi qu'une partie des locuteurs de *ncl* tombent vraisemblablement dans cette catégorie.

Dans le Chapitre 6, nous venons de proposer une description acoustique des systèmes vocaliques de 13 dialectes de l'anglais des Îles Britanniques à travers des représentations classiques dans l'espace des trois paramètres que sont F1, F2 et la durée. Notre attention s'est portée, en particulier, sur les cas où les voyelles de deux mots-test étaient remarquablement proches. Pour l'essentiel, l'organisation des systèmes, tel qu'il est possible de l'inférer à partir des espaces acoustiques, est cohérente avec la littérature du domaine. Cependant, la variation phonologique constatée entre locuteurs d'un même dialecte ré-

vèle, en partie, l'absence d'information sociologique sur les locuteurs. En ce qui concerne le potentiel discriminant des monophthongues, *hid*, *heard*, *Hudd* et *hood* semblent être les meilleurs traits discriminants, si on considère 13 catégories dialectales. Les diphtongues ne sont pas traitées de la même manière puisque, comme nous venons de l'expliquer, la représentation utilisée est très bruitée, et elle n'inclut pas explicitement d'information dynamique. Une expérience de classification des dialectes deux à deux montre qu'il est possible de séparer, à partir des pentes de F1 et F2, les locuteurs de deux dialectes qui ont une voyelle phonétiquement différente dans le mot-test *faces*. Dans la Chapitre 7, nous abordons la classification automatique et les représentations multidimensionnelles des voyelles du corpus ABI.

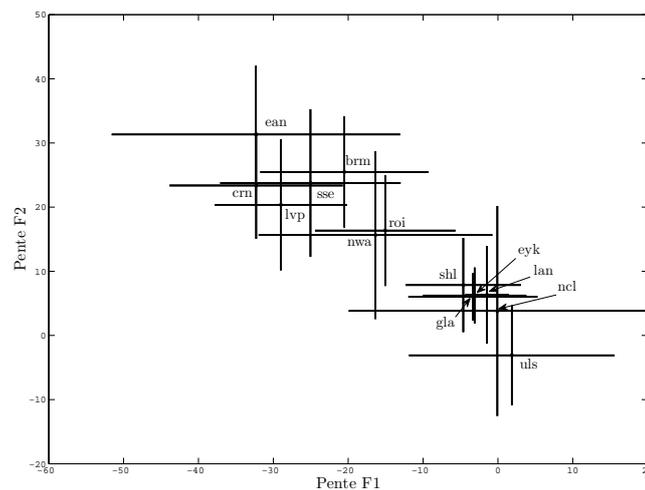


FIG. 6.154 – Médianes et écarts interquartiles des valeurs de pente pour F1 (abscisse) et F2 (ordonnée).

Chapitre 7

Classification automatique et représentations multidimensionnelles

Le Chapitre 6 avait pour but de présenter une description actuelle des systèmes vocaux de 13 dialectes du corpus ABI. Nous nous sommes jusque là contenté, pour l'essentiel, de méthodes d'investigations classiques en phonétique :

- prise en compte de trois paramètres physiques : F1, F2 et durée ;
- représentation des monophthongues par leurs coordonnées dans le plan F1/F2, mesurées au milieu temporel ;
- représentation des diphtongues par le biais des coordonnées dans F1/F2 du début et de la fin de la voyelle.

L'étude de ces trois paramètres physiques est légitimée par le fait qu'il est aisé de les interpréter dans un cadre descriptif traditionnel, et en raison de leur pertinence sur le plan de la perception. Se limiter à deux ou trois paramètres permet en outre de tous les représenter sur un même graphe, ce qui facilite l'analyse. Il est toutefois indéniable que cette méthode « classique » souffre d'inconvénients majeurs :

- l'estimation automatique des formants a des taux d'erreur parfois considérables. Nous introduisons dans ce chapitre d'autres paramètres fréquentiels permettant de contourner ce problème ;

- malgré la transformation de la fréquence objective en Bark, puis le centrage et la réduction de F1 et F2, on continue d’observer une variation acoustique individuelle (e.g. entre sexes) importante qui ne semble pas inclure de l’information dialectale. La Section 7.2 aborde une méthode de classification automatique qui permet de s’affranchir d’une partie de ces contraintes ;
- le fait de se cantonner à la prise en compte de 2 ou 3 dimensions offre, certes, une représentation canonique consensuelle des systèmes vocaliques, mais des moyens alternatifs existent, qui visent à apporter un éclairage supplémentaire aux données (voir notamment Section 7.4).

7.1 Extraction des paramètres et segmentation

La représentation fréquentielle du signal que nous avons employée jusqu’ici se limite aux deux premiers formants vocaliques. Les formants sont des paramètres utilisés très couramment et leur extraction obéit à des procédures standard ; nous n’avons, par conséquent, pas jugé utile de détailler ces procédures. En revanche, les technologies de la parole, notamment la reconnaissance automatique de la parole, utilisent des représentations plus élaborées. Les paramètres pris en compte dans la Section 7.2 sont appelés « Mel Frequency Cepstral Coefficients » (MFCC). Pour les obtenir, on procède d’abord à une analyse dite « cepstrale », qui vise à séparer les contributions de la source (la fréquence fondamentale) et du filtre (résonances caractéristiques du conduit vocal). En pratique, cette analyse consiste à prendre le logarithme du spectre d’un signal et à lui appliquer une transformation de Fourier inverse. Un filtrage effectué dans le domaine cepstral (liftrage) permet alors d’écartier la contribution de la source. La paramétrisation MFCC proprement dite s’obtient par le biais d’une échelle de fréquence non-linéaire (application de filtres dont la bande passante croît avec la fréquence, émulant la manière dont l’oreille perçoit la hauteur) et les coefficients sont calculés par une transformée en cosinus discrète. Ceci permet,

en outre, de travailler sur des paramètres décorrés¹⁵⁹.

Comme nous l'avons déjà indiqué dans la Section 6.2, les voyelles des mots à structure /hVd/ sont extraites à partir de la détection automatique de fréquence fondamentale. La segmentation du passage lu, quant à elle, a été réalisée par alignement forcé avec le Hidden Markov Model Toolkit (HTK)¹⁶⁰. L'apprentissage des modèles de phones n'a pas été réalisé par nos soins puisque nous ne disposons pas de corpus étiqueté en phones. Ces modèles, appris sur le corpus WSJCAM0 ([RFP⁺95]), nous ont été aimablement fournis par Mark Huckvale, qui les a lui-même utilisés pour la segmentation d'une partie du corpus ABI que nous n'avons pas employée pour nos recherches ([Huc04]). Nous avons ensuite manuellement établi un dictionnaire contenant, pour chaque entrée, un mot orthographique du passage lu avec sa transcription phonétique en anglais britannique standard. Il est courant d'enregistrer plusieurs variantes phonétiques en plus de la transcription canonique ; nous y avons cependant renoncé, ne connaissant pas *a priori* toutes ces variantes.

7.2 Méthode des matrices de distances (ACCDIST)

La méthode ACCDIST (Accent Characterisation by Comparison of Distances in the Inter-segment Similarity Table) a été mise au point par Huckvale [Huc04] ; elle s'inspire indirectement de Barry et collègues [BHN89]. La procédure consiste non pas à comparer chaque voyelle entre plusieurs locuteurs, mais à obtenir une représentation du système de chaque locuteur à travers la matrice de distances de ses voyelles prises deux à deux pour ensuite comparer ces matrices de distances entre locuteurs. Cette méthode possède deux points forts : l'utilisation des MFCC permet de contourner le problème des erreurs d'extraction automatique de formants, et, comme nous le verrons dans la description qui suit, la question de la normalisation du locuteur est implicitement résolue lors du calcul des distances entre deux matrices. Nous avons eu l'occasion de présenter nos résultats dans

¹⁵⁹Voir, par exemple, Haton et collègues ([HCF⁺06], 39-70) pour plus de détails.

¹⁶⁰<http://htk.eng.cam.ac.uk/> ; site consulté le 6 avril 2008. Voir aussi [Jel97, JM00] et [HCF⁺06] pour les aspects mathématiques.

[Fer06] et [FP07]; ce qui suit constitue un développement inspiré de ces deux publications.

Décrivons à présent les différentes étapes de la méthode ACCDIST, telle que nous l'avons appliquée aux mots à structure /hVd/. Nous avons donc délimité les frontières des voyelles en procédant à une détection automatique de la fréquence fondamentale. Certes le voisement dans les items /hVd/ se prolonge au-delà de la partie strictement vocalique du signal, mais ce biais n'affecte au final que notre estimation de la durée car le contenu spectral antérieur à 25 % et postérieur à 75 % de la durée vocalique n'a pas été retenu pour l'analyse. Douze coefficients cepstraux et l'énergie ont été calculés à 25, 50 et 75 % de la durée de la voyelle¹⁶¹. La dérivée de ces paramètres a également été calculée. Si on ajoute la durée, cela revient à dire que chaque voyelle est définie par un vecteur de 53 paramètres. Les coefficients cepstraux ont été calculés avec le programme `melfcc`¹⁶²; les options retenues sont celles recommandées par l'auteur du programme pour reproduire la manière dont HTK extrait les MFCC, à ceci près que la taille de la fenêtre d'analyse et le pas d'analyse ont été fixés à 20 et 10 ms, respectivement.

La Figure 7.1 illustre le processus de calcul des matrices de distances. L'influence du contexte consonantique étant de fait nulle, nous avons représenté toutes les occurrences d'un même timbre (phonème) chez un locuteur sous la forme de son vecteur moyen dans l'espace à 53 dimensions. On obtient donc, pour chaque individu, une matrice de paramètres de 19×53 (dans le cas des 19 mots en /hVd/). Cette matrice individuelle est ensuite transformée en matrice de distances de Minkowski entre les 19 items deux à deux (e.g. matrice m_1^A dans la Figure 7.1). Ensuite, à partir des matrices de locuteurs (m_1^A , m_2^A , m_3^B et m_4^B dans la Figure 7.1), une matrice moyenne est calculée pour chacun des 13 dialectes (matrices M^A et M^B de la figure). La procédure de classification (voir Figure 7.2) proprement dite s'appuie sur le principe du plus proche voisin : elle consiste à estimer le degré de similitude entre la matrice de chaque locuteur et les 13 matrices de dialectes, chaque locuteur étant classé dans le dialecte dont la matrice présente le plus

¹⁶¹Soit 13 paramètres calculés aux 3 premiers quarts de la voyelle : $13 \times 3 = 39$; auxquels on ajoute leur dérivée ($39 + 13 = 52$) ainsi que la durée de la voyelle : $52 + 1 = 53$ paramètres.

¹⁶²<http://www.ctr.columbia.edu/dpwe/resources/matlab/rastamat/mfccs.html>; site consulté le 6 avril

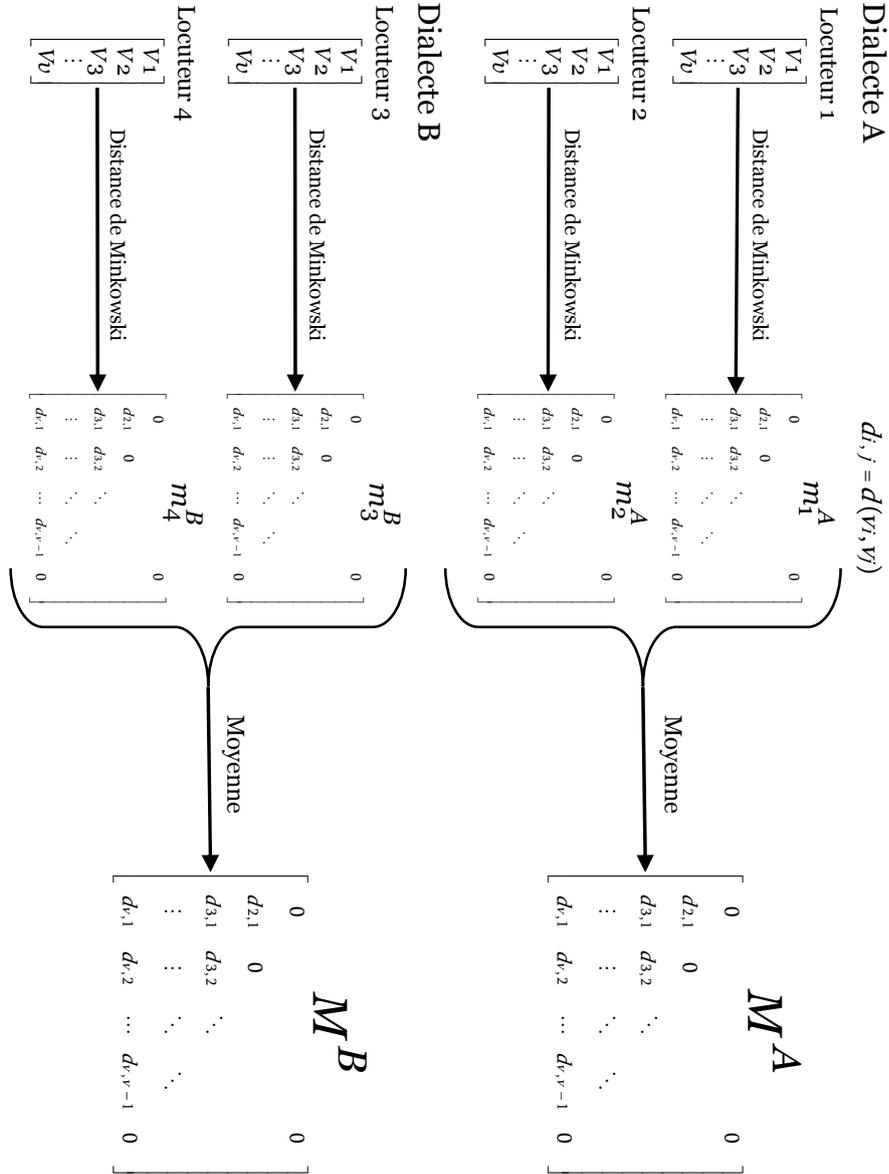


FIG. 7.1 – Schéma synoptique illustrant le calcul des matrices de distances dialectales. Exemple à partir de deux dialectes, A et B, et quatre locuteurs (1 à 4). Chaque voyelle ($V_1 \dots V_v$) est représentée par le vecteur de paramètres moyen de toutes les occurrences d'un même type. On calcule ensuite une matrice (e.g. matrice m_1^A) de distances de Minkowski (d) entre chacune des $v \times (v - 1)/2$ paires de voyelles. Enfin, pour chaque dialecte, la matrice moyenne de tous les locuteurs est calculée (e.g. matrice M^A).

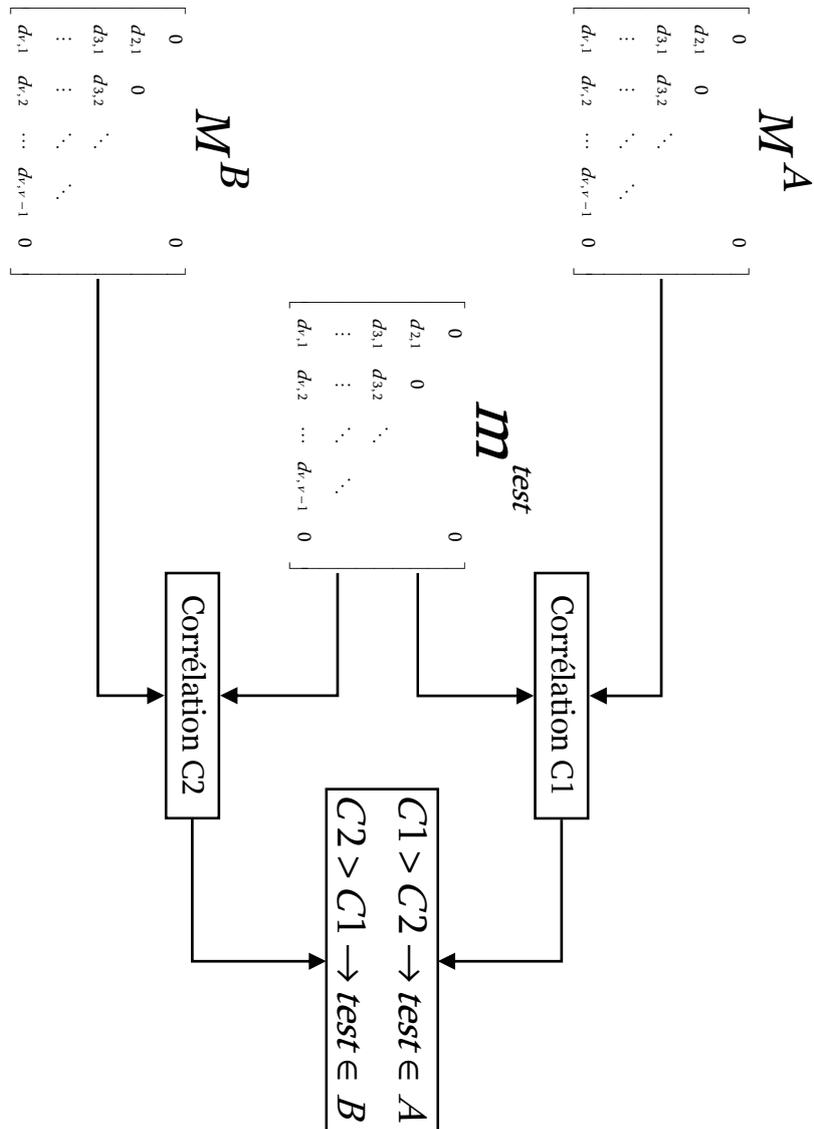


FIG. 7.2 – Schéma synoptique illustrant la procédure de classification. M^A et M^B sont les matrices de distances moyennes des dialectes A et B respectivement ; m^{test} est la matrice de distances du locuteur en cours de test.

fort degré de similitude avec la sienne. Nous utilisons la méthode du « leave-one-out » : la matrice du dialecte dont le locuteur en cours de test est issu est recalculée sans la matrice de ce dernier. Reste à déterminer comment mesurer la ressemblance entre deux matrices. Huckvale [Huc04] utilise un coefficient de corrélation de matrices. Nous avons également adopté cette méthode : la matrice de distances est « dépliée » en un vecteur, puis, on calcule le coefficient de corrélation de Pearson, dont nous rappelons la formule dans l'Équation 7.1. Nous avons choisi de présenter la formule détaillée du coefficient de Pearson, car elle fait clairement apparaître l'une des raisons qui en font un outil de choix dans notre procédure : on constate que la mesure est insensible aux différences d'amplitude de variation entre les deux variables (ici, les matrices de distances dépliées) puisque les valeurs sont « normalisées ». Le dialecte estimé du locuteur en cours de test est celui dont la matrice possède le coefficient de corrélation le plus élevé avec celle de ce locuteur.

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (7.1)$$

En excluant 2 locuteurs dont la liste de mots était incomplète, l'étude des mots /hVd/ se base donc sur la production de 261 locuteurs ; 138 femmes et 123 hommes. En appliquant la méthode ACCDIST à partir de distances euclidiennes entre les voyelles, les taux de classification correcte sont de 85,8 %, 89,8 % et 88,5 % pour les hommes, les femmes et tous sexes confondus, respectivement. Une légère amélioration apparaît dans toutes les conditions avec la distance de Manhattan : respectivement, 87,3, 90,6 et 89,3 % de classification correcte. La matrice de confusion de la tâche incluant les deux sexes (distance de Manhattan) est reproduite dans la Table 7.1.

Nous avons ensuite appliqué la méthode au passage lu du corpus ABI. Il convient ici d'identifier chaque voyelle de façon exclusive : par exemple, la voyelle accentuée de <salor> est phonologiquement identique à celle de <faces> en anglais standard, mais elles ne sont pas considérées comme une seule et même entité, les contextes consonantique et informationnel variables engendrant des réalisations phonétiques différentes. On sait par

TAB. 7.1 – Matrice de confusion de la classification des mots en /hVd/.

	brm	crn	ean	eyk	gla	lan	lvp	ncl	nwa	roi	shl	sse	uls
brm	18	-	1	-	-	1	-	-	-	-	-	-	-
crn	-	16	-	-	-	-	-	-	-	1	-	3	-
ean	1	-	14	-	-	-	-	-	-	-	-	4	-
eyk	2	-	-	22	-	-	-	-	-	-	-	1	-
gla	-	-	-	-	18	-	-	-	-	-	-	-	2
lan	-	-	-	-	-	21	-	-	-	-	-	-	-
lvp	-	-	-	-	-	-	19	-	-	-	-	-	-
ncl	-	-	-	1	-	-	-	17	1	-	-	-	-
nwa	1	-	1	-	-	-	1	-	17	-	-	-	-
roi	-	-	-	-	-	-	-	-	1	19	-	-	-
shl	1	-	-	-	1	-	-	-	-	-	19	-	1
sse	-	1	2	-	-	-	-	-	-	-	-	13	-
uls	-	-	-	-	-	-	-	-	-	-	-	-	20

exemple que <wait> et <weight> ([ei] et [ɛi], respectivement) ne riment pas chez tous les locuteurs dans certains dialectes du nord de l'Angleterre [Bea04], 123 ; autrement dit, ces mots appartiennent au même ensemble lexical en anglais standard, mais pas dans certains dialectes du nord. Il est nécessaire de segmenter au préalable le signal sonore en phones, chacun portant une étiquette l'identifiant de façon exclusive. L'attribution d'une étiquette à un phone découle de considérations théoriques conduisant à certains choix déterminants. En effet, l'alignement forcé fait appel à un dictionnaire dans lequel chaque mot orthographique est retranscrit phonologiquement selon la représentation canonique de l'anglais britannique standard telle qu'elle apparaît, par exemple, dans un dictionnaire de prononciation comme celui de Jones [Jon03]. Or il est évident qu'au-delà des différences de réalisation, les dialectes diffèrent également sur le plan de la représentation phonologique d'un même mot. Il s'agit ici d'un problème de catégorisation lié au concept d'ensembles lexicaux. Comme nous l'avons déjà noté plus haut (Section 2.1), les dialectologues britanniques utilisent des mots-clés pour désigner ce qui se rapproche des phonèmes dans les différents dialectes. Le nombre de ces mots-clés varie d'un dialecte à l'autre. Bien

qu'il existe, au moins implicitement, ce qui pourrait s'apparenter à un système panlectal – il suffirait de dresser la liste de tous les ensembles lexicaux des dialectes étudiés pour recenser tous les contrastes possibles au niveau des Îles Britanniques – décider de l'appartenance d'une voyelle donnée à un certain ensemble nécessite l'intervention d'un dialectologue expert. Rien ne garantit, au reste, que les connaissances de ce dernier lui permettent d'inférer sans erreur l'appartenance d'une voyelle dans un mot donné à un phonème. Plutôt que regrouper les sons en phonèmes et pratiquer des analyses à partir de valeurs moyennes pour chaque phonème, nous avons contourné la question linguistique en identifiant chaque phone de manière individuelle. Autrement dit, chaque phone – ou plus précisément chaque phone en position accentuée à l'intérieur d'un mot – est libellé de la manière suivante : le son du graphème <ai> dans le mot <sailor>, le son du graphème <a> dans le mot <faces>, etc.

Nous avons ensuite calculé le nombre de mots que tous les locuteurs avaient en commun. Les voyelles accentuées des 61 mots constituant le plus grand dénominateur commun entre les 264 locuteurs (135 hommes, 129 femmes) ont été extraites, et les mêmes paramètres que précédemment ont été calculés. Certes, tous les locuteurs ont lu le même texte, mais il arrive parfois qu'un sujet ne prononce pas un mot ; ce mot est donc écarté de l'analyse pour tous les sujets. Les scores de classification, en utilisant la distance euclidienne, sont de 89,6 %, 87,6 % et 90,5 % pour les hommes, les femmes et tous sexes confondus, respectivement. La distance de Manhattan permet une amélioration des performances seulement dans la condition « femmes » : 89,4 %. La matrice de confusion pour la condition « tous sexes confondus » est reproduite dans la Table 7.2.

Les matrices de confusion (Table 7.1 et 7.2) font apparaître des taux de classification correcte élevés, ce que ne laissait présager ni l'analyse auditive du corpus, ni les mesures effectuées dans les Chapitres 5 et 6. De plus, nous nous attendions à des scores relativement plus faibles pour la classification à partir du passage lu, compte tenu de la variation supplémentaire occasionnée par des contextes informationnels (contexte consonantique inclus) variables. On peut donc envisager soit que l'information spécifique au dialecte est

TAB. 7.2 – Matrice de confusion de la classification du passage lu.

	brm	crn	ean	eyk	gla	lan	lvp	ncl	nwa	roi	shl	sse	uls
brm	18	-	-	-	-	-	-	-	-	-	-	2	-
crn	-	14	1	1	-	-	-	-	-	1	-	3	-
ean	-	1	14	-	-	-	-	-	-	-	-	4	-
eyk	-	-	-	22	-	-	-	-	-	-	-	3	-
gla	-	-	-	-	20	-	-	-	-	-	-	-	-
lan	-	-	-	-	-	21	-	-	-	-	-	-	-
lvp	-	-	-	-	-	-	20	-	-	-	-	-	-
ncl	-	-	-	-	-	-	-	19	-	1	-	-	-
nwa	-	1	-	-	-	-	-	-	19	-	-	1	-
roi	-	1	-	-	-	-	-	-	1	18	-	-	-
shl	-	-	-	-	2	-	-	-	-	-	20	-	-
sse	-	1	-	-	-	-	-	-	-	-	-	15	-
uls	-	-	-	-	1	-	-	-	-	-	-	-	19

préservée malgré la variation supplémentaire, soit que la variation inhérente au passage lu apporte, outre du bruit, une information propre au dialecte que les mots en /hVd/ ne contiennent pas.

De plus, la plupart des erreurs de classification ne sont pas totalement incohérentes. En effet, si l'on inspecte la Table 7.2 plus en détail, on constate que la classe estimée contenant le plus grand nombre d'erreurs est *sse*, ce qui, en l'absence d'information complémentaire, semble prévisible eu égard au prestige exercé par cette variété (au moins en Angleterre). D'ailleurs, les dialectes dont certains locuteurs sont classés à tort dans la catégorie *sse* (*brm*, *crn*, *ean*, *eyk* et *nwa*) sont, à l'exception de *eyk* et peut-être *brm*, phonétiquement proches de *sse* en cela qu'ils sont, linguistiquement parlant, des dialectes du sud. La classification de deux locuteurs de *shl* et d'un locuteur de *uls* dans la catégorie *gla* est compatible avec la description de ces dialectes (Chapitre 3), et concorde avec les résultats obtenus au Chapitre 6 : ils font tous trois partie d'une supra-région, la zone Scots. Dans les grandes lignes, ces remarques sont également applicables à la Table 7.1.

Puisque les distances entre les voyelles d'un locuteur semblent être un indicateur très

TAB. 7.3 – Taux de classification maximal (%).

Corpus	Condition	Taux	Nombre de CP
/hVd/	H	91,04	36
/hVd/	F	90,55	30
/hVd/	H & F	93,87	31
passage	H	90,37	14
passage	F	93,80	36
passage	H & F	94,32	34

fiable de son dialecte, on peut utiliser ces distances comme paramètres d'entrée dans une analyse discriminante dans le but d'opérer une classification automatique. En pratique, cette approche comporte toutefois une restriction : le nombre de paramètres doit être inférieur au nombre d'individus. Or ce critère n'est rempli que dans le cas des mots /hVd/ dans la condition « sexes confondus »¹⁶³. Si l'on admet que ces distances obéissent très probablement à un nombre réduit de facteurs latents, l'analyse en composantes principales doit permettre de contourner le problème du nombre de paramètres. Une analyse en composantes principales préalable a donc été effectuée sur la matrice de distances des mots en /hVd/ et celle du passage lu. Puis une série d'analyses discriminantes ont été utilisées pour la classification avec comme paramètres d'entrée les i premières composantes principales (CP) pour i variant de 1 au nombre total de CP¹⁶⁴. Les taux maximum obtenus avec cette méthode sont récapitulés dans la Table 7.3, où l'on note une amélioration de la classification par rapport à la méthode impliquant le coefficient de corrélation.

Il conviendrait ici de mieux cerner ce que nous entendons par erreur de classification. Une erreur supposerait que tous les locuteurs d'un dialecte puissent être identifiés (par exemple, par le biais d'une expérience de perception ou par un dialectologue expert) de

¹⁶³En effet, les distances entre les 19 voyelles prises deux à deux sont au nombre de $19 \times 18/2 = 171$; ce nombre est donc inférieur au total des sujets (261), mais supérieur au nombre de sujets masculins (123) et féminins (138). Quant aux distances entre les 61 mots du passage lu pris deux à deux, elles sont au nombre de $61 \times 60/2 = 1830$.

¹⁶⁴Cette méthode comporte le défaut de générer, lorsque i est proche du nombre total de paramètres, des matrices de covariance qui ne sont plus définies positives, ce qui rend impossible le bon fonctionnement de l'analyse discriminante telle qu'elle est implémentée dans la fonction `classify` de Matlab.

manière non équivoque comme appartenant à ce dialecte. Or la déviance de certains locuteurs par rapport au prototype du dialecte qu'ils sont censés représenter¹⁶⁵ nous porte à croire que le score maximal de classification correcte possible avec le corpus ABI n'est pas de 100 %. Il est d'ailleurs probable que les taux que nous avons obtenus constituent un plafond, dans le sens où les locuteurs qui apparaissent comme mal classés dans les Tables 7.1 et 7.2 sont en réalité mieux classés, sur la base de leur système vocalique, qu'ils ne l'étaient à partir du critère géographique des lieux d'enregistrement. Ceci illustre parfaitement la difficulté inhérente à la constitution d'un corpus dialectal sur les Îles Britanniques. En effet, la définition de ce qu'est un dialecte est tributaire de l'objectif visé. Par exemple, si l'on souhaite étudier des locuteurs typiques, il est nécessaire de les sélectionner méticuleusement précisément parce qu'ils sont typiques. La définition du dialecte sera alors très normative. Or, dans une optique de classification automatique, il peut s'avérer intéressant de situer de nouveaux locuteurs par rapport aux pôles prototypiques du corpus d'entraînement, par exemple, pour l'indexation automatique d'acteurs imitant des accents.

Une solution alternative au problème de l'appartenance d'un locuteur à un dialecte consisterait, par exemple, à prendre en compte des facteurs socio-économiques et d'autres, tels que l'histoire linguistique, la mobilité géographique, etc. Cependant, ceci conduirait – si l'on souhaite conserver les 13 localités géographiques de ABI et les subdiviser en autant de sous-groupes qu'il y a de facteurs pertinents – à collecter une quantité de données démesurée. C'est d'ailleurs l'une des raisons probables qui conduisent les auteurs d'études empiriques (e.g. [FD99]) de dialectologie urbaine à ne se concentrer que sur une seule localité. En résumé, l'étiquette identifiant le dialecte de chaque locuteur ne lui a été attribuée que sur la base du lieu où il a grandi. Or, au vu de la complexité des phénomènes déterminant la prononciation d'un individu, ce critère d'attribution est trop réducteur.

¹⁶⁵Cette déviance est constatée dans la documentation du corpus ABI ainsi que dans l'analyse auditive de l'homogénéité et de la conformité au prototype décrite dans la Section 4.2.

7.3 Portée descriptive et diagnostic

Cette section est consacrée à l’exploration détaillée des résultats de la classification automatique et à des représentations dont le but est de faciliter l’interprétation phonétique de ces résultats.

Dans un premier temps, nous examinons les locuteurs mal classés à la lumière de l’analyse auditive décrite dans la Section 4.2. Puis, nous représentons les systèmes vocaliques des dialectes du corpus ABI afin de proposer une visualisation pertinente des distances entre voyelles (Section 7.4). Ensuite, la question des voyelles les plus discriminantes est abordée dans la Section 7.5. Enfin, nous proposons dans la Section 7.6 une brève interprétation de la distance objective entre plusieurs dialectes dans le cadre de la théorie des ensembles flous, que nous avons introduite dans la Section 2.3.3.

En comparant l’ensemble des 28 locuteurs (sur 261) mal classés de la tâche avec mots en /hVd/ et celui des 25 locuteurs (sur 264) de la tâche réalisée à partir du passage lu du corpus, on constate que ces deux ensembles ont 10 locuteurs en commun. La Table 7.4 présente ces 10 locuteurs et leur classe estimée dans chacune des deux tâches de classification. On remarque que la moitié des mauvaises classifications communes aux deux tâches concernent un locuteur ou une locutrice (surtout du sud) qui est identifié comme *sse*. Ceci n’a rien de surprenant car posséder un accent du type de *sse* est théoriquement compatible avec toutes les localités géographiques.

Si l’on confronte les données de la Table 7.4 avec l’analyse perceptive des sujets masculins décrite dans la Section 4.2, on note une bonne cohérence entre la tâche de classification automatique et la tâche de discrimination par l’expert humain¹⁶⁶. Sur les 10 locuteurs présentés dans la Table 7.4, 4 ont fait l’objet d’un faux rejet – i.e. ils appartenaient bien au dialecte en cours de test, mais l’expert a considéré que non – et deux autres, quoique identifiés comme membres du dialecte en cours de test, ont reçu la note de 1 sur 5 traduisant leur degré minimal de conformité au prototype. De plus, lorsque le locuteur `ean_M_kxk`

¹⁶⁶Pour rappel, l’expérience de la Section 4.2 consistait à entendre un extrait du corpus et décider s’il appartenait au dialecte en cours de test (l’expérience était divisée en 13 blocs) ou non, et avec quel degré de conformité au prototype (note sur 5).

TAB. 7.4 – Sujets mal classés dans les deux tâches.

Locuteurs	classe estimée (/hVd/)	classe estimée (passage)
crn_F_jmc	<i>sse</i>	<i>sse</i>
crn_F_slm	<i>sse</i>	<i>sse</i>
crn_M_acr	<i>sse</i>	<i>ean</i>
ean_M_htl	<i>sse</i>	<i>sse</i>
ean_M_kxk	<i>sse</i>	<i>sse</i>
eyk_F_sxc	<i>brm</i>	<i>sse</i>
eyk_M_axc	<i>sse</i>	<i>sse</i>
ncl_M_ggc	<i>eyk</i>	<i>roi</i>
roi_F_cmd	<i>nwa</i>	<i>nwa</i>
sse_M_kmp	<i>crn</i>	<i>crn</i>

a été utilisé comme distracteur dans le bloc testant le dialecte *sse*, il a fait l'objet d'une fausse acceptation (identifié comme *sse*), et c'est précisément la même « erreur » que commet la classification automatique. Dans la Section 6.3.9, nous avons déjà noté le caractère atypique du locuteur *ggc* de *ncl* : celui-ci a une voyelle de *heard* proche de celle de l'anglais standard et une diphtongue fermante dans *hade*. Nous avons également noté la faible conformité des productions du locuteur *kmp* au prototype de l'anglais standard (Section 6.3.1.3) ; celui-ci prononce le <r> graphique lorsqu'il y en a un dans les mots en /hVd/, et sa voyelle dans *hade* est très peu diphtonguée. Ce locuteur avait en outre reçu la note de 1 sur 5 censée mesurer son degré de conformité au prototype de notre expert (Section 4.2) ; il est donc certain que ce locuteur n'aurait pas dû figurer dans l'ensemble *sse*.

À ce stade, il est acquis que le coefficient de corrélation $r_{i,j}$ entre les matrices de distances phonétiques de deux locuteurs (i et j) est un excellent indice de leur similarité dialectale. Aussi, il est possible de générer une matrice de distances entre les locuteurs pris deux à deux en considérant que la dissimilarité entre les locuteurs i et j est égale à $1 - r_{i,j}$. Afin de représenter visuellement cette matrice de distances, nous avons utilisé la technique du *multidimensional scaling* (MDS). La Figure 7.3 représente les dialectes dans l'espace des 3 premières dimensions du MDS. Le centre de chaque ellipse correspond

aux coordonnées moyennes du dialecte ; le rayon dans chacune des dimensions représente un écart-type par rapport à la moyenne. Seuls les dialectes clairement identifiables sur le graphique sont notés ; *ncl*, *nwa* et *sse* n’y sont pas identifiés.

Notre interprétation nous permet d’isoler les 3 dialectes de la zone Scots (*gla*, *shl* et *uls*) du reste. On remarque la variation relativement faible de *gla*. Les dialectes du sud de l’Angleterre apparaissent au premier plan, ceux du nord, juste derrière. On peut noter le chevauchement entre *lan* et *eyk*. Le dialecte *roi* semble distinct des deux groupes que nous venons de mentionner, ce qui n’a rien d’étonnant. On relève également la proximité tout à fait cohérente de *lvp* et *roi*.

Afin d’obtenir une meilleure résolution des dialectes de Grande-Bretagne, le MDS a été recalculé en excluant *gla*, *roi*, *shl* et *uls* ; les distances sont représentées en 3 dimensions dans la Figure 7.4 (suivant le même principe que la Figure 7.3). La partition nord/sud apparaît clairement, et il est intéressant de noter que *nwa* semble être plus proche des dialectes du nord de l’Angleterre et *brm*, de ceux du sud. Les 10 premières valeurs propres associées aux deux figures que nous venons de commenter sont présentées dans la Figure 7.5 ; elles font apparaître que 3 dimensions constituent une reconstruction acceptable de la matrice de distances originale.

7.4 Représentations arborées

Afin de mieux comprendre la structure des systèmes vocaliques des 13 dialectes de ABI, la classification hiérarchique a été utilisée (voir Section 2.4.2). Les distances de Manhattan entre les 11 monophthongues¹⁶⁷ de chaque locuteur sont calculées pour ensuite obtenir, pour chaque dialecte, un locuteur médian¹⁶⁸. La Table 7.5 récapitule les coefficients de corrélation cophénétique – entre les distances des matrices et leur équivalent dans le dendrogramme – de chaque dialecte pour la méthode de représentation retenue.

¹⁶⁷Rappelons qu’il ne s’agit que d’une appellation conventionnelle.

¹⁶⁸Il ne s’agit pas d’un véritable locuteur, mais de la matrice des $11 \times 10/2 = 55$ distances médianes entre paires de voyelles calculées à partir de tous les locuteurs d’un dialecte.

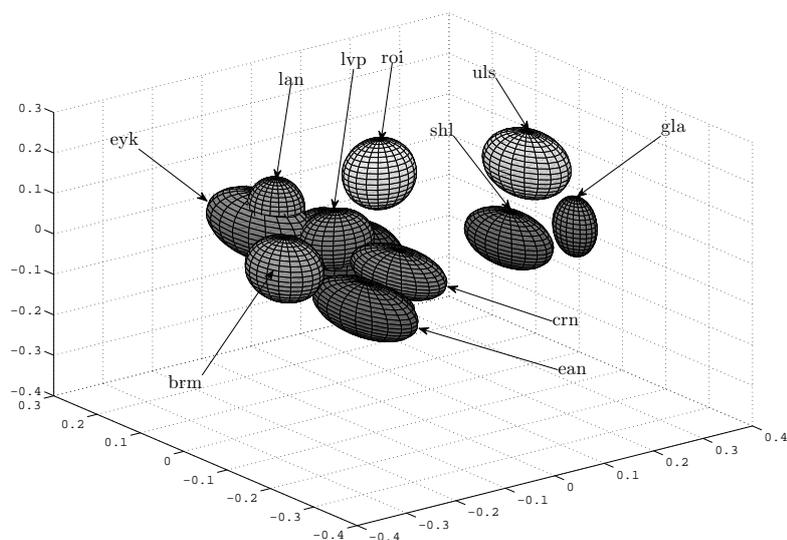


FIG. 7.3 – Ellipses des dialectes dans l'espace des 3 premières dimensions du MDS.

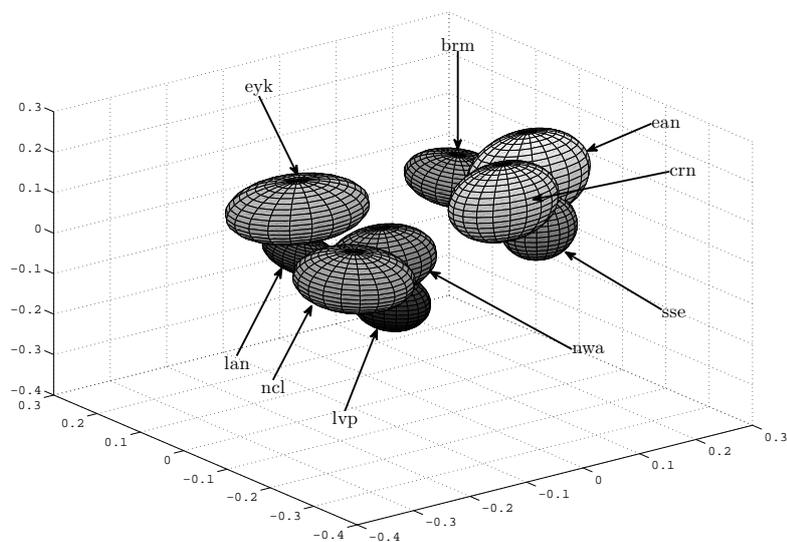
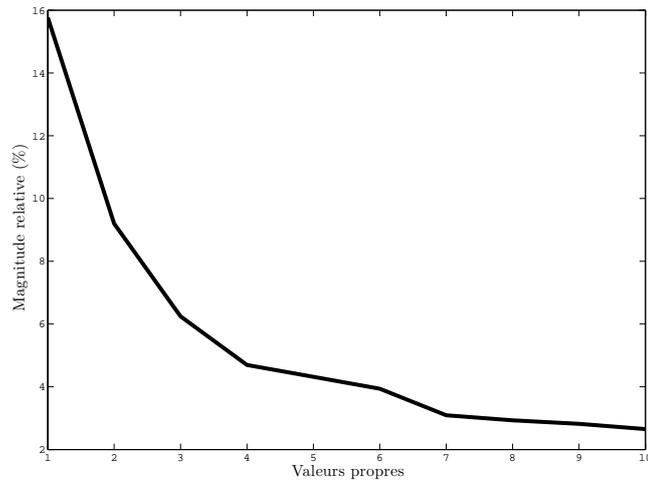
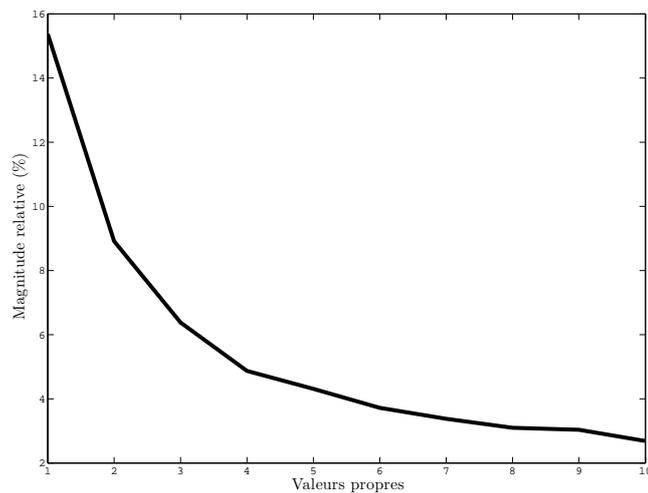


FIG. 7.4 – Ellipses des dialectes dans l'espace des 3 premières dimensions du MDS, *gla*, *roi*, *shl* et *uls* exclus.

La Figure 7.6 représente le dendrogramme des monophongues du locuteur médian de *brm*. Les graduations de l'axe des ordonnées correspondent aux distances de Manhattan originales, mais elles ne sont pas à la même échelle d'une figure à l'autre. Ceci est motivé



(a) 13 dialectes



(b) 9 dialectes

FIG. 7.5 – Valeurs propres du MDS des Figures 7.3 et 7.4.

par le fait que la méthode ACCDIST, du fait de l'utilisation de la corrélation, ne tient pas compte de la magnitude absolue des distances.

Dans la Figure 7.6, on constate la proximité de *hood* et *Hudd*, qui traduit très vraisemblablement l'absence de scission FOOT-STRUT en *brm*. On remarque également la relative proximité de *heed* et *hid*, qui corrobore le fait que la voyelle de KIT est particulièrement fermée dans ce dialecte (voir Section 3.2.1.3). Le regroupement de *who'* avec *heed* et *hid*

TAB. 7.5 – Coefficients de corrélation cophénétique.

Dialecte	Coefficient de corrélation
<i>brm</i>	0,774
<i>crn</i>	0,693
<i>ean</i>	0,756
<i>eyk</i>	0,711
<i>gla</i>	0,712
<i>lan</i>	0,827
<i>lvp</i>	0,826
<i>ncl</i>	0,748
<i>nwa</i>	0,706
<i>roi</i>	0,729
<i>shl</i>	0,734
<i>sse</i>	0,654
<i>uls</i>	0,733

n'est pas incohérent avec ce que nous avons pu observer dans le plan F1/F2 (Figures 6.38 et 6.39).

Concernant *crn*, la Figure 7.7 souligne la proximité de *hid* et *head* et celle entre *had* et *Hudd*. Pour la première paire mentionnée, nous n'avons aucune hypothèse de convergence ; il faut tout de même noter que, malgré l'aspect visuel, la jonction de *hid* et *head* se situe à une distance supérieure à celle de la jonction de *hood* et *Hudd* dans la Figure 7.6. Contrairement à la figure précédente, *who'd* est joint à *hood*, ce qui s'accorde avec la représentation à deux formants (Figures 6.49 et 6.50).

La Figure 7.8 fait apparaître la proximité de *hid* et *head* en *ean*. Puisque aucune hypothèse de différence systémique avec *sse* n'a été mentionnée, nous ne nous attarderons pas davantage sur cette figure.

La Figure 7.9 montre qu'il existe 4 paires de voyelles au-dessous de la valeur 300 en *eyk*. La distance la plus faible, entre *hood* et *Hudd*, correspond à l'absence de scission FOOT-STRUT ; la question de la distance entre *had* et *hard*, d'une part, et *head* et *heard*, d'autre part, a déjà été abordée dans la Section 6.3.5.1. On note d'ailleurs la remarquable

corrélation (au moins visuelle) entre les représentations dans F1/F2 (Figures 6.67 et 6.68) et le dendrogramme.

La distance la plus faible, dans la Figure 7.10, est celle qui sépare *hood* et *who'd*. Ceci traduit la convergence FOOT-GOOSE. On relève également la proximité de *hid* et *Hudd*, qui rappelle le caractère central de la voyelle de KIT en *gla* (voir une illustration dans la Figure 6.78).

Pour ce qui est de *lan*, la Figure 7.11 fait apparaître l'absence de scission FOOT-STRUT, caractérisée par la proximité de *hood* et *Hudd*. La proximité de *hard* et *hod* était déjà visible dans F1/F2 (Figures 6.67 et 6.67). On note que, à la différence de *eyk* (Figure 7.9), *who'd* est associé à des voyelles antérieures plutôt que postérieures (ou centrales).

Dans la Figure 7.12, on note la proximité de *hood* et *Hudd* (absence de scission FOOT-STRUT), de *hard* et *hod*, ainsi que de *hid* et *heard*; ces deux dernières paires ont déjà été remarquées lors de l'analyse formantique et font l'objet d'un commentaire dans la Section 6.3.8.1.

La Figure 7.13 fait apparaître la proximité de *hood* et *Hudd* (absence de scission FOOT-STRUT), le fait que *who'd* soit rattaché à cette dernière paire (ce qui confirme son caractère relativement postérieur), la proximité de *hid* et *head*, et leur association avec *heard*, qui prouve l'articulation antérieure de ce timbre, et la proximité de *hard* et *hod*. Tous ces points ont été mentionnés dans la Section 6.3.9.1.

La plus faible distance entre deux voyelles en *nwa* revient à la paire *Hudd* et *heard*. Leur proximité (environ 264) correspond pourtant à une différence de timbre facilement perceptible (voir Section 6.3.10.1). Il est en outre très probable qu'une différence de durée phonologique oppose ces deux voyelles : 150 contre 278 ms en moyenne pour chaque timbre, respectivement (voir Table 6.15).

Ceci soulève le problème de la pondération des variables dans la méthode ACCDIST : si la durée est un paramètre phonologiquement important, il est probable qu'il faille lui assigner un poids supérieur à celui des 52 autres paramètres dans le calcul des distances.

Nous avons exprimé la durée en centisecondes ($\mu \approx 26$ et $\sigma \approx 7$)¹⁶⁹, ce qui, étant donnée la sensibilité de la distance de Minkowski à la magnitude des paramètres, attribue de fait un poids plus important à la durée qu'aux autres paramètres en moyenne ($\mu \approx 3$ et $\sigma \approx 11$). Il est cependant probable que la pertinence de la durée sur le plan de la perception et de la phonologie nécessite une pondération supplémentaire. La proximité de *had* et *hard* a, quant à elle, était commentée dans la Section 6.3.10.1.

Pour le dialecte *roi* (Figure 7.15), le cas de la scission FOOT-STRUT a déjà été abordé dans la Section 6.3.11.1. Dans les Figures 6.116 et 6.117, on pouvait lire la proximité de *hard* et *hod*, ainsi que de *hood/Hudd* et *hoard* dans l'espace F1/F2. Les paramètres que nous utilisons ici donnent un schéma différent : en effet, la combinaison de coefficients cepstraux statiques en plusieurs points temporels de la voyelle doublée de coefficients dynamiques dérivés des premiers donne de fait plus de poids aux mouvements spectraux caractéristiques de la rhoticité que ne le faisait la représentation dans F1/F2 au milieu temporel. Ainsi, *hard* et *heard* ne sont plus appariés à *hod* et *hood/Hudd* ; la Figure 7.15 suggère en effet une première partition séparant les mots contenant un /r/ des autres.

Dans la Figure 7.16, la convergence FOOT-GOOSE est confirmée par la proximité de *hood* et *who'd*. Contrairement à *gla* (Figure 7.10), *hid* n'est pas associé à une voyelle centrale. On remarque également que, comme pour *roi* (Figure 7.15), le caractère rhotique de *shl* semble avoir prévalu dans la classification présentée.

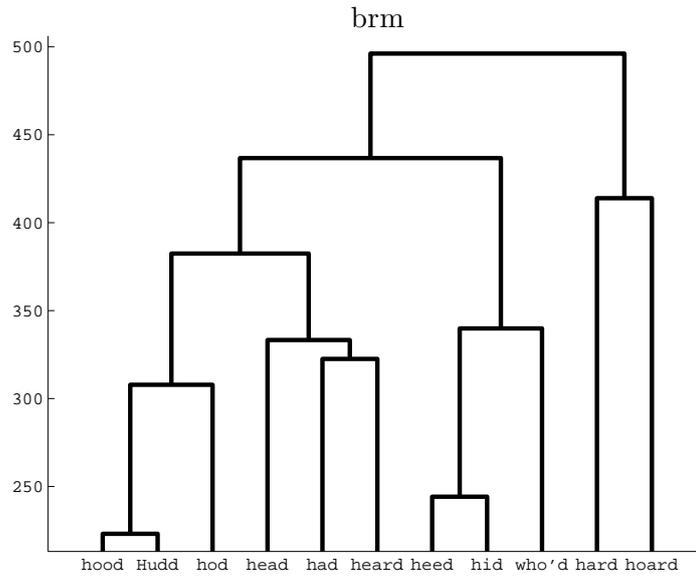
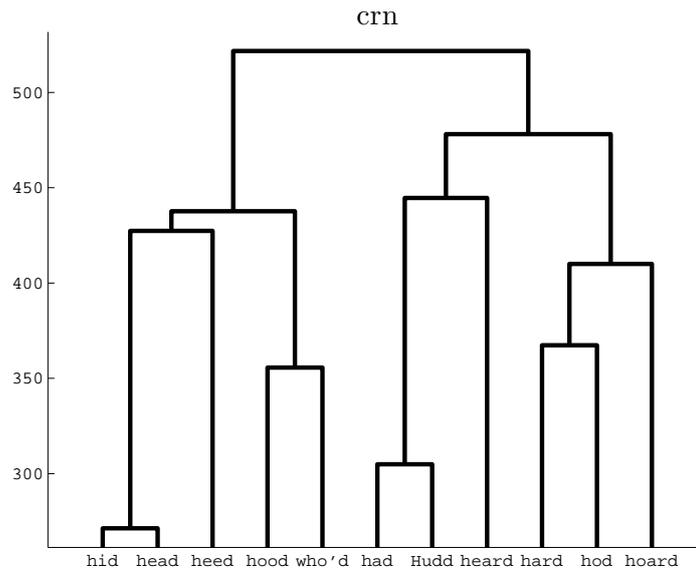
La Figure 7.17 représentant *sse* n'appelle pas de commentaire particulier.

Enfin, la Figure 7.18 confirme l'existence de la convergence FOOT-GOOSE et de la rhoticité dans le dialecte *uls*.

7.5 Détermination des voyelles discriminantes

Nous venons de le voir, notre représentation arborée des 11 monophthongues de l'anglais britannique standard pour chaque dialecte n'est pas très éloignée de ce que nous avons pu

¹⁶⁹Par convention ici, μ et σ représentent la moyenne et l'écart-type (respectivement) de l'échantillon.

FIG. 7.6 – Dendrogramme des distances entre monophthongues du locuteur médian de *brm*.FIG. 7.7 – Dendrogramme des distances entre monophthongues du locuteur médian de *crn*.

observer dans la Section 6.2 à partir de F1 et F2. Elle est cependant plus fiable que cette dernière en cela que la méthode d'extraction des paramètres est entièrement automatique (cf. tri manuel des voyelles dans la Section 6.2) et qu'elle ne nécessite aucune manipulation *post hoc* particulière visant à rectifier des erreurs de mesures (cf. Section 6.2 : utilisation de la régression).

Certains points restent cependant en suspens. En effet, la question de la pondération

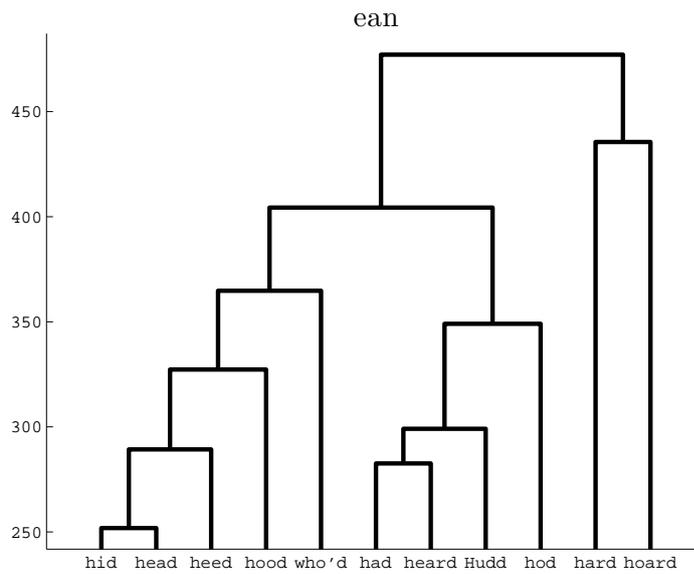


FIG. 7.8 – Dendrogramme des distances entre monophthongues du locuteur médian de *ean*.

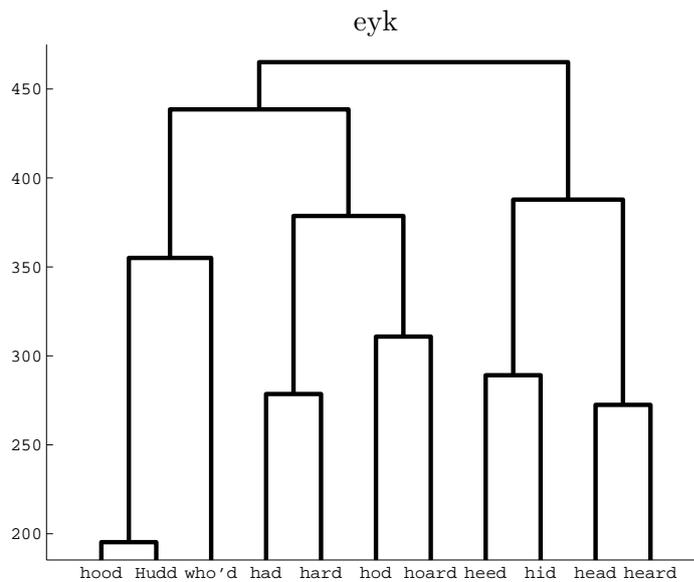


FIG. 7.9 – Dendrogramme des distances entre monophthongues du locuteur médian de *eyk*.

des variables, et notamment de la durée par rapport aux coefficients cepstraux, n'est pas traitée de façon satisfaisante. D'abord, comme nous l'avons mentionné un peu plus haut, la durée est un paramètre phonologiquement important dans la plupart des dialectes des Îles Britanniques, et elle nécessite sans aucun doute une pondération bien supérieure à n'importe lequel des paramètres cepstraux. Ensuite, certaines distances phonétiques sont plus pertinentes que d'autres pour caractériser l'appartenance d'un locuteur à un groupe

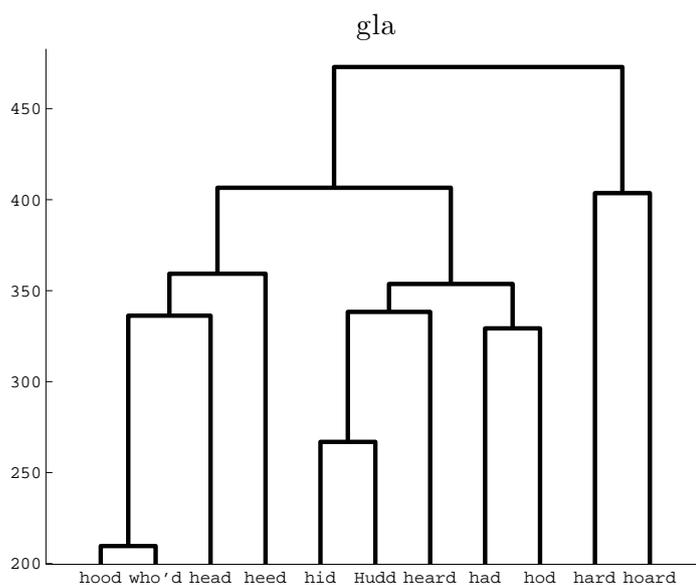


FIG. 7.10 – Dendrogramme des distances entre monophthongues du locuteur médian de *gla*.

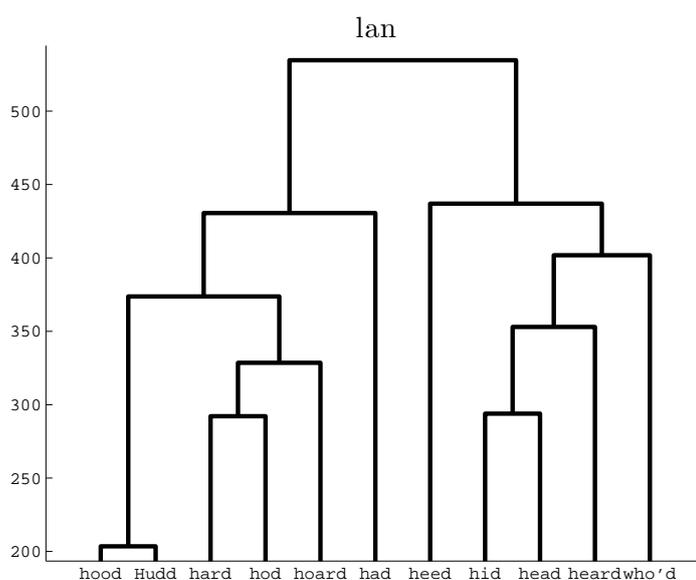


FIG. 7.11 – Dendrogramme des distances entre monophthongues du locuteur médian de *lan*.

dialectal. Ceci est illustré dans la Figure 7.19 : le locuteur *axc* de *eyk*, pourtant classé dans la catégorie *sse* dans nos deux tâches de classification automatique (Table 7.4), n'aurait pas dû l'être, car la très faible distance¹⁷⁰ entre *hood* et *Hudd* indique clairement

¹⁷⁰ Environ 166, ceci correspond tout à fait aux distances que nous avons pu observer entre *hood* et *Hudd* dans les dialectes où la scission FOOT-STRUT n'a pas eu lieu.

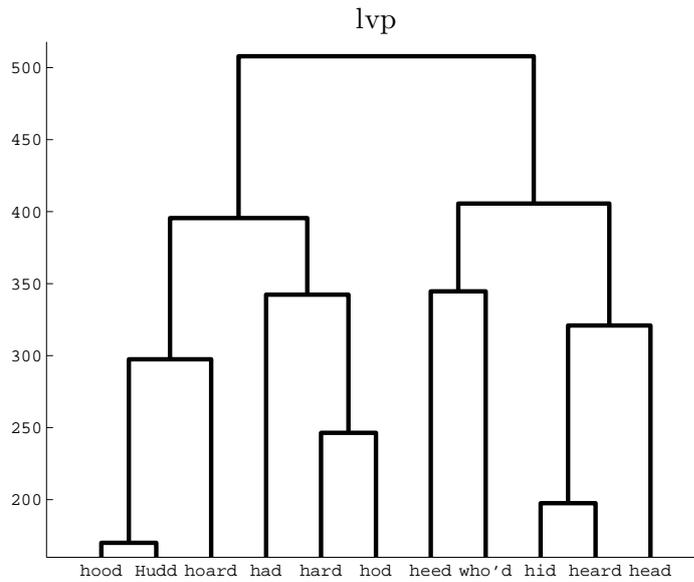


FIG. 7.12 – Dendrogramme des distances entre monophthongues du locuteur médian de *lvp*.

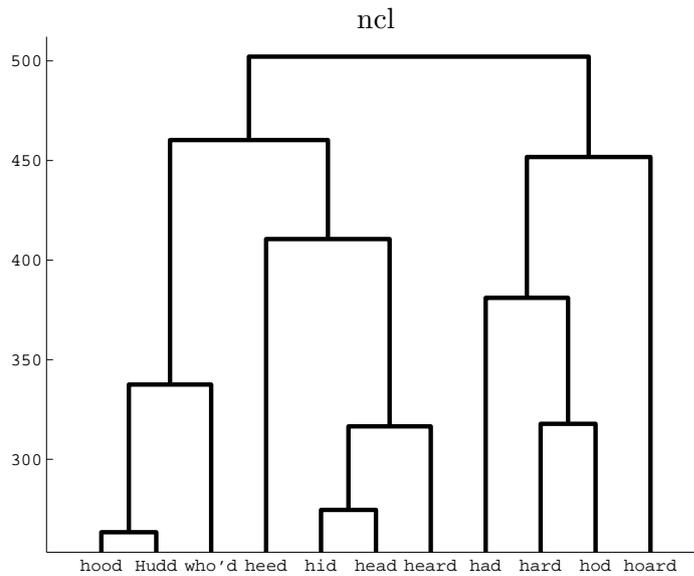


FIG. 7.13 – Dendrogramme des distances entre monophthongues du locuteur médian de *ncl*.

que ce locuteur n'a pas de scission FOOT-STRUT, ce qui aurait dû d'emblée l'identifier comme locuteur du nord de l'Angleterre. En effet, quelle que soit la similitude de ses autres distances entre voyelles avec celles de *sse*, la proximité de *hood* et *Hudd* aurait dû immédiatement le « disqualifier » en tant que candidat à la classe *sse*.

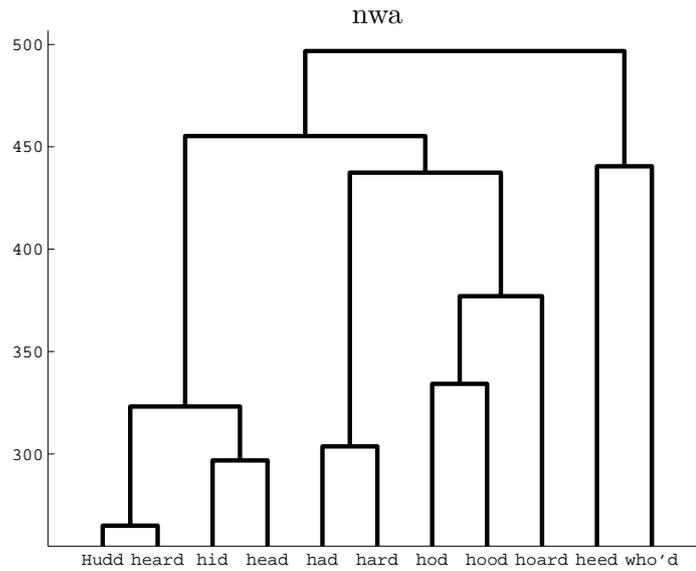


FIG. 7.14 – Dendrogramme des distances entre monophthongues du locuteur médian de *nwa*.

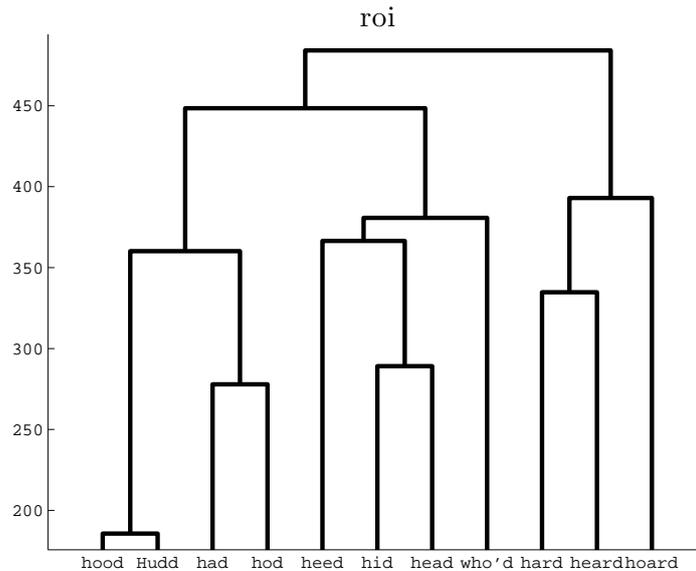


FIG. 7.15 – Dendrogramme des distances entre monophthongues du locuteur médian de *roi*.

Afin de tenter d'identifier les voyelles les plus discriminantes, la classification avec la méthode ACCDIST a été ré-estimée 19 fois (à partir des mots en /hVd/) en enlevant une voyelle à chaque fois (voir Figure 7.20). Notre hypothèse était que ceci pourrait faire apparaître des voyelles particulièrement discriminantes (et, *a contrario*, d'autres non), aboutissant ainsi à une grande variation du taux de classification correcte. Si les taux

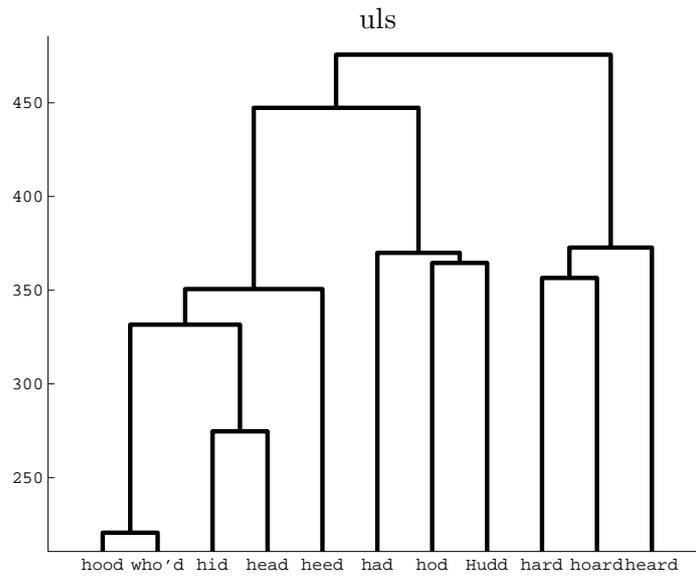


FIG. 7.18 – Dendrogramme des distances entre monophthongues du locuteur médian de *uls*.

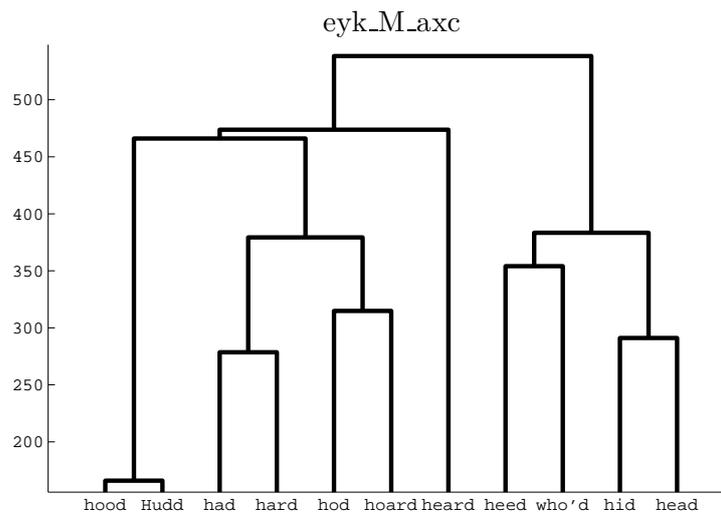


FIG. 7.19 – Dendrogramme des distances entre monophthongues du locuteur *axc* de *eyk*.

les voyelles une à une revient à ignorer la dimension systémique ; si l'on considère par exemple l'absence de scission FOOT-STRUT, il est certain que l'absence d'un phonème dans l'inventaire engendre une organisation différente du système et ceci a probablement une influence sur les distances entre toutes les voyelles.

Dans [FP06], nous avons déterminé les meilleures combinaisons de 3 à 19 voyelles en

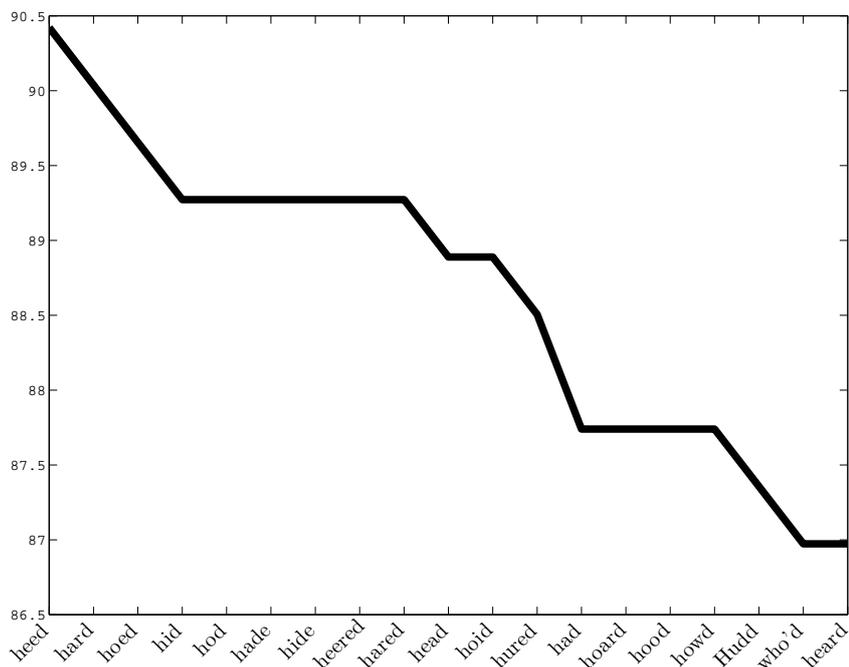


FIG. 7.20 – Pourcentage de classification correcte en fonction de la voyelle exclue.

estimant les meilleurs taux de classification (avec ACCDIST) pour toutes les combinaisons de 3 voyelles, 4 voyelles, etc. Les résultats sont rapportés dans la Table 7.6¹⁷¹. Ce tableau illustre le fait que les taux de classification correcte continuent de s'améliorer avec le nombre de voyelles, jusqu'à 15 pour les conditions hommes et femmes, et jusqu'à 16 pour la condition sexes confondus. Les meilleures combinaisons donnent une idée des voyelles les plus discriminantes, mais la méthode est imparfaite puisque toutes les combinaisons ne sont pas testées ; en effet, chaque combinaison de n voyelles est déterminée par la meilleure combinaison de $n - 1$ voyelles à laquelle on ajoute tour à tour l'une des voyelles restantes.

Jusqu'ici, nous n'avons traité que les distances entre les voyelles en contexte /hVd/. Or certaines différences systémiques ne peuvent pas être élicitées avec la liste des stimuli de forme /hVd/ ; c'est par exemple le cas de l'opposition TRAP-BATH. Dans la Figure 7.21, la distance de Manhattan moyenne entre les voyelles de *craft* et *fact* est représentée par

¹⁷¹Le tableau se lit comme suit : la combinaison de 3 voyelles conduisant au taux de classification le plus élevé (seulement chez les hommes) est *hid*, *hood* et *Hudd*, pour 4 voyelles, *hid*, *hood*, *Hudd* et *heard*, etc.

TAB. 7.6 – Meilleure combinaison de voyelles.

Nombre de voyelles	Hommes	Femmes	Sexes confondus
3	<i>hid, hood, Hudd</i>	<i>hid, heard, hide</i>	<i>hid, hood, Hudd</i>
4	<i>heard</i>	<i>had</i>	<i>heard</i>
5	<i>hide</i>	<i>Hudd</i>	<i>hide</i>
6	<i>who'd</i>	<i>hod</i>	<i>hade</i>
7	<i>hade</i>	<i>howd</i>	<i>who'd</i>
8	<i>hoed</i>	<i>head</i>	<i>hoed</i>
9	<i>hard</i>	<i>who'd</i>	<i>howd</i>
10	<i>heed</i>	<i>heered</i>	<i>hured</i>
11	<i>hoard</i>	<i>hoid</i>	<i>heered</i>
12	<i>hared</i>	<i>hured</i>	<i>heed</i>
13	<i>hod</i>	<i>heed</i>	<i>hod</i>
14	<i>head</i>	<i>hoard</i>	<i>head</i>
15	<i>had</i>	<i>hard</i>	<i>hoard</i>
16	*	*	<i>had</i>

dialecte. Les barres d'erreur représentent un intervalle de confiance de 95 %. La figure semble confirmer que seuls *ean* et *sse* ont l'opposition TRAP-BATH. La Figure 7.22, qui suit le même principe que la précédente, représente la distance entre la voyelle accentuée de *prefer* et celle de *thirty*. On note une variation importante ainsi qu'une ampleur variable (selon le dialecte) de cette variation. Par exemple, la taille limitée de la barre d'erreur de *sse* révèle que la distance entre les deux stimuli est assez homogène d'un locuteur à l'autre, alors que pour *roi*, c'est l'opposé¹⁷². Les valeurs moyennes montrent une distance élevée pour *gla*. À l'écoute, il apparaît qu'une voyelle centrale de type [ɜ] est utilisée pour *thirty*¹⁷³, et une voyelle antérieure fermée de type [ɛ] pour *prefer*. On peut émettre l'hypothèse que nous avons affaire à deux ensembles lexicaux différents. Ceci est appuyé par [SS99], qui reconnaît trois ensembles lexicaux pour l'anglais standard écossais (BIRTH, BERTH, NURSE) là où l'anglais standard n'en a qu'un. Cette différence semble également valable pour *roi*, *shl* et *uls*, mais l'analyse auditive révèle, comme le laissait

¹⁷²Un test d'homogénéité des variances ([WFH86], 182) révèle d'ailleurs que la probabilité que la variance de *roi* et celle de *sse* soient égales est $p < 0,00001$. Incidemment, c'est précisément dans des cas comme celui-ci que le test de l'ANOVA ne devrait pas être utilisé.

¹⁷³Ou peut-être s'agit-il de la voyelle de KIT qui, comme nous l'avons remarqué dans la Section 6.3.6, est très centralisée en *gla*.

supposer la variation autour de la moyenne pour *roi*, que tous les locuteurs ne font pas cette distinction. Ces résultats confirment que, malgré le bruit supplémentaire inhérent au passage lu (influence du contexte consonantique, informationnel, etc.), ce dernier contient des indices discriminants que les mots à structure /hVd/ ne permettaient pas de tester.

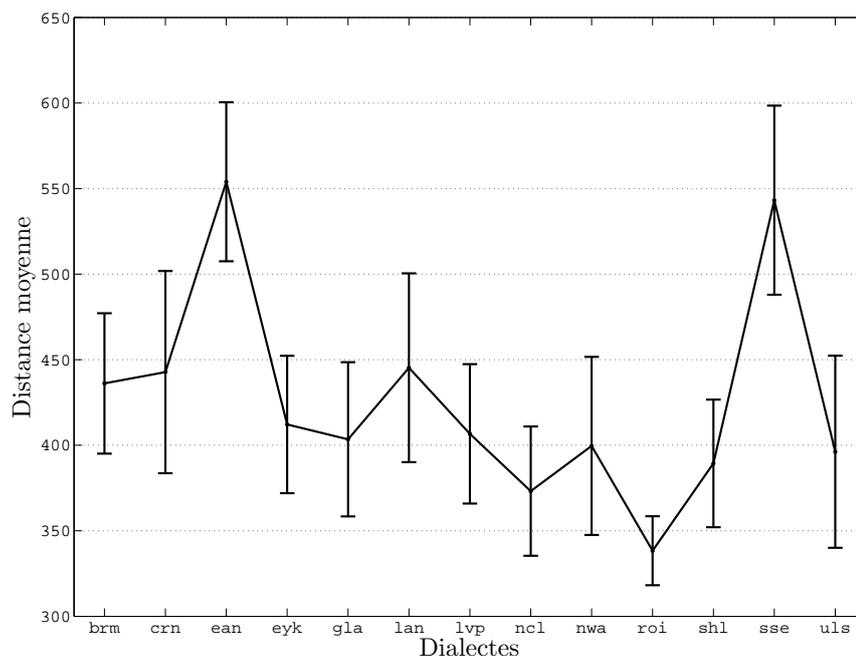


FIG. 7.21 – Distance de Manhattan moyenne par dialecte entre les voyelles de *craft* et *fact*.

7.6 Vers une interprétation dans le cadre de la théorie des ensembles flous

La classification ayant pour critère le coefficient de corrélation le plus élevé entre la matrice du locuteur en cours de test et les matrices des dialectes fonctionne sur le principe du « winner takes all » dans la mesure où, quelle que soit la valeur absolue de ce coefficient (et elle peut être relativement faible), et quelle que soit la valeur des autres coefficients (l'écart entre le premier et ces derniers varie d'un locuteur à l'autre), le locuteur est classé

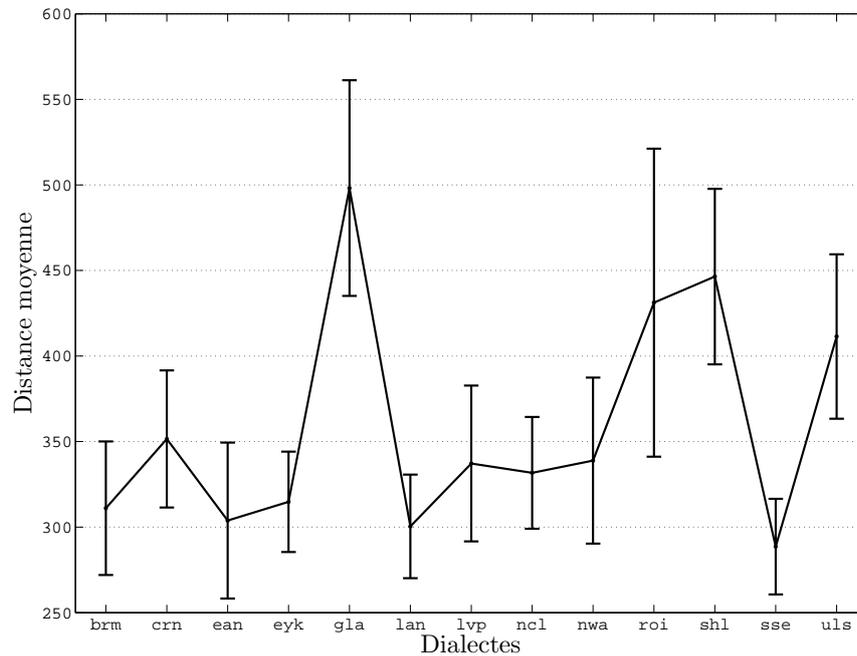


FIG. 7.22 – Distance de Manhattan moyenne par dialecte entre les voyelles accentuées de *prefer* et *thirty*.

de manière définitive dans le dialecte avec lequel il entretient la plus grande corrélation (positive). Or au vu de la variation constatée entre plusieurs locuteurs d'un même dialecte (notamment dans la Section 6.2), il serait sans doute utile de considérer l'appartenance à un dialecte dans le cadre de la théorie des ensembles flous (voir Section 2.3.3). Il est en effet possible qu'un locuteur possède des traits de deux dialectes différents. Et si ce locuteur est classé, selon la logique binaire, comme appartenant à une catégorie ou non, cela ne permet en rien de distinguer s'il s'agit d'un locuteur typique, d'un « exemplaire » central ; en d'autres termes, le principe du « winner takes all » attribue de fait le même degré de conformité au prototype à chaque locuteur présent dans une classe.

Si l'on conçoit que la matrice de distances entre les 19 voyelles en contexte /hVd/ constitue une bonne définition du dialecte de chaque locuteur, on peut obtenir une idée de l'homogénéité intra-classe par le biais d'un diagramme de silhouette. La Figure 7.23 représente la valeur de silhouette (voir Section 2.4) de chaque locuteur à partir des matrices

de distances des mots en /hVd/, les classes étant celles définies *a priori* dans le corpus. La distance entre matrices individuelles est celle utilisée pour la méthode ACCDIST, i.e. la corrélation.

On peut lire cette figure comme un ensemble de 13 diagrammes à bâton où chaque barre matérialise un locuteur. Dans chaque diagramme, les locuteurs sont classés en fonction de leur valeur de silhouette en ordre décroissant du haut vers le bas. Ce diagramme comporte, implicitement, plusieurs représentations de l'homogénéité d'une classe. D'abord, une valeur maximale relativement élevée dans chacun des 13 diagrammes indique la présence de locuteurs qu'on pourrait qualifier de typiques, puisqu'ils sont à la fois proches des autres membres de leur dialecte et distants de ceux du dialecte le plus proche. Incidemment, ces locuteurs peuvent constituer de bons centres de catégories, de bons prototypes ; la valeur de silhouette mesure en quelque sorte leur degré d'exclusivité. Ensuite, plus le diagramme est biseauté, moins le groupe est homogène. Enfin, la présence dans chaque diagramme de valeurs négatives signale que, sur la base des paramètres pris en compte, la classification serait plus cohérente si les individus présentant ces valeurs étaient classés dans une autre classe. On remarque par exemple l'homogénéité relativement bonne de *lan*, *lvp*, *roi* et *uls*. À l'inverse, *ean* comporte des valeurs positives comparativement faibles ainsi qu'une forte proportion de locuteurs ayant des valeurs négatives. Au vu de ces résultats, on peut émettre l'hypothèse que la classification n'est pas optimale. Il convient néanmoins de rester prudent : si les matrices de distances semblent être de bons descripteurs du dialecte, rien n'indique que l'information qu'elles contiennent soit exhaustive, et rien n'indique non plus que l'éventuelle information absente (e.g. le schéma intonatif) ne puisse pas réduire la distance d'un locuteur par rapport à son dialecte par un phénomène du même type que le « cue-trading »¹⁷⁴.

Afin de tester une meilleure répartition potentielle des individus dans 13 classes, nous avons procédé à une classification par le biais du *k-means clustering* (voir Section 2.4.5).

¹⁷⁴Par exemple, si la comparaison de deux locuteurs à partir de leurs systèmes vocaliques respectifs fait apparaître une grande distance entre eux, rien ne permet d'affirmer que des indices intonatifs ne pourraient pas contrebalancer (sur le plan physique ou perceptif) l'information de la distance telle qu'elle est fournie par les voyelles.

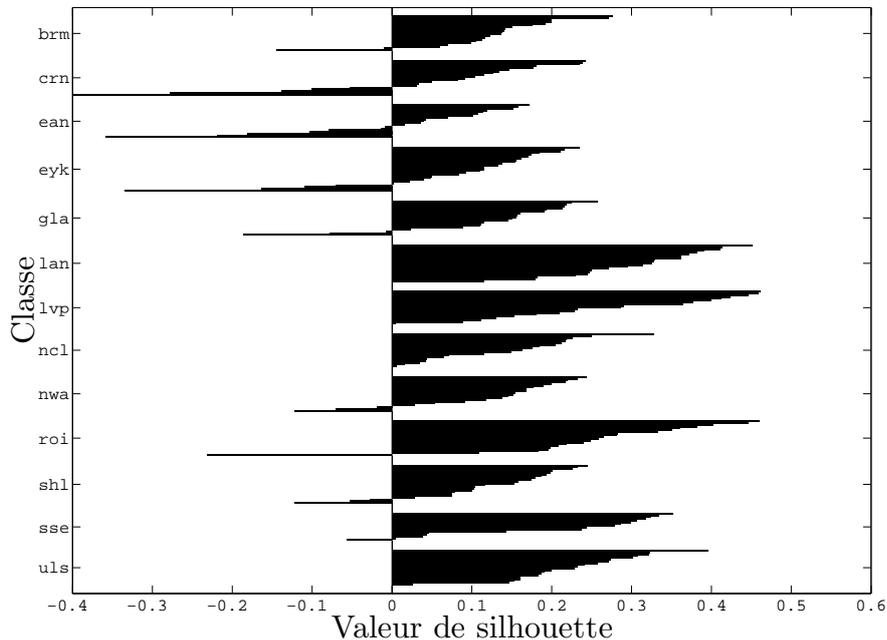


FIG. 7.23 – Diagramme de silhouette à partir des distances des mots en /hVd/ pour les classes données *a priori*.

La solution non supervisée renvoie une valeur de silhouette moyenne de 0,093, ce qui est inférieur à la valeur issue de la Figure 7.23, i.e. 0,156. Le *k-means clustering* ne trouve donc pas de partition meilleure que la partition initiale.

À ce stade, nous disposons de 3 moyens d'identifier un locuteur de référence pour chaque catégorie :

1. une mesure de la tendance centrale de chaque classe, par exemple, la matrice moyenne ou médiane des distances entre voyelles ;
2. le degré de conformité au prototype estimé par un phonéticien expert (Section 4.2), mais seulement pour les hommes ;
3. la valeur de silhouette maximale à l'intérieur de chaque dialecte.

Dans la Table 7.7, nous avons reporté les locuteurs prototypiques de chaque dialecte obtenus par les méthodes 1 et 3 de l'énumération précédente. Pour la méthode 1, c'est la matrice des distances médianes entre deux voyelles calculées à partir de tous les locuteurs

TAB. 7.7 – Locuteurs prototypiques potentiels. En exposant : $DCPE_{loc}/DCPE_{max}$.

Dialecte	Méthode médiane	Méthode silhouette
<i>brm</i>	brm_M_ars ^{2/5} /brm_F_mjd	brm_M_mrs ^{5/5}
<i>crn</i>	crn_F_dmc/crn_F_ejr	crn_M_kjr ^{4/5}
<i>ean</i>	ean_M_spd ^{1/5} /————	ean_F_ljb
<i>eyk</i>	eyk_F_sxc/————	eyk_F_yjs
<i>gla</i>	gla_M_wnh ^{3/5} /gla_M_sxm ^{4/5}	gla_M_wht ^{3/5}
<i>lan</i>	lan_F_acf/————	lan_M_jrh ^{1/5}
<i>lvp</i>	lvp_F_cxm/————	lvp_F_kxh
<i>ncl</i>	ncl_M_jxb ^{5/5} /————	ncl_F_pah
<i>nwa</i>	nwa_F_lxp/nwa_F_acl	nwa_F_bej
<i>roi</i>	roi_M_rme ^{5/5} /roi_M_aor ^{5/5}	roi_M_gwd ^{4/5}
<i>shl</i>	shl_F_jmr/shl_M_sxl ^{3/4}	shl_F_ccs
<i>sse</i>	sse_M_jph ^{1/5} /sse_M_rkm ^{5/5}	sse_F_njb
<i>uls</i>	uls_M_apc ^{3/5} /uls_F_omg	uls_M_gpd ^{3/5}

d'un dialecte. Nous avons également noté pour les hommes le degré de conformité au prototype ($DCPE_{loc}$) d'après la méthode 2 et le score maximal obtenu par un locuteur du même dialecte ($DCPE_{max}$).

La Table 7.7 fait apparaître que les prototypes potentiels suggérés par les deux méthodes sont différents. De plus, en ce qui concerne les locuteurs masculins, le parallèle avec l'analyse auditive montre que certains locuteurs avaient été jugés comme de mauvais exemplaires du dialecte concerné, ou encore, comme ean_M_spd et sse_M_jph, avaient fait l'objet de faux rejets.

Nous sommes face à deux conceptions opposées du prototype :

- le prototype probabiliste, qui, dans sa version la moins élaborée (comme dans la Figure 7.6 et les suivantes), est donné par une mesure de la tendance centrale d'une classe. Une version plus adéquate (basée sur la valeur de silhouette) prend en considération la proximité avec les objets des autres classes. Ce prototype présente cependant le défaut de masquer des différences à l'intérieur d'un dialecte : pour prendre l'exemple de *ncl* (voir Section 6.3.9.2), il est impossible de réduire les trois

réalisations attestées de la voyelle de FACE (monophthongue, diphtongue fermante et diphtongue centripète) à une valeur centrale.

- le prototype attendu de l’expert, qui a l’avantage d’être indépendant des données de l’échantillon puisqu’il s’appuie sur la représentation, quelle qu’en soit la forme, de prototypes par un phonéticien, dont le jugement est basé sur un corpus d’apprentissage¹⁷⁵ plus important et distinct du corpus d’évaluation. Dans le cas du prototype probabiliste, l’exemplaire le plus central obtient de fait un degré d’appartenance de 1 au prototype, ce qui induit un raisonnement circulaire, alors que le meilleur exemplaire d’une classe, dans le cas du prototype de l’expert, n’est meilleur qu’en termes relatifs, et peut donc présenter un degré d’appartenance au prototype inférieur à 1. Pour reprendre l’exemple de *ncl*, dans un échantillon (imaginaire) où tous les locuteurs auraient une monophthongue dans FACE, les valeurs acoustiques signalant une monophthongue seraient également attribuées à l’exemplaire central. Or cet exemplaire central serait probablement jugé par l’expert comme appartenant à *ncl* avec un degré inférieur à 1 puisque la variante typique de FACE en *ncl* est une diphtongue centripète.

Rappelons que nous tentons de définir un degré d’appartenance de chaque locuteur aux 13 classes dialectales de notre corpus sur la base de la distance d’un individu au prototype de chaque classe. Nous avons dans un premier temps déterminé, pour chaque classe, le locuteur dont la valeur de silhouette était supérieure aux autres (voir Table 7.7). Puis, chacun des 248 locuteurs restants a été comparé aux 13 locuteurs types avec un coefficient de corrélation. C’est à ce stade qu’il est légitime de se demander en quoi le coefficient de corrélation exprime la distance d’un locuteur au prototype. La relation entre le coefficient de corrélation et la distance perçue au prototype n’est probablement pas linéaire si l’on fait les hypothèses suivantes :

1. une très forte corrélation entre un individu et le prototype peut traduire une distance

¹⁷⁵La métaphore de l’apprentissage machine est conservée délibérément. Nuanceons toutefois : ce corpus d’apprentissage n’est probablement pas équilibré.

- dialectale acoustique infraliminale entre les deux ;
2. il n'est pas à exclure que l'effet attracteur du prototype se manifeste pour les classes dialectales comme il le fait, par exemple, dans les classes phonémiques ;
 3. il existe probablement un seuil en dessous duquel la corrélation ne reflète pas l'information dialectale, mais la relation qu'entretiennent de fait des matrices de distances entre les voyelles de locuteurs de l'anglais. Autrement dit, on conçoit l'existence d'une corrélation minimale entre les systèmes vocaliques des différents dialectes.

On peut donc anticiper une très faible variation perçue du degré de conformité au prototype lorsque les corrélations sont très fortes en vertu des points 1 et 2. Puis, la « prototypicalité » décroît avec la corrélation jusqu'à atteindre le seuil défini dans le point 3. Ce seuil peut par exemple correspondre à la corrélation moyenne des 248 locuteurs avec les 13 prototypes : $r \approx 0,46$. Cette relation non-linéaire est décrite par la fonction représentée dans le Figure 7.24. Les paramètres de la sigmoïde ont été ajustés de façon heuristique ; le calcul de cette fonction est donné dans l'Équation 7.2, où les deux paramètres ajustables ont été fixés à $a = 19$ et $c = 0,72$.

La valeur de $f(x)$ peut être interprétée, dans le cadre de la théorie des ensembles flous, comme le degré d'appartenance d'un locuteur au dialecte en fonction de sa corrélation au locuteur de référence (i.e. celui présentant la valeur de silhouette la plus élevée). Dans le but de simplifier le problème, les degrés d'appartenance de chaque locuteur ont été redimensionnés afin que leur somme soit égale à 1¹⁷⁶. Nous pouvons à présent illustrer le degré d'appartenance de certains locuteurs aux 13 classes et confronter ces illustrations au degré de conformité au prototype estimé par l'expert (DCPE).

$$f(x) = \frac{1}{1 + e^{-a(x-c)}} \quad (7.2)$$

¹⁷⁶Cette simplification est tout à fait contestable. En effet, il est théoriquement possible qu'un locuteur A soit très éloigné de toutes les classes du corpus et qu'un autre locuteur (B) soit très proche de toutes les classes. Redimensionner les degrés d'appartenance comme nous le faisons revient à faire disparaître l'information, pourtant capitale, qui différencie les locuteurs A et B. La recherche d'une méthode plus adéquate pourra faire l'objet de travaux ultérieurs.

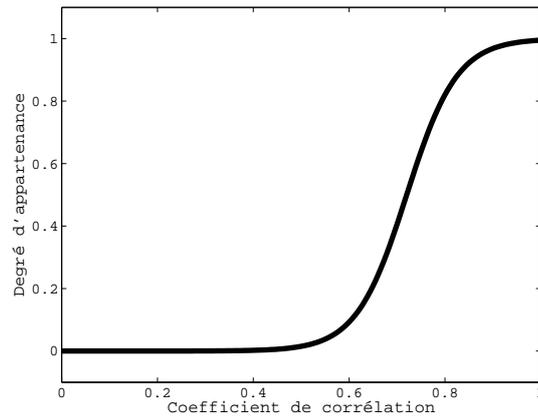


FIG. 7.24 – Fonction d'appartenance au dialecte.

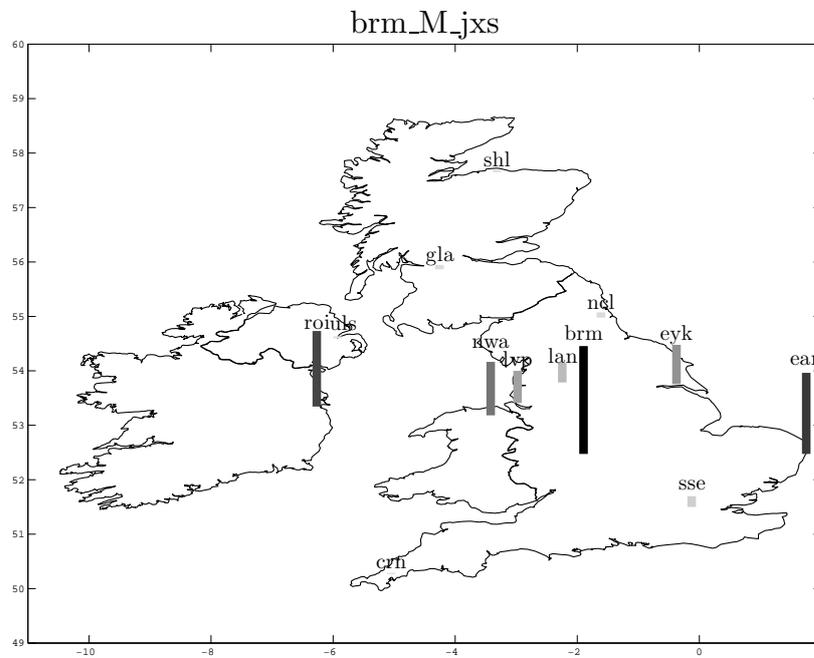


FIG. 7.25 – Locuteur *jxs* de *brm* ; DCPE : 1/5.

La Figure 7.25 (et les suivantes du même type) représente le degré d'appartenance d'un locuteur aux 13 dialectes du corpus ABI. La variété *sse* a été placée à Londres, le lieu des enregistrements, par convention. Le degré d'appartenance est représenté de façon redondante par la taille de la barre et par son niveau de gris (le noir traduit une valeur maximale). On constate que le locuteur *jxs* de *brm*, bien qu'ayant un degré d'appartenance

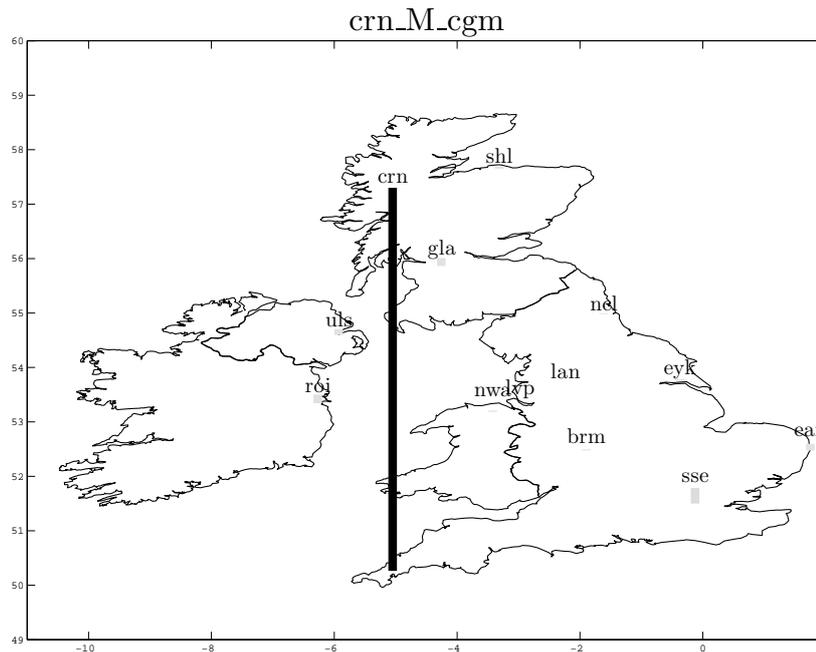
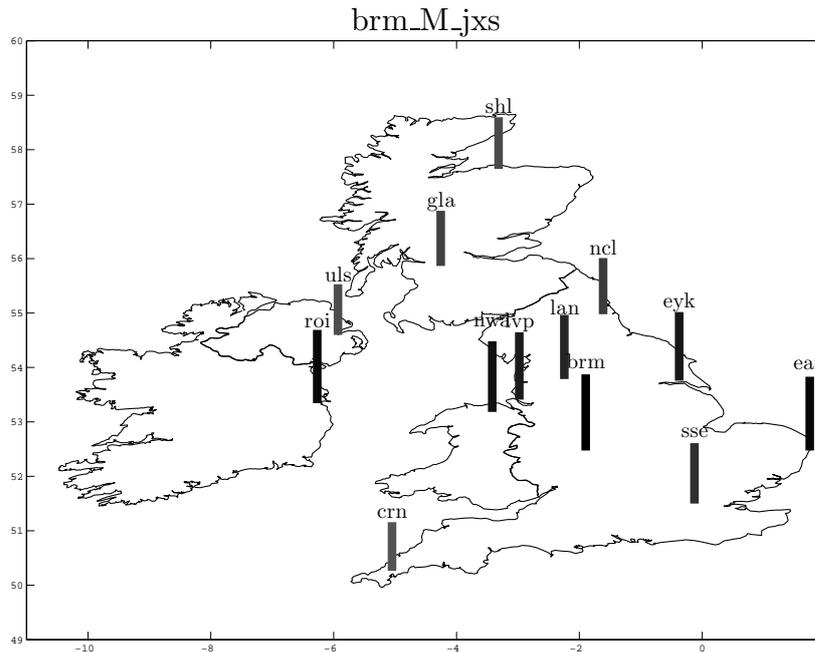


FIG. 7.26 – Locuteur *cgm* de *crn* ; DCPE : 4/5.

maximal à *brm*, présente des degrés d'appartenance relativement élevés dans les autres variétés ; ceci est conforme au DCPE de 1/5 qui lui a été attribué. À l'inverse, le locuteur *cgm* de *crn*, représenté dans la Figure 7.26 (DCPE : 4/5), a un degré d'appartenance maximal à *crn*, les autres étant très faibles.

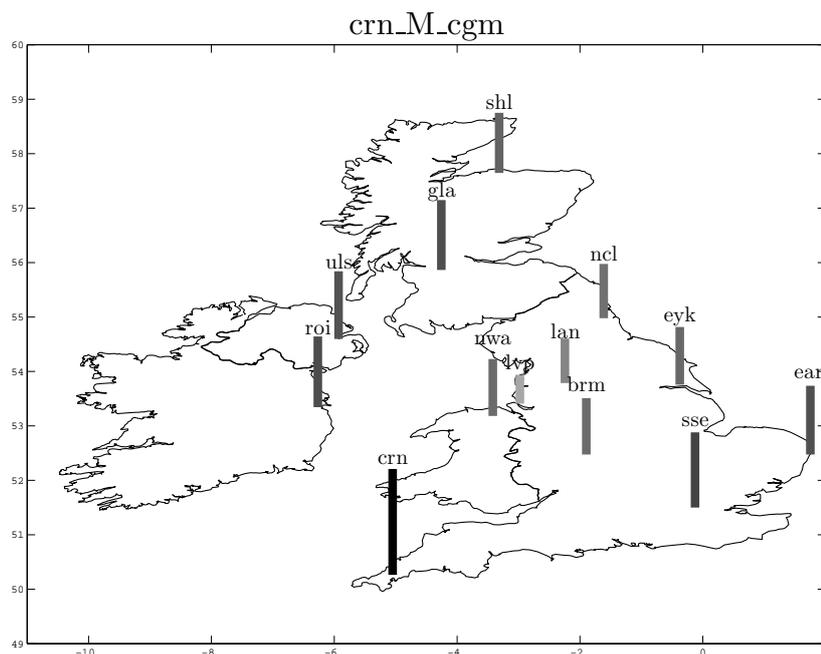
Les Figures 7.27 et 7.28 représentent les mêmes locuteurs que les Figures 7.25 et 7.26 respectivement. La différence réside dans le fait que pour les Figures 7.27 et 7.28 , la transformation sigmoïde n'a pas été appliquée aux coefficients de corrélation. On remarque instantanément que la la différence de DCPE entre les deux locuteurs apparaît clairement dans le cas de la transformation non-linéaire (Figures 7.25 et 7.26), mais pas lorsque le coefficient de corrélation brut est représenté (Figures 7.27 et 7.28).

Pour prendre un dernier exemple, la Figure 7.29 représente le locuteur *htl* de *ean*. Celui-ci à un très fort degré d'appartenance à *sse*. Il a en outre fait l'objet d'un faux rejet (considéré comme non *ean*) par l'expert, et il est systématiquement classé comme *sse* dans la classification par la méthode ACCDIST (voir Table 7.4).

FIG. 7.27 – Locuteur *jxs* de *brm*, corrélations brutes.

7.7 Synthèse des résultats et perspectives

La représentation des voyelles dans notre espace à 53 dimensions présente des avantages certains sur la représentation traditionnelle à deux formants. Outre le fait que la méthode est totalement automatique, les excellents scores de classification (Section 7.2) et la cohérence des représentations graphiques multidimensionnelles avec les descriptions récentes (Section 7.3) font de notre méthode un outil tout à fait adapté à la description des dialectes. Une des différences majeures vient du fait que, contrairement aux représentations bidimensionnelles abordées dans la Section 6.2, la caractérisation des monophthongues est également dynamique. En d'autres termes, la proximité entre deux monophthongues n'est pas seulement définie en termes de similitude spectrale en un instant donné, elle prend aussi en compte leur ressemblance sur le plan de la stabilité spectrale. Ce dernier aspect est cependant problématique : les points temporels où sont extraits les paramètres sont définis comme une fraction de la durée de la voyelle. Or les différentes phases de stabi-


 FIG. 7.28 – Locuteur *cgm* de *crn*, corrélations brutes.

lité spectrale d’une voyelle (transitions et état stable) ne semblent pas avoir une durée strictement proportionnelle à celle de la voyelle (voir [Gay78], ainsi que les différents spectrogrammes présentés dans la Section 6.2). Par exemple, si l’on compare la voyelle de *Hudd* dans la Figure 6.26(a) avec celle de *heard* de la Figure 6.27, il est évident que la coarticulation avec le [d] n’affecte pas la stabilité spectrale dans les mêmes proportions pour deux voyelles pourtant perçues comme des monophthongs. Ainsi, dans notre espace à 53 dimensions, *Hudd* est plus proche d’une diphtongue que ne l’est *heard*, ce qui ne fait pas sens sur le plan de la perception. Ce point constitue donc une piste intéressante à explorer dans des travaux futurs.

Nous avons exprimé le point de vue selon lequel chaque locuteur peut être caractérisé en mesurant son degré d’appartenance au prototype de tous les dialectes, ce degré d’appartenance étant une fonction non-linéaire du coefficient de corrélation donné par notre version de la méthode ACCDIST. Si ceci n’est probablement vérifiable qu’à travers une étude de perception, il existe toutefois de bonnes raisons de penser *a priori* que la réalité

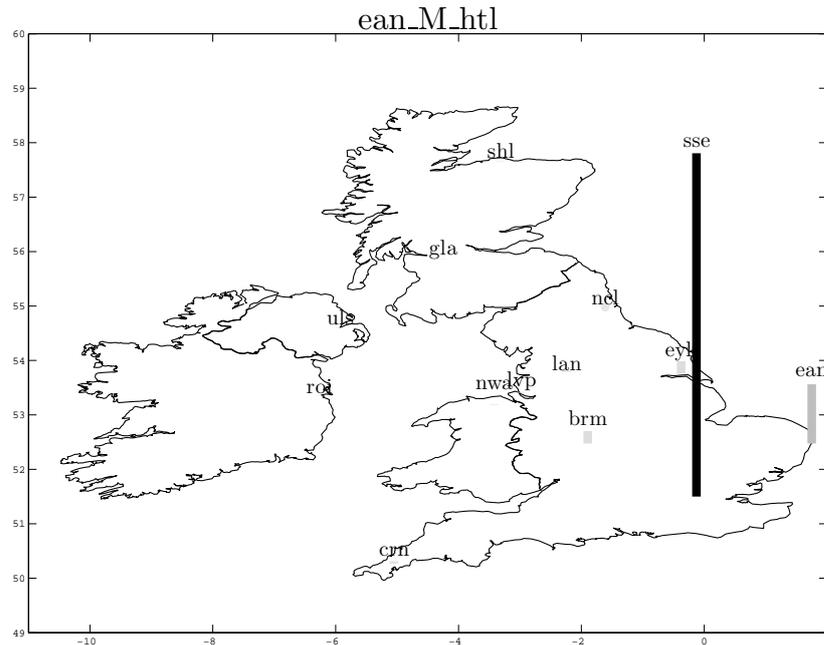


FIG. 7.29 – Locuteur *htl* de *ean* ; DCPE : faux rejet.

est plus complexe. Nous avons constaté plus haut, en nous appuyant sur la Figure 7.19, que certaines des distances entre paires de timbres (e.g. entre *hood* et *Hudd*) devraient avoir une importance bien supérieure aux autres dans l’attribution d’une classe à un locuteur. En effet, quelle que soit la proximité des voyelles du locuteur représenté dans la Figure 7.19 avec celles du prototype de *sse*, le premier n’aurait pas dû être classé dans *sse* étant donnée la faible distance séparant *hood* de *Hudd*. En d’autres termes, on peut certes considérer, comme nous l’avons fait dans les cartes du même type que la Figure 7.27, qu’un locuteur a un léger accent de Birmingham, un accent entre celui de Liverpool et de Birmingham, etc., mais il existe cependant des traits stigmatisants (c’est le cas de l’absence de scission FOOT-STRUT) qui disqualifient d’emblée certains candidats. Et malheureusement, cette information critique est probablement noyée dans la masse des 170 autres distances entre paires de mots en /hVd/. Ce défaut est vraisemblablement encore plus marqué dans le cas de la mauvaise classification du locuteur de la Figure 7.19 comme *sse* si l’on suit cette remarque de Trudgill (cité dans [Ker06]) : « it only takes one non-RP

feature for a speaker not to be a speaker of RP ».

Il est possible de contourner ce biais en fixant *a priori* des attentes intégrées dans le système de classification sous la forme de poids, comme l'ont fait Barry et collègues ([BHN89], voir Section 2.3.4). La façon dont les auteurs pondèrent la décision du classifieur est cependant quelque peu abrupte puisqu'ils n'utilisent que 3 valeurs (-1, 0, 1). On peut imaginer, pour la suite de nos travaux, le calcul de degrés de pondération plus graduels par le biais d'un système expert flou avec, en entrée, les distances entre voyelles traduisant des traits phonologiques diagnostiques et, en sortie, l'estimation d'une pondération comme aide à la décision dans l'attribution d'une classe à un locuteur.

Nous illustrons ceci avec une ébauche de système à 5 variables en entrée (traits phonologiques), 5 règles, et 3 variables de sortie (dialectes). Les variables d'entrée sont les distances entre les deux éléments de chacune des paires suivantes : TRAP-BATH, FOOT-GOOSE, NURSE-SQUARE, FOOT-STRUT et LOT-THOUGHT. Pour chaque variable d'entrée, l'univers du discours est constitué des distances acoustiques (redimensionnées entre 0 et 1) entre les deux membres d'une paire pour tous les locuteurs. Par exemple, la Figure 7.30 représente l'ensemble des distances entre TRAP et BATH. Cet ensemble est partitionné en 3 sous-ensembles : les distances relativement petites, moyennes, ou grandes. Dans la phase de projection des distances dans l'espace des sous-ensembles flous, une valeur d'entrée (en abscisse dans la Figure 7.30) est associée aux trois sous-ensembles avec des degrés d'appartenance divers. Les variables de sortie sont elles aussi subdivisées en 3 sous-ensembles, mauvais, moyen et bon, traduisant le fait que la pondération obtenue est en faveur (bon), en défaveur (mauvais) du dialecte concerné, ou qu'elle n'appuie la décision ni dans un sens, ni dans l'autre (voir Figure 7.31 pour une exemple de variable de sortie). Les règles sont au nombre de cinq :

1. SI NURSE-SQUARE¹⁷⁷ est petit, ALORS *gla* est mauvais, *lvp* est bon, *sse* est mauvais ;
2. SI FOOT-STRUT est petit, ALORS *gla* est mauvais, *lvp* est bon, *sse* est mauvais ;
3. SI FOOT-GOOSE est petit, ALORS *gla* est bon, *lvp* est mauvais, *sse* est mauvais ;

¹⁷⁷C'est-à-dire, la distance entre NURSE et SQUARE.

4. SI LOT-THOUGHT est petit, ALORS *gla* est bon, *lvp* est mauvais, *sse* est mauvais ;
5. SI TRAP-BATH est grand, ALORS *gla* est mauvais, *lvp* est mauvais, *sse* est bon.

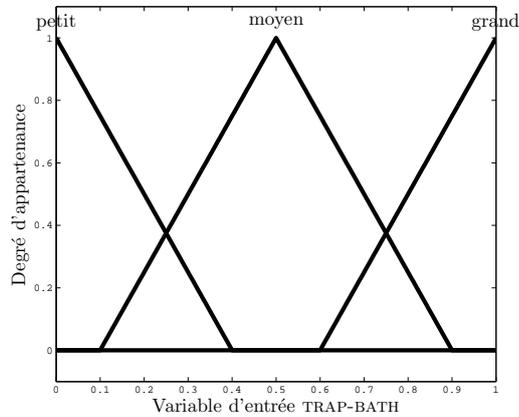
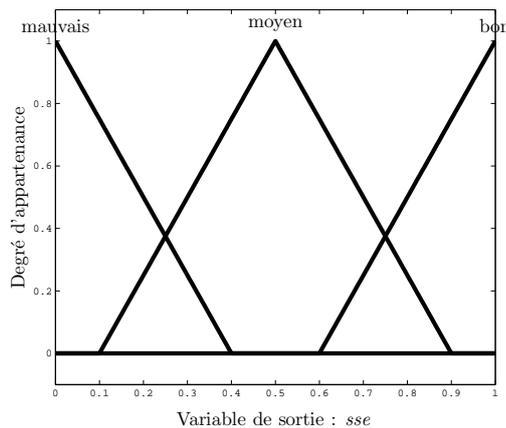


FIG. 7.30 – Exemple de variable d'entrée : la distance TRAP-BATH.

FIG. 7.31 – Exemple de variable de sortie : *sse*.

Il est ensuite possible d'accorder plus ou moins d'importance à ces règles, ce que nous avons fait, pour l'exemple, pour la règle 2, en tenant compte de la remarque de Wells mentionnée plus haut, que nous rappelons ici ([Wel82], 354) :

There are many educated northerners who would not be caught dead doing something so vulgar as to pronounce STRUT words with [ʊ], but who would feel it to be a denial of their identity as northerners to say BATH words with anything other than short [a].

Cette remarque nous a amené à donner moins de poids à la règle 2 afin que, par exemple, dans un conflit entre la règle 5 (TRAP-BATH) et la règle 2 (FOOT-STRUT), cette dernière

soit désavantagée. Une illustration de cette ébauche de système est présentée dans les Figures 7.32 et 7.33. Pour les deux figures, les 10 rectangles à gauche de l'image (ENTRÉES) représentent l'étape de la projection des variables continues initiales dans l'espace des sous-ensembles flous. Par exemple, la distance TRAP-BATH est concernée par la règle 5 ; dans la Figure 7.32, cette distance est de 0,89. Une fois projetée dans l'espace des ensembles flous (qui ont, pour l'exemple, une forme « triangulaire »), cette valeur a un degré d'appartenance de 0,725 aux sous-ensembles des grandes distances entre TRAP et BATH. À droite de la figure, le bloc de rectangles (3×5 , intitulé « ÉVALUATION », et excluant le bloc des 3 rectangles de SORTIE) évalue le conséquent de chacune des règles. Par exemple, en appliquant la règle 5, on constate que la valeur d'entrée correspond à un degré d'appartenance de 0,725 aux sous-ensembles $gla_{mauvais}$, $lvp_{mauvais}$ et sse_{bon} . Les résultats de l'évaluation des règles sont agrégés pour chaque variable de sortie, c'est ce qui correspond aux parties grisées dans les 3 rectangles de SORTIE en bas à droite. Enfin, la valeur de pondération w pour le système de classification du dialecte est obtenue en calculant le centre de gravité des zones grisées des 3 rectangles du bloc de SORTIE ; cette valeur est matérialisée sur les figures par une barre noire (pour plus de détails, voir [Neg02, JG99]). Tel qu'il est représenté dans les Figures 7.32 et 7.33, le système attribue une valeur de 0,5 à la variable de sortie dans les cas où la pondération ne penche ni en faveur, ni en défaveur du dialecte concerné. Plus la valeur tend vers 1, plus le dialecte est favorisé dans le processus de décision, et inversement pour une valeur tendant vers 0. L'exemple fictif de la Figure 7.32 montre l'activation des règles 3, 4 et 5, à des degrés divers. Les valeurs d'entrée pour FOOT-GOOSE et LOT-THOUGHT font pencher la décision en faveur de gla (et en défaveur des deux autres dialectes), et la valeur d'entrée relativement élevée de TRAP-BATH plaide en faveur de sse aux dépens des deux autres. En sortie, le poids le plus faible revient donc à lvp ($w = 0,138$) ; gla ($w = 0,415$) est légèrement pondéré à la baisse et sse ($w = 0,585$) reçoit une pondération favorable. Dans la Figure 7.33, les règles 1 et 2, qui concernent NURSE-SQUARE et FOOT-STRUT, sont également activées. Les valeurs d'entrée extrêmes pour ces deux règles font pencher la décision en faveur de

lvp ($w = 0,604$) et pondèrent à la baisse les deux autres dialectes ($w = 0,396$). Ce type de système expert a l'avantage d'être conçu très rapidement, et il fonctionne aussi bien à partir de valeurs acoustiques que d'un jugement auditif humain. Le calibrage d'un tel système à partir de données empiriques peut constituer un développement digne d'intérêt pour la suite de nos travaux.

Identifier l'origine géographique d'un locuteur et estimer son degré de ressemblance à la représentation prototypique d'un dialecte sont deux tâches bien différentes. Il est vrai que, comme nous avons pu le constater à maintes reprises au cours de ce travail, la variation phonético-phonologique pour un seul et même dialecte est très importante ; certains locuteurs ne sont en effet pas de bons représentants de leur catégorie. Le prototype probabiliste, auquel nous avons eu recours pour la classification automatique, tient compte des bons représentants comme des mauvais , ce qui conduit le système de classification à être particulièrement tolérant vis-à-vis de ces derniers. À l'inverse, le prototype attendu, celui de l'expert humain (et de l'ébauche de système expert flou que nous venons de présenter), autorise une variation plus limitée autour des prototypes. Par conséquent, faut-il considérer que ces deux approches sont inconciliables ? À notre avis, elle sont complémentaires en cela que l'une ou l'autre sera préférée en fonction de la tâche à accomplir. En effet, s'il s'agit, par exemple, d'indexer des voix publicitaires ou de comédiens selon l'accent, il va de soi que l'approche du système expert est adéquate, puisqu'elle permet un tri sur des critères pré-établis. En revanche, si le but est de prédire l'origine géographique d'un locuteur, l'approche probabiliste est plus adaptée puisqu'elle apprend les catégories à partir des données.

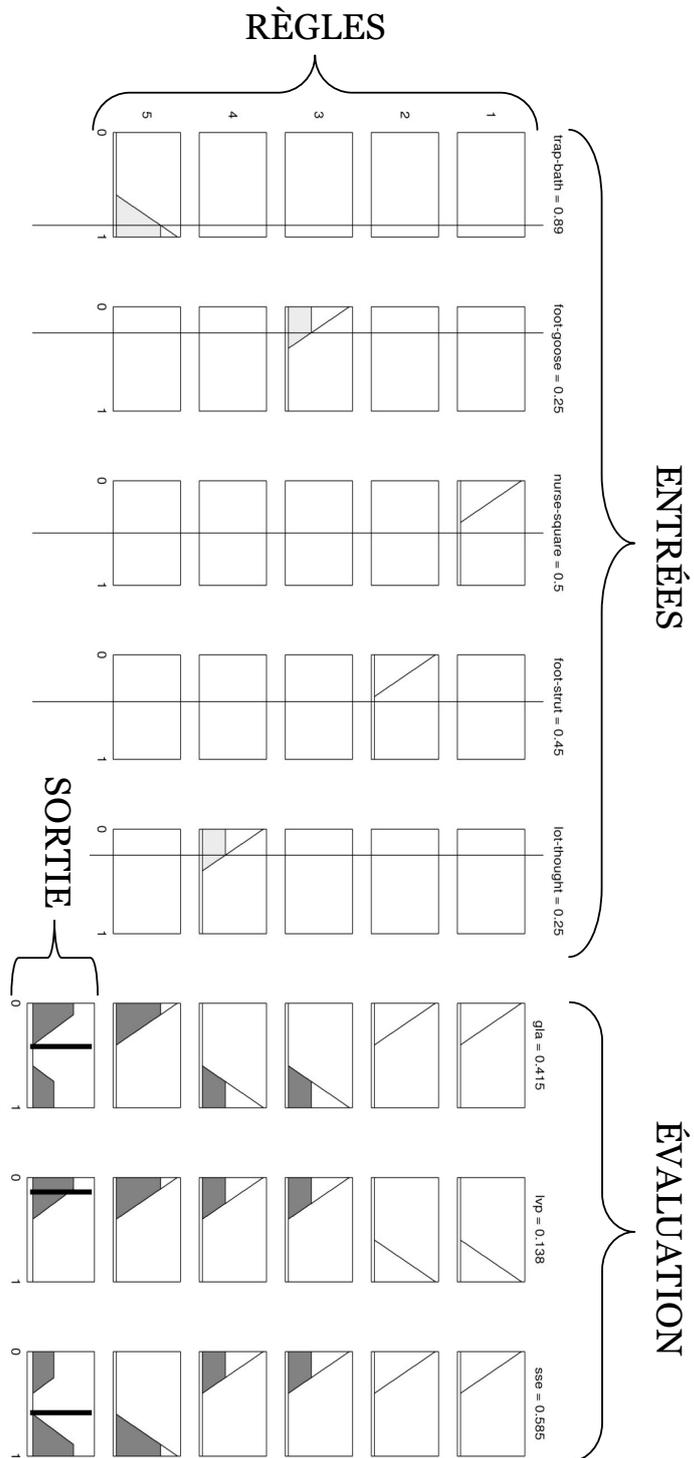


FIG. 7.32 – Exemple de système expert flou.

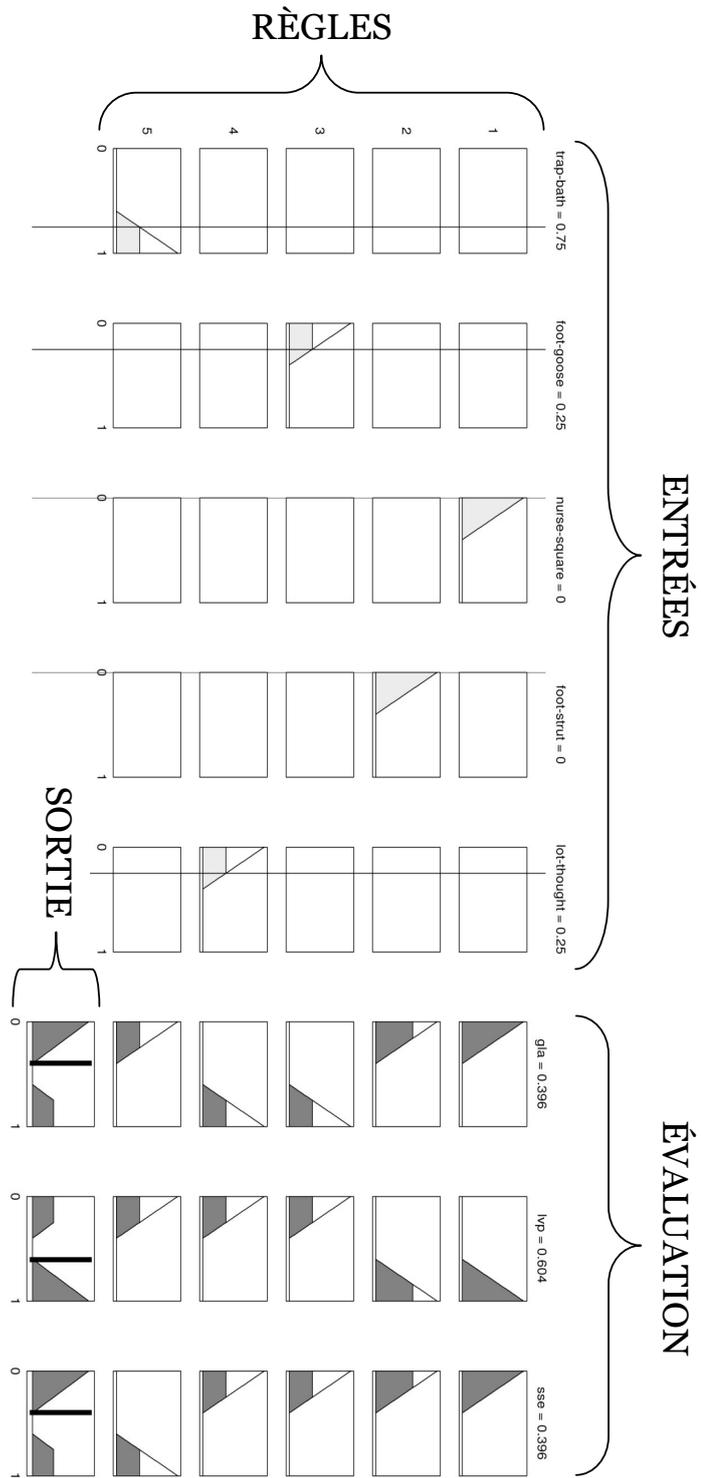


FIG. 7.33 – Exemple de système expert flou.

Chapitre 8

Conclusion générale

Notre étude avait pour double objectif de caractériser la prononciation de locuteurs et locutrices de 13 dialectes des Îles Britanniques¹⁷⁸ à partir de paramètres acoustiques, et d'évaluer la pertinence de ces paramètres à travers des expériences de classification automatique. Si l'on examine le type de paramètres utilisé, trois composantes émergent de nos travaux :

1. l'étude du rythme, qui implique la durée et l'intensité ;
2. l'analyse des voyelles dans un cadre phonético-acoustique classique, i.e. à partir des formants et de la durée ;
3. la classification automatique des locuteurs et la représentation des systèmes vocaux dans un espace multidimensionnel.

Pour l'analyse du rythme (Chapitre 5), nous avons appliqué des mesures de durée courantes dans les travaux multilingues. Si les études de dialectologie traditionnelle laissaient présager une pertinence certaine du rythme comme trait diagnostique des dialectes, nous ne disposions, *a priori*, d'aucune hypothèse forte concernant ce phénomène en raison du caractère évasif de la plupart des références sur ce thème. En effet, l'unique caractéristique fiable et mesurable sur le long terme semblait être l'absence théorique d'opposition

¹⁷⁸Corpus Accents of the British Isles (ABI) : enregistrements de listes de mots à structure /hVd/ et d'un passage lu par 264 sujets.

de quantité dans les dialectes de la zone Scots (*gla*, *shl* et *uls*)¹⁷⁹. Une première étape de l'analyse a consisté à observer les corrélations entre indices du rythme et, plus particulièrement, la corrélation entre ces indices et le débit de parole. Les indices ont été calculés à partir de la segmentation manuelle d'une partie du passage lu du corpus ABI et de la segmentation automatique de l'intégralité du même passage. Les résultats montrent que les trois espaces bidimensionnels de référence – $\%V/\Delta C$, $rpvic/npviv$ et $\%V/VarcoV$ – sont constitués de variables corrélées d'un espace à l'autre. En outre, les indices dits « normalisés » – type *Varco* ou *npvi* – sont nettement moins corrélés au débit que leurs pendants non normalisés ; ceci indique, à condition que le débit lui-même ne soit pas un trait discriminant, qu'il est préférable d'employer la version normalisée des indices du rythme calculés sur la durée.

En ce qui concerne la projection des dialectes dans les espaces bidimensionnels de référence, on remarque un chevauchement important entre classes, et, contrairement à nos attentes, la zone Scots ne forme pas un tout cohérent, distant des autres dialectes.

La pertinence des indices rythmiques de chacun de ces espaces a ensuite été testée au moyen d'une analyse discriminante. Les taux de classification correcte issus de cette analyse – en moyenne, environ 16 % et 12 % pour les segmentations manuelle et automatique, respectivement – sont, certes, très vraisemblablement supérieurs au hasard, mais ils sont cependant trop faibles pour que les indices étudiés soient considérés comme fiables. Lorsque tous les paramètres calculés à partir de la segmentation manuelle (à l'exception du débit) sont inclus dans l'analyse discriminante, le taux de classification correcte passe à presque 31 %.

Enfin, partant du constat que les mesures de type *pvi* paraissent mieux fondées que les autres, et en tenant compte du fait que les phénomènes accentuels en anglais – probablement impliqués dans la sensation de rythme – ont l'intensité parmi leurs corrélats acoustiques, nous avons introduit une nouvelle mesure : le *pvi* d'intensité. Une analyse

¹⁷⁹Rappel des abréviations désignant les dialectes : Birmingham (*brm*), Cornwall (*crn*), East Anglia (*ean*), East Yorkshire (*eyk*), Glasgow (*gla*), Lancashire (*lan*), Liverpool (*lvp*), Newcastle (*ncl*), North Wales (*nwa*), Republic of Ireland (*roi*), Scottish Highlands (*shl*), Standard Southern English (*sse*) et Ulster (*uls*).

discriminante incluant les *pvi* d'intensité, consonantiques et vocaliques, bruts et normalisés, permet d'atteindre un taux de classification correcte proche de 34 %. Les taux de classification observés dialecte par dialecte sont très variables, ce qui pourrait indiquer que certains dialectes sont plus facilement identifiables (e.g. *shl*) sur la base du rythme, tel que nous l'avons mesuré, que d'autres (e.g. *lan*).

Notre étude du rythme présente deux lacunes flagrantes : elle souffre de l'absence de données phonologiques précises sur les caractéristiques suprasegmentales de tous les dialectes comparés, d'où la faiblesse des hypothèses de départ ; et le fait que le corpus n'ait pas été recueilli pour la circonstance induit une absence de contrôle sur les facteurs tels que le débit ou les compétences de lecteurs des sujets. En ce qui concerne le premier point, il nous apparaît *a posteriori* qu'une analyse systématique et détaillée des propriétés suprasegmentales des dialectes des Îles Britanniques fait encore défaut. Si l'on exclut les aspects intonatifs ([GPNF00, GPN01, GP02, KGCR05]), il n'existe, à notre connaissance, aucune étude de grande ampleur des caractéristiques accentuelles et – plus généralement – des propriétés phonético-phonologiques susceptibles d'influer sur l'impression de rythme dans les dialectes des Îles Britanniques. Une telle étude, s'appuyant éventuellement sur des expériences de perception, représente un développement ultérieur potentiel de nos travaux. Par exemple, une analyse de la réalisation des voyelles phonologiquement réduites, qui varie très probablement d'un dialecte à l'autre, constitue une perspective de recherche digne d'intérêt.

Dans le Chapitre 6, nous avons présenté une analyse acoustique des systèmes vocaliques à travers une méthodologie classique s'appuyant sur les valeurs de F1 et F2 estimées au milieu temporel de voyelles présentées dans des mots-test à structure /hVd/. Pour chaque dialecte, et indépendamment pour les hommes et les femmes, les systèmes ont été analysés en trois temps. En suivant la taxinomie des voyelles de l'anglais standard, nous avons d'abord traité les monophthongues, puis les diphtongues fermantes et, enfin, les diphtongues centripètes. Cette partition avait pour seul but de ne pas surcharger les représentations graphiques, mais ne préjugait en rien des réalisations phonétiques des voyelles concernées

dans les différents dialectes. Les valeurs de durée incluses dans l'analyse regroupent les hommes et les femmes puisque aucune hypothèse ne justifiait un traitement différencié pour ce paramètre. Cette description comprend à la fois des représentations par dialecte pour lesquelles les formants sont transformés en Bark et centrés-réduits afin d'émuler la manière dont l'oreille les perçoit ([KPA89, KPW94]) et d'éliminer les caractéristiques spectrales individuelles ([Lob71, ASvH04]), et des spectrogrammes où la fréquence est exprimée en Hertz, ce qui autorise des comparaisons directes avec d'autres études.

La question de l'anglais standard a été abordée dans la Section 3.1, et les caractéristiques acoustiques des voyelles de cette variété ont été détaillées dans la Section 6.3.1. La description de l'anglais standard a des implications directes pour l'enseignement de l'anglais comme langue étrangère, car cette variété constitue la norme dans l'enseignement en Europe. Deux points de vue s'opposent dans l'élaboration de cette norme. En effet, on peut soit juger souhaitable d'enseigner à des étudiants la phonétique et la phonologie de l'anglais standard tel qu'il est parlé par des anglophones de leur âge, soit, comme le soutient Nolan ([Nol98], 92), considérer qu'enseigner une prononciation légèrement « passée de mode » présente les avantages suivants :

- son prestige est largement reconnu et consensuel ;
- elle est décrite dans les manuels disponibles ;
- elle présente une certaine stabilité, et reflète les innovations qui ont fini par s'imposer.

Sans véritablement trancher en faveur de l'une de ces deux options, nous suggérons que la description des voyelles de Upton ([Upt04]) et celle de Olausson et Sangster ([OS06]) sont à privilégier sur celle du dictionnaire de prononciation de Jones ([Jon03]) dans son édition de 2003. En effet, nous avons constaté des différences remarquables entre les études phonétiques récentes (dont la nôtre) et la représentation des voyelles adoptée dans [Jon03], que nous avons reproduite dans la Section 3.1.1 (Figures 3.1, 3.2 et 3.3). En ce qui concerne les monophthongues, à condition que l'on postule une parfaite correspondance entre les trapèzes vocaliques articulatoire et acoustique, on relève que la voyelle de GOOSE est

de nos jours¹⁸⁰ plus antérieure que ne le laisse supposer [Jon03]. Pour ce qui est du degré d'aperture de la voyelle de TRAP, nous rejoignons [Jon03], qui considère que le symbole /a/ serait peut-être plus approprié; [Jon03] conserve néanmoins /æ/ parce que ce dernier symbole est phonétiquement juste pour la voyelle américaine. Concernant les diphtongues fermantes, les dictionnaires de prononciation de Wells ([Wel90]) et Jones ([Jon03]) continuent de noter (dans leur représentation du trapèze vocalique) un premier élément plus antérieur dans PRICE que dans MOUTH, bien que cette différence n'apparaisse pas dans leur transcription phonétique. Nos données¹⁸¹ nous autorisent à penser que c'est précisément l'inverse qui se produit : le timbre initial de PRICE est plus postérieur que celui de MOUTH. Par conséquent, dans le cadre d'une transcription supposée guider la prononciation (par opposition à une transcription phonologique minimaliste), la solution retenue par Olausson et Sangster ([OS06]) semble être la mieux fondée : /ʌ/ et /a/ pour le premier élément de PRICE et MOUTH, respectivement. Les dictionnaires de Wells et Jones notent des diphtongues centripètes pour SQUARE, NEAR et CURE. S'il nous a été impossible de tester de façon fiable cette dernière voyelle (pour les raisons exposées dans la Section 6.3.1.3), nous pouvons néanmoins affirmer que la variante monophthongale semble être la norme pour SQUARE. En revanche, NEAR continue d'être une diphtongue, quoique très légère pour certains locuteurs. Nous ne pouvons cependant pas exclure que le contexte consonantique ou le caractère en partie artificiel des listes de mots à structure /hVd/ aient pu affecter le degré de stabilité des formants. Notre étude s'est concentrée sur les voyelles accentuées; nous n'avons donc pas analysé les voyelles réduites, les consonnes ou le schéma intonatif. Ces aspects constituent des intérêts de recherche potentiels pour la suite de nos travaux.

L'analyse des voyelles de tous les dialectes confirme une caractéristique du corpus ABI que nous ne pouvons que déplorer : nous ne disposons pas des informations individuelles nécessaires pour expliquer la variation phonétique et phonologique à l'intérieur d'un seul et même dialecte. Par exemple, il va sans dire que si la moitié des locuteurs de *brm* ont

¹⁸⁰D'après nos résultats : Sections 6.3.1.1 et 6.3.1.3, et Figure 6.28.

¹⁸¹Voir en particulier les Figures 6.16 et 6.29.

la scission FOOT-STRUT et l'autre moitié, non (Section 6.3.2.1), ces différences ne sont pas le fruit du hasard, mais bien le résultat de facteurs sociolinguistiques au sens large (i.e. incluant l'histoire linguistique, la mobilité, l'âge, etc.). Cette remarque s'applique également aux réalisations intra-dialectales variables de *who'd*, tant sur le plan du degré de stabilité formantique que sur celui de l'antériorisation (e.g. *sse*, *crn* et *nwa*), à l'absence de consensus concernant la convergence NURSE-SQUARE en *uls*, aux réalisations variables de FACE en *ncl*, etc.

Il serait rébarbatif de récapituler ici les observations faites dans le Chapitre 6 ; l'exposé présenté dans ce chapitre est délibérément concis afin que l'essentiel des caractéristiques marquantes des systèmes vocaliques de chaque dialecte soit rapidement accessible. Pour tous les dialectes, un nombre volontairement restreint de phénomènes phonétophonologiques a été présenté ; une description exhaustive de chaque timbre aurait abouti à un catalogue illisible. La synthèse des résultats du Chapitre 6 met en évidence, entre autres points marquants, une certaine ambivalence de *brm*, qui corrobore les propos de Wells ([Wel82], 364) sur le statut de zone intermédiaire de ce dialecte dans la dichotomie nord/sud en Angleterre. En effet, la représentation des valeurs médianes de chaque dialecte pour la voyelle de *Hudd* dans le plan F1/F2 fait ressortir une partition très nette entre les dialectes présentant la scission FOOT-STRUT et les autres. Or les locutrices de *brm* sont plus proches des systèmes à un phonème pour FOOT-STRUT alors que les locuteurs sont plus proches des dialectes où deux phonèmes sont attestés pour FOOT et STRUT.

L'estimation du potentiel discriminant des monophthongues, à travers la statistique F de l'ANOVA et le *NIR*, fait apparaître que *hid*, *hood*, *Hudd* et *heard* semblent comporter les voyelles les plus fiables pour une tâche de classification à 13 dialectes. Concernant plus précisément la voyelle de *Hudd*, c'est un très bon diagnostic des accents pour lesquels l'absence de scission FOOT-STRUT est consensuelle dans nos échantillons, mais le cas de *brm* prouve que sa fiabilité varie d'un dialecte à l'autre.

Pour ce qui est des diphtongues, en plus des représentations traditionnelles, qui caractérisent les voyelles par leurs valeurs de départ et d'arrivée dans F1/F2, nous avons évalué

un classifieur des dialectes deux à deux ayant comme paramètres d'entrée les valeurs de pentes de F1 et F2 (obtenues par une régression linéaire) pour la voyelle de FACE. Parallèlement à cette expérience, nous avons insisté sur la nécessité d'une approche cinématique, qui prendrait en compte des paramètres tels que la vitesse des mouvements formantiques ou la distance (par opposition au déplacement) parcourue.

Le Chapitre 7 traite de la classification automatique des locuteurs et des représentations des systèmes vocaliques dans un espace acoustique multidimensionnel. Notre version de la méthode ACCDIST (Section 7.2), qui vise à classer les locuteurs par dialectes, permet d'atteindre des taux de classification correcte d'environ 90 %, aussi bien à partir des mots à structure /hVd/ que du passage lu. De légères améliorations ont été observées lorsque nous avons procédé à une analyse linéaire discriminante ayant pour paramètres d'entrée les distances entre voyelles projetées dans un espace réduit issu d'une analyse en composantes principales. Les erreurs de classification ont été mises en relation avec l'analyse auditive du phonéticien ; cette comparaison révèle que les locuteurs mal classés par la machine avaient souvent été jugés comme de mauvais représentants de leur dialecte. Les corrélations entre matrices de distances des voyelles deux à deux ont ensuite été employées comme mesure de la proximité entre les locuteurs. La représentation graphique de ces distances par le biais des trois premières dimensions du MDS semble être un bon indicateur de la distance acoustique entre dialectes. Cette représentation (Figure 7.3) montre le regroupement de *gla*, *shl* et *uls* en une supra-région, la zone Scots, distante des autres dialectes. Le dialecte *roi* constitue, quant à lui, un groupe à part. La Figure 7.4 permet de visualiser deux groupes de dialectes, qui pourraient correspondre à la partition linguistique nord/sud de l'Angleterre. Il est intéressant de remarquer que *brm* est plus proche du groupe des dialectes du sud, et que *nwa* s'approche davantage du nord. Concernant *brm*, la variation constatée au niveau de la scission FOOT-STRUT pourrait expliquer cette proximité avec le sud linguistique. En revanche, *Hudd* et *hood* sont bien distincts en *nwa* ; ce phénomène n'explique donc pas la proximité de *nwa* avec les dialectes du nord (linguistique) de l'Angleterre. Parmi les raisons possibles de la similitude de l'espace vocalique de

nwa avec les dialectes du nord, on peut avancer le caractère postérieur de *who'd*, typique de *eyk* et *ncl* (e.g. Figure 6.144), le caractère postérieur de *hood*, constaté également à *brm*, *eyk*, *lan*, *lvp* et *ncl* (Figure 6.143), ou encore la position relativement antérieure de *hard* dans l'espace vocalique, comme en *eyk* (Figure 6.140).

Nous avons ensuite représenté les distances entre monophthongues en utilisant des dendrogrammes. Ces graphes sont susceptibles d'améliorations (e.g. pondération des variables) qui permettraient de s'approcher de la réalité perceptive ou phonologique ; mais ils prouvent néanmoins qu'une représentation arborée des systèmes vocaliques à partir de paramètres MFCC (plus nombreux que les formants utilisés en phonétique classique) constitue un moyen de visualisation alternatif tout à fait adapté à l'étude phonétique. Nous avons enfin, dans un cadre plus conjectural et heuristique, tenté d'interpréter nos résultats dans le contexte de la théorie des ensembles flous. En faisant l'hypothèse d'un effet attracteur du prototype, et en considérant que la corrélation moyenne entre locuteurs traduisait un degré d'appartenance du locuteur à un dialecte très faible, nous avons ajusté les paramètres d'une fonction sigmoïde afin de transformer les coefficients de corrélation en degrés d'appartenance. Cette représentation d'un locuteur comme membre de tous les dialectes, mais à des degrés divers, mérite, à notre avis, d'être testée expérimentalement en perception ; c'est une ligne de recherche que nous envisageons de poursuivre.

La méthode de classification automatique que nous avons employée est dépendante du texte. Elle présuppose en effet que le fragment de parole utilisé soit identique pour tous les locuteurs et qu'il soit transcrit phonétiquement, ou au moins orthographiquement pour ensuite procéder à un alignement forcé. Nous présumons qu'il est possible de transgresser la première exigence (fragment identique) sous certaines conditions tout en conservant une efficacité acceptable du système. Par exemple, si l'on dispose d'échantillons suffisamment longs pour comporter les voyelles des ensembles lexicaux les plus discriminants, on peut envisager une étape supplémentaire pendant laquelle les mots sont classés en ensembles lexicaux (ce qui implique la création *a priori* d'un dictionnaire faisant correspondre chaque mot à un ensemble lexical). Il ne s'agirait donc plus de calculer la distance

acoustique entre les voyelles de deux mots, mais plutôt la distance entre le vecteur de paramètres moyens des mots d'un ensemble avec celui des mots d'un autre ensemble. Cette approche comporte deux risques : les contextes informationnel et consonantique¹⁸² génèrent de la variation supplémentaire non pertinente et, comme nous l'avons répété à maintes reprises, l'appartenance d'un mot à un ensemble lexical peut varier d'un dialecte à l'autre. On peut néanmoins espérer que des échantillons individuels de grande taille seraient susceptibles de minimiser l'impact de ces deux faiblesses. La seconde exigence (transcription phonétique du fragment) est, quant à elle, plus essentielle que la première. En effet, l'intérêt de la méthode ACCDIST pour le dialectologue ou le linguiste repose sur le fait qu'elle est « motivée » phonétiquement, voire phonologiquement. La connaissance phonético-phonologique des systèmes vocaliques est explicitement incluse dans la méthode de classification ; et les résultats de cette méthode sont de ce fait interprétables dans un cadre de linguistique classique. Dans le cas où les échantillons sont différents et aucune transcription n'est disponible, il reste possible, théoriquement, d'appliquer la méthode ACCDIST après avoir, en amont, procédé à une reconnaissance automatique de la parole. Ces deux lignes de recherche – module intermédiaire de regroupement des mots en ensembles lexicaux et reconnaissance automatique de la parole avant classification – seront à explorer plus avant. En l'absence de transcription phonétique, des méthodes s'appuyant sur des caractéristiques acoustiques « à long terme » sont également concevables, comme la modélisation des courbes de F0 ([HYHI04]) ou l'analyse de la qualité de la voix et des « articulatory settings » ([Lav94, SS99, CR07]¹⁸³).

La question du type de paramètres acoustiques permettant une bonne caractérisation des voyelles mérite d'être abordée. Les deux ou trois premiers formants constituent une définition phonétique parcimonieuse des voyelles. Les représentations bidimensionnelles de la Section 6.2 témoignent d'ailleurs de l'intérêt de disposer d'un espace à peu de di-

¹⁸²Cette formulation est redondante : le contexte informationnel au sens large inclut le contexte consonantique.

¹⁸³Ce dernier aspect n'a été que peu étudié dans un contexte dialectal. Par exemple, Laver ([Lav94], 411) mentionne une qualité de voix « vélarisée » pour Birmingham et Liverpool, et Stuart-Smith ([SS99], 220) fait allusion à une pharyngalisation perçue comme typique de la classe ouvrière à Édimbourg.

mensions dans une optique descriptive, et constituent dans tous les cas la norme dans les publications de phonétique acoustique. De plus, par définition, les formants correspondent à des maxima spectraux, ce qui les rend théoriquement robustes au bruit. Au vu des avantages incontestables de la représentation formantique des voyelles, il est légitime de se demander pourquoi nous avons employé une représentation moins économique, et, plus généralement, pourquoi les formants ne sont pas employés dans des domaines tels que la reconnaissance automatique de la parole. Nous avons explicitement répondu à cette question dans la Section 6.2 : le peu de fiabilité des algorithmes d'estimation des formants conduit à renoncer à obtenir ces valeurs automatiquement. De Wet et collègues ([dWWB⁺04], 1782) réaffirment qu'il n'existe à ce jour aucun outil susceptible de calculer automatiquement des fréquences de formants de façon fiable. Le tri manuel des voyelles et le traitement *post hoc* par le biais de droites et courbes de régression visant à minimiser l'impact des erreurs d'estimation génèrent des contraintes rédhibitoires dans le traitement automatique de corpus de très grande taille. La première contrainte est d'ordre opérationnel : si l'on souhaite indexer des locuteurs en fonction de leur dialecte, par exemple pour des bases de données de voix publicitaires, il est impératif que le système fonctionne quasiment en temps réel ; sinon, le jugement auditif d'un expert étant plus rapide que l'extraction manuelle de formants, le système n'aurait aucun intérêt. La seconde contrainte se situe au niveau de la démarche scientifique : certes, qu'il s'agisse de seuillage, de lissage, ou d'estimateurs dits « robustes » (e.g. médiane, moyenne tronquée, etc.), les méthodes visant à écarter les valeurs déviantes sont couramment employées en analyse de données. Cependant, il nous est difficile d'évaluer le biais induit par les méthodes que nous avons employées pour corriger les erreurs d'estimation (ajustement de droites et courbes de régression aux formants), et nous ignorons par conséquent si ce biais est systématique ou s'il varie d'un dialecte à l'autre. Il est par conséquent certain que les formants ne sont pas de bons candidats pour le traitement automatique de corpus de grande taille, encore moins pour le développement d'applications fonctionnant en temps réel.

Mais qu'en est-il de l'utilisation des formants dans un cadre de phonétique acoustique descriptive traditionnelle : pour un corpus de taille modeste, l'estimation automatique des paramètres présente-t-elle un quelconque avantage ? En réalité, c'est le fait de ne pas estimer les paramètres automatiquement qui constitue un défaut épistémologique majeur car cela va à l'encontre du principe de reproductibilité (voir e.g. Popper [Pop02], 23-24). Les MFCC offrent une représentation beaucoup moins parcimonieuse et les vecteurs de paramètres spectraux que nous utilisons sont probablement surdimensionnés. Néanmoins, il est possible de les extraire de manière entièrement automatique, et Huckvale ([Huc04]) a montré, en utilisant des mesures de l'enveloppe spectrale comparables, que les taux de classification pour les dialectes de ABI avec la méthode ACCDIST étaient bien supérieurs à ceux obtenus avec des formants, et, contrairement à ces derniers, n'étaient pas sensibles aux différences de sexe. Le seul obstacle potentiel de ce type de représentation pour le phonéticien réside dans l'impossibilité de représenter graphiquement plus de 2 ou 3 dimensions. En réalité, nous espérons avoir démontré (Section 7.3) qu'il était possible de synthétiser la pléthore de paramètres de nos espaces acoustiques dans le but de produire des représentations graphiques informatives pour le phonéticien. Par conséquent, nous croyons que les MFCC peuvent constituer un outil avantageux pour la description phonétique, et espérons voir se développer leur utilisation.

Les faiblesses du corpus ABI sont multiples, et elles amènent à tirer certaines leçons concernant la constitution d'un corpus d'enregistrements pour l'étude acoustique des dialectes. Premièrement, il aurait été instructif de disposer d'échantillons de parole spontanée à partir desquels nous aurions pu effectuer, par exemple, une analyse du rythme – puisque notre méthode est intégralement automatique et indépendante du texte. Nous tenons néanmoins à réaffirmer la nécessité absolue de la parole lue, et notamment des mots isolés, dans une analyse de phonétique acoustique car le contrôle du contexte informationnel est primordial. En effet, comme nous l'avons décrit dans la Section 2.3, la réalisation acoustique d'un phonème est le résultat de contraintes imposées à une forme cible idéale. Or toute tentative d'inférer cette forme à partir de la substance observable impose que les

mêmes contraintes (notamment informationnelles) aient été appliquées aux entités comparées. La critique mettant en avant l'absence de validité écologique des listes de mots isolés est certes fondée, mais elle s'applique à toutes les études *in vitro*. Ensuite, le degré d'exhaustivité des mots-test est déterminant pour révéler les particularités allophoniques ou systémiques d'un dialecte. Nous avons abordé la question à plusieurs reprises, nous nous contenterons donc de donner l'exemple suivant : si <pair>, <per> et <purr> sont susceptibles, pour certains locuteurs, de correspondre chacun à trois ensembles lexicaux distincts en Irlande du Sud (Section 3.2.4.1), disposer de mots-test adéquats pour éliciter ces oppositions potentielles est capital. Enfin, comme nous avons pu le déplorer en plusieurs occasions, il est primordial de réunir les informations individuelles nécessaires à la compréhension de la variation intra-dialectale.

Plus généralement, cette remarque peut être reliée à la méthodologie de l'échantillonnage. Idéalement, une démarche probabiliste saine consiste à inférer les propriétés d'une population à partir d'un échantillon aléatoire. De plus, la sociolinguistique a depuis longtemps démontré l'importance d'un échantillonnage stratifié : pour chaque zone géographique, il est pertinent de partitionner les locuteurs en plusieurs sous-ensembles établis à partir de critères tels que l'âge, le niveau d'étude, l'activité professionnelle, la religion (et ses corrélats sociaux ; e.g. les différences de prononciation entre catholiques et protestants en Irlande [Hic04a]), l'appartenance à une ethnie, etc. En l'absence d'information individuelle dans le corpus ABI, la différence observée entre deux dialectes est une conséquence du facteur géographique et de facteurs sociolinguistiques. Le corpus ABI ne permet donc qu'une étude observationnelle au sens fort ; l'emploi de l'inférence statistique est exclu car il conduirait, par exemple, d'après nos résultats sur *roi*, à conclure que l'absence de scission FOOT-STRUT constitue la norme à Dublin alors qu'elle caractérise en réalité l'accent populaire.

La théorie des ensembles flous a été mentionnée à trois reprises ; d'abord dans le but d'illustrer un scénario de la projection des valeurs acoustiques des voyelles dans l'espace des catégories phonémiques (Section 2.3.3), ensuite pour modéliser les degrés d'appartenance

d'un locuteur aux 13 dialectes du corpus (Section 7.6) et enfin, pour poser les bases d'un système expert flou ayant pour objectif la classification automatique des locuteurs par dialectes (Section 7.7). Il nous a paru important d'introduire brièvement le formalisme de cette théorie car, à notre connaissance, si les phonéticiens et les phonologues parlent parfois de « fuzziness » et de « fuzzy categories », ce qu'ils entendent par là reste vague¹⁸⁴. Le raisonnement flou a également été abordé parce qu'il présente un côté très intuitif, notamment dans le parallèle qu'il est possible d'établir entre ce type de raisonnement et celui auquel l'être humain a recours au quotidien. Il serait instructif d'employer ce cadre théorique dans le but de déterminer la forme des classes dialectales. Un locuteur peut-il être membre d'un dialecte à des degrés divers ? Si oui, au fur et à mesure que le degré d'appartenance d'un locuteur à une classe s'éloigne de 1, cela implique-t-il que ce locuteur présente un degré d'appartenance croissant à une autre classe ? Quels sont les corrélats acoustiques de la proximité d'un locuteur au prototype d'une classe ? Existe-t-il un effet attracteur des prototypes ? Pour les locuteurs non prototypiques, au moins deux cas de figure sont envisageables.

Dans le premier cas, le locuteur présente les traits phonético-phonologiques de deux (voire plusieurs) accents différents, et il les présente pleinement. On peut alors suggérer que l'on a affaire à un locuteur « hybride ». Si, conceptuellement, les ensembles flous restent adéquats¹⁸⁵ pour situer ce type de locuteur dans l'espace des dialectes, les techniques de classification qui autorisent le chevauchement des groupes (« overlapping clusters », voir [ELL01], 145-154) semblent particulièrement adaptées à l'analyse de locuteurs hybrides. En effet, ces techniques permettraient, par exemple, d'identifier toutes les combinaisons de faisceaux de traits phonético-phonologiques effectivement attestées chez les locuteurs (scission FOOT-STRUT ou non, FACE réalisé comme une monophthongue ou pas, etc.), et de regrouper ces derniers en fonction de ce critère. Le nombre de catégories et les traits qui les définissent ne seraient donc plus donnés *a priori* ; cet outil pourrait se révéler très utile dans l'étude de la formation de nouveaux accents liée au nivellement dialectal.

¹⁸⁴Exception faite des travaux en psycholinguistique de Massaro ([Mas87]).

¹⁸⁵Il est en effet toujours possible d'évaluer la proximité d'un locuteur au prototype d'un dialecte.

Le second cas correspond aux locuteurs dont les traits de prononciation, systématiques ou réalisationnels, tels qu'ils sont mesurés dans l'espace continu des paramètres acoustiques, ne sont typiques d'aucune classe dialectale (anglais standard inclus). Deux interprétations sont alors possibles : soit la distance importante entre l'espace acoustique d'un locuteur et celui de tous les dialectes traduit un artefact (dû aux mesures ou à la technique de classification) qui ne reflète pas la réalité phonétique, soit on peut envisager l'existence d'une classe supplémentaire qui n'était pas identifiée *a priori*.

Si le modèle flou se révèle apte à répondre à ces questions, des expériences de perception pourraient permettre d'ajuster les paramètres des fonctions d'appartenance, voire de calibrer un système expert d'inférence floue. Nous envisageons de développer un tel système pour la classification des dialectes, qui reposerait donc sur ce que nous avons appelé jusqu'ici le prototype « de l'expert », par opposition au prototype « probabiliste ». Ce dernier est en effet totalement déterminé par les caractéristiques de l'échantillon de chaque classe disponible pour l'apprentissage du modèle. Or, comme nous l'avons déjà mentionné, à moins de recueillir des échantillons stratifiés sur critères sociolinguistiques, ou un échantillon aléatoire de très grande taille¹⁸⁶, le modèle résultant sera tributaire du biais introduit lors du recueil de données et ne présentera donc pas les propriétés de généralisation propres à un échantillon statistique fiable. C'est pour cette raison que, en l'absence de bases de données de très grande taille et d'échantillons fiables, nous pensons qu'il est possible de mettre au point un système expert flou pour la classification automatique des dialectes, ce qui aurait pour avantages :

- de s'affranchir de la phase coûteuse du recueil des données pour l'apprentissage d'un modèle statistique ; les dialectes des Îles Britanniques sont en effet assez bien documentés ;
- d'écarter les biais d'échantillonnage ;

¹⁸⁶Et même dans ce cas-là, un échantillon demeure très imparfait : l'inférence statistique pose qu'il est possible de prédire les paramètres d'une population avec un certain degré de confiance à partir d'un échantillon aléatoire de cette population ; certes, mais les locuteurs de dialectes nous semblent assimilables aux cygnes de Popper : « no matter how many instances of white swans we may have observed, this does not justify the conclusion that *all* swans are white » ([Pop02], 4).

-
- de mieux correspondre à certaines applications, telles que l’indexation de voix d’acteurs ou publicitaires en fonction de l’accent. En effet, ces tâches nécessitent la détection d’accents stéréotypiques, voire caricaturaux.

En plus de l’analogie éventuelle des classes dialectales avec des ensembles flous et de la mise en place d’un système expert pour la classification des dialectes, nous envisageons pour la suite de nos travaux un effort de théorisation de l’interface phonétique/phonologie dans le cadre de la théorie des ensembles flous. Dans la Section 6.4, la Figure 6.152 représentait un arbre de décision illustrant un scénario potentiel du lien entre les catégories phonémiques et le signal acoustique. La simplicité et le caractère très intuitif de cette représentation en font un outil qui, selon nous, pourrait assister le linguiste dans sa recherche. D’ailleurs, il est possible d’inclure le raisonnement flou dans la conception de tels arbres ([OW03]). La projection des données acoustiques dans l’espace des catégories phonémiques est généralement malaisée en raison de la grande variation des paramètres physiques. Nous considérons cependant, comme nous l’avons démontré dans la Section 2.3.1, que les connaissances actuelles permettent en théorie de prédire les fluctuations des paramètres physiques induites par la variation du contexte informationnel (voir les travaux de [FLM99, JM00, Ay100, AT06]), ainsi que celles résultant de propriétés individuelles des locuteurs (e.g. [ASvH04] et méthode ACCDIST) : il semble donc opportun d’explorer le potentiel d’une phonologie empirique construite autour de la notion d’ensembles flous.

Annexe A

Annexe : passage lu du corpus ABI

Nous avons tenté de rétablir la ponctuation, mais rien ne garantit qu'elle corresponde au texte que les sujets ont réellement lu.

Fichier 1 When a sailor in a small craft faces the might of the vast Atlantic Ocean today, he takes the same risks that generations took before him. But, in contrast to them, he can meet any emergency that comes his way with a confidence that stems from a profound trust in the advances of science. Boats are stronger and more stable, protecting against undue exposure. Tools and instruments are more accurate and more reliable, helping in all weather and conditions. Food and drink are better researched and easier to cook than ever before.

Fichier 2 The Vikings looked for new lands with fleets of small ships, each powered by thirty to forty men pulling on sixteen-foot oars. They were guided by nothing but their knowledge of the stars and an unshakeable faith in their gods. Over half a millennium later, Spaniards, Portuguese, and Elizabethan freebooters had improved their knowledge a little, reduced their gods to one, and dispensed with oarsmen in favour of a total reliance on the wind. But despite their visions of great deeds, they were loathe to venture out far by themselves.

Fichier 3 Were they less adventurous or more modest? Did they prefer the reassuring thought of help from sister-ships to the heightened glory of single-handed exploits? But was their glory in fact any less? Did any member of the crew receive one iota of thanks, recognition or fame? We have no means of measuring of course; but the truth is, none of the commanders of the ships which accompanied Francis Drake are remembered today. No more than the type of sail, the make of radio or navigation instrument, the self-steering equipment or the medicines and food supplied to our modern adventurers will be remembered in four hundred years time.

Annexe B

Annexe : rythme et voisement

Dellwo et collègues ([DFA07]) postulent que si le nourrisson est capable de distinguer des langues ayant des rythmes différents, il s'appuie probablement sur des unités de base encore moins complexes que les consonnes ou les voyelles. C'est pourquoi les auteurs proposent d'utiliser les intervalles voisés et les intervalles non voisés pour mesurer le rythme, ou plus exactement, le coefficient de variation de la durée des intervalles non voisés sur une portion de signal dénuée de pause et le pourcentage de voisement sur cette même portion. Les auteurs obtiennent une séparation assez nette entre, d'un côté, l'allemand et l'anglais, et de l'autre, le français et l'italien, à partir de ces deux dimensions.

En utilisant le passage lu du corpus ABI, nous avons dans un premier temps procédé à une détection des pauses à l'aide d'un programme en Tcl/Tk utilisant les fonctions du Snack Toolkit. Puis, une fois les pauses exclues, le logiciel Praat nous a permis de détecter les trames voisées. Enfin, le coefficient de variation de la durée des intervalles non voisés, *VarcoNoVo*, et le pourcentage de voisement *%Vo* ont été calculés sur chaque portion de parole s'étendant d'une pause à la suivante. Dans la Figure B.1, chaque dialecte est représenté par sa médiane et son écart interquartile dans les deux dimensions. Cette figure fait clairement apparaître que ces indices ne sont pas fiables pour la discrimination inter-dialectale ; nous n'avons donc pas exploré cette piste plus avant.

Index

A

acquisition de la phonologie, 22–23

Aitken, loi d', voir *Scottish Vowel Length*

Rule

Akaike, critère d', 177–178

analyse discriminante, 62–63, 148, 153–155,
159, 319, 358–359

analyse en composantes principales, 62, 168,
319, 363

ANOVA, 59, 139, 293–297, 337

B

Bark (formule de conversion), 30

BATH-broadening, 20–21, 83, 86–87, 90, 92,
97, 103

Beatles, 5

binomial, test, 148, 153, 154

Blair, Tony, 94

C

classification hiérarchique, 61, 323–328

coefficients cepstraux, voir MFCC

Connery, Sean, 4

cornique (langue celtique), 91

corrélation

cophénétique, 61, 323

de Pearson, 141–144, 146, 149, 153, 157,
160, 286, 289, 306, 315, 322, 325,
338, 340, 343–344, 346, 348, 358, 363,
364

de Spearman, 286

covert contrasts, 22–23

critère d'Akaike, voir Akaike

D

dendrogramme, voir classification hiérarchique

dérhoticisation, 23, 103

distance

de Mahalanobis, 297

de Manhattan, 59

de Minkowski, 59–60

euclidienne, 59

E

Élizabeth II, 4, 80

emprunts, 19, 76

ensembles flous, voir théorie des ensembles
flous

ensembles lexicaux (définitions et exemples),
12–14

entropie, 34–36

erse (langue celtique), 104

F

fou, système expert, voir système expert fou

fous, théorie des ensembles, voir théorie des ensembles fous

G

gallois (langue celtique), 99, 141

Garde, principe de, 17–18

Grand Changement Vocalique, 15–16, 76, 101

The Grand Old Duke of York, 81

H

H-dropping, 84, 90, 106

happy-tensing, 79, 84, 88, 94–96

Herzog, principe de, 17

homogénéité des variances, test de l', 337

hyper-/hypo-speech, voir théorie H&H

hypercorrection, 17, 20, 55, 85

hyperspace effect, 34

K

k-means clustering, 63, 158, 340–341

Kruskal-Wallis, test de, 139

L

L-vocalization, 95, 98

Loach, Ken, 87

locus, 32

M

MFCC, 310–312, 364, 367

multidimensional scaling, 168–169, 172, 322–323, 363–364

N

Neighborhood Activation Model, 35–36

NG-coalescence, 92

nivellement dialectal, 54, 369

normal information radius, 297–299

normalisation du locuteur, 31–32, 55, 57, 169, 182, 311, 315

P

P-centers, 127

perception catégorielle, 30, 44, 47

polynômes, voir régression

prestige, 4, 10, 20, 55, 69, 81, 87, 93, 97, 105, 318, 360

prototype

effet attracteur, 43–44, 46

focal, 40–42

Q

quantique, théorie, voir théorie quantique

quasi-convergences, 21–22

R

r de liaison, 79, 90, 97, 138

r intrusif, 79, 90, 97

r labiodental, 82, 90, 97

Received Pronunciation, 4, 18, 21, 53, 68–69, 73–76, 78, 79, 81, 82, 92, 96, 106, 134–136, 349

redondance, 36–38, 225, 300

régression, 64–65, 172–181

S

saillance (*markedness*), 53

scissions et convergences

FOOT-GOOSE, 17, 55, 84, 103, 108, 257–258, 287, 300, 327, 328, 334, 350–352

FOOT-STRUT, 20–21, 55–57, 83, 85–90, 92, 93, 97, 106, 207–209, 230, 243, 257, 271, 281, 288, 296–297, 325–328, 332, 334–335, 349–352, 362, 368

LOT-THOUGHT, 39, 103, 108, 276, 350–352

NEAR-SQUARE, 95

NORTH-FORCE, 103, 106

NURSE-NORTH, 97

NURSE-SQUARE, 90, 92, 93, 106, 108, 229, 251, 284, 350–352, 362

PRICE-CHOICE, 17, 21, 93, 211, 226

STRUT-Schwa, 100

Scottish Vowel Length Rule, 25, 102–103, 135, 276

seuil différentiel de perception

de durée, 127

des formants, 29–30

sigmoïde, fonction, 49, 344, 346, 364

silent centers, 27–28

silhouette, 63–64, 158

Smooth Signal Redundancy Hypothesis, 37

Smoothing, 79, 95, 101

système expert flou, 348–353

T

TH-fronting, 82, 90

TH-stopping, 93, 107, 108

Thatcher, Margaret, 97

théorie des ensembles flous, 44, 46–51, 339–346, 348–353, 368–371

théorie H&H, 34–35, 37

théorie quantique, 24, 41–42

V

variance, analyse de la, voir ANOVA

variance, homogénéité, voir homogénéité des variances, test de l'

VOICES (BBC project), 10, 91

Y

Yod-coalescence, 138

Yod-dropping, 84, 93–95

Bibliographie

- [Abe64] David ABERCROMBIE : Syllable quantity and enclitics in English. *In* David ABERCROMBIE, D. B. FRY, P. A. D MACCARTHY, N. C. SCOTT et J. L. M. TRIM, éditeurs : *In Honour of Daniel Jones*, pages 216–222. Longmans, Londres, 1964.
- [Abe65] David ABERCROMBIE : *Studies in Phonetics and Linguistics*. Oxford University Press, Londres, 1965.
- [AH02] Pongtep ANGKITITRAKUL et John H. L. HANSEN : Stochastic trajectory model analysis for accent classification. *In ICSLP*, pages 493–496, Denver, 2002.
- [AH03] Pongtep ANGKITITRAKUL et John H. L. HANSEN : Use of trajectory models for automatic accent classification. *In Interspeech-Eurospeech 2003*, pages 1353–56, Genève, 2003.
- [ASvH04] P. ADANK, R. SMITS et R. van HOUT : A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5):3099–3107, 2004.
- [AT06] Matthew AYLETT et Alice TURK : Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119(5):3048–58, 2006.
- [AT07] Jalal-eddin AL-TAMIMI : *Indices dynamiques et perception des voyelles : étude translinguistique en arabe dialectal et en français*. Thèse de doctorat, Université Lyon 2, 2007.

- [AW04] Ulrike ALTENDORF et Dominic WATT : The dialects in the South of England : Phonology. *In* Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 178–203. Mouton de Gruyter, Berlin, 2004.
- [Ay100] Matthew AYLETT : *Stochastic Suprasegmentals. Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech*. Thèse de doctorat, University of Edinburgh, 2000.
- [BBGB95] F. BELL-BERTI, C. E. GELFER et M. BOYLE : Utterance-final lengthening : The effect of speaking rate. *In ICPHS*, pages 162–165, Stockholm, 1995.
- [BD86] André-Pierre BENGUEREL et Janet D’ARCY : Time-warping and the perception of rhythm in speech. *Journal of Phonetics*, 14(2):231–246, 1986.
- [Bea04] Joan BEAL : English dialects in the North of England : Phonology. *In* Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 113–133. Mouton de Gruyter, Berlin, 2004.
- [BF85] Z. S. BOND et J. FOKES : Non-native patterns of English syllable timing. *Journal of Phonetics*, 13(1):407–420, 1985.
- [BH04] Caroline BOUZON et Daniel HIRST : Isochrony and prosodic structure in British English. *In Speech Prosody*, pages 223–226, Nara, Japon, 2004.
- [BHN89] W. J. BARRY, C.E. HOEQUIST et F. J. NOLAN : An approach to the problem of regional accent in automatic speech recognition. *Computer Speech and Language*, 3:355–356, 1989.
- [Bil03] Michael BILTON : *Wicked Beyond Belief. The Hunt for the Yorkshire Ripper*. HarperCollins, Londres, 2003.
- [BJFL⁺03] Alan BELL, Daniel JURAFSKY, Eric FOSLER-LUSSIER, Cynthia GIRANT, Michelle GREGORY et Daniel GILDEA : Effects of disfluencies, predictabi-

-
- lity, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113(2):1001–24, 2003.
- [Bla83] Anthony BLADON : Two-formant models of vowel perception : Shortcomings and enhancements. *Speech Communication*, 2(4):305–313, 1983.
- [Blo70] Bernard BLOCH : *Bernard Bloch on Japanese*. Yale University Press, New Haven, 1970.
- [Boi80] Claude BOISSON : *L'Accentuation des composés en anglais contemporain, avec quelques contributions à l'accentologie générale*. Thèse de doctorat, Université Paris 7, 1980.
- [BR03] William J. BARRY et Michela RUSSO : Measuring rhythm. Is it separable from speech rate? *In Interfaces prosodiques*, pages 15–20, Nantes, 2003.
- [Bri02] David BRITAIN : Phoenix from the ashes? : The death, contact and birth of dialects in England. *Essex Research Reports in Linguistics*, 41:42–73, 2002.
- [Bro76] David J. BROAD : Toward defining acoustic phonetic equivalence for vowels. *Phonetica*, 33(6):401–424, 1976.
- [Bus67] Clara N. BUSH : Some acoustic parameters of speech and their relationships to the percpetion of dialect differences. *TESOL Quarterly*, 1(3):20–30, 1967.
- [BVK93] C. S. BLACKBURN, Julie VONWILLER et Robin W. KING : Automatic accent classification using artificial neural networks. *In Eurospeech*, pages 1241–44, Berlin, 1993.
- [Byb01] Joan BYBEE : *Phonology and Language Use*. Cambridge University Press, Cambridge, 2001.
- [Byr99] Gregory BYRON : Young favour Americanisms. *The Guardian*, 25 juin 1999.
- [BZVC98] Kay M. BERKLING, Marc A. ZISSMAN, Julie VONWILLER et Chris CLEIRIGH : Improving accent identification through knowledge of English syllable structure. *In ICSLP*, pages 89–92, Sydney, 1998.

- [Cal89] CALLIOPE : *La Parole et son traitement automatique*. Masson, Paris, 1989.
- [CCis] Ioana CHITORAN et Abigail C. COHN : Complexity in phonetics and phonology : Gradience, categoriality, and naturalness. In François PELLEGRINO, Egidio MARSICO, Ioana CHITORAN et Christophe COUPÉ, éditeurs : *Approaches to Phonological Complexity*. soumis.
- [CH68] Noam CHOMSKY et Morris HALLE : *The Sound Pattern of English*. Harper & Row, New York, 1968.
- [CH06] R. J. G. B. CAMPELLO et E. R. HRUSCHKA : A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157(21):2858–75, 2006.
- [Che91] François CHEVILLET : *Les Variétés de l'anglais*. Nathan, s.l., 1991.
- [Chi99] Deborah CHIRREY : Edinburgh : Descriptive material. In Paul FOULKES et Gerry DOCHERTY, éditeurs : *Urban Voices : Accent Studies in the British Isles*, pages 223–229. Arnold, Londres, 1999.
- [Cla39] André CLASSE : *The Rhythm of English Prose*. Basil Blackwell, Oxford, 1939.
- [CLP06] Cynthia G. CLOPPER, Susannah V. LEVI et David B. PISONI : Perceptual similarity of regional dialects of American English. *Journal of the Acoustical Society of America*, 119(1):566–574, 2006.
- [CM99] Beverley COLLINS et Inger M. MEES : *The Real Professor Higgins. The Life and Career of Daniel Jones*. Mouton de Gruyter, Berlin, 1999.
- [CP04] Cynthia G. CLOPPER et David B. PISONI : Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32(1):111–140, 2004.
- [CR07] Marion COADOU et Abderrazak ROUGAB : Voice quality and variation in English. In *ICPhS*, pages 2077–80, Sarrebruck, 2007.

-
- [Dan90] H. R. DANIELS : Accent recognition, listening positions and the case for the archiphone as a dynamic identity in speech perception. In J.-L. DUCHET, J.-M. FOURNIER, J. HUMBLEY et P. LARREYA, éditeurs : *5^e Colloque d'avril sur l'anglais oral*, pages 27–49, Villetaneuse, 1990. Université Paris 13.
- [Dau83] Rebecca M. DAUER : Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11:51–62, 1983.
- [Del65] Pierre DELATTRE : *Comparing the Phonetic Features of English, French, German and Spanish : An Interim Report*. Julius Groos Verlag, Heidelberg, 1965.
- [Del06] Volker DELLWO : Rhythm and speech rate : A variation coefficient for ΔC . In *38th Linguistics Colloquium*, pages 231–241, Piliscsaba, Hongrie, 2006.
- [Det97] David DETERDING : The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, 27:47–55, 1997.
- [Det01] David DETERDING : The measurement of rhythm : A comparison of Singapore and British English. *Journal of Phonetics*, 29(2):217–230, 2001.
- [DFA07] Volker DELLWO, Adrian FOURCIN et Evelyn ABBERTON : Rhythmical classification of languages based on voice parameters. In *ICPhS*, pages 1129–32, Sarrebruck, 2007.
- [DFP06] Volker DELLWO, Emmanuel FERRAGNE et François PELLEGRINO : The perception of intended speech rate in English, French, and German by French speakers. In *Speech Prosody*, Dresde, 2006.
- [DH00] R. I. DAMPER et S. R. HARNAD : Neural network models of categorical perception. *Perception and Psychophysics*, 62(4):843–867, 2000.
- [DHS01] Richard O. DUDA, Peter E. HART et David G. STORK : *Pattern Classification*. Wiley, New York, 2001.

- [dJMHN07] Gea de JONG, Kirsty McDOUGALL, Toby HUDSON et Francis NOLAN : The speaker discriminating power of sounds undergoing historical change : A formant-based study. *In ICPHS*, pages 1813–16, Sarrebruck, 2007.
- [DL02] Jacques DURAND et Bernard LAKS : Phonology, phonetics, and cognition. *In Jacques DURAND et Bernard LAKS, éditeurs : Phonetics, Phonology, and Cognition*, pages 10–50. Oxford University Press, Oxford, 2002.
- [DLC55] Pierre DELATTRE, Alvin M. LIBERMAN et Franklin S. COOPER : Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27(4):769–773, 1955.
- [DMC02] John A. DIXON, Berenic MAHONEY et Roger COCKS : Accents of guilt : Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology*, 21(2):162–168, 2002.
- [DMS+02] G. DREYFUS, J.-M. MARTINEZ, M. SAMUELIDES, M. B. GORDON, F. BADRAN, S. THIRIA et L. HÉRAULT : *Réseaux de neurones. Méthodologie et applications*. Eyrolles, Paris, 2002.
- [DO69] Pierre DELATTRE et Carroll OLSEN : Syllabic features and phonic impression in English, German, French and Spanish. *Lingua*, 22:160–175, 1969.
- [DO97] S. J. DRINKWATER et N. C. O’LEARY : Unemployment in Wales : Does language matter ? *Regional Studies*, 31(6):583–591, 1997.
- [DRBT04] S. D’ARCY, M. J. RUSSELL, S. R. BROWNING et M. J. TOMLINSON : The Accents of the British Isles (ABI) corpus. *In MIDL*, pages 115–119, Paris, 2004.
- [DSA+04] Volker DELLWO, Ingmar STEINER, Bianca ASCHENBERNER, Jana DANKOVICOVA et Petra S. WAGNER : BonnTempo-Corpus and BonnTempo-Tools : A database for the study of speech rhythm and rate. *In ICSLP*, pages 777–780, Jeju, Corée, 2004.

-
- [DW03] Volker DELLWO et Petra S. WAGNER : Relationships between speech rate and rhythm. *In ICPHS*, pages 471–474, Barcelona, 2003.
- [dWWB⁺04] Febe de WET, Katrin WEBER, Louis BOVES, Bert CRANEN, Samy BENGIO et Boulard HERVÉ : Evaluation of formant-like features on an automatic vowel classification task. *Journal of the Acoustical Society of America*, 116(3):1781–92, 2004.
- [EI04] Bronwen G. EVANS et Paul IVERSON : Vowel normalization for accent : An investigation of best exemplar locations in Northern and Southern British English sentences. *Journal of the Acoustical Society of America*, 115(1):352–361, 2004.
- [EI07] Bronwen G. EVANS et Paul IVERSON : Plasticity in vowel perception and production : A study of accent change in young adults. *Journal of the Acoustical Society of America*, 121(6):3814–26, 2007.
- [Ell94] Stanley ELLIS : The Yorkshire Ripper enquiry : Part I. *Forensic Linguistics*, 1(2):197–206, 1994.
- [ELL01] Brian S. EVERITT, Sabine LANDAU et Morven LEESE : *Cluster Analysis*. Arnold, Londres, 2001.
- [Elm05] Simon ELMES : *Talking for Britain. A Journey through the Nation's Dialects*. Penguin, Londres, 2005.
- [Fan60] Gunnar FANT : *Acoustic Theory of Speech Production*. Mouton, La Haye, 1960.
- [FD99] Paul FOULKES et Gerry DOCHERTY, éditeurs. *Urban Voices : Accent Studies in the British Isles*. Arnold, Londres, 1999.
- [FD00] Paul FOULKES et Gerry DOCHERTY : Another chapter in the history of /r/ : ‘Labiodental’ variants in British English. *Journal of Sociolinguistics*, 4(1):30–59, 2000.

- [Fen01] Barbara A. FENNELL : *A History of English. A Sociolinguistic Approach*. Blackwell, Oxford, 2001.
- [Fer06] Emmanuel FERRAGNE : Quand le professeur Higgins s'invite sur votre ordinateur : une approche moderne de la dialectologie. *Anglophonia*, 20:133–148, 2006.
- [FJ07] Edward FLEMMING et Stephanie JOHNSON : Rosa's roses : Reduced vowels in American English. *Journal of the International Phonetic Association*, 37(1):83–96, 2007.
- [FKN91] Gunnar FANT, Anita KRUCKENBERG et Lennart NORD : Durational correlates of stress in Swedish, French, and English. *Journal of Phonetics*, 19:351–365, 1991.
- [Fla55] James L. FLANAGAN : A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America*, 27(3):613–617, 1955.
- [FLM99] Eric FOSLER-LUSSIER et Nelson MORGAN : Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, 29(2-4):137–158, 1999.
- [Fou91] Marios FOURAKIS : Tempo, stress, and vowel reduction in American English. *Journal of the Acoustical Society of America*, 90(4):1816–27, 1991.
- [FP04a] Emmanuel FERRAGNE et François PELLEGRINO : A comparative account of the suprasegmental and rhythmic features of British English dialects. In *MIDL*, pages 121–126, Paris, France, 2004.
- [FP04b] Emmanuel FERRAGNE et François PELLEGRINO : Diphthongization as a cue for the automatic identification of British English dialects. *Journal of the Acoustical Society of America*, 116(4):2630, 2004.
- [FP04c] Emmanuel FERRAGNE et François PELLEGRINO : Rhythm in read British English : Interdialect variability. In *ICSLP*, pages 1573–76, Jeju, Corée, 2004.

-
- [FP06] Emmanuel FERRAGNE et François PELLEGRINO : Les systèmes vocaliques des dialectes de l'anglais britanniques. *In Journées d'étude sur la parole*, pages 411–414, Dinard, 2006.
- [FP07] Emmanuel FERRAGNE et François PELLEGRINO : Automatic dialect identification : A study of British English. *In Christian MÜLLER, éditeur : Speaker Classification*, volume 2, pages 243–257. Springer, New York, 2007.
- [Fry55] D. B. FRY : Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27(4):765–768, 1955.
- [Fud77] Erik FUDGE : Long and short [æ] in one Southern British speaker's English. *Journal of the International Phonetic Association*, 7(2):55–65, 1977.
- [GA03] Stephen D. GOLDINGER et Tamiko AZUMA : Puzzle-solving science : The quixotic quest for units in speech perception. *Journal of Phonetics*, 31(3-4):305–320, 2003.
- [GAD05] Cédric GENDROT et M. ADDA-DECKER : Impact of duration on F1/F2 formant values of oral vowels : An automatic analysis of large broadcast news corpora in French and German. *In Eurospeech*, pages 2453–56, Lisbonne, 2005.
- [Gas68] George GASCOIGNE : Certain notes of instruction concerning the making of verse or rhyme in English. *In Edward ARBER, éditeur : The Steele Glas*, pages 31–40. Arber, Birmingham, 1868.
- [Gay78] Thomas GAY : Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 63(1):223–230, 1978.
- [GCS07] Louis GOLDSTEIN, Ioana CHITORAN et Elizabeth SELKIRK : Syllable structure as coupled oscillator modes : Evidence from Georgian vs. Tashlhiyt Berber. *In ICPHS*, pages 241–244, Sarrebruck, 2007.
- [Gim80] A. C. GIMSON : *An Introduction to the Pronunciation of English*. Arnold, Londres, 1980.

- [GL86] J. C. GOWER et P. LEGENDRE : Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3(1):5–48, 1986.
- [GL02] Esther GRABE et Ee Ling LOW : Durational variability in speech and the rhythm class hypothesis. In Carlos GUSSENHOVEN et N. WARNER, éditeurs : *Papers in Laboratory Phonology VII*. Cambridge University Press, Cambridge, 2002.
- [Gop90] H. S. GOPAL : Effects of speaking rate on the behavior of tense and lax vowel durations. *Journal of Phonetics*, 18:497–518, 1990.
- [Gor04] Matthew J. GORDON : The West and Midwest : Phonology. In Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 338–350. Mouton de Gruyter, Berlin, 2004.
- [GP02] Esther GRABE et Brechtje POST : Intonational variation in the British Isles. In *Speech Prosody*, pages 343–346, Aix-en-Provence, 2002.
- [GPL96] Stephen D. GOLDINGER, David B. PISONI et Paul A. LUCE : Speech perception and spoken word recognition. In Norman J. LASS, éditeur : *Principles of Experimental Phonetics*, pages 277–327. Mosby, St. Louis, 1996.
- [GPN01] Esther GRABE, Brechtje POST et F. J. NOLAN : Modelling intonational variation in English. The IViE system. In *Prosody 2000*, pages 51–57, Poznan, Pologne, 2001.
- [GPNF00] Esther GRABE, Brechtje POST, F. J. NOLAN et Kimberley FARRAR : Pitch accent realization in four varieties of British English. *Journal of Phonetics*, 28(2):161–185, 2000.
- [Gre96] Steven GREENBERG : Auditory processing of speech. In Norman J. LASS, éditeur : *Principles of Experimental Phonetics*, pages 362–407. Mosby, St Louis, 1996.

-
- [GS95] William A. GALE et Geoffrey SAMPSON : Good-Turing frequency estimation without tears. *Journal of Quantitative Linguistics*, 2(1):217–237, 1995.
- [Hag95] Robert HAGIWARA : *Acoustic Realizations of American /r/ as Produced by Women and Men*. Thèse de doctorat, UCLA, 1995.
- [Har06] Jonathan HARRINGTON : An acoustic analysis of ‘happy-tensing’ in the Queen’s Christmas broadcasts. *Journal of Phonetics*, 34(4):439–457, 2006.
- [Hay00] Katrina HAYWARD : *Experimental Phonetics*. Pearson, Harlow, 2000.
- [HCF⁺06] Jean-Paul HATON, Christophe CERISARA, Dominique FOHR, Yves LAPRIE et Kamel SMAÏLI : *Reconnaissance automatique de la parole. Du signal à son interprétation*. Dunod, Paris, 2006.
- [Hen83] C. G. HENTON : Changes in the vowels of Received Pronunciation. *Journal of Phonetics*, 11:353–371, 1983.
- [Hen00] Philip HENSHER : Don’t be fooled : The Queen is not speaking our language. *The Independent*, 22 décembre 2000.
- [HG03] Wilbert HEERINGA et Charlotte GOOSKENS : Norwegian dialects examined perceptually and acoustically. *Computers and the Humanities*, 37:293–315, 2003.
- [HGCW95] James HILLENBRAND, Laura A. GETTY, Michael J. CLARK et Kimberlee WHEELER : Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5):3099–3111, 1995.
- [Hic99] Raymond HICKEY : Dublin English : Current changes and their motivation. In Paul FOULKES et Gerry DOCHERTY, éditeurs : *Urban Voices : Accent Studies in the British Isles*, pages 265–281. Arnold, Londres, 1999.
- [Hic04a] Raymond HICKEY : *A Sound Atlas of Irish English*. Mouton de Gruyter, Berlin, 2004.
- [Hic04b] Raymond HICKEY : Irish English : Phonology. In Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON,

éditeurs : *A Handbook of Varieties of English*, volume 1, pages 68–97. Mouton de Gruyter, Berlin, 2004.

- [HKR07] Jonathan HARRINGTON, Felicitas KLEBER et Ulrich REUBOLD : /u/-fronting in RP : A link between sound change and diminished perceptual compensation for coarticulation? *In ICP_hS*, pages 1473–76, Sarrebruck, 2007.
- [HM05] Sarah HAWKINS et Jonathan MIDGLEY : Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, 35(2):183–199, 2005.
- [HN01] Wilbert HEERINGA et John NERBONNE : Dialect areas and dialect continua. *Language Variation and Change*, 13(3):375–400, 2001.
- [Hon97] John HONEY : Sociophonology. *In* Florian COULMAS, éditeur : *The Handbook of Sociolinguistics*, pages 92–106. Blackwell, Malden, 1997.
- [Hon07] Patrick HONEYBONE : New-dialect formation in nineteenth century Liverpool : A brief history of Scouse. *In* A. GRANT, C. GREY et K. WATSON, éditeurs : *The Mersey Sound : Liverpool's Language, People and Places*, pages 106–140. Open House Press, Liverpool, 2007.
- [HPW00] Jonathan HARRINGTON, Sallyanne PALETHORPE et Catherine I. WATSON : Does the Queen speak the Queen's English? *Nature*, 408(6815):927–928, 2000.
- [HR04] Benjamin HALBERSTAM et Lawrence J. RAPHAEL : Vowel normalization : The role of fundamental frequency and upper formants. *Journal of Phonetics*, 32:423–434, 2004.
- [HTW05] Arthur HUGHES, Peter TRUDGILL et Dominic WATT : *English Accents and Dialects. An Introduction to Social and Regional Varieties of English in the British Isles*. Hodder Arnold, Londres, 2005.

-
- [Huc04] Mark HUCKVALE : ACCDIST : A metric for comparing speakers' accents. *In ICSLP*, pages 29–32, Jeju, Corée, 2004.
- [Huc07a] Mark HUCKVALE : ACCDIST : An accent similarity metric for accent recognition and diagnosis. *In* Christian MÜLLER, éditeur : *Speaker Classification*, volume 2, pages 258–275. Springer, Berlin, 2007.
- [Huc07b] Mark HUCKVALE : Hierarchical clustering of speakers into accents with the ACCDIST metric. *In ICPHS*, pages 1821–24, Sarrebruck, 2007.
- [HYHI04] John H. L. HANSEN, Umit YAPANEL, Rongqing HUANG et Ayako IKENO : Dialect analysis and modeling for automatic classification. *In ICSLP*, pages 1569–72, Jeju, Corée, 2004.
- [IK95] Paul IVERSON et Patricia K. KUHL : Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97(1):553–562, 1995.
- [Jak63] Roman JAKOBSON : *Essais de linguistique générale*. Éditions de Minuit, Paris, 1963.
- [Jan04] Esther JANSE : Word perception in fast speech : Artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, 42(2):155–173, 2004.
- [JBFL⁺98] Daniel JURAFSKY, Alan BELL, Eric FOSLER-LUSSIER, Cynthia GIRAND et William RAYMOND : Reduction of English function words in Switchboard. *In ICSLP*, pages 3111–14, Sydney, 1998.
- [Jel97] Frederick JELINEK : *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, [Mass.], 1997.
- [JFW93] Keith JOHNSON, Edward FLEMMING et Richard WRIGHT : The hyperspace effect : Phonetic targets are hyperarticulated. *Language*, 69(3):505–528, 1993.

- [JG99] J. S. R. JANG et Ned GULLEY : *Fuzzy Logic Toolbox User's Guide*. The MathWorks, 1999.
- [JM00] Daniel JURAFSKY et James H. MARTIN : *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, Upper Saddle River, 2000.
- [JNQ03] Esther JANSE, Sieb NOOTEBOOM et Hugo QUENÉ : Word-level intelligibility of time-compressed speech : Prosodic and segmental factors. *Speech Communication*, 41:287–301, 2003.
- [Joh00] Keith JOHNSON : Adaptive dispersion in vowel perception. *Phonetica*, 57(2-4):181–188, 2000.
- [Joh05] Keith JOHNSON : Speaker normalization in speech perception. In David B. PISONI et Robert E. REMEZ, éditeurs : *The Handbook of Speech Perception*, pages 363–389. Blackwell, Malden, 2005.
- [Jol95] Geneviève JOLY : *Précis de phonétique historique du français*. Armand Colin, Paris, 1995.
- [Jon63] Daniel JONES : *The Pronunciation of English*. Cambridge University Press, Cambridge, 1963.
- [Jon66] Daniel JONES : *The Pronunciation of English*. Cambridge University Press, Cambridge, 1966.
- [Jon67] Daniel JONES : *An Outline of English Phonetics*. Heffer, Cambridge, 1967.
- [Jon03] Daniel JONES : *English Pronouncing Dictionary*. Cambridge University Press, Cambridge, 2003.
- [Joo48] Martin JOOS : *Acoustic Phonetics*. Linguistic Society of America, Baltimore, 1948.
- [JST99] James J. JENKINS, Winifred STRANGE et Sonja A. TRENT : Context-independent dynamic information for the perception of coarticulated vowels. *Journal of the Acoustical Society of America*, 106(1):438–448, 1999.

-
- [Jur03] Daniel JURAFSKY : Probabilistic modeling in psycholinguistics : Linguistic comprehension and production. In Rens BOD, Jennifer HAY et Stefanie JANNEDY, éditeurs : *Probabilistic Linguistics*, pages 39–95. MIT Press, Cambridge, [Mass.], 2003.
- [Jus96] Peter. W. JUSCZYK : Developmental speech perception. In Norman J. LASS, éditeur : *Principles of Experimental Phonetics*, pages 328–361. Mosby, St Louis, 1996.
- [KC75] D. H. KLATT et W. E. COOPER : Perception of segment duration in sentence context. In A. COHEN et S. G. NOOTEBOOM, éditeurs : *Structure and Process in Speech Perception*, pages 69–89. Springer-Verlag, New York, 1975.
- [KDL46] W. KOENIG, H. K. DUNN et L. Y. LACY : The sound spectrograph. *Journal of the Acoustical Society of America*, 18(1):19–49, 1946.
- [KDL96] Ray D. KENT, James DEMBOWSKI et Norman J. LASS : The acoustic characteristics of American English. In Norman J. LASS, éditeur : *Principles of Experimental Phonetics*, pages 185–225. Mosby, St Louis, 1996.
- [Ker01] Paul KERSWILL : Mobility, meritocracy and dialect levelling : The fading (and phasing) out of Received Pronunciation. In P. RAJAMÄE et K. VOGELBERG, éditeurs : *British Studies in the New Millenium : the Challenge of Grassroots*, pages 45–58. Université de Tartu, Tartu, 2001.
- [Ker03] Paul KERSWILL : Dialect levelling and geographical diffusion in British English. In D. BRITAIN et J. CHESHIRE, éditeurs : *Social Dialectology. In Honour of Peter Trudgill*, pages 223–243. Benjamins, Amsterdam, 2003.
- [Ker06] Paul KERSWILL : RP, Standard English and the standard/non-standard relationship. In D. BRITAIN, éditeur : *Language in the British Isles*, pages 34–51. Cambridge University Press, Cambridge, 2006.
- [KGCR05] Greg KOCHANSKI, Esther GRABE, John COLEMAN et B. ROSNER : Loud-

- ness predicts prominence : Fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118:1038–54, 2005.
- [Kla76] D. H. KLATT : Linguistic uses of segmental duration in English : Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59(5): 1208–21, 1976.
- [Kno78] Gerry KNOWLES : The nature of phonological variables in Scouse. In Peter TRUDGILL, éditeur : *Sociolinguistic Patterns in British English*, pages 80–90. Arnold, Londres, 1978.
- [Koh91] K. J. KOHLER : Isochrony, units of rhythmic organization and speech rate. In *ICPhS*, pages 257–261, Aix-en-Provence, 1991.
- [Kor06] Jacques KOREMAN : Perceived speech rate : The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America*, 119(1):582–596, 2006.
- [KPA89] Diane KEWLEY-PORT et Bishnu S. ATAL : Perceptual differences between vowels located in a limited phonetic space. *Journal of the Acoustical Society of America*, 85(4):1726–40, 1989.
- [KPW94] Diane KEWLEY-PORT et Charles S. WATSON : Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, 95(1):485–496, 1994.
- [Kuh91] Patricia K. KUHLE : Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50(2):93–107, 1991.
- [LA95] Leigh LISKER et Arthur S. ABRAMSON : A long-short vowel dichotomy in fluent English ? In *ICPhS*, pages 226–229, Stockholm, 1995.
- [Lab94] William LABOV : *Principles of Linguistic Change : Internal Factors*, volume 1. Blackwell, Cambridge, [Mass.], 1994.

-
- [Lab06] William LABOV : A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, 34(4):500–515, 2006.
- [Lad67] Peter LADEFOGED : *Three Areas of Experimental Phonetics*. Oxford University Press, Londres, 1967.
- [Lag98] Jean de LAGARDE : *Initiation à l'analyse de données*. Dunod, Paris, 1998.
- [Lak87] George LAKOFF : *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. University of Chicago Press, Chicago, 1987.
- [Lav94] John LAVER : *Principles of Phonetics*. Cambridge University Press, Cambridge, 1994.
- [LCSSK67] A. M. LIBERMAN, F. S. COOPER, D. P. SHANKWEILER et M. STUDDERT-KENNEDY : Perception of the speech code. *Psychological Review*, 74(6):431–461, 1967.
- [LG95] Ee Ling LOW et Esther GRABE : Prosodic patterns in Singapore English. *In ICPhS*, pages 636–639, Stockholm, 1995.
- [LGN00] Ee Ling LOW, Esther GRABE et Francis J. NOLAN : Quantitative characterizations of speech rhythm : Syllable-timing in Singapore English. *Language & Speech*, 43(4):377–401, 2000.
- [Lin83] Björn LINDBLOM : Economy of speech gestures. *In* P. MACNEILAGE, éditeur : *The Production of Speech*, pages 217–245. Springer, New York, 1983.
- [Lin90] Björn LINDBLOM : Explaining phonetic variation : A sketch of the H&H theory. *In* W. J. HARDCASTLE et Alain MARCHAL, éditeurs : *Speech Production and Speech Modelling*, pages 403–439. Dordrecht, Kluwer, 1990.
- [Lob71] B. M. LOBANOV : Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49(2):606–608, 1971.
- [LP97] R. B. LE PAGE : The evolution of a sociolinguistic theory of language. *In*

- Florian COULMAS, éditeur : *The Handbook of Sociolinguistics*, pages 15–32. Blackwell, Malden, 1997.
- [LSK67] Björn LINDBLOM et STUDDERT-KENNEDY : On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, 42(4):830–843, 1967.
- [Mal71] Bertil MALMBERG : *Les Domaines de la phonétique*. Presses Universitaires de France, Paris, 1971.
- [Mar05] André MARTINET : *Économie des changements phonétiques*. Maisonneuve et Larose, Paris, 2005.
- [Mar06] Sharon MARSDEN : A sociophonetic study of labiodental /r/ in Leeds. *Leeds Working Papers in Linguistics & Phonetics*, 11:153–170, 2006.
- [Mas87] Dominic William MASSARO : *Speech Perception by Ear and Eye : A Paradigm for Psychological Inquiry*. Lawrence Erlbaum, Hillsdale, 1987.
- [Mat99] Ignatius G. MATTINGLY : A short history of acoustic phonetics in the U.S. In John J. OHALA, Arthur J. BRONSTEIN, M. GRAZIA BUSÀ, Julie A. LEWIS et William F. WEIGEL, éditeurs : *A Guide to the History of the Phonetic Sciences in the United States*, pages 1–6. University of California, Berkeley, 1999.
- [McC77] J. Derrick MCCLURE : Vowel duration in a Scottish accent. *Journal of the International Phonetic Association*, 7(1):10–16, 1977.
- [McD06] Kirsty MCDUGALL : Dynamic features of speech and the characterisation of speakers : Towards a new approach using formant frequencies. *International Journal of Speech, Language, and the Law*, 13(1):89–126, 2006.
- [Mil84] M. MILLER : On the perception of rhythm. *Journal of Phonetics*, 12:75–83, 1984.
- [Mil89] James D. MILLER : Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85(5):2114–34, 1989.

-
- [ML01] V. MAKARENKOV et P. LEGENDRE : Optimal variable weighting for ultrametric and additive trees and k-means partitioning : Methods and software. *Journal of Classification*, 18:245–271, 2001.
- [MM78] James MILROY et Lesley MILROY : Belfast : Change and variation in an urban vernacular. In Peter TRUDGILL, éditeur : *Sociolinguistic Patterns in British English*, pages 19–36. Arnold, Londres, 1978.
- [MN07] Kirsty MCDUGALL et F. J. NOLAN : Discrimination of speakers using the formant dynamics of /u:/ in British English. In *ICPhS*, pages 1825–28, Sarrebruck, 2007.
- [MWW94] William MARSLÉN-WILSON et Paul WARREN : Levels of perceptual representation and process in lexical access : Words, phonemes, and features. *Psychological Review*, 101(4):653–675, 1994.
- [Nab02] Ian T. NABNEY : *Netlab. Algorithms for Pattern Recognition*. Springer, Londres, 2002.
- [Neg02] Michael NEGNEVITSKY : *Artificial Intelligence. A Guide to Intelligent Systems*. Pearson, Harlow, 2002.
- [Nol98] F. J. NOLAN : The shifting sands of English pronunciation. In P. RAJAMÄE, éditeur : *2nd International Tartu Conference on British Studies*, pages 76–93, Tartu, 1998. Université de Tartu.
- [O’C73] J. D. O’CONNOR : *Phonetics*. Penguin, Londres, 1973.
- [OED07] *The Oxford English Dictionary (en ligne)*. Oxford University Press, Oxford, 2007.
- [ON99] M. O’DELL et T. NIEMINEN : Coupled oscillator model of speech rhythm. In *ICPhS*, pages 1075–78, San Francisco, 1999.
- [OS06] Lena OLAUSSON et Catherine SANGSTER : *Oxford BBC Guide to Pronunciation*. Oxford University Press, Oxford, 2006.

- [OSW78] Harold ORTON, Stewart SANDERSON et John WIDDOWSON, éditeurs. *The Linguistic Atlas of England*. Croom Helm, Londres, 1978.
- [OW03] Cristina OLARU et Louis WEHENKEL : A complete fuzzy decision tree technique. *Fuzzy Sets and Systems*, 138(2):221–254, 2003.
- [PAO00] François PELLEGRINO et Régine ANDRÉ-OBRECHT : Automatic language identification : An alternative approach to phonetic modelling. *Signal Processing*, 80:1231–1244, 2000.
- [Par98] Gudrun PARSONS : *From “RP” to “Estuary English” : The Concept ‘Received’ and the Debate about British Pronunciation Standards*. M.A., Hambourg, 1998.
- [PB52] Gordon E. PETERSON et Harold L. BARNEY : Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2):175–184, 1952.
- [PB99] Antonio PAMIES BERTRÁN : Prosodic typology : On the dichotomy between stress-timed and syllable-timed languages. *Language Design*, 2:103–130, 1999.
- [Pen04] Robert PENHALLURICK : Welsh English : Phonology. In Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 98–112. Mouton de Gruyter, Berlin, 2004.
- [PFR00] T. PFAU, R. FALTLHAUSER et G. RUSKE : A combination of speaker normalization and speech rate normalization for automatic speech recognition. In *ICSLP*, pages 362–365, Pékin, 2000.
- [PFR04] François PELLEGRINO, Jérôme FARINAS et Jean-Luc ROUAS : Automatic estimation of speaking rate in multilingual spontaneous speech. In *Speech Prosody*, Nara, Japon, 2004.

-
- [Pic99] J. M. PICKETT : *The Acoustics of Speech Communication. Fundamentals, Speech Perception Theory, and Technology*. Allyn & Bacon, Boston, 1999.
- [Pik45] Kenneth L. PIKE : *The Intonation of American English*. University of Michigan Press, Ann Arbor, 1945.
- [PIR06] Aniruddh D. PATEL, John R. IVERSEN et Jason C. ROSENBERG : Comparing the rhythm and melody of speech and music : The case of British English and French. *Journal of the Acoustical Society of America*, 119(5):3034–47, 2006.
- [PL60] Gordon E. PETERSON et Ilse LEHISTE : Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32(6):693–703, 1960.
- [PLN99] Aniruddh D. PATEL, Anders LÖFQVIST et Walter NAITO : The acoustics and kinematics of regularly timed speech : A database and method for the study of the P-center problem. *In ICP*h*S*, pages 405–408, San Francisco, 1999.
- [Pop02] Karl POPPER : *The Logic of Scientific Discovery*. Routledge, Londres, 2002.
- [PP48] Ralph K. POTTER et Gordon E. PETERSON : The representation of vowels and their movements. *Journal of the Acoustical Society of America*, 20(4): 528–535, 1948.
- [PR98] T. PFAU et G. RUSKE : Estimating the speaking rate by vowel detection. *In ICASSP*, volume 2, pages 945–948, Seattle, 1998.
- [RDM03] Franck RAMUS, Emmanuel DUPOUX et Jacques MEHLER : The psychological reality of rhythm classes : Perceptual studies. *In ICP*h*S*, pages 337–342, Barcelone, 2003.
- [RFP⁺95] Tony ROBINSON, Jeroen FRANSEN, David PYE, Jonathan FOOTE et Steve RENALS : WSJCAM0 : A British English speech corpus for large vocabulary continuous speech recognition. *In ICASSP*, pages 81–84, Détroit, 1995.

- [RFP03] Jean-Luc ROUAS, Jérôme FARINAS et François PELLEGRINO : Automatic modelling of rhythm and intonation for language identification. *In ICPPhS*, pages 567–570, Barcelone, 2003.
- [RFP04] Jean-Luc ROUAS, Jérôme FARINAS et François PELLEGRINO : Évaluation automatique du débit de la parole sur des données multilingues spontanées. *In Journées d'études sur la parole*, Fès, Maroc, 2004.
- [RG06] Priya RAGHUBIR et Eric GREENLEAF : Ratios in proportion : What should be the shape of the package ? *Journal of Marketing*, 70(2):95–107, 2006.
- [RM99] Franck RAMUS et Jacques MEHLER : Language identification with suprasegmental cues : A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105(1):512–521, 1999.
- [RNM99] Franck RAMUS, Marina NESPOR et Jacques MEHLER : Correlates of linguistic rhythm in the speech signal. *Cognition*, 73:265–292, 1999.
- [Roa82] Peter ROACH : On the distinction between 'stress-timed' and 'syllable-timed' languages. *In David CRYSTAL, éditeur : Linguistic Controversies*, pages 73–79. Arnold, Londres, 1982.
- [Ros72] Mario ROSSI : Le seuil différentiel de durée. *In A. VALDMAN, éditeur : Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, pages 435–450. Mouton, La Haye, 1972.
- [Rou01] Jean-Pierre ROUSSELOT : *Principes de phonétique expérimentale*. Welter, Paris, 1897-1901.
- [RP94] B. S. ROSNER et J. B. PICKERING : *Vowel Perception and Production*. Oxford University Press, Oxford, 1994.
- [Sap90] Gilbert SAPORTA : *Probabilités, analyse de données et statistique*. Technip, Paris, 1990.
- [Sau95] Ferdinand de SAUSSURE : *Cours de linguistique générale*. Payot & Rivages, Paris, 1995.

-
- [SBDM97] Harvey M. SUSSMAN, Nicola BESSELL, Eileen DALSTON et Tivoli MAJORS : An investigation of stop place of articulation as a function of syllable position : A locus equation perspective. *Journal of the Acoustical Society of America*, 101(5):2826–38, 1997.
- [SBK⁺04] Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs. *A Handbook of Varieties of English*, volume 1. Mouton de Gruyter, Berlin, 2004.
- [Sch01] Felix SCHAEFFLER : Measuring rhythmic deviation in second language speech. In *Eurospeech*, pages 819–822, Aalborg, 2001.
- [Sco07] James M. SCOBIE : Biological and social grounding of phonology : Variation as a research tool. In *ICPhS*, pages 225–228, Sarrebruck, 2007.
- [Scr35] E. W. SCRIPTURE : Film tracks of English vowels. *Journal of the Acoustical Society of America*, 6(3):169–172, 1935.
- [Sen95] W. F. SENDLMEIER : Feature, phoneme, syllable or word : How is speech mentally represented ? *Phonetica*, 52(3):131–143, 1995.
- [SGHF00] James M. SCOBIE, Fiona GIBBON, W. J. HARDCASTLE et Paul J. FLETCHER : Covert constrat as a stage in the acquisition of phonetics and phonology. In Michael B. BROE et Janet B. PIERREHUMBERT, éditeurs : *Papers in Laboratory Phonology V*, pages 194–207. Cambridge University Press, Cambridge, 2000.
- [SH03] Katherine SHOBBROOK et Jill HOUSE : High rising tones in Southern British English. In *ICPhS*, pages 1273–76, Barcelone, 2003.
- [SHT99] James M. SCOBIE, Nigel HEWLETT et Alice TURK : Standard English in Edinburgh and Glasgow : The Scottish Vowel Length Rule revealed. In Paul FOULKES et Gerry DOCHERTY, éditeurs : *Urban Voices : Accent Studies in the British Isles*, pages 230–245. Arnold, Londres, 1999.
- [SS99] Jane STUART-SMITH : Glasgow : Accent and voice quality. In Paul

- FOULKES et Gerry DOCHERTY, éditeurs : *Urban Voices : Accent Studies in the British Isles*, pages 203–222. Arnold, Londres, 1999.
- [SS04] Jane STUART-SMITH : Scottish English : Phonology. In Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 47–67. Mouton de Gruyter, Berlin, 2004.
- [SS07] Jane STUART-SMITH : A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. In *ICPhS*, pages 1449–52, Sarrebruck, 2007.
- [SSS06] James M. SCOBIE et Jane STUART-SMITH : Quasi-phonemic contrast and the fuzzy inventory : Examples from Scottish English. *QMUC Speech Science Research Centre Working Papers*, 8, 2006.
- [Sté97] Colette STÉVANOVITCH : *Manuel d'histoire de la langue anglaise des origines à nos jours*. Ellipses, Paris, 1997.
- [Ste34] John C. STEINBERG : Application of sound measuring instruments to the study of phonetic problems. *Journal of the Acoustical Society of America*, 6(1):16–24, 1934.
- [Ste69] Joshua STEELE : *An Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols*. Scolar Press, Menston, 1969.
- [Ste89] Kenneth N. STEVENS : On the quantal nature of speech. *Journal of Phonetics*, 17(1):3–45, 1989.
- [Ste98] Kenneth N. STEVENS : *Acoustic Phonetics*. MIT Press, Cambridge, [Mass.], 1998.
- [STH99] James M. SCOBIE, Alice E. TURK et Nigel HEWLETT : Morphemes, phonetics and lexical items : The case of the Scottish Vowel Length Rule. In *ICPhS*, pages 1617–1620, San Francisco, 1999.

-
- [SW63] Claude E. SHANNON et Warren WEAVER : *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1963.
- [Tay95] John R. TAYLOR : *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford University Press, Oxford, 1995.
- [tB00] Louis ten BOSCH : ASR, dialects, and acoustic/phonological distance. *In ICSLP*, pages 1009–12, Pékin, 2000.
- [Tra90] Hartmut TRAUNMÜLLER : Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88(1):97–100, 1990.
- [Tru83] Peter TRUDGILL : *On Dialect. Social and Geographical Perspectives*. Basil Blackwell, Oxford, 1983.
- [Tru90] Peter TRUDGILL : *The Dialects of England*. Basil Blackwell, Cambridge, [Mass.], 1990.
- [Tru04] Peter TRUDGILL : The dialect of East Anglia : Phonology. *In* Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 142–153. Mouton de Gruyter, Berlin, 2004.
- [Upt04] Clive UPTON : Received Pronunciation. *In* Edgar W. SCHNEIDER, Kate BURRIDGE, Bernd KORTMANN, Rajend MESTHRIE et Clive UPTON, éditeurs : *A Handbook of Varieties of English*, volume 1, pages 217–230. Mouton de Gruyter, Berlin, 2004.
- [vB93] Dick R. van BERGEM : Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, 12(1):1–23, 1993.
- [vB02] Gerald van BELLE : *Statistical Rules of Thumb*. Wiley, New York, 2002.
- [VM96] Jan P. VERHASSELT et Jean-Pierre MARTENS : A fast and reliable rate of speech detector. *In ICSLP*, pages 2258–61, Philadelphie, 1996.

- [VR86] R. R. VERBRUGGE et B. RAKERD : Evidence of talker-independent information for vowels. *Language and Speech*, 29(1):39–57, 1986.
- [WA03] Dominic WATT et William ALLEN : Illustrations of the IPA : Tyneside English. *Journal of the International Phonetic Association*, 33(2):267–271, 2003.
- [Wal00] Katie WALES : North and South : A linguistic divide? *English Today*, 16(1):4–15, 2000.
- [Wat00] Dominic WATT : Phonetic parallels between the close-mid vowels of Tyneside English : Are they internally or externally motivated? *Language Variation and Change*, 12(1):69–101, 2000.
- [Wat06] Kevin WATSON : Phonological resistance and innovation in the North-West of England. *English Today*, 22(2):55–61, 2006.
- [WD04] Petra S. WAGNER et Volker DELLWO : Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. *In Speech Prosody*, pages 227–230, Nara, Japon, 2004.
- [Wel62] John Christopher WELLS : *A Study of the Formants of the Pure Vowels of British English*. M.A., University of London, 1962.
- [Wel82] John Christopher WELLS : *Accents of English*. Cambridge University Press, Cambridge, 1982.
- [Wel90] John Christopher WELLS : *Longman Pronunciation Dictionary*. Longman, Harlow, 1990.
- [Wel91] John Christopher WELLS : The Cockneyfication of RP? *In* Gunnel MELCHERS et Nils-Lennart JOHANNESSON, éditeurs : *Nonstandard varieties of language*, pages 11–13, Stockholm, 1991. Almqvist & Wiksell.
- [Wel99] John Christopher WELLS : Pronunciation preferences in British English : A new survey. *In ICPHS*, pages 1245–48, San Francisco, 1999.

-
- [WFH86] Anthony WOODS, Paul J. FLETCHER et Arthur HUGHES : *Statistics in Language Studies*. Cambridge University Press, Cambridge, 1986.
- [WFS01] Britta WREDE, Gernot A. FINK et Gerhard SAGERER : An investigation of modelling aspects for rate-dependent speech recognition. *In European Conference on Speech Communication and Technology*, pages 2527–30, Aalborg, 2001.
- [WH94] Briony WILLIAMS et Steven M. HILLER : The question of randomness in English foot timing : A control experiment. *Journal of Phonetics*, 22:423–439, 1994.
- [WI00] Dominic WATT et Catherine INGHAM : Durational evidence of the Scottish Vowel Length Rule in Berwick English. *Leeds Working Papers in Linguistics & Phonetics*, 8:205–228., 2000.
- [Wii91] Kalevi WIIK : On a third type of speech rhythm : Foot timing. *In ICPHS*, pages 298–301, Aix-en-Provence, 1991.
- [WK99] Ann WILLIAMS et Paul KERSWILL : Dialect levelling : Change and continuity in Milton Keynes, Reading and Hull. *In Paul FOULKES et Gerry DOCHERTY, éditeurs : Urban Voices : Accent Studies in the British Isles*, pages 141–162. Arnold, Londres, 1999.
- [WM99] Dominic WATT et Lesley MILROY : Patterns of variation and change in three Newcastle vowels : Is this dialect levelling? *In Paul FOULKES et Gerry DOCHERTY, éditeurs : Urban Voices : Accent Studies in the British Isles*, pages 25–46. Arnold, Londres, 1999.
- [WM07] Laurence WHITE et Sven L. MATTYS : Calibrating rhythm : First language and second language studies. *Journal of Phonetics*, 35(4):501–522, 2007.
- [WMP⁺04] D. H. WHALEN, Harriet S. MAGEN, Marianne POUPLIER, A. MIN KANG et Khalil ISKAROUS : Vowel production and perception : Hyperarticulation without a hyperspace effect. *Language & Speech*, 47(2):155–174, 2004.

- [WMSG07] Laurence WHITE, Sven L. MATTYS, Lucy SERIES et Suzi GAGE : Rhythm metrics predict rhythmic discrimination. *In ICPhS*, pages 1009–12, Sarrebruck, 2007.
- [Wri03] Richard WRIGHT : Factors of lexical competition in vowel articulation. *In* John LOCAL, Richard OGDEN et Rosalind TEMPLE, éditeurs : *Papers in Laboratory Phonology VI*, pages 75–87. Cambridge University Press, Cambridge, 2003.
- [WS98] Adalbert WILHELM et Meike SANDER : Interactive statistical analysis of dialect features. *The Statistician*, 47(3):445–455, 1998.
- [WS05] Dagen WANG et Narayanan SHRIKANTH : Speech rate estimation via temporal correlation and selected sub-band correlation. *In ICCASP*, pages 413–416, Philadelphie, 2005.
- [WW91] Thomas H. WONNACOTT et Ronald J. WONNACOTT : *Statistique*. Economica, Paris, 1991.
- [YV02] Qin YAN et Saeed VASEGHI : A comparative analysis of UK and US English accents in recognition and synthesis. *In ICASSP*, pages 413–417, Orlando, Floride, 2002.
- [Zad65] Lofti A. ZADEH : Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.
- [Zad95] Lofti A. ZADEH : Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, 37(3):271–276, 1995.
- [Zel98] Brigitte ZELLNER : Fast and slow speech rate : A characterisation for French. *In ICSLP*, pages 3159–63, Sydney, 1998.