

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université

Le 07 juillet 2022 par

Marie REBOURG

Evaluation de l'intelligibilité après un cancer ORL : Approche perceptive par décodage acoustico-phonétique et mesures acoustiques

Discipline

Sciences du Langage

École doctorale

ED 356 Cognition, Langage et Education

Laboratoire/Partenaires de recherche

Laboratoire Parole et Langage - LPL

La Ligue contre le Cancer



Composition du jury

Rudolph SOCK Rapporteur

LILPA - Université de Strasbourg

Lorraine BAQUE Rapporteuse

Universitat Autònoma de Barcelona

Lise CREVIER-BUCHMAN Examinatrice

LPP – Université Paris 3

Didier DEMOLIN Président du jury

LPP – Université Paris 3

Nicolas FAKHRY Directeur de thèse

APHM CHU la Conception – LPL – Aix Marseille Université

Muriel LALAIN Co-directrice de thèse

LPL – Aix Marseille Université

Affidavit

Je soussignée, Marie Rebourg, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Nicolas Fakhry et Muriel Lalain, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Aix-en-Provence, le 22 avril 2022



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Affidavit

I, undersigned, Marie Rebourg, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Nicolas Fakhry et Muriel Lalain, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place Aix-en-Provence, date 22 avril 2022



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Liste des publications

1. **Rebourg, M.**, Lalain, M., Ghio, A., Fredouille, C., Monestier, L., Fakhry, N., & Woisard, V. (2022). Evaluation de l'intelligibilité des segments vocaliques de patients traités pour un cancer des VADS et de locuteur contrôles. 32^{ème} édition des *Journées d'Etudes sur la Parole*, Noirmoutier, France
2. Lalain, M., Monestier, L., Ghio, A., Fredouille, C., **Rebourg, M.**, Fakhry, N., & Woisard, V. (2022). Prédiction du degré d'altération de l'intelligibilité chez des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. 32^{ème} édition des *Journées d'Etudes sur la Parole*, Noirmoutier, France
3. **Rebourg, M.**, Lalain, M., Ghio, A., Monestier, L., Fredouille, C., Woisard, V., & Fakhry, N. (2021). Evaluer la réalisation acoustique des voyelles dans le cadre de la prise en charge de patients ORL. *Journées de Phonétique Clinique*, 1.
4. Ghio, A., Lalain, M., Rebourg, M., Marczyk, A., Fredouille, C., & Woisard, V. (2021). Validation of an Intelligibility Test Based on Acoustic-Phonetic Decoding of Pseudo-Words : Overall Results from Patients with Cancer of the Oral Cavity and the Oropharynx. *Folia Phoniatica et Logopaedica*. <https://doi.org/10.1159/000519427>
5. Marczyk, A., Ghio, A., Lalain, M., **Rebourg, M.**, Fredouille, C., & Woisard, V. (2021). Optimizing linguistic materials for feature-based intelligibility assessment in speech impairments. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01610-9>
6. **Rebourg M.**, Lalain, M., Ghio, A., Monestier, L., Fakhry, N., Fredouille, C., & Woisard, V (2021). Poster : Evaluer la réalisation acoustique des voyelles dans le cadre de la prise en charge de patients ORL. *Journée de Linguistique Expérimentale JLex*, Aix-en-Provence, Mars 2021
7. **Rebourg, M.**, Lalain, M., Ghio, A., Fredouille, C., Fakhry, N., & Woisard, V. (2020). Évaluer l'intelligibilité, mots ou pseudo-mots? Comparaison entre deux groupes d'auditeurs. In C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla, & S. Schneider (Éds.), *6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)* (p. 543-551). ATALA. <https://hal.archives-ouvertes.fr/hal-02798584>
8. Ghio, A., Lalain, M., **Rebourg, M.**, Fredouille, C., & Woisard, V. (2020). *Evaluation de l'intelligibilité de patients avec traitement du cancer des cavités orales et pharyngales*. In C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla, & S. Schneider (Éds.), *6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)* (p. 271-279). <https://hal.archives-ouvertes.fr/hal-02798549>
9. Marczyk, A. K., Ghio, A., Lalain, M., **Rebourg, M.**, Fredouille, C., & Woisard, V. (2020). Have a Cake and Eat it Too : Assessing Discrimination Performance of an Intelligibility Index Obtained from a Reduced Sample Size. *12th Conference*

- on *Language Resources and Evaluation (LREC 2020)*, 1784–1788. <https://hal.archives-ouvertes.fr/hal-02611678>
10. Ghio, A., Lalain, M., Giusti, L., Robert, D., **Rebourg, M.**, André, C., Fredouille, C., Balaguer, M., & Woisard, V. (2019). Application d'un test d'intelligibilité à partir de pseudo- mots dans le cas de patients post traitement de cancers des VADS. *Journées de Phonétique Clinique*. <https://hal.archives-ouvertes.fr/hal-02098836>
 11. Lalain, M., Ghio, A., Fredouille, C., **Rebourg, M.**, Belleville, L., Duval, M., Balaguer, M., & Woisard, V. (2019, mai). Mots ou pseudo-mots? Quel impact sur l'évaluation perceptive de l'intelligibilité? *Journées de Phonétique Clinique*. <https://hal.archives-ouvertes.fr/hal-02482566>
 12. **Rebourg, M.**, Lalain, M., Ghio, A., Fredouille, C., & Woisard, V. (2019). Pertinence de l'utilisation de non mots pour évaluer l'intelligibilité. *Journées de Phonétique Clinique*, 172. <https://hal.archives-ouvertes.fr/hal-02098845>
 13. Ghio, A., Lalain, M., Giusti, L., Pouchoulin, G., Robert, D., **Rebourg, M.**, Fredouille, C., Laaridh, I., & Woisard, V. (2018). Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique. *XXXIle Journées d'Études sur la Parole*, 285-293. <https://doi.org/10.21437/JEP.2018-33>
 14. **Rebourg M.** (2018). Poster : Validation d'une tâche de Décodage Acoustico Phonétique : lexicalisation, mémorisation, familiarisation. *Journée de Linguistique Expérimentale JLex*, Aix-en-Provence, Avril 2018
 15. Ghio, A., Lalain, M., Giusti, L., Pouchoulin, G., **Rebourg, M.**, André, C., Fredouille, C., & Woisard, V. (2017). *Du décodage acoustico-phonétique pour mesurer l'intelligibilité de locuteurs atteints de troubles de production de la parole*. 3.

Participation aux conférences

1. Journées d'Etudes sur la Parole, Juin 2022, Noirmoutier, France, Communication orale
2. Journées de Phonétique Clinique, 2021, Toulouse, France, Communication affichée
3. Journée Annuelle des Doctorants, 2021, Aix-en-Provence, France, Communication affichée
4. Journée du Rectorat, 2020, Aix-en-Provence, Communication orale
5. Journée Jeunes&Chercheurs, Ligue Contre le Cancer, Paris, Octobre 2020, Communication affichée
6. Journée d'Etudes sur la Parole, 2020, Nancy, France, Communication orale
7. Journées de Phonétique Clinique, 2019, Mons, Belgique, Communication affichée
8. Journée Annuelle des Doctorants, 2019, Aix-en-Provence, Communication orale

Résumé

La perte d'intelligibilité représente l'une des plaintes principales des patients traités pour un cancer de la cavité buccale ou de l'oropharynx et son évaluation perceptive est primordiale pour la prise en charge des troubles de la parole en contexte clinique. Cependant, la définition de la notion d'intelligibilité ne fait pas consensus dans la littérature et est généralement présentée au sens large comme « le degré de précision avec lequel un message est compris par un auditeur ». L'intelligibilité ainsi définie néglige la complexité de la communication parlée et ne permet pas de préciser le niveau linguistique considéré par l'évaluation. Cette imprécision est susceptible de perturber l'évaluation clinique.

Cette thèse se propose d'apporter des éléments de réflexion linguistiques pour questionner la notion d'intelligibilité et préciser sa définition dans le cadre de l'évaluation des troubles de la parole chez ces patients traités pour un cancer de la cavité buccale ou de l'oropharynx. En explorant les relations complexes entre production et perception de la parole, et l'interdépendance des niveaux phonétique et phonologique, au service de l'évaluation de l'intelligibilité segmentale en contexte clinique, nous proposons de définir l'intelligibilité comme « le degré de précision avec lequel les unités linguistiques d'un signal acoustique de parole, produit par un locuteur, sont décodées par un auditeur ». Cette définition reflète mieux la complexité qui réside dans cette notion : l'évaluation de la qualité de production des unités linguistiques produites par un locuteur au travers de la perception d'un auditeur. Nous avons mis à l'épreuve cette conception de l'intelligibilité en proposant l'évaluation des productions de patients enregistrés dans le cadre du projet C2SI, et ce, selon 2 approches.

La première approche est une évaluation perceptive par décodage acoustico-phonétique (DAP) de pseudo-mots. Le principe de construction de ce test ainsi que la méthode de calcul du résultat permettent d'obtenir un score de Déviation Phonologique Perçue (score PPD) qui rend compte du degré d'altération de la parole (score global) et qui permet d'identifier précisément les segments de parole altérés (score local). Les trois expériences conduites pour l'évaluation perceptive de 40 locuteurs (20 patients et 20 contrôles) montrent que les pseudo-mots constituent un matériau linguistique pertinent pour évaluer l'intelligibilité segmentale, puisqu'ils permettent de neutraliser les biais couramment identifiés d'apprentissage et de restauration lexicale ainsi que l'effet d'expertise auditive clinique. Enfin, une analyse des scores locaux permet de préciser le degré d'altération des différents segments vocaliques.

La deuxième approche consiste en des analyses acoustiques des productions des locuteurs et le lien potentiel entre les évaluations perceptives et acoustiques. Ces analyses acoustiques portent sur les productions de 127 locuteurs (87 patients et 40 contrôles). Focalisées sur les segments vocaliques elles sont basées sur les valeurs de formants et permettent le calcul de métriques acoustiques. Les résultats montrent que les mesures

acoustiques (Vowel Space Area (VSA) et Indice de Distinctivité (ID)) ne permettent pas de prédire les scores d'intelligibilité segmentale (globale et locale), mais que l'ID constitue une métrique acoustique complémentaire aux scores PPD obtenus par évaluation perceptive.

L'évaluation de l'intelligibilité des patients traités pour un cancer de la cavité buccale ou de l'oropharynx est améliorée par une mesure segmentale et au moyen d'une évaluation perceptive par DAP complétée par une estimation du système vocalique. Ces travaux contribuent ainsi à l'amélioration de l'évaluation clinique des troubles de la parole ; l'évaluation perceptive par DAP en particulier montre un fort potentiel d'évaluation qui permet d'envisager son transfert vers une application clinique.

Mots clés : Intelligibilité, Cancer ORL, Production de la parole, Perception de la parole, Phonétique clinique, Acoustique

Abstract

Loss of intelligibility is one of the main complaints of patients treated for ENT cancer and its perceptual assessment is essential for the management of speech disorders in the clinical setting. However, the definition of the notion of intelligibility does not have a consensus in the literature and generally presented in a broad sense as "the degree of precision with which a message is understood by a listener". Intelligibility thus defined neglects the complexity of spoken communication and does not allow to specify the linguistic level considered by the evaluation. This imprecision is likely to disturb the clinical evaluation.

This thesis proposes to bring linguistic elements of reflection to question the notion of intelligibility and to specify its definition within the framework of the evaluation of speech disorders in ENT cancer. By exploring the complex relationships between speech production and perception, and the interdependence of the phonetic and phonological levels, in the service of the evaluation of segmental intelligibility in a clinical context, we propose to define intelligibility as "the degree of precision with which the linguistic units of an acoustic speech signal, produced by a speaker, are decoded by a listener". This definition better reflects the complexity that lies in this notion: the evaluation of the production quality of linguistic units produced by a speaker through the perception of a listener. We have tested this conception of intelligibility by proposing the evaluation of the productions of patients recorded in the framework of the C2SI project, and this, according to 2 approaches.

The first approach is a perceptive evaluation by acoustic-phonetic decoding (APD) of pseudowords. The principle of construction of this test as well as the method of calculation of the result make it possible to obtain a score of Perceived Phonological Deviation (PPD score) which gives an account of the degree of alteration of the speech (global score) and which makes it possible to identify precisely the segments of speech altered (local score). The three experiments conducted for the perceptual evaluation of 40 speakers (20 patients and 20 controls) show that pseudowords constitute a relevant linguistic material to evaluate segmental intelligibility, since they neutralize the commonly identified biases of learning and lexical restoration as well as the effect of clinical auditory expertise. Finally, an analysis of local scores allows us to specify the degree of impairment of the different vowel segments.

The second approach consists of acoustic analyses of the speakers' productions and the potential link between perceptual and acoustic evaluations. These acoustic analyses were performed on the productions of 127 speakers (87 patients and 40 controls). Focused on vowel segments, they are based on formant values and allow the calculation of acoustic metrics. The results show that acoustic measures (Vowel Space Area (VSA) and Distinctiveness Index (DI)) do not predict segmental intelligibility scores (global and

local), but that DI is a complementary acoustic metric to PPD scores obtained by perceptual assessment.

The intelligibility assessment of patients treated for oral cavity or oropharyngeal cancer is improved by segmental measurement and by means of perceptual assessment by DAP complemented by an estimation of the vowel system. This work contributes to the improvement of the clinical evaluation of speech disorders; the perceptual evaluation by DAP in particular shows a strong potential of evaluation which makes it possible to envisage its transfer towards a clinical application.

Keywords : Intelligibility, Head and Neck Cancer, Speech production, Speech perception, Clinical phonetic, Acoustic

Remerciements

En premier lieu je tiens à remercier particulièrement la Ligue Contre la Cancer qui a apporté son soutien financier à ce projet de recherche au moyen d'un contrat doctoral, sans lequel ces travaux n'auraient pu être menés à bien. Ce soutien a été possible grâce à des dons de citoyens soucieux de s'engager dans la recherche contre le cancer, merci à tous les donateurs.

Je remercie les membres du jury qui me font l'honneur d'avoir accepté d'évaluer ces travaux de recherche, Mesdames Lorraine Baqué et Lise Crevier Buchman et Messieurs Rudolph Sock et Didier Demolin. En espérant que la lecture de ce manuscrit vous soit plaisante.

Je remercie mon directeur de recherche, Professeur Nicolas Fakhry, pour son soutien, sa bienveillance et sa réactivité. Nos regards croisés sur ces travaux ont été enrichissants. J'espère vous avoir apporté de nouvelles perspectives sur le handicap communicationnel des patients.

Je remercie Muriel Lalain, co-directrice de recherche, tu as su me guider dans l'apprentissage du métier de chercheur avec pédagogie, tu as partagé avec moi tes qualités professionnelles, mais également personnelles et humaines. Nos échanges m'ont apporté bien plus que ce que tu peux imaginer. Tes conseils avisés, justes et bienveillants ont retenti jusque dans ma vie personnelle. J'ai beaucoup appris, merci infiniment.

Merci au laboratoire Parole et Langage de m'avoir accueillie pendant ces trois années, la structure du laboratoire et sa plateforme expérimentale constituent un environnement de travail privilégié. Merci à la direction pour l'attention portée au bien-être des doctorants durant cette période particulière. Merci à Stéphanie et Cathy, toujours disponibles et réactives, Claudia et Joëlle, Cyril et Sébastien, avec vous tout problème trouve sa solution.

Merci à l'équipe Physiologie et Organisation de la Parole, dirigée par Caterina Petrone et Pr Antoine Giovanni, ainsi que l'ensemble de ses membres, pour avoir offert un cadre de présentation et de soutien bienveillant aux doctorants en leur permettant de s'exprimer sur les enjeux scientifiques de l'équipe, et de présenter leurs travaux.

Je remercie les membres des équipes des projet recherche C2SI et RUGBI, vos compatibilités ont permis de développer un environnement de travail collaboratif riche et stimulant, au service d'un objectif commun. Je remercie chaleureusement la porteuse de ces projets, le Dr Virginie Woisard, toujours juste et objective. Merci également à nos

collègues du LIA, et plus particulièrement à Corinne Fredouille pour sa collaboration dans l'alignement et la phonétisation des données de corpus et de perception. Corinne et Virginie, merci pour votre bienveillance.

Je remercie l'ensemble des orthophonistes qui ont participé aux tests de perception. Votre engagement auprès des patients est exemplaire. Merci pour votre disponibilité et pour le temps que vous avez consacré à la recherche.

Je remercie mes camarades doctorants, nous sommes sur le même navire et l'entraide est la clef de l'arrivée à bon port de chacun de nous.

Airelle, petit fruit des bois, continue d'illuminer ceux qui ont la chance de te connaître.

Lydia, tu es ma plus belle rencontre, un master et un doctorat ont scellé notre amitié irrévocablement.

Alain, merci pour toutes tes qualités, que tu partages sans réserve. Ton expérience et ta sagesse sont inestimables.

Anna, merci pour ta bienveillance et ta disponibilité à mon égard.

Mes amis en *or qui brille*, Marie, Pierre-Marie, Alexandre et ceux qui se reconnaîtront dans cette description, votre soutien, vos rires et nos moments ensemble sont aussi précieux que vous dans ma vie. Merci d'être vous. Les amis sont la famille que l'on choisit.

Je remercie l'homme de ma vie, toujours à mes côtés. Ces années ensemble m'emmènent chaque jour vers une meilleure version de moi-même.

Merci maman, tu as été, es et seras toujours là pour moi. Ton soutien indéfectible et ton amour inconditionnel m'ont portée chaque jour. Sans toi, je ne suis pas.

Je vous suis à tous reconnaissante pour votre soutien permanent.

Je tenais également à remercier les sponsors officieux de la rédaction de ce document, les chocolateries Puyricard et Stéphane Roux. Vos savoureuses douceurs sont un moment d'évasion.

A ceux qui veillent sur nous,

Table des matières

Affidavit	2
Affidavit	3
Liste des publications	4
Participation aux conférences	6
Résumé	7
Abstract	9
Remerciements	11
Table des matières	14
Introduction	16
Panorama scientifique et clinique	19
1. Intelligibilité et cancérologie ORL	20
1.1. Intelligibilité	20
1.1.1. La notion d'intelligibilité	20
1.1.1. La mesure de l'intelligibilité	24
1.2. Cancérologie ORL et séquelles fonctionnelles	28
1.2.1. Cancers ORL généralités	28
1.2.2. Les traitements	31
1.2.3. Les séquelles fonctionnelles	33
2. Processus et concepts linguistiques	37
2.1. Production et perception de la parole	37
2.1.1. La production de la parole	37
2.1.2. La perception de la parole	42
2.1.3. Production et perception de la parole	45
2.2. Concepts de phonétique et phonologie	56
2.2.1. Phonétique articulatoire	58
2.2.2. Phonétique acoustique et perceptive	61
3. Evaluation de l'intelligibilité	66
3.1. Evaluation perceptive	67
3.2. Evaluation perceptive segmentale	69
3.3. Evaluation de l'intelligibilité segmentale	72
3.3.1. Méthode de Décodage acoustico-phonétique	73
3.3.2. Contribution des variables acoustiques à intelligibilité	80
Comment évaluer l'intelligibilité segmentale après un cancer ORL ?	88
4. Approche par Décodage Acoustico Phonétique	91
4.1. Corpus	91
4.1.1. Corpus BECD et DAP_m	91
4.1.2. Corpus DAP_131	97

4.1.3.	Corpus : synthèses	101
4.1.4.	Traitement des enregistrements	101
4.2.	Méthodologie	102
4.2.1.	Design expérimental « BECD DAP »	102
4.2.2.	Design expérimental « Quantité Qualité »	105
4.2.3.	Méthodologie d'un test de perception	108
4.2.4.	Traitement des données de perception	108
4.3.	Résultats	111
4.3.1.	BECD – DAP	111
4.3.2.	Naïf – Expert	113
4.3.3.	Quantité – Qualité	117
5.	Scores Perceptifs Segmentaux	121
5.1.	Corpus C2Si	121
5.1.1.	Population	121
5.1.2.	Métadonnées cliniques	122
5.2.	Méthodologie	125
5.2.1.	Contexte	125
5.2.2.	Principes	126
5.3.	Résultats	130
5.3.1.	PPD locaux	130
5.3.2.	Transcriptions des auditeurs	138
6.	Approche par mesures acoustiques	148
6.1.	Corpus	148
6.2.	Méthodologie	149
6.2.1.	F1/F2	149
6.2.2.	Métriques	160
6.3.	Résultats	164
6.3.1.	Statistiques	164
6.3.2.	Visualisation	167
	Discussion	173
	Conclusion	187
	Bibliographie	190
	Table des figures	201
	ANNEXES	205
A.	Exemple de spectrogramme	205
B.	Projet C2Si	206
	Locuteurs et tâches de production	207
C.	Questionnaire auditeurs experts	209
A.	Scripts Praat	211
B.	Script R	215

Introduction

L'acte de communication parlée implique au moins deux interlocuteurs : un locuteur qui transmet sa pensée au moyen de sons articulés, à un auditeur qui en saisit le sens. Au cours de l'acte communicationnel, leurs rôles sont interchangeables. Au moment où vous débutez la lecture de cette communication écrite, vous vous engagez, en tant que destinataire du message, dans un acte de communication, et participez ainsi à sa réussite. L'étymologie du terme communiquer, « mettre en commun », « partager » suppose en effet que cet acte implique la coopération des différents protagonistes qui y sont engagés. Lecteur, vous y prenez part, dans ce contexte particulier, avec vos rôles *de rapporteur et d'examineur*, vos connaissances et vos spécialités. Ainsi, la réussite de cette communication ne dépend pas seulement de nous, elle dépend aussi de vous. Il s'agira de vous transmettre *par des mots* notre réflexion au sujet de l'intelligibilité.

Dans le cadre de la mesure de la qualité d'un échange communicationnel, les termes « compréhensibilité » et « intelligibilité » sont souvent associés, voire confondus ; cela s'entend si l'on considère l'origine latine du terme *intelligibilis* « que l'on peut comprendre ». Pourtant, ces termes ne sont pas synonymes, et s'ils sont liés, ils le sont par la question de la mesure et de l'évaluation, qui permettent d'apprécier la qualité de l'échange communicationnel.

Dans le contexte clinique des troubles de la parole, l'appréciation de la sévérité du trouble est possible par l'évaluation perceptive de l'intelligibilité, mesure clé du handicap communicationnel. Cette évaluation doit permettre en bilan initial, de préciser la perte d'intelligibilité afin de proposer un plan thérapeutique adapté. Au cours du suivi, elle permet d'apprécier l'efficacité des traitements (rééducation orthophonique, prothèses) et de les modifier en conséquence. La précision de la mesure, cruciale en clinique, fait pourtant défaut et peut conduire à une mésestimation de l'intelligibilité, puis à une prise en charge peu satisfaisante. La qualité de la prise en charge des troubles de la parole est pourtant indispensable pour réduire le handicap communicationnel. Dans le contexte des cancers ORL, les patients traités pour un cancer de la cavité buccale ou de l'oropharynx, s'ils voient leur espérance de vie augmenter, présentent en suite des traitements, des séquelles fonctionnelles importantes qu'il convient de prendre en charge pour améliorer leur qualité de vie.

C'est cette préoccupation pour la qualité de vie des patients qui est à l'origine du projet « Carcinologic Speech Severity Index, C2SI » proposé par Virginie Woisard (2014). L'objectif de ce projet consistait en la création d'un indice automatique de sévérité de la parole applicable en cancérologie des voies aéro-digestives supérieures. Pour développer cet indice, les équipes partenaires du projet ont proposé un certain nombre de tâches linguistiques destinées à la création d'un important corpus de parole utilisé ensuite pour des analyses perceptives et automatiques. C'est dans ce cadre que la tâche de décodage

acoustico-phonétique pour une évaluation perceptive de l'intelligibilité a été développée. Durant notre Master, nous avons contribué à ce projet en participant au recueil du corpus, à la passation des tests de jugement perceptif, au traitement et à l'analyse des données. Nos premiers travaux de recherche dans ce domaine nous ont alors conduite à nous intéresser aux méthodes d'évaluation de l'intelligibilité. Ce projet est ainsi le berceau de cette thèse dont les principes méthodologiques s'appuient sur une partie du corpus constitué : la tâche de production de pseudo-mots dédiée à l'évaluation perceptive par décodage acoustico-phonétique.

Ainsi, l'évaluation de l'intelligibilité chez les patients traités pour un cancer de la cavité buccale et de l'oropharynx est au cœur de ces travaux. Nous avons voulu ici questionner la précision de la mesure de l'intelligibilité, comprendre ce qui faisait défaut à l'évaluation perceptive, gold standard de l'évaluation clinique, afin de proposer éventuellement une mesure plus fiable.

Questionner la précision de la mesure nous a irrémédiablement conduite à interroger la définition même du concept d'intelligibilité. Dans la première section de ce manuscrit, nous présenterons le panorama scientifique et clinique sur lequel s'est appuyée notre réflexion. Nous montrerons que le cadre définitoire relatif à la notion d'intelligibilité, en ce qu'il a d'imprécis, impacte nécessairement sa mesure. A travers la présentation des différentes définitions du terme, nous montrerons qu'il est en réalité un terme générique dans lequel cohabitent difficilement les concepts de compréhensibilité et de compréhension. Cet amalgame fâcheux de termes proches mais non synonymes constitue selon nous la première cause identifiable aux problèmes de mesure. Cette confusion des termes, utilisés dans le cadre de l'évaluation perceptive, trouve probablement son origine dans le fait que lorsqu'il s'agit d'évaluer l'intelligibilité, seul l'auditeur est considéré. En d'autres termes, évaluer l'intelligibilité revient à quantifier ce que l'auditeur comprend de l'échange. Cette conception rudimentaire de l'évaluation de l'intelligibilité néglige la complexité de l'échange communicationnel. Nous proposons de considérer quelques modèles explicatifs des mécanismes de production et de perception de la parole, en mettant en lumière la relation complexe qui les unit, et ce quelles que soient les unités linguistiques considérées. Nous verrons que l'évaluation de l'intelligibilité, essentiellement perceptive dans le domaine clinique pâtie de l'utilisation de batteries de tests qui ne tiennent pas compte de la variabilité de la production, de l'interdépendance des niveaux linguistiques impliqués, ni de l'adaptabilité de la perception.

Ceci nous amènera à reconsidérer spécifiquement l'évaluation de l'intelligibilité des troubles de la parole séquellaires des traitements des cancers ORL : Nous expliciterons les éléments qui selon nous doivent être pris en compte dans l'évaluation d'un trouble qui touche l'articulation de la parole. En nous appuyant sur les concepts de phonétique et de phonologie, et plus particulièrement des phonèmes et des traits distinctifs, nous défendrons notre conception de l'évaluation de l'intelligibilité segmentale de la parole. Elle sera basée sur une approche perceptive par décodage acoustico-phonétique et sur des mesures acoustiques des segments vocaliques.

Dans la seconde section de cette thèse, nous présenterons notre proposition pour une évaluation de l'intelligibilité segmentale pertinente et objective au moyen de deux approches. En premier lieu, nous aborderons l'approche perceptive de l'évaluation de l'intelligibilité. Il sera question d'évaluer la pertinence de l'utilisation de pseudo-mots pour neutraliser les biais perceptifs couramment rencontrés dans l'évaluation perceptive clinique. Nous présenterons les éléments méthodologiques mis en place, comprenant la description des corpus, plusieurs des designs expérimentaux basés sur des tests de perception, le traitement et l'analyse des données ainsi que les résultats obtenus sur la base de calculs de scores d'intelligibilité globale. Nous proposerons également une analyse des scores d'intelligibilité locale que la méthode d'évaluation perceptive par DAP permet d'obtenir. Cette analyse nous permettra d'évaluer la robustesse et la précision d'évaluation que permet cette méthode, et d'identifier les segments d'intérêt pour la suite de nos analyses.

La dernière partie dédiée à l'approche par mesures acoustiques des segments vocaliques comportera les éléments méthodologiques mis en place pour cette étude comprenant le corpus, l'analyse de l'extraction des mesures acoustiques, la description des méthodes de calcul des métriques, ainsi que le traitement et l'analyse statistique des données. Enfin, nous présenterons ainsi que les résultats concernant le potentiel de prédiction des indices calculés et leurs liens avec les évaluations perceptives.

Nous discuterons enfin nos résultats, au regard des données théoriques, des enjeux de l'évaluation clinique des troubles de la parole et des limitations méthodologiques. Les éléments de discussions nous conduiront à conclure autour de perspectives de travaux de recherche et d'applications cliniques possibles.

Panorama scientifique et clinique

« L'évaluation de l'intelligibilité après un cancer ORL », l'expression, dans ce qu'elle a de simple ou de presque banal, pourrait évoquer une routine clinique classique, aux enjeux clairement définis et aux méthodes nécessairement robustes. Il n'en est pas ainsi, les relations entre les enjeux et la méthode sont autrement plus complexes.

Nous présenterons dans cette première section le paysage théorique sur lequel s'est appuyée notre réflexion concernant cette problématique de l'évaluation de l'intelligibilité. Nous verrons que le terme « intelligibilité », tant sur la scène de la recherche que dans le paysage clinique est susceptible de recouvrir différentes réalités. Cette imprécision, en ce qu'elle se répand sur le choix des méthodes d'évaluation, des batteries de tests, des modalités de réponses des évaluateurs, ou encore des unités à évaluer, impacte nécessairement l'évaluation qui consiste alors en une mesure dont la fiabilité est jugée moyenne, fragile, variable. Cliniciens et patients sont les premiers à pâtir de cette imprécision. En présentant notre cadre clinique de la cancérologie ORL nous montrerons qu'une réflexion sur l'évaluation de l'intelligibilité ne peut faire l'économie de la prise en compte de la spécificité du trouble de la parole dont il est question et de ce que l'on souhaite précisément évaluer.

Nous évoquerons ici les origines de la notion d'intelligibilité et le lien qui très tôt s'est établi avec la notion de mesure. Nous positionnerons les problématiques liées à la question de « l'évaluation de l'intelligibilité » dans la relation d'interdépendance qu'entretiennent le locuteur et l'auditeur au sein de l'échange communicationnel. Plus précisément, nous essaierons de montrer que lorsqu'il est question de communication parlée, locuteur et auditeur jouent ensemble une partition complexe où des unités linguistiques de différents niveaux sont mises en sons par le système de production du premier, pour être perçues et décodées par le système de perception du second. Ce voyage au cœur de la communication parlée nous entraînera vers une présentation de quelques grands concepts de phonétique et de phonologie. Nous nous appuyerons sur ces concepts dans notre cadre de réflexion et les réinvestirons dans les approches perceptives et acoustiques que nous proposerons dans cette thèse comme des méthodes possibles d'évaluation de l'intelligibilité chez des patients traités pour un cancer ORL.

1. Intelligibilité et cancérologie ORL

Il est presque étonnant de constater que le terme « intelligibilité » est devenu si courant si commun, tant sa définition questionne, interroge. Dans le domaine de la communication parlée, la notion d'intelligibilité est quasi indissociable de la nécessité de mesure, d'évaluation. Nous verrons que les différentes acceptions que revêt l'expression « évaluation de l'intelligibilité » rendent l'exercice aussi complexe qu'il est essentiel lorsqu'il s'agit d'évaluer la sévérité d'un trouble de la parole séquellaire d'un traitement pour un cancer ORL.

1.1. Intelligibilité

Ce qui fait partie du domaine de « l'intelligible » était discuté chez les philosophes de l'Antiquité grecque. Platon (427-347 av. JC), dans sa théorie des idées, oppose le monde qu'il appelle sensible au monde qu'il nomme intelligible. Le monde intelligible, ou monde des Idées ou des Formes, représente le seul monde réel. Le monde sensible n'a de réalité que de par sa « participation » au monde intelligible (Longeart, 2019b). En revanche Aristote (384 – 322 av JC) s'oppose à cette séparation qui selon lui ôte toute intelligibilité au monde, celle-ci étant présente en chaque être (Longeart, 2019a). Débattre et discuter de ce qui est intelligible ou non était déjà en réflexion dans l'esprit des philosophes de l'Antiquité grecque. Ce domaine de réflexion et de recherche a été réinvesti à l'époque moderne grâce à l'essor des moyens de télécommunications et le développement de tests d'évaluation de la parole et continue d'être étudié de nos jours.

1.1.1. La notion d'intelligibilité

Dans les années 1870 A. Graham Bell, dans une tentative d'améliorer le télégraphe, découvre que les sons peuvent être transmis grâce au courant électrique. Consécutivement à cette découverte il mène des expériences permettant l'amélioration de la qualité des transmissions. C'est lors des tests d'amélioration de cette nouvelle machine, qui deviendra le téléphone, que Bell fait de l'intelligibilité une composante scientifique techno-acoustique. L'essor des télécommunications et la nécessité d'évaluer et d'améliorer les systèmes, ont fait de l'intelligibilité une notion centrale pour la mesure de l'efficacité des systèmes. Développée au service de l'évaluation des moyens de télécommunications et de l'amélioration de la transmission de signaux acoustiques (donc de la qualité du système de transmission), l'intelligibilité a rapidement investi le champ de l'évaluation de la parole.

L'intelligibilité de la parole devient alors une notion complexe, intrinsèquement liée à la question de l'évaluation/mesure, et dont la définition ne fait pas consensus, entre les auteurs, et selon qu'on s'y intéresse du point de vue du locuteur, de la transmission ou de l'auditeur. C'est en effet la particularité de l'intelligibilité de la parole : elle est dépendante

d'un système de production, d'un canal de transmission et d'un système de perception. Schiavetti (1992) indique déjà « que toute mesure de l'intelligibilité de la parole est une mesure de l'interaction entre le locuteur, un système de transmission et un auditeur ». L'intelligibilité concerne de nombreux domaines, lorsqu'il est question de transmettre une information avec efficacité : communication parlée, apprentissage des langues secondes, conversation quotidienne, transmission d'information (gare, train, aéroport...), évaluation de parole, mesure de l'impact d'un traitement sur la parole, mesure de déficits articulatoires, ... Elle est employée pour évaluer la performance des systèmes de transmissions de la parole, mais également, si l'on s'en réfère à la proposition de Schiavetti, pour mesurer l'efficacité de la communication parlée.

1.1.1.1. Communication parlée

L'organisation du système de production de la parole est une spécificité humaine. Les organes mobilisés dans cette action répondent avant tout à des fonctions primaires : la respiration et l'alimentation. Au cours de l'évolution ils ont été réemployés de façon inédite, au service d'une faculté unique : la communication parlée. Abitbol et Vallancien dans « Voix, physiologie » nous rapportent la définition de la communication parlée selon Abraham Moles « La communication parlée requiert l'utilisation des organes de la respiration, dans le but d'engendrer des séquences de phénomènes sonores, transitoires, opposables, assemblés dans un tout, suivant certaines règles et dont l'objet est de véhiculer l'expression de la pensée ». Il s'agit donc d'utiliser les moyens offerts par notre système physiologique dans un but communicatif.

Aux prémices de la communication parlée, on retrouve la communication acoustique qui est le fait de transmettre des informations (communiquer) grâce à des signaux sonores. Cela concerne la communication animale par l'émission de vocalisations ou infra/ultra-sons, tels que le chant des oiseaux ou celui des baleines. Bien sûr, la communication acoustique est un acte à la base de la communication parlée humaine, mais les unités d'encodage de cette dernière sont probablement plus complexes, sûrement plus nombreuses et très diversifiées (unité minimale de son (phonèmes), unité lexicale minimale (morphèmes), unité lexicale (mots), phrase, sémantique, grammaire, syntaxe...). La communication a été étudiée et théorisée avant même le développement technologique et technique de mesures des propriétés acoustiques. Nous n'avons pu étudier les propriétés spectrales ou formantiques d'un son avant d'avoir un outil de capture et de représentation du signal acoustique, en revanche nous avons pu étudier la grammaire ou réaliser un inventaire phonémique d'une langue sans outil technologique particulier. En d'autres termes, certaines propriétés des sons ont pu être capturées et décrites au seul moyen du système perceptif humain quand d'autres ont nécessité un certain degré de développement technologique pour être formalisés.

La communication parlée s'organise autour de deux acteurs : le locuteur et l'auditeur. Le locuteur, lorsqu'il produit un signal acoustique bénéficie d'un rétro-contrôle sur sa production au moyen de son propre système perceptif. Le signal acoustique produit à destination de l'auditeur est perçu par celui-ci au moyen de son système perceptif également. Ce sont donc deux systèmes de perception et un système de production qui

sont en jeu dans cette relation. Celle-ci repose sur la mobilisation de plusieurs niveaux linguistiques qui participent à son aboutissement. Une représentation schématique de la chaîne de communication parlée est illustrée dans la Figure 1.

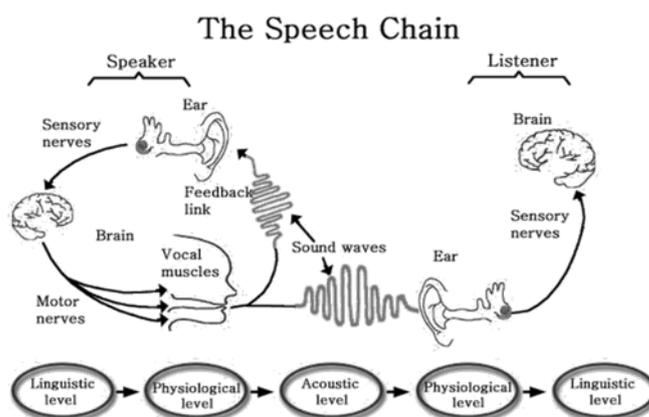


Figure 1 : Schéma de la chaîne de communication parlée (Denes & Pinson, 1993)

L'objectif de la communication réside dans la transmission d'informations au moyen d'unités sonores. Il s'agit de produire et transmettre un signal de parole qui est perçu et décodé par un auditeur au service d'un but commun : communiquer. Dans ce contexte, la question de l'efficacité de la communication émerge rapidement puisque c'est son enjeu même. Elle dépend des mécanismes complexes de production et de perception de la parole, lesquels reposent sur différents niveaux linguistiques dont l'articulation n'est pas plus simple.

Considérée comme élément essentiel dans le cadre de l'échange communicationnel, l'intelligibilité devrait être appréhendée au regard de la complexité des mécanismes et des niveaux linguistiques qu'implique cet échange. Cette première lacune explique en partie le manque de précision terminologique de la notion.

1.1.1.2. Définitions

La quantité de définitions proposées par les différents auteurs reflète l'absence de consensus mais surtout la difficulté définitoire que nous impose le terme :

- « Degré de compréhension d'un message verbal, ou d'une forme de parole (naturelle ou synthétique) déterminé à l'aide de tests de perception » (Larousse, 2019)
- « L'intelligibilité est le degré de précision avec lequel le message est compris par l'auditeur » (Yorkston, 1992)
- « L'intelligibilité de la parole est une expression décrivant la qualité d'écoute et de compréhension d'un échange dans une pièce » (Lambourg, 2002)
- « Dans le domaine de la parole, l'intelligibilité désigne la qualité de compréhension d'un message linguistique contenu dans un signal de parole et relève des domaines linguistique, acoustique et/ou médical. » (Woisard, 2011a)
- « L'intelligibilité de la parole [...] représente le degré de compréhension des signaux vocaux lorsqu'ils sont déformés par des interférences systématiques, environnementales et internes. » (Feng & Chen, 2022)

L'ensemble de ces définitions lie intrinsèquement la notion de compréhension à celle d'intelligibilité ainsi qu'à la notion de 'mesure' sur laquelle nous reviendrons dans un second temps. La relation entre l'intelligibilité et la compréhension est discutable, et dépend de ce que l'on entend par 'intelligibilité', de qui, du locuteur ou de l'auditeur on adopte le point de vue, et du niveau linguistique que l'on considère ou que l'on souhaite évaluer. Si pour Woisard (2011) l'intelligibilité est liée à la compréhensibilité « l'intelligibilité est la capacité à être compris alors que la compréhension est la capacité de comprendre », alors l'intelligibilité est associée au versant de la production de la parole, au service de la communication. En revanche Fontan (2012) ne considère que les aspects perceptifs, puisque pour lui la compréhensibilité représente la capacité de compréhension du sens global du message par l'auditeur. En ce sens, Barefoot et al, (1993) définissent la compréhensibilité « comme la mesure dans laquelle un auditeur comprend des énoncés produits par un locuteur dans un contexte de communication ». Ils proposent toutefois une distinction entre compréhensibilité, qui, selon eux, relève des domaines de la parole et du langage et au sein de laquelle « des énoncés significatifs sont produits par les locuteurs et traités par les auditeurs », et intelligibilité qui « relève principalement du domaine de la parole » et qu'ils ne cantonnent presque qu'au domaine « auditif-acoustique ». Or, c'est la notion même d'intelligibilité qui reflète l'interface production – perception de la parole.

Les auteurs ne s'accordent pas sur la définition de ces notions et leur distinction n'est pas très claire. Si l'intelligibilité « est la capacité à être compris », soit à être compréhensible, et que la compréhensibilité est « la capacité de compréhension », alors intelligibilité et compréhension sont indissociables et quasiment mêmes. Or la compréhensibilité, capacité à être compréhensible et compris relève du domaine du locuteur, mais la capacité de comprendre, la compréhension, tient au rôle de l'auditeur. C'est donc la relation entre les capacités de production d'un locuteur et les capacités perceptives d'un auditeur qui définit l'intelligibilité de la parole. En d'autres termes, la 'compréhension' représente la capacité d'un auditeur à percevoir et décoder un signal de parole, en revanche, la 'compréhensibilité' représente la capacité d'un locuteur à émettre un signal acoustique perceptible et décodable. Enfin, l'intelligibilité représente le rapport entre les capacités de production et perceptives d'individus en situation de communication. Il advient donc réducteur de ne la définir qu'à travers le prisme perceptif.

C'est en effet l'autre critique que l'on peut émettre à propos de ces définitions : elles abordent le concept d'intelligibilité de façon unilatérale, en ne considérant que le pan perceptif de la communication parlée en se plaçant seulement du côté de l'auditeur (« message est compris par l'auditeur », « qualité d'écoute », « qualité de compréhension », « degré de compréhension des signaux vocaux »). L'émetteur du signal est présupposé mais ni sa présence ni sa fonction ne sont explicités alors que c'est bien de sa production dont il est question. Comme nous venons de l'évoquer, la notion d'intelligibilité se définit par la relation qu'entretiennent les niveaux de production et de perception de la parole. C'est là que réside sa complexité ; de plus la réussite de l'échange communicationnel dépend également de nombreux systèmes et niveaux linguistiques partagés entre les interlocuteurs. Eux-mêmes doivent être efficaces et coordonnés.

D'autres auteurs se sont essayés à définir la notion d'intelligibilité de façon plus globale en intégrant à la fois les aspects perceptifs mais également productifs de la communication, en faisant explicitement référence à l'émetteur (le locuteur) et au récepteur (l'auditeur). Comme Schiavetti (1992) selon qui « L'intelligibilité de la parole peut être définie comme la concordance entre l'intention du locuteur et la réponse de l'auditeur à la parole transférée à travers le système de transmission ».

Cependant, aucune définition n'est pleinement satisfaisante : la précision fait défaut et la relation avec la notion de compréhension reste ambiguë. Pourtant, leur dichotomie peut être abordée au travers de conceptualisations finalement assez simples. Prenons par exemple la célèbre phrase de Chomsky (2009) « d'incolores idées vertes dorment furieusement » : elle est intelligible mais pas compréhensible, son sens est incohérent. La phrase « Il fait beau dehors » produite avec cinq marshmallows dans la bouche n'est cette fois ni intelligible ni compréhensible. Des éléments contextuels de haut niveau peuvent tout de même aider l'auditeur à reconstruire le sens d'un message mais les unités phonologiques sont mal décodées, et c'est toute la perception qui est perturbée. En ce sens, on peut être intelligible sans être compréhensible. Les capacités de compréhension en termes d'accès au sens sont perturbées quand l'intelligibilité de la production en termes de qualité est dégradée. Cela suggère que la compréhension entretient une relation de dépendance vis-à-vis de l'intelligibilité puisqu'elle lui constitue un prérequis. Alors que l'intelligibilité elle n'est ni dépendante ni subordonnée à la compréhension. Selon cette idée, la compréhension permet l'intégration des informations de haut niveau alors l'intelligibilité peut être circonscrite à l'intégration des informations de bas niveau. En d'autres termes, l'intelligibilité de la parole constitue un prérequis pour la compréhension alors que l'intelligibilité peut être définie et évaluée de façon autonome, sans recours à la notion de compréhension. Dans cette conception, elle peut être circonscrite aux niveaux segmentaux de la production d'unités phonétiques et de la perception d'unités phonologiques.

Les auteurs abordent également l'intelligibilité en lien avec la notion de mesure. En effet, le concept de 'mesure de l'intelligibilité' n'est pas nouveau puisque la façon dont est abordée cette notion complexe, tant en science qu'en clinique, implique de déterminer une 'mesure' pour qu'elle puisse être quantifiée, notée, et évaluée au regard d'un taux ou d'un score représentatif du déficit communicationnel. La première définition énoncée dans cette section (Larousse, 2019) conçoit finalement, elle aussi, l'intelligibilité comme une mesure, puisqu'elle est « déterminée à l'aide de tests de perception », ce qui implique une évaluation ou un jugement de la part d'un auditeur, et donc l'attribution d'une note ou le calcul d'un taux. Nous verrons par la suite comment les problématiques liées à la notion de mesure contribuent à rendre critique la question de l'évaluation de l'intelligibilité.

1.1.1. La mesure de l'intelligibilité

L'intelligibilité peut être mesurée de deux façons : Les méthodes directes et indirectes (Lambourg, 2002). La méthode directe fait référence à l'oreille humaine, premier organe

mobilisé dans un test d'écoute (jugement perceptif) dans lequel un auditeur écoute et transcrit ce qu'il perçoit (test perceptif). Ses réponses peuvent être ouvertes (transcription libre) ou fermées (choix multiples dans une liste de proposition). Si l'intelligibilité est mesurée au moyen d'outils techniques (dits plus objectifs), comme des mesures d'amplitude, de fréquence, d'énergie, sur le signal acoustique, la méthode de mesure est dite indirecte.

La méthode de mesure directe est privilégiée par les auteurs et constitue une référence en contexte clinique, notamment pour l'évaluation de la voix (GRBAS) ou de la parole. Par exemple, dans leur étude, Bradlow et al, (1996) exploitent les enregistrements de 20 locuteurs (10 hommes et 10 femmes) produisant chacun 100 phrases (corpus Indiana Multi-Talker Sentences Database (Karl & Pisoni, 1994)). Dans chaque phrase se trouvent 5 mots clefs et des mots de fonction. Les données d'intelligibilité de la parole sont basées sur des transcriptions de phrases par 10 auditeurs par locuteurs, soit 200 auditeurs : Chaque groupe de 10 auditeurs écoute le set complet de 100 phrases produites par un seul locuteur. Les auditeurs écoutent chaque phrase et transcrivent ce qu'ils ont entendu sur un clavier d'ordinateur. Ces transcriptions sont scorées sur un critère 'mots clefs' qui considère une phrase correctement transcrite si et seulement si ses 5 mots clefs sont correctement transcrits. S'il y a une erreur de transcription d'un mot clef, c'est toute la phrase qui est considérée comme une erreur. Le score d'intelligibilité globale représente le score moyen obtenu pour les 100 phrases produites par le locuteur.

Ishikawa et al. (2019), dans leur étude sur la perception et l'acoustique de l'intelligibilité des voyelles de l'anglais américain, utilisent également la méthode directe pour déterminer l'intelligibilité. 20 locutrices féminines (10 dysphoniques et 10 contrôles) ont produit 11 voyelles cibles de l'anglais américain, prises dans la structure /h/ voyelle /d/. Cette structure CVC est enchâssée dans une phrase porteuse « I'm going to say ... », soit « I'm going to say /h/ voyelle /d/ » (« Je vais dire /h/ voyelle /d/ »). 10 auditeurs naïfs écoutent les productions et « sélectionnent le mot qu'ils entendent à partir des propositions disponibles sur l'écran », soit effectuent un choix forcé et restreint parmi une liste de propositions de mots affichés à l'écran. Dans ces travaux, l'intelligibilité est mesurée par le pourcentage de mots cibles et donc de voyelles correctement identifiées.

La méthode indirecte se veut plus objective car l'extraction d'informations acoustiques à l'aide d'outils automatiques permettrait de s'affranchir de la variabilité perceptive et serait donc plus régulière. C'est pourquoi, certains auteurs explorent les capacités de mesures acoustiques indirectes à prédire l'intelligibilité de la parole, au travers des corrélats mesures acoustiques et scores d'intelligibilité. Dans la suite de leurs études, Ishikawa et al. (2019) effectuent par exemple des mesures acoustiques sur les signaux de parole enregistrés par les locuteurs. Ils choisissent des mesures de formants (F0, F1 et F2) et cherchent ensuite des corrélations avec les scores d'intelligibilité déterminés au travers d'une méthode d'évaluation directe. Nous notons que cette méthode indirecte est peu courante en clinique et que l'objectivité de la mesure obtenue reste discutable.

Ces deux méthodes d'évaluation, directes et indirectes, ne sont pas équivalentes : la perception humaine est variable et les études de (Jarzé et al., 2017a) ont montré de

grandes variabilités inter et intra individuelles dans l'évaluation de l'intelligibilité de la parole. C'est donc le rapport entre la capacité de production d'un locuteur et la capacité de décodage d'un auditeur qui est mis en jeu. Schiavetti (1992) conçoit ainsi la « mesure » de l'intelligibilité de la parole comme « une mesure de l'interaction entre le locuteur, un système de transmission et un auditeur ». Le choix du terme 'interaction' implique la notion de réciprocité entre l'action du locuteur et la réaction de l'interlocuteur. On comprend ainsi que la méthode directe puisse être relativement subjective.

De plus, Miller (2013) souligne la grande variabilité de pratiques lorsqu'il s'agit de mesurer l'intelligibilité : la tâche assignée au locuteur peut ainsi être de production de phrases, de mots, en condition de lecture ou de répétition. La mesure, ou plus précisément le calcul des scores est susceptible d'être calculé sur l'ensemble d'une production, ou bien sur des mots-clefs ou encore sur des segments. Le score lui-même peut être exprimé en note (2/10) ou bien en taux (pourcentage d'unités reconnues). Enfin, la tâche de l'auditeur diffère d'un test à l'autre. Ce dernier peut être amené à scorer l'intelligibilité à partir d'une échelle visuelle analogique, ou bien à transcrire les productions perçues, ou encore à les répéter.

On observe tout de même une constante : mesurer l'intelligibilité revient toujours à mesurer la performance d'un locuteur / émetteur au travers d'un système de perception humain ou automatique. Cette mesure concerne l'efficacité de la transmission d'un signal acoustique, résultant de l'interaction entre un locuteur et un auditeur, avec une attention particulière sur le premier. Ce sont ses capacités et habilités productives qui sont évaluées au travers d'une mesure de l'intelligibilité, mais cette évaluation reste dépendante d'outils de perception. Mesurer ou évaluer l'intelligibilité implique ainsi un processus complexe de « discrimination et d'identification » (G. A. Miller et al., 1951) des unités linguistiques.

La problématique de la mesure concerne ainsi également les unités ciblées dans celles-ci. Dans « Articulation testing methods », Fletcher & Steinberg (1929), définissent la mesure d'intelligibilité « comme le pourcentage du nombre total de mots et phrases parlées, respectivement, qui sont correctement interprétés »¹. Cette mesure ainsi définie représente un taux de lexique correctement identifié. Dans l'étude de Bradlow et collègues, ce sont les 'mots-clefs' déterminants dans des phrases qui seront l'objet des mesures ; c'est le degré d'intelligibilité des mots qui détermine le taux d'intelligibilité des phrases. Ils évaluent donc deux niveaux différents de l'intelligibilité de la parole. Un niveau lexical et niveau phonémique : un niveau macro et un niveau micro linguistique. Ces différences de niveaux considérés et donc de cibles évaluées reflètent le flou du cadre définitoire. Cela souligne l'importance de préciser la définition de l'intelligibilité en y intégrant la notion de niveau linguistique. En effet, dans une conception large de l'évaluation l'intelligibilité au moyen d'une mesure, certains effets perceptifs, comme la restauration lexicale, sont permis par le recours aux informations de haut niveau linguistique (lexical, sémantique et contextuelle). Relativement au niveau qui est

¹ « as the percentage of the total number of spoken words and sentences, respectively, that are correctly interpreted »

concerné par l'évaluation ces mécanismes perceptifs sont susceptibles de perturber l'évaluation s'ils ne sont pas considérés dans la mesure établie. En revanche, dans une conception précisée/affinée de l'intelligibilité, en termes segmental (comme ce sera le cas dans ce travail), ce recours peut être contrôlé et l'évaluation objectivée en précisant les unités dont il est question. La précision des unités évaluées doit donc faire partie intégrante de sa définition, afin d'en préciser la mesure, au regard du cadre clinique considéré.

En clinique, la mesure de l'intelligibilité permet de mesurer le handicap à la communication induit par un trouble de la production de la parole. Son évaluation permet de déterminer quelles sont les composantes linguistiques qui sont dégradées et préservées, et donc d'adapter la rééducation et le parcours de soin du patient. Elle permet également de mesurer l'effet d'un traitement sur la production de la parole en contexte pathologique. Dans ce contexte, la perte d'intelligibilité est définie comme le déficit fonctionnel de la parole réduisant la capacité à agir avec autrui (Lindblom, 1990b). Mesurer le déficit fonctionnel d'un locuteur se réalise au travers d'une évaluation fonctionnelle de la parole dont l'intelligibilité est la mesure clef (Kent et al., 1999).

Fang et al. (2017) trouvent d'autant plus complexe d'évaluer l'intelligibilité de la parole pathologique, en effet « l'évaluation de l'intelligibilité de la parole est difficile car elle est généralement non stationnaire et mutationnelle », mais également par ce que c'est une notion multiple. La difficulté à définir une notion reflète sa complexité. Cela traduit finalement différentes conceptions de l'intelligibilité, d'une part au sens large, dans une acception globale, qui intègre tous les niveaux linguistiques mis en jeu dans la communication, mais également dans une acception plus étroite en référence à chaque niveau linguistique.

Dès 1996, Yorkston et al partent du principe fondamental que le trouble de la parole induit un handicap de communication, et que le paramètre fondamental de ce handicap est l'intelligibilité de la parole. Cette perte d'intelligibilité représente un handicap de communication qui impacte la qualité de vie d'un locuteur au travers de sa capacité à évoluer dans son environnement linguistique. Cet environnement est constitué de plusieurs niveaux qui apparaissent hiérarchisés au regard des théories linguistiques de production et de perception de la parole. La perturbation articulaire de la production des unités phonologiques de bas niveaux impacte la production des unités supérieures et par ricochet la transmission et la perception du signal acoustique. En ce sens l'intelligibilité segmentale contribue à l'intelligibilité globale et elles visent ensemble à satisfaire l'enjeu de la communication.

La notion d'intelligibilité peut alors s'aborder et être évaluée de manière générale, en intégrant l'ensemble des niveaux linguistiques nécessaires à la réussite de l'échange communicationnel. Elle est alors « l'intelligibilité de la parole ». Toutefois, on peut également la concevoir de manière restreinte en se focalisant sur un niveau linguistique particulier. On peut envisager de préciser sa définition au regard du niveau linguistique évalué. L'intelligibilité peut alors se concevoir en termes segmentaux en référence aux

niveaux linguistiques de production et de perception. Cette conception élémentaire, à la fois claire et adaptable précise sa définition sans pour autant occulter sa complexité.

Les cancers de la cavité buccale et de l'oropharynx affectent directement les structures anatomiques impliquées dans la production de la parole et par répercussion la production des unités phonémiques de bas niveau. L'évaluation de l'intelligibilité segmentale revêt un caractère particulièrement intéressant, puisque la restauration des unités de bas niveau participe à l'efficacité des niveaux supérieurs.

1.2. Cancérologie ORL et séquelles fonctionnelles

L'évaluation de l'intelligibilité en contexte clinique est régulièrement conduite auprès de patients présentant un trouble de la parole dont l'origine peut varier : neurologique, traumatique, congénitale, neurodégénérative... Cette problématique mériterait sans doute d'être repensée pour ces différentes étiologies. Dans cette thèse, notre réflexion est ancrée dans le cadre clinique spécifique de la cancérologie ORL dont la description est proposée ci-après. En effet, regroupés sous les termes « cancers des VADS », ces cancers et leurs traitements sont susceptibles d'impacter la production de la parole, au-delà des impacts sur les fonctions vitales (respiration, alimentation). Localisés dans la sphère ORL, ils sont regroupés sous les termes « cancers des voies aérodigestives supérieures » (VADS).

1.2.1. Cancers ORL généralités

Les données globales d'épidémiologie des cancers communiquées par l'Institut National du Cancer (INC) indiquent, pour l'année 2018, plus de 380 000 cas de cancers tous confondus (54 % chez l'homme, 46 % chez la femme). Il est estimé, pour la même année, que plus de 150 000 décès sont attribuables aux cancers (57 % chez l'homme, 43 % chez la femme), (*Données globales d'épidémiologie des cancers - Epidémiologie des cancers*, s. d.).

Pour les cancers des VADS, en 2008, la Haute Autorité de Santé (HAS) compte 16 000 nouveaux cas par an, et presque 10 ans plus tard, pour l'année 2017, l'INC recense plus de 15 000 cas, l'incidence de ces cancers reste relativement stable au cours du temps. Elle reste toutefois élevée, ces cancers se situent au 5^{ème} rang des cancers les plus fréquents en France. Les hommes sont plus touchés que les femmes par ces cancers, en revanche, on relève une évolution de l'incidence : en 2008, la HAS indique que les hommes représentent 80% des nouveaux cas alors qu'en 2017 ils représentaient environ 70% des cas. Ainsi, l'incidence masculine tend à baisser alors que celle des femmes augmente (environ 10% en 10 ans). Cette tendance est directement liée à l'évolution de la consommation de tabac et d'alcool dans la population féminine. Ces cancers surviennent le plus souvent entre 50 et 64 ans. Environ 5000 patients décèdent chaque année, soit

environ 33% des nouveaux cas. Enfin, la HAS rapporte qu'en 2007 les Affections Longues Durées (ALD) pour les cancers des VADS concerne près de 60 000 patients en France.

Les facteurs de risque principaux sont la consommation de tabac et d'alcool. Pris ensemble, les effets de ces substances se démultiplient, c'est l'effet synergique sur le risque de cancer des VADS. L'intoxication alcoolo-tabagique concerne 90% des patients atteints. D'autres facteurs de risques sont avérés ; dans une moindre mesure, l'infection par certains types de papillomavirus humain (HPV), qui peut favoriser un cancer de l'oropharynx, ou encore l'exposition professionnelle aux poussières de bois ou de cuir sont reconnus comme facteurs de risque responsables de cancers des sinus et des fosses nasales.

Dans cette large catégorie de cancers des VADS on retrouve notamment les cancers de la cavité buccale et de l'oropharynx. Ils représentent respectivement 20 à 25% et 10 à 15% des cancers des VADS (HAS). Les cancers du larynx, qui représentent 30 à 35% de cas de cancers des VADS, et leurs traitements entraînent des altérations spécifiques susceptibles de désorganiser le système phonatoire. Les voies respiratoires et les voies digestives, qui se séparent au niveau du larynx, peuvent être amenées à ne plus communiquer entre elles. Les perturbations laryngées entraînent des troubles de la voix alors que des pathologies de la cavité buccale ou de l'oropharynx résultent plutôt des troubles de la production de parole.

Les troubles de la voix dans les cancers du larynx ont fait l'objet de nombreuses études et les cliniciens disposent d'outils de rééducation adaptés. En revanche, les études sur les troubles de la parole après un cancer de la cavité buccale ou de l'oropharynx sont moins nombreuses et peu d'outils d'évaluation et de rééducation sont disponibles pour les cliniciens qui s'orientent alors vers des bilans composites créés par leurs soins. Notons que c'est en partie pour pallier ce manque que le projet C2SI a été élaboré.

1.2.1.1. Localisations

Dans cette thèse, nous nous intéressons aux cancers de la cavité buccale (CB) et de l'oropharynx (OP), régions anatomiques que nous décrivons ci-après.

La cavité buccale est délimitée, sur sa partie supérieure, par le palais dur qui est immobile. Sur sa partie antérieure c'est une rangée de dents supérieures (immobile) et inférieures (mobile), ainsi que les lèvres qui la circonscrivent. Sur sa partie inférieure c'est la mandibule et le plancher buccale qui la délimitent. La mandibule est mobile, elle entraîne la rangée de dents inférieures et la langue dans sa dynamique. La rangée de dents inférieures n'a de mobilité que grâce aux mouvements verticaux de la mandibule, elle y est fixée. En revanche, la mobilité de la langue est plus indépendante, elle peut s'élever dans la cavité buccale, alors que la mandibule s'abaisse. Les mouvements linguaux permettent de moduler l'espace de la cavité créée par l'abaissement de la mandibule. Les organes inclus dans la cavité buccale sont la langue (apex et dos), le plancher, la mandibule et la région rétromolaire.

Les parties arrière basses de la cavité nasale et arrière de la cavité buccale sont reliées au larynx par un tube musculo membraneux, le pharynx, il forme la cavité pharyngée. L'oropharynx représente la partie centrale de la cavité pharyngée, à l'arrière de la cavité

buccale. Il est délimité par la partie postérieure des cavités buccale et nasale jusqu'à la base de la langue, juste au-dessus de l'épiglotte. L'oropharynx comprend la base de langue, le palais mou ou voile du palais, les amygdales et la paroi postérieure de l'oropharynx. La coupe sagittale Figure 2² illustre les limites de la cavité buccale et de l'oropharynx.

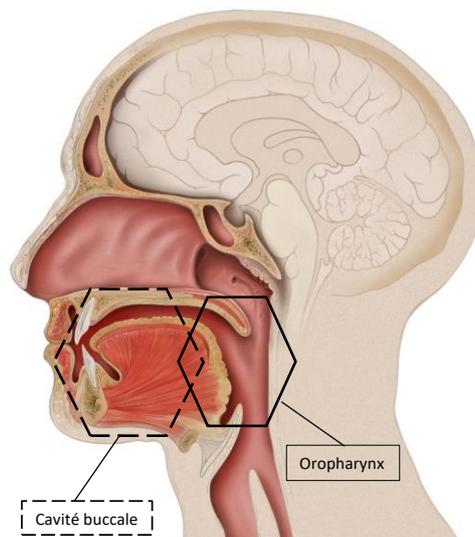


Figure 2 : Coupe sagittale des localisations cavité buccale et oropharynx

Les cancers de la cavité buccale et de l'oropharynx peuvent être de différents types histologiques, l'histologie la plus fréquente pour les cancers de la cavité buccale est le carcinome épidermoïde (HAS).

1.2.1.2. Histologie

Les cancers sont classés en fonction du type de cellules à partir de laquelle la tumeur semble s'être développée. C'est l'histologie. Dans notre corpus se trouvent des patients atteints de carcinomes épidermoïdes, histologie la plus fréquente dans les cancers de la cavité buccale. D'autres étaient atteints par des cancers développés sur les glandes salivaires accessoires présentes dans la muqueuse (carcinome adénoïde kystique par exemple).

Les organes de la cavité buccale et de l'oropharynx sont recouverts de muqueuse, c'est cette membrane qui est en contact avec le milieu extérieur : l'air respiré et les particules qu'il contient. La majorité (9/10) des cancers des VADS sont dits épidermoïdes, c'est-à-dire qu'ils naissent au sein de l'épithélium, dans une cellule du tissu qui tapisse les organes. D'autres types de carcinomes, plus rares, apparaissent au sein des glandes salivaires accessoires présentes dans la muqueuse (carcinome adénoïde kystique, adénocarcinome, etc ...) Dans le cas des cancers des VADS les affections sont localisées dans les glandes salivaires. D'après l'INCa, « L'identification du type histologique aidera

² © Patrick J. Lynch; illustrator; C. Carl Jaffe; MD; cardiologist Yale University Center for Advanced Instructional Media Medical Illustrations by Patrick Lynch, generated for multimedia teaching projects by the Yale University School of Medicine, Center for Advanced Instructional Media, 1987-2000. Patrick J. Lynch, <http://patricklynch.net> Creative Commons Attribution 2.5 License 2006; <https://creativecommons.org/licenses/by/2.5/> Les dessins et les légendes (en noir) ont été ajoutés sur la figure originale.

au choix des traitements appropriés ». Les traitements sont également orientés en fonction du stade du cancer au moyen de la classification TNM qui complète le diagnostic.

1.2.1.3. Classification des cancers

Le stade d'un cancer définit son extension, il est exprimé au travers d'une classification qui rend compte de plusieurs caractéristiques du cancer. C'est la classification TNM « Tumor, Node, Metastasis » (Tumeur, Ganglions, Métastase ». Elle permet de rendre compte de la taille de la tumeur, du degré d'envahissement éventuel des ganglions lymphatiques et de la présence, ou absence, de métastases. Ceci au travers d'échelles numériques.

Le stade de la tumeur (T) rend principalement compte de sa taille sur une échelle de 1 à 4. Une tumeur de petite taille (moins de 2 cm) représente le stade T1, avec une taille moyenne T2, elle s'étend de 2 à 4 centimètres. Les tumeurs supérieures à 4 cm appartiennent à la catégorie T3. Enfin, T4 ne représente plus la taille de la tumeur mais son extension aux structures anatomiques adjacentes. Le stade 'node' détermine l'envahissement ganglionnaire (adénopathies) au travers d'une échelle de 0 à 3 qui indique le nombre de ganglions touchés et l'extension extra-ganglionnaire. Cet envahissement ganglionnaire concerne les ganglions de drainage de la région de la tumeur. Dans le cadre des cancers des VADS ce sont tous les ganglions cervicaux. Le stade M détermine si des organes à distance de la tumeur sont envahis, au travers du terme Métastase, sur une échelle binaire 0 (absence de métastases à distance) ou 1 (présence de métastases à distance). Le terme 'à distance' fait référence à la localisation de l'envahissement des organes dans une région différente de celle de la tumeur. Dans le cadre d'une tumeur des VADS, s'il est constaté un envahissement tumoral dans la région intra-abdominale, par exemple, on parle de métastases. Lorsqu'une tumeur est repérée, l'équipe médicale recherche l'étendue du cancer et évalue la dispersion des cellules cancéreuses dans le corps du patient. Si elle relève une invasion ganglionnaire, dont l'étendue locale est définie par N, elle recherche la présence de métastases ou si d'autres structures ganglionnaires plus éloignées sont envahies (M). Pour définir le « stade » TNM d'un patient les trois composantes « tumeur », « node » et « métastase » sont à considérer. « cT2N0M0 » signifie que le patient a une tumeur T2, sans envahissement ganglionnaire régional et sans métastases à distance détectable lors du bilan d'extension pré-thérapeutique. Cette classification permet à l'équipe médicale de déterminer le protocole de soin le mieux adapté au patient en considérant son profil clinique et son environnement (soutien familial, qualité de vie, état psychologique, ...).

1.2.2. Les traitements

L'équipe médicale détermine le ou les traitements dont le patient va bénéficier relativement à la localisation de la tumeur, à son histologie et à l'évaluation des stades T, N et M. Trois traitements sont principalement administrés : la chirurgie, la radiothérapie et la chimiothérapie. Ils peuvent être prodigués seul ou combinés les uns avec les autres, dans l'ordre jugé le plus adéquat.

La chirurgie lorsqu'elle est réalisée, consiste à retirer tout ou partie de la tumeur. C'est l'exérèse de la tumeur ainsi que des ganglions de drainage. Le maximum de cellules cancéreuses visibles est retiré. Une marge d'environ 1 centimètre est prise autour de la tumeur afin de s'assurer du retrait de la totalité des cellules cancéreuses. La chirurgie peut parfois nécessiter une reconstruction en cas de tumeur volumineuse. C'est un traitement très localisé. La radiothérapie est un traitement locorégional qui consiste à traiter les organes cibles en préservant au maximum les structures anatomiques avoisinantes. Elle utilise des rayons dirigés vers ou directement implantés dans la zone à traiter. Elle vise à détruire les cellules cancéreuses localisées près de la tumeur initiale et bloque la capacité de multiplication des éventuelles cellules cancéreuses restantes. La chimiothérapie est un traitement médicamenteux qui consiste en l'administration d'un ou plusieurs médicaments, et qui permet de détruire les cellules cancéreuses et de les empêcher de se multiplier. C'est un traitement par injection qui agit par voie générale, c'est-à-dire qu'il se diffuse dans tout le corps. C'est le traitement privilégié lors de la présence de métastases.

Si un traitement par radiothérapie et / ou chimiothérapie est nécessaire ils peuvent être administrés à différents moments du parcours de soin pour servir au mieux les objectifs thérapeutiques. S'ils sont préconisés préalablement au traitement chirurgical, ils visent à limiter la croissance ou réduire la taille de la tumeur pour limiter la gêne du patient et en faciliter l'exérèse, le retrait. S'ils sont administrés après l'opération chirurgicale de la tumeur alors ils visent plutôt à détruire et neutraliser la multiplication des éventuelles cellules cancéreuses subsistantes. Ils visent aussi à prévenir le risque de récurrence. Afin de limiter les séquelles post-opératoires, et en fonction de l'étendue de la résection, il peut être préconisé au patient une reconstruction de la partie retirée dans l'intervention chirurgicale. Elle peut intervenir dans le même temps opératoire. Une reconstruction consiste à combler l'espace laissé par l'exérèse, elle se réalise généralement au travers de lambeaux prélevés sur le patient (auto-greffe). Dans d'autres cas l'espace laissé par la résection de la tumeur n'a pas été comblé.

L'équipe médicale privilégie les traitements locaux ou locorégionaux quand cela est possible, néanmoins, l'objectif de survie est primordial et nécessite le traitement le mieux adapté. Chaque cancer est différent et nécessite le traitement approprié. Quel que soit le traitement dont le patient bénéficie, les objectifs de ce parcours de soin sont multiples. La guérison est primordiale, comme le maintien des fonctions vitales. L'objectif de l'équipe médicale est de mesurer la balance bénéfique / risque entre les traitements prodigués et la qualité de vie du patient. Cela passe par la préservation du maximum de structures anatomiques pour maintenir les fonctions vitales et communicationnelles.

Dans le cadre de la prise en charge des séquelles fonctionnelles dans les cancers des VADS, il est important de considérer le délai qui sépare la rémission du patient, et donc la fin de son traitement, de sa participation à une tâche de production de la parole. En effet, un délai suffisant (6 mois) permet de s'assurer de la stabilité de sa production de parole. Si le traitement dont a bénéficié le patient impacte sa production de parole, quelques mois après la fin du traitement, ces troubles de productions devraient être stabilisés et c'est à ce moment qu'il est pertinent d'évaluer les composantes linguistiques dégradées et

préservées. En effet, malgré le soin apporté au choix des traitements, des séquelles fonctionnelles peuvent subsister. L'équipe médicale qui prend en charge le patient, qui élabore le protocole de soin le plus adapté et qui suit le patient tout au long de son parcours de soin, est composée de professionnels de santé issus de différentes disciplines. C'est la complémentarité de chaque spécialité qui permet au patient d'avoir le meilleur suivi possible.

1.2.3. Les séquelles fonctionnelles

Les cancers des VADS (CBO) et leurs traitements sont connus pour perturber l'organisation physiologique et fonctionnelle de la cavité buccale et de l'oropharynx. Ils entraînent des troubles de la déglutition et de la production de la parole. Ces pathologies ORL impactent le patient dans sa capacité à répondre de façon autonome à un besoin vital (déglutition) et dans ses habilités communicationnelles (parole), pourtant essentielles à la qualité de vie (Balaguer, Boisguérin, et al., 2019; Clarke et al., 2016; Woisard, 2011b)

Une étude de Giraud & Tessier (2018), tend à montrer les bénéfices d'un protocole de travail quotidien de la déglutition et de la parole (voix) au cours des 7 semaines de radiothérapie. Les patients pratiquent des exercices de mobilité linguale (différentes séries de mouvements), de force linguale, de déglutition et de production de séquences sonores mobilisant les structures phonatoires. Les résultats sont collectés au travers d'auto-évaluations le Deglutition Handicap Index (DHI) (Woisard et al., 2006) et le Voice Handicap Index (VHI) (Jacobson et al., 1997; Johnson et al., 1997), validés en français. Ces questionnaires sont complétés par une évaluation perceptive par l'échelle GRBAS (Hirano, 1981). Ces questionnaires d'auto-évaluation apportent un autre regard au clinicien. Ils le renseignent sur le patient à propos de ses propres perceptions sur sa qualité de vie et son handicap communicationnel. Le patient exprime ainsi ses gênes, ses difficultés et ses limites, au regard des séquelles de sa pathologie.

Ces questionnaires ne demeurent pas totalement satisfaisants puisqu'ils interviennent après le traitement. Une enquête multicentrique (Allemagne, Suisse, Autriche) basée sur des questionnaires d'auto-évaluation par le patient (Suarez-Cunqueiro et al., 2008) suggère que les patients souhaiteraient être mieux informés des risques de séquelles associées aux différents modalités et stade du cancer et à chaque traitement. Par exemple, un cancer avec une tumeur de stade avancé (T3 ou T4) implique qu'une zone relativement grande est touchée, le traitement dont bénéficie le patient est généralement corrélé avec le degré d'agressivité et de dangerosité de la tumeur. Le lien de cause à effet entre la taille de la tumeur et les séquelles fonctionnelles sur la déglutition et la parole est établi (Balaguer, Boisguérin, et al., 2019; Monestier et al., 2021; Suarez-Cunqueiro et al., 2008). Concernant les traitements, la radiothérapie est connue pour entraîner une perte de mobilité au travers de la rigidification des tissus mous. Ainsi la langue, le voile du palais, le conduit vocal lui-même sont directement impactés par la diminution de la souplesse des organes impliqués dans la déglutition et la phonation. De plus, lorsque la radiothérapie est associée à la résection chirurgicale de la tumeur les séquelles et perturbations sont d'autant plus importantes. Les bénéfices des traitements se

combinent, les impacts fonctionnels aussi. Des éléments de littérature suggèrent que certaines localisations de tumeurs sont davantage impliquées dans les séquelles fonctionnelles handicapantes de l'alimentation et la communication, comme celles localisées sur le plancher buccal ou la mandibule inférieure (Suarez-Cunqueiro et al., 2008).

La majorité des cancers de la cavité buccale et de l'oropharynx et leurs traitements impactent donc fonctionnellement la déglutition et la parole. La prise en charge de ces patients et des séquelles fonctionnelles nécessite des équipes pluridisciplinaires. La collaboration entre chaque spécialité est une des clefs majeures du parcours de soin. Ce panel de professionnels suit toutes les étapes du parcours de soin, l'établissement du diagnostic, les traitements, l'après traitement, le suivi au long court. Oncologues, médecins ORL, chirurgiens oncologues, radiothérapeutes oncologues, chimiothérapeutes oncologues, anesthésistes, infirmiers, aides-soignants et orthophonistes sont les principaux intervenants auprès du patient atteint d'un CCBO. Des professionnels de santé alliés qui peuvent également être amenés à intervenir ponctuellement : dentiste, kinésithérapeute, ergothérapeute...

La prise en charge des séquelles fonctionnelles implique particulièrement certains professionnels : les médecins ORL, phoniâtres et orthophonistes sont responsables de la réhabilitation fonctionnelle de la parole et de la déglutition. Ils réalisent des évaluations cliniques de la sensori-motricité des articulateurs (lèvres, langue, mandibule), de leur force et tonus. Ils évaluent également les mécanismes de déglutition et la qualité de la phonation. Les troubles de la déglutition et de la parole sont abordés et traités conjointement car ils impliquent les mêmes structures anatomiques. Leurs prises en charge s'apportent un bénéfice mutuel, la réhabilitation de l'un participe à l'amélioration de l'autre.

En 2016, Clarke et collègues proposent de poser les lignes directrices de la prise en charge et de la réhabilitation de la parole et de la déglutition dans les cancers de la face et du cou, au travers d'un Guideline multidisciplinaire national, dans leur articles "Speech and swallow rehabilitation in head and neck cancer: United Kingdom National Multidisciplinary Guidelines" (Clarke et al., 2016). Les auteurs apportent des recommandations et notamment³ : Tout patient atteint d'un cancer de la tête et du cou – doit être entouré par une équipe pluridisciplinaire pour suivre chaque étape du parcours de soin, - devrait bénéficier d'une évaluation de la parole et de la déglutition avant le traitement, - que la continuité du suivi et de l'apport de l'orthophoniste sont importants pour maintenir la parole et une déglutition effective et sans danger après le traitement. Ils recommandent également qu'une réduction des déficiences, un maintien de la fonction et un rétablissement plus rapide sont permis par des programmes d'exercices prophylactiques et par l'enseignement des manœuvres de déglutition tout au long du

³ "• All multidisciplinary teams should have rehabilitation patient pathways covering all stages of the patient's journey including multidisciplinary and pre-treatment clinics. • All head and neck cancer patients should have a pre-treatment assessment of speech and swallowing. • Continued speech and language therapist input is important in maintaining voice and safe and effective swallow function following head and neck cancer treatment. • A programme of prophylactic exercises and the teaching of swallowing manoeuvres can reduce impairments, maintain function and enable a speedier recovery."

parcours de soin (avant, pendant et après le traitement). Le rôle spécifique notamment de l'orthophoniste au travers de sa prise en charge est souligné. En effet, ce professionnel prend en charge les troubles de la déglutition et les troubles de la production de la parole, il participe à la réalisation d'un bilan clinique approfondi, il mesure la sensibilité et motricité des organes, la gestion du bol alimentaire et de la salive, la tonicité musculaire, il s'intéresse à la physiopathologie consécutive aux traitements du cancer. Il intervient tout au long du parcours de soin (Lombard, 2009) notamment au travers de l'évaluation de la parole, il aide le patient à mettre en place des stratégies articulatoires de compensation. Il propose des exercices rééducatifs sur le plan fonctionnel au travers de différentes techniques. En France, le Décret n°2002-721 du 2 mai 2002 relatif aux actes professionnels et à l'exercice de la profession d'orthophoniste, (2002) spécifie les missions de l'orthophoniste auprès des patients atteints de pathologies ORL et notamment la rééducation : - « des fonctions oro-faciales entraînant des troubles de l'articulation et de la parole ; - et la conservation de la voix, de la parole et du langage, la démutisation et l'apprentissage de la lecture labiale - des troubles de la déglutition... »

L'orthophoniste évalue instrumentalement et perceptivement la production de parole, la thérapie qu'il propose vise à maximiser les fonctions résiduelles de la parole après le traitement et à réhabiliter les déficits aux travers d'exercices de mobilité des articulateurs mobiles (Perry & Frowen, 2006). La participation active du patient à la thérapie est une autre clef fondamentale pour la réhabilitation de la parole chez le patient (Clarke et al., 2016).

Ci-après, un exemple d'intervention clinique auprès d'un patient après un cancer de la langue : « *Prononcez-moi le mot 'kangourou' » « Quand vous prononcez le 'k' et le 'g' ça oblige votre langue à bien reculer, et moi ça me donne un indice sur votre capacité à avaler, ou pas, correctement : Si votre langue recule mal vous allez avoir du mal à envoyer les aliments en arrière quand vous avalez » (AP-HM - Hôpitaux Universitaires de Marseille, 2013), intervention de Camille Galant, orthophoniste, auprès d'une patiente après un cancer de la langue. Et de poursuivre que le rôle de l'orthophoniste est « *d'évaluer les sons que le patient a du mal à articuler de façon à travailler et rééduquer particulièrement ces sons, et dans le même temps de regarder comment fonctionne la déglutition* ». « *La langue joue un rôle primordial dans la déglutition* » c'est donc particulièrement important d'évaluer son fonctionnement, d'autant plus chez cette patiente. En effet, une chirurgie de la langue a un impact direct sur la phase orale de la déglutition et de la parole, mais le degré d'atteinte dépend largement de l'étendue du tissu lingual réséqué (Perry & Frowen, 2006).*

Ainsi, la prise en charge des patients après traitement s'articule autour de l'évaluation et de la réhabilitation fonctionnelle des organes impliqués dans les réponses aux besoins vitaux de respiration et d'alimentation, mais également autour du maintien des fonctions de communication parlée en ciblant le niveau articulatoire. L'évaluation de l'intelligibilité constitue ici une mesure clef du déficit fonctionnel, puisque sa réhabilitation participe à l'amélioration de la qualité de vie.

Au terme de cette première partie, il apparaît que la question de l'intelligibilité et de son évaluation est d'un intérêt majeur dans le cadre clinique de la cancérologie ORL.

Cependant, il apparaît aussi qu'une mesure précise et adaptée aux besoins des patients ne peut être envisagée dans le cadre d'une définition imprécise du concept. L'évaluation de l'intelligibilité entendue dans le cadre d'un échange communicationnel implique que les processus de production et de perception de la parole sont nécessairement à considérer, et ce, conjointement. Ces processus, dont il convient de préciser le degré d'implication, sont définis en partie par des modèles construits autour d'étapes et d'unités, et sont liés par les niveaux phonologique et phonétique. Dans la partie suivante, nous tâcherons de montrer en quoi le recours aux concepts de la phonétique et de la phonologie permet de penser l'amélioration de l'évaluation segmentale chez des patients traités pour un cancer ORL.

2. Processus et concepts linguistiques

L'évaluation de l'intelligibilité, question centrale de cette thèse, ne peut être abordée ni avec simplicité, ni avec légèreté. La communication parlée, en ce qu'elle implique au moins 2 individus, tour à tour locuteurs et auditeurs, dans un contexte particulier, une relation particulière, revêt une complexité que son nom ne laisse envisager. Il ne peut être question non plus de légèreté tant l'enjeu clinique est important pour les patients à qui cet acte apparemment simple fait soudainement défaut. Les travaux en linguistique ont interrogé les mécanismes de production, de perception de la parole, de représentations phonologiques articulo-acoustique et de ces travaux ont émergé des modélisations, des théories, des concepts sur lesquels il nous semble indispensable de prendre appui pour interroger et améliorer l'évaluation de l'intelligibilité.

2.1. Production et perception de la parole

La perception est une activité biologique commune et vitale à tout organisme en contact avec son milieu. Chez l'humain, il est communément admis que 5 sens servent la perception. C'est Aristote (384-322 av JC) dans *Traité de l'Âme* (Aristote, 1846) qui caractérise les sens, aujourd'hui admis comme les 5 sens externes, que sont le toucher, la vue, l'ouïe, le goût et l'odorat. Plus récemment (XXème siècle) 4 nouveaux sens humains ont été proposés et font consensus : la proprioception, l'équilibriception, la thermoception et la nociception. Ils sont définis comme des sens internes. Il est essentiel à tout organisme biologique que ses sens perceptifs soient efficaces puisqu'il produit une réponse à ce qu'il a perçu dans son environnement. En lien avec la perception, c'est de la production de comportements adaptés que dépend sa survie. Ainsi, c'est l'harmonie entre ce qui est perçu par les sens et ce qui est produit par le corps qui détermine les capacités adaptatives d'un organisme. La production et la perception de la parole humaine, activités relevant de la communication parlée, restent un rapport entre ce qui est produit par le corps et ce qui est perçu par les sens : Dans la production de la parole c'est la proprioception qui est engagée ; dans la perception de la parole c'est l'ouïe qui est principalement mobilisée, mais aussi la vue. Cette brève introduction illustre la complexité des relations qui unissent les activités de production et de perception de la parole. Ainsi, la production de la parole est dépendante de l'intégrité du système physiologique, d'un processus d'acquisition et de l'entraînement de ces systèmes, mais aussi de la perception de la parole.

2.1.1. La production de la parole

La production de la parole humaine et la production d'un comportement biologique adapté à son environnement relèvent de domaines différents. Différents mécanismes sont à l'œuvre dans la relation biologique entre ce qui est perçu par les sens et ce qui produit

par le corps : une biche ressent un danger au travers d'une émotion élaborée à partir de ce qu'elle perçoit avec ses sens, par exemple la peur, et produit un comportement adapté à la situation, par exemple la fuite. Ce comportement relève de l'émotionnel. La production de la parole a ceci de spécifique, qu'elle relève d'une intention particulière, elle est le fruit du domaine de l'intentionnel. La production de parole se réalise au moyen du système de phonation, elle repose sur l'intention communicationnelle et l'organisation physiologique du système de production.

2.1.1.1. Physiologie production

La production de la parole repose sur une intention de communiquer un signal acoustique porteur d'information. C'est dans le cerveau que cette intention est formulée, et c'est également dans le cerveau que sont encodées l'ensemble des unités linguistiques mises en jeu dans la production de la parole. C'est aussi lui qui commande les organes moteurs impliqués dans la phonation, au travers des nerfs. La production de la parole implique plusieurs organes dans sa réalisation (Figure 3) : le cerveau, les nerfs, les muscles, les poumons, la trachée, le larynx, les cordes vocales, les cavités pharyngale, buccale et nasale, le voile du palais, les dents, les lèvres, les mandibules, ...

La production d'un signal acoustique résulte du passage de l'air dans le conduit vocal. Dans cette description nous suivrons le trajet de l'air pour visiter le système de phonation (Figure 4). Les poumons constituent la soufflerie qui impulse l'air dans ce système. L'air traverse la trachée et le larynx qui accueille les cordes vocales, ou plis vocaux. Celles-ci sont mobiles, elles peuvent être ouvertes, en abduction, pour laisser libre le passage de l'air ou fermées, en adduction, contribuant ainsi à la protection des voies respiratoires pendant la déglutition. En effet les parties supérieures des voies respiratoires et des voies digestives partagent des organes, c'est à un niveau inférieur, au carrefour de la trachée et de l'œsophage, qu'elles se divisent, pour rejoindre respectivement les poumons et l'estomac. Nous poursuivons le chemin de l'air au travers des cordes vocales qui peuvent vibrer ou non à son passage. Il continue son avancement à travers le pharynx, qui constitue le résonateur pharyngé. Il se divise en trois parties, de bas en haut, le laryngopharynx à l'arrière du larynx, l'oropharynx à l'arrière de la cavité buccale et le nasopharynx à l'arrière des fosses nasales. L'air atteint ensuite les résonateurs de la cavité buccale et éventuellement la cavité nasale dépendamment de la position du voile du palais. Il est enfin projeté à l'extérieur et forme un signal acoustique. Ce sont les modifications fines de la configuration du conduit vocal et donc de la forme et du volume des résonateurs, qui modifient les propriétés acoustiques des sons.

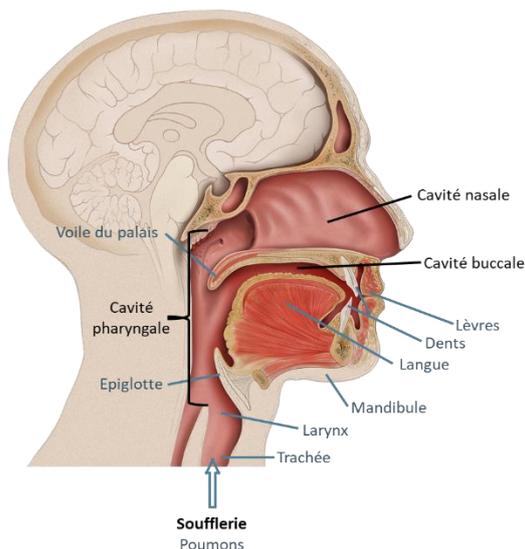


Figure 3 : Organes impliqués dans la production de parole

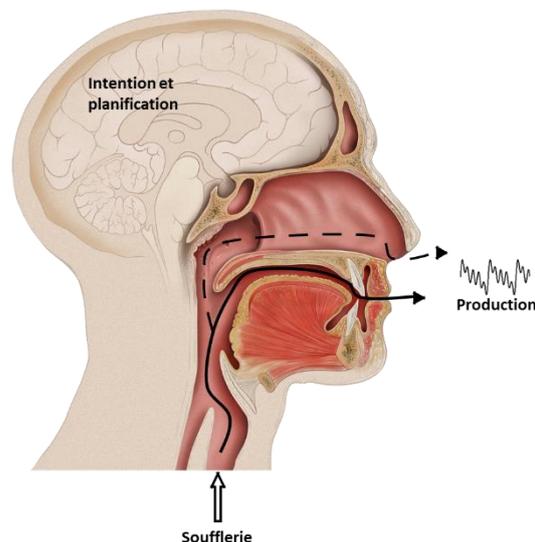


Figure 4 : Schéma du trajet de l'air à travers le conduit vocal et les résonateurs

Le modèle « source-filtre » (Fant, 1970), développé dans *Acoustic theory of speech production*, constitue le modèle de référence pour expliquer la production de la parole. C'est un modèle simple qui conceptualise le système de production acoustique de la parole comme une combinaison entre un générateur d'un son, la « source » et un ensemble de 2 à 4 résonateurs couplés, les « filtres », qui viennent moduler certaines propriétés du son source en sortie. C'est un modèle plutôt physiologique de la production de la parole : l'air impulsé dans le conduit vocal constitue la source des sons de base et les différents résonateurs qui modifient les propriétés acoustiques du son de base constituent les filtres.

L'air progresse à travers le conduit vocal et traverse différents résonateurs. Les origines latines du mot 'production' reflètent partiellement ces phénomènes : '*productio*' allongement et '*producere*' faire avancer. La production physiologique de la parole repose sur la progression de l'air dans le conduit vocal contraint par les modulations anatomiques de ce dernier. Les résonateurs du conduit vocal sont la cavité buccale, la cavité nasale et la cavité pharyngale. Ils permettent d'amplifier et de modifier le son. La modification de la forme des résonateurs est permise grâce aux articulateurs mobiles que sont la langue, la mandibule, le voile du palais et les lèvres.

La cavité buccale est délimitée par un plafond, constitué, d'avant en arrière, d'une rangée de dents, des alvéoles du palais, d'un palais dur et d'un palais mou. Sa partie inférieure est articulée et accueille une rangée de dents, la langue qui est elle-même mobile et une partie sublinguale ; cela constitue son plancher. La mandibule inférieure est un articulateur mobile osseux. Ses mouvements sont limités à la dimension verticale « haut-bas » et elle permet les mouvements d'ouverture et de fermeture (aperture) de la cavité buccale et entraîne la langue et la mâchoire inférieure dans cette dynamique. La langue est un articulateur mobile musculaire. Elle se décompose en trois parties : l'apex, le corps et la base. Elle est composée de 17 muscles qui lui permettent une grande agilité.

Elle peut effectuer des mouvements complexes dans les 3 dimensions, « avant-arrière » s'avancer et se reculer, « haut-bas » s'élever et s'abaisser, « latéralité » droite et gauche, et d'autres mouvements tels que se tourner, s'enrouler ou se replier sur elle-même. Ses mouvements verticaux sont étroitement liés à ceux de la mandibule. Les mouvements de la langue et de la mandibule peuvent être coordonnés, elles peuvent s'abaisser ou s'élever ensemble, mais aussi asymétriques, la mandibule s'abaisse et la langue s'élève dans la cavité buccale. Sur le plan phonatoire, elles permettent de moduler la forme de la cavité buccale. Pour la mandibule nous retiendrons les mouvements verticaux qui entraînent un agrandissement de la cavité buccale par l'abaissement du plancher buccal. Le degré d'ouverture de la cavité buccale permet à la langue de réaliser des mouvements plus ou moins amples et précis en fonction de la qualité consonantique ou vocalique des segments. Le point de constriction maximal entre la langue et le plafond de la cavité buccale détermine le lieu d'articulation, essentiel à la réalisation des voyelles. Les lèvres modifient également la longueur et le degré d'aperture de la cavité buccale. Ce sont des articulateurs mobiles musculaires. Au repos elles sont au contact l'une de l'autre, mais dans la production de parole elles effectuent des mouvements complexes d'arrondissement, d'étirement et d'avancement (protrusion). Ces mouvements labiaux sont caractéristiques des sons vocaliques et de certains sons consonantiques.

La cavité nasale est constituée par les fosses nasales, son volume ne peut être modulé. Seul un abaissement du voile du palais permet le passage de l'air en son sein, ce qui caractérise les sons possédant le trait de nasalité. Le palais se compose de deux parties : une grande partie avant, c'est le palais dur, qui est circonscrit par le toit osseux de la cavité buccale et une petite partie, à l'arrière, appelée palais mou ou voile du palais. Ce dernier est constitué d'une membrane fibreuse et de muscles, recouverts d'une muqueuse. C'est une cloison mobile capable de s'abaisser pour laisser passer l'air dans la cavité nasale ou s'élever pour bloquer le passage de l'air qui n'atteint plus la cavité nasale et est contraint de poursuivre son avancée dans la cavité buccale. Ce mouvement d'abaissement ou d'élévation du voile du palais est dirigé par un ensemble de muscles releveurs du voile. Ces muscles, ainsi que les muscles élévateurs du larynx entraînent aussi un mouvement vertical (allongement – raccourcissement) de la cavité pharyngale. Elle est constituée par un tube musculo-membraneux qui relie le larynx à l'arrière de la cavité buccale et au bas de la cavité nasale. Des contractions musculaires lui permettent d'effectuer des mouvements constrictifs sur le plan antéro-postérieur qui modifient son diamètre.

Les cordes vocales constituent elles aussi un articulateur mobile, sur le plan latéral ou horizontal. Ce sont les muscles intrinsèques qui sont responsables de leur mouvement d'abduction (écartement) et d'adduction (rapprochement). Lors de la respiration elles sont ouvertes et détendues au maximum. Lors de la production de la parole elles peuvent être ouvertes, pour laisser libre le passage de l'air ou fermées. La vibration des cordes vocales est créée par un jeu de pression de l'air entre l'étage sous glottique et le résonateur que constitue la cavité pharyngée. Les constriction du conduit vocal modifient les propriétés acoustiques du son qui le traverse. En position fermée, les cordes vocales entraînent l'augmentation de la pression sous-glottique. Ce sont les différentiels

de pression entre les niveaux sous et supra glottiques qui permettent la mise en vibration des cordes vocales.

Par exemple lors de la production de la consonne orale, occlusive, bilabiale, sourde [p] l'air traverse la cavité pharyngée sans faire vibrer les cordes vocales qui restent relâchées, c'est une consonne sourde. La fermeture du conduit vocal au niveau des lèvres entrave le passage de l'air, cela entraîne l'augmentation de la pression intra-orale. Le relâchement de l'occlusion bilabiale entraîne un bruit d'explosion. Les mêmes principes s'appliquent à la consonne nasale, occlusive, bilabiale, sonore [m] pour ce qui concerne l'occlusion et la bilabialité. En revanche, pour produire ce phonème sonore, les cordes vocales entrent en vibration et le flux d'air se divise pour traverser à la fois la cavité buccale et la cavité nasale, au moyen de l'abaissement du voile du palais. En revanche, pour produire la consonne orale, fricative, post-alvéolaire, sonore [ʒ], la progression de l'air dans le conduit vocal rencontre des constriction dans la cavité pharyngale, produisant un bruit de friction et les cordes vocales sont en vibration. Puis lors de son passage dans la cavité buccale le dos de la langue se rapproche du palais sur la partie post-alvéolaire produisant une nouvelle constriction qui génère une turbulence de l'air. Ce sont donc des modifications physiologiques fines qui permettent de moduler le flux d'air afin de produire des sons de parole. Ainsi, l'ensemble des unités sonores, phonèmes, peuvent être décrites sur la base d'éléments physiologiques, notamment par un faisceau de traits articulatoires. Les exemples proposés ici relèvent de productions ordinaires, voire prototypiques, c'est-à-dire ne subissant aucune perturbation. Toutefois, aux vues de la finesse des gestes articulatoires à réaliser pour produire des sons de parole coordonnés, la moindre perturbation des structures anatomiques impliquées est susceptible d'impacter la production de parole.

2.1.1.2. Perturbations de la production de la parole

Plusieurs pathologies sont reconnues pour entraîner des troubles de la production de la parole. L'origine de ces troubles permet de les distinguer. Ainsi, les pathologies de source neurologique regroupées sous l'appellation 'dysarthries' sont considérées différemment des pathologies impactant directement la physiologie du système de production de la parole.

2.1.1.3. Trouble neurologique

Les dysarthries représentent une catégorie de troubles de la production de la parole dont l'origine est neurologique. Ils peuvent être causés par une maladie neurodégénérative comme la maladie de Parkinson ou la Sclérose Latérale Amyotrophique (SLA) ou consécutives de lésions cérébrales comme un Accident Vasculaire Cérébral (AVC) ou un traumatisme crânien. Elles impactent les fonctions nerveuses et musculaires des organes impliqués dans la phonation, et peuvent perturber, entre autres, l'exécution motrice de la parole.

Les pathologies neurologiques peuvent aider à comprendre l'organisation de la production de la parole dans le cerveau. L'association entre une aire du cerveau et une fonction cognitive prend en partie racine dans la découverte de Paul Broca (1824-1880).

En 1861, après l'examen post mortem d'un cerveau de patient atteint de son vivant de trouble de la parole, Broca localise une lésion sur la partie arrière et inférieure du lobe frontal de l'hémisphère gauche. Il en déduit que cette zone joue un rôle spécifique dans la production de la parole. Une aphasie de Broca (motrice) est alors caractérisée par une réduction quantitative et qualitative des capacités langagières du patient, mais ses capacités de perception et de compréhension ne sont pas atteintes. Même si ces découvertes s'inscrivent dans la théorie localisationniste aujourd'hui rediscutée, elles ont constitué un apport majeur pour la compréhension des mécanismes langagiers. Par ailleurs, concernant leur prise en charge, on ne note pas de grande différence concernant l'évaluation de l'intelligibilité, avec les troubles dont l'origine est physiologique.

2.1.1.4. Trouble physiologique

Un trouble physiologique de la production de la parole résulte de perturbations anatomiques des organes impliqués dans sa production. Des pathologies dites de naissance, comme par exemple les fentes labioalvéolopalatovélaires (ou fentes palatines), sont des malformations anatomiques congénitales de la structure du palais. Ce dernier constitue la séparation entre la cavité buccale (plafond) et la cavité nasale (plancher). Ces malformations entraînent des perturbations dans les mécanismes de la déglutition et de production de parole. Les résonateurs nasal et oral communiquent entre eux, la structure, la forme, la mobilité du voile du palais peuvent être impactées. Le traitement privilégié de ces pathologies est la chirurgie reconstructrice. Néanmoins, une des séquelles de cette pathologie reste l'insuffisance vélopharyngée (Vazquez et al., 2007). Le velum ou voile du palais ne se ferme pas complètement, lors de la production de parole le flux d'air pénètre dans la cavité nasale. La voix s'en trouve nasalisée. Cette nasalité permanente impacte la production de la parole.

Nous l'avons souligné dans la partie 1.2.3 Les séquelles fonctionnelles, les cancers de la cavité buccale ou de l'oropharynx sont susceptibles de perturber la production de la parole, de par leur localisation dans le conduit vocal et/ou consécutivement aux traitements dont les patients ont bénéficié. Ces cancers et leurs traitements perturbent le fonctionnement et l'intégrité des structures anatomiques impliquées dans la production de parole. En ce sens, c'est le niveau articulatoire qui est primitivement impacté.

Pour satisfaire l'enjeu communicationnel, l'intégrité du système de production constitue un prérequis. En effet, les perturbations de ce dernier sont susceptibles d'impacter la production de la parole de façon défavorable. Ainsi, les sources de perturbations de la communication sur le versant perceptif ont deux origines potentielles : d'une part la production d'un signal acoustique altéré, et d'autre part un système physiologique de perception défaillant.

2.1.2. La perception de la parole

Les mécanismes de perception de la parole ont réinvesti les sens perceptifs à leur dessein. La perception mobilise en effet des sens biologiques, l'ouïe et parfois la vue, mais repose sur des facultés d'interprétation supérieures. Les mots latins à l'origine du sens des mots percevoir, '*percipere*' prendre ensemble, récolter et perception, '*perceptio*' action

de saisir par l'esprit, connaissance, peuvent également être mis en perspective avec la perception de la parole. En effet, on y récolte un ensemble d'informations de différentes natures que l'on saisit par l'esprit et interprète par la connaissance linguistique. Nous verrons en quoi ce phénomène d'intégration d'informations peut tantôt constituer un avantage ou un inconvénient dans l'échange communicationnel.

2.1.2.1. Physiologie perception

Le signal acoustique entre dans l'oreille externe, fait vibrer le tympan qui actionne le marteau, l'enclume et l'étrier. Ce dernier transmet un message de nature mécanique à l'oreille interne, par le vestibule de la cochlée. La vibration ainsi transmise continue son avancée jusqu'à atteindre le limaçon de la cochlée. Dans celui-ci se trouve la lymphe, un liquide dans lequel baignent les cellules cillées. L'onde de vibration se propage à travers la lymphe et vient stimuler les cellules cillées. Ces dernières sont chargées de transformer le stimulus mécanique en impulsion électrique et de la transmettre au nerf auditif. Celui-ci est relié au cerveau qui décode ces impulsions électriques. Ainsi, la perception de la parole repose sur plusieurs organes physiologiques, les oreilles, la cochlée, les nerfs et le cerveau (Figure 5⁴).

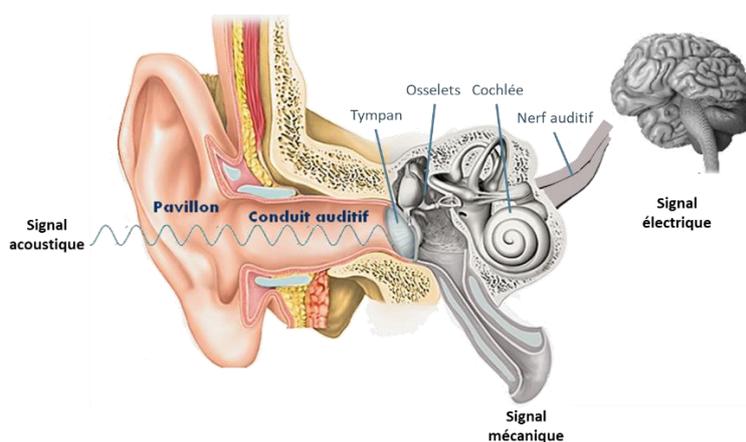


Figure 5 : Anatomie et physiologie de l'oreille

2.1.2.2. Perturbations de la perception

L'enjeu communicationnel implique également la perception d'un auditeur. Contrairement aux pathologies impactant la production de la parole il n'existe pas de terme générique pour définir les pathologies qui perturbent spécifiquement et uniquement la perception de la parole et qui auraient une origine neurologique. L'intégrité du système neurophysiologique de celui-ci est nécessaire au succès de l'échange.

2.1.2.3. Trouble neurologique

Les capacités de perception et de compréhension d'un message linguistique peuvent être perturbées par des troubles de la perception de la parole d'origine neurologique.

⁴ © Dessin Michel Saemann - Archives Larousse, les éléments de description ont été ajoutés sur la figure par nos soins.

C'est le cas notamment dans certaines aphasies. Par exemple, une aphasie de Wernicke se caractérise par des déficits au niveau de la production du patient, qui se manifestent par une perte de contrôle (anosognosie, logorrhée, jargonophasie...), le débit de parole peut être fluide mais incompréhensible. Mais aussi par des déficits majeurs de la compréhension du langage, cette zone serait donc plutôt impliquée dans la perception des mots et des symboles linguistiques. C'est Carl Wernicke (1848-1905) qui identifie que la lésion cérébrale de son patient est située à l'intersection entre le lobe temporal et le lobe pariétal de l'hémisphère gauche. Cette région porte désormais le nom d'aire de Wernicke' et qualifie également l'aphasie qui s'y localise, aphasie de Wernicke (sensorielle). L'atteinte de la structure complexe que constitue le cerveau est susceptible d'avoir des conséquences sur la perception d'un signal acoustique de parole, en termes de traitement de l'information. Le système physiologique est cependant le point d'entrée du signal acoustique.

2.1.2.4. Trouble physiologique

La baisse ou la perte de l'audition peut être handicapante dans la vie quotidienne, elle est susceptible d'impacter l'échange communicationnel parlé en perturbant la perception de la parole. L'Assurance Maladie propose une catégorisation des niveaux de surdité, elle utilise une représentation basée sur la perception de la parole pour illustrer la gêne occasionnée. Ainsi, dans une surdité 'légère', représentant une perte auditive de 20 à 39 décibels, « la personne fait répéter son interlocuteur dès la perte de 30dB, sur les sons aigus ». Le seuil de 40 décibels représente « le premier niveau majeur de handicap ». Une perte auditive de 40 à 69dB caractérise une surdité 'moyenne', fourchette dans laquelle « la personne ne comprend que si son interlocuteur élève la voix, elle est « dure d'oreille » ». Entre -70 et -89dB la surdité est sévère, « la gêne quotidienne est majeure » sur le plan communicationnel. En effet, l'auditeur perçoit un signal acoustique dégradé, dans lequel il ne peut plus percevoir certaines fréquences constitutives des sons de la parole. La surdité est dite profonde au-delà de 90dB de perte auditive et « la personne n'entend plus du tout la parole ». L'impact communicationnel de la surdité est relatif au degré de dégradation de la perception du signal acoustique.

La surdité peut se manifester dès la naissance, être consécutive d'un traumatisme, ou plus simplement être une conséquence du vieillissement naturel du corps au cours de la vie. La surdité congénitale, généralement d'origine génétique, résulte d'une malformation des structures anatomiques de l'oreille moyenne ou interne. Elle peut toucher les osselets, mais aussi tout ou partie de la cochlée. Elle peut causer des retards dans l'acquisition du langage car la perception de la parole y tient un rôle majeur. Les auditeurs atteints de surdité congénitale totale ou partielle peuvent bénéficier de traitements chirurgicaux (implant cochléaire) et de suivi orthophonique. Dans ce parcours clinique la mesure de l'intelligibilité est utilisée pour mesurer l'acuité auditive du patient au travers de son appareillage mais aussi pour mesurer sa production (Laccourreya et al., 2015).

Le tympan est une membrane souple qui reçoit le signal acoustique aérodynamique et le transforme en signal mécanique au travers de sa mise en vibration. S'il est endommagé, par une perforation par exemple, la transmission du signal depuis l'oreille externe vers l'oreille moyenne est bloquée. D'autres traumatismes peuvent impacter les structures de

l'oreille essentielles à la perception de la parole. Une exposition prolongée à un haut niveau sonore ou à un bruit bref mais intense peuvent entraîner des lésions de l'oreille interne. Les cellules ciliées de la cochlée sont fragiles, une fois détériorées elles ne se régénèrent pas. Elles ne peuvent plus assurer leurs fonctions de transformation du signal mécanique en signal électrique et de transmission de ce signal aux nerfs auditifs. Cela entraîne une baisse de l'acuité auditive. La baisse d'acuité auditive consécutive au vieillissement résulte de la dégradation des cellules ciliées ou d'une usure des osselets qui sont fragilisés et moins performants. La transmission du signal acoustique au cerveau est perturbée par l'usure de sa partie mécanique, le système de perception de la parole est dégradé. Les capacités à maintenir un échange communicationnel efficace sont fortement impactées. L'atteinte de l'une des structures de l'oreille est susceptible d'entraîner une perte de l'acuité auditive et en conséquence d'impacter la perception de la parole. L'intégrité du système perceptif est donc, tout comme celle du système de production, essentielle à la communication.

2.1.3. Production et perception de la parole

Produire et percevoir de la parole impliquent les mêmes organes, mobilisés à des degrés différents selon que l'on se place sur le versant de la production ou sur le versant de la perception. Ces deux compétences relèvent d'un processus d'acquisition, et sont étroitement liées puisqu'elles sont essentielles à la communication parlée et sont pratiquées simultanément par des individus engagés dans un échange communicationnel. Chacune est essentielle et bénéfique à l'autre : percevoir est essentiel à l'acquisition de la production et produire est essentiel pour être perçu.

2.1.3.1. Un seul cerveau

Tout d'abord, la production de la parole repose sur l'intention de communiquer. C'est dans le cerveau que naît cette intention. Le cerveau est un organe unique et complexe, divisé en deux hémisphères reliés par fibres nerveuses.

Les modèles neurophysiologiques (Ackermann, 2008; Guenther, 2006) de la production et de la perception de la parole postulent que différentes zones du cerveau seraient dédiées à une tâche plutôt spécifique. Ils mentionnent l'implication du réseau « fronto-temporo-pariétal » gauche pour ces deux activités. Il est communément admis que le lobe frontal, accueille l'aire motrice de la parole, ou aire de Broca, et l'aire motrice supplémentaire impliquées dans le contrôle de la production volontaire. L'aire de Wernicke qui contrôle la perception des mots et des symboles du langage réside dans le lobe temporal gauche. Le cortex temporal est impliqué dans l'attribution de la signification des sons. Il accueille également le cortex auditif primaire, contrôle du feedback et des inputs auditifs, et associatif impliqués dans la construction sonore des mots. Enfin, le lobe pariétal permet d'intégrer des informations sensorimotrices et visuospatiales au travers de l'implication du cortex somato-sensoriel associé à la proprioception. Ce lobe est également impliqué dans la planification du mouvement via le cortex moteur primaire. Dans le cerveau ce sont des régions communes qui servent à la

fois à la perception de la parole (décodage des sons, accès lexical, compréhension) et à la production (planification, encodage phonologique, encodage de la motricité).

Les modèles neurobiologiques de la production et de la perception de la parole, basés sur des arguments de neuroimagerie (Golfinopoulos et al., 2010; Price, 2012), montrent une modulation de l'activité du cortex auditif lors de la production de la parole par rapport à la perception de la parole (traitement de feedback auditif vs simple entrée auditive extérieure). Cela permet au locuteur d'opérer une comparaison entre les séquences réellement produites et celles prédites. Ils montrent également une activation du cortex moteur primaire lors de la perception de la parole. L'activation du cortex auditif est moins forte quand on produit que quand on perçoit, de même que l'activation du cortex moteur primaire est moins forte quand on perçoit que quand on produit. Néanmoins, l'activation conjointe des cortex moteur et auditif lors de la production et de la perception de la parole montre que ces mêmes zones participent à ces deux activités distinctes, mais intrinsèquement liées.

Le cerveau est un organe complexe dont on ne sait finalement peu de chose qui fasse réellement consensus au regard de ces capacités et facultés. Il n'est pas l'objet d'intérêt de ces travaux, mais cette brève présentation nous permet de mieux saisir les enjeux et la complexité de la relation neurofonctionnellement attestée entre les mécanismes de production et de perception de la parole, mécanismes nécessairement impliqués dans l'évaluation de l'intelligibilité. En effet, produire et percevoir de la parole sont des activités complexes qui impliquent d'encoder et de décoder plusieurs niveaux linguistiques contenus dans un signal acoustique. Les auteurs ont proposé différentes théories pour décrire les processus d'encodage et de décodage des unités linguistiques en jeu dans la production et la perception de la parole. Nous aborderons ces questions relativement à leurs aspects fonctionnels. L'intégrité des systèmes de production et de perception de la parole, et des organes sensori-moteurs impliqués, constitue le socle de la communication parlée.

La relation entre la production et la perception de la parole est étroite. Dans l'acte de la communication parlée, les places de chacun évoluent au cours de l'échange. Tantôt locuteur, tantôt auditeur, dans l'échange communicationnel émetteur et récepteur sont mêmes. Les individus sont simultanément impliqués dans des actes de production et de perception. Ainsi, comme évoqué précédemment, il faut produire de la parole pour être perçu. La production de la parole repose sur la conceptualisation et l'encodage de plusieurs niveaux linguistiques, matérialisés par un signal acoustique de parole. Les étapes qui constituent cette phase de production ont fait l'objet de plusieurs théories, et notamment de modèles qui tentent de représenter ces différents niveaux et leur enchaînement. Les modèles théoriques de l'organisation de la production de la parole sont nombreux, nous en présenterons un échantillon pertinent pour servir notre propos.

2.1.3.2. Modèles de production

Le premier modèle que nous abordons est celui de Levelt (1990). Dans son ouvrage *Speaking : from intention to articulation*, il propose un modèle de production de la parole,

qui, comme le titre de l'ouvrage l'indique, débute à l'intention de produire un message communicationnel à l'articulation de ce dernier. Il propose un modèle sériel dont la progression est ordonnée de façon « intuitivement naturelle ». Il peut être résumé comme suit « conceptualization > message > sentence form > articulation > monitoring and control ». On y trouve donc une phase de conceptualisation de l'intention de communiquer, suivie d'une phase d'élaboration du contenu de message communicationnel. S'en suit la préparation de la forme de la phrase et enfin la phase d'articulation du message conceptualisé. Après l'étape articulatoire, le locuteur opère un contrôle de l'adéquation entre la phase conceptuelle et la phase de production, c'est une autosurveillance qui permet au locuteur de corriger sa production ou son message si ceux-ci ne correspondent pas à son intention. C'est l'étape articulatoire qui nous intéresse ici, en effet, elle est essentielle puisque c'est dans celle-ci que le message à transmettre est mis en son. C'est dans cette phase que le signal acoustique est produit et transmis.

Quelque temps plus tard, Levelt et al., (1999) affinent et étoffent le modèle de Levelt 1990 en proposant un modèle d'accès lexical de la production de la parole basé sur les différents niveaux d'encodage linguistique nécessaires à la production d'un message à but communicationnel. Ils conservent la phase de conceptualisation du modèle de 1990 et ajoutent que c'est dans celle-ci que sont sélectionnés les concepts lexicaux. S'en suit une étape de sélection lexicale ou 'lemma'. Ensuite ils proposent des étapes d'encodage plus fines et plus élaborées que dans le modèle précédent. Notamment l'ajout d'une étape d'encodage morphologique (i.e : encodage des morphèmes les plus petites unités de sens) des items lexicaux sélectionnés. Ensuite une phase d'encodage phonologique dans laquelle le locuteur prépare le mot phonologique le plus approprié s'opère. Enfin l'encodage phonétique des sons à produire, représentés comme une partition de gestes articulatoires, précède l'articulation qui produit le signal acoustique. En parallèle le locuteur contrôle et surveille si sa production correspond en tout point à son intention. Cette version révisée du modèle est notamment étoffée par l'introduction de niveaux d'encodage phonologique et phonétique des unités linguistiques produites. Cette étape correspond à la sélection des unités phonologiques et à leur encodage sur le plan articulatoire. La notion de 'geste articulatoire', introduite par Browman & Goldstein (1986), représentée comme une partition, est essentielle à la réalisation de la phase articulatoire. Ces modèles ont en commun une phase de « self-monitoring » (ou 'autosurveillance') dans laquelle le locuteur opère un contrôle, ou un rétrocontrôle, sur sa propre production. Cette autosurveillance du locuteur sur sa production souligne davantage l'étroitesse de la relation entre la production et la perception. Le locuteur est, pour partie, également le destinataire de son propre signal acoustique : il perçoit sa propre production. Ces modèles, initialement prévus pour expliquer les mécanismes de production, font émerger l'interdépendance des mécanismes de production et de perception, ainsi que celle des niveaux phonologique (représentation, conceptualisation) et phonétique (articulation).

Nous tenons à préciser qu'un modèle est une illustration simplifiée qui tente de rendre compte d'un processus complexe, qui revêt parfois un caractère prédictif. Ces modèles tentent donc de représenter le processus complexe de production de la parole.

Néanmoins, ils occultent plusieurs aspects de la production de la parole, et notamment sa grande variabilité. En effet, un même message linguistique peut être porté par différentes réalisations, dépendantes du contexte, de la situation dans laquelle l'énoncé est produit, du locuteur, de son état émotionnel, de l'intégrité de son système de production, ... Pourtant la variabilité est omniprésente, tant entre les locuteurs (variabilité inter-individuelle), que chez un même locuteur (variabilité intra-individuelle).

2.1.3.3. Variabilité contextuelle

Le contexte de l'échange communicationnel dans lequel le message est produit influence la qualité de production du signal acoustique. Pour illustrer une adaptation articulatoire de la production de parole, qui selon nous souligne sa variabilité intra-individuelle, Lindblom (1990), propose le modèle « Hyper et Hypo Speech » (H&H) dans lequel le contexte interactionnel influence la production du signal acoustique. Le locuteur produit la juste quantité nécessaire d'information acoustique, quand il sait l'information préalablement partagée, le message serait hypo-articulé ; quand le message partagé contient de nouvelles informations, il postule que le message serait hyper-articulé. En s'appuyant sur les connaissances supposément partagées par l'interlocuteur, le locuteur adapte donc sa production du signal acoustique au niveau articulatoire. L'encodage des unités de bas niveau est influencé par les processus cognitifs de plus haut niveau. De même, lors d'un échange communicationnel le locuteur modifie sa production en fonction de l'environnement et du contexte situationnel de l'échange.

Dans le sens de la théorie H&H, quand un locuteur produit de la parole dans un contexte bruité, il modifie les caractéristiques acoustiques de son signal de parole. La production du locuteur est hyper-articulée lorsqu'elle est produite dans un environnement bruité (Garnier et al., 2006). Garnier et collègues postulent que ces adaptations articulatoires ont plutôt pour objectif de faire émerger des caractéristiques spécifiques (spectrales) du signal acoustique « plutôt qu'à augmenter prioritairement le rapport signal sur bruit », soit augmenter le volume de la production. C'est une théorie intéressante puisque dans un environnement bruité certaines caractéristiques acoustiques de la parole sont masquées, occultées, par ce bruit. Ce phénomène perturbe la communication parlée. En réponse à cette perturbation, le locuteur compense en modifiant l'articulation des sons. Le locuteur, également auditeur de sa propre production, considère les composantes acoustiques dégradées par l'environnement et adapte sa production pour compenser ces dégradations. A nouveau, ce sont les informations de haut niveau (« perçues par les sens », contexte) qui influencent la matérialisation du signal acoustique au bas niveau, au travers des caractéristiques acoustiques des sons produits (« par le corps »).

Comme nous l'avons évoqué les informations de haut niveau interactionnelles (connaissances partagées) et contextuelles (environnement) influencent l'encodage et la réalisation des sons de parole au niveau articulatoire. Bas et haut niveaux, production des sons de parole et connaissances du locuteur, entretiennent ensemble une relation d'interdépendance. Cela souligne d'avantage la relation étroite qu'entretiennent la production 'par le corps' et la perception 'par les sens', mais également l'adaptabilité de la production de la parole. Celle-ci se manifeste au travers de la variabilité de la

production. En effet, la variabilité est adaptative, elle considère un ensemble de circonstances qui modifient la forme du signal acoustique pour servir la communication.

2.1.3.4. Variabilité physiologique

D'autres modalités introduisent de la variabilité dans la production de la parole. Ce sont des caractéristiques individuelles qui conditionnent la production. Elles sont dépendantes de l'individu en tant qu'être humain unique. Elles sont de haut niveau comme l'état émotionnel et de bas niveau comme des perturbations physiologiques des organes de la production de la parole.

Ainsi, l'état émotionnel du locuteur est également un exemple de variation intra-individuelle dans la production de la parole. Une étude portant sur la caractérisation de la parole de locuteur atteint de dépression (Cummins et al., 2015) a montré que la parole de ces locuteurs est « moins animée », qu'elle présente une réduction du nombre de sons produits. On note également une réduction de la variabilité spectrale d'amplitude équivalente à celle observée chez des locuteurs dysarthriques. Un trouble émotionnel peut donc entraîner des conséquences sur la production de parole qui suivent des tendances observées chez des locuteurs atteints de trouble de la production de parole pathologique d'origine neurologique.

La production de la parole se réalise au travers du système physiologique impliqué dans cet acte. Une altération fonctionnelle du système de production de la parole introduit également une grande variabilité de production puisque la dynamique du geste articulatoire est perturbée. Les cancers de la cavité buccale et de l'oropharynx, et leurs traitements, perturbent le fonctionnement du système phonatoire. Ils impactent directement les structures anatomiques impliquées dans la production du signal acoustique. Ce sont les aspects articulatoires qui sont impactés dans ces pathologies. La production est perturbée, les sons sont dégradés.

La variabilité de la production de la parole peut aussi être consécutive, de l'état psycho-émotionnel et/ou physico-fonctionnel du locuteur qui influencent les dimensions acoustiques des sons produits. L'état psychique du locuteur est susceptible d'impacter la qualité des sons de parole produits dans ses aspects spectraux en termes d'intensité. L'altération physiologique du système de production impacte directement l'articulation des sons de parole, le signal acoustique est dégradé. Les configurations articulatoires en jeu dans la production ne sont plus intègres.

Nous précisons qu'au regard de la grande variabilité intra-individuelle de la production de la parole, il apparaît naturel que la variabilité inter-individuelle soit d'autant plus importante. Nous ne l'aborderons pas ici car la variabilité intra-individuelle, à elle seule, nous permet d'entrevoir les enjeux que la variabilité de la parole soutient. Nous retiendrons : qu'elle résulte de l'adaptation du locuteur au contexte ou qu'elle soit consécutive d'une pathologie, la variabilité de la production de la parole sert un but communicationnel. Elle vise à satisfaire les enjeux de la communication parlée et est partagée par l'interlocuteur, également engagé dans l'acte communicationnel.

La perception de la parole repose sur l'intégrité du système perceptif. Il est essentiel de comprendre le fonctionnement de ce dernier puisqu'il conditionne la façon dont nous

percevons et appréhendons les phénomènes linguistiques. Une perturbation dans son fonctionnement est susceptible d'altérer la qualité de réception du message et donc d'impacter la façon dont *on saisit par l'esprit et dont on interprète par la connaissance linguistique*, dont on perçoit la parole. Engagés à parts égales dans l'échange communicationnel, locuteur et auditeur échangent ensemble. Nous avons laissé apercevoir la complexité de la tâche de l'auditeur au travers de l'autosurveillance du locuteur sur sa production, qui fait également de lui le récepteur de son signal acoustique. Sa mission : l'auditeur doit décoder la variabilité inhérente aux signaux acoustiques perçus. C'est aussi de lui que dépend la réussite de l'échange communicationnel. Nous rappelons ici qu'*un même message linguistique peut être porté par différentes réalisations, dépendantes du contexte, de la situation dans laquelle l'énoncé est produit, du locuteur, de son état émotionnel, de l'intégrité de son système productif...* Chacun de ces aspects concerne également l'auditeur, en effet, au travers de son système perceptif, l'auditeur tient compte de la situation dans laquelle le message est émis, il intègre la variabilité. En plus des informations portées par le signal acoustique, il considère un ensemble d'informations de haut niveau interactionnel et contextuel, toujours en visant la réussite de l'échange.

2.1.3.5. Modèles de perception

Nous présenterons ici les différentes hypothèses émises concernant les processus de traitement de la parole sur le plan fonctionnel. Les études sur la perception de la parole se sont notamment questionnées sur les unités perceptives de base, en d'autres termes quelles sont les unités linguistiques prédominantes sur lesquelles s'appuie un auditeur pour percevoir et décoder un signal acoustique de parole.

Les premières études sur l'identification phonémique se sont questionnées sur la façon dont le signal perçu est mis en correspondance avec une forme stockée en mémoire et sous quelle forme il est encodé pour être stocké. Liberman et collègues, s'intéressent principalement à la relation entre le signal acoustique et le niveau de perception phonémique. Ils postulent que la production et la perception de la parole sont étroitement liées, puisqu'elles « partagent le même ensemble d'invariants ». Ils développent la Théorie motrice de la perception de la parole (A. M. Liberman et al., 1967; A. Liberman & Mattingly, 1985) dans laquelle ce qui est perçu par l'auditeur est représenté par une succession de gestes phonétiques réalisés intentionnellement par le locuteur. Ces gestes sont « représentés dans le cerveau comme des commandes motrices invariantes qui appellent des mouvements des articulateurs à travers certaines configurations linguistiquement significatives ». En ce sens, les arguments de neuro-imagerie qui montrent une activation du cortex moteur lors de la perception de la parole soutiennent cette idée. Liberman propose alors que la perception des consonnes et des voyelles fasse intervenir différents modes de traitement. Selon cette théorie, les consonnes auraient un mode de perception catégorielle. Les frontières entre consonnes sont bien établies puisqu'elles impliquent différentes configurations articulatoires effectives pour passer de l'une à l'autre. En revanche, les voyelles seraient perçues comme un continuum, puisqu'elles peuvent être produites les unes à la suite des autres en modifiant progressivement la forme du conduit

vocal et la configuration des articulateurs. Selon Liberman (Féry, 1999), si les sons de parole montrent de la variabilité acoustique, alors l'invariabilité de la parole réside dans ses aspects articulatoires. C'est cette pensée qui sous-tend la notion de gestes articulatoires.

Les travaux de Liberman et collègues concernant les consonnes ont porté sur les consonnes occlusives. Mais d'autres études ne corroborent pas complètement l'effet de perception catégorielle pour les consonnes en incluant des consonnes fricatives et liquides dans leur expérimentation (Fujisaki & Kawashima, 1970). Toutefois, les protocoles expérimentaux employés demandent aux auditeurs d'effectuer un choix parmi un petit nombre de réponses prédéfinies. Ce choix expérimental pourrait, en partie, induire cet effet de perception dite catégorielle. Ainsi, l'effet de perception dite catégorielle serait partiellement imputable au design expérimental qui contraint l'auditeur à faire un choix parmi un set de réponses discrètes préétablies. Il est alors postulé que c'est la réponse de l'auditeur au travers de ce choix forcé qui est catégorielle et non la perception elle-même.

Sur la base de cette distinction, Massaro & Cohen (1983), proposent un nouveau protocole expérimental dans lequel les auditeurs doivent porter un jugement continu, plutôt que discret, à l'aide d'une échelle numérique. Les résultats montrent un faisceau de preuves convergentes en faveur de la perception continue, en opposition à la perception strictement catégorielle. Massaro et collègues, proposent alors que l'auditeur emploie la stratégie perceptive la plus avantageuse dépendamment des circonstances et des informations dont il dispose. Ainsi, en contexte expérimental l'auditeur peut employer une stratégie perceptive catégorielle (s'il y est contraint) alors que lors de la perception de la parole les informations perçues et utilisées par les auditeurs « conservent dans tous les cas un caractère continu et graduel » (Nguyen, 2005). De plus, Nguyen (2005) suggère qu'« il n'est pas exclu de penser que l'auditeur puisse procéder à un double codage de l'information, continue (sous la forme d'une sorte de spectrogramme auditif par exemple) et discret (sous forme d'un ensemble d'étiquettes phonémiques) ». Dans ce sillage, on pourrait envisager que la perception dite catégorielle soit une perception plutôt prototypique du segment cible, alors que la perception dite continue pourrait être associée à une liste d'exemplaires. C'est-à-dire que la forme de base du phonème détermine sa catégorie, et que ses différentes réalisations en contexte constituent une liste d'occurrences mobilisables dans la perception.

Ces études s'interrogent donc sur la nature, la forme des unités phonémiques de la perception. Néanmoins, que l'identification phonémique soit réalisée au travers d'une perception dite catégorielle, ou dite continue, elle relève de la compréhension fine du fonctionnement du système perceptif. En revanche, quelle que soit la stratégie perceptive prédominante dans la perception phonémique, c'est une stratégie d'efficacité qui est privilégiée. L'efficacité et l'intégrité du système perceptif conditionnent la réussite de l'échange communicationnel. Les auditeurs se montrent capables de décoder un signal acoustique de parole à partir de caractéristiques acoustiques spécifiques, mais ils sont également capables de décoder un signal altéré, dans lequel certaines caractéristiques acoustiques sont susceptibles de ne pas être 'conformes' à ce qui est attendu. Cette

adaptabilité de la perception de la parole fait partie intégrante de la gestion de la variabilité de la parole au moyen de l'intégration de plusieurs niveaux d'information.

Dans la théorie motrice de la perception de la parole le son n'est pas le seul objet de la perception. Il constitue le résultat du geste articulatoire qui fournit également le support pour la perception dudit geste articulatoire, base des catégories phonétiques. Ce geste est produit par le locuteur et saisi par l'auditeur au moyen de sa réalisation, le son. Le son ne constitue pas le seul objet de la perception, ce qui souligne davantage encore la complexité qui réside entre les versants de production et de perception de la parole. Elles partagent un large panel d'unités interconnectées. En ce sens, la perception de la parole implique la production à certains niveaux. Les études de neurosciences et imageries (2.1.3.1) confirment le lien indissociable qui les unit, et de leur influence mutuelle. L'activation des régions dédiées à la production lors de la perception de la parole, et réciproquement, l'activation des régions dédiées à la perception lors de la production ont cours de façon intra-individuelle, le locuteur opère un contrôle perceptif de sa production, et inter-individuelle, l'auditeur perçoit la production. Ces relations interdépendantes sont réinvesties par (Schwartz et al., 2012) dans un modèle qui intègre le contrôle moteur dans la perception de parole. « The Perception for Action Control Theory » (PACT) considère que la perception de la parole repose sur l'ensemble des mécanismes qui permettent de contrôler la parole sur les plans moteur, cognitif et sensoriel. Les unités de parole sont caractérisées à la fois par « une cohérence motrice et perceptive » (Schwartz et al., 2012, p. 18). Dans cette théorie, l'unité de communication est un « geste perceptuellement façonné » représentant une unité perceptuo-motrice. Elle suppose que la perception implique la connaissance de l'action, et que l'action vise des gestes façonnés par la perception.

Ohala (1970) a montré que des contraintes mécaniques du système de production de parole sont susceptibles de produire des perturbations non intentionnelles du signal acoustique et pouvaient aboutir à un décodage erroné ('mis-perception') de la part de l'auditeur. En ce sens, l'échange à but communicationnel est un compromis satisfaisant les contraintes articulatoires et les contraintes perceptives. De plus, la production de parole par un locuteur « est le produit combiné de son état organique et de tout ajustement phonétique de l'appareil vocal⁵ » (Beck, 2005). Cette affirmation de Beck, (2005), est d'autant plus vraie dans le cas de la parole pathologique consécutive à un traumatisme. Par exemple, les cancers de la cavité buccale et de l'oropharynx et leurs traitements, peuvent entraîner des séquelles fonctionnelles qui désorganisent 'l'état organique', soit le système physiologique, et qui impactent directement la production de parole en perturbant les 'ajustements' de l'appareil phonatoire. Ces perturbations produisent de la variabilité dans la production de la parole, l'adaptabilité de la perception de la parole permet de l'intégrer. Les systèmes de production et de perception sont hautement adaptatifs.

⁵ "is the combined product of his or her organic state and of any phonetic adjustments to the vocal apparatus"

2.1.3.6. Intégration perceptive

La perception de la parole est malléable, l'auditeur présente la faculté d'adapter sa perception de la parole en intégrant autant d'informations que nécessaire pour décoder le signal acoustique. La relation qu'il entretient avec son interlocuteur influence sa perception et son interprétation du message. Il est communément admis que l'on perçoit mieux quelqu'un qu'on connaît ou que l'on a déjà entendu auparavant. Au-delà des considérations que l'auditeur porte aux connaissances partagées avec son interlocuteur, l'environnement dans lequel le signal linguistique est transmis est susceptible d'impacter la perception de l'auditeur. Dans un échange communicationnel en milieu bruité le locuteur fait émerger des caractéristiques spectrales particulières à destination de l'auditeur. Ce dernier considère ces spécificités articulatoires dans sa perception du signal acoustique. Néanmoins, le locuteur ne peut pas compenser totalement la dégradation du signal au travers du bruit, le signal acoustique perçu demeure donc dégradé. L'auditeur participe alors à la compensation puisqu'il a la faculté de restaurer partiellement ce signal acoustique.

Le phénomène de restauration est une habileté de l'auditeur qui, à partir de ses connaissances linguistiques de haut niveau, se montre capable de reconstruire les séquences sonores dégradées dans un signal acoustique de parole. Pour cela il s'appuie sur un ensemble d'interactions entre les informations de haut niveau et les informations de bas niveau portées par le signal acoustique, qui lui permettent de décoder le signal de parole, et de le reconstruire si celui-ci est altéré, par un environnement bruité ou une perturbation d'ordre physiologique par exemple. Les informations de haut niveau font référence à l'ensemble de connaissances linguistiques d'un individu, de l'environnement dans lequel le signal acoustique est produit et des connaissances partagées entre les interlocuteurs. Les informations de bas niveau font référence aux sons de parole, au niveau articulatoire et au codage / décodage de ces sons.

La conjonction de ces informations constitue des indices sur lesquels s'appuie l'auditeur pour percevoir un message et lui attribuer du sens. C'est l'interaction entre ces niveaux d'informations qui permet les processus de restauration. Cela suppose que ces effets puissent apparaître à différents niveaux linguistiques. Par exemple, en guise d'avertissement, A dit à B « Qui sème le vent ... », c'est à B de recouvrer la fin de la célèbre maxime « ... récolte la tempête », ainsi, l'auditeur reconstruit la fin de la phrase à partir de ces connaissances de haut niveau.

Dans le même esprit, l'auditeur est préférentiellement attiré par la reconstruction d'une séquence lexicale même si ce qui est perçu n'est pas un véritable mot. C'est ce que montre l'expérience de Ganong (1980) sur la perception de séquences ambiguës [t/d] en anglais. C'est la séquence qui suit le phonème cible qui oriente le choix de l'auditeur : [t] a principalement été associé à la séquence [ask] qu'à la séquence [af], elle-même majoritairement associée à [d] en référence aux mots « dash » ou « task ». En français, cela pourrait se traduire par une préférence pour [d] si le segment perçu est [ig] par rapport à [iʒ] qui devrait plutôt être associé à [t] en référence aux mots « digue » et « tige ». Ces phénomènes relèvent d'un « effet lexical » dans lequel les auditeurs ont tendance à

s'orienter vers des catégorisations phonétiques qui produisent des mots par rapport à celles qui n'en produisent pas (Ganong, 1980). De plus, ces effets d'attraction lexicale se retrouvent également dans les réponses des auditeurs à une tâche de transcription de pseudo-mots (Grataloup et al., 2009; Rebourg, 2018; Samuel, 1996, 2001). Après inventaire et catégorisation de ces réponses, environ 6% ont été relevés comme des séquences appartenant au lexique. Cette étude suggère que l'attraction lexicale demeure inévitable puisque c'est une stratégie de perception ancrée. Ainsi, Ganong (1980), propose de définir la restauration lexicale comme « un processus qui opère lors de l'accès lexical, lorsque l'on reconstruit un mot à partir de nos connaissances lexicales, même si le signal acoustique perçu ne correspond, en réalité, pas à un mot ». Cette définition rejoint la vision de Warren (1970), qui suggère que l'on tire parti de nos connaissances lexicales pour interpréter ce qu'on entend et pallier l'absence d'informations lors de la perception d'un signal de parole altéré. C'est notamment sur la base d'un test de perception qu'il en propose cette définition. Il montre par ailleurs que lorsque le phonème cible est remplacé par du bruit (bruit blanc ou toux par exemple) les auditeurs ne parviennent pas à localiser le son altéré, alors que s'il est remplacé par du silence ils l'identifient plus facilement.

De la restauration lexicale à la restauration phonémique il n'y a qu'un pas, car même si la restauration intervient sur un item lexical, cela s'apparente largement à une restauration phonémique puisque ce sont précisément des unités sonores de bas niveaux qui sont reconstruites. Ainsi, Bashford et al. (1996) présentent la restauration phonémique comme un « processus sophistiqué et hautement adaptatif qui peut utiliser des portions intactes de parole entourant une interruption pour identifier et reconstruire perceptuellement le fragment de signal oblitéré⁶ ».

Les phénomènes de restauration phonémique dans la perception de la parole sont essentiels à l'enjeu communicationnel. Ils participent à l'intelligibilité au sens large, mais aussi au sens restreint puisque la restauration des unités de bas niveau représentées par les sons de parole, concerne directement un niveau segmental. Ainsi, les phénomènes bien connus de restauration phonémique soulignent la relation étroite entre les informations de haut et de bas niveaux, et le bénéfice mutuel qu'ils s'apportent.

Des effets de contextes illustrent également l'intégration de différents niveaux d'information pour décoder un signal acoustique de parole. Par exemple, les effets de coarticulation sont définis par la répercussion d'informations acoustiques entre différents phonèmes produits au sein d'une séquence. C'est un transfert d'information de proche en proche aux phonèmes adjacents. Ces effets inhérents à la production ont des répercussions sur la perception de la parole. Ils constituent une aide pour décoder un signal acoustique et permettent à l'auditeur d'anticiper la suite de la séquence en fonction des informations acoustiques fines qu'il aura préalablement perçues. C'est une idée décrite par Liberman, (1970) (McClelland et al., 1987) pour qui « le contexte dans lequel un phonème se produit restructure les indices de l'identité de ce phonème⁷ ». En effet, les

⁶ "Phonemic restoration, is a highly adaptive and sophisticated process that can use intact portions of speech surrounding an interruption to identify and perceptually reconstruct the obliterated signal fragment."

⁷ "the context in which a phoneme occurs restructures the cues to the identity of that phoneme"

gestes articulatoires réalisés sont différents selon que l'on prononce [si] ou [su], des caractéristiques acoustiques et articulatoires de la voyelle se répercutent sur la consonne. Ce chevauchement des indices acoustiques dans la séquence perçue lui confère un caractère prédictif sur lequel s'appuie l'auditeur pour décoder le signal. Lors de la production de la séquence [si], l'étirement des lèvres (configuration articulatoire nécessaire à la réalisation de /i/ antérieur, haut, étiré) est présent dès la production de /s/. De même, lors de la production de la séquence [su], l'arrondissement des lèvres (/u/ postérieur, haut, arrondi) est également présent dès la réalisation de /s/. Ainsi, si l'on présente à des auditeurs la production de la consonne /s/ seulement, ils se montrent capables de 'prédire' le phonème vocalique suivant, soit une partie de la structure dans laquelle ce phonème a été produit. Le transfert d'informations acoustiques entre les phonèmes produits au sein d'une séquence sert l'auditeur puisqu'il recueille des indices acoustiques fins pour anticiper et distinguer des séquences proches ou semblables.

C'est également l'intégration de ces niveaux d'informations qui permet à un auditeur de faire face à la variabilité de la parole. Plusieurs auteurs ont tenté de modéliser la perception de la parole, notamment au travers de modèles d'accès au lexique. Dans les modélisations à exemplaires (Kuhl, 1991), la forme de surface (un signal d'entrée) et sa forme sous-jacente (un signal perçu) entretiennent une relation étroite. Les formes de surface sont associées à une séquence déjà perçue et stockée en mémoire. Les formes sous-jacentes sont emmagasinées en mémoire, et ce répertoire est perpétuellement mis à jour. Ces formes sous-jacentes n'ont pas un caractère durable mais elles peuvent être réactivées quand cela est nécessaire. Prenons l'exemple du souvenir pour illustrer ce phénomène ; je ne me souviens plus consciemment de ce souvenir, mais quelque chose, un élément de l'environnement (perçu par les sens) m'y fait penser et il revient à ma mémoire alors qu'il n'avait pas été 'réactivé' depuis plusieurs années par exemple. Pour une forme linguistique on peut imaginer que ce soit similaire, toutes les formes rencontrées sont représentées quelque part dans la mémoire et peuvent être mobilisées dans la perception et le décodage des unités de parole. Ces modèles, au moyen de listes d'exemplaires, considèrent la variabilité de la parole. Dans les modèles abstractionnistes (Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002) l'auditeur se base sur des informations phonémiques et intra-phonémiques (les traits) pour décoder le signal de parole. Les représentations de surface sont indépendantes du contexte et abstraites. En se montrant robustes aux bruits et aux variations de surfaces elles intègrent la variabilité de la parole.

Nous avons exposé ici différents modèles qui tentent de rendre compte des niveaux impliqués dans la production et dans la perception de la parole, ainsi que de déterminer quelle est l'unité perceptive de base. Aucun modèle ni aucune unité ne font totalement consensus, en revanche ils soulèvent une même difficulté, l'intégration de la variabilité de ces mécanismes. En réponse à cette contrainte les auteurs suggèrent plusieurs stratégies d'adaptation. Dans le versant de la production de la parole, la variabilité relève d'adaptation au contexte et « au corps ». Dans le versant de la perception l'appréhension de la variabilité repose sur les facultés du système perceptif à décoder le signal acoustique

et à le restaurer si nécessaire. En revanche, les auteurs s'accordent sur la nécessité de l'encodage – décodage de plusieurs niveaux linguistiques pour la production et la perception. C'est la somme de ces informations combinées qui détermine la réussite de l'acte communicationnel. L'auditeur mobilise et utilise probablement autant d'informations que nécessaires et disponibles pour décoder un signal acoustique de parole. Il apprend constamment lors de ses échanges communicationnels et maintient à jour ses connaissances linguistiques. Les auteurs s'accordent également dans la conceptualisation d'un niveau d'intégration segmental qui porte sur l'encodage-décodage des unités phonémiques de bas niveau. En effet, l'articulation constitue la dernière étape de la production, celle de la mise en son. Le décodage de ces unités sonores constitue la première étape de la perception. Si l'on souhaite évaluer l'intelligibilité dans le cadre de troubles de la parole qui touchent l'articulation, le recours aux concepts de phonétique et de phonologie peut s'avérer utile : ils offrent en effet un moyen de conceptualiser et de représenter les sons aux travers de leurs propriétés articulatoires et acoustiques.

2.2. Concepts de phonétique et phonologie

Historiquement, la phonétique et la phonologie sont abordées et considérées comme deux disciplines indépendantes. L'évolution de la recherche a peu à peu permis d'envisager ces disciplines comme deux facettes, deux versants, d'une même problématique. Constituées de la même racine 'phon', du grec ancien *fonè* ou *phōné* qui signifie 'son articulé' ou 'son de la voix humaine', elles ont en commun de porter sur l'étude des sons de parole de la langue parlée, de leur conception à leur réalisation, en passant par leurs propriétés acoustiques.

Le champ de la phonologie émerge dans les années 1930, notamment par les travaux du cercle de Prague et plus particulièrement de Trubetzkoy (1939). « L'étude des sons appartenant aux actes de parole (la phonétique), qui est concernée par les phénomènes physiques concrets, devra utiliser les méthodes des sciences naturelles, tandis que l'étude des sons appartenant au système du langage (la phonologie) utilisera uniquement les méthodes de la linguistique, ou des sciences sociales. » De fait, phonétique et phonologie sont considérées comme deux disciplines qui utilisent des méthodes différentes, et appartiendraient presque à deux champs de recherche distincts. Toutefois, cette dichotomie peut aussi être abordée au travers de la notion de versant qui neutraliserait l'opposition entre ces champs de recherche. Les « sciences naturelles », apparentées à la phonétique selon Trubetzkoy, peuvent s'interpréter comme des sciences du corps, et être relatives à la physiologie du système de production de la parole et à l'acoustique des sons produits. La phonologie relèverait des « sciences linguistique et sociale » et peut s'interpréter au travers des unités qui constituent l'objet langue, en termes d'opposition et de distinctivité dans un système.

Pierrehumbert (1990) souligne que « une théorie englobant la phonologie, la phonétique et leur relation entre elles est nécessaire », suggérant qu'elles sont étroitement liées. Ceci est appuyé par Fraser, §5, p109 (Hardcastle & Beck, 2014) selon

qui « l'intérêt principal ne réside pas dans les représentations phonétique ni phonologique mais dans les relations entre elles⁸ ». L'idée de Ohala (1990) selon laquelle phonétique et phonologie sont des modules non-autonomes suggèrent que l'une et l'autre impliquent l'intégration d'éléments communs pour exister en tant qu'unité. Elles entretiennent donc une relation d'interdépendance en ayant le même objet d'étude : les sons de parole. Toutefois, elles l'abordent sous un prisme différent. La phonétique concerne l'aspect matériel des sons au travers de la production, de la transmission et de la réception des sons de parole. La phonologie se distingue de la phonétique car elle s'intéresse aux unités constitutives, contrastives et opposables du système dont les sons sont les éléments. Elle traite des unités linguistiques distinctives qui composent le système sonore d'une langue, les phonèmes. Un phonème existe en opposition avec un autre phonème, au travers de la valeur distinctive qu'il occupe dans le système. Phonétique et phonologie s'articulent entre elles dans le système de représentation de l'Alphabet Phonétique International dans lequel à un phonème correspond un symbole phonétique. Un phonème est une représentation abstraite d'unités sonores, il se distingue des aspects articulatoires et acoustiques concrets transcrits par le symbole phonétique.

En substance, la phonétique s'intéresse à la réalisation concrète et perceptive des sons de parole. Elle se divise en trois branches qui recouvrent les 3 phases essentielles à l'échange de parole en situation conversationnelle : - la phonétique articulatoire portée sur les mécanismes physiologiques de production de la parole et sur l'inventaire des gestes articulatoires impliqués dans la production ; - la phonétique acoustique tournée vers des mesures physiques et acoustiques des propriétés des sons et de leur transmission dans l'air ou au moyen de télécommunications ; - la phonétique perceptive ou auditive qui se concentre sur les aspects liés à la réception des sons aux niveaux neurologiques et physiologiques, auditif et perceptif.

La production d'unités phonétiques est variable et la perception intègre cette variabilité. Cutler and Broersma, §4, p71 (Hardcastle & Beck, 2014) traduisent cette variabilité par l'adaptabilité des systèmes de production et perception, ainsi « le traitement phonétique natif est très précis quand cela est nécessaire, imprécis lorsqu'il est suffisant et même flexible lorsque c'est approprié⁹. » « Le traitement phonétique (...) est un équilibre qui vise à atteindre la plus grande précision quand cela est nécessaire, mais qui permet d'abandonner cette précision quand elle n'est pas nécessaire¹⁰ ». C'est donc la quantité d'information phonétique jugée pertinente qui est produite par un locuteur ou utilisée par un auditeur dans sa perception.

Les unités d'intérêt de la phonétique et de la phonologie résident dans les sons de parole et leurs traitements proviennent des mêmes systèmes. Le système phonémique du français peut être divisé en deux catégories principales de sons : consonantiques et

⁸ "The key interest has been in neither the phonetic nor the phonological representation, but the relationships among them"

⁹ "Given that the native phonetic processing is so precise when needed, imprecise when sufficient, and even flexible when appropriate" (Beck, 2005)

¹⁰ "Phonetic processing (...) is a balancing act in which the aim is to achieve the highest precision where it does good, and abandon precision where it is not necessary" (Beck, 2005)

vocaliques. Nous rappelons ici à la mémoire du lecteur quelques éléments de physiologie essentiels à la compréhension des mécanismes articulatoires en jeu dans la production phonémique. Le flux d'air expulsé traverse le conduit vocal (trachée > larynx > la glotte (cordes vocales) > cavité pharyngée > cavité buccale / nasale). Ces articulateurs produisent des constriction qui modifient la forme du conduit et modulent le flux d'air afin de produire des sons. Chaque son résulte d'un geste articulatoire défini en une série de traits articulatoires nécessairement mis en jeu pour atteindre une cible phonémique.

2.2.1. Phonétique articulatoire

Deux types de sons caractérisent le français, les sons consonantiques et les sons vocaliques. Dans cette section nous présenterons ces deux catégories de phonèmes, associés au éléments physiologiques nécessairement mobilisés pour leur réalisation.

2.2.1.1. Consonnes

Les sons consonantiques sont caractérisés et se distinguent les uns des autres par trois traits spécifiques : le lieu et mode d'articulation et le mode phonatoire. Le mode phonatoire fait référence à la vibration des cordes vocales. Elles ne vibrent pas lors de la production de consonnes sourdes (non voisées) mais vibrent lors de la production de consonnes sonores (voisées). Le mode articulatoire fait lui référence à la manière dont l'air est contraint pour produire des sons. Si l'air passe dans la cavité nasale, il produit un son de type nasal, s'il passe par la cavité buccale, les sons produits sont de type oral. Le passage de l'air peut être obstrué par l'accolement des articulateurs, le type de son produit est dit occlusif. Il peut également être perturbé par le rapprochement des articulateurs qui freinent le passage de l'air, cela produit un bruit de frottement, caractéristique des sons fricatifs. Enfin, le lieu d'articulation caractérise l'endroit où les articulateurs vont se rapprocher dans le conduit vocal. Par exemple, l'apex de la langue entre en contact avec les alvéoles du palais lors de la production de la consonne [t] (occlusive apico-alvéolaire sourde). Les caractéristiques acoustiques des sons sont définies par la configuration du conduit vocal.

Les sons consonantiques peuvent donc s'opposer entre eux par le lieu d'articulation, le mode articulatoire et/ou le mode phonatoire. Nous pouvons illustrer les oppositions 'simples', par exemple, les phonèmes [m], nasale, bilabiale, sonore et [n], nasale, apico-alvéolaire, sonore, s'opposent sur le lieu d'articulation. Dans la production du phonème [m] c'est le mouvement des lèvres qui est caractérisé, la langue est abaissée dans la cavité buccale, alors que pour [n] l'apex de la langue entre en contact avec les alvéoles du palais. Les phonèmes [d], occlusif, apico-alvéolaire, sonore et [z], fricative, apico-alvéolaire, sonore, s'opposent sur le mode articulatoire ; dans [d] le passage de l'air est bloqué jusqu'à « explosion » alors que pour [z] le passage de l'air est contraint par une constriction critique du conduit vocal. Pour illustrer une opposition sur le mode phonatoire prenons [p], occlusive, bilabiale, sourde et [b] occlusive bilabiale, sonore ; lors de la réalisation de [p] les cordes vocales ne vibrent pas contrairement à la réalisation de [b] lors de laquelle les cordes vocales vibrent.

D'autres oppositions sont plus 'complexes' lorsqu'elles concernent des oppositions sur au moins 2 traits spécifiques. Par exemple, [d] et [p] s'opposent sur le lieu d'articulation et le mode phonatoire : pour [d] l'apex de la langue entre en contact avec les alvéoles du palais et les cordes vocales vibrent lors du passage de l'air, en revanche, lors de la production de [p] c'est le mouvement des lèvres qui est caractérisé, la langue n'est pas mobilisée, et les cordes vocales ne vibrent pas. Un dernier exemple nous permet d'illustrer une opposition relative aux 3 traits spécifiques des sons consonantiques. [m] et [ŋ], fricative, post-alvéolaire, sourde, s'opposent sur le lieu et le mode articulo-phonatoire et sur le mode phonatoire. Dans la production de [m], l'air passe dans la cavité nasale, les lèvres sont en mouvement et les cordes vocales vibrent, en opposition, lors de la production de [ŋ], le passage de l'air dans le conduit vocal est contraint par une constriction, l'apex de la langue entre en contact avec l'arrière des alvéoles du palais et les cordes vocales ne vibrent pas.

2.2.1.2. Voyelles

Les sons vocaliques, communément regroupés sous l'appellation de 'voyelles', seront décrits dans cette section. Meunier (2007) suggère que « les voyelles représentent un ensemble de production plutôt homogènes » puisqu'elles « supposent une libre circulation de l'air dans le tractus vocal ». Ce point commun majeur implique que toute la dimension acoustique du son est donnée par les articulateurs mobiles de la cavité buccale, et la cavité nasale. Ainsi, les sons vocaliques de la langue française sont caractérisés par quatre traits distinctifs : la position de la langue sur l'axe d'aperture et sur l'axe antéro-postérieur, la position des lèvres ou protrusion et la nasalité. Un autre point commun fondamental réside dans le voisement. La production de toute voyelle implique la vibration des cordes vocales. Généralement, les voyelles sont décrites à travers le prisme de l'oralité, les voyelles nasales sont décrites comme des voyelles orales auxquelles le trait de nasalité, supposant le passage de l'air dans la cavité buccale permis par l'abaissement du voile du palais, est ajouté, mais cette position est débattue au sein de la littérature. Les voyelles du français sont /i/, /u/, /y/, /e/, /ɛ/, /œ/, /ø/, /o/, /ɔ/, /a/, /ɑ/, /ã/, /ẽ/, /õ/, et /œ̃/.

La mandibule effectue un mouvement vertical qui permet de moduler l'espace de la cavité buccale, elle s'abaisse dans une configuration ouverte et s'élève pour fermer et réduire la taille de la cavité. Son mouvement d'aperture entraîne également un rétrécissement de la cavité pharyngée et un abaissement de la langue. C'est le degré d'aperture de la mandibule qui détermine la catégorie de la voyelle dans une représentation en deux dimensions. Son ordonnée indexe les voyelles ouvertes (/a/, /ɑ/ et /ã/), mi-ouvertes (/ɔ/, /ɛ/, /œ/, /ẽ/, /õ/ et /œ̃/), mi-fermées (/o/, /e/ et /ø/), fermées (/i/, /u/ et /y/). L'abscisse de cette représentation est déterminée par la position de la langue dans la cavité buccale, sur l'axe antéro-postérieur. Elle peut être avancée, son apex peut s'approcher des incisives ou de la partie antérieure de la cavité buccale. Lorsque la partie dorsale de langue va au contact de l'arrière du palais, la masse linguale est dans une configuration postérieure dans la cavité buccale. Selon son degré d'élévation dans la cavité buccale la langue réalise une constriction plus ou moins large. Notons que sa

position est parallèlement entraînée par la mandibule et que son degré d'ouverture impacte la taille des résonateurs, lorsqu'elle est ouverte la cavité buccale présente une amplitude plus grande mais ce mouvement rétrécit la cavité pharyngale. A l'inverse une bouche fermée présente une cavité buccale plus petite mais l'espace dans la cavité pharyngale est maximisé. Les voyelles antérieures sont /i/, /y/, /a/, /e/, /ɛ/, /œ/, /ø/, /ɛ̃/, et /œ̃/. Les voyelles postérieures sont /u/, /o/, /ɔ/, /ɑ/, /ɔ̃/ et /ɑ̃/. Le lieu d'articulation est déterminé par le point de rapprochement maximal entre la langue et le plafond de la cavité buccale (Ladefoged, 1995; Meynadier, 2010). La position des lèvres, qui représente les derniers articulateurs mobiles avant la transmission du signal acoustique, est également représentée. Au travers de l'axe d'antériorité, sont placés à l'avant de l'axe les voyelles non arrondies (ou étirées) et à l'arrière les voyelles arrondies. Les lèvres peuvent être en position arrondie ou étirée, c'est la protrusion. Les voyelles arrondies sont /u/, /y/, /o/, /ɔ/, /œ/, /ø/, /ɔ̃/, et /œ̃/ et les voyelles non arrondies ou étirées sont /i/, /a/, /ɑ/, /e/, /ɛ/, /ɑ̃/ et /ɛ̃/. Trois dimensions sont représentées dans le schéma du triangle vocalique du français : l'aperture de la mandibule, l'antériorité de la position de la langue et l'arrondissement des lèvres. Cette représentation est très intéressante puisqu'elle a été établie à partir des propriétés et configurations physiologiques des articulateurs de la cavité buccale, auxquelles correspondent des caractéristiques acoustiques intégrables dans la même représentation schématique. Par écho, elle en rend donc compte. Pour réaliser la voyelle /i/ les lèvres sont fermées, étirées et la langue est avancée dans la cavité buccale, à l'inverse, pour réaliser la voyelle /o/ les lèvres sont arrondies, la mandibule est mi-fermée et la langue est en arrière de la cavité buccale.

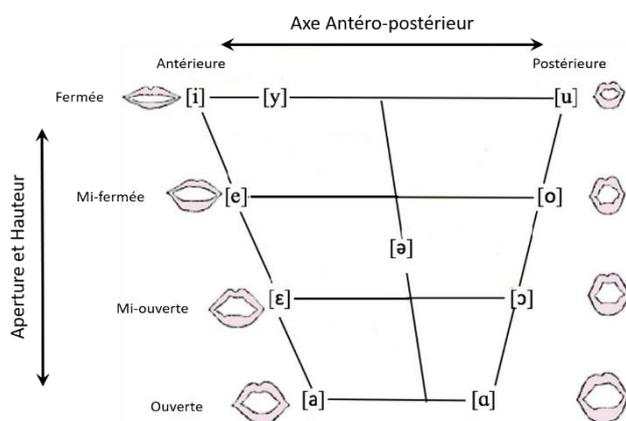


Figure 6 : Illustration du système vocalique du français

Enfin, une dernière dimension ne concernant que quatre voyelles en français (/ɑ̃/, /ɛ̃/, /ɔ̃/, et /œ̃/), est représentée par un trait spécifique : la nasalité. Cette propriété acoustique est produite au travers de l'abaissement du voile du palais permettant le passage du flux d'air dans la cavité nasale. Néanmoins leur description en font pas consensus dans la littérature, tantôt décrite comme l'équivalent des voyelles orales correspondantes (/ɑ/, /ɛ/, /ɔ/, /œ/) (Meunier, 2007; Meynadier, 2010), d'autres auteurs postulent qu'elles relèvent de configurations systématiques qui leurs sont propres (Delvaux, 2009).

D'autre part, l'ensemble des segments vocaliques du français entraînent la vibration des cordes vocales et sont donc tous voisés. Cependant, certaines situations de parole sont susceptibles de neutraliser cette propriété de voisement. Par exemple, la voix ou phonation chuchotée se caractérise par « un bruit laryngé » supposant une production intelligible, sans vibration des cordes vocales (Lagier et al., 2009). Certains effets de coarticulations peuvent également avoir un impact sur le voisement (Meynadier, 2010). Un dernier phénomène peu répertorié dans la littérature concerne le dévoisement des voyelles finales (Paternostro, 2008), ce dévoisement se traduit par une voyelle 'soufflée' présentant un bruit de frottement fricatif.

L'ensemble des différentes propriétés articulatoires présentées permettent de regrouper certaines voyelles ensemble sous un ensemble de traits communs. Ainsi, la notion d'archiphonème désigne un symbole ou un label qui regroupe plusieurs voyelles. Cette notion présente un intérêt particulier puisqu'au travers de ces regroupements elle restreint le nombre de catégories vocaliques ce qui peut se révéler adéquat dans le cadre d'analyse de très grands corpus. Ces regroupements font écho à l'usage des sons en fonction de caractéristiques particulières. En effet, /a/, /ɑ/ et /o/, /ɔ/ ont une valeur contrastive dans le système phonologique du français mais cette distinction n'est pas produite par tous les locuteurs et n'est pas non plus perçue par tous les auditeurs, principalement en raison de caractéristiques linguistiques régionales dans lesquelles certaines distinctions ne sont pas pertinentes puisque non contrastives. Cela signifie qu'elles reposent sur des réalisations physiques différentes mais qui n'assume pas de rôle fonctionnel dans la langue. Il en va de même pour les segments /e/, /ɛ/ et /ø/, /œ/. Ainsi, les segments postérieurs mi-ouvert /ɔ/ et mi-fermé /o/ peuvent être rassemblés sous le symbole /o/, les voyelles ouvertes antérieure /a/ et postérieure /ɑ/ sous l'archiphonème /a/ et les segment antérieurs mi-ouverts /e/ et /ø/ et mi-fermés /ɛ/ et /œ/ se regroupent sous l'archiphonème /e/. Ainsi, le nombre de segment vocaliques peut être réduit à six : /i/, /u/, /y/, /a/, /o/ et /e/.

Chacune des configurations nécessairement impliquées dans la réalisation d'un son de parole relève de gestes articulatoires. Les différentes positions des articulateurs, apportent des caractéristiques acoustiques spécifiques au son produit. Ces gestes s'articulent autour de continuum physiologique que se traduit sur le plan acoustique, le passage d'une configuration à l'autre est opéré par des modifications fines des positions des articulateurs qui entraînent des modifications des composantes spectro-acoustiques des sons. La phonétique acoustique considère des mesures physiques et acoustiques des propriétés des sons. Nous décrirons ci-après ce domaine en prenant appui sur les sons vocaliques.

2.2.2. Phonétique acoustique et perceptive

Un son périodique simple se caractérise par une onde sinusoïdale simple régulière et répétée, de fréquence f dans une fenêtre temporelle t , également appelée 'période', ou $F = 1/T$ avec une fréquence exprimée en Hertz. Les sons de parole sont des sons complexes, qui se composent donc d'ondes périodiques simples combinées ensemble appelées 'harmoniques'. Les relations des harmoniques d'un son s'articulent en termes de

fréquence et amplitude. Un ensemble d'harmoniques est directement et systématiquement associé à un son complexe spécifique. Les liens mathématiques entre les sons périodiques simples et les sons complexes sont explicités par le théorème de Fourier (ou Fast Fourier Transformation – FFT). Celui-ci se base sur la durée (t) de l'harmonique le plus long, soit celui avec la fréquence la plus basse, dont l'ensemble des autres harmoniques représentent des multiples entiers. La fréquence d'oscillation d'un son complexe est déterminée par l'harmonique le plus grave, elle représente la Fréquence fondamentale, ou F_0 . La F_0 est la seule composante spectrale d'un son périodique simple, un son complexe dépendant de la F_0 , la fréquence des autres harmoniques étant déterminée à partir de celle-ci, puisqu'ils en sont des multiples entiers (2^{nd} harmonique = $2 \times F_0$, ...). C'est ainsi un rapport entre la fréquence, l'amplitude et le temps qui est exprimé. Les différents spectres FFT ainsi obtenus peuvent être combinés pour obtenir une représentation graphique de la réalisation acoustique, le spectre. Cette représentation est largement privilégiée dans l'étude des sons de la langue, elle offre la possibilité de segmenter, d'annoter, d'étiqueter le signal acoustique et d'effectuer des mesures sur ce signal. En plus de matérialiser le spectre acoustique au moyen d'une représentation, celui-ci permet de visualiser les pics d'énergie contenus dans un signal acoustique au travers du spectre à bande large qui représente les bandes de formants. Elles sont en ordre croissant sur l'axe des ordonnées (de F_1 à F_n), en fonction de la fréquence, de la plus basse à la haute. Elles représentent des zones d'énergie fréquentielles principalement portées par les segments vocaliques. Un exemple de spectrogramme est disponible en annexe (A. Exemple de spectrogramme)

Les sons de parole sont donc représentés au travers de la relation entre la fréquence et l'amplitude spectrale couplées au phénomène de résonance articulaire. Ces représentations acoustiques font écho aux caractéristiques énergétiques des sons de parole. L'énergie contenue dans le signal acoustique est étroitement liée aux configurations du conduit vocal et aux gestes articulaires réalisés pour produire le son de parole, les formants en sont le reflet. Nous présenterons ici l'intérêt spécifique porté aux 1^{er} et 2^{nd} formants (F) des voyelles du français. Le $3^{\text{ème}}$ formants F_3 est traditionnellement associé à la protrusion des lèvres (Georgeton et al., 2012), mais nous ne l'aborderons pas davantage. En effet, la comparaison des valeurs de F_1 et F_2 permet de déterminer la voyelle émise (Calliope (Firm) & Fant, 1989). Chaque son possède des formants caractéristiques qui lui confèrent une empreinte spectrale unique. Celle-ci sert la classification et constitue des propriétés acoustiques saisies par la perception.

Ainsi, le plan F_1 - F_2 permet de représenter les segments vocaliques dans un espace bi-dimensionnel en fonction de la fréquence des deux premiers formants. Cet espace reflète les configurations articulaires au même titre que le triangle vocalique. Ainsi, F_1 reflète l'aperture de la mandibule et F_2 la position de la langue sur l'axe antéro-postérieur. Les valeurs (en Hertz) de F_1 sont plus élevées lorsque la cavité buccale est ouverte que lorsqu'elle est fermée, à l'inverse, les valeurs de F_2 sont plus élevées lorsque la mandibule et la cavité buccale sont fermées. En effet, le mouvement d'aperture de la mandibule entraîne mécaniquement un agrandissement de la taille de la cavité buccale, provoquant un rétrécissement de la cavité pharyngale. Une grande cavité buccale rime donc avec un

F1 élevé et une grande cavité pharyngale rime avec un F2 élevé. En plus de cet effet impactant la taille des résonateurs buccal et pharyngé, l'aperture de la cavité buccale entraînant mécaniquement un léger recul de la langue dans la cavité buccale, rendant une voyelle antérieure ouverte (/a/) plus en arrière qu'une voyelle antérieure fermée (/i/). Ainsi, le F2 d'une voyelle antérieure fermée est plus élevé que le F2 d'une antérieure ouverte. Celui-ci étant lui-même plus élevé que le F2 des voyelles postérieures. On observe donc une dynamique inverse entre les valeurs de F1 et de F2. Ces subtilités liées à la physiologie de l'appareil phonatoire sont également représentées dans le design du triangle vocalique, sur l'axe d'aperture les voyelles antérieures ne sont pas alignées perpendiculairement à l'axe des abscisses, contrairement aux voyelles postérieures.

Notons ici que les différences anatomiques entre les hommes et les femmes ont un impact sur la longueur du conduit vocal et donc sur la taille du larynx et des résonateurs qu'il héberge. Le conduit vocal féminin est généralement plus court que son équivalent masculin. Ces différences caractéristiques impliquent des valeurs moyennes de formant différentes dépendamment du sexe du locuteur : les valeurs des formants vocaliques sont plus élevées chez les femmes que chez les hommes. Les valeurs moyennes des formants F1 et F2, pour les sexes féminin et masculin, sont représentés dans le tableau suivant, d'après Gendrot & Adda-Decker (2005) à partir de la lecture phrases isolées.

Voyelles	i	y	e	ɛ	a	œ	ø	ɔ	o	u
F1 _{male}	300	300	350	450	650	500	400	550	400	350
F2 _{male}	2050	1800	1950	1700	1300	1450	1450	1050	900	850
F1 _{female}	350	350	450	650	750	550	450	600	450	350
F2 _{female}	2400	2050	2300	2000	1550	1650	1650	1200	950	850

Tableau 1 : Valeurs formantiques de référence utilisées pour le français (Gendrot & Adda-Decker, 2005a)

Ces valeurs constituent une référence pour une analyse acoustique des voyelles orales du français. Les valeurs formantiques sont généralement extraites de façon automatique, au moyen d'un logiciel de traitement de son, Praat (Boersma & Weenink, 1992) par exemple. Celui-ci nécessite l'implémentation d'un script pour extraire les valeurs de formants sur un ou plusieurs points par rapport à la durée du segment ciblé. De nombreux chercheurs ont contribué aux développements de ces scripts librement disponibles et mis au service de la communauté scientifique (il s'agit ici de scripts destinés à Praat), nous citerons Cédric Gendrot (<https://cedricgendrot.wordpress.com/scripts-praat/>) et Mietta Lenne (<https://lennes.github.io/spect/>) pour leurs contributions significatives.

La phonétique acoustique se distingue de la phonétique perceptive en ce sens qu'elle s'intéresse aux composantes spectrales et énergétiques du signal acoustique alors que cette dernière s'intéresse aux propriétés perçues de ces sons. Néanmoins, il n'est pas évident de déterminer si certains éléments perçus sont plus saillants que d'autres, si les auditeurs ont des stratégies de décodage commune ou individuelle, ni comment cet ensemble s'organise. Nous laissons de côté ces questionnements perceptifs qui, selon

nous, s'inscrivent davantage dans le champ de la neurolinguistique. En revanche, chaque son possède une empreinte spectrale unique qui le caractérise et le distingue des autres (Jemaa, 2013). Ces propriétés acoustiques peuvent être décrites en termes de traits distinctifs. Deux faisceaux de traits sont décrits dans la littérature, les traits articulatoires (phonétique) et les traits distinctifs (phonologie). Cette représentation en termes de trait permet d'illustrer les contrastes et les analogies entre les sons et offre une description précise des éléments acoustiques dont ils sont constitués.

2.2.2.1. Traits distinctifs

L'idée de distinguer les phonèmes sur la base d'un faisceau d'indices acoustiques en nombre restreint est à attribuer à plusieurs figures scientifiques majeures de la linguistique du XX^{ème} siècle. En effet, les prémices de cette théorie émergent au travers de la notion d'opposition phonémique (Trubetzkoy, 1939), puis sont ensuite réinvestis par Jakobson (1941) qui postule que les unités phonémiques peuvent être décomposées en un ensemble de plus petites unités encore, faisant directement référence aux contrastes phonologiques. Jakobson, Fant et Halle, (1952) le formalisent en premier, ils précisent d'ailleurs qu'ils considèrent « la représentation en liste de traits distinctifs, en particulier leurs définitions des différents niveaux, comme une esquisse provisoire ouverte à la discussion et qui requiert des vérifications expérimentales et une élaboration supplémentaire ». Puis ce sont Chomsky & Halle (1968) qui se réapproprient ces théories et en repensent le formalisme. Ils précisent que la « composante phonologique » représente un « système de règles qui s'applique à une structure de surface et lui attribue une certaine représentation phonétique ». Ensemble, ces linguistes posent les bases et définissent la notion de « traits distinctifs » qui permet de distinguer et caractériser l'ensemble des phonèmes présent dans les langues, au moyen d'une description permettant de classer les phonèmes par opposition mutuelle à l'aide de faisceaux de traits (Jakobson et al., 1952). Notons que le tableau de référence de l'Alphabet Phonétique International indique les productions irréalisables (en gris) et que les cases laissées libres offrent un potentiel de sons pour l'instant non employés dans les langues du monde répertoriées.

La notion de 'trait' est une représentation formelle, chacun d'entre eux est « construit sur la base d'un faisceau d'indices en correspondance biunivoque avec certaines propriétés du signal » (Rossi, 1981). Ces traits sont donc basés sur les contrastes phonologiques qui permettent de distinguer les phonèmes. Ils caractérisent les propriétés distinctives des unités phonémiques au moyen de traits pertinents, qui permettent d'opposer deux phonèmes, soit deux unités d'information sémantique distinctes (Ghio, 1997, p. 199). C'est en effet de cela qu'il s'agit, qualifier les distinctions phonémiques, en tant que plus petite unité distinctive, puisqu'elles sont susceptibles d'impacter directement le sens d'un énoncé oral. Par exemple, dans les énoncés 'J'ai pris un *bain*' ([bɛ̃]) vs 'J'ai pris un *pain*' ([pɛ̃]), les phonèmes consonantiques [p] et [b] s'opposent sur le trait de voisement (vibration de cordes vocales). Une mauvaise transmission ou perception de cette vibration et c'est le sens de l'énoncé qui s'en trouve impacté. Il en est de même pour une opposition sur le lieu d'articulation par exemple, les énoncés « drôle

de *saveur* » ([savœʁ]) vs « drôle de *faveur* » ([favœʁ]) ou /s/ est alvéolaire et /f/ est labiodental s'opposent sur le trait distinctif + et – coronal, respectivement. Ainsi, on imagine assez facilement les conséquences que pourrait avoir une confusion sur le mode d'articulation, entre une nasale et une occlusive, dans des énoncés comme « il *nie* être complice » ([ni]) vs « il *dit* être complice » ([di]) ou /n/ est + nasale et /d/ + coronal en termes de traits phonologiques. Il est de même bien moins amusant d'avoir fait une « faute » ([fot]) qu'une « fête » ([fɛt]) ou /ɛ/ et /o/ s'opposent sur le lieu d'articulation, position de la langue sur l'axe antéro-postérieur, /ɛ/ est antérieure et /o/ postérieure. Ils sont distingués par le trait phonologique 'back' (/ɛ/ - back et /o/ + back).

Ces exemples permettent d'illustrer que ce ne sont pas les sons qui sont altérés mais bien les traits dont ils sont composés, puisque dans les déviations suscitées, un ensemble de traits tend à être préservé. En effet, face à la perte d'un trait, des mécanismes de compensation apparaissent, ils relèvent d'une adaptation (allongement de la durée par exemple) qui permet de renforcer certains contrastes. En d'autres termes, ce n'est pas le son dans sa globalité qui est altéré, c'est une de ses composantes phonologiques. Le couplage en traits distinctifs pour qualifier un phonème permet une analyse linguistique fine des bas niveaux en termes de production et de perception des unités phonémiques.

Bien que les représentations linguistiques soient statiques par nature, différemment de la perception qui revêt un caractère dynamique, elles permettent de mieux comprendre et appréhender les différents niveaux et composantes linguistiques. Ainsi, la notion de trait distinctif a beaucoup apporté au formalisme de la phonétique et de la phonologie fondamentales, et apparaît également pertinente pour qualifier les altérations de ces unités dans la parole saine tout comme dans la parole pathologique. Ainsi, les unités phonémiques décrites en termes de traits phonétiques (articulatoires) et phonologiques (acoustiques) sont des outils linguistiques qui peuvent être mises au service de l'évaluation de la parole et notamment dans le cadre de la parole pathologique. Celle-ci est caractérisée par des troubles qui relèvent de la mobilité et de la coordination de gestes articulatoires et donc de l'intégrité du système de phonation. En effet, la parole d'un locuteur est « le produit combiné de sons état organique et de tout ajustement phonétique de l'appareil vocal » (Beck, 2005). En cas de perturbations physiologiques de ce système, les outils de représentation linguistique sont adéquats pour évaluer la parole pathologique au moyen d'analyses acoustiques du signal de parole mais aussi de descriptions des unités produites et perçues de bas niveau que représentent les sons de parole. Dans un contexte clinique, elles peuvent participer à l'identification et à la caractérisation des composantes de parole préservées et dégradées, à condition de les intégrer à la stratégie d'évaluation de l'intelligibilité.

3. Evaluation de l'intelligibilité

Nous l'avons souligné dans la section précédente, les mécanismes de production et de perception de la parole, opérés au travers du système physiologique, sont très efficaces. Leur intégrité est essentielle à la réussite de l'échange communicationnel, mais des mécanismes de compensation peuvent entrer en jeu si elle fait défaut. La réussite de l'échange repose alors sur l'opérationnalité de deux systèmes, physiologique et fonctionnel, ainsi que sur leurs interactions. En conséquence, la question de l'évaluation de l'intelligibilité n'est pas triviale. En effet, mesurer la qualité de cet échange est complexe, c'est une gymnastique qui s'articule autour de l'évaluation d'une production au travers de la perception, production dépendante du système physiologique qui la permet et de l'intégrité de son fonctionnement. Un auditeur utilise simultanément des informations portées par des unités plus ou moins larges. Nous l'avons vu, la perception du signal acoustique passe par l'intégration de plusieurs niveaux d'informations (phonémique, lexicale, sémantique ...). Ces mécanismes sont particulièrement efficaces, la production varie pour s'adapter au contexte et à l'état organique du locuteur. La perception considère cette variabilité et l'intègre au moyen du recours aux informations de haut niveau permis par les phénomènes de restauration lexicales et phonémique.

En plus du locuteur et de l'auditeur, le contexte dans lequel se déroule l'évaluation en définit les enjeux. Par exemple, dans le contexte de l'apprentissage d'une nouvelle langue, l'évaluation portera d'abord sur des unités plutôt larges telle que des mots de vocabulaire par exemple. Les connaissances de haut et bas niveaux de l'apprenant sur sa langue maternelle seront réinvestis dans l'apprentissage de cette nouvelle langue. Cela permet notamment à l'auditeur de développer des connexions entre des informations déjà connues et de nouvelles informations. Il peut en résulter des mécanismes facilitateurs, par exemple il est plus facile de retenir des nouveaux mots en L2 quand ceux-ci ont le même genre que dans la maternelle (Zappa et al., 2022). Dans ce cadre, l'évaluation vise à mesurer l'apprentissage. Elle se matérialise généralement par des contrôles des connaissances par une modalité écrite. C'est alors plutôt la compréhension qui est ciblée. L'évaluation de la production dans une langue seconde se fera dans un second temps, et plus particulièrement en contexte. Le locuteur éprouvera ses niveaux de compréhension et d'intelligibilité. Selon que l'on souhaite évaluer l'apprentissage, la compréhension ou l'intelligibilité, ce sont différents niveaux linguistiques qui seront ciblés.

En définitive, l'évaluation de l'intelligibilité nécessite de prendre en compte l'ensemble de la situation interactionnelle : le locuteur, l'auditeur, le contexte ou plus précisément la spécificité de l'objectif visé par l'évaluation. En contexte clinique, il s'agit d'évaluer le handicap communicationnel. Nous l'avons souligné précédemment, la difficulté d'évaluation relève à la fois de l'imprécision de la notion d'intelligibilité et de la non prise

en compte des mécanismes de production et de perception en jeu. Ces difficultés sont augmentées si dans le cadre de l'évaluation on ne détermine pas avec précision le niveau ou l'unité linguistique qu'il convient d'évaluer, et si on ne choisit pas les bons outils de description. L'ensemble de ces aspects ne sont pourtant généralement pas considérés dans les évaluations classiques, qui cumulent déjà les difficultés évoquées en 1.1.2 LA mesure de l'intelligibilité.

3.1. Evaluation perceptive

C'est la méthode perceptive qui est utilisée dans l'évaluation de la parole en contexte clinique. Dans cette thèse, il s'agit d'évaluer le trouble de la parole chez des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. Cette évaluation des aspects productifs du locuteur est classiquement réalisée au travers du système perceptif de l'auditeur au moyen de différents tests et batteries d'évaluations cliniques.

Les premiers tests d'évaluation de l'intelligibilité dédiés au contexte clinique ont été développés pour l'anglais, au travers de la « *Word intelligibility lists* » (Tikofsky & Tikofsky, 1964) et du « *Single word intelligibility test* » (Tikofsky, 1970) composé de 50 mots uni ou bisyllabiques. A peu près au même moment, Voiers et al. (1973) développent le « Diagnostic Rhyme Test ». C'est un test de rime basé sur des paires minimales, en anglais. Ce test propose une évaluation « diagnostique et comparative » de l'intelligibilité des consonnes à l'initiale de mot. Il sera adapté au Néerlandais (Steeneken, 1982) et au Français, en tant que « Test des paires minimales » de Peckels & Rossi (1973). Ce dernier est le premier test disponible pour le français. C'est un test très abouti mais qui est peu utilisé car sa passation est longue : elle requiert la production de 216 mots. C'est un exercice potentiellement laborieux et contraignant pour un locuteur atteint de troubles de la production de la parole.

Par la suite, d'autres tests ont été développés, notamment « Le Single Word Intelligibility Test » (Kent et al., 1989) qui a été adapté et traduit en français par Gentil (1992) puis repris par Auzou et al. (1998) dans l'ECD puis par Crochemore & Vannier (2001) pour devenir le Test Phonétique d'Intelligibilité de la Batterie d'Evaluation Clinique de la Dysarthrie (BECD) réalisé par Auzou & Rolland-Monnoury (2006) » (Ghio et al., 2016). Cette Batterie d'Evaluation Clinique de la Dysarthrie (ci-après BECD) fournit un score d'intelligibilité déterminé à partir de 3 exercices différents visant à évaluer 3 niveaux linguistiques différents : Lecture de mots, lecture de phrases et production de parole spontanée. L'évaluateur sélectionne 10 mots et 10 phrases, parmi 50 disponibles dans la batterie, que le locuteur produit. L'évaluateur propose un score pour chaque épreuve, sur une échelle de 0 à 8. Les scores d'intelligibilité représentent la somme de ces 3 notes.

Le matériel linguistique de cette batterie d'évaluation clinique est en fait peu contrôlé. Concernant les mots, tout d'abord, le nombre de syllabes par mots n'est pas pris en compte, on retrouve environ 20% de mots bisyllabiques et 80% de mots monosyllabiques, ces derniers sont sur-représentés par rapport aux mots bisyllabiques. Ensuite, on ne

retrouve pas d'équilibre phonétique dans la liste de mots : par exemple, les consonnes fricatives sourdes du français /ʒ/ /z/ and /v/ n'apparaissent pas dans la liste et ne sont donc pas évaluées. Certains phonèmes sont sous-représentés, /d/ et /y/ n'apparaissent qu'une seule fois (respectivement dans /doktør/ - 'docteur' et /gry/ - 'grue'). Tout comme le phonème vocalique /u/, dans le mot 'mouche' (/muʃ/). Cette unicité de phonème facilite la reconnaissance du mot cible. D'autres phonèmes ne sont représentés qu'au travers d'un groupe consonantique (ou cluster), à l'initiale de la syllabe, /f/ et /g/. Cela suggère que la distribution et la fréquence d'occurrence des phonèmes ne sont pas contrôlées. De plus la fréquence d'occurrence des mots dans la langue (ici dans les films, par million d'occurrences – source Lexique.org) n'est pas non plus contrôlée. Certains sont rares ('clenche' /klãʃ/ - 0, 'porc' /por/ - 41) d'autres sont plus fréquents ('café' /kafe/ - 164, 'docteur' /doktør/ - 233.86). De surcroit, le nombre de voisins orthographiques n'a pas non plus été contrôlé, 'brin' /brẽ/ en a 9, alors que 'thym' /tẽ/ n'en a pas. Enfin, cette liste de mots contient des noms ('pain' /pẽ/, 'paquet' /pake/), des pronoms ('toi' /twa/, 'moi' /mwa/) et des mots qui peuvent être des noms ou des adjectifs ('franc' /frã/, 'nain' /nẽ/) ou des adverbes ('grand' /grã/) : la catégorie grammaticale des mots cibles n'est pas contrôlée. Pour résumer, la fréquence d'apparition des mots dans la langue, la fréquence et la position des phonèmes dans le mot, le nombre de voisins orthographiques et la catégorie grammaticale des mots ne sont pas contrôlés. Concernant les 50 phrases disponibles dans ce test, elles ont toutes les mêmes amorces (« l'enfant »), suivies d'un verbe conjugué au présent de l'indicatif. C'est donc le verbe conjugué qui est la cible de l'évaluation, celui-ci étant pris dans une phrase porteuse. La répétition des amorces apporte un caractère de prédictibilité à la phrase (Ghio, Giusti, et al., 2020). Ce sont des phrases courtes (sujet + verbe) qui ont toutes la même structure syntaxique et la même modalité (affirmative). Il n'y pas de modalités interrogative ou exclamative. Il n'y a pas non plus de complément d'objet. Ce sont des phrases simples et peu diversifiées.

Dans ces deux tests visant à mesurer l'intelligibilité, le nombre d'items disponibles dans la liste de mots est faible (BECD) et/ou restreint (Test paires minimales) et est donc facilement mémorisable. Ceci a un effet délétère sur l'évaluation du patient, puisque l'évaluateur, grâce à l'adaptabilité des mécanismes de perception, restaure les items perçus ce qui entraîne une sous-évaluation du déficit du patient. Ce sont les effets inhérents aux mécanismes en jeux dans la perception de la parole ; nous l'avons souligné les processus de production et de perception de la parole sont nécessairement impliqués et sont à considérer conjointement (2.1.3 Production et perception de la parole).

L'examen de ces tests soulève également la question du niveau linguistique évalué. La BECD propose une évaluation de l'intelligibilité au sens large, puisque c'est un score global calculé à partir des scores obtenus pour 3 tâches de production. Pourtant, les tâches de production spontanée et lecture de phrase évaluent l'intelligibilité de la parole au sens large, alors que la tâche de production de mots évalue l'intelligibilité lexicale. Le test des paires minimales évalue l'intelligibilité dans un sens restreint : en ciblant les consonnes initiales il atteint le niveau segmental de l'intelligibilité. Ces deux tests n'évaluent donc pas les mêmes niveaux linguistiques, pourtant ils prétendent tout deux évaluer

« l'intelligibilité ». Il est donc d'une absolue nécessité de déterminer la finalité et l'objet de la mesure pour conduire une évaluation de l'intelligibilité méthodologiquement adaptée.

Ces différences entre les évaluations au regard des différents niveaux linguistiques qu'elles considèrent, soulignent davantage encore l'importance de préciser le cadre définitoire de la notion d'intelligibilité. En effet, la considération du niveau linguistique visé par l'évaluation clinique permet de caractériser l'objet de la mesure. Ce manque de précision rend les différents tests peu satisfaisants au regard du niveau linguistique évalué. Le choix du matériel linguistique impacte considérablement l'évaluation. Dans les batteries classiques, les items lexicaux en nombre limité laissent s'introduire des biais à l'évaluation perceptive, tels que les effets de restauration lexicale permis pas le recours aux informations de haut niveau (2.1.2 La perception de la parole). D'autres part, les différents déficits présentent des évaluations spécifiques. Par exemple, dans une évaluation du manque du mot les cliniciens emploient les tests dédiés de dénomination et fluence qui visent le niveau lexical. Dans le cas de l'évaluation des déficits fonctionnels après un cancer ORL c'est la précision articulatoire qui est visée. On peut alors questionner l'intérêt de l'emploi de formes lexicales pour évaluer ce niveau phonémique. On est en effet face à un problème d'exécution motrice : dans les cancers ORL, il n'y a pas d'atteinte neurologique, donc on présuppose que les unités linguistiques sont encodées correctement et les commandes d'exécution motrices sont aussi intègres. Néanmoins, sur le plan physiologique les structures du conduit vocal (cavité buccale et oropharynx) sont lésées ce qui entraîne des difficultés d'exécution du geste articulatoire. Dans une évaluation de la parole pathologique, après un cancer traité de la cavité buccale ou de l'oropharynx, il convient d'évaluer les aspects articulatoires grâce à l'examen des unités pertinentes : les segments. Cette idée n'est pas nouvelle ; Certains travaux sur la mesure de l'intelligibilité, hors contexte clinique, se sont déjà intéressés au niveau segmental.

3.2. Evaluation perceptive segmentale

Les travaux menés dans ce domaine avaient pour objectif de mesurer la contribution des différents segments de parole à l'intelligibilité.

Les études de Cole et al., (1996); Fogerty & Kewley-Port, (2009); Kewley-Port et al., (2007) présentent un intérêt particulier puisqu'elles mettent toutes à l'épreuve le même corpus de productions de phrases en anglais américain, le TIMIT database (Garofolo et al., 1993). C'est un corpus conséquent de productions de 450 phrases différentes pour un total de 6300 phrases produites par 669 locuteurs (Zue & Seneff, 1996). Les phrases produites dans ce corpus ont été construites pour être équilibrées en termes d'occurrences entre les consonnes et les voyelles.

La première étude conduite, Cole et al., (1996), propose d'évaluer « la contribution des consonnes versus des voyelles dans la reconnaissance de mots » dans la parole. A partir des phrases enregistrées dans le corpus TIMIT, les auteurs modifient le signal acoustique de façon à conserver des proportions plus ou moins importantes des segments vocaliques ou consonantiques. Ils déterminent plusieurs conditions : consonnes seulement, voyelles

seulement, et des proportions de segments conservés 10, 20, 30 et 40%, ils y ajoutent l'étude de la contribution des frontières de segments. Le reste du signal sonore est remplacé par du bruit. Après écoute et transcription du plus de mots possibles dans la phrase ciblée par différents jurys d'écoute entre chaque expérience, les taux de transcriptions sont mesurés au moyen d'un algorithme. Un taux par phrase et un taux par mot sont ainsi calculés. Leurs résultats montrent que c'est toujours la condition dans laquelle les voyelles sont dominantes qui présente les taux de reconnaissance les plus élevés. En effet, dans la condition voyelles seulement, 87.4% de mots et 60% de phrases sont correctement identifiées, contre 47,9% de mots et 11,9% de phrases dans la condition sans voyelle. Les conditions qui conservent les plus grandes proportions de segment vocalique sont mieux identifiées que les conditions à dominances consonantiques. Ils concluent que « les voyelles sont plus importantes pour la reconnaissance que les consonnes malgré le fait qu'elles soient également représentées dans les phrases tests ». Ainsi, pour la reconnaissance des mots et des phrases, les voyelles présentent un bénéfice supérieur aux consonnes, leur contribution est plus importante dans l'intelligibilité (au sens large) de la parole.

Ces résultats sont en accord avec ceux de Kewley-Port et al., (2007). Plus d'une décennie plus tard, ils répliquent la première expérience de Cole et al., (1996), avec le même corpus (TIMIT), mais auprès de deux groupes d'auditeurs, jeunes et âgés. 3 conditions sont proposées aux auditeurs : Full, la phrase n'a pas été modifiée, Consonne seulement, ou seuls les segments consonantiques ont été conservés sur le signal acoustique, et Voyelle seulement, phrases dans lesquelles seules les voyelles sont conservées. Dans les phrases les segments 'masqués' sont remplacés par du bruit. « L'intelligibilité des phrases a été notée comme le nombre total de mots corrects et convertie en pourcentage des valeurs¹¹ » (Kewley-Port et al., 2007). Les phrases dans lesquelles seules les consonnes sont conservées sont moins bien identifiées que les phrases ou les voyelles seulement ont été retenues, et ce pour les deux groupes d'auditeurs. Leurs résultats montrent un « avantage pour les phrases ne contenant que des voyelles par rapport aux phrases qui ne contiennent que des consonnes¹² » (Kewley-Port et al., 2007). Cela montre que la contribution des voyelles est supérieure à celles des consonnes dans l'intelligibilité des phrases en anglais américain.

Une autre étude, Fogerty & Kewley-Port (2009), portant sur le même corpus TIMIT, et plus précisément sur les mêmes phrases que Kewley-Port et al., (2007), corrobore les résultats précédents. Elle est également conduite auprès de deux groupes d'auditeurs, jeunes et âgés. En plus des conditions Consonnes seulement et Voyelles seulement ils ajoutent 4 conditions par segment dans lesquels ils conservent des proportions plus ou moins importantes des segments ciblés, sur un modèle similaire à l'expérience de Cole et al. (1996). Les résultats montrent que, quel que soit le groupe d'auditeur, c'est dans la condition Voyelle seulement et dans celle comprenant les plus grandes proportions de Voyelle conservées que les scores d'intelligibilité en termes de taux de reconnaissance sont les meilleurs. Ils concluent en faveur d'un « avantage de deux pour

¹¹ "Sentence intelligibility was scored as the total number of words correct and converts to % values"

¹² "Advantage for sentences containing only vowels compared to sentences that contained only consonants"

un pour les phrases contenant uniquement des voyelles par rapport aux phrases contenant uniquement des consonnes¹³ » (Fogerty & Kewley-Port, 2009). Ces trois études montrent qu'en anglais américain, ce sont les segments vocaliques qui contribuent davantage à l'intelligibilité de phrases et de mots, par rapport aux segments consonantiques. Ainsi, plus la proportion de 'voyelles' conservée est importante, meilleure l'intelligibilité est. Par ailleurs, la mise à l'épreuve d'un même corpus, dans plusieurs expériences, pendant une période de presque quinze ans représente un modèle pour la reproductibilité scientifique.

D'autres auteurs se sont intéressés à la contribution des voyelles versus des consonnes dans d'autres langues, par exemple Chen et al. (2013), explore cette contribution en Mandarin. Au moyen du corpus 'Madarin Speech Perception' (MPS) (Fu et al., 2011), ils sélectionnent 100 phrases qui sont toutes constituées de 7 mots monosyllabiques. A l'image des trois expériences précédentes, ils proposent deux conditions principales Consonnes seulement et Voyelles seulement, dans lesquelles la durée totale du segment ciblé est conservée. Les parties de segments 'masqués' sont remplacées par du bruit. A ces deux conditions s'ajoutent 4 sous-conditions pour chaque segment (consonnes vs. voyelles). Dans la condition Consonne, ils conservent la durée totale de la consonne et ajoutent une certaine proportion de la voyelle suivante (10, 20, 30 ou 40%), soit C+V^{10%} etc. A l'inverse dans la condition Voyelle c'est la durée de la voyelle cible qui est réduite (de 10, 20, 30 ou 40%), à chaque extrémité (soit 5% au début et 5% à la fin dans la condition où la durée de la voyelle masquée est égale à 10% de la durée totale du segment). Dans le protocole expérimental, les auditeurs écoutent et répètent ce qu'ils ont perçus. Les scores d'intelligibilité de phrase sont calculés en divisant le nombre de mot correctement reconnu par le nombre total de mots dans chaque phrase ciblée. Leurs résultats montrent que dans la condition Consonnes, les taux d'identification augmentent lorsque la consonne est suivie par une proportion de voyelle. Plus cette proportion est importante, meilleur le taux d'identification est. Dans la condition Voyelle, ce sont les segments les moins tronqués, donc les plus longs, qui présentent les meilleurs taux d'identification. De plus, la condition Voyelle seulement est mieux identifiée (99%) que la condition Consonne seulement (34,1%), et aucune différence significative n'est mise en exergue dans la comparaison entre la condition Voyelle seulement et V^{-10%}. Les auteurs concluent en faveur d'un avantage des voyelles, par rapport aux consonnes, avec un bénéfice estimé à trois pour un, pour les voyelles, dans l'intelligibilité de phrases en Mandarin.

Ces différentes études concluent que ce sont les voyelles qui contribuent davantage à l'intelligibilité de phrases en Anglais américain et en Chinois mandarin. Les auteurs relèvent toutefois une différence de contribution pour les voyelles : deux pour un en anglais américain et trois pour un en mandarin. Celle-ci peut s'expliquer par des caractéristiques linguistiques propres à la langue ciblée. En effet, le Mandarin est une langue à ton, cette tonalité porte sur les segments vocaliques, qui possèdent donc des propriétés acoustiques particulières, fines et porteuses de sens. En effet, les tons sont

¹³ "A two-to-one advantage for sentences containing only vowels compared to sentences that contained only consonants"

distinctifs, à l'image des traits acoustiques et articulatoires, et sont donc contrastifs au niveau lexical.

Les modèles d'expériences conduits dans les études présentées ci-dessus, démontrent la pertinence du segment comme unité d'évaluation de l'intelligibilité. Pour évaluer la contribution des consonnes versus des voyelles dans l'intelligibilité de la parole, les auteurs proposent de conserver tout (ou partie) des segments ciblés. Ils opèrent des modifications sur le signal acoustique préalablement enregistré. Cette méthodologie semble pertinente puisqu'elle permet d'atteindre le niveau segmental de la parole, à partir d'un enregistrement de phrase. Bien que les modifications sur le signal acoustique permettent d'atteindre le niveau segmental, le matériel linguistique de type phrastique est à l'origine pensé pour une évaluation globale de la parole. L'intelligibilité est alors entendue dans son sens large, puisque les scores sont calculés à partir de la transcription d'une phrase (contexte sémantique). Même lorsque les auteurs proposent des scores d'intelligibilité au niveau lexical (mots), ceux-ci sont produits enchâssés dans des phrases. Dans ces travaux, le recours aux informations de haut niveau est toujours possible et susceptible de parasiter l'évaluation. En outre, les effets de contextes et de coarticulation ne sont pas considérés.

La mobilisation des informations de haut niveau ainsi que les processus de restauration phonémique, sont toujours susceptibles de perturber l'évaluation de la parole, même au niveau segmental. Dans le contexte clinique qui est le nôtre, il est tout à fait pertinent de proposer une évaluation du niveau segmental, c'est-à-dire des unités de bas niveau qui le composent. Cependant, il est crucial de contrôler les possibles biais inhérents à l'évaluation perceptive. L'enjeu est ici de contraindre l'auditeur à s'appuyer principalement sur des informations de bas niveau, portées uniquement par le signal acoustique et de limiter le recours aux informations de haut niveau. Une évaluation de l'intelligibilité segmentale serait alors possible et adaptée à notre contexte clinique.

3.3. Evaluation de l'intelligibilité segmentale

Selon nous, deux approches peuvent être considérées pour évaluer l'intelligibilité segmentale. Une première approche repose sur une nouvelle méthode d'évaluation perceptive. Suivant le postulat selon lequel le recours aux informations de haut niveau contextuelles et lexicales peut être contrôlé, un nouveau test d'évaluation a été proposé dans le cadre du projet C2SI. Pour les besoins de ce test, un répertoire de pseudo-mots a été élaboré. Leur production par 40 locuteurs contrôle 87 patients traités pour un cancer ORL a permis la constitution d'un corpus de production de pseudo-mots. Celui-ci a fait l'objet de plusieurs études impliquant des tests de perception qui ont permis le recueil de transcriptions auprès d'auditeurs naïfs. Ces transcriptions permettent le calcul d'un score de Déviation Phonologique Perçue qui représente la distance inter-phonémique entre deux chaînes de caractères. Ce score considère l'écart entre une production et sa perception au niveau segmental. Ce matériel linguistique associé à cette méthode de calcul devrait permettre d'obtenir une évaluation fiable en termes de déficits

articulatoires en s'affranchissant des biais perceptifs susceptibles de perturber l'évaluation.

La seconde approche repose sur des mesures acoustiques réalisées sur le signal acoustique. Ces mesures portent sur les valeurs des premiers et seconds formants et ciblent spécifiquement les segments vocaliques. Elles permettent également le calcul de différentes métriques représentatives de certains aspects du système vocalique des locuteurs. Nous présenterons ci-après les principes de ces approches.

3.3.1. Méthode de Décodage acoustico-phonétique

Dans le cadre du projet de recherche Carcinologic Speech Severity Index¹⁴ (C2SI) une nouvelle méthode d'évaluation perceptive de l'intelligibilité a été développée. Celle-ci est basée sur la production de pseudo-mots évalués au moyen de tests de jugement perceptif de l'intelligibilité. Le recours à l'utilisation de pseudo-mots est motivé par la nécessité de contrôler les biais perceptifs précédemment explicités. La transcription de pseudo-mots place les auditeurs en situation de décodage acoustico-phonétique qui les contraint à évaluer uniquement les informations de bas-niveau i.e. les segments.

Dans cette partie, nous précisons le matériel linguistique élaboré en décrivant la constitution des pseudo-mots et les motivations qui soutiennent la création d'un tel répertoire ainsi que le corpus qu'ils permettent de constituer. Nous détaillerons ensuite la méthode de calcul des scores perceptifs de déviation phonologique perçue (PPD) sur lesquels seront basées une partie de nos analyses.

3.3.1.1. Stimuli

Le choix de générer un répertoire de pseudo-mots repose sur leurs spécificités. Ce matériel linguistique non lexical devrait permettre d'inhiber les biais fréquemment rencontrés dans les batteries de tests classiques d'évaluation de l'intelligibilité. La plupart des tests existants propose du matériel linguistique de nature lexicale. Nous l'avons souligné précédemment, les items lexicaux induisent un effet d'attraction lexicale qui se traduit par la restauration phonémique des items. Ainsi, si un patient produit le mot « château », mais que l'évaluateur sait que la liste contient le mot « gâteau », il va perceptivement reconstruire la séquence qu'il sait être juste, son évaluation sera alors sous-cotée. De plus, ces évaluations sont basées sur des listes, de mots ou de phrases, courtes et fermées dans lesquelles une centaine d'items est généralement disponible pour constituer une liste plus réduite de travail. Or, la répétition de la tâche par le clinicien conduit à l'apprentissage du matériel linguistique. Cela incrémente les effets de restauration lexicale. L'ensemble de ces effets perceptifs conduisent à une sous-évaluation des déficits linguistiques du patient et donc à lui proposer une évaluation qui reflète mal son handicap communicationnel.

Le parti pris de l'emploi de pseudo-mots apparaît pertinent pour neutraliser les effets de restauration lexicale puisqu'ils rendent le matériel linguistique dépourvu de correspondance, lexicale et sémantique, avec une forme stockée en mémoire. De plus, ils

¹⁴ Un descriptif plus détaillé du projet et tâches proposées est disponible en Annexe 2 (B, p196).

offrent, au travers du choix des règles phonotactiques fait pour les constituer, un potentiel immense en termes de quantité de matériel linguistique équilibré disponible pour une évaluation. Un grand nombre de formes possibles devrait permettre de neutraliser les effets d'apprentissage du matériel linguistique par l'évaluateur. A partir de l'inventaire des phonèmes d'une langue, il est possible de créer de nouvelles formes, inédites, qui respectent ses contraintes phonotactiques mais qui sont dépourvues de signification. C'est en quelque sorte un répertoire de potentielles formes sonores qui pourraient être réalisées par des locuteurs en contexte, lesquels feraient émerger une nouvelle forme lexicale en y associant un sens. Mais leur intérêt, quand ils sont employés dans une évaluation de parole, réside dans leur absence de signification, et la très grande quantité et diversité de formes qu'ils permettent de créer. En neutralisant les effets d'apprentissage et de restauration lexicales ils contraignent l'évaluateur à se focaliser sur des informations segmentales dans le signal de parole. L'évaluation devrait gagner en fiabilité, en termes, à la fois méthodologique et de prédiction du handicap communicationnel.

Le matériel linguistique sélectionné pour une évaluation de la parole supporte plusieurs enjeux. Il doit remplir des critères cliniques en termes de durée, de faisabilité, et de prédiction du handicap communicationnel, mais aussi des critères linguistiques et acoustiques : « Les items tests doivent avoir une base acoustique suffisamment détaillée pour permettre une explication d'un déficit d'intelligibilité en termes de déficits articulatoires spécifiques¹⁵ », (Weismer et al., 1988). En d'autres termes, les phonèmes doivent être pris dans des structures qui respectent les règles phonotactiques de la langue en termes de fréquence d'occurrence et de position afin de conserver les effets articulatoires pertinents et inhérents à la production de la parole. Ainsi, ils conservent leur capacité à représenter des déficits articulatoires rencontrés en contexte lexical.

Ainsi, en conséquence à ces nombreux biais perceptifs et cliniques l'équipe en charge de la constitution de cette tâche a choisi de créer un répertoire de pseudo-mots, pour le français, disponibles pour l'évaluation de l'intelligibilité segmentale. A partir de l'inventaire des phonèmes du français et du contrôle de leur fréquence d'occurrence et de leur position dans la langue, une sélection a été opérée. Ces unités phonétiques ont été combinées au sein d'unités plus larges composées de deux syllabes, selon la structure CVCV ou C(C)VC(C)V

Les phonèmes et la place qu'ils peuvent occuper sont les suivants :

C1 ∈ {p t k b d g v z ʒ f s ʃ r l m n ñ yod} / **CC1** ∈ {pr tr kr gr br fr pl kl fl st bl sk sp gl dr ps}

V1 ∈ {a i y u o e ã ë}

C2 ∈ {p t k b d g v z ʒ f s ʃ r l m n ñ yod} / **CC2** ∈ {st ks rd rs kt rn pl gr dr kl rj lt rv vr gz rp tr rt bl rm pr kr sk br sp rk fr fl rb gl ps pt}

V2 ∈ {a i y u o ã ë}

¹⁵ "Ultimately, the test items should have a sufficiently detailed acoustic basis to permit an explanation of an intelligibility deficit in terms of specific articulatory deficits"

Ainsi, 92 200 formes ont été générées. 5854 correspondaient à des formes lexicales sur le plan phonologique et ont été éliminées. Il est resté 89 346 formes validées en tant que pseudo-mots. Plusieurs publications détaillent la création de ces pseudo-mots (Lalain et al., 2020; Rebourg, 2017). Voici un exemple non exhaustif de pseudo-mots : *bitant, pluniou, liba, brana, favo, lefro, supant...* Davantage d'exemples sont disponibles dans la section 3.3.1.1 Stimuli. Cet ensemble de pseudo-mots, destinés à l'évaluation de la parole en contexte clinique, offre un grand potentiel pour la création de listes de travail différentes et en très grand nombre, de sorte qu'elles ne puissent pas être apprises. Pour cela, et grâce à l'automatisation du processus, chaque liste est générée de façon aléatoire, en suivant des contraintes de fréquence d'occurrence des phonèmes dans chaque liste pour qu'un ensemble de phonèmes pertinents soient systématiquement évalués. Une liste est constituée de 52 pseudo-mots, dans la version initiale du test.

En contexte expérimental, pour que les locuteurs puissent être évalués, ils doivent préalablement être enregistrés. Ainsi, les locuteurs du corpus DAP ont été enregistrés dans un environnement calme. Les stimuli leur ont été présentés via la station PercEval (André et al., 2003) selon deux modalités simultanées : visuelle, affichée sur un écran d'ordinateur et auditive, dans un casque audio. Cette double modalité devait permettre de pallier les éventuels troubles de la vision, de l'audition, de lecture ou de l'attention qui pourraient affecter les patients, souvent âgés. Tous les locuteurs ont été enregistrés par un micro TLM102 et un enregistreur Zoom Ph4. Dans le cadre du projet C2SI, ces listes d'items ont été produites par des locuteurs patients, traités pour un cancer de la cavité buccale ou de l'oropharynx, et contrôles. Soit 52 pseudo-mots x 127 locuteurs = 6 604 stimuli enregistrés. La liste de pseudo-mots produite par un locuteur a été enregistrée sur une unique piste sonore. Pour les besoins expérimentaux, chaque production a été associée avec son pseudo-mot cible. La préparation des stimuli en vue de leur évaluation a requis qu'ils soient étiquetés un à un. Pour cela, nous avons utilisé le logiciel d'annotation Praat. Nous avons aligné chaque production avec sa transcription cible, et extrait chaque production de façon que chaque pseudo-mot constitue un fichier son individuel. Les détails sur les listes de pseudo-mots enregistrées seront précisés en parties 4.1.1.6 et 4.1.4.

Ces stimuli (6 604 pseudo-mots) ont été utilisés dans des tests de jugement perceptif de l'intelligibilité auprès d'auditeurs. De façon très générale, ces tests perceptifs se déroulent selon un protocole qui consiste à recruter des auditeurs qui reçoivent une consigne sur la tâche à réaliser et qui, au travers de leur réponse à la tâche, permettent de réaliser des études sur la parole. Dans un test de jugement perceptif de l'intelligibilité, l'auditeur est équipé d'un casque audio, et d'un clavier. Placé devant un écran d'ordinateur sur lequel est installée la station d'évaluation PercEval, il écoute les productions et est chargé de transcrire le plus fidèlement possible ce qu'il a perçu au moyen d'une transcription orthographique (4.2.3 Méthodologie d'un test de perception).

Les stimuli ont été présentés aux auditeurs selon un ordre préétabli, de façon à limiter les biais expérimentaux, ils sont transcrits par plusieurs auditeurs différents et chaque stimulus est présenté à différents moments de l'expérience pour contrebalancer un effet de fatigabilité auditive par exemple. De plus, nous procédons à un brassage des stimuli

entre les groupes de locuteurs (patients vs contrôles) pour que l'auditeur soit confronté à plusieurs locuteurs. Au cours de notre master, nous avons notamment participé au recueil de données par l'enregistrement de sujets contrôles, à la préparation des stimuli (Rebourg, 2017) ainsi qu'à la passation des tests de perception et à l'analyse des résultats (Rebourg, 2018).

En plus de ce matériel linguistique riche, diversifié et contrôlé, l'équipe du LPL a élaboré une méthode de calcul innovante, au travers d'une matrice de confusion phonémique. Cette matrice repose sur la théorie des traits articulatoires et attribue un coût en termes de distance, entre la cible qui devait être produite et la cible perçue, en nombre moyen de traits altérés par phonème (Ghio et al., 2018a).

3.3.1.2. Méthode de calcul des scores perceptifs

La notion de trait distinctif a été réinvestie dans la méthode de calcul des scores de Déviation Phonologique Perçue. En effet, elle permet une analyse fine des unités phonémiques. Au moyen d'une matrice de confusion basée sur la théorie des traits distinctifs, les auteurs proposent une mesure objective et pondérée en termes de nombre de traits moyens altérés par phonème.

L'idée de pouvoir mesurer une distance « concrète » entre une cible linguistique attendue et produite, et sa perception n'est pas nouvelle, et peut être abordée sous plusieurs axes différents. Des mesures des différences spectrales observables sur le signal acoustique, aux mesures élaborées à partir de théories linguistiques, l'idée émergente définit cette mesure au moyen d'un index. Dès les années 50, Miller et al. (1951), évoquent l'utilisation des « différences connues dans les spectres des sons pour construire un indice de la distance entre les sons de la parole, mais cet indice n'est pas encore disponible ». Dans cet esprit d'index à dessein de mesure de distance entre unités, la méthode de calcul des scores de Déviation Phonologique Perçue développée dans le cadre de la tâche de DAP du projet de recherche C2SI est innovante. Elle est le fruit d'une réflexion de plusieurs années, puisque dès ses travaux de thèse Ghio (1997) mûrit l'idée de l'élaboration d'une matrice de coût entre unités linguistiques. Ce projet de recherche a ainsi offert un cadre au développement d'une méthode de calcul basée sur la théorie des traits articulatoires (Chomsky & Halle, 1968). Au moyen d'une matrice de coût et de l'algorithme de Wagner-Fisher qui permet de calculer la distance de Levenshtein entre deux chaînes de caractères, ce calcul offre un score de distance en nombre de traits moyen altérés par phonème. Ce score analogique, bien plus précis que les scores binaires classiquement obtenus, permet des nuances et une finesse d'analyse tout à fait pertinentes notamment pour identifier précisément les altérations chez un patient donné. Il s'agira donc ici de comparer la cible qui devait être produite par un locuteur avec la transcription d'un auditeur au niveau phonétique.

L'utilisation de la matrice de coût implique que les réponses des auditeurs soient phonétisées, ceci afin d'établir une correspondance adéquate entre la cible phonétique qui devait être produite et la séquence phonologique qui a été perçue. Ainsi, nous ne pouvons faire l'impasse sur l'explication du processus de phonétisation employé dans ces travaux et illustré dans le schéma Figure 7. Les stimuli élaborés pour la tâche de DAP sont

basés sur les phonèmes du français, ils respectent les règles phonotactiques de la langue et sont contraints en termes de fréquences d'occurrence et de position de chacun des phonèmes. Ces stimuli phonétiques ont dû être transformés en une forme orthographique pour pouvoir être produits par les locuteurs (modalité de présentation des stimuli à produire audio-visuelle). Pour cela, une double phonétisation a lieu, la première transcrit les stimuli selon les conventions de transcription de Lexique.org qui permettent un codage sur un seul caractère et sans caractères spéciaux, facilitant le traitement au travers de plusieurs logiciels ayant des systèmes d'exploitation différents. Ensuite, les données sont envoyées à nos collaborateurs du LIA (Avignon) afin d'être phonétisées selon les conventions de transcriptions LIA_Phon (Bechet, 2001). Nous récupérons cette première phonétisation puis y appliquons les conventions de transcription Lexique.org en vue de la seconde phonétisation de la réponse de l'auditeur. Ainsi, nous avons d'un côté la cible attendue phonétisée et de l'autre la réponse de l'auditeur phonétisée. A présent, c'est au moyen de la matrice de coût, présentée dans la section suivante, que nous comparons la cible qui devait être produite avec la cible qui a été perçue.

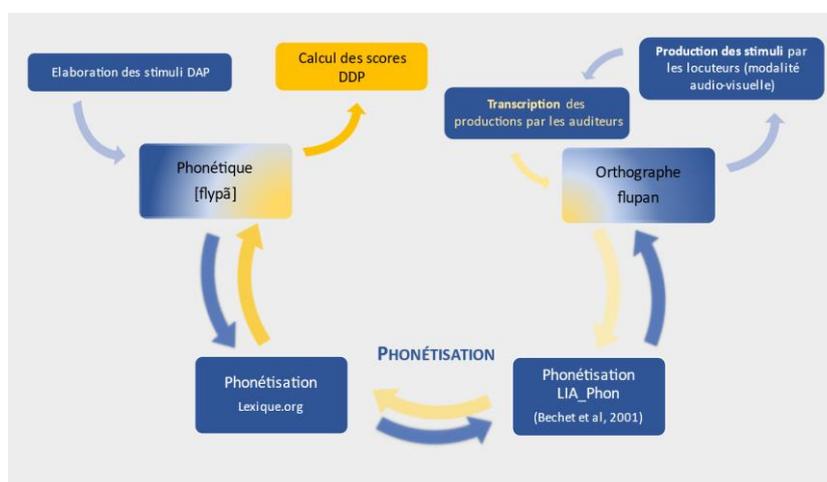


Figure 7 : Schéma du processus de phonétisation des stimuli et des réponses de auditeurs

Deux matrices de coût (une pour les segments vocaliques et la seconde pour les segments consonantiques) rendent compte du poids de chaque segment les uns envers les autres. Le poids est mesuré par la distance en termes de traits articulatoires qui séparent la cible produite de la cible perçue. Les deux matrices (segments vocaliques et consonantiques) sont construites selon les mêmes principes, et opèrent selon le même fonctionnement. Elles permettent de mesurer la distance en nombre de traits articulatoires moyens, entre deux cibles. Nous présentons plus en détails la matrice de coût des segments vocaliques car nous leurs accorderons un intérêt particulier dans ces travaux. Ainsi, nous l'utiliserons pour présenter les principes généraux de fonctionnement de la méthode de calcul de score de Déviation Phonologique Perçue. Pour obtenir une description détaillée de l'élaboration et du fonctionnement de la méthode de calcul des scores nous recommandons au lecteur de se référer à la publication consacrée (Ghio et al., 2018a).

La décomposition en trait distinctifs des phonèmes du français, d'après Chomsky & Halle (1968), permet de « dresser une matrice de distances » pour les phonèmes

vocaliques et consonantiques « par comptage du nombre de traits différents entre chaque phonème » (Ghio et al., 2018a). Les matrices attribuant un coût à chaque différence sont symétriques et celle construite pour les segments vocaliques est représentée dans la Figure 8. Ainsi, lorsque la cible produite est correctement identifiée, la distance entre la cible produite et la cible perçue est nulle. Plus le segment perçu est éloigné de sa cible plus la distance augmente. Cette matrice présente les coûts de chaque substitution entre segments vocaliques.

	a	i	u	o	e	y	ø	ɛ	ɔ	œ	ã	ẽ	õ	Ê	Ô	Û	μ	&	
a	0	3	3	2	2	4	3	1	1	2	1	2	2	3	1	1	2	2	1
i	3	0	2	3	1	1	2	2	4	3	4	3	5	4	1	3	2	3	1
u	3	2	0	1	3	1	2	4	2	3	4	5	3	4	3	1	2	4	2
o	2	3	1	0	2	2	1	3	1	2	3	4	2	3	2	0	1	3	1
e	2	1	3	2	0	2	1	1	3	2	3	2	4	3	0	2	1	2	0
y	4	1	1	2	2	0	1	3	3	2	5	4	4	3	2	2	1	3	1
ø	3	2	2	1	1	1	0	2	2	1	4	3	3	2	1	1	0	2	0
ɛ	1	2	4	3	1	3	2	0	2	1	2	1	3	2	0	2	1	1	0
ɔ	1	4	2	1	3	3	2	2	0	1	2	3	1	2	2	0	1	2	1
œ	2	3	3	2	2	2	1	1	1	0	3	2	2	1	1	1	0	1	0
ã	1	4	4	3	3	5	4	2	2	3	0	1	1	2	2	2	3	1	2
ẽ	2	3	5	4	2	4	3	1	3	2	1	0	2	1	1	3	2	0	1
õ	2	5	3	2	4	4	3	3	1	2	1	2	0	1	3	1	2	1	2
œ̃	3	4	4	3	3	3	2	2	2	1	2	1	1	0	2	2	1	0	1
Ê	1	1	3	2	0	2	1	0	2	1	2	1	3	2	0	2	1	1	0
Ô	1	3	1	0	2	2	1	2	0	1	2	3	1	2	2	0	1	2	1
Û	2	2	2	1	1	1	0	1	1	0	3	2	2	1	1	1	0	1	0
μ	2	3	4	3	2	3	2	1	2	1	1	0	1	0	1	2	1	0	1
&	1	1	2	1	0	1	0	0	1	0	2	1	2	1	0	1	0	1	0

Figure 8 : Matrice de coût des segments vocaliques, en nombre de traits différents entre les segments. Ê, Û et Ô sont des archiphonèmes correspondant respectivement à {e, ɛ}, {œ, ø} et {o, ɔ} (Ghio et al., 2018a).

Pour opérer la comparaison entre la cible qui devait être produite et la cible qui a été perçue c'est l'algorithme de Wagner-Fischer qui a été utilisé. Celui-ci permet de prendre en compte les phénomènes d'élision, d'insertion et de substitution. Une élision est repérée lorsque aucun phonème n'a été transcrit à la place d'un phonème attendu. A l'inverse, une insertion est qualifiée par l'ajout d'un phonème là ou un seul phonème est attendu, en d'autres termes, deux phonèmes sont transcrits alors que la cible est unique. Enfin, une substitution se caractérise par le choix d'un phonème différent de celui attendu. La considération de ces phénomènes au travers de l'algorithme permet de leur attribuer un poids.

L'algorithme est employé pour calculer une distance de Levenshtein, entre deux chaînes de caractères. Il s'agit ici d'unités phonétiques, non d'unités orthographiques, comme cela est classiquement le cas pour cette mesure (Levenshtein). En effet, dans la comparaison entre phonèmes il est possible d'apporter des nuances entre le poids de chaque segment les uns par rapports aux autres, ainsi la comparaison entre /i/ et /y/, segments antérieurs, fermés et respectivement étiré et arrondi aura moins de poids que la comparaison entre /i/ antérieur, fermé et étiré et /õ/ segment postérieur, mi ouvert, arrondi et nasalisé, puisque plus de traits articulatoires les opposent. /i/ et /y/ s'opposent sur la position des lèvres, soit un trait d'écart et /i/ et /õ/ s'opposent également sur la position de lèvre (1 trait), mais aussi sur l'antériorité (1 trait), l'aperture (2 traits) et la nasalité (1 trait), soit 5 traits d'écart au total. La considération du niveau phonémique permet également d'attribuer plus de poids à une comparaison entre un segment

vocalique et un segment consonantique qu'à une comparaison entre deux segments vocaliques. Ainsi, le mécanisme de comparaison entre deux chaînes de caractères phonémiques (conventions de transcriptions Lexique.org) est illustré dans la Figure 9. Sur l'axe des abscisses on retrouve la cible attendue et sur l'axe des ordonnées, la réponse, sous forme de transcription phonétisée, de l'auditeur. L'exemple ici proposé illustre l'ensemble des coûts locaux calculés.

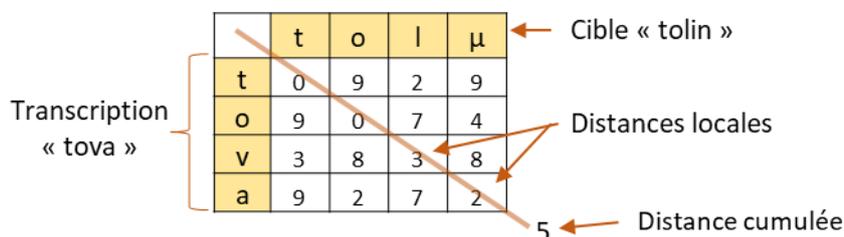


Figure 9 : Exemple de comparaison entre les deux chaînes de caractères que constituent la cible et la transcription de l'auditeur (/μ/ = /ɛ/)

Cette comparaison entre la cible qui devait être produite et la cible qui a été perçue permet d'obtenir une distance cumulée, celle-ci est divisée par le nombre de phonèmes qui constituent la cible pour obtenir le score global par item produit. Le score de Déviation Phonologique Perçue global de chaque locuteur représente la moyenne des scores obtenue par chaque production. Ainsi, les scores globaux par auditeurs sont calculés comme suit (se référer à la section méthodologie d'un test de perception dans lequel différents auditeurs transcrivent le même stimuli) : à partir des 3 transcriptions des auditeurs pour un stimuli produit par le même locuteur un score de Déviation Phonologique Perçue moyen est obtenu pour chaque stimulus produit. A partir du score moyen de chaque stimulus produit (n=52), un score global moyen est calculé par locuteur. Cette méthode de calcul a été automatisée et permet d'obtenir un score global de Déviation Phonologique Perçue pour chaque locuteur, exprimé en nombre moyen de traits altérés par phonème et noté sPPD global.

Ce score global calculé sur l'ensemble des productions d'un locuteur a permis le calcul d'un score local, c'est-à-dire d'une distance phonème à phonème qui rend possible l'examen particulier de chaque segment. Ainsi, nous avons manifesté un intérêt particulier pour ces scores locaux dont le calcul n'a été possible qu'après des modifications dans le code du système de traitement automatique par Alain Ghio, ingénieur développeur de ce calcul. Les scores de Déviation Phonologique Perçue Locale permettent d'obtenir en détail les scores de distances locales phonème à phonème ainsi que les correspondances des transcriptions phonétisées des auditeurs avec le pseudo-mots phonétisé ciblé pour chaque phonème. Le niveau segmental qu'ils permettent de considérer revêt un intérêt majeur dans le cadre de l'évaluation clinique de la parole pathologique.

Cette tâche de production et sa méthode de calcul des scores ont fait l'objet de plusieurs évaluations dans le cadre de l'évaluation et de la validation des critères de pertinences et d'objectivité de la tâche de DAP. Ghio et al. (2020), ont montré que ce matériel linguistique

permet de distinguer deux groupes de locuteurs patients et control, ce qui montre la capacité des pseudo-mots à discriminer deux populations. De plus, ils ont défini un seuil au-delà duquel la parole produite par le locuteur est considérée comme dysfonctionnelle : à 0,6 traits moyens par phonème la déviation produite est telle qu'elle peut être considérée comme pathologique. Une étude, dont l'objectif était d'évaluer le maintien de la pertinence des scores attribués au moyen de la tâche de DAP, a montré que le nombre d'item par liste peut être réduit de 52 à 16 en continuant de proposer des scores hautement pertinents. Cette réduction est opérée au travers de critères phonologiques appliqués aux listes d'items (Marczyk et al., 2021; Marczyk et al., 2020). Dans un mémoire explorant les réponses des auditeurs à deux sessions de tests de jugement perceptifs de l'intelligibilité, Rebourg (2018) montre un effet d'attraction lexicale faible (~6%), mais inévitable, dans les transcriptions des auditeurs. Celle-ci ne porte pas préjudice à l'évaluation de la parole proposée au travers de la tâche de DAP. Dans un protocole visant à tester l'effet de mémorisation des items, aucun effet de mémorisation n'a été démontré ce qui montre que ce matériel linguistique neutralise l'effet dans cette condition. L'ensemble de ces résultats accrédite la pertinence de ce matériau linguistique non lexical, les pseudo-mots, disponible en grande quantité, proposé pour une évaluation de la parole, spécifique au contexte clinique.

Cette proposition pour une évaluation au moyen d'une méthode de décodage acoustique-phonétique, basée sur des pseudo-mots et sur une méthode de calcul de scores de Déviation Phonologique Perçue, devrait permettre de neutraliser les effets de restauration lexicale en contrôlant le recours au haut niveau (lexical). Néanmoins, elle ne permet pas de s'affranchir de l'évaluation perceptive, qui nous l'avons souligné, constitue le gold standard. Cependant, ce standard n'exclut pas qu'une autre approche soit envisageable. Suivant l'idée selon laquelle on peut s'affranchir de la perception d'un auditeur pour proposer une évaluation qui permet de caractériser le système vocalique du locuteur, les variables acoustiques apparaissent comme de bons candidats.

Dans la partie suivante, nous rendrons compte de plusieurs études qui examinent la relation entre des scores d'intelligibilité segmentale vocalique et des variables acoustiques. Les valeurs de F1 et F2 peuvent être considérées en tant que telles ou réinvesties dans le calcul de métriques.

3.3.2. Contribution des variables acoustiques à l'intelligibilité

Les études évaluant la contribution des consonnes versus des voyelles à l'intelligibilité en anglais et en mandarin (voir 3.2 Evaluation perceptive segmentale) suggèrent que cette conclusion pourrait être étendue au français. Ainsi, il n'est pas exclu de considérer que les segments vocaliques demeurent de bons candidats pour caractériser l'intelligibilité des locuteurs. Cette caractérisation devrait permettre de préciser les déficits articulatoires. Nous soulignons ici que le protocole expérimental employé influence la façon dont l'auditeur perçoit et décode le signal acoustique de parole. Il doit être contrôlé pour que les méthodes expérimentales permettent d'appréhender les

structures fines de la langue. De plus, Bradlow et al. (1996) suggèrent que l'association « des mesures acoustiques-phonétiques et des données perceptives pourraient conduire à une meilleure compréhension des principales caractéristiques acoustiques et phonétiques auxquelles les auditeurs réagissent lors de la perception de la parole, et aiderait donc à différencier la parole hautement intelligible de la parole moins intelligible¹⁶ ».

Les études présentées dans cette partie confrontent des scores d'intelligibilité avec des mesures acoustiques portant sur le signal de parole en mesurant l'interaction entre des résultats perceptifs (scores intelligibilité) et des mesures objectives (mesures acoustiques). Tout d'abord, nous observons un accord de consensus sur les mesures effectuées : des mesures de valeurs formantiques, prises à un point particulier de la durée du segment (au milieu par exemple), représentant une valeur moyenne à partir de 3 points particuliers ciblés sur le signal (au milieu du premier tiers, puis de second tiers, puis du troisième tiers de la durée total du segment ciblé), ou encore au moyen de représentations plus dynamiques qui dressent des courbes de l'évolution des valeurs formantiques au cours de la durée du segment. Les valeurs formantiques s'imposent comme de bons représentants pour une étude acoustique des voyelles.

3.3.2.1. F1/F2

Dans cette section nous présenterons plusieurs études dont l'objectif était de confronter des scores d'intelligibilité avec des mesures acoustiques de valeurs de formants. Les valeurs formantiques entretiennent une relation étroite avec la physiologie du système de production de la parole, leur fréquence est influencée par les dimensions physiques, la configuration du conduit vocal et particulièrement celle de la cavité buccale dont l'aperture modifie la dimension de la cavité pharyngale. Les formants représentent les bandes d'énergie ou les régions spectrales amplifiées sur le signal acoustique. Leurs dimensions sont associées à la position des articulateurs, la valeur de F1 est associée avec la hauteur de la langue dans la cavité buccale et donc l'aperture de la mandibule et celle de F2 correspond à l'antériorité/postériorité de l'apex ou du corps de la langue dans la cavité buccale (Figure 6 : Illustration du système vocalique du français). Leurs valeurs de références présentent des différences entre les hommes et les femmes (Tableau 1 : Valeurs formantiques de référence utilisées pour le français (Gendrot & Adda-Decker, 2005a)).

La corrélation entre une mesure de score d'intelligibilité et une variable acoustique peut s'interpréter de deux façons, d'après Weismer et al. (2001) : (i) les variables acoustiques impactent directement l'intelligibilité de la parole, de façon que la modification des premiers change les seconds, dans cette posture les variables acoustiques peuvent être considérées comme une composante intégrale du déficit d'intelligibilité ; (ii) les variables acoustiques sont un indice des mécanismes globaux

¹⁶ "combining both acoustic-phonetic measurements with perceptual data might lead to a better understanding of the salient acoustic-phonetic characteristics that listeners respond to during speech perception, and would therefore help to differentiate highly intelligible speech from less intelligible speech"

impliqués dans parole, en ce sens elles ne peuvent être considérés comme un élément de valeur dans les déficits d'intelligibilité. En d'autres termes ils suggèrent qu'une variable acoustique a soit une valeur de représentation d'indices très fins, de bas niveau acoustique, ou soit une valeur de représentation d'indices plus généraux. Il est alors envisageable d'attribuer une valeur de prédiction des déficits d'intelligibilité à (i) mais on ne pourrait en faire autant avec (ii) qui représenterait plutôt une tendance globale. Toutefois, nous postulons que ces deux considérations ne sont pas incompatibles, les variables acoustiques peuvent revêtir à la fois un caractère général, quand les mesures portent sur des cohortes de locuteurs importantes, et très fins, lorsqu'elles sont appliquées à des études de cas, et plus particulièrement de cas cliniques.

Les études conduites dans la parole saine s'incrémentent davantage dans la vision globale (ii) (Weismer et al., 2001). Préalablement à cette vision dualiste Weismer et al. (1988) ont conduit une étude portant la signature acoustique des mots dans un test d'intelligibilité. 30 locuteurs, 15 hommes et 15 femmes, (âgés de plus de 65 ans sans problème médicaux connus) ont été enregistrés sur la production d'une liste de 98 mots. A partir des enregistrements, ils évaluent la trajectoire des formants F1 et F2, soit l'évolution de la fréquence (en Hz) au cours du temps de la durée de la voyelle (en ms). Ils tracent les courbes représentant ces trajectoires, mot à mot pour chaque locuteur. Ils regroupent ensuite les locuteurs en fonction de leur sexe, homme et femme, et tracent une représentation moyenne par mot pour chaque groupe. Ils comparent les trajectoires de formants mot par mot en fonction de chaque groupe de locuteur. Ils observent que les femmes produisent des pentes de F2 plus grandes que les hommes et il semble que F1 ne soit que peu affecté par le sexe. Les résultats de cette étude constituent probablement un prélude à la dichotomie que les auteurs développeront une décennie plus tard (Weismer et al., 2001).

D'autres part, Fogerty & Humes, (2012) optent pour d'autres variables acoustiques (la fréquence fondamentale (F0), l'enveloppe et la structure fine temporelle) pour établir leur contribution à l'intelligibilité. Les propriétés de ces variables sont communes aux consonnes et aux voyelles, et les auteurs s'intéressent à leurs spécificités dans les mots et les phrases. Leurs résultats montrent notamment que l'amplitude de la F0 représente un indice important pour la contribution des voyelles par rapport aux consonnes dans l'intelligibilité des phrases, mais pas dans le contexte de production de mots isolés. Cette étude souligne à nouveau la contribution des voyelles à l'intelligibilité dans la parole saine. Elle s'inscrit également dans la vision plus globale caractérisant l'intelligibilité au sens large.

En revanche, les études portant sur les troubles de la production de la parole s'inscrivent davantage dans la conception selon laquelle les variables acoustiques possèdent un caractère de prédiction vis-à-vis de l'intelligibilité, ou du moins de précision.

Par exemple, Ishikawa et al. (2019), ont mené une étude portant sur des locuteurs atteints de troubles de la production de la parole d'origine pathologique. 20 locutrices, dont 10 contrôles et 10 patientes atteintes de dysphonie ont été enregistrées. La

dysphonie est un trouble de la voix séquellaire, suite à la présence ou à la résection d'un nodule ou d'un polype sur une corde vocale par exemple. Les 20 locuteurs ont produit 11 voyelles de l'anglais américain, enchâssées dans la structure /h/ voyelle /d/, le tout produit dans la phrase porteuse « I'm going to say /h/ V /d/ » (ndl : « je vais dire ... »). 10 auditeurs naïfs, natifs de anglais américain ont été recrutés. Ils ont réalisé une tâche d'écoute avec une réponse sous forme de choix forcé. Il leur était demandé de « sélectionner le mot qu'ils entendent à partir des propositions disponibles sur l'écran ». Les auteurs calculent un score d'intelligibilité à partir du nombre de voyelles correctement identifiées, en pourcentage. D'autre part, ils calculent les valeurs de F1 et F2 sur la portion « stable » de la voyelle, toutes les 6,25 millisecondes.

Ils soumettent deux conditions aux auditeurs, calme versus bruité. Leurs résultats indiquent que les différences entre F1 et F2 dans les deux groupes de locuteurs (dysphonique et control) prédisent significativement la différence d'intelligibilité en condition calme (mais pas en condition bruitée). Ils montrent également que F2 prédit significativement l'intelligibilité des voyelles. Ils suggèrent que « plus la différence entre F2 est grande, plus la différence d'intelligibilité sera petite¹⁷ » (Ishikawa et al., 2019). Ils explorent également l'impact de la dysphonie sur l'axe d'aperture de la mandibule et montrent que les voyelles basses sont plus sensibles à la dysphonie que les voyelles hautes en anglais américain. En effet, les scores d'intelligibilité se sont révélés presque parfait (99%) pour les voyelles hautes en condition calme, en revanche « la voyelle basse d'arrière /ɑ/ avait l'intelligibilité la plus faible (58%), et la voyelle /ɔ/ avait la deuxième intelligibilité la plus faible (66%)¹⁸ » (Ishikawa et al., 2019). Ceci peut, pour partie, être expliqué par le fait qu'en anglais américain, les voyelles hautes sont plus distinctes (/i/ et /u/). La plage de F2 est donc maximisée et la confusion entre les deux presque inexistante. De plus, les configurations « ouverte » et « mi-ouverte » comprennent plus de voyelles (n = 5) ce qui entraîne probablement plus de confusion puisqu'elles sont plus proches sur le plan acoustique. Ceci, conformément aux propositions de la théorie Dispersion-focalisation (Liljencrants & Lindblom, 1972) serait susceptible d'impacter les scores d'intelligibilité.

De plus, cette étude porte sur un trouble fonctionnel de la production de la voix, différent des dysarthries (troubles neurologiques) plus couramment étudiées. Elle met en lien des mesures d'intelligibilité sur les voyelles avec des mesures acoustiques sur le signal de parole. Les auteurs suggèrent en effet que les indices acoustiques portés par F2 contribuent davantage à l'intelligibilité des voyelles que ceux portés par F1. Ils montrent également que des indices acoustiques très fins peuvent revêtir un caractère prédictif. Cette étude se situe davantage dans la vision selon laquelle les variables acoustiques constituent une « composante intégrale du déficit d'intelligibilité ».

Dans cette partie, nous avons montré que les valeurs de F1 et F2 constituent des représentations acoustiques pertinentes pour préciser une mesure de l'intelligibilité au

¹⁷ "the larger the difference in F2 is, the smaller the intelligibility difference would be"

¹⁸ "the low-back vowels /ɑ/ had the lowest intelligibility (58%), and the vowel /ɔ/ had the second-lowest intelligibility (66%)"

regard du signal acoustique. Dans la partie suivante nous présenterons quelques exemples de métriques qui peuvent être calculées sur la base de ces valeurs de F1 et F2.

3.3.2.2. Indices

Les valeurs des F1 et F2 relevées sur le signal acoustique peuvent être employées dans le calcul de métriques reflétant différents degrés d'information. Ces indices permettent de proposer des représentations du système dans sa globalité ou au regard de chaque catégorie de segment vocalique. L'espace vocalique renseigne sur la forme et la dimension du système ; l'indice de distinctivité représente la distinction entre les différentes catégories de voyelles. Celles-ci sont d'abord établies à partir de parole saine et peuvent ensuite être mises à l'épreuve sur différents types de parole (pathologie, comparaisons inter-langues). Dans cette partie nous présenterons deux métriques, d'abord l'aire du triangle vocalique (VSA) qui permet de caractériser le système dans sa globalité, puis l'indice de distinctivité qui permet de caractériser la distinction entre chaque catégorie de voyelle.

3.3.2.3. Vowels space area

Un exemple de l'utilisation des mesures acoustiques pour caractériser des locuteurs sains réside dans l'étude de Bradlow et al. (1996). Ils ont sélectionné 100 phrases différentes), chacune produite par 20 locuteurs, 10 hommes et 10 femmes (corpus Indiana Multi-Talker Sentences Database (Karl & Pisoni, 1994)), pour un total de 2000 phrases produites. Chacune des phrases comprend 5 mots clefs associés à des mots de fonction. Ce corpus intègre les données d'intelligibilité de la parole sous forme de transcription des phrases par 10 auditeurs par locuteurs, soit un total de 200 auditeurs différents (natifs de l'anglais américain, étudiants à l'université d'Indiana). Autrement dit, chaque groupe de 10 auditeurs écoute le set complet de 100 phrases produit par un seul locuteur. Ils transcrivent ce qu'ils ont entendu sur un clavier d'ordinateur. Les transcriptions de phrases sont scorées sur un critère 'mots clefs' qui considère une phrase correctement transcrite si et seulement si les 5 mots clefs sont correctement transcrits. Toute erreur de transcription d'un mot clef et l'ensemble de la phrase est considérée comme mal transcrite. Le score d'intelligibilité globale représente le score moyen obtenu pour les 100 phrases produites par le locuteur. Leurs résultats montrent que les femmes sont généralement plus intelligibles que les hommes, et qu'une meilleure intelligibilité corrèle avec une amplitude de F0 plus large. En revanche, elle ne montre aucune corrélation avec le débit de parole et la moyenne de la F0. Les auteurs relèvent toutefois une corrélation négative entre l'intelligibilité et la réduction de l'espace vocalique : ainsi un espace vocalique réduit entraîne une baisse d'intelligibilité contrairement à un espace vocalique plus large. En ce sens, ils suggèrent que l'espace vocalique représenterait un indicateur de l'intelligibilité globale. Cela suggère que la dimension de l'espace vocalique impacte la distinction entre les phonèmes, plus l'espace est large plus les catégories pourront être distinctes les unes des autres. Dans un espace plus réduit moins de distinction et plus de recouvrement sera probablement observé entre les différentes catégories de voyelles. Cela semble particulièrement vrai pour la dimension F1 qui présente des scores d'intelligibilité globalement plus élevés.

D'autres auteurs proposent d'aborder la relation entre la dimension du système vocalique et l'intelligibilité dans une étude comparative de la production de voyelles dans des phrases, auprès de 3 groupes de locuteurs : un groupe contrôle et deux groupes de locuteurs dysarthriques, des patients atteints de sclérose latérale amyotrophique (SLA) et de la maladie de Parkinson (Park). Weismer et al. (2001) ciblent les valeurs de F1 et F2 des quatre voyelles cardinales de l'anglais (/i/, /u/, /ɑ/, et /æ/) et réalisent des mesures de formants. A partir de ces mesures ils proposent des représentations de l'aire des espaces vocaliques de chaque locuteur. Ils regroupent les représentations en fonction du sexe des locuteurs hommes versus femme. Ils réalisent également un test de perception auprès de 10 auditrices natives de l'anglais, qui ont pour consigne de répéter la phrase perçue. Ces mesures sont complétées par une évaluation de l'intelligibilité focalisée sur le critère de 'précision articulatoire' et une évaluation de la sévérité plus globale. Leurs résultats montrent que globalement, les hommes présentent des espaces vocaliques plus réduits que ceux des femmes. Cette différence se retrouve dans le groupe SLA, les femmes présentent un espace vocalique plus large que les hommes (Seulement une locutrice dans le groupe Park). D'autre part, les locuteurs du groupe contrôle présentent un espace vocalique plus ample, et plus de distinction entre les catégories de voyelles par rapport aux locuteurs dysarthriques. Ces derniers tendent à avoir un espace vocalique plus compressé, mais cette compression n'est pas toujours une réduction simple et symétrique de l'espace vocalique et relève de phénomènes plus subtils et plus complexes. Enfin, concernant l'évaluation de l'intelligibilité, c'est le groupe contrôle qui obtient la meilleure intelligibilité par rapport aux deux groupes de locuteurs dysarthriques. Parmi eux, le groupe SLA obtient une meilleure intelligibilité que le groupe Park. La corrélation entre l'espace vocalique et l'échelle d'intelligibilité ($r = 0.68$ et $p < 0.05$) est positive. Les auteurs suggèrent tout de même que les mesures acoustiques prises à partir de mots isolés peuvent être un prédicteur de l'intelligibilité au niveau de la phrase et inversement.

Un autre indice pourrait se révéler complémentaire, à la fois des mesures d'intelligibilité et de la dimension du système vocalique, c'est l'indice de distinctivité. Nous évoquerons d'abord l'indice Phi duquel il s'inspire.

3.3.2.4. Distinctivité

L'indice Phi (Huet & Harmegnies, 2000) caractérise la distance entre un point dressé dans le plan F1/F2, pour chaque voyelle ou pour le centre de gravité de la catégorie ciblée, et le centre de gravité du système vocalique. L'indice ainsi calculé représente le rapport entre le point central du système et un autre point représentant une voyelle ou la moyenne de la catégorie de voyelle. Cet indice a notamment été mis à l'épreuve par Philippart et al. (2020) dans une étude longitudinale (1 an, 4 séances) en contexte bilingue pour des enfants de 21 à 36 mois. Ils montrent que l'indice Phi évolue au cours des différentes sessions de production, il fluctue, ce qui montre que l'acquisition des gestes phonologiques et du langage n'est pas linéaire. Cet indice offre, entre autres, la possibilité d'étudier l'évolution des processus d'acquisition de la parole mais également d'éclairer les processus de parole au niveau segmental pour des locuteurs sains mais aussi pour des

locuteurs atteints de troubles de la production de la parole. En contexte clinique cet indice pourrait permettre d'évaluer l'évolution de la parole post-traitement et ainsi constituer un outil de l'évaluation clinique.

Les travaux sur cet indice ont offert un cadre de réflexion à (Meunier & Ghio, 2018) les conduisant à développer l'indice de distinctivité. Basé sur l'indice Phi, il permet « de rendre compte de la relation dynamique entre la dispersion des catégories de voyelles et la dispersion de l'ensemble du système ». L'ID représente le rapport entre la variation du système et la variation intra-catégorie. Cet indice étant récent dans la littérature il n'a pas encore fait l'objet d'études complémentaires. Il nécessite d'être davantage mis à l'épreuve pour différents types de parole. Ils suggèrent tout de même que 4 profils croisés de locuteurs émergent relatifs à la distinctivité entre les catégories et à la taille du système. Deux paramètres de dispersion sont ainsi considérés, celui de chaque catégorie de voyelle et celui du système. Dans un système dispersé, les voyelles peuvent être concentrées et présenter un degré de distinction élevé les unes envers les autres, ou chaque catégorie de voyelle peut être plus dispersée et présenter un degré de distinction moyen entre elles. Dans un système concentré, les catégories de voyelles lorsqu'elles sont concentrées présentent un degré de distinction qui est également moyen. En revanche, si les catégories vocaliques sont dispersées, le degré de distinction est faible entre chacune d'entre elles. Le degré de dispersion / concentration des catégories vocaliques en fonction de la taille du système suggère qu'il existe entre les locuteurs une confusion plus ou moins importante des catégories. Cet indice permet d'envisager de dégager des profils de locuteurs sains, mais aussi de caractériser plus précisément les altérations de la parole chez des patients atteints de troubles de la production de la parole.

La production des voyelles chez un locuteur est susceptible de varier au cours du temps, Elisabeth II, Reine d'Angleterre, n'y échappe pas (Harrington et al., 2000). Les facteurs liés à l'âge, au genre, au milieu social, ... et probablement aux 'tendances' de l'époque, impactent la production vocalique d'un locuteur au cours de sa vie. Mais ces changements peuvent se manifester de façon plus rapide dans le cadre des pathologies ORL, notamment les cancers, dont les traitements bouleversent l'état organique, la physiologie du conduit vocal et sont susceptibles de perturber la production de la parole sur le plan articulatoire.

Elles demeurent des unités perceptives aux propriétés acoustiques spécifiques permettant de les distinguer sur la base d'un faisceau d'indices particuliers.

Concernant l'évaluation, plusieurs critiques peuvent être émises. Concernant les méthodes perceptives, le matériel linguistique utilisé, les méthodes de calcul, les tâches proposées aux auditeurs sont variées. Les méthodes utilisées pour déterminer le score d'intelligibilité présentent plus de variation : Le matériel linguistique ciblé peut être transcrit orthographiquement, ou répété par l'auditeur, puis transcrit par un autre annotateur, souvent le chercheur qui dirige la session de test. Nous soulignons ici que l'auditeur transcrit ou répète ce qu'il a perçu, ses aptitudes productives sont donc aussi en jeu. Ces différences introduisent de la variabilité entre les méthodes d'évaluation perceptive

De plus, les méthodes de calcul du score d'intelligibilité diffèrent, elles peuvent porter sur le nombre de mots correctement identifiés, le nombre de mots clefs (mots de contenu) correctement identifiés, ou encore être simplement une 'note' sur 10 ou sur le nombre de phrases cibles. Des taux peuvent aussi être calculés, représentant généralement le nombre de mots correctement identifiés divisé par le nombre de mots dans la phrase ciblée. Les unités ciblées dans le calcul de ces scores sont relativement larges (mots, phrases), et présentent plusieurs biais perceptifs connus, notamment les effets de restauration lexicale, ou les effets liés au contexte sémantique, qui ne sont que rarement considérés. Les méthodes de calcul du score d'intelligibilité ne font pas consensus (note, taux...) et impactent, de même que la tâche demandée à l'auditeur (évaluation par échelle visuelle, ou par transcription) nécessairement le résultat (Miller, 2013). Les mesures indirectes, par exemple, acoustiques, sont vues comme plus objectives. Cependant, les choix méthodologiques concernant les mesures elles-mêmes (valeur moyenne ou valeur prise en un point) peuvent être variés. Enfin, il reste difficile d'établir un lien entre ces deux types de mesures qui ne visent pas le même niveau linguistique.

Les auteurs ne proposent pas d'évaluation perceptive du niveau segmental en termes de phonèmes correctement identifiés mais les analyses acoustiques proposées portent sur le niveau segmental. Probablement qu'une meilleure homogénéité des unités visées par l'évaluation perceptive et acoustique serait bénéfique à la mesure.

Ainsi, aborder le concept d'intelligibilité au travers d'analyse phonétique des représentations phonologiques permettrait de l'envisager à un niveau segmental. Une analyse à un tel niveau nécessite des outils d'analyses fiables et objectifs, basés sur des théories linguistiques solides et reconnues dans la communauté scientifique. Ces éléments soulignent davantage encore la nécessité de préciser la définition de l'intelligibilité au regard de la complexité de l'échange communicationnel et du niveau linguistique ciblé par l'évaluation, pour proposer de nouvelles approches.

Comment évaluer l'intelligibilité segmentale après un cancer ORL ?

Au regard des éléments scientifiques, théoriques et cliniques précédemment exposés, il est clair que l'évaluation de l'intelligibilité, en tant que mesure clef du handicap communicationnel mérite l'attention avisée des chercheurs et des cliniciens.

Cette mesure, en ce qu'elle s'actualise dans le cadre d'un échange communicationnel entre le patient et son clinicien, implique nécessairement que soient pris en compte la participation de chacun à cet échange et le contexte de ce dernier. Il convient alors de tenir compte des mécanismes de production du locuteur, dans leur spécificité liée à la pathologie, ici, le trouble de la parole, exprimé au niveau de la réalisation articulatoire des segments, par un système physiologique lésé. Les mécanismes de perception de l'auditeur, ici également évaluateur, doivent également être considérés, en particulier dans toute leur complexité d'intégration et d'adaptation. Ainsi, locuteur, auditeur, contexte communicationnel et niveau linguistique a minima définissent les enjeux de l'évaluation. Le recours aux concepts et théories linguistiques sont alors éclairants pour proposer un cadre définitoire plus précis à la problématique de l'évaluation de l'intelligibilité chez des patients traités pour un cancer de la cavité buccale et de l'oropharynx.

Nous proposons de définir l'intelligibilité, entendue en termes segmentaux, comme « le degré de précision articulatoire qu'un locuteur est capable de produire évalué au travers de la capacité d'écoute et de décodage dont un auditeur peut faire preuve ». Cette définition, plus précise et adaptée à notre contexte pathologique (cancérologie ORL, troubles articulatoires de la production de parole, niveau segmental), pourrait ainsi être adaptée à d'autres contextes (autre trouble, acquisition d'une L2...), et d'autres niveaux linguistiques (lexique, prosodie ...).

En nous inscrivant dans ce cadre définitoire, nous avons souhaité dans ces travaux, préciser l'évaluation de l'intelligibilité dans les cancers ORL au moyen de deux approches. La première approche, est une approche perceptive par décodage acoustico-phonétique et s'inscrit dans le champ des mesures directes de l'intelligibilité. La seconde approche consiste en des mesures acoustiques des segments vocaliques et s'inscrit cette fois dans le champ des mesures indirectes de l'intelligibilité. L'avantage de ces deux approches et des méthodes qui seront ici proposées est qu'elles permettent une évaluation de l'intelligibilité segmentale.

L'approche perceptive par décodage acoustico-phonétique permet en effet de mesurer le degré d'altération de l'intelligibilité grâce à un score exprimé en nombre moyen de traits altérés par phonème. Ce score de Déviation Phonologique perçue (Score PPD) peut être un score global, calculé à partir de l'ensemble des productions d'un locuteur, et qui

rend compte de son intelligibilité globale. Il peut être aussi un score local, calculé pour chaque segment produit au sein des pseudo-mots produits par un locuteur, et qui permet de repérer ses segments préservés et ses segments les plus altérés. Dans les deux cas, c'est le degré d'altération d'intelligibilité segmentale qui est exprimé.

Les mesures acoustiques des segments vocaliques sont également des mesures segmentales. Il s'agit de caractériser, au moyen de métriques calculées à partir des valeurs de F1 et F2, l'espace vocalique des locuteurs.

Ces deux approches, parce qu'elles évaluent toutes deux le niveau segmental, à l'inverse des données de la littérature, pourront être confrontées. Elles seront présentées dans la seconde section de ce document.

Dans cette deuxième section, au total 4 corpus seront exploités. Un corpus de production de mots, issus de la BECD (corpus BECD), un corpus de pseudo-mots multi-listes (corpus DAP_m), un corpus de pseudo-mots mono-liste (DAP_131) et enfin, le corpus DAP enregistré dans le cadre du projet de recherche C2SI (corpus DAP_C2SI).

Dans l'approche perceptive par décodage acoustico-phonétique, deux séries d'analyses seront conduites. Par la première série d'analyses, basées sur les scores d'intelligibilité « PPD Global » nous interrogeons en premier lieu la pertinence de cette méthode d'évaluation perceptive pour contrôler les biais perceptifs d'intégration d'informations de haut-niveau qui, nous l'avons vu, peuvent conduire à une sous-évaluation de l'intelligibilité. Nous mobiliserons ici les trois premiers corpus suscités (mots de la BECD, pseudo-mots du DAP x 2). Ils feront l'objet de tests de perception (n = 3) pour répondre à plusieurs questions de recherches. Ainsi, nous comparerons les scores globaux moyens et leurs évolutions au cours de l'expérimentation pour deux matériaux linguistiques (mots versus pseudo-mots). Pour les pseudo-mots plus particulièrement, ces comparaisons viseront deux catégories d'auditeurs (naïfs versus experts). Enfin, une dernière expérience évaluera les précédents résultats au moyen d'un nouveau paradigme expérimental visant à déterminer s'ils sont davantage imputables à la qualité (mots vs. pseudo-mots) ou à la quantité (50 vs. 1000) du matériel linguistique.

Dans une seconde série d'analyses, la décomposition des scores PPD global en scores PPD local, phonème à phonème permettra de caractériser plus précisément encore l'altération segmentale. Ici, nous emploierons le corpus DAP C2SI. Il a déjà fait l'objet de tests de perception dans le cadre d'autres études (Ghio et al., 2018a; Ghio, Lalain, et al., 2020; Rebourg, 2018), nous disposons donc des résultats de ces tests en termes de scores PPD globaux et locaux et des transcriptions des auditeurs. Nous proposons d'examiner ces scores locaux au moyen d'une analyse statistique qui permettra l'identification des segments de parole préservés et dégradés. Une analyse descriptive des réponses des auditeurs, en termes de transcriptions locales permettra de caractériser les segments substitués aux segments cibles.

La seconde approche repose sur des mesures acoustiques (F1 et F2) et le calcul de métriques qui concernent le système vocalique dans sa globalité (Aire de l'espace

vocalique) et la distinction entre chaque catégorie de segments vocaliques (Indice de Distinctivité). Ces mesures porteront sur le corpus DAP du projet C2SI. Elles seront utilisées pour préciser l'altération des segments vocaliques et pour distinguer les deux groupes de locuteurs. Nous évaluerons également la relation entre ces mesures et les scores perceptifs. A partir des métriques pertinentes nous proposerons une visualisation des systèmes vocaliques des locuteurs. Nous proposerons également d'enrichir les représentations visuelles en intégrant, en plus des éléments acoustiques, des éléments perceptifs, i.e les transcriptions des auditeurs.

Cette seconde section sera divisée en trois parties, chaque partie comprendra la description des données de corpus en termes de population, de matériel linguistique et de traitements des enregistrements. Ensuite nous présenterons l'ensemble des aspects méthodologiques des tests de perception d'une part et des analyses descriptives d'autre part, et préciserons les modalités des traitements des données. Enfin, nous présenterons les résultats des études conduites pour répondre à nos questions de recherches.

4. Approche par Décodage Acoustico Phonétique

Afin de pouvoir répondre à notre première question de recherche, nous avons mobilisé plusieurs ressources. Dans cette section nous présenterons les corpus de mots et de pseudo-mots qui ont permis de conduire les tests de perception, dont nous détaillerons le protocole expérimental. Nous préciserons les populations d'auditeurs impliqués dans ces tests, ainsi que la méthode de traitement des données de perception collectées.

4.1. Corpus

Dans cette partie nous présenterons deux corpus différents mais complémentaires. Dans un premier temps nous présenterons le corpus composé des productions d'une liste de mots issue de la BECD et de plusieurs listes de pseudo-mots provenant des stimuli générés dans le cadre du projet C2SI. Dans un second temps nous présenterons un corpus comprenant la production d'une seule et unique liste de pseudo-mots par 10 locuteurs patients et 10 contrôles. Enfin nous proposerons un bref récapitulatif de ces 3 corpus différents.

4.1.1. Corpus BECD et DAP_m

Les corpus mobilisés dans ces travaux de recherches sont présentés ci-après. Après avoir apporté des précisions sur les groupes de locuteurs enregistrés dans ces corpus nous détaillerons les matériaux linguistiques produits par les locuteurs en développant les tâches de production de mots et de pseudo-mots.

4.1.1.1. Locuteurs et métadonnées

Les deux corpus utilisés ici sont constitués des productions des mêmes locuteurs. Ils ont été enregistrés sur deux tâches de production de la parole : production de mots (Corpus BECD) et production de pseudo-mots (Corpus DAP_m). Dans ces corpus, les patients ont été enregistrés à l'Oncopole de Toulouse, par M. Balaguer, orthophoniste et docteur en Sciences, dans le cadre de leur suivi clinique. Les locuteurs contrôles ont été enregistrés à Aix-en-Provence par l'équipe du LPL. L'ensemble de ces enregistrements ont été intégrés dans le corpus C2SI.

4.1.1.2. Locuteurs Contrôles et Patients

Ces petits corpus comprennent les enregistrements d'un groupe de 20 locuteurs. Ce groupe se divise en deux sous-groupes de locuteurs dans lesquels on retrouve 10 patients et 10 contrôles. Les groupes sont appareillés en sexe et chaque groupe inclut 5 hommes et 5 femmes. Les groupes de femmes sont bien appareillés en âge (μ contrôles = 52,2 ans

et μ patientes = 54,2 ans). On observe une plus grande différence entre les groupes d'hommes avec 63 ans en moyenne pour les patients et 46 ans en moyenne pour le groupe contrôle. Ce dernier est en moyenne 17 ans plus jeune que le groupe patient. Les détails du genre et de l'âge des locuteurs est résumé dans les Tableau 2 et Tableau 3.

Contrôles	
Genre	Âge
M	28
M	46
M	49
M	51
M	56
F	48
F	48
F	51
F	57
F	57
Moyenne M	46
Moyenne F	52,2
Moyenne	49,1

Tableau 2 : Âge et sexe des locuteurs contrôles des corpus BECD et DAP_m

Patients	
Genre	Âge
M	60
M	60
M	61
M	66
M	68
F	42
F	45
F	58
F	62
F	64
Moyenne M	63
Moyenne F	54,2
Moyenne	58,6

Tableau 3 : Âge et sexe des locuteurs patients des corpus BECD et DAP_m

Les patients dont il est question dans ce corpus sont des patients atteints, puis traités, pour un cancer de la cavité buccale ou de l'oropharynx. Dans la partie suivante nous apporterons des précisions afin de mieux caractériser cette population en présentant des métadonnées sur la localisation du cancer et son stade selon la classification TNM, ainsi que les traitements dont ont bénéficié les patients.

4.1.1.3. Métadonnées des Patients

Les patients inclus dans ce corpus feront l'objet d'un descriptif détaillé au regard des métadonnées cliniques dont nous disposons.

(a) Localisation

Dans ce corpus 60% des patients ont une atteinte localisée au niveau de l'oropharynx, et pour 40% des patients l'atteinte était localisée dans la cavité buccale (1.2.1.1. Localisations)

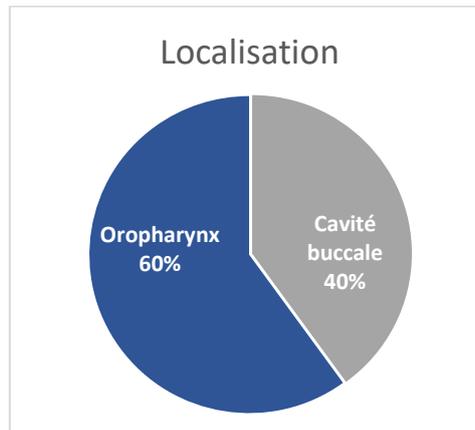


Figure 10 : Localisation de l'atteinte physiologique, divisée en deux groupes, la cavité buccale (39% des patients) – plancher, langue, rétromolaire et mandibule - et l'oropharynx (61% des patients) – amygdale, base de langue, oropharynx et voile du palais -.

(b) Stade TNM

Dans ce corpus, les patients ont été traités pour un cancer de la cavité buccale ou de l'oropharynx. Les cancers de ces patients sont classés selon la classification internationale TNM qui permet de rendre compte de l'étendue d'un cancer (1.2.1.3 Classification des cancers).

Pour le stade T (Figure 11), déterminant la taille de la tumeur sur une classification allant de 1 à 4, il y a 4 patients au stade 2, et 3 patients au stade 4. Les patients aux stades 1 et 3 ne sont pas représentés dans ce corpus. Pour 3 patients du corpus, la taille de la tumeur n'est pas renseignée, cependant, 2 d'entre eux ont bénéficié d'une chirurgie de la tumeur dans le protocole de soin établi par l'équipe médicale. Cela signifie qu'un seul patient n'a pas eu de tumeur traitée par chirurgie.

Concernant le stade N (Figure 12), qui détermine l'étendue de l'invasion ganglionnaire locale, 1 patient présente les stades 1 et 3 et 2 patients au stade 3. 3 patients n'ont pas eu d'invasion ganglionnaire (N0), cependant parmi eux, 1 a bénéficié d'une chirurgie des ganglions. Pour 3 patients la métadonnée N n'est pas renseignée, néanmoins 2 ont bénéficié d'une chirurgie N. Pour les deux patients « Non renseigné » mais ayant bénéficié d'une chirurgie de la tumeur et des ganglions, on peut en déduire que qu'il y a eu une atteinte, mais dont l'ampleur n'est pas précisée. Enfin, concernant le stade M, il est de 0 pour 7 patients, et inconnu pour 3 patients.

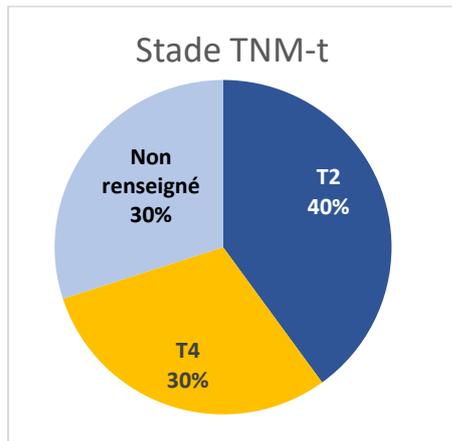


Figure 11 : Répartition des patients en fonction de la taille de la tumeur (en %)

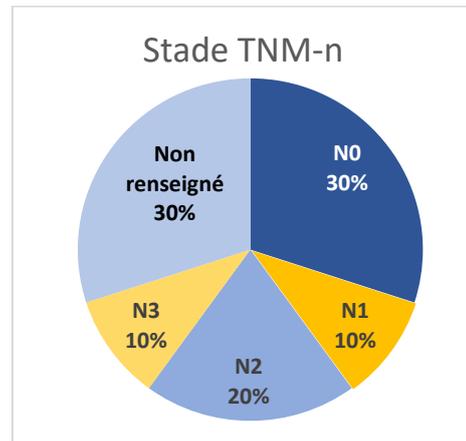


Figure 12 : Répartition des patients en fonction de l'envahissement des ganglions lymphatiques (en %)

(c) Traitements

Dans ce corpus, 8 patients ont bénéficié d'une chirurgie de la tumeur dans leur protocole de soin, et pour 2 patients ce traitement n'a pas été prescrit par l'équipe médicale. Le second secteur de la Figure 13 illustre le type de chirurgie dont les patients ont bénéficié.

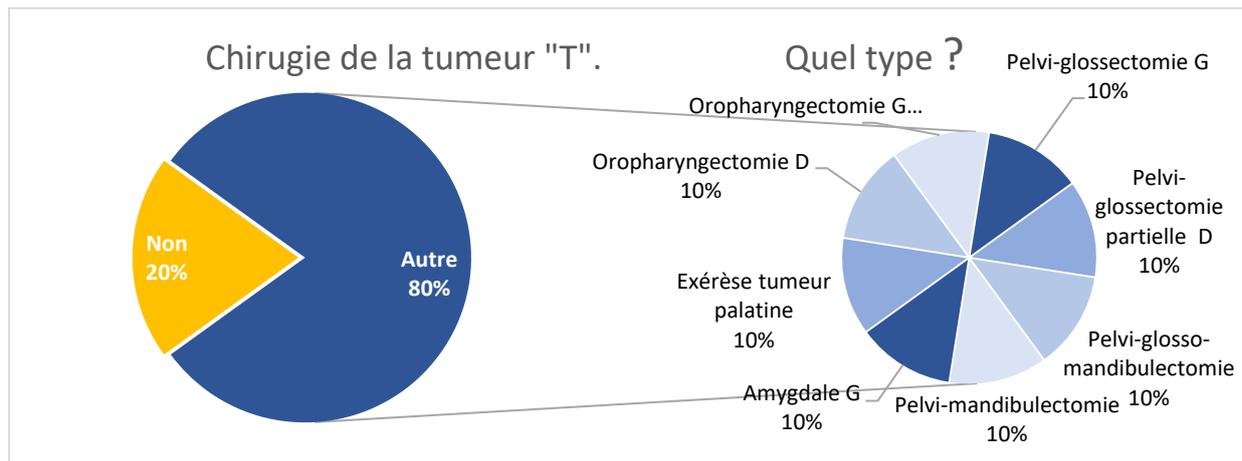


Figure 13 : Proportion de patients traités par chirurgie et type d'intervention chirurgicale pratiquée (en %)

La Figure 14 montre les différents traitements et leur combinaison ordonnée dont ont bénéficié les patients. 80% d'entre eux ont bénéficié d'une chirurgie de la tumeur, et 6 d'entre eux ont ensuite bénéficié d'une chirurgie des ganglions. 80% des patients ont bénéficié d'une radiothérapie, et la moitié d'entre eux ont également bénéficié d'une chimiothérapie. Pour 2 patients seulement l'équipe médicale n'a pas inclus de chirurgie dans le protocole de soin, ils ont bénéficié, respectivement, d'une radiothérapie seule et d'une radiothérapie associée à une chimiothérapie.

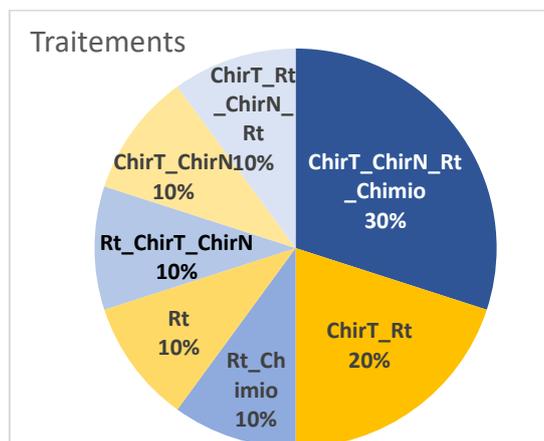


Figure 14 : Type de traitement reçu par les patients, exprimé en pourcentage. « RT » Radiothérapie ; « Chimio » Chimiothérapie ; « ChirT » Chirurgie de la tumeur ; « ChirN » Chirurgie des ganglions ; lorsque le patient a reçu une combinaison de traitement, l'ordre est indiqué par l'ordre des abréviations (« ChirT_Rt » signifie que le patient a bénéficié d'une chirurgie de la tumeur, puis d'une radiothérapie) (en %)

(d) Délais Post traitement

Dans ce corpus, le délai post traitement n'a pas été contrôlé. Les patients se situent dans une fourchette très large : 2 mois 276 mois (23 années) avec une moyenne à 100,5 mois (un peu plus de 8 années).

4.1.1.4. Tâches enregistrées

Dans cette partie nous présenterons les tâches de production de la parole enregistrés auprès des 20 locuteurs présentés précédemment.

Tous les enregistrements ont été réalisés avec le même matériel technique, soit un ordinateur sur lequel est installée l'application PercEval (et donc Lancelot) (André et al., 2003), pour la présentation des stimuli (modalité audio-visuelle), un micro TLM102 pour la prise de son et un enregistreur ZOOM Ph4 pour le stockage des fichiers sonores générés. L'ensemble des locuteurs ont donc été enregistrés suivant le même protocole.

Les locuteurs sont installés dans un environnement calme. Devant eux se trouve un micro-trépied et un écran d'ordinateur. Ils sont également équipés d'un casque audio. Les stimuli à produire sont donnés aux locuteurs selon une double modalité (audio-visuelle). Ils sont affichés à l'écran (modalité visuelle) et entendus dans le casque, une voix de synthèse vocale produit les stimuli (modalité auditive). Cette double modalité est issue du protocole de la tâche de DAP du projet C2SI, les patients étant souvent âgés elle permet de compenser les effets du vieillissement, comme une perte d'audition ou d'acuité visuelle.

Les stimuli sont présentés à l'écran par la station de tests de perception auditive et visuelle PercEval et son application Lancelot. Les participants produisent les stimuli perçus et sont enregistrés. Deux types de matériaux linguistiques sont produits D'une part une liste de mots, issus de la BECD, et des pseudo-mots issus de la tâche de DAP du projet C2SI.

4.1.1.5. Tâche de production de mots – Corpus BECD

Les 20 locuteurs, 10 patients et 10 contrôles présentés dans la partie 4.1.2.2 ont été enregistrés sur la production de la liste de 50 mots de la Batterie Evaluation Clinique de la Dysarthrie BECD. C'est une tâche unique de production de mots. Chaque locuteur produit 50 mots, nous obtenons ainsi 20 locuteurs x 50 mots = 1000 mots.

abri	chaîne	grain	nain	pont
appris	chèvre	grand	noix	porc
arbre	clenche	gras	pain	prend
boxeur	cloche	groin	paire	sport
brin	craie	gros	paon	tasse
caché	crin	grue	paquet	taxe
café	croche	herbe	parquet	thym
calé	docteur	main	peur	toi
carré	franc	moi	plein	toile
cassé	frein	mouche	pleur	train

Tableau 4 : Liste des mots de la BECD

4.1.1.6. Tâches de production de pseudo-mots – Corpus DAP_m

Les mêmes locuteurs ont également produit deux listes de 52 pseudo-mots issues de la tâche de DAP. Les deux listes produites par les locuteurs sont différentes et sont également différentes entre chaque locuteur. Chaque locuteur produit 2 listes différentes générées spécifiquement, à partir du corpus de pseudo-mots du projet de recherche C2SI. Chaque locuteur produit donc 2 x 52 pseudo-mots = 104 pseudo-mots x 20 locuteurs = 2080 pseudo-mots différents.

blouco	fania	ludin	psupo	squirbu
boubo	finja	mapa	purtu	stolou
bouflin	fletin	mofu	quevi	togli
brena	fregant	nabri	quinscou	touchou
champri	ganvant	nenia	quinzin	trampsant
chanri	gliga	nianti	riyo	vintru
cliyain	grospou	niourcant	rucha	vouso
crirmou	jali	panzo	saptant	yafro
dimu	jomi	plinfu	simblant	yissant
drejou	lerin	prucu	spocrou	zada
dunin				zoubou

Tableau 5 : Exemple de liste de pseudo-mots

banja	drati	judrou	plipa	spovu
bavrant	finrsi	lanin	pourdant	stoucho
blafi	flactin	limu	progri	tisso
brinrno	fouplu	marpant	psero	toclou
checa	frebin	mouilli	quinchin	tromin
chenia	gliga	nioultant	randa	vogzi
clouzou	grugou	nioustu	ricou	vurvo
crilo	gueilla	nolu	sanjo	yainsant
cutin	guinfi	nubou	sazi	yankusu
danda	jinu	peniou	scanvant	zinrjant
derin				zupou

Tableau 6 : Exemple de liste de pseudo-mots

Pour résumer : Chaque locuteur a été enregistré sur la production de la liste de 50 mots issue de la BECD et sur la production de 2 listes de 52 pseudo-mots issues du matériel linguistique constitué pour la tâche de DAP du projet de recherche C2Si. Soit 154 productions par locuteurs. Nous avons donc enregistré 154 stimuli x 20 locuteurs, soit 3080 stimuli en tout et respectivement 1000 stimuli pour le corpus BECD et 2080 stimuli pour le corpus DAP_m. Ces listes de mots et de pseudo-mots ont fait l'objet du même traitement en vue de leur préparation pour les tests de perception (4.1.4. Traitement des enregistrements) que les enregistrements du corpus présenté dans la partie suivante.

4.1.2. Corpus DAP_131

Pour les besoins de notre étude un second corpus a été constitué. Il s'agit d'un corpus de la production de la même liste de pseudo-mots par tous les locuteurs.

4.1.2.1. Locuteurs et métadonnées

20 locuteurs ont été enregistrés sur la production d'une seule et même liste de 52 pseudo-mots tirés du protocole de la tâche de DAP, la liste 131. Dans ce corpus, les patients ont été enregistrés à l'Oncopole de Toulouse dans le cadre de leur suivi clinique, par M. Balaguer, orthophoniste et docteur en Sciences, dans le cadre de son projet de thèse. Les locuteurs contrôles ont été enregistrés à Aix-en-Provence, par M. Rebourg, pour les besoins de cette étude. L'ensemble de ces enregistrements ont été intégrés au corpus C2SI.

4.1.2.2. Patients et contrôles

10 patients, traités pour un cancer de la cavité buccale ou de l'oropharynx ont été enregistrés à Toulouse et sujets 10 contrôles, appareillés en âge et en sexe avec le groupe patient, ont été enregistrés à Aix-en-Provence.

Ce petit corpus comprend les enregistrements d'un groupe de 20 locuteurs. Ce groupe se divise en deux sous-groupes de locuteurs dans lesquels on retrouve 10 patients et 10 contrôles. Les groupes sont parfaitement appareillés en sexe et chaque groupe inclut 5 hommes et 5 femmes. Les groupes de femmes sont bien appareillés en âge (μ contrôles = 52,2 ans et μ patientes = 54,2 ans). On observe une plus grande différence entre les groupes d'hommes avec 63,2 ans en moyenne pour les patients contre 73 ans en moyenne pour le groupe contrôle. Ce dernier est en moyenne 10 ans plus jeune que le groupe

patient. Il est également plus difficile de recruter des locuteurs âgés. Les détails du genre et de l'âge des locuteurs est résumé dans les Tableau 4 et Tableau 7.

DAP contrôles	
Genre	Age
M	50
M	51
M	58
M	77
M	80
F	49
F	58
F	61
F	65
F	72
Moyenne M	63,2
Moyenne F	61
Moyenne	62,1

Tableau 7 : Âge et sexe des locuteurs contrôles du corpus DAP_131

DAP patients	
Genre	age
M	61
M	65
M	74
M	82
M	83
F	53
F	64
F	65
F	70
F	75
Moyenne M	73
Moyenne F	65,4
Moyenne	69,2

Tableau 8 : Âge et sexe des locuteurs patients du corpus DAP_131

Les patients dont il est question dans ce corpus sont des patients atteints, puis traités, pour un cancer de la cavité buccale ou de l'oropharynx. Dans la partie suivante nous apporterons des précisions afin de mieux caractériser cette population en présentant des métadonnées sur la localisation du cancer et son stade selon la classification TNM, ainsi que les traitements dont ont bénéficié les patients.

4.1.2.3. Patients

Les patients inclus dans ce corpus feront l'objet d'un descriptif détaillé au regard des métadonnées cliniques dont nous disposons

(a) Localisation

Dans ce corpus 60% des patients ont une atteinte localisée dans la cavité buccale et pour 40% des patients l'atteinte est localisée au niveau de l'oropharynx

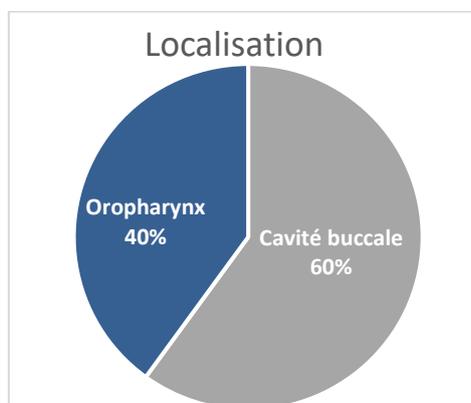


Figure 15 : Localisation de l'atteinte physiologique, divisée en deux groupes, la cavité buccale (60% des patients) – plancher, langue, rétromolaire et mandibule - et l'oropharynx (40% des patients) – amygdale, base de langue, oropharynx et voile du palais-

(b) Stade TNM

Dans ce corpus, les patients sont pour 5 d'entre eux classés T4 selon la classification internationale TNM qui permet de rendre compte de l'étendue d'un cancer (1.2.1.3). « T » rend compte de la taille de la tumeur. 2 patients sont au stade 3 et 1 classé T2. Ceci est résumé dans la Figure 16. S'agissant de l'envahissement ganglionnaire, N (Figure 17), pour 2 patients aucun envahissement n'a été relevé par l'équipe médicale (N0) et pour 2 autres c'est le stade 2 qui est constaté. Enfin, 1 patient est classé N3 avec un envahissement à distance et 3 patients sont classés N1. Enfin, les 8 patients pour lesquels le stade M est disponible sont classés M0. Pour deux patients les données concernant les stades T, N et M ne sont pas renseignées.

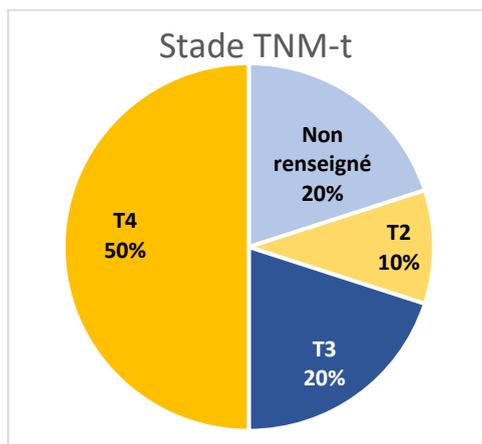


Figure 16 : Répartition de la population de patients en fonction de la taille de la tumeur (en %)

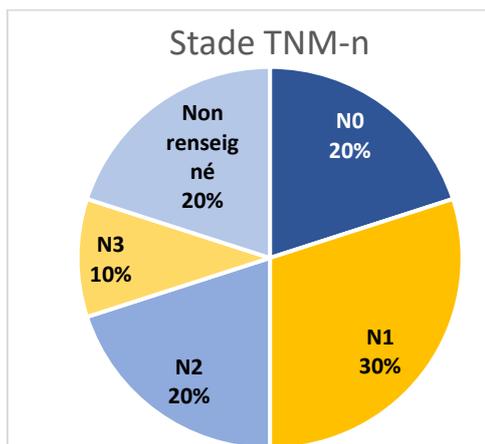


Figure 17 : Répartition de la population de patients en fonction de l'envahissement des ganglions lymphatiques (en %)

(c) Traitements

Dans ce corpus, 8 patients ont bénéficié d'une chirurgie de la tumeur dans leur protocole de soin, et pour 2 patients ce traitement n'a pas été prescrit par l'équipe médicale. Le second secteur de la Figure 18 illustre le type de chirurgie dont les patients ont bénéficié. Notons que pour 3 d'entre eux cette donnée est manquante.

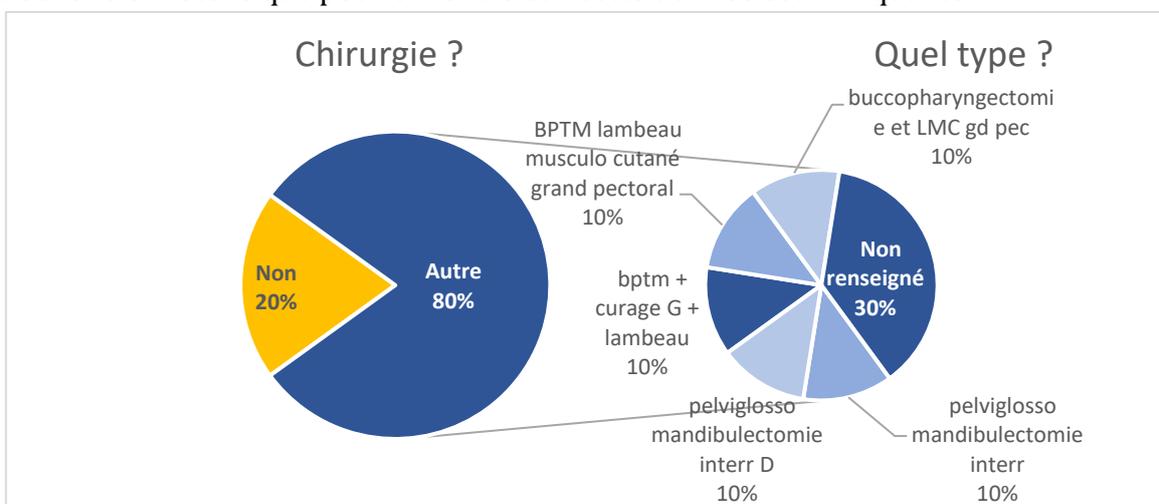


Figure 18 : Proportion de patients traités par chirurgie et type d'intervention chirurgicale pratiquée (en %)

La Figure 19 montre les différents traitements dont ont bénéficié les patients. 8 d'entre eux ont bénéficié d'une chirurgie, néanmoins, dans ce corpus il n'est pas précisé si cette dernière concerne la tumeur et/ou les ganglions. Tous les patients ont en revanche bénéficié d'une radiothérapie, et seulement 4 d'entre eux ont bénéficié d'une chimiothérapie. Dans ce corpus, il n'est pas précisé l'ordre dans lequel les patients ont reçu les traitements. Les combinaisons de traitement dont ils ont bénéficié sont représentées dans la Figure 19

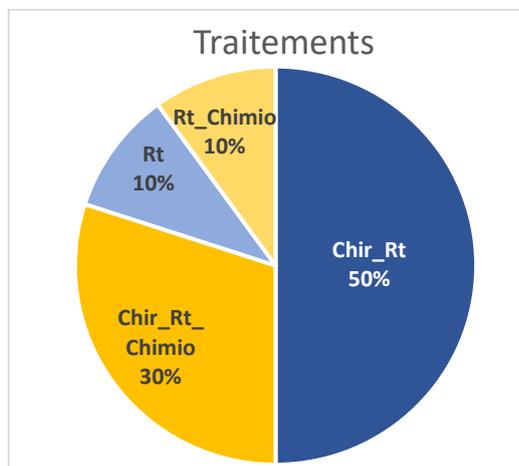


Figure 19 : Type de traitement reçu par les patients, exprimé en pourcentage. « RT » Radiothérapie ; « Chimio » Chimiothérapie ; « Chir » Chirurgie de la tumeur et/ou des ganglions (en %)

(d) Délai post traitement

Dans ce corpus, le délai post traitement présente une grande amplitude. Les patients se situent dans un intervalle relativement large : 5 mois 226 mois (presque 19 années) avec une moyenne à 77 mois (un peu plus de 6 années).

4.1.2.4. La tâche de production : DAP liste 131

Les 20 locuteurs présentés dans la partie 4.1.2.1 ont été enregistrés sur la production d'une seule et même liste de 51 pseudo-mots, la liste 131. Ce sont donc 20 locuteurs x 52 pseudo-mots = 1040 pseudo-mots qui ont été enregistrés dans ce corpus.

biniou	flimbin	lansou	pligui	spava
bitant	fochi	lefro	prospu	staji
blango	foru	mebi	psimpra	tipso
brana	frorba	mouglou	quinla	tofin
choudant	glusca	nebrou	quinrtant	tralin
chupant	gouillou	nianfo	rancou	vatrin
cluchant	greptu	niouqui	ruflu	vava
crerquin	gueru	noussi	sazo	yezo
dimin	jimpu	pancri	scumi	yublu
dinou	jorma	pantant	souillou	zinjo
dridin				zouniant

Tableau 9 : Liste de pseudo-mots DAP n° 131

Les modalités d'enregistrements sont identiques à celles décrites précédemment (même protocole, même micro, environnement calme, ...). Dans la partie suivante nous proposons une brève synthèse des corpus présentés dans cette section.

4.1.3. Corpus : synthèses

3 corpus différents ont été constitués. Un groupe de 20 locuteurs (10 patients et 10 contrôles) a été enregistré sur la production de la liste de 50 mots issue de la BECD (corpus BECD) et sur la production de différentes listes de 52 pseudo-mots (corpus DAP_m). Un second groupe de 20 locuteurs (10 patients et 10 contrôles) a été enregistré sur la production d'une seule et même liste de 52 pseudo-mots, la liste 131 (corpus DAP_131).

L'ensemble des patients ont été enregistrés à Toulouse, dans le cadre de leur suivi clinique. Les locuteurs contrôles ont été enregistrés à Aix-en-Provence par les membres du LPL associés au projet de recherche C2SI.

Le tableau (Tableau 10) suivant résume le nombre de stimuli enregistrés dans chacun des corpus présentés dans cette partie.

Nom du corpus	Nature des stimuli	Nombre
BECD	Mots	1000
DAP_m	Pseudo-mots	2080
DAP_131	Pseudo-mots	1040

Tableau 10 : Récapitulatif des corpus utilisés dans ces travaux comprenant la nature du matériel linguistique (lexical, non lexical) et le nombre de stimuli par corpus

Ces enregistrements ont fait l'objet de traitements pour être préparés en vue des tests de perception.

4.1.4. Traitement des enregistrements

Les enregistrements collectés font l'objet d'un prétraitement qui permet leur utilisation dans les tests de perception. Chaque item produit est d'abord aligné avec sa transcription de façon automatique. Ensuite, nous vérifions manuellement cet alignement et le corrigeons si nécessaire. Les frontières posées permettent d'annoter et de délimiter le stimulus. A partir de ce bornage, nous extrayons les stimuli produits de façon à obtenir un fichier son (.wav) par production et par locuteur.

Ce découpage permet de brasser les stimuli dans les tests de perception, entre les locuteurs et les groupes de locuteurs. Ainsi, nous pouvons présenter les stimuli aux auditeurs dans l'ordre de notre choix. Les détails sur les modalités de randomisation et de présentation des stimuli aux auditeurs seront présentés dans les sections « Designs expérimentaux » dédiées à cet effet. Ces aspects appartenant à la genèse des stimuli (enregistrement, étiquetage, alignement, extraction) constituent la base de la préparation des stimuli, en vue de leur intégration au corpus, mais également pour la préparation des tests de perception.

4.2. Méthodologie

Dans cette section nous présenterons les designs expérimentaux mis en place dans nos tests de jugement perceptif de l'intelligibilité. En premier lieu nous présenterons le design expérimental permettant d'évaluer la pertinence des pseudo-mots par rapport aux mots pour contrôler le recours aux informations de haut niveau, et plus particulièrement des effets de restauration lexicale. Celui-ci a été conduit auprès de deux groupes d'auditeurs afin d'évaluer également l'effet d'expertise auditive des cliniciens, par rapport aux auditeurs naïfs.

Nous présenterons ensuite le design expérimental permettant de comparer les effets d'apprentissage du matériel linguistique au regard de sa qualité (mots versus pseudo-mots). En comparant ces résultats avec les résultats précédents, nous pourrions obtenir une appréciation de ces effets.

4.2.1. Design expérimental « BECD DAP »

Le protocole expérimental décrit dans cette section a été élaboré afin de mener une campagne de tests de jugement perceptif de l'intelligibilité. Ces tests ont été conduits pour collecter les transcriptions orthographiques de 18 auditeurs dits « naïfs » et 18 auditeurs dits « experts ». Les corpus BECD (1000 stimuli) et DAP_m (2080 stimuli) ont été mobilisés.

4.2.1.1. Design

Au vu de la très grande quantité d'items, nous avons choisi de limiter le nombre de stimuli transcrits par chaque auditeur à 1540. Les locuteurs ont été divisés en deux groupes (A et B) de 10 locuteurs dans lesquels on retrouve 5 patients et 5 témoins, impliquant que les corpus BECD et DAP_m ont eux aussi été divisés en deux sous corpus composés chacun de 500 stimuli BECD et 1040 stimuli DAP_m.

A partir de chaque sous corpus, des listes de stimuli ont été constituées. Les 2 x 500 stimuli des corpus BECD ont été divisés en 2 x 3 listes et les 2 x 1040 stimuli du corpus DAP ont été divisés en 2 x 6 listes. Chaque liste BECD est constituée de 167 productions et chaque liste DAP est constituée de 174 productions. En tout, nous avons donc 6 listes de productions pour les corpus BECD avec 167 stimuli par liste et 12 listes pour les corpus DAP avec 174 stimuli par liste.

Les listes 1 à 3 du corpus BECD sont produites par le groupe A composé de 5 patients et 5 témoins, les listes 4 à 6 ont été produites par l'autre groupe (B) de locuteurs dans lequel il y a également 5 patients et 5 témoins. Dans le corpus DAP_m, les listes 1 à 3 et 7 à 9 ont été constituées à partir des productions du groupe A, et correspondent pour les listes 1 à 3 à la production d'une première liste de DAP1 et pour les listes 7 à 9 à la production de la seconde liste de DAP2. Il en va de même pour les listes 4 à 6 et 10 à 12, composées des productions du groupe B, et qui correspondent respectivement aux productions d'une première liste de DAP1 puis d'une seconde liste de DAP2.

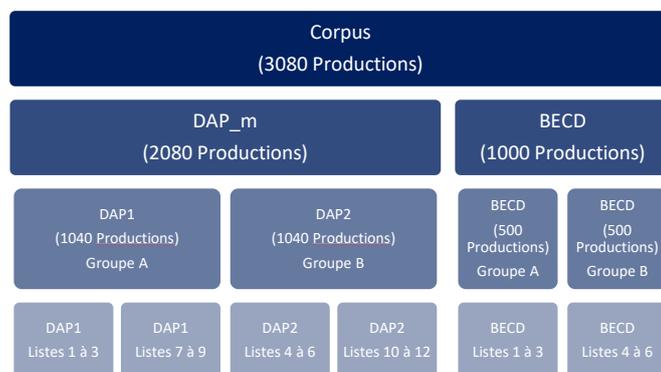


Tableau 11 : Résumé de la répartition des corpus

Ces listes ont été utilisées pour effectuer des tests de jugement perceptif de l'intelligibilité auprès de plusieurs populations. Dans ces tests, l'auditeur écoute dans un casque les productions de patients (traités pour un cancer des VADS) et de témoins, et transcrit orthographiquement, sur un clavier d'ordinateur, ce qu'il entend.

Chaque auditeur écoute et transcrit 3 listes BECD et 6 listes DAP, produites par un groupe de locuteurs. Chaque liste est transcrite par trois auditeurs différents dans chaque ordre de passation possible, c'est-à-dire que chaque production est transcrite 9 fois par 3 auditeurs différents x 3 positions. Nous obtenons ainsi 3 transcriptions, proposées par 3 auditeurs différents, et ce dans chacune des positions (T1, T2 et T3). Soit 9 transcriptions par listes et donc par stimuli.

L'ordre de passations des listes est résumé dans le tableau de design expérimental suivant :

		T1	T2	T3
3 auditeurs ≠	BECD	1	2	3
	DAP 1	1	2	3
	DAP 2	7	8	9
3 auditeurs ≠	BECD	4	5	6
	DAP 1	4	5	6
	DAP 2	10	11	12
3 auditeurs ≠	BECD	3	1	2
	DAP 1	2	3	1
	DAP 2	8	9	7
3 auditeurs ≠	BECD	6	4	5
	DAP 1	5	6	4
	DAP 2	11	12	10
3 auditeurs ≠	BECD	2	3	1
	DAP 1	3	1	2
	DAP 2	9	7	8
3 auditeurs ≠	BECD	5	6	4
	DAP 1	6	4	5
	DAP 2	12	10	11

Tableau 12 : Résumé du design expérimental BECD-DAP

Ce protocole expérimental a permis de réaliser des tests de jugement perceptif de l'intelligibilité auprès de deux groupes d'auditeurs dits « Naïfs » et « Experts ».

4.2.1.2. Auditeurs

Deux groupes d'auditeurs, appelés « Naïfs » et « Experts » ont participé à cette étude. Au cours des tests de perception, ils ont écouté dans un casque les productions de locuteurs patients et contrôles, et devaient transcrire, sur un clavier d'ordinateur, ce qu'ils entendaient. Les généralités concernant la méthodologie d'un test de perception sont décrites dans la partie 4.2.3. Méthodologie d'un test de perception.

18 auditeurs naïfs, natifs de langue française, sans problème de vue ou d'audition non corrigés, ayant un bon niveau en orthographe ont participé à cette campagne de tests. Les 18 auditeurs naïfs ont tous réalisé la tâche en débutant par la transcription des mots issus de la BECD puis ont poursuivi l'exercice en transcrivant les listes de pseudo-mots du DAP. Les auditeurs ont été dédommagés à hauteur de 10€ en ticket Kadeos.

18 auditeurs experts en écoute de la parole pathologique, tous orthophonistes ont participé à cette étude. Ce groupe se compose de 16 femmes et 2 hommes. Aucun critère d'ancienneté d'exercice n'a été déterminé, le critère principal d'inclusion concernait la prise en charge de patient dysarthriques au cours de leur exercice professionnel.

Notre population d'auditeurs experts est composée de 4 orthophonistes travaillant à 100% en libéral et 5 travaillant à 100% en Centre Hospitalier. 7 orthophonistes mènent une double activité répartie en taux variable entre la profession libérale et le salariat en CH (50% et 50%, 30% et 70%, 20% et 80%). Enfin, 2 orthophonistes sont étudiants en doctorat. L'un d'entre eux exerce un petit volume horaire en hôpital.

Nous avons proposé un questionnaire aux auditeurs experts suite à leur passation (Annexe C), ils nous ont permis d'en savoir plus sur leur pratique clinique et plus particulièrement sur l'évaluation de l'intelligibilité dans ce contexte. Ainsi, 100% des orthophonistes a été amené à évaluer l'intelligibilité. 14 d'entre eux ont déjà utilisé la BECD pour évaluer l'intelligibilité, et 10 utilisent la BECD en tant qu'outil de bilan de façon régulière. 8 utilisent d'autres batterie d'évaluation : des bilans composites, créés par leur soin, ou élaborés au sein de leur unité de travail (CH).

La passation de l'expérience étant relativement longue (\approx 2h) nous avons dû nous adapter à l'activité professionnelle des paramédicaux. Certains auditeurs experts ont pu se déplacer au LPL pour passer l'expérience, pour d'autres, nous nous sommes déplacée sur leur lieu de travail. Ainsi, 8 passations se sont déroulées dans les différents CH (Toulouse, Marseille), 9 passations se sont déroulées exclusivement au LPL, et 1 passation s'est déroulée entre le LPL et le cabinet libéral de l'orthophoniste. Certaines passations ont dû être divisées en deux sessions pour des raisons d'urgences hospitalières et pour s'adapter au mieux aux contraintes de ces professionnels paramédicaux. Ainsi, 11 auditeurs ont réalisé l'expérience en une seule session, et pour 7 auditeurs les passations ont été divisées en deux sessions (le plus souvent espacées d'une semaine). Pour des questions d'organisation les auditeurs experts ont réalisé la tâche en débutant par la transcription des listes de pseudo-mots du DAP, puis ont poursuivi l'exercice en

transcrivant les mots issus de la BECD. Ces auditeurs ont été dédommagés à hauteur de 40€ en ticket Kadeos.

Ce design expérimental a été élaboré pour répondre à un double objectif, il permet de comparer les scores d'intelligibilité obtenus par la production et la transcription de deux types de matériaux linguistiques : des mots et des pseudo-mots. Il permet également de comparer les scores obtenus par les transcriptions de deux catégories d'auditeurs : des auditeurs naïfs et des auditeurs experts.

4.2.1.3. Objectif

L'objectif soutenant cette étude réside dans la validation du critère de pertinence de l'utilisation (i) du matériel linguistique proposé dans le DAP et (ii) de la tâche de DAP auprès de deux groupes d'auditeurs, ce, en montrant que le DAP permet de réduire l'intégration des informations de haut niveau majorées par la pratique clinique et que les deux catégories d'auditeurs proposent donc des scores comparables.

Pour valider le critère de pertinence de l'utilisation de pseudo-mots, comme proposé dans la tâche de DAP, nous comparerons les scores PPD moyens attribués par une évaluation basée sur des mots (BECD) à ceux attribués par une évaluation basée sur des pseudo-mots (DAP). Nous confronterons également l'évolution des scores moyens au cours du temps de la tâche pour les deux matériaux linguistiques (Mots vs Pseudo-mots). Les scores attribués par une évaluation avec des pseudo-mots devraient conserver leur stabilité au cours du temps de la tâche car les effets de restauration et d'apprentissage du matériel linguistique devraient être neutralisés. Ces mêmes scores devraient être plus élevés (donc la dégradation de l'intelligibilité jugée plus sévère) que ceux attribués par une évaluation de type BECD, car ils concourent pour être de meilleurs représentants des déficits et handicap des patients, et ce pour les mêmes raisons.

4.2.2. Design expérimental « Quantité Qualité »

Dans cette section nous présenterons le design expérimental du test de jugement perceptif de l'intelligibilité conduit afin d'évaluer si c'est la qualité du matériel linguistique (pseudo-mots) ou la quantité de celui-ci qui permet de contrôler les effets d'apprentissage.

Pour rappel, les corpus mobilisés dans cette étude sont le corpus BECD (4.1.1.5) et le corpus DAP_131 (4.1.1.6). Les locuteurs des groupes patients et contrôles ne sont pas les mêmes groupes locuteurs, en tout, 40 locuteurs différents ont été enregistrés. 20 patients, 10 enregistrés sur la production de 50 mots issus de la BECD et 10 autres enregistrés sur la production de la liste 131 (52 pseudo-mots) issue de la tâche de DAP. Il en va de même pour le groupe contrôle dans lequel sont inclus 20 locuteurs, 10 enregistrés sur la production de 50 mots issus de la BECD et 10 autres enregistrés sur la production de la liste 131 (52 pseudo-mots) issue de la tâche de DAP. Pour rappel, ce sont donc 20 locuteurs x 50 mots = 1000 stimuli pour les mots de la BECD et 20 locuteurs x 52 pseudo-mots = 1040 stimuli pour les pseudo-mots du DAP qui sont employés ici.

Dans cette étude, nous avons répliqué le design expérimental de l'étude présentée précédemment.

4.2.2.1. Design

Afin de proposer un protocole expérimental qui suive les mêmes conditions que le design précédent, les locuteurs ont été divisés en deux groupes (A et B) de 10 locuteurs dans lesquels on retrouve 5 patients et 5 témoins, impliquant que les corpus BECD et DAP131 ont eux aussi été divisés en deux sous corpus composés chacun de 500 stimuli BECD et 520 stimuli DAP.

A partir de chaque sous corpus, des listes de stimuli ont été constituées. Les 2 x 500 stimuli du corpus BECD et les 2 x 250 stimuli du corpus DAP ont respectivement été divisés en 2 x 3 listes. Chaque liste BECD est constituée de 167 productions et chaque liste DAP est constituée 174 productions. En tout, nous avons donc 6 listes de productions pour les corpus BECD avec 167 stimuli par liste et 6 listes pour les corpus DAP avec 174 stimuli par liste.

Les listes 1 à 3 du corpus BECD sont produites par le groupe A composé de 5 patients et 5 témoins, les listes 4 à 6 ont été produites par l'autre groupe (B) de locuteurs dans lequel il y a également 5 patients et 5 témoins. Dans le corpus DAP, les listes 1 à 3 ont été constituées à partir des productions du groupe A. Il en va de même pour les listes 4 à 6 composées des productions du groupe B.

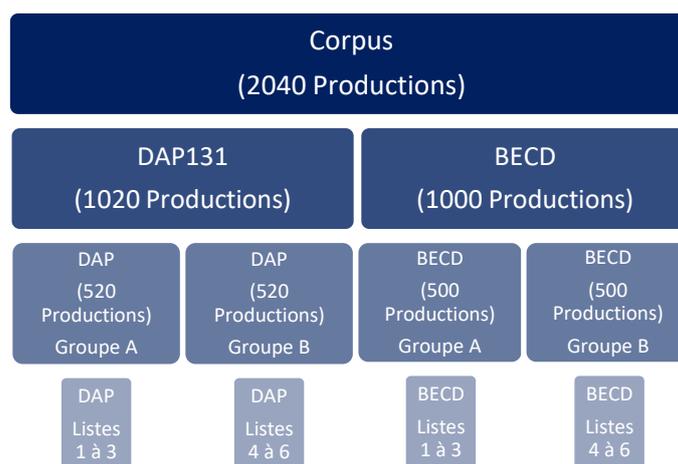


Table 1 : Résumé de la répartition des corpus

Chaque auditeur écoute et transcrit 3 listes BECD et 3 listes DAP, produites par un groupe de locuteur.

Chaque liste est transcrite par trois auditeurs différents dans chaque ordre de passation possible, c'est-à-dire que chaque production est transcrite 9 fois par 3 x 3 auditeurs différents.

Nous obtenons donc 3 transcriptions par position, proposées par 3 auditeurs différents, et ce dans chacune des positions (T1, T2 et T3)

		T1	T2	T3
3	BECD	1	2	3
auditeurs ≠	DAP131	1	2	3
3	BECD	2	3	1
auditeurs ≠	DAP131	2	3	1
3	BECD	3	1	2
auditeurs ≠	DAP131	3	1	2
3	BECD	4	5	6
auditeurs ≠	DAP131	4	5	6
3	BECD	5	6	4
auditeurs ≠	DAP131	5	6	4
3	BECD	6	4	5
auditeurs ≠	DAP131	6	4	5

Tableau 13 : Résumé du design expérimental de l'expérience « Quantité – Qualité »

Ce design expérimental permet de comparer l'évolution des scores PPD au cours de la répétition de 50 mots et de 52 pseudo-mots. Il a permis de réaliser des tests de perception auprès d'auditeurs naïfs.

4.2.2.2. Auditeurs

18 auditeurs naïfs, natifs de langue française, sans problème de vue ou d'audition non corrigés, ayant un bon niveau en orthographe ont été recrutés grâce à la base de données de participants aux expériences scientifiques du Centre d'Expérimentation sur la Parole (CEP) du LPL. Ils ont tous réalisé la tâche en débutant par la transcription des mots de la BECD puis ont poursuivi l'exercice en transcrivant les listes de pseudo-mots du DAP. 1 auditeur évalue 20 locuteurs en tout, dont 10 patients et 10 contrôles. Soit 1020 stimuli, dont 500 mots de la BECD et 520 pseudo-mots du DAP. Les auditeurs ont été dédommagés à hauteur de 20€ en ticket Kadeos.

4.2.2.3. Objectif

L'objectif de cette étude concerne l'évaluation du matériel linguistique. Nous avons précédemment postulé que l'emploi de pseudo-mots devrait permettre de contrôler les effets de restauration lexicale, nous souhaitons, par cette expérience, statuer sur les effets d'apprentissage du matériel linguistique.

En confrontant l'évolution des scores PPD au cours de la tâche pour le même nombre de mots versus de pseudo-mots, nous devrions mettre en exergue des éventuels effets d'apprentissage du matériel linguistique (comme c'est le cas pour les mots de la BECD, dans les expériences précédentes). Ainsi, si les scores restent stables au cours de la tâche cela montrera que les effets d'apprentissage sont neutralisés par l'emploi de pseudo-mots. En revanche, si ces scores évoluent au cours de l'expérimentation cela montrera que les auditeurs sont capables d'apprendre les pseudo-mots. Au regard des résultats de l'expérience précédente nous serons alors en mesure de déterminer si la neutralisation est imputable à la quantité de pseudo-mots ou à leur qualité, c'est à dire en tant que matériel linguistique non lexical.

4.2.3. Méthodologie d'un test de perception

Un test de perception de la parole suit plusieurs grands principes dans son organisation et dans sa passation. Ces principes seront présentés dans cette section, en revanche, les détails méthodologiques du design expérimental développé dans ces travaux sont à retrouver dans les sections méthodologiques consacrées à chacune des expériences conduites (4.2 et 5.2)

Les éléments nécessaires à un test de perception s'articulent autour des stimuli, du matériel et des participants. Nous préparons les stimuli, les enregistrements sont étiquetés et découpés comme souhaité sur le signal sonore. Nous déterminons et testons le protocole expérimental, et la fonctionnalité du matériel (casque, souris, écran, logiciel). Nous recrutons des auditeurs et ensuite s'en vient la passation. Tous les auditeurs sont dans des conditions d'écoute optimales, dans un environnement calme, par exemple en salle multipostes au sein du laboratoire ou dans un bureau isolé. Ils sont équipés d'un casque audio leur permettant de déterminer un niveau sonore confortable qu'ils conserveront pendant la durée de la passation. Ils sont installés devant un poste d'ordinateur, ils contrôlent le lancement et la répétition des stimuli audio à l'aide de la souris et transcrivent ce qu'ils perçoivent au moyen du clavier d'ordinateur.

Les stimuli sont présentés et les transcriptions sont récoltées au travers de la station automatisée de tests de perception auditive et visuelle, PercEval (André et al., 2003), développée au Laboratoire Parole et Langage. Cette station d'évaluation permet de programmer à l'avance la modalité et l'ordre de présentation des stimuli, ainsi que les modalités de recueil des réponses des auditeurs. Ce peut être un espace permettant à l'auditeur de taper ce qu'il souhaite dans la case de réponses, ou l'auditeur peut être amené à sélectionner une réponse parmi un choix affiché à l'écran. Dans les tests de perception, PercEval permet de présenter les stimuli aux participants et de collecter leurs réponses de manière organisée.

Afin d'obtenir des résultats fiables, les mêmes stimuli sont présentés à plusieurs auditeurs différents, cela permet de neutraliser un éventuel effet lié à l'auditeur. En plus d'être évalués par plusieurs auditeurs, les stimuli sont mélangés de façon qu'ils se retrouvent dans plusieurs positions différentes au sein de la passation. Ainsi, un même stimulus est présenté à plusieurs auditeurs dans le premier tiers de la passation, puis à d'autres auditeurs dans le second tiers, et enfin à d'autres auditeurs dans le dernier tiers de la passation. On s'assure ainsi qu'aucun effet de fatigabilité lié à l'effort que demande la tâche n'impacte les résultats. Ainsi, la randomisation de l'ordre de présentation des stimuli, le brassage entre les différents groupes de locuteurs et transcription, par plusieurs auditeurs différents permettent de s'assurer que les effets observés sont bien liés aux paradigmes de recherche et non imputables aux conditions de passation et de réalisation de la tâche (fatigabilité, apprentissage, familiarisation...).

Nous ferons référence à l'ensemble de ces aspects méthodologiques tout au long des travaux ici présentés.

4.2.4. Traitement des données de perception

Pour calculer les scores PPD et réaliser l'analyse statistique des données collectées lors des tests de jugement perceptif de l'intelligibilité, il est nécessaire d'opérer différents traitements sur les données brutes afin de résoudre des problèmes liés aux systèmes

d'encodage des différents outils numériques utilisés au cours de l'expérience. Les données brutes sont les données obtenues après les tests de jugement perceptif de l'intelligibilité, telles que organisées par PercEval, interface utilisée pour la collecte de ces données.

Ainsi, pour obtenir les scores d'intelligibilité au test de jugement perceptif de l'intelligibilité, objet des analyses statistiques, nous utilisons l'algorithme de calcul de distance entre deux chaînes de caractères Wagner Fischer. Pour que nos données puissent être traitées par l'algorithme, et compatibles avec la matrice de confusion, nous devons résoudre des problèmes d'encodage des données et pour cela nous opérons des transformations selon le traitement suivant.

Tout d'abord, ce traitement nécessite l'utilisation de « CygWin », collection de logiciels libres et d'outils Open Source, qui permet de simuler un environnement Unix, sous Windows en rendant possible l'exécution de logiciels, normalement non compatibles avec ce système d'exploitation. CygWin permet d'appliquer une suite de commandes et de traitement contenus dans un fichier à un autre fichier contenant des données sur lesquelles on souhaite réaliser des opérations. L'application d'une règle prédéfinie, un très grand nombre de fois, de manière automatique permet de limiter le nombre d'erreur.

En premier lieu, nous résolvons un problème d'encodage des données brutes, en effet, lorsqu'un auditeur propose une transcription qui contient une lettre accentuée, elle est codée par un symbole dans notre fichier de réponse. Par exemple, la transcription « é » apparaît « %E9 ». Nos conventions de correction visent à restaurer « é » à la place de « %E9 ». Nous remplaçons donc le symbole du fichier réponse par la lettre correspondante transcrite initialement par l'auditeur. CygWin nous permet d'appliquer les commandes de corrections de façon automatique sur l'ensemble du fichier de réponses brutes qui contient 56 390 valeurs. Nous vérifions manuellement qu'il ne reste pas de caractère non restauré dans notre fichier de réponse. Si nous identifions un caractère mal encodé, nous ajoutons une ligne de commande sur notre fichier de traitement, et nous ré-appliquons le traitement sur le fichier de réponse, jusqu'à ce que chaque problème de conversion soit résolu.

Nous obtenons ainsi des transcriptions corrigées, qui correspondent aux transcriptions initiales des auditeurs. Ces réponses corrigées sont les transcriptions que nous validons pour l'analyse des données. Ces réponses validées sont transmises au Laboratoire Informatique d'Avignon (LIA) qui phonétise les réponses valides à l'aide du LIAPhon (Bechet, 2001). A partir des transcriptions renvoyées par le LIA et à l'aide de CygWin, nous appliquons une nouvelle phonétisation basée sur les conventions de transcriptions de Lexique.org qui permettent d'encoder les transcriptions sur un seul caractère. La phonétisation de type Lexique.org, sur un seul caractère, est nécessaire pour que les données puissent être compatibles avec la matrice de confusion utilisée pour calculer les scores d'intelligibilité PPD.

Après ces étapes de pré-traitement, nous disposons de données compatibles avec la matrice des coûts en termes de distance (Levenshtein), et l'algorithme de Wagner Fischer. Nous comparons donc les transcriptions phonétiques des réponses des auditeurs aux tests de jugement perceptif de l'intelligibilité avec les transcriptions phonétiques des cibles qui devaient être produites par les locuteurs. Nous utilisons les transcriptions phonétisées pour l'étape de traitement suivante. Nous récupérons les données et à l'aide de CygWin, nous appliquons l'algorithme de calcul entre deux chaînes de caractères Wagner Fischer.

L'algorithme de Wagner Fischer permet le calcul de distance d'édition entre deux chaînes de caractères. Il propose un calcul général auquel il faut attribuer une distance entre chaque caractère. Pour cela, nous utilisons une matrice de confusion qui attribue un coût en termes de distance basé sur la théorie des traits, développé par A. Ghio dans le cadre de l'analyse des données DAP du projet de recherche C2Si (Ghio et al., 2018b). L'algorithme intègre la distance de Levenshtein, qui considère trois opérations d'éditations élémentaires : la suppression, l'insertion ou la substitution d'un caractère. Le nombre minimal d'opération d'éditations élémentaires nécessaire pour transformer une chaîne de caractère en une autre chaîne de caractère définit la distance d'édition entre deux chaînes de caractères (3.3.1 et 3.3.1.2 Méthode de calcul des scores perceptifs).

Ainsi, à l'issue du passage dans la matrice de confusion, nous obtenons les distances moyennes cumulées entre la cible qui devait être produite et la cible qui a été perçue, par stimulus. Soit, une distance comparative entre les transcriptions des cibles qui devaient être prononcées par les locuteurs et les transcriptions effectives des réponses des auditeurs.

Pour les besoins des calculs de scores de distances cumulées, nous comptons le nombre de caractères de la cible phonétique. Les scores de distance cumulée sont obtenus en divisant le score donné par la matrice de confusion, qui intègre l'algorithme de Wagner Fischer, obtenu pour chaque stimulus par le nombre de caractères de la cible phonétique. En moyennant les scores des 52 productions pour chaque locuteur, nous obtenons les scores d'intelligibilité, ou scores PPD Global, en termes de distance cumulée à la cible, qui représentent le nombre moyen de traits altérés par phonème (3.3.1.2. Méthode de calcul des scores perceptifs). Ces scores sont les variables dépendantes (VD) des analyses statistiques. Pour les besoins de cette expérience en particulier, dans laquelle il s'agit de mesurer l'évolution du score, un numéro a été assigné à chaque item produit par chaque locuteur, en fonction de l'ordre de la passation ; Ainsi, le premier stimulus transcrit est numéroté 1, le second stimulus transcrit est numéroté 2, et ainsi de suite jusqu'à 500 pour les items lexicaux BECD et jusqu'à 1040 pour les items non-lexicaux DAP.

Pour réaliser l'analyse statistique des résultats obtenus à l'expérimentation « BECD DAP », nous disposons d'un fichier contenant les scores d'intelligibilité (VD), le type d'auditeur (VI : Expert vs Naïf) et le type de matériel linguistique (VI : Mots de la BECD vs Pseudo-mots du DAP), ainsi que l'ordre de passation de chaque item. Pour l'analyse statistique, nous comparerons les scores de distance cumulée d'intelligibilité à l'aide d'une analyse de variance qui sera appliquée afin de tester les différences significatives des scores d'intelligibilité (VD) en fonction du matériel d'évaluation (VI : Mots VS pseudo-mots) et en fonction du groupe d'auditeurs (VI : Naïfs VS Experts).

Pour réaliser l'analyse statistique, au terme de notre expérience « Quantité Qualité », nous disposons d'un fichier contenant les scores de Déviation Phonologique Perçue (PPD) (VD) et le type de matériel linguistique (VI : Mots de la BECD vs Pseudo-mots du DAP), ainsi que l'ordre de passation de chaque item. Pour l'analyse statistique, nous comparerons les scores de distance cumulée d'intelligibilité à l'aide d'une analyse de variance qui sera appliquée afin de tester les différences significatives des scores d'intelligibilité (VD) en fonction du matériel d'évaluation (VI : Mots VS Pseudo-mots).

4.3. Résultats

Dans la section suivante nous présenterons en premier lieu les résultats de l'étude visant à évaluer la pertinence des deux matériaux linguistiques, mots et pseudo-mots conduite auprès de 18 auditeurs naïfs. Dans un second temps nous présenterons les résultats de l'expérience conduite auprès de 18 auditeurs experts en perception de la parole pathologique (orthophonistes). Enfin, nous exposerons les résultats de l'étude visant à déterminer si le contrôle des éventuels effets d'apprentissage est davantage imputable à la quantité ou à la qualité du matériel linguistique.

4.3.1. BECD – DAP

Une analyse de variance (ANOVA) a été conduite, afin de tester les effets simples et les effets d'interaction entre les différents facteurs et ainsi de déterminer les liens entre les variables testées. Nous avons utilisé le logscore PPD en variable dépendante (VD) et en variables indépendantes (VI) le groupe d'auditeur (facteur à 2 niveaux : Experts vs Naïfs), le type de matériel linguistique (facteur 2 niveaux : Mots BECD vs Pseudo-mots DAP), et le groupe de locuteurs (Patient et Témoins) en facteur. Les résultats montrent un effet simple des groupes de locuteurs, d'auditeurs et du matériel linguistique ($p < 0.001$). Cela signifie que, dans nos données, les groupes de locuteurs sont distingués, tout comme les groupes d'auditeurs et le type de matériel linguistique. Les résultats révèlent également un effet d'interaction significatif entre Locuteur et Auditeur ($p = 0.027$) et entre Matériel linguistique et Auditeur ($p = 0.047$). Cela montre que les locuteurs sont évalués différemment par les deux groupes d'auditeurs et que les groupes d'auditeurs ne proposent pas des scores équivalents avec les deux matériaux linguistiques. Cependant, les résultats ne montrent pas d'effet d'interaction Locuteur et Matériel linguistique ($p = 0.44$), donc les deux groupes de locuteurs ne sont pas évalués différemment par les matériaux linguistiques. En revanche, ils révèlent également un effet d'interaction Locuteur, Auditeur et Matériel linguistique ($p < 0.001$) ; les groupes d'auditeurs évaluent différemment les groupes de locuteurs en fonction du matériel linguistique utilisé.

Dans cette section, l'ensemble de nuages de points présentés illustrent le même type de données. On trouvera, en ordonnée les scores de déviation phonologique perçue (PPD) obtenus grâce aux transcriptions des auditeurs au test de jugement perceptif de l'intelligibilité ; en abscisse le numéro attribué à l'item en fonction de son ordre dans la passation, qui représente donc le décours temporel du temps de cette dernière. Le premier item transcrit est numéroté 1, le second item transcrit est numéroté 2... Ces numéros d'items vont de 1 à 500 pour le matériel linguistique lexical BECD et de 1 à 1040 pour le matériel linguistique non lexical DAP. Chaque point représente la moyenne des scores moyens attribués aux items en fonction de leur ordre dans la passation. La droite de régression linéaire illustre l'évolution globale des scores PPD au cours du temps de la tâche.

Les évaluations réalisées à partir de matériel linguistique lexical (Mots) montrent une grande variabilité dans l'attribution des scores, imputable aux effets de restauration et à la grande variabilité de la perception des troubles de la production de la parole (Jarzé et al., 2017b).

Pour tester la pertinence de l'utilisation de pseudo-mots, par rapport à des mots, pour évaluer l'intelligibilité nous avons comparé les scores PPD moyens attribués au cours d'une évaluation conduite avec des items lexicaux (Mots – BECD) avec les scores attribués au cours d'une évaluation avec des items non lexicaux (Pseudo-mots – DAP), par un groupe de 18 auditeurs naïfs (4.2.1. Design expérimental « BECD DAP »). Les Figure 20 et Figure 21 illustrent graphiquement les résultats obtenus.

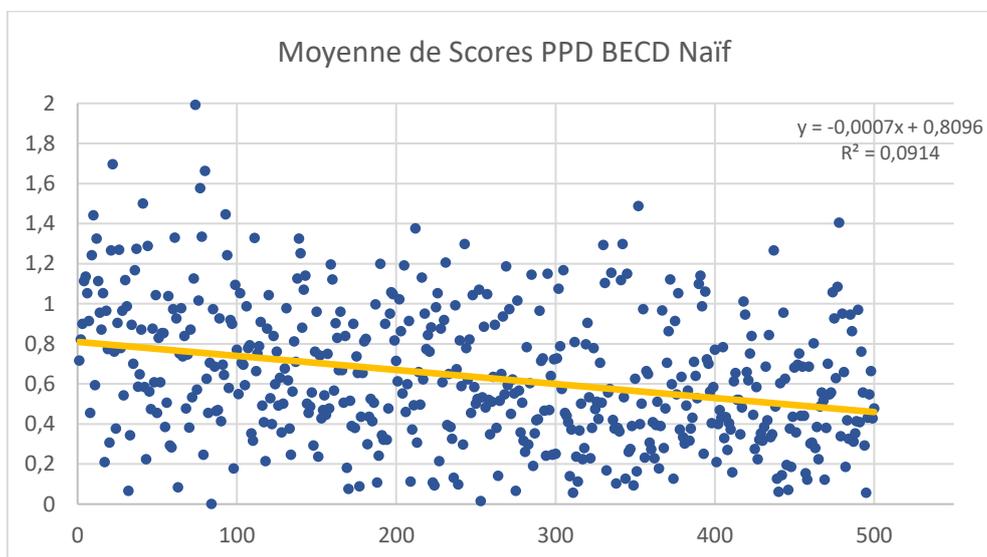


Figure 20 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item lexical BECD dans la passation

La Figure 20 montre une relation affine modérée et négative entre les scores PPD moyens et la position de l'item dans la passation. Cela signifie que les scores moyens de PPD baissent au cours de la tâche et que les auditeurs, à force d'être exposés aux stimuli, les apprennent et donc les restaurent. Cela se traduit par un score PPD plus faible à la fin de tâche, au 500^{ème} mot, par rapport aux 1^{ers} mots transcrits. Cette figure montre un effet d'amélioration de 0,0007 trait par phonème par essai, soit un abaissement de 0,35 trait par phonème pour 500 essais. Ainsi, les scores moyens passent de 0,85 pour l'essai 1 (première transcription) à 0,47 au 500^{ème} essai, ce qui représente une baisse de 44% du score moyen. Ce résultat révèle un effet d'apprentissage du matériel linguistique au cours de la tâche. Nos analyses statistiques confirment un effet de l'ordre des items dans la passation pour le matériel linguistique lexical (Mots BECD) Naïf ($F(1,498) = 50.12$, $r = -0.302$, $p < 0.001$)

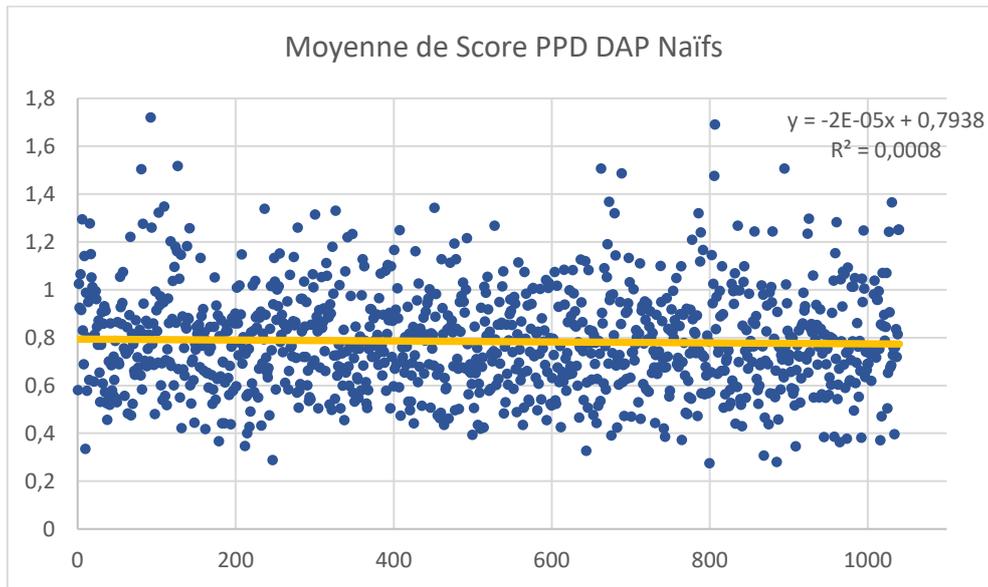


Figure 21 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item non lexical DAP dans la passation

La Figure 21 montre une relation affine forte et neutre entre les scores moyens d'intelligibilité et la position de l'item dans la passation. Cela signifie que les scores PPD moyens restent constants entre la transcription du premier pseudo-mot et celle du 1040^{ème} pseudo-mot. Cela ne révèle aucun effet lié au matériel linguistique non lexical – Pseudo-mots DAP – au cours de la tâche. Ceci est confirmé par les analyses statistiques qui montrent qu'aucun effet n'a été mis en exergue dans une évaluation basée sur l'utilisation de pseudo-mots (DAP) Naïf ($F(1,1038) = 0.7931$, $r = -0.028$, $p = 0.37$)

Cette étude confirme notre hypothèse selon laquelle l'utilisation de pseudo-mots permet d'abroger l'effet d'apprentissage du matériel linguistique. Ces résultats permettent de valider le critère de pertinence de l'emploi de pseudo-mots, par rapport aux mots dans l'évaluation de l'intelligibilité segmentale.

4.3.2. Naïf – Expert

Pour tester la fiabilité de l'utilisation de pseudo-mots pour évaluer l'intelligibilité nous avons comparé les scores PPD moyens attribués par deux groupes d'auditeurs – Naïfs et Experts (4.2.1. Design expérimental « BECD DAP »).

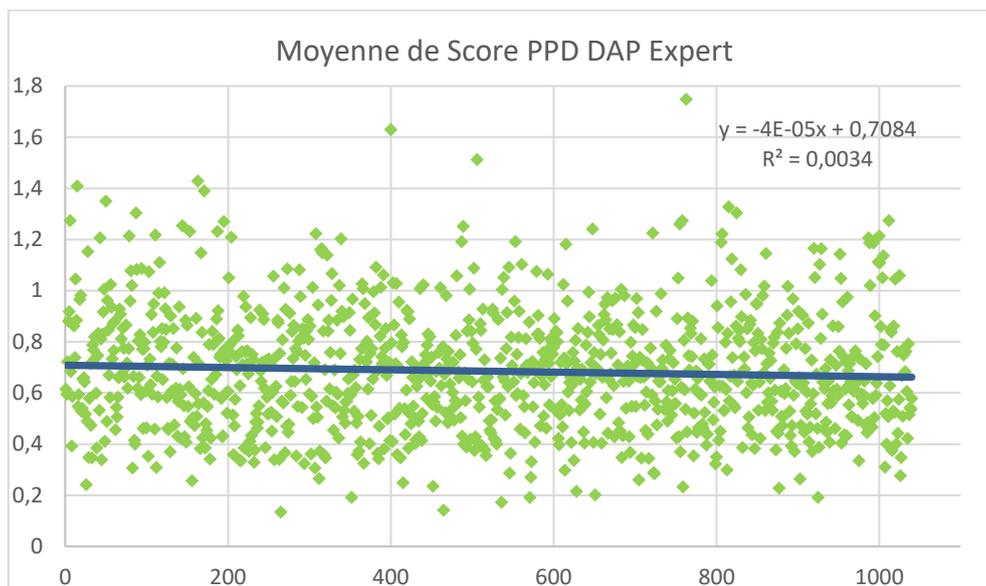


Figure 22 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item non lexical DAP dans la passation, avec des auditeurs experts

La Figure 22 montre une relation affine forte et neutre entre les scores moyens d'intelligibilité et la position de l'item dans la passation, tout comme la Figure 21. Cela indique que les scores moyens restent constants entre la transcription du pseudo-mot 1 et celle du 1040^{ème} pseudo-mot pour le groupe d'auditeurs experts. Ceci est confirmé par les analyses statistiques qui ne montrent aucun effet dans une évaluation basée sur l'utilisation de pseudo-mots (DAP) Expert ($F(1,1038) = 3.54$, $r = 0.034$, $p = 0.06$)

Après avoir testé l'effet d'apprentissage au cours de la tâche de DAP auprès d'une population d'auditeurs naïfs, qui a montré la pertinence de l'utilisation d'un tel matériau linguistique, nous avons testé cet effet auprès d'une population d'auditeurs cliniciens, spécialistes en écoute de la parole pathologique. Ceci afin de valider pleinement la pertinence de cette tâche. Les résultats n'ont révélé aucun effet d'apprentissage du matériau linguistique pour les deux groupes d'auditeurs, naïfs et experts. L'utilisation d'un très grand nombre de pseudo-mots dans une tâche d'évaluation de l'intelligibilité permet donc de neutraliser l'effet d'apprentissage du matériel linguistique et d'assurer la stabilité des scores au cours des évaluations.

4.3.2.1. DAP – BECD et Experts – Naïfs

Pour tester l'effet d'expertise auditive des cliniciens nous comparons l'évolution globale des scores au cours du déroulement de la tâche entre les deux groupes d'auditeurs, naïfs et experts. Nous nous intéressons donc à la différence observée entre les droites de régression linéaire.

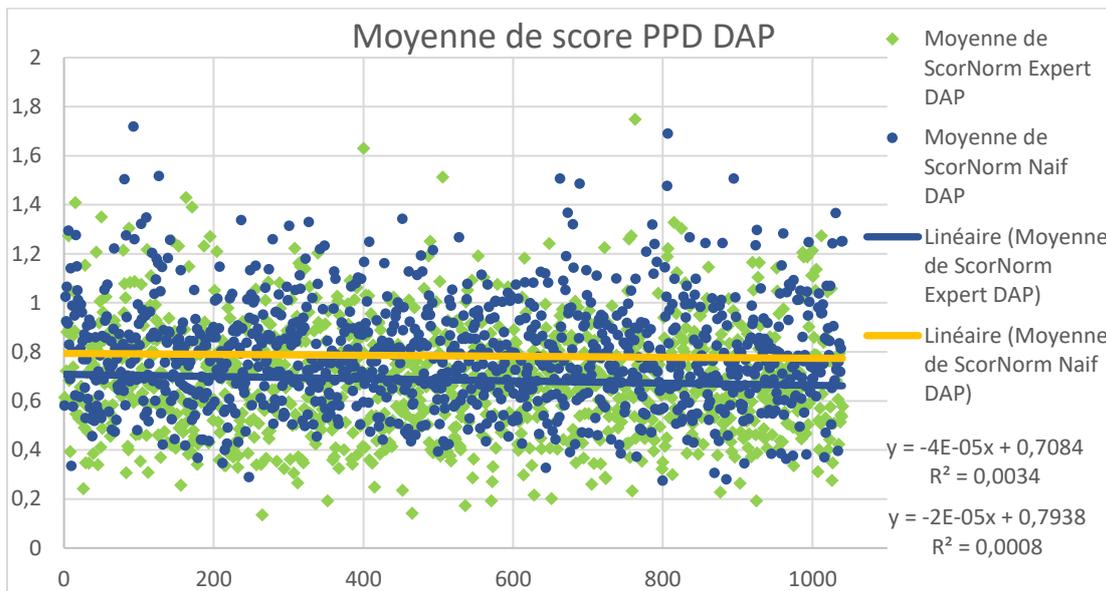


Figure 23 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type d'auditeur, naïfs et experts, pour le matériel linguistique non lexical DAP

La Figure 23 montre des relations affines fortes et neutres, entre les scores PPD moyens et la position de l'item dans la passation. Cela montre que les scores PPD moyens restent constants avec une évaluation par DAP (1040 items), quel que soit le degré d'expertise de l'auditeur.

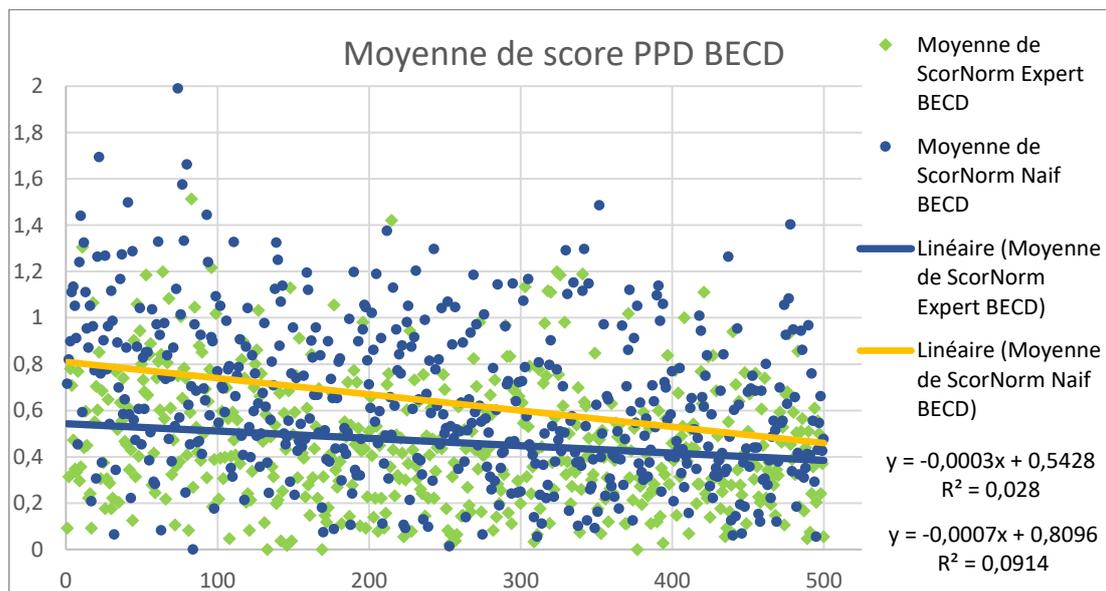


Figure 24 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type d'auditeur, naïfs et experts, pour le matériel linguistique lexical BECD

La Figure 24 montre des relations affines modérées et négatives entre les scores moyens d'intelligibilité et la position de l'item dans la passation. Cela montre que les scores de PPD moyens baissent avec une évaluation par la BECD (500 items), cette tendance s'observe quelle que soit la qualité d'expertise de l'auditeur.

La Figure 23 montre des relations affines fortes et neutres pour le DAP (Figure 21, Figure 22 et Figure 23) contrairement aux relations affines modérées et négatives observées pour la BECD (Figure 24 et Figure 20), entre les scores moyens d'intelligibilité et la position de l'item dans la passation. Cela montre que graphiquement les scores moyens l'intelligibilité restent constants avec une évaluation pour le DAP (1040 items) (cf : section précédente) et baissent avec une évaluation par la BECD (500 items), et ce quel que soit le degré d'expertise des auditeurs. Nos analyses statistiques ont confirmé un effet de l'ordre des items dans la passation pour le matériel linguistique lexical (Mots BECD) et ce pour les deux groupes d'auditeurs, Expert ($F(1,498) = 14.36, r = -0.197, p < 0.001$) et Naïf ($F(1,498) = 50.12, r = -0.09, p < 0.001$). Aucun effet n'a été relevé pour le matériel linguistique non lexical (Pseudo-Mots DAP).

En considérant la différence observée entre les droites de régression linéaire des deux groupes d'auditeurs pour des évaluations basées sur la tâche de DAP (Figure 23), on observe une différence de scores moyens entre les deux groupes d'auditeurs, Experts ($\mu = 0.685$) et Naïfs ($\mu = 0.787$), les experts proposant des scores plus faibles ce qui traduit une intelligibilité moins dégradée. Ces scores indiquent que les auditeurs cliniciens sont de meilleurs décodeurs que les auditeurs naïfs, probablement du fait de leur expertise auditive en écoute de la parole pathologique. En d'autres termes, cela signifie que les cliniciens décodent mieux les stimuli non lexicaux. Pour mesurer l'effet d'expertise sur la tâche de DAP, nous avons comparé les scores PPD moyens attribués par les deux groupes d'auditeurs. La différence observée s'est révélée significative ($p < 0.001$). Les experts proposent des scores PPD significativement plus bas que les auditeurs naïfs. Cette différence moyenne correspond à une différence moyenne de 0.1 trait d'écart moyen par phonème, ce qui est très faible. Nous avons donc cherché à évaluer la taille de cet effet en calculant un coefficient *d* de Cohen (« cohen.d » in R). Un 'd' aux alentours de 0.2 est considéré comme « faible », à 0.5 « médium » et à 0.8 « fort ». Le *d* de Cohen estimé pour le DAP, entre les auditeurs naïfs et experts est ($d : 0.102$) dit négligeable. La différence de scores PPD moyens attribués par les auditeurs naïfs et les auditeurs experts ne perturbe donc pas l'évaluation car sa faiblesse montre que les deux groupes d'auditeurs n'évaluent finalement pas différemment les locuteurs avec le DAP. Cela indique que l'effet d'expertise auditive des cliniciens est neutralisé lors d'une évaluation de l'intelligibilité par la tâche de DAP, même s'ils apparaissent indéniablement comme de « meilleurs décodeurs ».

Pour établir le sens du lien entre nos variables nous avons établi une matrice de corrélation qui intègre l'ordre de passation des items (VI : Order) à la place du groupe de locuteur. Cette matrice propose des résultats identiques à ceux des modèles linéaires établis pour les mêmes variables.

Les résultats confirment un effet de l'ordre des items dans la passation pour le matériel linguistique lexical (Mots BECD) et ce pour les deux groupes d'auditeurs, Experts ($F(1,498) = 14.36, r = -0.197, p < 0.001$) et Naïfs ($F(1,498) = 50.12, r = -0.09, p < 0.001$) (Figure 24). En revanche, aucun effet n'est mis en exergue dans une évaluation basée sur l'utilisation de pseudo-mots (DAP) et ce pour les deux groupes d'auditeurs, Experts ($F(1,1038) = 3.53, r = -0.058, p = 0.06$) et Naïfs ($F(1,1038) = 0.79, r = -0.028, p = 0.37$) (Figure 23).

Ces résultats indiquent que, dans une évaluation utilisant du matériel linguistique lexical (Mots BECD) il existe un effet d'apprentissage du matériel linguistique au cours de la tâche. L'accès au lexique mental permis pas l'emploi de mots favorise la restauration perceptive et leur mémorisation lorsque leur production est dégradée. Les listes de mots courtes et fermées classiquement utilisées (Auzou & Rolland-Monnoury, 2006; Blanc et al., 2014) facilitent ces effets. Défavorables dans le cadre de l'évaluation clinique de l'intelligibilité, car conduisant à la sous-évaluation des déficits, ces effets se trouvent neutralisés dans une évaluation s'appuyant sur du matériel linguistique non lexical (Pseudo-mots DAP). En effet, les pseudo-mots DAP permettent d'obtenir une évaluation stable au cours du temps, critère qui valide la fiabilité de la tâche et du matériel linguistique, et de neutraliser l'effet d'apprentissage et de mémorisation, et donc les effets de restauration perceptive liés au matériel linguistique.

4.3.3. Quantité – Qualité

Dans cette étude, nous avons cherché à déterminer si le contrôle de la restauration lexicale était davantage imputable à la qualité du matériau linguistique (pseudo-mots vs mots) ou à la quantité d'items présentés. Une analyse de variance (ANOVA) a été conduite, afin de tester les effets simples entre le score PPD (VD) et les groupes de locuteurs (patient et contrôle) pour chaque type de matériau linguistique (mots vs pseudo-mots). Les résultats montrent que les deux matériaux linguistiques permettent de distinguer les groupes de locuteur en termes de score PPD, de façon significative. (BECD ($F(1,9498) = 1723.4$, $r = 0.153$, $p < 0.001$) et DAP ($F(1,19368) = 2907$, $r = 0.103$, $p < 0.001$))

Dans cette section, l'ensemble de nuages de points présentés illustrent le même type de données. On trouvera, en ordonnée les scores de déviation phonologique perçue (PPD) obtenus grâce aux transcriptions des auditeurs au test de jugement perceptif de l'intelligibilité ; en abscisse le numéro attribué à l'item en fonction de son ordre dans la passation, qui représente donc le décours temporel du temps de cette dernière. Le premier item transcrit est numéroté 1, le second item transcrit est numéroté 2... Ces numéros d'items vont de 1 à 500 pour le matériel linguistique lexical BECD et de 1 à 520 pour le matériel linguistique non lexical DAP. Chaque point représente la moyenne des scores moyens attribués aux items en fonction de leur ordre dans la passation. La droite de régression linéaire illustre l'évolution globale des scores PPD au cours du temps de la tâche.

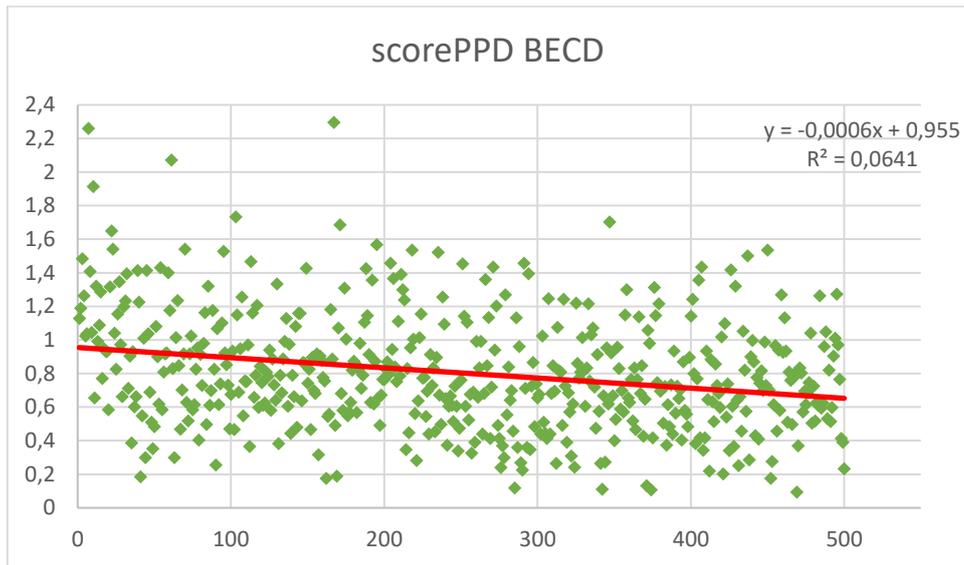


Figure 25 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item lexical BECD dans la passation

La Figure ci-dessus montre une relation affine forte et négative entre les scores moyens d'intelligibilité et la position de l'item dans la passation. Cela signifie que les scores moyens de PPD baissent au cours de la tâche et que les auditeurs, à force d'être exposés aux stimuli, les apprennent et donc les restaurent. Cela se traduit par un score PPD plus faible à la fin de tâche, au 500^{ème} mot, par rapport aux 1^{ers} mots transcrits. Cette figure montre un effet d'amélioration de 0,00066 trait par phonème par essai, soit un abaissement de 0,33 traits par phonème pour 500 essais. Ainsi, les scores moyens passent de 0,97 pour l'essai 1 (première transcription) à 0,64 au 500^{ème} essai, cela représente une baisse de 34% du score moyen. Cela révèle un effet d'apprentissage du matériel linguistique au cours de la tâche. Nos analyses statistiques confirment un effet de l'ordre des items dans la passation pour le matériel linguistique lexical (Mots BECD) Naïf ($F(1,498) = 38.72, r = 0.004, p < 0.001$). Nous répliquons donc ici les résultats observés en 4.3.1.

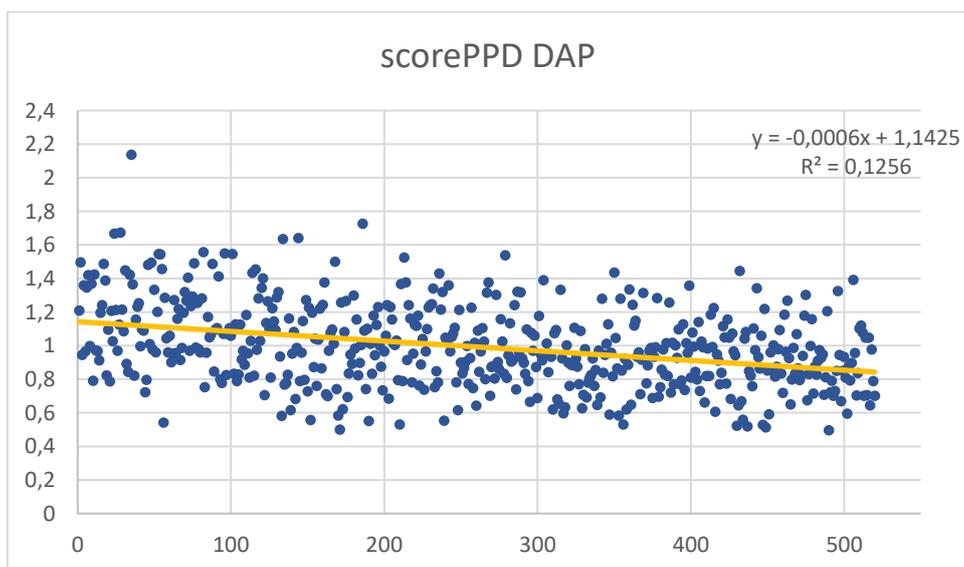


Figure 26 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item non lexical DAP dans la passation

La Figure ci-dessus montre une relation affine forte et négative entre les scores moyens d'intelligibilité et la position de l'item dans la passation. Cela signifie que les scores moyens de PPD baissent au cours de la tâche et que les auditeurs, à force d'être exposés aux stimuli, les apprennent et donc les restaurent. Cela se traduit par un score PPD plus faible à la fin de tâche, au 500^{ème} pseudo-mot, par rapport aux 1^{ers} pseudo-mots transcrits. Cette figure montre un effet d'amélioration de 0,00066 trait par phonème par essai, soit un abaissement de 0,33 trait par phonème pour 500 essais. Ainsi, les scores moyens passent de 1,19 pour l'essai 1 (première transcription) à 0,86 au 500^{ème} essai, cela représente une baisse de 28% du score moyen. Cela révèle un effet d'apprentissage du matériel linguistique au cours de la tâche. Nos analyses statistiques confirment un effet de l'ordre des items dans la passation pour le matériel linguistique non lexical (pseudo-mots) ((F(1,518) = 64.78, r = 0.006, p < 0.001)

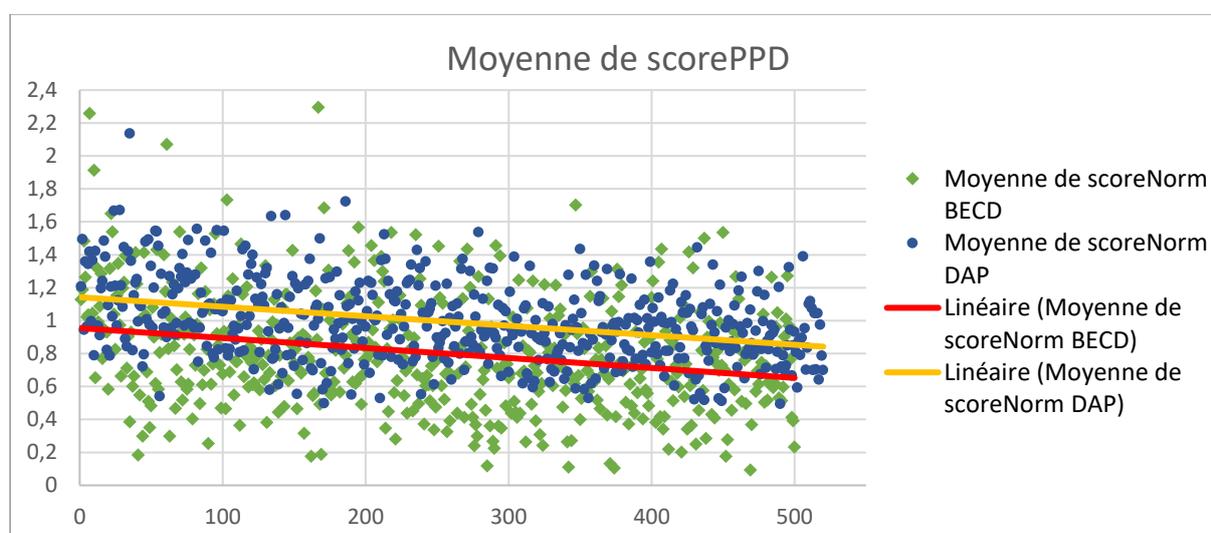


Figure 27 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type de matériel linguistique, mots de la BECD et pseudo-mots du DAP

La figure précédente montre des relations affines modérées et négatives entre les scores moyens d'intelligibilité et la position de l'item dans la passation. Cela montre que les scores de PPD moyens baissent avec une évaluation avec un nombre restreint d'item (50), cette tendance s'observe quelle que soit la qualité du matériel linguistique.

Ce design expérimental permet de comparer l'évolution des scores PPD au cours de la répétition de 50 mots et de 52 pseudo-mots. Les scores moyens PPD obtenus par DAP ($\mu = 0.99$) sont significativement plus élevés ($p < 0.001$) que ceux obtenus avec une évaluation avec les mots de la BECD ($\mu = 0.80$), cela signifie qu'une évaluation basée sur des pseudo-mots est plus sévère qu'une évaluation basée sur des mots de lexique. La taille de cet effet mesurée avec un coefficient d de Cohen ($d = -0.153$) nous indique que cette différence est faible. Elle représente 0.2 trait moyen d'écart par phonème. La différence de scores moyens observée entre ces deux types d'évaluation est significative. Ces résultats montrent que les scores calculés à partir d'une évaluation perceptive basée sur des mots sont moins élevés, ce qui peut être expliqué par le recours aux informations de haut niveau lexicales, l'auditeur a déjà rencontré la majorité de ces mots dans son

expérience linguistique. Ces items sont connus et stockés dans le lexique mental de l'auditeur. En revanche, les scores calculés à partir d'une évaluation perceptive basée sur des pseudo-mots sont plus élevés probablement car l'auditeur n'a jamais rencontré ces formes non lexicales auparavant ce qui neutralise tout effet de restauration.

Cependant, les deux matériaux linguistiques montrent la même évolution de score PPD au cours du temps de la tâche : on observe une baisse moyenne de 0,33 trait entre la 1^{ère} et la 500^{ème} transcription, cela représente une baisse de score moyen de 31% au cours de la tâche. Ces résultats montrent un effet d'apprentissage des pseudo-mots au cours de la tâche, lorsque c'est une liste fermée de 52 items non lexicaux qui est présentée aux auditeurs. Cela suggère que l'auditeur maintient à jour ses connaissances linguistique en intégrant les pseudo-mots à son répertoire de formes linguistiques. Il mobilise et réinvesti ces nouvelles connaissances dans son évaluation perceptive. Il est envisageable que les mécanismes perceptifs en jeu soient analogues à ceux des effets de restauration lexicale, mais appliqués pour des formes non lexicales. Cet effet d'apprentissage des pseudo-mots soulève la question de l'emploi d'un très grand nombre de mots pour une évaluation perceptive, mais ceux-ci ne permettraient pas de contrôler les effets de restauration lexicale, ni ne résoudraient la problématique de l'équilibre phonétique. D'autre part, les pseudo-mots permettent de contrôler la position et la fréquence d'occurrence de chaque phonème, ils présentent également l'avantage de n'avoir jamais été entendus auparavant par l'auditeur.

La comparaison de ces résultats, qui montrent une évolution à la baisse des scores PPD au cours de la tâche (Figure 26), avec ceux de l'expérience précédente, qui montrent une évolution stable des scores PPD au cours de la tâche (Figure 21), indique que c'est la quantité de pseudo-mots disponible qui permet de contrôler les effets d'apprentissage du matériel linguistique.

L'ensemble de ces résultats suggère donc qu'un très grand répertoire de pseudo-mots est nécessaire pour une évaluation perceptive de la parole, puisqu'ils permettent de contrôler les effets de restauration lexicale et d'apprentissage du matériel linguistique.

Ces premières expériences destinées à valider l'approche perceptive par Décodage Acoustico-Phonétique de pseudo-mots montrent des résultats concluants à partir des scores PPD Globaux. La suite de nos analyses portera sur l'identification des segments altérés que la méthode de calcul des scores locaux permet d'entreprendre.

5. Scores Perceptifs Segmentaux

Cette seconde étude porte sur le corpus de production de pseudo-mots du projet de recherche C2SI qui a été évalué par des tests de jugement perceptif de l'intelligibilité par décodage acoustico-phonétique. Les scores obtenus (scores PPD global et local), dont le calcul est basé sur la théorie des traits, caractérisent l'atteinte au niveau segmental. Il est donc possible de proposer une analyse fine de ces altérations pour examiner la contribution des différents segments à l'intelligibilité et ce sans proposer de nouvelles expériences de perception impliquant des modifications du signal acoustique (Chen et al., 2013; Cole et al., 1996; Fogerty & Kewley-Port, 2009; Kewley-Port et al., 2007). Dans un premier temps nous prendrons soin de décrire le corpus et la population (40 contrôles et 87 patients). Ensuite, nous présenterons la méthodologie développée et employée pour l'analyse descriptive des scores perceptifs locaux et des transcriptions des auditeurs en termes de phonèmes. Enfin nous exposerons les résultats de cette analyse.

5.1. Corpus C2Si

Dans le cadre du projet de recherche C2SI (Astésano et al., 2018; Woisard et al., 2020), un corpus de production de pseudo-mots a été enregistré auprès d'une cohorte de 127 locuteurs. Ces enregistrements ont déjà fait l'objet d'études reposant sur des tests de perception qui ont permis de recueillir des transcriptions d'auditeurs naïfs, de montrer la sensibilité de la tâche, ainsi que d'éprouver la méthode de calcul des scores PPD globaux.

5.1.1. Population

40 sujets contrôles, sans trouble de la production de la parole, sans trouble visuel ou auditif non corrigé, n'ayant jamais suivi une rééducation orthophonique et 87 patients traités pour un cancer de la cavité buccale ou de l'oropharynx, soit 127 locuteurs ont été enregistrés sur différentes tâches de production de la parole.

Parmi les locuteurs du groupe patients, 7 locuteurs ont été enregistrés 2 fois, avec un minimum de 6 mois entre les deux enregistrements. Dans les graphiques descriptifs suivants, ces mêmes patients seront traités comme 2 patients différents, compte tenu du délai (min 6 mois) entre les deux recueils de production. De plus, les analyses de ces productions sont réalisées de manière indépendante, pour le DAP, par exemple, deux scores seront attribués au même locuteur en fonction, respectivement, des enregistrements n°1 et n°2. Cela porte donc notre cohorte de locuteurs à 134.

Dans ce corpus, 55% des locuteurs sont de sexe masculin et 45% féminin. Dans le groupe patient, 39 femmes et 55 hommes ont été enregistrés, soit 87 locuteurs, pour un total de 94 enregistrements (7 locuteurs enregistrés deux fois, cf : §1), et dans le groupe témoin ce sont 21 femmes et 18 hommes qui ont été enregistrés, ainsi qu'un locuteur dont le sexe n'est pas renseigné dans les métadonnées, soit un total de 40 locuteurs témoins

(Figure 28). Dans chaque groupe, patient et témoin, les locuteurs sont appariés en âge (Tableau 14).

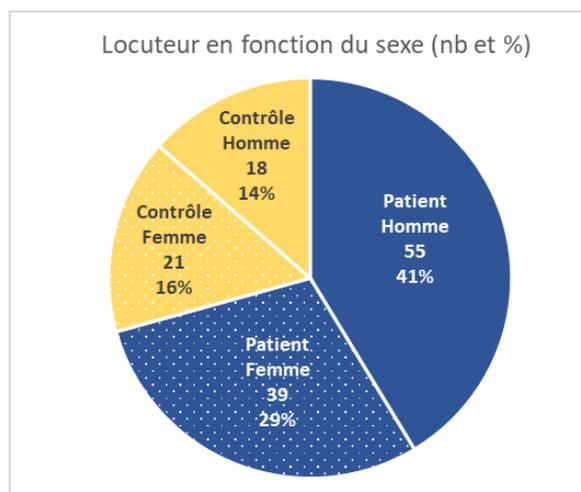


Figure 28 : Répartition des locuteurs en fonction du groupe (patient vs contrôle) et du sexe (homme et femme), en nombre par catégorie et en pourcentage du total des locuteurs

Age	Contrôles		Patients	
	Femme	Homme	Femme	Homme
Min	35	30	51	36
Max	76	79	87	86
Moy	59,1	61,3	66,7	65,3

Tableau 14 : Tableau de l'âge des locuteurs, min, max et moyenne par sexe et par groupe de locuteur

Plus d'hommes que de femmes ont été inclus dans ce corpus car l'incidence des cancers de la cavité buccale et de l'oropharynx est supérieure pour ceux-ci. De même, la moyenne d'âge est plus élevée pour les patients car l'incidence de survenue de ces cancers se situe après 50 ans (1.2.1 Cancers ORL généralités).

Les 40 sujets contrôles inclus dans ce corpus répondent aux caractéristiques suivantes : - ne pas avoir eu un cancer localisé dans les VADS, - ne pas avoir eu d'AVC avec séquelles, - ne pas souffrir de pathologies neurodégénératives, - ne pas bégayer, - ne pas souffrir de trouble i. de la vision (habileté à la lecture), ii. auditif grave et non corrigé, iii. cognitif (eg : attentionnel), - savoir lire, écrire et compter (bon niveau de lecture et niveau certificat d'étude).

Pour pouvoir être inclus dans ce corpus, les 87 patients devaient remplir une liste de critères spécifiques, i.e. critères d'inclusion : - avoir eu un cancer de la cavité buccale et/ou de l'oropharynx, - avoir été traité par chirurgie et/ou radiothérapie et/ou chimiothérapie, - être à plus de 6 mois de la fin du traitement, pour s'assurer de la stabilité des troubles de la production, qu'ils soient audibles ou non.

5.1.2. Métadonnées cliniques

Concernant le groupe de patients, plusieurs métadonnées cliniques ont été collectées. L'ensemble des données cliniques présentées dans la section suivante sont précisément décrites dans la partie 1.2. Nous rapporterons ici les valeurs qui concernent

spécifiquement la population de patients inclus dans ce corpus. Pour une description générale et détaillée des données cliniques ici quantifiées, se référer aux parties 1.2.1 Cancers ORL généralités et 1.2.2 Les traitements. Nous présenterons les données relatives à la localisation, en termes de macro-classes et sous-classes, nous décrirons les différents stades TNM représentés dans ce corpus. Nous évoquerons les traitements dont ont bénéficié les patients, et les délais après ceux-ci.

L'histologie la plus fréquente pour les cancer de la cavité buccale est le carcinome épidermoïde (HAS), qui a concerné 95% des patients inclus dans ce corpus. Les 5% restant étaient atteints d'un carcinome adénoïde kystique. Dans ce corpus, deux macro-classes de localisations ont été renseignées : la cavité buccale, à l'avant du conduit vocal, et l'oropharynx situé plus en arrière du conduit vocal (Figure 2). Pour 39% des patients de ce corpus, l'affection était localisée dans la Cavité Buccale (atteintes localisées sur le plancher buccal, la langue, la mandibule et la région rétro-molaire). Pour 61% des patients, l'affection était située dans l'Oro-Pharynx, (atteintes localisées sur les amygdales, la base de la langue et le voile du palais). Si l'atteinte est plus étendue, c'est-à-dire qu'elle touche plusieurs organes, par exemple les amygdales et le voile du palais, elle est appelée « oropharynx » et est classée dans la localisation du même nom. Les critères de sélection des patients (post-traitement) ne permettent pas de connaître précisément la localisation de la tumeur, celle-ci a été estimée aux vues des séquelles physiologiques observées. Les taux (en %) de patient dans chaque sous-classe de localisation tumorale estimée sont reportés dans la Figure 29 et rendent compte de l'hétérogénéité du groupe.

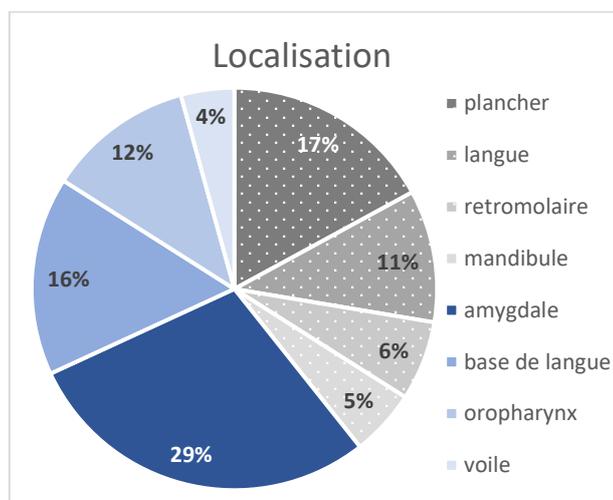


Figure 29 : Répartition des patients (en %) selon la localisation estimée de la tumeur, en gris les organes de la cavité buccale et en bleu ceux de l'oropharynx

En termes de classification TNM (qui permet de rendre compte du stade d'un cancer), les métadonnées étaient bien renseignées pour le stade T, mais moins bien pour le stade N (18% de données manquantes). En revanche, le stade M n'a pas été renseigné. Nous rappelons ici brièvement la définition de cette classification : « La lettre T est l'initiale de tumeur et correspond à la taille de la tumeur ; la lettre N est l'initiale de 'node', ganglion en anglais, et indique si des ganglions lymphatiques ont été ou non envahis ainsi que le degré d'envahissement ; la lettre M est l'initiale de métastase et signale la présence ou l'absence de métastases. » (*Définition classification TNM, 2019*).

Ce corpus inclut des patients atteints des quatre stades TMN-t de la classification. Les tumeurs des patients ont été classées T1 pour 12% d'entre eux et T2 pour 40%. Les tumeurs de 'petites' tailles T1 et T2 représentent environ la moitié du corpus de patient, 52%. Les tumeurs de tailles plus conséquentes concernent 48% des patients de ce corpus. 14% ont été atteints d'une tumeur classée T3 et 34% d'une tumeur classée T4 (Figure 30). Les 3 stades de la classification TNM-n sont également représentés, ainsi que le stade 0. En effet, un cancer implique 'forcément' une tumeur, mais pas toujours un envahissement ganglionnaire. Pour 26% des patients de ce corpus les cellules cancéreuses n'ont pas atteint les ganglions, soit environ un quart de la population. Les patients classés dans le stade N1 représentent 20% de ce corpus. 31% des patients ont été diagnostiqués au stade N2 et enfin 5% au stade N3.

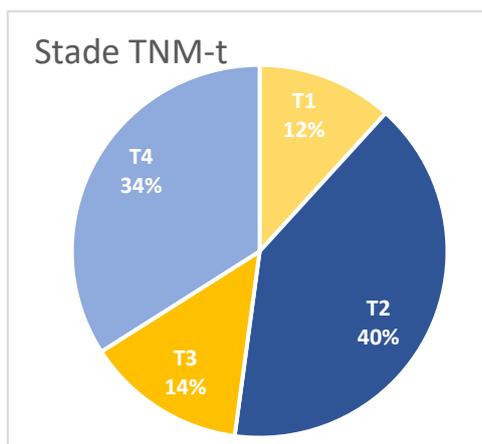


Figure 30 : Répartition de la population de patients en fonction de la taille de la tumeur (en %)

Dans ce corpus, les patients ont bénéficié de plusieurs combinaisons de traitement différentes illustrées dans la Figure 31. Les codages « ChirT » et « ChirN » désignent respectivement la chirurgie de la tumeur « T » et des ganglions « N ». La codage « RT » fait référence à la radiothérapie et l'abréviation « Chimio » désigne la chimiothérapie. Ces différents traitements peuvent être prodigués seuls ou être combinés les uns avec les autres, selon le protocole de soin établi par l'équipe médicale. Ainsi, dans ce corpus, 81% des patients ont bénéficié d'un traitement par chirurgie de la tumeur et des ganglions. Ils ne sont que 6% à avoir bénéficié d'une chirurgie des ganglions sans que celle-ci ne soit combinée à une chirurgie de la tumeur (ChirN_RT et ChirN_Rt_Chimio). Pour seulement 10% des patients du corpus, l'équipe médicale n'a pas inclus de chirurgie dans le protocole de soin (RT et RT_Chimio). En plus des traitements chirurgicaux, 35% des patients ont bénéficié d'une radiothérapie et 41% ont bénéficié d'une radiothérapie et d'une chimiothérapie. Ils sont 1% à avoir bénéficié seulement d'un traitement par chimiothérapie complémentaire aux chirurgies. 3% des patients ont bénéficié d'une chirurgie de la tumeur non combinée avec une chirurgie des ganglions (ChirT_RT_Chimio). Enfin, la radiothérapie est le seul traitement à avoir été prodigué seul, pour 1% des patients.

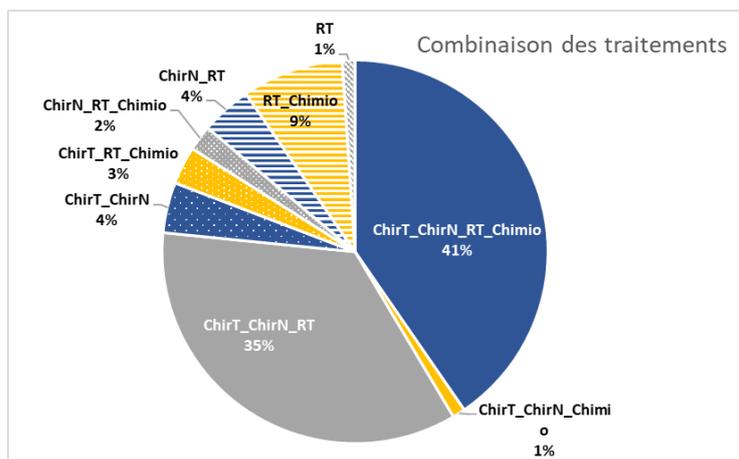


Figure 31 : Type de traitement reçus par les patients, exprimé en pourcentage. « RT » Radiothérapie ; « Chimio » Chimiothérapie ; « ChirT » Chirurgie de la tumeur ; « ChirN » Chirurgie des ganglions ; lorsque le patient a reçu une combinaison de traitement, l'ordre est indiqué par l'ordre des abréviations (« ChirN_Rt » signifie que le patient a bénéficié d'une chirurgie des ganglions, puis d'une radiothérapie)

Dans ce corpus, les métadonnées du type de chirurgie pratiquée, du nombre et du type de reconstruction dont les patients ont bénéficié, de la survenue ou non d'une récurrence du cancer, sont renseignées. Néanmoins, ces aspects ne présentent pas des métadonnées exploitables pour notre propos, nous ne détaillerons donc pas ces aspects du corpus.

Enfin, dans le cadre du recueil de parole pathologique impliquant des séquelles fonctionnelles, il est important de s'assurer de la stabilité du trouble de parole. Dans le cadre de la cancérologie ORL, un délai d'au minimum 6 mois après la fin du traitement est adéquate pour garantir cette stabilité. C'est le délai minimum retenu pour inclure un patient dans l'étude (C2SI). En moyenne, dans ce corpus, les patients ont terminé leurs traitements depuis 63,4 mois (5,2 années), dans l'intervalle de [6 mois – 239 mois] ou [0,5 année – 19,9 années]. Plus de 37% des patients se trouvent dans un délai post-traitement supérieur à la moyenne, ce taux conséquent témoigne de l'augmentation du taux de survie et souligne donc l'importance des questions liées à l'évaluation de la parole et de la prise en charge clinique du handicap communicationnel.

Dans la partie suivante nous présenterons la méthodologie employée pour l'analyse descriptive des scores PPD locaux et des transcriptions des auditeurs.

5.2. Méthodologie

Dans cette étude nous comparons les scores locaux (phonème à phonème) entre deux groupes de locuteurs (patients et contrôles) pour différentes catégories de phonèmes, et plus particulièrement les phonèmes vocaliques. Pour ces derniers segments, nous regardons les phonèmes de substitution transcrits en cas d'erreur.

5.2.1. Contexte

La tâche de Décodage Acoustico-Phonétique basée sur l'utilisation de pseudo-mots, développée dans le cadre du projet de recherche C2SI (Annexe B, p206), a déjà fait l'objet

de plusieurs études visant à démontrer la fiabilité et la sensibilité de la méthodologie de la tâche, du matériel linguistique employé et de la méthode de calcul des scores de Déviation Phonologique Perçue. Ces études emploient un test de jugement perceptif, dans lequel un auditeur écoute et transcrit ce qu'il perçoit. Ces tests se basent sur la méthodologie d'un test de perception qui suit plusieurs grands principes : les items perçus sont randomisés entre eux et entre les groupes de locuteurs, chaque item produit est transcrit par plusieurs auditeurs, les scores calculés représentent la moyenne des différentes transcriptions, comme explicité en (4.2.3 Méthodologie d'un test de perception)

Deux campagnes de tests majeures ont permis de collecter plus de 19000 transcriptions, selon le protocole d'un test de perception de la parole. Les scores d'intelligibilité globale (score PPD global) ont été calculés au travers de la matrice de coût, basée sur la théorie de traits articulatoires, et intégrant l'algorithme de Wagner-Fischer. La décomposition de ces scores permet d'atteindre le niveau segmental en récupérant les scores PPD locaux phonème à phonème. Ainsi, le coût, en termes de distance entre la cible à produire et la cible perçue, en nombre de traits d'écart, est calculé pour chaque phonème.

Une première étude explorant les scores PPD globaux et locaux a été conduite en collaboration avec le docteur Laura Monestier, médecin ORL, dans le cadre de son master en linguistique expérimentale. Ces travaux, portés sur le groupe patient, concernaient les scores PPD globaux en lien avec les métadonnées des patients collectées lors de la création du corpus. Les résultats ont montré un effet significatif de la taille de la tumeur sur l'altération des segments de parole, en défaveur des tumeurs de grosse taille. L'examen des scores PPD locaux des segments consonantiques et vocaliques au travers de leurs caractéristiques articulatoires a montré que globalement, les segments réalisés à l'arrière du conduit vocal (nasalité, postériorité) et présentant un fort degré de constriction (occlusion, hauteur) sont les segments les plus dégradés (Lalain et al., 2022; Monestier et al., 2021).

Dans cette présente étude, nous nous intéresserons aux scores PPD locaux de l'ensemble des segments produits par les deux groupes de locuteurs. Nous examinerons les potentielles différences entre les deux groupes de locuteurs, patients et contrôles. Nous analyserons la nature des segments dégradés et préservés pour les deux groupes afin de faire émerger les différences et similitudes. Enfin, nous considérerons les réponses des auditeurs afin de dégager d'éventuelles tendances communes.

5.2.2. Principes

Nous avons conduit cette analyse à partir des réponses (transcriptions) des auditeurs, collectées dans le cadre des études de (Ghio et al., 2018b; Ghio, Lalain, et al., 2020; Rebourg, 2018) visant à tester la tâche de DAP du projet C2SI et à éprouver la méthode de calcul des scores PPD globaux. Grâce à Alain Ghio, ingénieur de recherche au LPL, nous avons pu « pénétrer » la matrice de coût et ainsi obtenir les scores locaux phonèmes à

phonèmes, sur lesquels portent nos analyses. Celles-ci concernent les deux groupes de locuteurs inclus dans le corpus, des patients traités pour un cancer de la cavité buccale ou de l'oropharynx et des sujets contrôles. La mise en perspective des observations faites sur le groupe patient au travers de celles réalisées sur le groupe contrôles devrait permettre d'éclairer les aspects fondamentaux de la production et de la perception de la parole au travers d'une analyse fine des segments, et plus particulièrement des segments vocaliques. Les différents segments analysés sont présentés dans le paragraphe suivant.

Les groupes consonantiques, appelés également clusters, sont des unités composées d'au moins deux consonnes / sons consonantiques, le mot « spectacle » par exemple, possède trois groupes consonantiques différents « sp », « ct », « cl ». Ces mêmes groupes peuvent se rencontrer dans d'autres mots « spectre », « clef », « cactus ». Il en existe une grande variété en français, ce qui justifie leur apparition dans la sélection du matériel linguistique de ce corpus. 16 formes ont été retenues en position CC1 (pr tr kr gr br fr pl kl fl st bl sk sp gl dr ps) et 32 formes en position CC2 (st ks rd rs kt rn pl gr dr kl rj lt rv vr gz rp tr rt bl rm pr kr sk br sp rk fr fl rb gl ps pt) pour générer des pseudo-mots contenant des groupes consonantiques : « spibla », « lucti », « clantu », « blardu », « decto » ... Les consonnes, dont la production produit les sons consonantiques, sont des unités phonémiques autonomes. Lors de l'élaboration des pseudo-mots, les consonnes avec une correspondance phonémique ont été sélectionnées (exception faite du « h » qui n'a de correspondance phonologique). Ainsi, 19 phonèmes consonantiques ont été retenus pour les positions C1 et C2 (p t k b d g v z ʒ f s ʃ r l m n ñ j).

Les phonèmes vocaliques retenus pour l'élaboration des pseudo-mots sont au nombre de 8 pour la position V1 (a i y u o e ã ě) et 7 en position V2, (a i y u o ã ě), l'équipe a choisi de ne pas faire apparaître le /e/ en finale de pseudo-mot pour éviter une trop grande variabilité de production. Dans notre étude, les symboles utilisés pour illustrer les segments vocaliques sont issus du système de transcription de Lexique.org. Ils diffèrent légèrement de ceux employés dans l'API. Ainsi, les phonèmes /i/, /y/, /u/ et /a/ sont identiques. En revanche, nous utilisons les symboles /E/ pour illustrer l'archiphonème qui regroupe les voyelles moyennes antérieures /e/, /ə/, /ɛ/ et /ø/, et l'archiphonème /Ô/ qui regroupe les voyelles moyennes postérieures /o/ et /ɔ/. Enfin, pour les phonèmes nasalisés /ã/ et /ě/ nous employons les symboles /@/ pour illustrer /ã/ et /μ/ pour illustrer /ě/ et /œ/. Les différents systèmes de traitements de données utilisés ne supportent pas tous les caractères spéciaux spécifiques de l'API, les convention Lexique.org utilisent des symboles plus facilement compatibles entre les logiciels. Les symboles que nous retenons dans cette présente étude sont donc /i/, /y/, /u/, /a/, /E/, /Ô/, /@/ et /μ/. Pour les détails concernant l'élaboration des pseudo-mots et la sélection des phonèmes, se référer à la partie 3.3.1 Méthode de Décodage acoustico-phonétique et aux publications (Lalain et al., 2020).

Nous portons une attention particulière à la description des segments vocaliques puisqu'ils feront l'objet d'un intérêt spécifique pour l'approche par mesures acoustiques. En effet, ces segments représentent un ensemble de productions relativement homogènes (Meunier, 2007), notamment puisqu'ils impliquent la vibration des cordes vocales,

propriété acoustique appelée voisement. De plus, les propriétés articulatoires de ces segments peuvent être décrites au moyen d'un ensemble de traits. Ces propriétés peuvent être représentées dans un triangle vocalique. La place de chaque segment produit dans ce triangle est déterminée par des mesures des propriétés acoustiques, qui permettent d'extraire des informations numériques quantifiables, valeurs employées pour le calcul de différentes métriques. Ces spécificités, en complément des données de la littérature (cf 2.2.1.2) constituent des atouts qui suggèrent que les segments vocaliques sont pertinents pour conduire une étude approfondie au travers des scores PPD locaux, des transcriptions effectives des auditeurs et d'analyses acoustiques.

Dans ces études, nous nous sommes concentrée sur les substitutions (quand un caractère unique est substitué à un autre) en laissant de côté les phénomènes d'insertion et d'élision. L'étude menée sur (i) les PPD locaux considère toutes les substitutions présentes dans le corpus, en revanche l'étude (ii) des transcriptions effectives des auditeurs ne considère que les substitutions par une même catégorie de segments. Nous détaillerons la sélection de ces substitutions dans la section méthodologique dédiée.

L'étude (ii) des transcriptions effectives des auditeurs a porté sur les segments vocaliques. Les huit segments vocaliques cibles du corpus ont été retenus (/i/, /y/, /u/, /a/, /E/, /Ô/, /@/ et /μ/). Pour chaque segment produit le score PPD local a été calculé, soit 36815 scores tous segments vocaliques et groupes de locuteurs confondus. Un tiers des réponses concernent le groupe contrôle (33,94%, n = 12495) et deux tiers concernent les patients (66,06%, n = 24320).

Nous avons identifié et catégorisé trois types de réponses ayant un impact direct sur les scores PPD locaux. Tout d'abord les réponses correctes, c'est-à-dire les transcriptions qui correspondent à la cible attendue pour lesquelles les scores PPD locaux sont de 0, ne montrant aucune déviation entre la cible attendue et la cible perçue (en moyenne 64,62% des réponses, n = 23791). Ramené au nombre total de réponses transcrites dans chaque groupe de locuteur, le taux de réponses correctes dans le groupe contrôle (80,53%, n = 10062) est supérieur à celui du groupe patient (56,45%, n = 13729). Ensuite nous avons porté notre attention sur les réponses présentant une déviation entre la cible attendue et la cible perçue. Nous avons relevé que des cibles avaient été transcrites par plusieurs caractères se traduisant par plusieurs phonèmes, et d'autres par un seul caractère, soit un seul phonème de substitution a été transcrit. Ces phénomènes représentent, respectivement, des insertions et des substitutions.

Les insertions sont donc caractérisées par la transcription de plusieurs phonèmes perçus là où un phonème unique est attendu. La déviation entre la cible attendue et la cible perçue est importante et représente, en moyenne, 6,77 traits d'écart moyen par phonème. Les insertions représentent 6,59% des réponses (n = 2427) tous segments et groupes de locuteurs confondus. Dans chaque groupe de locuteurs, par rapport au nombre de total de réponses transcrites, les insertions représentent 3,45% des données pour le groupe contrôle (n = 431), et 8,21% pour le groupe patient (n = 1996). Les substitutions sont, elles, caractérisées par la transcription d'un seul phonème à la place

d'un autre phonème unique attendu et représentent une part importante des réponses transcrites, 28,78% en moyenne, (n = 10597). En moyenne, la déviation phonologique perçue par rapport à la cible attendue est de 1,56 traits d'écart moyen par phonème. Par rapport au nombre total de réponses transcrites dans chaque groupe de locuteurs, la part de substitutions est plus importante dans le groupe patient (35,34%, n = 8595) que dans le groupe contrôle (16,02%, n = 2002). Nous constatons donc qu'il y a plus de diversité de réponses, et donc plus de déviations entre la cible attendue et la cible perçue dans le groupe patient par rapport au groupe contrôle, ce qui était attendu puisque les analyses des scores PPD globaux avaient montré des différences significatives entre les deux groupes de locuteurs, moins élevés en faveur des contrôles.

Nous avons choisi d'analyser les substitutions en perspective avec les réponses correctes et de nous affranchir des insertions. Celles-ci présentent une déviation moyenne importante mais représentent un faible pourcentage des données disponibles (segments vocaliques, sPPD local). D'autre part, la grande diversité d'insertions relevées (n = 272, tous segments et groupes de locuteurs confondus) rend les données difficiles à synthétiser et ne permet pas de conduire des analyses fiables, en effet plus de la moitié des segments insérés (n = 137) n'ont qu'une seule et unique occurrence.

En revanche, sur le total des substitutions (n = 10597) pour les huit segments vocaliques ciblés, nous relevons 22 segments substitués différents. Nous avons observé la fréquence d'occurrence, en pourcentage, de chaque segment substitué. Nous avons sélectionné ceux dont la fréquence d'occurrence totale moyenne était supérieure à 0,2%, en s'assurant que la fréquence d'occurrence des segments exclus dans chaque catégorie de segment individuel ciblé n'était pas supérieure à 0,75%. Ceci pour obtenir suffisamment d'occurrences des phonèmes de substitutions afin d'avoir du poids dans nos analyses et d'assurer la fiabilité des résultats. Cela nous conduit à écarter 0,89% des substitutions, soit 94 transcriptions. Nous retenons alors 10503 substitutions, dont 1987 pour le groupe contrôle et plus de quatre fois plus pour le groupe patient (n = 8516). Notons que le nombre d'erreurs en termes de substitutions dans le groupe contrôle est significativement inférieur à celui du groupe patient, ce qui entraîne un déséquilibre dans la quantité de données disponibles entre les deux groupes locuteurs. Cette différence reflète à la fois le déséquilibre présent dans le corpus, et la 'réalité' des deux groupes, un groupe contrôle ne présentant pas de trouble de la production de la parole contrairement au groupe de patients pour qui les séquelles fonctionnelles du cancer et de ses traitements impactent les mécanismes physiologiques de la production de la parole. Ceci produit un effet perceptif sur l'auditeur pour qui le nombre d'erreur reflète les difficultés de transcriptions.

Nous retenons alors neuf segments de substitutions principaux : (/i/, /y/, /u/, /a/, /E/, /Ô/, /@/, /μ/ et /§/). Les différents symboles utilisés sont décrits dans le paragraphe 2 de la présente section, à l'exception du phonème /ô/ symbolisé à présent par /§/. Nous étudierons donc les transcriptions les plus fréquentes pour les différents segments vocaliques ciblés et présenterons les analyses descriptives de ces transcriptions effectives dans la partie 5.3.2 Transcriptions des auditeurs.

Dans cette étude notre objectif est d'exploiter de manière optimale les données obtenues à l'issue des tests de jugement perceptif par DAP. En premier lieu, (i) nous testons les différences de scores PPD locaux entre les macro-classes phonémiques des clusters (ou groupes consonantiques), des consonnes et des voyelles, pour chaque groupe de locuteur, patient et contrôle, au moyen d'analyses statistiques. Nous proposons également une investigation approfondie des segments vocaliques en relation avec leurs caractéristiques articulatoires : la nasalité, l'antériorité et la hauteur. Nous proposons de comparer les différences et les similitudes entre les deux groupes de locuteurs. Nous soutenons l'hypothèse selon laquelle les segments les moins bien perçus présenteront des similitudes entre les deux groupes de locuteurs et, qu'en comparaison avec le groupe contrôle, mettront en évidence un effet délétère de la pathologie sur la production du groupe patient. Après avoir analysé les scores PPD locaux moyens au moyen d'une analyse statistique, (ii) nous proposons une analyse descriptive des transcriptions effectives des auditeurs pour les segments vocaliques. Nous comparons la cible qui devait être produite avec les transcriptions effectives des auditeurs les plus fréquentes, au niveau phonémique (phonétisation Lexique.org).

5.3. Résultats

5.3.1. PPD locaux

Nos analyses statistiques sur les scores PPD locaux ont été conduites avec le logiciel RStudio. Une représentation graphique de la normalité, pour chaque classe de phonème testée, montre que nos données n'ont pas une distribution gaussienne. Nous avons dû nous orienter vers des tests non paramétriques. Le test de Wilcoxon a été utilisé pour tester les comparaisons deux à deux. Pour estimer la taille de chacun des effets nous avons calculé le coefficient D de Cohen.

Une première série d'analyses concerne les macro-classes phonémiques (Clusters, Consonnes et Voyelles) avec une comparaison intergroupe (Patient vs Contrôle) des scores PPD locaux. Une seconde série d'analyses propose d'examiner les différences intergroupes (Patient vs Contrôle) pour les segments vocaliques selon leurs caractéristiques articulatoires (résonance, position de la langue sur les axes d'aperture et antéro-postérieur).

5.3.1.1. Macro-classes phonémiques

5.3.1.2. Cluster, consonnes, voyelles

L'objectif de cette partie de l'étude est d'évaluer quels segments, en termes de macro-classes phonémiques, sont les moins dégradés, et ce entre les deux groupes de locuteurs. Nous espérons pouvoir dégager des unités plus robustes face aux dégradations induites par la pathologie. Ainsi, nous comparons les scores moyens PPD locaux entre les macro-classes de cluster (CLU), de consonnes (CSN) et de voyelles (VOY), pour les deux groupes de locuteurs patient et contrôle (Figure 32).

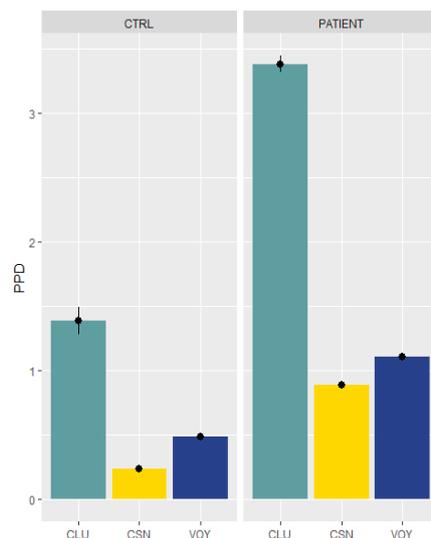


Figure 32 : Scores PPD locaux moyens selon les macro-classes phonémiques pour les groupes de locuteurs patients et contrôles

Nos résultats montrent que tous les contrastes sont significatifs. Les clusters, ou groupes consonantiques, sont les segments les plus dégradés, de façon significative pour les deux groupes de locuteurs, patients ($p < 0.001$) et contrôles ($p < 0.001$). De plus, nous remarquons que le score PPD local CLU moyen des patients ($\mu = 3,38$, $sd = 1,61$) est plus de deux fois supérieur au score PPD local moyen CLU des contrôles ($\mu = 1,38$, $sd = 1,6$), ceci nous permet de rappeler au lecteur que, dans l'ensemble de cette étude, les scores des patients sont toujours significativement plus élevés que ceux des contrôles, et ce pour tous les contrastes testés.

La macro-classe phonémique regroupant les consonnes seules apparaît comme la classe la plus préservée. Les scores PPD locaux CSN moyen des patients ($\mu = 0,88$, $sd = 1,62$) et des contrôles ($\mu = 0,23$, $sd = 0,96$) sont les plus bas. Les différences de scores entre les CSN et les deux autres classes (CLU et VOY) sont significatives pour les patients ($p < 0.001$) et les contrôles ($p < 0.001$). Cela suggère que ce sont les segments les moins altérés dans la production de parole.

Enfin, la classe de phonèmes vocaliques VOY est significativement moins dégradée que les CLU, mais est significativement plus altérée que les CSN, pour les patients ($p < 0.001$) et pour les contrôles ($p < 0.001$). Néanmoins, les scores PPD locaux VOY moyens sont plus proches de ceux des consonnes que des clusters, pour les patients ($\mu = 1,11$, $sd = 2,07$) et pour les contrôles ($\mu = 0,48$, $sd = 1,51$).

A ce stade de l'étude, nos résultats montrent que ce sont les segments consonantiques (CSN) qui sont les mieux préservés, quel que soit le groupe de locuteurs. Ils sont suivis de près par les segments vocaliques (VOY). Quant aux groupes consonantiques (CLU), ce sont les segments les plus altérés, et donc également les plus difficiles à identifier. Nous choisissons de les écarter de notre étude notamment parce qu'une étude a déjà montré leur statut particulier pour l'évaluation de l'intelligibilité (A. K. Marczyk et al., 2020). En revanche, ces résultats nous mènent à explorer la faible, mais parfois significative, différence entre les macro-classes de consonnes et de voyelles. Résultats que nous présenterons dans la partie suivante.

5.3.1.3. Consonnes, voyelles

Cette partie de l'étude a pour objectif d'évaluer la différence de score PPD local moyen entre les classes phonémiques des consonnes et des voyelles. Nous nous sommes demandé si certaines caractéristiques acoustiques des phonèmes vocaliques pouvaient expliquer les différences de scores PPD locaux CSN moyens et PPD locaux VOY moyens, les scores de ces derniers étant supérieurs. Nous avons donc divisé le groupe de phonèmes vocaliques en séparant les segments vocaliques (VOYorales) de ceux comportant le trait de nasalité (VOYnasales). Nos résultats nous conduiront à étudier la relation entre CSN et VOYorales, dans un second temps.

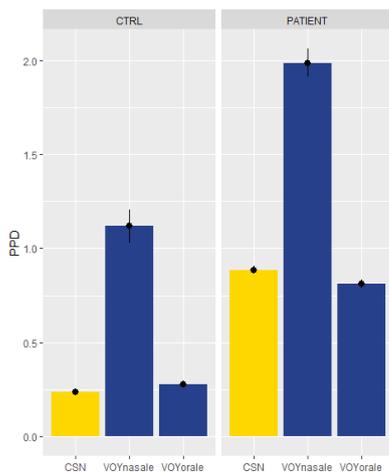


Figure 33 : Scores PPD locaux moyens selon les macro-classes phonémiques consonnes et les sous classes voyelles nasales et voyelles orales, pour les groupes de locuteurs patients et contrôles

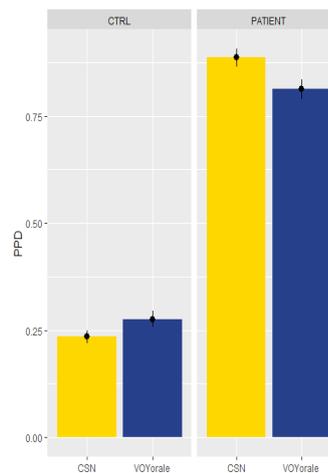


Figure 34 : Scores PPD locaux moyens selon la macro-classe phonémique consonnes et la sous classe voyelles orales, pour les groupes de locuteurs patients et contrôles

Nos résultats montrent que dans la classe VOY, le score PPD local VOYnasales est plus élevé que le sPPD local VOYorales, pour les patients (VOYnasales $\mu = 1,96$, $sd = 2,92$; VOYorales $\mu = 0,81$, $sd = 1,57$), et pour les contrôles (VOYnasales $\mu = 1,09$, $sd = 2,44$; VOYorales $\mu = 0,28$, $sd = 0,95$). Cet effet est significatif pour les deux groupes de locuteurs patients ($p < 0.001$) et contrôles ($p < 0.001$). Cela signifie que les différences de scores PPD locaux CSN moyens et PPD locaux VOY moyens sont majoritairement imputables aux VOYnasales. Les scores PPD locaux CSN moyens patients ($\mu = 0,88$, $sd = 1,62$) et contrôles ($\mu = 0,23$, $sd = 0,96$), sont relativement proches des scores PPD locaux VOYorales moyen. Ainsi, la comparaison entre ces deux catégories (CSN et VOYorales) se révèle pertinente puisqu'elle permettra de déterminer si elles sont dégradées de façon équivalente ou si l'une se montre plus robuste que l'autre.

Dans la comparaison de la différence de score PPD locaux moyens entre les consonnes CSN et les voyelles orales VOYorales, nous observons une tendance inverse entre les scores PPD locaux CSN moyens et PPD locaux VOYorales moyens (Figure 34). Dans le groupe contrôle les voyelles orales ($\mu = 0,28$) ont un PPD local moyen supérieur de 0.05 par rapport aux consonnes (PPD local CSN moyen $\mu = 0,23$). Cela signifie que, dans ce

groupe, ce sont les consonnes qui sont les moins altérées car elles présentent les scores moyens les moins élevés. Cette différence, bien que très faible, s'est révélée significative ($p < 0.001$). À l'inverse du groupe contrôle, dans le groupe patient, les VOYorales ($\mu = 0,81$) ont un score PPD local moyen inférieur de 0,07 par rapport aux CSN ($\mu = 0,88$) ; ce sont donc les voyelles orales qui sont les moins altérées, par rapport aux consonnes, dans le groupe de patients. Cette différence n'est pas significative ($p = 0,57$). Il apparaît donc que, pour les contrôles, les consonnes sont les segments les plus préservés, et ce significativement par rapport aux voyelles orales, alors que dans le groupe patient, nos résultats n'indiquent pas de différence.

Dans la suite de cette étude, nous choisissons d'investiguer plus amplement la macro-classe des voyelles. En effet, nos premiers résultats montrent les voyelles orales et les consonnes ne sont pas perçues différemment pour le groupe de patient. En revanche, dans le groupe contrôle, ce sont les voyelles qui présentent le degré d'altération le plus important. Ces arguments, en complément de ceux de la littérature les placent au rang de segments d'intérêt pour la suite de ces travaux.

5.3.1.4. Classes articulatoires vocaliques

Nous présenterons en détails les analyses conduites sur la relation entre les traits articulatoires vocaliques et les scores PPD locaux. Nous comparerons les voyelles orales et les voyelles nasales, nous nous intéresserons à l'antériorité en divisant le groupe de voyelles orales en fonction de la position de la langue dans la cavité buccale, antérieure ou postérieure, et nous diviserons à nouveau le groupe de voyelles orales en trois, relativement à la hauteur en considérant les voyelles hautes, moyennes et basses.

5.3.1.5. Voyelles Nasales vs Orales

L'observation des scores PPD locaux moyens pour chaque voyelle (Figure 35) montre que deux archiphonèmes se détachent du lot, avec des scores plus élevés indiquant qu'ils sont les plus dégradés. /@/ et /μ/ ont pour trait commun la nasalité. Partant de ce constat, nous avons divisé la classe de voyelles en fonction de cette caractéristique. Ainsi, deux groupes de phonèmes se sont distingués, les segments vocaliques oralisés (/i/, /y/, /u/, /E/, /Ô/ et /a/) et les segments vocaliques nasalisés (/@/ et /μ/). Une fois regroupées, on observe une grande différence de score PPD locaux moyens entre ces deux catégories (Figure 36).

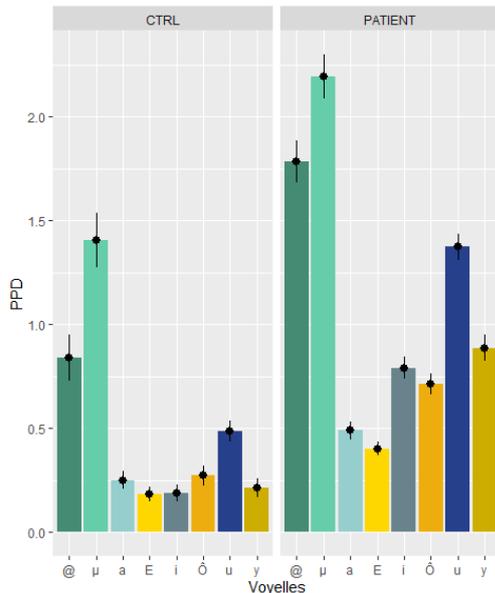


Figure 35 : Scores PPD locaux moyens par segment vocalique, pour les groupes de locuteurs patients et contrôles

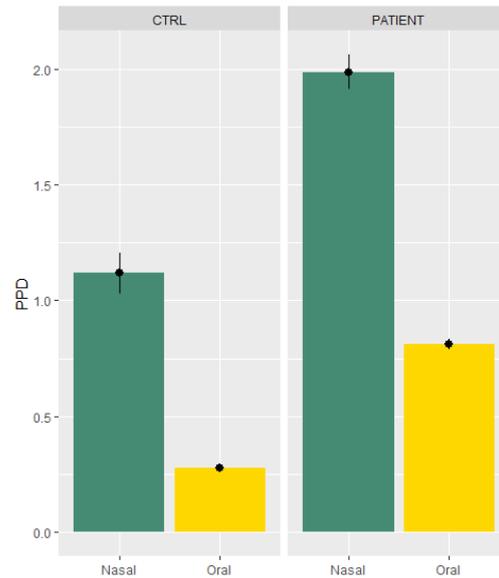


Figure 36 : Scores PPD locaux moyens selon les sous classes phonémiques nasales et orales, pour les groupes de locuteurs patients et contrôles

Nos résultats montrent que les voyelles orales sont toujours mieux réalisées, par rapport aux voyelles nasales, puisqu'elles obtiennent un sPPDlocal moyen systématiquement inférieur à celles-ci. Dans le groupe patient, le score PPD local moyen des voyelles nasales (Nasal, $\mu = 1,96$, $sd = 2,92$) est plus de deux fois supérieur à celui des voyelles orales (Oral, $\mu = 0,81$, $sd = 1,57$). Dans le groupe contrôle, l'écart se creuse davantage, les score PPD local moyen Nasal ($\mu = 1,09$, $sd = 2,44$) est plus de trois fois supérieur à celui des Oral ($\mu = 0,28$, $sd = 0,95$). Cette différence est significative pour les deux groupes de locuteurs patients ($p < 0.001$) et contrôles ($p < 0.001$). Afin d'évaluer la taille de l'effet nous avons calculé un coefficient 'd' de Cohen qui montre en effet 'moyen' pour les patients ($d = 0,59$) et pour les contrôles ($d = 0,57$).

Cela signifie que, quel que soit le groupe de locuteurs, patient ou contrôle, les voyelles orales sont toujours moins dégradées que les voyelles nasales. Les voyelles orales semblent plus robustes face à la dégradation pathologique. Les deux analyses suivantes ne portent que sur cette classe de phonèmes.

5.3.1.6. Antériorité

Le groupe de phonèmes constitué par les sons vocaliques oraux, ou regroupés sous l'appellation voyelles orales, peut être divisé selon des critères articulatoires. Dans cette analyse, nous avons considéré la position de la langue dans la cavité buccale sur l'axe antéro-postérieur. Ainsi, deux catégories se distinguent, les voyelles antérieures, dites d'avant ('Front' : /i/, /y/ et /E/) et les voyelles postérieures, dites d'arrière ('Back' : /u/ et /Ô/).

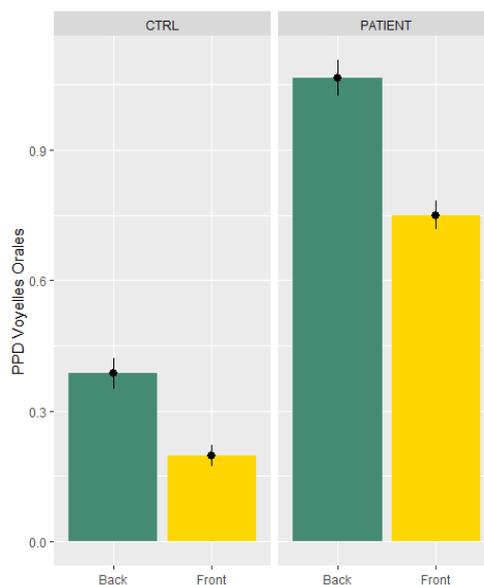


Figure 37 : Scores PPD locaux moyens selon la position avant (Front) vs arrière (Back) pour les groupes de locuteurs patients et contrôles

Nos résultats montrent que les voyelles d'avant Front sont moins altérées que les voyelles d'arrière Back, pour les deux groupes de locuteurs. Les score PPD local Front moyen est moins élevé pour les patients ($\mu = 0,75$, $sd = 1,52$), et les contrôles ($\mu = 0,19$, $sd = 0,82$), par rapport aux sPPD local Back moyen pour les patients ($\mu = 1,06$, $sd = 1,7$) et les contrôles ($\mu = 0,39$, $sd = 1,05$). La différence de scores PPD local moyen Front vs Back est significative pour les patients ($p < 0.001$) et les contrôles ($p < 0.001$). Pour estimer la taille de cet effet nous avons calculé le coefficient 'd' de Cohen, l'effet observé est faible pour les patients ($d = 0,19$) et aussi pour les contrôles ($d = 0,20$). Bien que faible, l'effet d'antériorité est confirmé ; les voyelles d'avant sont moins altérées que les voyelles d'arrière, dans les deux groupes de locuteurs.

5.3.1.7. Hauteur

Dans cette seconde partie d'analyse, nous avons considéré la position de la langue dans la cavité buccale, sur l'axe d'aperture, au travers de la hauteur. Nous distinguons donc trois catégories, les voyelles hautes ('High' : /i/, /y/ et /u/), les voyelles moyennes ('Medium' : /E/ et /Ô/) et les voyelles basses ('Low' : /a/). Nous testerons chaque contraste pour chaque groupe de locuteurs.

La Figure 38 illustre les scores PPD locaux moyens par catégorie de hauteur. Dans le groupe contrôle, on observe peu de différence de score PPD locaux moyen entre les voyelles basses (Low : $\mu = 0,25$, $sd = 0,92$) et les voyelles moyennes (Medium : $\mu = 0,24$, $sd = 0,83$). Ce sont ces dernières qui présentent le sPPD local moyen le moins élevé. Les voyelles hautes semblent être les plus dégradées dans ce groupe, puisqu'elles ont les scores PPD local moyen les plus élevés (High : $\mu = 0,30$, $sd = 0,98$). Dans le groupe de patients, on observe des tendances différentes. En effet, ce sont les voyelles basses qui présentent le score PPD local moyen le moins élevé (Low : $\mu = 0,49$, $sd = 1,36$). De plus la différence de sPPD local moyen entre les voyelles basses et les voyelles moyennes (Medium : $\mu = 0,61$, $sd = 1,21$) est inverse et plus élevée que chez les contrôles. Enfin, ce

sont les voyelles hautes qui présentent le score PPD local moyen le plus élevé ($\mu = 1,02$, $sd = 1,75$), indiquant que ce sont ces segments qui sont les plus altérés.

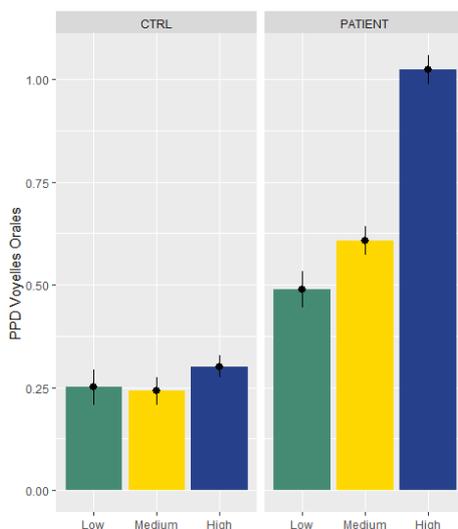


Figure 38 : Scores PPD locaux moyens en fonction de la hauteur (High, Medium, Low), pour les groupes de locuteurs patients et contrôles

Dans le groupe patient, tous les contrastes sont significatifs. Les voyelles basses sont significativement mieux réalisées que les voyelles moyennes et hautes ($p < 0.001$). Les voyelles moyennes sont donc plus altérées que les voyelles basses, mais significativement moins que les voyelles hautes ($p < 0.001$). Les voyelles hautes sont significativement ($p < 0.001$) plus altérées que les voyelles moyennes et basses. Le calcul du coefficient 'd' de Cohen pour chaque contraste a montré que les effets repérés étaient faibles ($d = [0,26 ; 0,32]$), mais néanmoins présents. Dans le groupe contrôle tous les contrastes testés ne sont pas significatifs. En effet, aucun effet n'est mis en lumière entre les voyelles moyennes et les voyelles hautes ($p = 0,15$). En revanche, bien que la différence de score moyen soit très faible, les voyelles basses sont significativement plus dégradées que les voyelles moyennes ($p < 0.001$). Les voyelles basses sont également significativement moins altérées que les voyelles hautes ($p = 0.005$). Pour les contrastes significatifs, le calcul du coefficient 'd' de Cohen a montré des effets faibles (Low – Med 'd' = 0,25 et Low – High 'd' = 0,30) mais présents. Ces résultats signifient que ce sont les voyelles hautes qui sont les plus dégradées, par rapport aux voyelles moyennes ou basses. Elles présentent les scores PPD moyen locaux les plus élevés et les contrastes avec les autres catégories de voyelles sont significatifs.

Pour résumer l'ensemble des résultats concernant les classes articulatoires vocaliques présentés dans cette partie, nous retiendrons plusieurs éléments essentiels. Tout d'abord, les scores PPD locaux moyens des patients sont significativement supérieurs à celui des contrôles, et ce pour tous les contrastes testés. De plus, les deux groupes montrent plusieurs effets similaires, avec un effet majoré par la pathologie dans le groupe patient. Les voyelles orales sont moins altérées que les voyelles nasales. Les voyelles d'avant sont moins altérées que les voyelles d'arrière. Les voyelles moyennes et basses sont moins altérées que les voyelles hautes. Les résultats de ces analyses indiquent que ce sont les segments réalisés à l'arrière et en haut dans la cavité buccale qui sont les plus altérés. En effet, si on observe les scores PPD locaux moyens de chaque voyelle (Figure 39), la voyelle

haute et postérieure /u/ présente le score le plus élevé. C'est le segment vocalique le plus altéré. Ces résultats font écho aux résultats obtenus par L. Monestier qui, dans son étude, identifiait la région postérieure haute comme la région de réalisation articulaire la plus délicate pour les patients, que ce soit pour les voyelles ou pour les consonnes (Lalain et al., 2022).

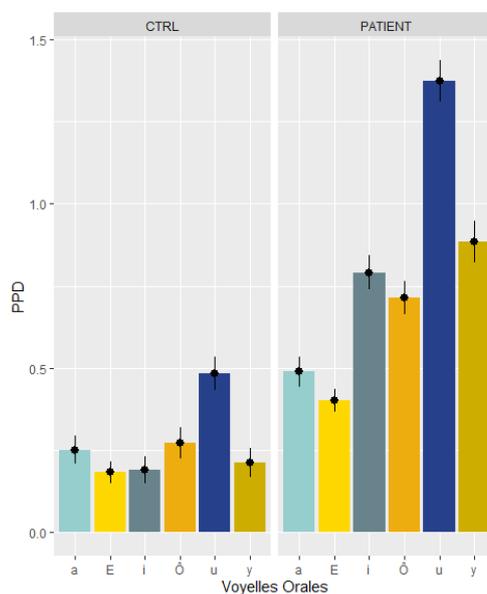


Figure 39 : Scores PPD locaux par voyelle orale, pour les groupes de locuteurs patients et contrôles

Tout groupe de locuteurs confondus, les scores PPD locaux moyens des substitutions par segment vocalique montrent que ce sont les voyelles cardinales qui présentent les scores les plus élevés (/a/ = 1,72 ; /i/ = 1,63 et /u/ = 1,82). Ce sont les segments vocaliques moyens qui présentent les scores PPD les moins élevés (/E/ = 1,07 et /O/ = 1,24).

Pour résumer les résultats de l'ensemble des analyses menées sur les PPD locaux, nous retiendrons que, dans toutes les comparaisons (macro-classes et classes articulatoires), les scores PPD locaux sont différents entre les deux groupes de locuteurs. Les patients obtiennent des scores PPD locaux significativement plus élevés que ceux obtenus par les contrôles ($p < 0.001$).

Les comparaisons en macro-classes phonémiques ont montré que ce sont les clusters qui sont systématiquement les plus altérés, viennent ensuite les voyelles et enfin les consonnes qui sont les mieux identifiées par les auditeurs. En revanche, nous avons relevé que la différence de score entre les consonnes et les voyelles est partiellement imputable aux voyelles nasales lorsque l'on divise la macro-classes voyelles en sous classes voyelles orales vs voyelles nasales. Ainsi, consonnes et voyelles orales présentent des scores plus proches.

Ensuite, nos résultats montrent que les voyelles orales sont significativement plus préservées que les voyelles nasales ($p < 0.001$). Au sein de la sous classe des voyelles orales, en termes d'antériorité, ce sont les voyelles antérieures qui sont les plus préservées ($p < 0.001$), par rapport aux voyelles postérieures ; en termes de hauteur, les voyelles basses sont toujours plus préservées que les voyelle hautes ($p < 0.001$). Ce sont

donc les voyelles hautes et postérieures qui sont les plus altérées, c'est-à-dire le phonème /u/ dans le système vocalique du français. On constate en effet que c'est la voyelle orale la plus altérée pour les deux groupes de locuteurs. Nos résultats montrent que ce sont les mêmes traits vocaliques qui sont les moins bien identifiés dans les deux groupes de locuteurs, et cet effet est majoré par la pathologie. Ces segments les moins bien reconnus sont les plus altérés dans la production.

L'examen des transcriptions des auditeurs nous permettra de préciser cette analyse segmentale.

5.3.2. Transcriptions des auditeurs

Dans cette section nous présenterons les résultats des analyses descriptives des réponses effectives des auditeurs pour les segments vocaliques. Nous comparerons la cible qui devait être produite avec la transcription effective de l'auditeur, au niveau phonémique. Nous illustrerons ces résultats au moyen de graphiques en barre et de tableaux rapportant les valeurs des occurrences de chaque phonème substitué transcrit. Nous avons choisi de présenter les taux de substitutions en perspective avec les réponses correctes afin de garantir l'objectivité des analyses. Elles constituent une base de référence face aux taux de substitutions mesurés.

Pour rappel : 9 segments présentant les fréquences d'occurrences les plus élevés en pourcentage, par rapport à l'ensemble des segments de substitutions identifiés (n = 22) ont été sélectionnés. Ainsi, n = 9 segments moins le segment cible = 8 segments de substitutions retenus par segment ciblé. Les huit segments vocaliques cibles sont /i/, /y/, /u/, /a/, /E/, /Ô/, /@/ et /μ/. Les neuf segments de substitutions les plus fréquents retenus sont /i/, /y/, /u/, /a/, /E/, /Ô/, /@/, /μ/ et /§/. Pour chaque segment vocalique cible nous présenterons les résultats des taux de substitution et d'identifications correctes. Nous porterons notre intérêt sur les différents phonèmes de substitutions les plus fréquemment transcrits et sur les différences entre les deux groupes de locuteurs, contrôle et patient. Nous proposons également un tableau rapportant les taux d'occurrence en pourcentage pour chaque phonème substitué, en fonction du groupe de locuteurs.

Les analyses des scores PPD locaux indiquent que les voyelles nasales obtiennent des scores significativement plus élevés que ceux des voyelles orales. Nous avons donc choisi de présenter les résultats des segments nasalisés séparément de ceux des segments vocaliques oraux. Nous commenterons principalement les segments de substitutions communs aux deux groupes de locuteurs.

5.3.2.1. Segments cibles nasalisés

- /@/ = /ã/ et /μ/ = /ẽ/ et /œ/

Les segments /@/ et /μ/ impliquent un passage de l'air dans la cavité nasale lors de leur réalisation, cette configuration d'abaissement du voile du palais permet la réalisation acoustique de la propriété de nasalité. Le segment /@/ est réalisé avec une position de la langue moyenne dans la cavité buccale, une bouche ouverte et des lèvres arrondies. Le segment /μ/ lui se réalise avec une langue avancée dans la cavité buccale, à hauteur moyenne et avec des lèvres plus ou moins étirées puisqu'il représente un archiphonème

(/ẽ/ et /œ/). Cette notion même permet d'estomper cette dernière distinction. Le segment nasal /@/ obtient plus de réponses correctes (82,29%) que le segment /μ/ (64,97%) dans le groupe contrôle. C'est le segment le moins bien identifié parmi les 8 segments vocaliques ciblés dans le groupe contrôle. Le segment /μ/ présente deux fois plus de réponses substituées (35,03%) que le segment /@/ (17,71%) dans ce même groupe. Dans le groupe patient, les taux d'identifications correctes des segments nasalisés sont bas, 52,92% pour le segment /@/ et 46,66% pour le segment /μ/. Ils font partie des 3 segments les moins bien identifiés, parmi les 8 segments vocaliques ciblés, pour ce groupe de locuteur. Les taux de substitutions relevés sont, en conséquence, élevés avec 47,08% de substitutions pour le segment /@/ et 53,34% pour le segment /μ/.

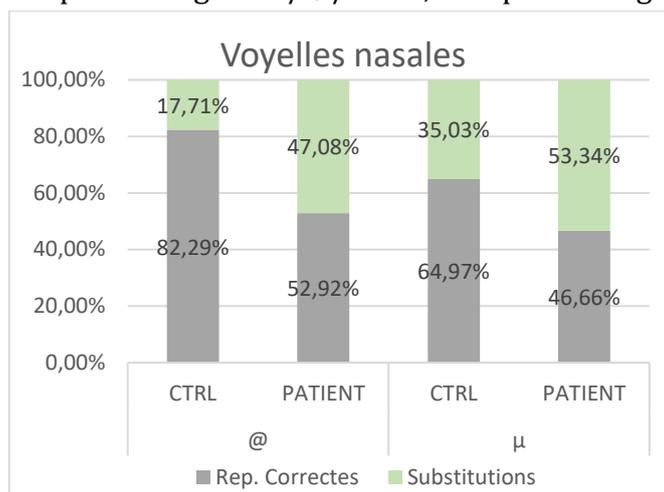
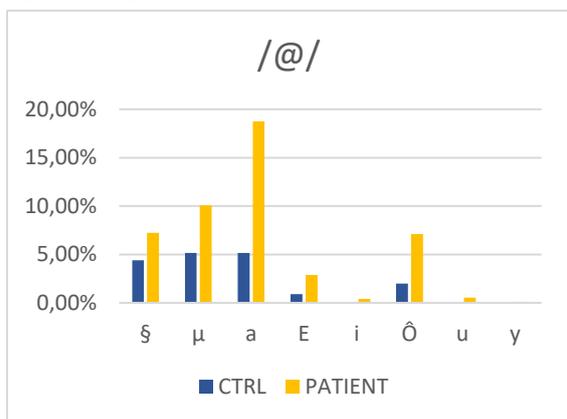


Figure 40 : Taux de réponses correctes et de substitutions pour les segments vocaliques comportant le trait de nasalité /@/ et /μ/, dans les groupes contrôle et patient

Les taux d'occurrences de chaque segment substitué sont reportés dans la Figure et table 1 pour le segment /@/ et dans la Figure et table 2 pour le segment /μ/. La tendance principale commune aux deux groupes de locuteurs et aux deux segments ciblés concerne la 'dénasalisation', les auditeurs proposent majoritairement le segment /a/ comme segment de substitution. Les segments /a/ et /@/ partagent un faisceau de traits articulatoires communs à l'exception de la nasalité, un trait les sépare l'un de l'autre. La proximité des configurations articulatoires impliquées dans ces deux réalisations éclaire cette tendance. En revanche, /a/ et /μ/ sont plus éloignés, en plus de la nasalité, l'aperture de la mandibule constitue le second trait d'écart entre ces deux segments.

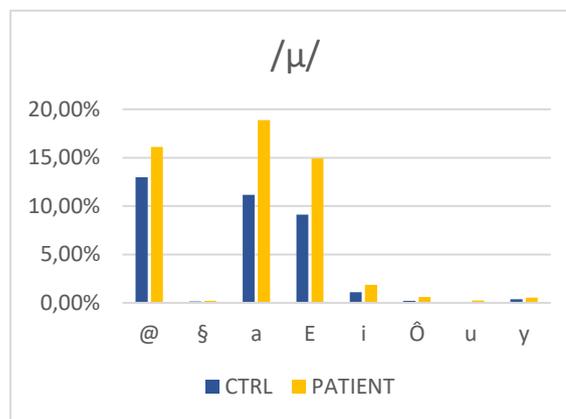
Pour le segment /μ/, la seconde transcription de substitution principalement proposée par les auditeurs est le segment nasalisé /@/. C'est le segment nasalisé le plus proche dans le système vocalique du français. En revanche, pour le segment /@/ les auditeurs ont partagé leurs réponses entre les segments /μ/ et /ʒ/. Ces deux segments sont également proches de /@/ dans le système vocalique et sont séparés par un trait articulatoire d'écart avec la cible, l'aperture de la mandibule. Les auditeurs semblent ainsi privilégier des candidats de substitutions qui soient proches de la cible au sein du système vocalique. Le segment /ʒ/ n'a qu'une occurrence négligeable en substitution du segment /μ/, ils sont séparés par deux traits articulatoires l'antériorité-postériorité de la position de la langue et l'étirement-arrondissement des lèvres. Il apparaît qu'en substitution de /μ/, /ʒ/ n'est pas un candidat pertinent à la perception des auditeurs.

Enfin, le segment /μ/ présente une occurrence élevée de substitution par le segment /E/, soit la version orale du segment nasalisé /μ/. Ainsi, seul le trait de nasalité sépare ces deux segments. Les auditeurs ont donc plutôt tendance à proposer des transcriptions de substitutions 'dénasalisé' pour le segment /μ/. Enfin, une faible occurrence du segment de substitutions /i/ est présente dans les deux groupes de locuteurs. Pour le segment /@/ c'est le segment /Ô/ qui présente une occurrence de transcriptions de substitutions notable et commune aux deux groupes de locuteurs. Le segment /E/ est également proposé en substitution de /@/, dans une moindre mesure, mais dans les deux groupes de locuteurs.



@	§	μ	a	E	i	Ô	u	y
CTRL	4,41%	5,17%	5,17%	0,90%	0,07%	2,00%		
PATIENT	7,23%	10,07%	18,74%	2,88%	0,39%	7,12%	0,54%	0,12%

Figure et table 1 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient



μ	@	§	a	E	i	Ô	u	y
CTRL	12,96%	0,15%	11,14%	9,10%	1,09%	0,22%		0,36%
PATIENT	16,12%	0,20%	18,89%	14,91%	1,85%	0,60%	0,24%	0,52%

Figure et table 2 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient

Les auditeurs proposent préférentiellement la transcription de substitution /a/ pour les deux segments nasalisés /@/ et /μ/, et dans les deux groupes de locuteurs. Ensuite, la tendance montre que les auditeurs choisissent des candidats de substitutions comprenant le trait de nasalité, caractéristique des segments ici ciblés. On relève une attirance mutuelle entre les segments /@/ et /μ/, majorée pour le segment /μ/ en faveur du segment /@/, alors que pour le segment /@/ on observe une attraction moindre envers /μ/, au profit du segment /§/. Enfin, le segment /μ/ est fréquemment substitué par /E/ alors que cet effet est très faible pour le segment /@/, en revanche dans ce dernier, /Ô/ apparaît comme un candidat de substitution alors qu'il est absent pour le segment /μ/. Les tendances principales montrent que les auditeurs s'orientent vers des segments comprenant le trait de nasalité ou vers des segments qui ne le possède pas mais qui restent proches dans le système vocalique de la langue, à 1 trait d'écart.

5.3.2.2. Segments cibles vocaliques

- /i/ et /y/

Le segment vocalique /i/ se réalise à l'avant de la cavité buccale, avec une position de la langue haute dans cet espace et des lèvres étirées. De même, le segment vocalique /y/ se réalise à l'avant de la cavité buccale, avec une position de la langue haute et des lèvres arrondies. Un trait articulatoire sépare ces deux configurations, l'arrondissement-

étirement des lèvres. Ces segments d'avant, hauts et étirés-arrondis présentent les taux d'identification correctes les plus élevés dans le groupe contrôle avec /i/ = 93,24% de réponses correctes et /y/ = 92,14%. Les taux de substitutions dans le groupe contrôle sont les moins élevés parmi tous les segments vocaliques ciblés dans cette étude (/i/ = 6,76% de substitutions et /y/ = 7,89%). Ces segments apparaissent comme les mieux identifiés pas les auditeurs pour le groupe contrôle.

Dans le groupe patient, les taux de réponses correctes sont nettement moins élevés et constituent 67,84% des réponses pour le segment /i/ et 64,55% pour le segment /y/. Les taux de substitutions sont donc bien plus conséquents, 32,16% de segments substitués pour /i/ et 35,45% pour /y/. Différemment du groupe contrôle, dans lequel ils représentent les 1^{er} et 2^{ème} segments les mieux reconnus, les segments /i/ et /y/ sont respectivement les 3^{ème} et 4^{ème} segments les mieux identifiés, dans le groupe patient, parmi les 8 segments vocaliques ciblés.

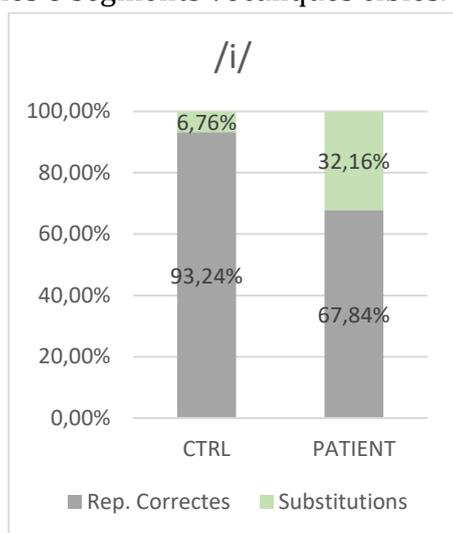


Figure 41 : Taux de réponses correctes et de substitutions pour le segment /i/, dans les groupes contrôle et patient

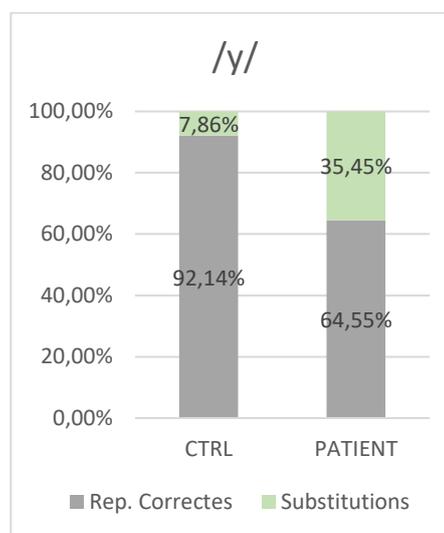
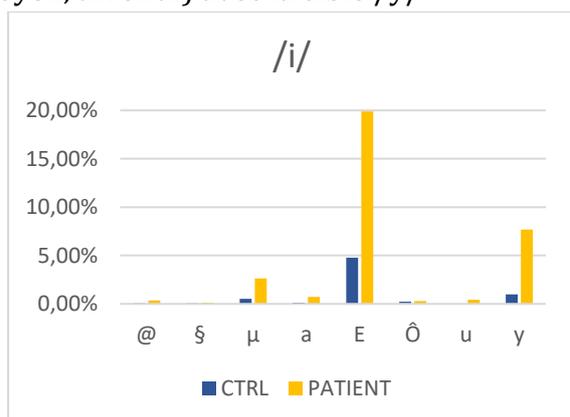


Figure 42 : Taux de réponses correctes et de substitutions pour le segment /y/, dans les groupes contrôle et patient

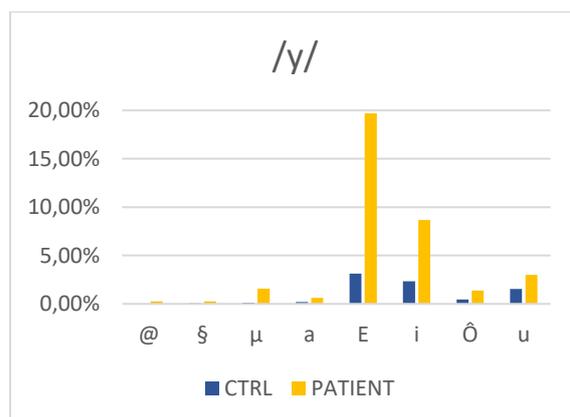
Les principaux segments de substitutions observés dans les transcriptions des auditeurs sont rapportés dans la Figure et table 3 pour le segment /i/ et dans la Figure et table 4 pour le segment /y/. Des points communs entre les réponses de substitutions proposées par les auditeurs sont observables entre les segments /i/ et /y/. En effet, c'est le segment /E/ qui est majoritairement transcrit, dans les deux groupes de locuteurs contrôle et patient. Ce segment se réalise à l'avant de la cavité buccale, avec une bouche mi ouverte/fermée et une position de la langue moyenne dans la cavité buccale. Il représente ici un archiphonème dit moyen sur l'axe d'aperture et a un trait articulatoire d'écart avec les cibles /i/ et /y/. C'est le segment le plus 'moyen', 'centralisé', le laissant apparaître comme un bon candidat de substitution. Les seconds candidats de substitutions sont /y/ pour le segment /i/ et /i/ pour le segment /y/, un trait articulatoire les sépare l'un de l'autre c'est l'étirement-arrondissement ou /i/ est étirée et /u/ est arrondi. Nous constatons qu'ils constituent l'un pour l'autre des candidats de substitutions employés relativement fréquemment, et que leurs similitudes articulatoires pourraient expliquer cette attraction mutuelle.

De plus, pour le segment /i/ c'est le phonème nasalisé /μ/ qui est préférentiellement transcrit, d'avantage pour les patients que pour les contrôles. Cette attirance est aussi relevée pour le segment /y/ mais concerne presque exclusivement le groupe patient. En revanche, une tendance particulière au segment /y/ s'observe puisque, pour les deux groupes de locuteurs, les auditeurs proposent le segment de substitution /u/, qui partage un trait d'écart avec la cible, c'est la position de la langue dans la cavité buccale, antérieur pour la cible /y/ et postérieur pour le segment substitué /u/. Pour le segment /y/ on observe également une très faible occurrence du segment de substitution /Ô/ commune aux deux groupes de locuteurs, il s'avère plus éloigné de la cible que les autres segments de substitutions puisqu'il ne partage que le trait d'arrondissement (/Ô/ postérieur, moyen, arrondi) avec la cible /y/.



i	@	§	μ	a	E	Ô	u	y
CTRL	0,06%	0,06%	0,52%	0,12%	4,79%	0,23%		0,98%
PATIENT	0,37%	0,12%	2,65%	0,72%	19,90%	0,28%	0,44%	7,67%

Figure et table 3 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient



y	@	§	μ	a	E	i	Ô	u
CTRL		0,06%	0,13%	0,19%	3,12%	2,34%	0,45%	1,56%
PATIENT	0,25%	0,25%	1,59%	0,63%	19,68%	8,68%	1,38%	3,00%

Figure et table 4 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient

Les auditeurs proposent donc des segments de substitutions relativement peu éloignés de la cible en termes de traits articulatoires, la tendance principale pour les segments /i/ et /y/ s'oriente vers la 'centralisation' (/E/), avec une bouche perçue comme plus ouverte. La principale tendance secondaire concerne l'attirance mutuelle des segments /i/ et /y/ qui constituent l'un pour l'autre de bons candidats de substitutions de par leur proximité dans le système vocalique. Enfin, la dernière tendance s'articule autour de l'ajout d'un trait (/μ/ nasalité) ou un trait perçu divergeant de la cible au niveau du mouvement de la langue (/u/ postérieur) et de l'aperture (/Ô/moyenne).

- /E/ = archiphonème /ε/, /e/ et /Ô/ = archiphonème /o/, /ɔ/

Dans notre étude, le segment /E/ est un archiphonème regroupant les segments cibles /e/ et /ε/. Il apparaît deux fois moins que les autres phonèmes vocaliques ciblés puisqu'il n'a été placé qu'en position V1 dans la syllabe initiale du pseudo-mot. Nous le considérons comme un segment qui se réalise à l'avant de la cavité buccale, antérieur, et dans lequel la position de la langue est moyenne. La notion d'archiphonème affecte ici l'aperture de la mandibule et efface la distinction mi-ouvert mi-fermé, au profit d'une conception plus large sous le terme moyenne. Le segment cible /Ô/ représente également un archiphonème regroupant les phonèmes vocaliques /o/ et /ɔ/. C'est donc un segment

postérieur, et moyen dans lequel la distinction mi-ouvert mi-fermé est également gommée. Dans le groupe contrôles les taux de substitutions pour les segments /Ô/ et /E/ sont vraiment proches, (15,83% pour /E/ et 16,22% pour /Ô/), il est en donc de même pour les taux de réponses correctes (84,17% pour /E/ et 83,78% pour /Ô/). Dans le groupe contrôle, les taux d'indentification correcte et de substitutions sont équivalents entre les deux archiphonème moyen /E/ et /Ô/. En revanche, dans le groupe patient, le segment /Ô/ est moins bien identifié (60,87%) que le segment /E/ (69,33%). Ce dernier présente donc un taux de substitution inférieur (30,67%) à celui du segment /Ô/ (39,13%) dans le groupe patient. C'est de deuxième segment le mieux identifié dans ce groupe de locuteur.

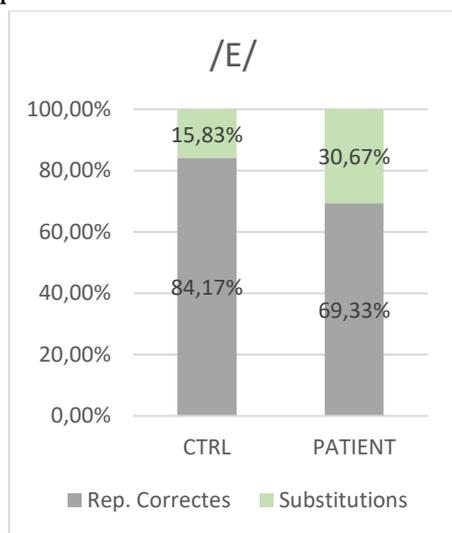


Figure 43 : Taux de réponses correctes et de substitutions pour le segment /E/, dans les groupes contrôle et patient

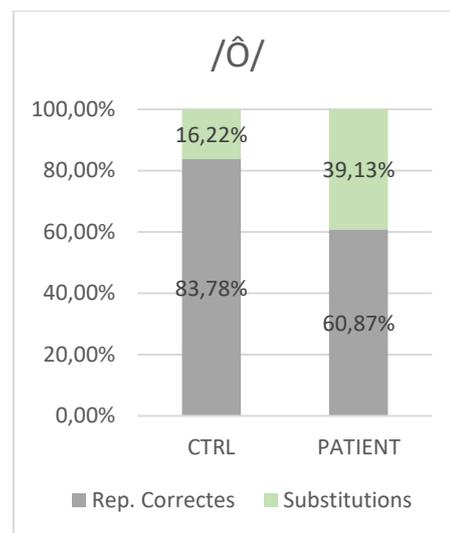


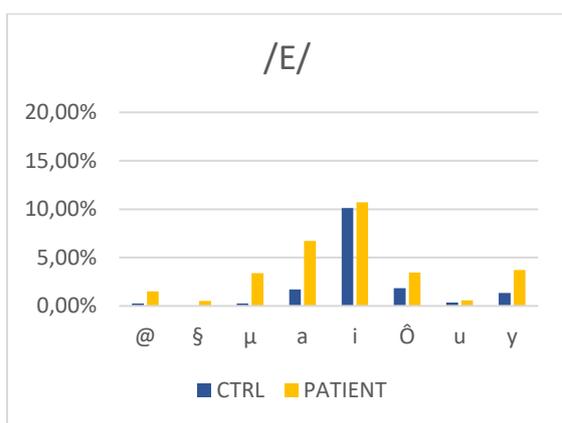
Figure 44 : Taux de réponses correctes et de substitutions pour le segment /Ô/, dans les groupes contrôle et patient

Les principaux segments de substitutions observés dans les transcriptions des auditeurs sont rapportés dans la Figure et table 5 pour le segment /E/ et dans la Figure et table 6 pour le segment /Ô/. Les segments moyens antérieur /E/ et postérieur /Ô/ ne montrent pas de véritable tendance commune, si ce n'est qu'ils apparaissent comme les segments pour lesquels il y a une plus grande variété de réponses substitués.

Pour le segment /E/, c'est le segment /i/ qui est le plus proposé en tant que transcription de substitution, pour les deux groupes de locuteurs. C'est un segment antérieur distant d'un trait articulatoire avec l'archiphonème /E/, sur l'axe d'aperture ou /i/ est fermé et /E/ moyennement ouvert-fermé. A l'opposé sur l'axe d'aperture, le segment /a/ ouvert représente le segment de substitution le plus proposé par les auditeurs dans le groupe patient. Dans le groupe contrôle le taux d'occurrence du segment /a/ est comparable à ceux des segments /Ô/ postérieur et /y/ fermé éloignés de la cible par un trait articulatoire. Ces derniers sont proportionnellement plus importants dans le groupe patient et leurs taux d'occurrences sont comparables à celui du segment /μ/ nasalisé. Ce dernier n'a que très peu été transcrit en tant que substitution dans le groupe contrôle.

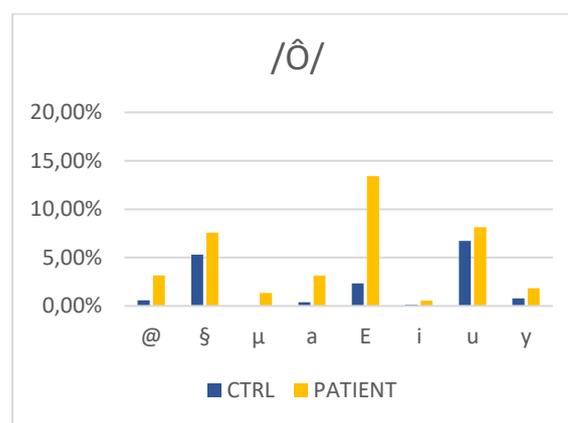
Le segment /Ô/ est principalement substitué par des segments postérieurs, en effet, les segments /u/ et /§/ apparaissent comme les segments, communs aux deux groupes

de locuteurs, les plus transcrits. Ils sont distants d'un trait articulatoire avec l'archiphonème /Ô/, ou /u/ est fermé et /Ô/ moyen et /§/ comporte le trait de nasalité, contrairement à la cible. De plus, le segment /E/ représente le segment le plus transcrit en substitution à /Ô/ mais seulement dans le groupe patient. Proportionnellement aux segments /u/ et /§/ le segment /E/ est moins transcrit dans le groupe contrôle, mais sa fréquence d'occurrence est quand même supérieure à celle de /y/. C'est également un segment moyen mais il est antérieur contrairement à sa cible. Enfin, les segments /@/ et /a/ ont des fréquences d'occurrences comparables entre elles dans le groupe patient comme les segments /μ/ et /y/. En revanche ils n'apparaissent que dans des proportions négligeables dans le groupe contrôle.



E	@	§	μ	a	i	Ô	u	y
CTRL	0,24%		0,24%	1,71%	10,11%	1,83%	0,37%	1,34%
PATIENT	1,50%	0,52%	3,40%	6,74%	10,73%	3,47%	0,59%	3,73%

Figure et table 5 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient



Ô	@	§	μ	a	E	i	u	y
CTRL	0,58%	5,30%		0,39%	2,33%	0,13%	6,72%	0,78%
PATIENT	3,16%	7,58%	1,34%	3,12%	13,41%	0,55%	8,13%	1,85%

Figure et table 6 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient

Une seule tendance qui émerge des transcriptions substituées, à la fois communes aux deux groupes de locuteurs et aux deux segments moyens /E/ antérieur et /Ô/ postérieur : les réponses des auditeurs tendent à rester sur l'axe antéro-postérieur et c'est la position de la langue qui est principalement maintenue, /Ô/ est majoritairement substitué par /u/ et /§/ et /E/ par /i/. A nouveau les transcriptions des auditeurs s'éloignent le moins possible de la cible, la diversité des segments peut être éclairée par la position moyenne de ces segments dans le système vocalique, les plaçant dans des positions plus 'centrales' avec plus de voisins que les segments se situant à la périphérie du système. En effet, en segment antérieur fermé, par exemple, ne peut pas être perçu comme plus fermé ou plus antérieur puisque aucun segment vocalique ne correspond à cette configuration articulatoire dans le système vocalique. Ces différents éléments suggèrent que les segments /E/ et /Ô/ feraient de bons candidats de substitutions aux segments qui les entourent dans le système vocalique.

- Le segment /a/

Le segment /a/ est ici le seul représentant des phonèmes vocaliques bas (archiphonème /a/ et /ɑ/), il se réalise avec une position de langue abaissée dans la cavité buccale, une bouche (et mandibule) ouverte et des lèvres arrondies. Ce segment présente les meilleurs taux d'identifications correctes dans les deux groupes de locuteurs, avec 88,74% de bonnes réponses dans le groupe contrôle et 83,53% dans le groupe patient. C'est le segment qui présente le taux de substitution le plus faible dans le groupe patient avec 16,47%. Dans le groupe contrôle, le taux de substitution relevé est supérieur à ceux des segments /i/ et /y/ avec 11,26% pour le segment /a/.

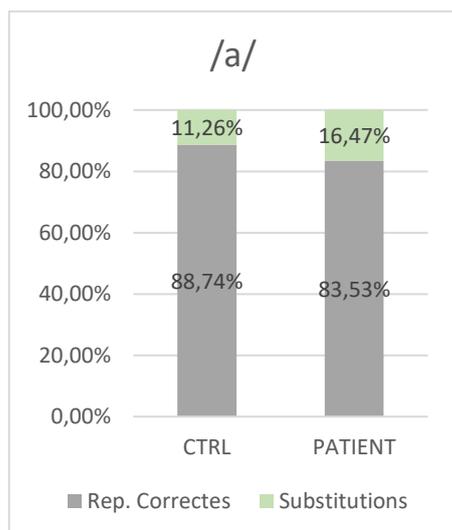


Figure 45 : Taux de réponses correctes et de substitutions pour le segment /a/, dans les groupes contrôle et patient

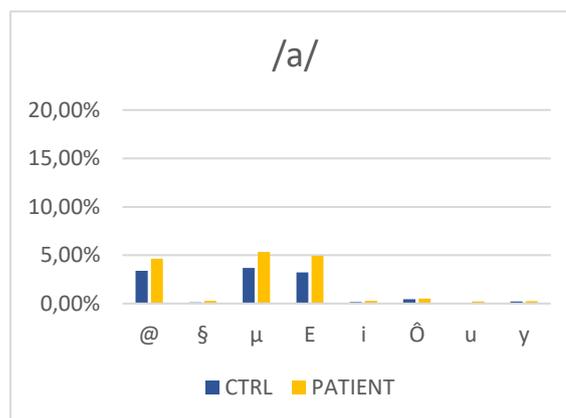


Figure et table 7 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient

La Figure et table 7 montre bien que ce sont les mêmes segments qui sont principalement substitués au segment cible /a/, les auditeurs semblent attirés par les mêmes candidats de substitutions, entre les deux groupes de locuteurs. C'est le segment /μ/ qui est le plus employé en phonème de substitution et 2 traits articulatoires séparent la cible /a/ de la substitution /μ/, l'aperture de la mandibule (de ouverte à mi-ouverte) et la nasalité. L'attraction vers une cible nasalisée est également observée au travers de la transcription /@/, seul le trait articulatoire de nasalité sépare la cible /a/ de la transcription de substitution /@/. Le segment /E/ est également proposé par les auditeurs comme candidat de substitution au segment /a/, il est éloigné de 1 (mi-ouvert) à 2 (mi-fermé) traits d'écart avec la cible, sur l'axe d'aperture.

Les tendances principales des choix des candidats de substitutions s'articulent autour de la nasalité et de l'aperture de la mandibule. Le trait de nasalité est perçu alors qu'il est absent de la cible et la mandibule est perçue comme plus fermée (mi-ouvert ou mi-fermé) que dans la cible /a/. Là encore, les phonèmes de substitutions préférentiellement transcrits par les auditeurs sont peu éloignés de la cible en termes de traits articulatoires (1 à 2 traits d'écart). Ils suivent les mêmes tendances entre les deux groupes de locuteurs et s'orientent vers l'ajout d'un trait articulatoire, la nasalité, et/ou vers un trait mal perçu, l'aperture.

- Le cas particulier du segment /u/

Le segment vocalique /u/ se réalise à l'arrière de la cavité buccale, avec une position de la langue haute dans cet espace et des lèvres arrondies. L'analyse des scores PPD locaux indique que c'est le segment qui présente les scores les plus élevés dans les deux groupes de locuteurs. L'analyse descriptive des transcriptions des auditeurs montre que c'est le segment qui présente les taux de substitutions les plus importants, pour les deux groupes de locuteurs. C'est donc le segment le moins bien identifié et transcrit par les auditeurs.

Dans le groupe contrôle seulement 76,05% des segments /u/ ciblés ont été transcrits correctement, contre 45,26% dans le groupe patient. Les taux d'identification correcte de ce segment sont les plus faibles, indiquant que produire ou percevoir ce segment est moins évident, par rapport aux autres segments vocaliques ciblés dans cette étude. Par conséquent, les taux de substitutions du segment /u/ sont les plus élevés pour les deux groupes de locuteurs et représentent 23,95% des réponses dans le groupe contrôle et plus de la moitié des réponses totales dans le groupe patient (54,74%) (Figure 46).

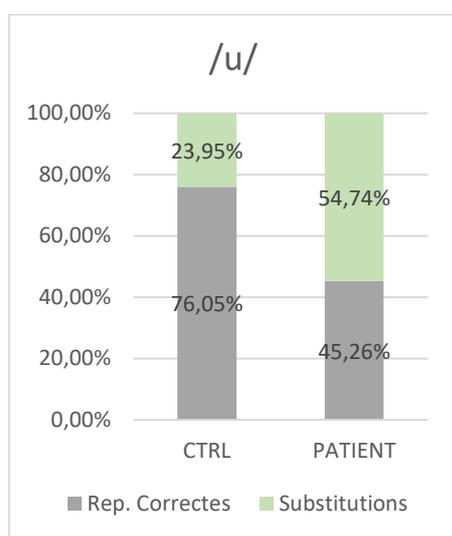
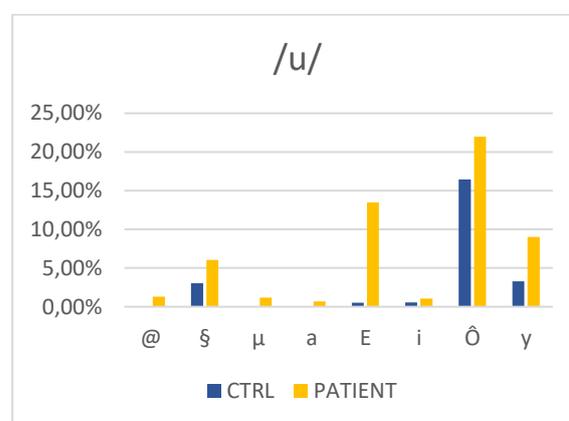


Figure 46 : Taux de réponses correctes et de substitutions pour le segment /u/, dans les groupes contrôle et patient



u	@	§	μ	a	E	i	Ô	y
CTRL		3,02%	0,06%	0,06%	0,50%	0,56%	16,45%	3,30%
PATIENT	1,30%	6,06%	1,17%	0,68%	13,48%	1,05%	21,98%	9,03%

Figure et table 8 : Taux d'occurrence de chaque phonème de substitution pour chaque groupe de locuteur contrôle et patient

L'histogramme (Figure et table 8) comporte une échelle à 25% car c'est le segment qui présente le taux de substitutions, et par conséquent d'erreurs, le plus élevé pour les deux groupes de locuteurs. Le segment /u/ est préférentiellement transcrit par le segment de substitution /Ô/, postérieur, arrondi, moyen. Le trait articulaire qui différencie /u/ de /Ô/ concerne l'aperture de la mandibule plus ouverte dans /Ô/. Le segment /§/ est également choisi par les auditeurs comme candidat de substitution et 2 traits articulaires le séparent de la cible, l'aperture (comme pour /Ô/) et la nasalité. Le second segment de substitution le plus transcrit est /y/ qui se réalise au moyen d'une position de la langue antérieure, ou avancée, et haute dans la cavité buccale et des lèvres arrondies. Ce segment ne présente qu'un trait articulaire d'écart avec la cible /u/, c'est l'antériorité (/u/ postérieur et /y/ antérieur). Cette tendance à sélectionner des phonèmes proches du phonème cible en termes articulaires dans le système vocalique est commune aux deux groupes de locuteurs. Toutefois, on observe une tendance supplémentaire propre au

groupe patient, pour lequel le segment de substitutions /E/ apparaît de façon conséquente. Le segment antérieur et moyen /E/ présente 2 à 3 traits d'écart avec la cible, l'antériorité, l'aperture et l'arrondissement. Cette transcription de substitution dominante dans le groupe patient est quasiment absente dans le groupe contrôle.

Les résultats de cette analyse montrent que pour les deux groupes de locuteurs, ce sont les mêmes segments qui sont dégradés, et que la pathologie accentue les dégradations. Ils montrent également que les auditeurs, lorsqu'un phonème n'est pas correctement identifié, ont tendance à privilégier les mêmes candidats en tant que phonèmes de substitution. De plus, ces résultats montrent que les auditeurs ont tendance à proposer une transcription de substitution peu éloignée de la cible initiale en termes de traits articulatoires. Cela suggère que certaines propriétés acoustiques sont préservées et que l'auditeur s'appuie sur celles-ci.

Enfin, l'ensemble de ces analyses ont été permises par l'approche perceptive de l'intelligibilité, par décodage acoustico-phonétique de pseudo-mots. Les scores obtenus par cette évaluation ont permis notamment de montrer que les voyelles constituent des segments d'intérêt dans l'examen de l'intelligibilité. Le résultat majeur qui émerge de l'examen des scores locaux est l'importante altération de la voyelle orale postérieure fermée arrondie /u/ et des voyelles nasales. L'examen des transcriptions a montré que les segments vocaliques dégradés sont majoritairement identifiés comme des voyelles moyennes, ce qui pourrait témoigner d'une centralisation des segments, en particulier chez les patients. L'examen des données par mesures acoustiques pourra rendre compte de cette tendance.

6. Approche par mesures acoustiques

Nous l'avons souligné dans notre première section, l'approche par mesures acoustiques devrait permettre de compléter l'évaluation segmentale en apportant un autre regard sur les productions vocaliques des locuteurs.

Les études associant l'aire du triangle vocalique (VSA) avec des scores d'intelligibilité ont montré que celle-ci est réduite dans les dysarthries (Weismer et al., 1988, 2001) et dans la dysphonie (Ishikawa et al., 2019; Ishikawa & Webster, 2020). Suggérant qu'une compression de cet espace corrèle avec un score d'intelligibilité peu élevé, associé à une altération importante de la production. Les profils de locuteurs soulevés par Meunier & Ghio (2018) associant les paramètres de dispersion du système avec celui des catégories vocaliques suggèrent alors que le degré d'intelligibilité pourrait être associé avec la dispersion/concentration du système et des catégories de voyelles. Ainsi, un espace vocalique maximisé, associé à un indice de distinctivité élevé devrait être associé avec des scores perceptifs locaux bas montrant peu d'altération et au contraire, un espace vocalique compressé associé à un indice de distinctivité bas montrant un chevauchement entre les catégories devrait être associé avec des scores perceptifs locaux élevés montrant une altération importante des segments.

Dans cette partie, nous préciserons le corpus sur lequel portent ces mesures ainsi que la méthodologie employée pour relever et extraire les valeurs de F1 et F2. Nous examinerons ces valeurs et proposerons une analyse descriptive des valeurs qui apparaissent comme très éloignées des standards. Nous calculerons ensuite l'aire de l'espace vocalique et la valeur de l'indice de distinctivité. La réduction de l'espace vocalique devrait être associée à un indice de distinction entre les catégories de voyelles moins important alors qu'un indice de distinction élevé devrait montrer un espace vocalique plus ample. Nous examinerons la relation de ces métriques avec les scores PPD globaux et locaux. Enfin, nous proposerons une représentation visuelle des espaces vocaliques des groupes de locuteurs dans le plan F1/F2.

6.1. Corpus

Dans cette étude, nous utilisons le corpus « DAP » du projet de recherche C2SI présenté dans la section 5.1. Les enregistrements de parole font ici l'objet d'analyses acoustiques, lesquelles seront mises en relation avec les scores d'intelligibilité issus des différents tests de jugement perceptif de l'intelligibilité présentés en amont.

6.2. Méthodologie

Les analyses acoustiques présentées dans cette section sont basées sur les analyses des formants des segments vocaliques du corpus. Les valeurs de formants revêtent un caractère particulièrement intéressant à la fois puisqu'elles permettent le calcul d'indices représentatifs du degré de distinctivité ou de recouvrement d'un système vocalique (VSA et ID) mais également le calcul de métriques offrant une valeur de représentation graphique. En effet, ces métriques permettent de placer les voyelles dans un système bidimensionnel représentatif des aspects articulatoires impliqués dans leur production. Elles offrent ainsi une interface entre le signal acoustique et les mécanismes de production. Les représentations des systèmes vocaliques sont ici exploitées pour contrôler la qualité des mesures acoustiques dans un premier temps, puis dans le cadre d'une analyse descriptive des systèmes en fonction des groupes de locuteurs. Les indices issus des métriques font l'objet d'analyses statistiques pour évaluer la relation entre les scores perceptifs et les productions des locuteurs.

6.2.1. F1/F2

Dans cette étude en production de la parole, nous travaillons à partir des enregistrements des locuteurs. A partir de ces enregistrements, une segmentation du signal acoustique au niveau segmental, ou phonémique, a été réalisé par nos collègues du Laboratoire Informatique d'Avignon. En particulier, C. Fredouille a utilisé l'aligneur automatique du LIA (Bechet, 2001) pour segmenter et annoter le signal acoustique de chaque pseudo-mot au niveau segmental de façon automatique. (Annexes 1 : Un exemple de signal acoustique étiqueté et aligné avec sa transcription LIAPhon, p205)

A partir du signal acoustique et des fichiers de transcriptions phonétiques étiquetées et alignées nous avons extrait les valeurs de F1 et F2 pour chaque segment vocalique. Les valeurs reportées de F1 et F2 correspondent à la valeur moyenne du premier et du second formant sur le second tiers (ou tiers-milieu) de la durée totale du segment vocalique ciblé. Les segments d'une durée inférieure à 30ms et supérieure à 400ms ont été exclus. Nous avons réalisé cette extraction au moyen d'un script Praat (Mietta Lennes (<http://www.helsinki.fi/~lennes/praat-scripts/>) version modifiée Esther Le Grézause (elg1@uw.edu) in May 2016, dernières modifications par Anna Marczyk (ana-katarzyna.MARCZYK-BUKLAHA@univ-amu.fr) Juin 2020).

A partir des valeurs automatiquement extraites de F1 et F2, nous avons visualisé, au moyen de nuages de points, la répartition et l'agencement de ces valeurs dans l'espace vocalique, par groupe de locuteurs (Patients et Contrôles). Cette représentation a permis d'avoir une idée de la dimension des espaces vocaliques ainsi représentés. La Figure 56 : Illustration de la visualisation de j donnera un aperçu des représentations proposées en page 163. Chaque point représente la conjoncture des valeurs de F1 et de F2 dans un espace bidimensionnel, et chaque couleur représente une catégorie de segment vocalique, ou voyelle.

Concernant le contrôle de la qualité des mesures, les plots sur les données brutes ont permis de mettre en évidence des valeurs de formants peu conformes à la cible attendue. En effet, cette représentation en nuage de points permet de repérer visuellement les valeurs de formants distantes du nuage global (représentées par des ellipses dans la figure 11), et donc de repérer les valeurs qui sont en dehors des standards communément admis. Ces valeurs peuvent constituer des *outliers* dont l'origine, le nombre et l'étendue doivent être précisés. Classiquement, les études sur les mesures de formants proposent de filtrer les valeurs selon un cadre de référence, par exemple, Gendrot & Adda-Decker (2005) définissent des seuils de valeurs en deçà et au-delà desquels les valeurs ne sont plus considérées comme réalistes, car elles sont en désaccord avec les caractéristiques acoustiques attestées des segments en question.

Nous avons appliqué dans un premier temps ce filtre pour trier nos données et écarter les valeurs considérées comme aberrantes. Les valeurs de référence ont été implémentées dans un script R (A. Marczyk, Juin 2020). Toutefois, ce filtre nous a conduit à écarter un trop grand volume de données (~20%). Nos questionnements ont concerné la pertinence de l'usage de ce filtre sur des données de parole pathologique. En effet, les valeurs de ce filtre se basent sur de la parole produite par des sujets sains dans la diffusion de nouvelles à la radio (Gendrot & Adda-Decker, 2005b). Appliquer ce cadre de références sans nuance risquerait de masquer et d'écarter des données qui correspondent à des variations liées à la pathologie, qui au contraire, nous intéressent tout particulièrement.

Voyelles	/i/ min-max	/y/ min-max	/u/ min-max	/a/ min-max	/o ɔ/ min-max	/e ε/ min-max
F1 _{male}	750	800	900	1000	900	800/1000
F2 _{male}	1500 - 2500	1300 - 2200	400 - 1550	800 - 2300	600 - 1600/1800	1100 - 2400
F1 _{female}	900	900	1000	1100	1000	900/1100
F2 _{female}	1600 - 3100	1400 - 2800	400 - 1500	900 - 2300	600 - 1600/2000	1400 - 3000

Tableau 15 : Valeurs de formants minimales et maximales de F1 et F2 pour les hommes et les femmes, d'après Gendrot & Adda-Decker (2005)

Nous avons donc opté pour un filtrage manuel des valeurs de F1 et de F2. Les valeurs du filtre Gendrot, trop restrictives pour une étude sur la parole pathologique, ont néanmoins servi de cadre de référence mais nous l'avons élargi pour considérer les effets liés à la pathologie. Nous avons ajouté +/- 100 à chaque intervalle. Par exemple, pour /i/ nous avons accepté F1_{male} jusqu'à 850 et F2_{male} entre 1400 et 2600, il en est de même pour tous les autres segments et pour les femmes.

Avant de poursuivre sur les résultats de ces filtrages manuels, il est essentiel de clarifier la source des données mobilisées dans cette étude. Il s'agit des valeurs acoustiques présentement extraites, des données des tests de perception de la parole objets de l'étude précédente (5. et 5.3), soit les Scores PPDLocaux et les transcriptions des auditeurs. Ainsi, la mise en relation des valeurs de formants (9565 valeurs de F1 et de F2 extraites) et des réponses transcrites des auditeurs associés aux PPD locaux, phonèmes à phonèmes offre une interface entre la production et la perception de la parole. Pour

rappel, chaque segment a été transcrit par trois auditeurs différents (soit 27618 scores PPDLocal et autant de transcriptions d'auditeurs). Ainsi, chaque valeur de F1 et F2 par segment et par locuteur a été associée avec les trois réponses transcrites par les auditeurs, ainsi qu'avec le score PPDLocalVoy qui représente la distance entre la cible attendue et la cible perçue en termes de traits articulatoires. Ces données nous ont permis de catégoriser les 'bonnes' réponses, quand la cible correspond à la transcription (PPDLocal = 0) et les 'mauvaises' réponses, quand la cible et la réponse transcrite sont différentes (PPDLocal ≠ 0). Au total, nous avons retenu 27538 valeurs¹⁹.

6.2.1.1. Filtrage des valeurs de F1/F2

Notre filtrage a permis de déterminer que 93,81% des valeurs de F1 et F2 correspondent aux valeurs de référence (Gendrot +/- 100Hz) que nous avons déterminées. En effet, lorsque les valeurs de formants relevées entrent dans le cadre des valeurs possibles de la cible attendue ([cible]), voir Tableau ci-dessus, les valeurs de formants sont d'emblées acceptées. Dans ce cas, les valeurs de formants correspondent à la cible (F1 et F2 = [cible]).

Cas des valeurs de formant dans le cadre établi		
Segment ciblé	Valeurs de formant	Transcription auditeur
x	x	x
x	x	y

> Bonne réponse

> Distance Cible - Transcription > PPDLoc

Tableau 16 : Conditions d'acceptation des valeurs de F1 et F2

Ce filtrage a également permis de repérer 315 valeurs hors cadre de F1 et 1390 valeurs hors cadre de F2. Soit respectivement 1,14% (n=315) et 5,04% (n=1390) du total des valeurs (n = 27538). Nous avons investigué isolément ces 6,19% de valeurs de F1 et F2 écartées par notre filtrage manuel, afin de comprendre à quel phénomène associer ces valeurs hors du cadre prédéterminé. Nous avons abordé le contrôle de ces valeurs au regard de l'ensemble des données à notre disposition dans cette étude soit : le signal acoustique, les locuteurs (groupe patient et contrôle), les segments, les transcriptions des auditeurs (phonétisées) et les scores PPD Global et LocalVoy. Dans ce cas, les valeurs de formant ne correspondent pas à la cible (F1 et F2 ≠ [cible]).

Nous procéderons en une analyse descriptive de l'ensemble de ces valeurs hors du cadre de référence, au regard de leur correspondance avec certains segments ou certains locuteurs.

Nos premières observations sur les segments ont montré que les segments hauts (/i/, /y/ et /u/) et bas (/a/) concentrent davantage de valeurs hors du cadre de référence pour F1. En revanche, les segments moyens /e/ et /o/ présentent peu de valeurs hors du cadre de référence. La tendance est un peu différente pour les valeurs hors cadre de F2, puisque

¹⁹ 120 transcriptions n'ont pas trouvé de correspondance acoustique car les segments d'une durée inférieure à 30ms et supérieure à 400ms ont été exclus de l'extraction des valeurs de formants puisqu'elles ne permettent pas de calculer une valeur fiable.

ce sont quasi-exclusivement les segments hauts (/i/, /y/ et /u/) qui sont concernés. La concentration du nombre de valeurs en dehors du cadre de référence pour tous les segments et pour F1 et F2 sont reportés dans le tableau suivant, en taux et en nombre.

Voyelle	F1		F2	
	Taux	n	Taux	n
a	14,60%	46	2,16%	30
e	1,90%	6	4,39%	61
i	25,71%	81	14,10%	196
o	2,86%	9	7,63%	106
u	14,29%	45	61,44%	854
y	40,63%	128	10,29%	143

Tableau 17 : Taux et nombre de valeur de F1 et F2 hors du cadre de référence pour chaque voyelle

Nous avons relevé plus d'erreurs (en termes de valeurs en dehors du cadre de référence sus-expliqué) pour les F2 que pour les F1. Les erreurs de F2 concernent également plus de locuteurs (n = 110), et apparaissent plus sporadiques (entre 3 et 50 par locuteur) que les erreurs sur F1 (n = 52 locuteurs). De plus, nous avons repéré un locuteur qui concentre à lui seul 20% des erreurs de F1 (n = 63), et 3,6% des erreurs sur F2 (n = 50). Ce locuteur masculin de 67 ans au moment de l'enregistrement a été atteint d'un cancer du plancher de la cavité buccale douze ans plus tôt. Il a bénéficié d'une chirurgie des ganglions et d'une radiothérapie pour une tumeur de stade 2. Les caractéristiques du signal acoustique de ce locuteur ne permettent pas d'extraire des valeurs de formants fiables. En effet, les trois extraits suivants illustrent le signal acoustique, notamment au moyen d'un spectromètre à bandes larges, qui permettent de repérer l'énergie présente sur le signal acoustique.

Le premier extrait (« *ganlou* ») montre que le signal n'a pu être correctement segmenté et aligné. Pourtant la production du patient est peu altérée, ce qui a priori aurait dû se traduire par une extraction correcte des valeurs acoustiques de F1 et F2, bien présents et visibles même s'ils montrent peu de variation. L'erreur d'alignement trouve peut-être son origine dans le dévoisement systématique des secondes syllabes dans les productions de ce patient, qui pourrait « décaler » l'alignement automatique.

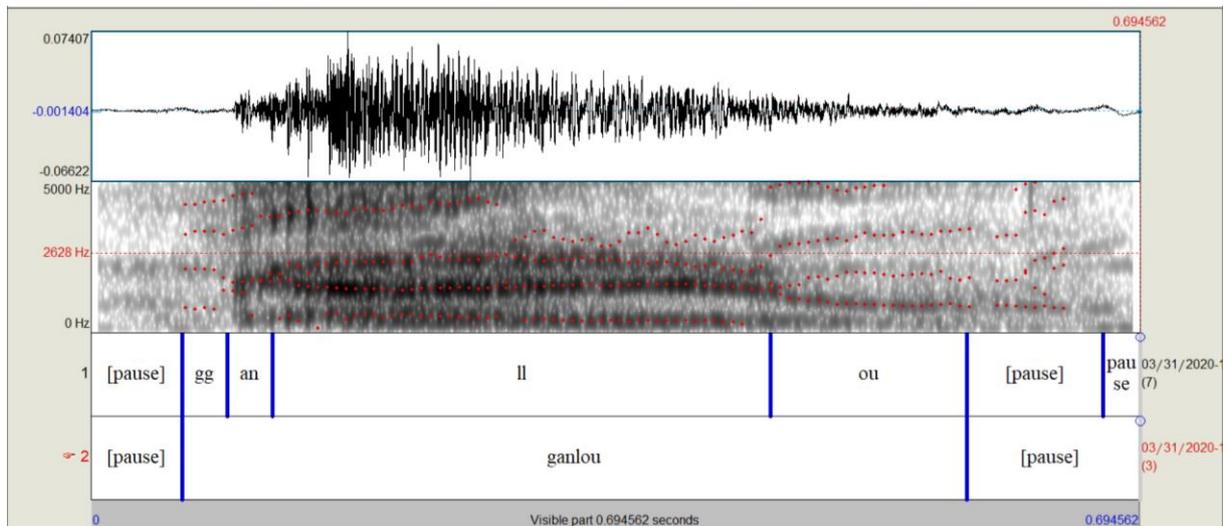


Figure 47: Illustration du signal acoustique d'un locuteur patient prononçant le pseudo-mot « ganlou ». En premier l'oscillogramme du signal produit, suivit par le spectrogramme à bande large sur lequel apparaissent les formants (en rouge). L'annotation segmentale en phonème (convention LIAPhon) se retrouve dans la première tire d'annotation, la seconde représente le pseudo-mot qui devait être produit.

Un second exemple illustre le signal acoustique atypique de ce locuteur, lors de sa production du pseudo-mot « prega », pour lequel la segmentation et l'alignement n'ont pu être réalisés convenablement. La fréquence de F1 est fort perturbée et contient peu d'énergie. De plus, sur la première partie du signal on observe une confusion entre le second et le troisième formant, qui se répercute sur la seconde partie du signal où l'hypothétique F1 est confondu avec F2. Ces caractéristiques ne permettent pas de relever des valeurs fiables.

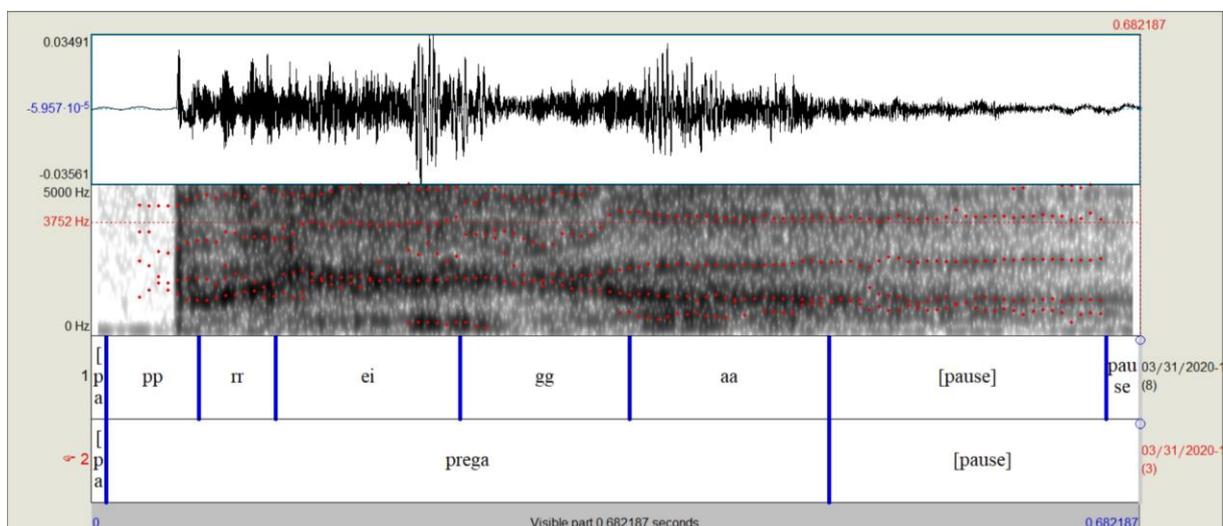


Figure 48 : Illustration du signal acoustique d'un patient prononçant le pseudo-mot « prega »

Enfin, un dernier exemple du signal acoustique de ce locuteur montre une erreur d'alignement sur la première syllabe du pseudo-mot « chovu », ce qui reporte une valeur de F1 et F2 hors du cadre de référence, malgré des formants visibles et caractéristiques de 'o' sur la portion de signal. En revanche, sur la seconde partie ou syllabe du pseudo-mot le premier formant est absent et les second et troisième formants sont confondus.

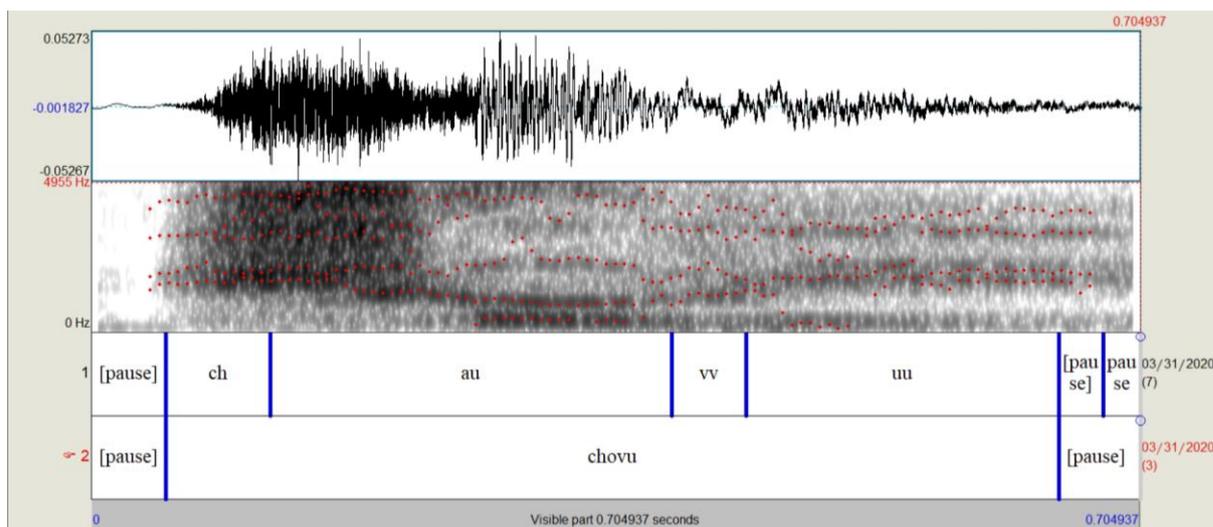


Figure 49 : Illustration du signal acoustique d'un patient produisant le pseudo-mot « chovu ».

Un second locuteur concentre 8,9% des valeurs hors cadre de F1 ($n = 28$) mais reste proche de la moyenne dans le nombre de valeurs hors cadre de F2 ($n = 14$). Aucun autre locuteur ne se démarque des autres, exception faite de ces deux locuteurs. En moyenne, les valeurs hors cadre de F1 représentent 4,5 'erreurs' par locuteurs concerné et 12,3 valeurs aberrantes de F2. Nous retenons donc qu'un locuteur se détache particulièrement des autres, mais que l'examen de son signal acoustique explique le taux élevé de valeurs hors du cadre de référence, puisque ces caractéristiques ne permettent pas de réaliser un alignement sans erreur et d'extraire des valeurs de formants fiables.

De plus, nous avons relevé plus de valeurs de formants hors cadre dans le groupe patient, par rapport au groupe contrôle. En moyenne, F1 et F2 confondus, 76,5% des valeurs hors cadre sont relevées dans le groupe patient contre 23,5% dans le groupe contrôle. La parole pathologique semble représenter une contrainte pour l'extraction de valeurs de formants. Pour le groupe patient plus particulièrement, nous avons regardé si les métadonnées cliniques pouvaient avoir une incidence sur ces valeurs. La taille de la tumeur, la localisation de celle-ci et le type de traitement dont les patients ont bénéficié impactent-ils la détection des formants ?

Notre examen montre que la taille de la tumeur ne semble pas impacter cette détection puisque tous les stades_T sont concernés par ces valeurs hors cadre. En revanche, au regard de la localisation de celle-ci, les tumeurs localisées sur le plancher de la cavité buccale ou sur l'amygdale sont sur-représentées, par rapport aux autres localisations. Enfin, concernant les traitements dont ont bénéficié les patients, il apparaît que la combinaison de chirurgie de la tumeur et des ganglions associées à de la radiothérapie et de la chimiothérapie ou à de la radiothérapie seulement, a un impact plus important que les autres combinaisons de traitements sur la détection des formants. En effet, ces deux combinaisons concentrent plus de la moitié des valeurs hors cadre pour F1 et F2. Notre analyse descriptive suggère donc que les localisations plancher de la cavité buccale et amygdale, ainsi que certaines associations de traitements (Chirurgie tumeur et ganglions + radiothérapie et chimiothérapie ou + radiothérapie) pourraient avoir un effet délétère sur la capacité des systèmes à aligner puis à extraire des valeurs de formants. Dans la suite

de cette analyse, nous présenterons des extraits de signaux acoustiques qui appuieront ces résultats.

En plus des segments, des locuteurs et des métadonnées des locuteurs patients, nous avons poursuivi cette analyse descriptive des valeurs hors du cadre de référence préétabli en observant les signaux acoustiques d'autres locuteurs. Nous avons repéré plusieurs phénomènes-sources susceptibles de reporter des valeurs non conformes. Parmi ces phénomènes, on note des erreurs de segmentation et d'alignement automatiques, des signaux acoustiques dégradés et des erreurs de production. Dans les exemples de signaux acoustiques que nous présenterons nous préciserons quelques informations au sujet du locuteur. De plus, chaque fois qu'une illustration présentera trois tires d'annotation, la dernière aura été ajoutée par nos soins pour proposer une correction de la segmentation automatique des tires un et deux.

Nous avons repéré des erreurs d'alignement sur le signal acoustique, les frontières des segments vocaliques ont été mal positionnées ce qui a entraîné une mauvaise extraction des valeurs de formant. Rien, ni dans la visualisation du signal acoustique, ni à l'écoute de la production ne permet d'expliquer cette erreur technique. Les deux exemples suivants illustrent ce type d'erreur, nous avons proposé une correction de la segmentation automatique dans la troisième tire d'annotation.

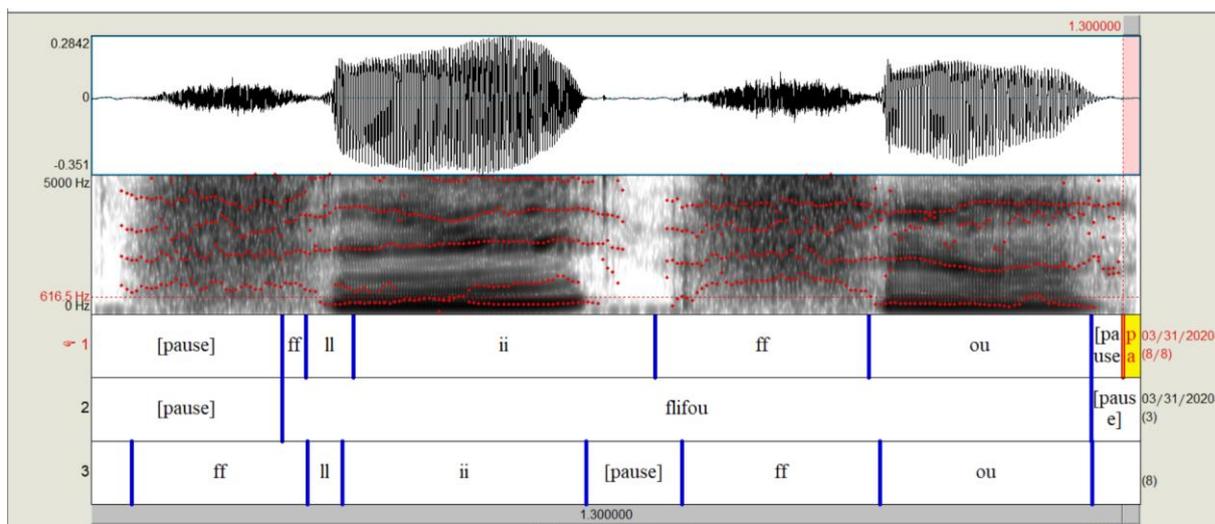


Figure 50 : Signal acoustique d'un patient produisant le pseudo-mot « flifou ». En premier l'oscillogramme du signal produit, suivit par le spectrogramme à bande large sur lequel apparaissent les formants (en rouge). L'annotation segmentale en phonème (convention LIAPhon) se retrouve dans la première tire d'annotation, la seconde représente le pseudo-mot qui devait être produit. La troisième tire a été ajoutée par nos soins pour proposer une correction de l'alignement phonémique.

Ce signal « *flifou* » a été produit par un homme de 51 ans au moment de l'enregistrement, atteint d'un cancer de la mandibule 16 auparavant. Il a bénéficié d'une chirurgie pour une tumeur T2 et d'une chirurgie des ganglions associée à une radiothérapie. Le signal acoustique est de bonne qualité et n'explique pas l'erreur d'alignement dès le début du pseudo-mot à partir de la fricative [f]. En revanche, la non prise en considération de la pause inter syllabique impacte nécessairement l'extraction des mesures formantiques qui ne peut se faire au tiers milieu.

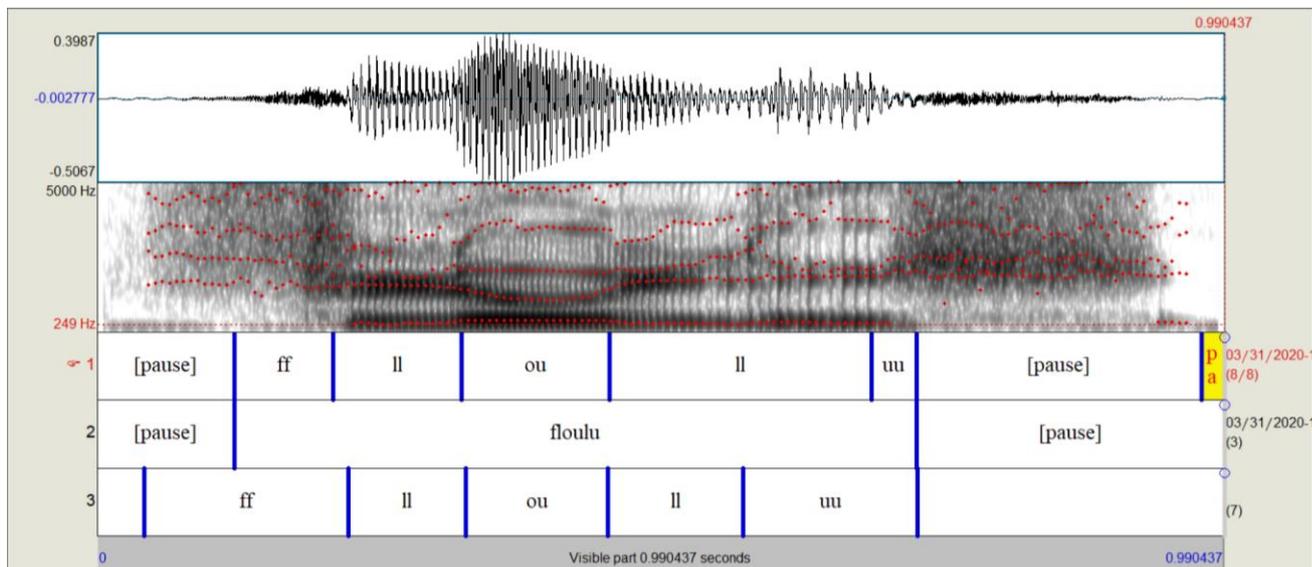


Figure 51 : Signal acoustique d'un patient produisant le pseudo-mot « floulu »

Ce signal acoustique représente la production du pseudo-mot « *floulu* » par un locuteur atteint d'un cancer de l'amygdale T1N2 douze ans avant l'enregistrement, au moment duquel il était âgé de 71 ans. Il a bénéficié d'une chirurgie des ganglions et d'une radiothérapie. Les formants F1 et F2 des segments vocaliques /u/ et /y/ présentent des valeurs dans le cadre de référence, cependant, les valeurs extraites à partir d'un alignement erroné apparaissent comme étant hors cadre.

Ces deux exemples illustrent des erreurs de segmentation et d'alignement des phonèmes avec le signal acoustique. Elles impliquent que la fenêtre du tiers milieu de la durée totale de la voyelle, à partir de laquelle est calculée la valeur moyenne de F1 et de F2 reportée ne soit pas correctement positionnée et les valeurs prises en compte dans le calcul de F1 et F2 sont donc erronées. Les résultats le sont consécutivement. Une hypothèse possible pour expliquer ces erreurs d'alignement pourrait être les particularités de production de ces deux locuteurs pour les exemples visés. Une pause inter syllabique, pour le premier exemple et une voix craquée sur la deuxième syllabe du deuxième exemple pourraient expliquer les erreurs d'alignement.

Un second type d'erreur d'alignement a été relevé dans notre analyse et pourrait confirmer notre explication ci-dessus. Certains locuteurs patients produisent des pauses au sein du pseudo-mot. Celles-ci sont mal interprétées par l'aligneur automatique, la segmentation ne s'opère pas sur la totalité du segment non lexical mais seulement sur la première syllabe. Au même titre que dans les exemples précédents, lorsque la segmentation du signal acoustique et l'alignement des phonèmes sont erronés, les portions de signal sur lesquels sont calculées les valeurs de F1 et F2 ne permettent pas de relever une valeur fiable.

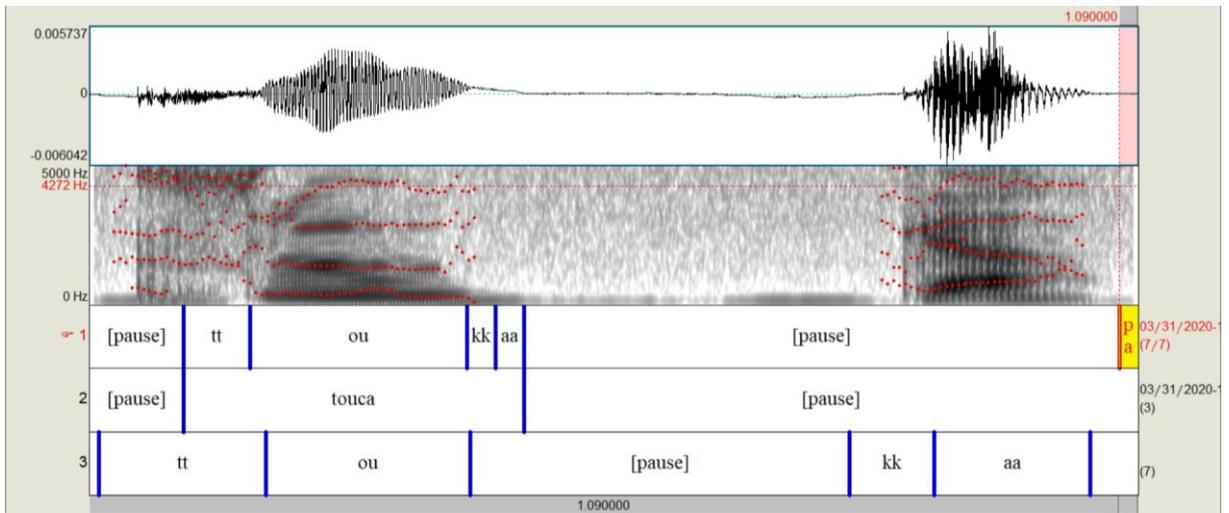


Figure 52 : Signal acoustique de la production du pseudo-mot « touca » par une patiente

Cet exemple de signal acoustique segmenté seulement sur la première syllabe du pseudo-mot « *touca* » illustre une erreur de segmentation due à une pause produite au sein de pseudo-mot, à moins qu'il ne s'agisse d'un allongement de la tenue de l'occlusive. La locutrice, âgée de 87 ans au moment de l'enregistrement, réalisé 2 ans et 2 mois après sa rémission d'un cancer de la mandibule, a bénéficié de chirurgies de la tumeur et des ganglions. La pause (ou allongement) produite au sein du pseudo-mot perturbe la segmentation automatique et l'alignement des phonèmes avec les portions de signal acoustique qui lui correspondent réellement.

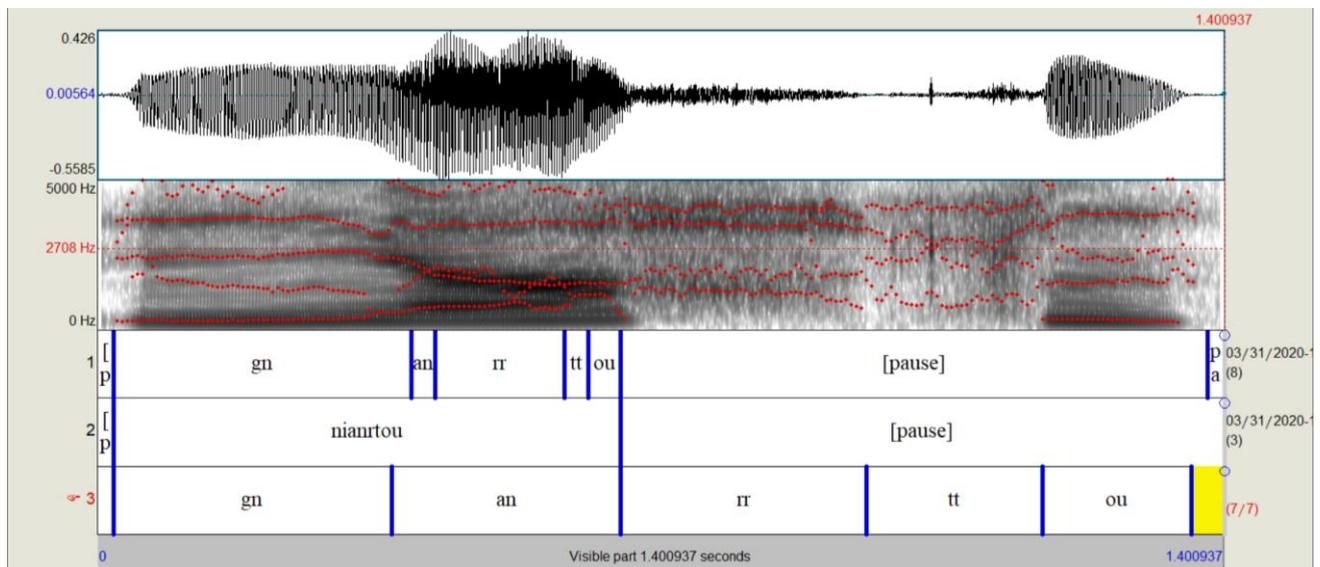


Figure 53 : Signal acoustique de la production du pseudo-mot « nianrtou » par un patient

Cet exemple de signal acoustique illustre une segmentation focalisée sur la première syllabe du pseudo-mot « *nianrtou* ». Le segment consonantique vélaire /r/ produit du bruit qui n'est pas interprété comme de la parole pour le logiciel de segmentation automatique. L'alignement des phonèmes avec le signal acoustique est perturbé. La troisième tire propose une correction de l'alignement phonémique. Il s'agit ici du même locuteur que pour le pseudo-mot « *flifou* » suscité. Dans ces deux exemples (« *touca* » et

« nianrtou »), la segmentation automatique n'a pas su faire face à la variabilité de la production introduite par une perturbation du signal en position intra-pseudo-mot.

D'autres signaux acoustiques comprenant des valeurs de F1 et F2 en dehors du cadre de référence ont montré des erreurs du logiciel Praat dans la détection des formants. Celles-ci sont imputables à la production du locuteur, qui dans certain cas, ne permet tout simplement pas de distinguer les différents segments produits sur le signal acoustique. Par exemple, le bruit produit par le locuteur au cours de sa production est représenté par des trajectoires de F1 et F2 constantes et parallèles tout au long du signal produit. Elles ne présentent pas les variations qui les caractérisent habituellement. L'exemple suivant illustre la production du pseudo-mot « zimi » par un locuteur âgé de 74 ans au moment de l'enregistrement, réalisé 10 mois après sa rémission d'un cancer de l'amygdale classé T4N2 traité par chirurgie de la tumeur, radiothérapie et chimiothérapie. Ce signal acoustique ne permet pas de différencier les différents segments produits. Le signal est parasité par du bruit qui efface les variations acoustiques.

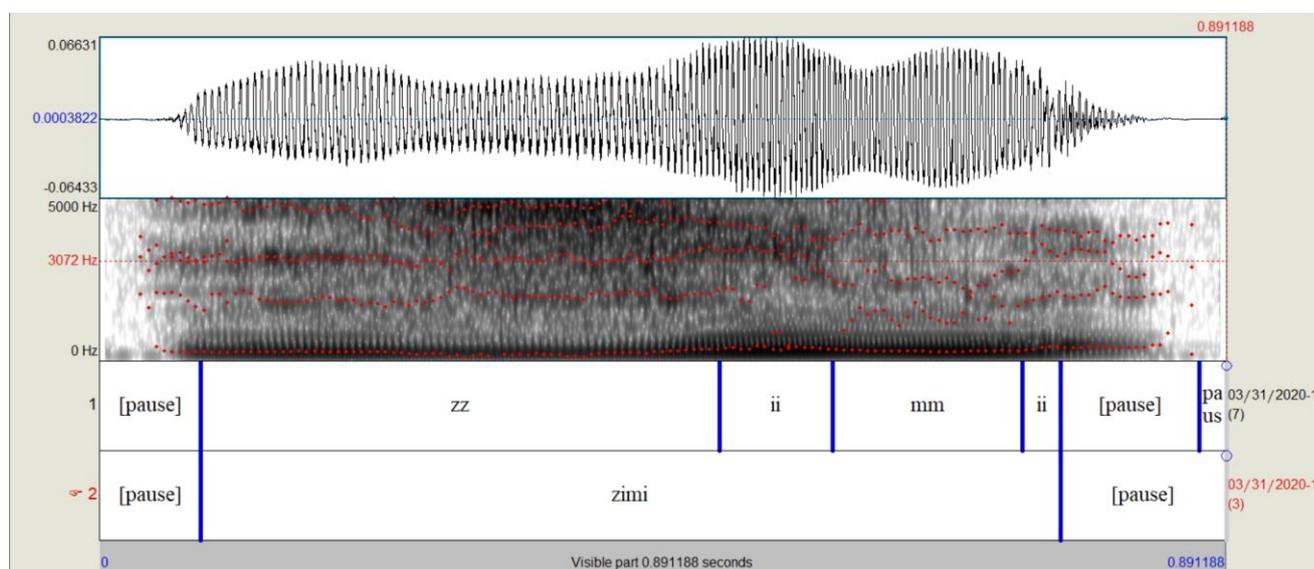


Figure 54 : Signal acoustique de la production du pseudo-mot « zimi » par un patient

Enfin, nous avons relevé que certains locuteurs ont commis des erreurs dans la production de pseudo-mots. Dans l'exemple suivant, la locutrice contrôlée de 56 ans a produit le segment /o/ à la place du segment /y/. Les trois transcriptions des auditeurs montrent que le segment a été identifié comme produit, soit comme un /o/. Les valeurs de F1 et F2 relevées sont proches des valeurs moyennes de /o/ pour une femme (Gendrot & Adda-Decker, 2005b). Ce segment a donc été mal produit par rapport à la cible attendue mais a été correctement transcrit par les 3 auditeurs par rapport à la production effectivement réalisée ce qui explique que les valeurs attendues d'un /y/ soient hors cadre.

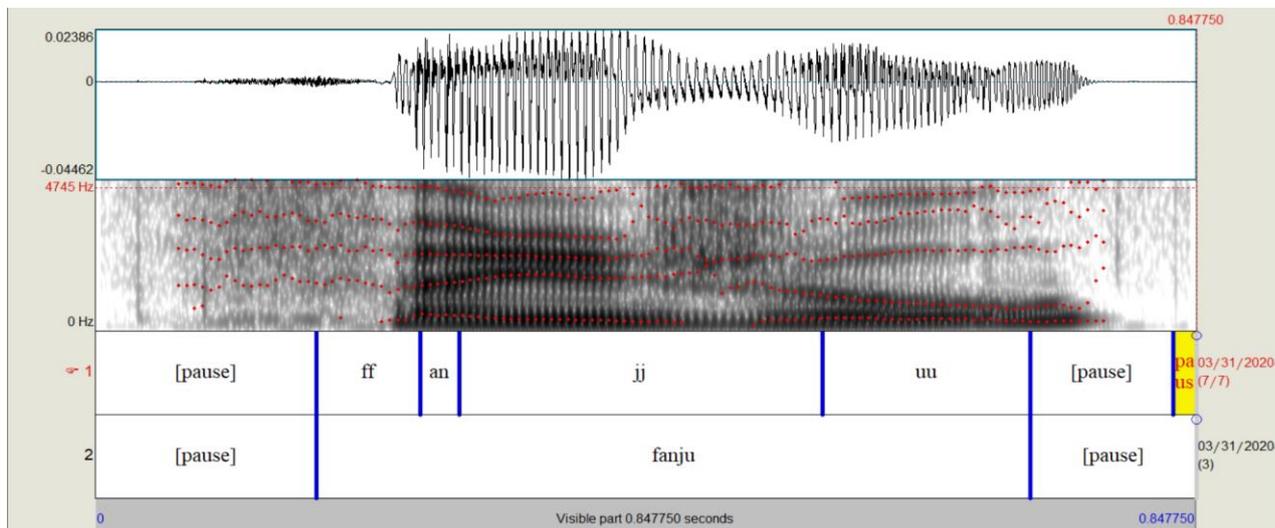


Figure 55 : Signal acoustique d'un patient qui aurait dû produire le pseudo-mot « fanju »

Pour résumer cette analyse descriptive, la majorité des valeurs de F1 et F2 hors du cadre de référence établit sont dues à des problèmes techniques de segmentation et de l'aligneur automatique. La source secondaire de ces valeurs est représentée par la production du locuteur dont le signal acoustique ne permet pas de distinguer les différents segments produits.

Concernant la première source d'erreur dite « d'analyse automatique », l'objet de cette thèse et nos compétences ne permettent pas d'entrevoir de solution. Cependant, notre examen détaillé des exemples ci-dessus permet d'identifier des imperfections, de proposer un certain nombre d'hypothèses (par exemple, voix craquée, pause inter mot) et ainsi apporter des éléments qui pourront être repris par les spécialistes du traitement automatique.

Les différents phénomènes potentiellement explicatifs de l'extraction de valeurs de formant hors cadre peuvent être résumés dans le tableau suivant :

<i>Cas des valeurs de formants en dehors du cadre établi</i>		
<i>Segment ciblé</i>	<i>Valeurs de formant</i>	<i>Transcription auditeur</i>
<i>x</i>	<i>y</i>	<i>x</i>
<i>x</i>	<i>y</i>	<i>y</i>
<i>x</i>	<i>y</i>	<i>z</i>

- > Erreur de détection de formant ou alignement
- > Erreur production, erreur détection,
- > Signal acoustique très dégradé

Tableau 18 : Conditions de considération des valeurs de F1 et F2

À l'issue de ce contrôle de la qualité de nos mesures acoustiques, la question se pose de l'avenir des valeurs hors cadre. Dans un tel travail, la suppression de valeurs dites aberrantes, sans connaître leur origine ne constitue pas une option méthodologique. Nous avons montré que la grande majorité des valeurs en dehors du cadre de référence établi est directement imputable à des erreurs de segmentation qui proposent une fenêtre erronée. Ces erreurs de segmentation sont issues des analyses automatiques. L'analyse de grands corpus entraîne l'utilisation de systèmes automatiques qui peuvent présenter, nous l'avons vu, certaines limites. En contrepartie, leur utilisation permet de travailler sur une quantité de données considérables. Ce qui nous semble primordial, c'est de

s'interroger sur l'origine de ces valeurs et de les quantifier avant de prendre une quelconque décision. C'est ce que ce contrôle nous a permis de faire.

Ainsi, nos observations des valeurs hors cadre de références nous ont conduites à exclure un locuteur qui cumule à lui seul 20% des valeurs non conformes de F1. Cela ramène l'estimation du taux relevé à 5,86 % de valeurs non conformes par rapport à l'ensemble de nos données, dont 0,93% de F1 et 4,93% de F2. Du reste nous avons choisi de conserver l'ensemble des valeurs en considérant une marge d'erreur. Mis en perspective avec la quantité de données dont il est question les modèles statistiques demeurent puissants et les représentations graphiques permettent de repérer aisément les quelques segments se détachant du nuage de point. Au regard de ce volume, nous acceptons le risque d'erreur au bénéfice de la puissance statistique.

Hormis pour le locuteur exclu des analyses, toutes les valeurs de formants ont été utilisées dans le calcul des différentes métriques. Nous avons choisi de ne pas convertir nos valeurs et de conserver la mesure des valeurs initialement extraites en Hertz, puisqu'elles présentent l'avantage d'être des mesures réelles qui sont interprétables au regard du signal acoustique. Elles présentent néanmoins l'inconvénient de surévaluer la valeur des écarts sur F2 (et davantage sur F3) par rapport à celle de F1. Nous l'avons en effet constaté, il y a plus de valeurs de F2 en dehors du cadre de référence établit que de valeur de F1. Cette problématique n'en est pas une ici puisque ce volume a été estimé.

6.2.2. Métriques

Grâce aux valeurs de formants ainsi extraites nous calculons différentes métriques, qui vont servir différents objectifs. Dans un premier temps, nous calculerons les métriques destinées aux analyses statistiques, puis nous décrirons celles employées dans les représentations graphiques. Ces représentations, telles que conceptualisées ici, constituent une interface entre des valeurs issues de la production et les données relatives à la perception de la parole.

6.2.2.1. Métriques Statistiques

L'ensemble des différents calculs présentés dans cette section ont été réalisés au moyen du logiciel statistique Rstudio, à l'aide d'un script produit et partagé par A. Marcyzk. Les métriques sont calculées à partir des valeurs de F1 et F2 dont elles constituent le socle.

6.2.2.2. Vowel Space Area (VSA) ou Aire du triangle vocalique

L'estimation de la dimension de l'aire du triangle vocalique, ou « Vowel Space Area » (VSA) revêt un intérêt particulier puisqu'elle permet d'évaluer la taille du système vocalique d'un locuteur. Les voyelles cardinales, en français /a/, /i/ et /u/, représentent les 'angles' de ce système. Les valeurs de F1 et F2 pour ces trois segments sont impliquées dans le calcul de cette métrique. Elle se calcule comme suit : $VSA = 0,5 * \{ (/i/F2 * /a/F1 + /a/F2 * /u/F1 + /u/F2 * /i/F1) - (/i/F1 * /a/F2 + /a/F1 * /u/F2 + /u/F1 * /i/F2) \}$ d'après (Kent & Kim, 2003), adapté au français. Cette métrique nous apparait comme

potentiellement adaptée au cadre pathologique puisque des éléments de littérature suggèrent que la compression de l'espace vocalique est une propriété retrouvée dans plusieurs dysarthries, et il est également admis qu'elle serait liée à l'intelligibilité (Kent et al., 1997). Nos travaux s'intéressent à la perte d'intelligibilité après un cancer des VADS, l'impact fonctionnel séquellaire est susceptible d'impacter la dimension du système vocalique.

6.2.2.3. Indice de Distinctivité (ID)

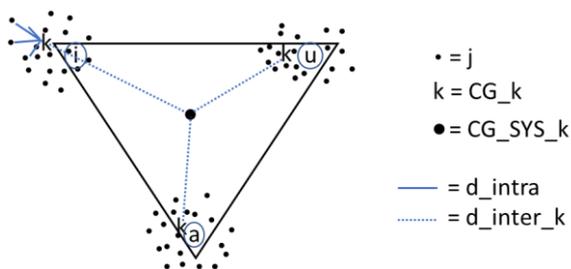
L'Indice de Distinctivité (ID) (Meunier & Ghio, 2018) représente le rapport entre la variation du système et la variation au sein d'une catégorie vocalique (variation intra-catégorie). Cet indice permet de caractériser la distinctivité entre chacune des catégories vocaliques. Il implique le calcul de plusieurs métriques intermédiaires. Nous mettrons en perspectives chacune d'entre elles et proposerons un récapitulatif de celles-ci.

Tout d'abord, notons que la conjonction des valeurs de F1 et F2 représente un point j dans le plan F1/F2 (donc chaque point j correspond à une voyelle). L'ensemble des points j forment nuage de points dont la moyenne est traduite par k , qui représente le centre de gravité de la catégorie (CG_k). La moyenne des centres de gravité de chaque catégorie représente le centre de gravité du système (CG_SYS_k). L'écart entre un point j et la moyenne du nuage k représente la variation inter-catégorie (d_{intra}). L'écart entre le centre de gravité du système CG_SYS_k et k représente la variation du système (d_{inter}).

Pour le calcul de l'Indice de Distinctivité, le principe est de calculer les écarts entre CG_k et chaque point j , pour obtenir d_{intra} . Puis chaque écart est mis au carré pour obtenir des valeurs en valeur absolue. Puis la racine carrée de l'ensemble est calculée pour s'affranchir des valeurs au carré (valeurs réalistes) nous donnant une valeur de $d_{intra_ID} = \text{Racine Carrée} ((F(j) - F(k))^2)$. Ensuite, nous devons calculer les écarts entre CG_SYS_k et chaque k pour obtenir d_{inter_k} qui représente la variation du système. Ce calcul représente les écarts entre CG_SYS_k et chaque k . A nouveau, les écarts sont mis au carré puis la racine carrée de l'ensemble est calculée. Ainsi, $d_{inter_k_ID} = \text{Racine Carrée} ((F(k) - F(CG_SYS_k))^2)$. Enfin, nous pouvons procéder au calcul de l'indice de distinctivité qui représente le rapport entre la variation du système et la variation intra-catégorie, soit $ID = d_{inter_k_ID} / d_{intra_ID}$, un indice qui exprime la distinctivité entre chaque catégorie de segment du système vocalique. Ainsi plus l'ID est élevé plus les catégories de voyelles produites sont distinctes les unes des autres.

Notons que nous proposons ici la version « prototype » de ce calcul, mais qu'il lui existe un équivalent par exemplaire, où k est remplacé par j dans l'ensemble des calculs réalisés. Cette version à exemplaires est adaptée pour être utilisée lorsque le déséquilibre des données est une variable pertinente pour l'étude (comparaison de corpus de langues différentes par exemple). Pour plus de précisions sur cet indice et de détails sur son calcul nous orientons le lecteur vers la publication (Meunier & Ghio, 2018).

j = un point dans le plan F1/F2
 k = CG_k = moyenne du nuage de point j
 CG_SYS_k = moyenne de CG_k
 d_intra = écart entre CG_k et j
 d_inter_k = écart entre CG_SYS_k et k
 $ID_Prototype$ = $d_inter_k_ID / d_intra_ID$



6.2.2.4. Objectif

Plusieurs études associent des mesures acoustiques (F0, F1, F2, trajectoire de formants) avec des mesures scorées d'intelligibilité, ou établissent un lien entre la dimension de l'espace vocalique et la parole dysarthrique (Kent et al., 1997; Weismer et al., 2001). Ici, nous proposons d'évaluer la relation entre les métriques décrites précédemment (VSA et ID) et les scores de Déviation phonologique Perçue (PPD) Globaux et Locaux. Ainsi, nous associons une mesure de l'intelligibilité segmentale, PPD Global, distance cumulée, et PPD LocalVoy, distance locale, avec des mesures représentatives de l'espace vocalique et de la distinctivité de ces segments.

L'objectif des analyses statistiques est de déterminer le pouvoir discriminant de chaque métrique vis-à-vis du groupe de locuteur. Puis de déterminer si ces métriques sont congruentes et complémentaires ou plutôt concurrentes par rapport aux scores PPD. Si le score PPDLocalVoy se révèle plus discriminant que le PPDGlobal les corrélations avec les métriques acoustiques devraient être fortes, en revanche, si le PPDGlobal est plus discriminant que le PPDLocalVoy, les corrélations de ce dernier avec les métriques devraient être réduites. Si les scores PPDGlobaux et LocauxVoy se révèlent équivalents on peut s'attendre à ce que les effets observés convergent. Nous avons employé la méthode de régression linéaire afin d'évaluer dans quelle mesure les métriques acoustiques peuvent prédire une évaluation perceptive de l'intelligibilité segmentale.

En termes de prédictions, nous émettons l'hypothèse selon laquelle les scores PPD élevés, exprimant un degré d'altération important, devraient refléter une réduction de la dimension de l'aire du triangle vocalique. Les études (Kent et al., 1997; Weismer et al., 2001) suggèrent une compression de cet espace dans les dysarthries, on peut imaginer que cette conclusion puisse être étendue à d'autres troubles de la production de la parole, comme ici les troubles fonctionnels après un cancer de la cavité buccale et/ou de l'oropharynx.

Concernant l'indice de distinctivité, nous suggérons qu'un indice de distinctivité élevé exprimant une distinction importante entre les catégories de segments vocaliques devrait refléter des scores PPD peu élevés montrant peu d'altération de l'intelligibilité. En d'autres termes, plus le déficit d'intelligibilité segmental est important plus la distinction entre les catégories de voyelle devrait être réduite. La pathologie, en compressant l'espace vocalique affecte la distinction des catégories de segments. Néanmoins, Meunier & Ghio, (2018) suggèrent qu'il existe deux types de systèmes, concentré versus dispersé, au sein desquels les voyelles peuvent elles aussi être concentrées ou dispersées. Ce qui, au regard

de la diversité des profils de patients, nous conduit à nuancer la relation entre les scores PPD et l'ID, puisqu'un système concentré peut montrer une bonne distinctivité et qu'à l'inverse un système plus dispersé est susceptible d'exprimer moins de distinctivité.

Parallèlement à ces métriques statistiques nous avons recours aux métriques de visualisation exploitées dans le cadre d'analyses descriptives.

6.2.2.5. Métriques de visualisation

Certaines des métriques préalablement utilisées dans le calcul de l'Indice de Distinctivité peuvent être réemployées au service de la visualisation puisqu'elles peuvent être placées dans la plan F1/F2. Le centre de gravité de la catégorie (CG_k) et le centre de gravité du système (CG_SYS_k) seront placés sur la représentation. Ces mesures permettent d'obtenir une image des productions des locuteurs et de caractériser leurs systèmes vocaliques au regard de trois informations. Dans un espace bidimensionnel, la conjonction de la valeur de F1 et de F2 permet d'obtenir un point pour chaque production. Le calcul de CG_k offre un point de représentation pour chaque catégorie de segment et le CG_SYS_k un point pour l'ensemble du système. Ces derniers permettent d'avoir une représentation pour chaque groupe de locuteurs (ou tout autre groupe jugé utile (homme/femme, CB/OP, etc..)) et de comparer des tendances en conservant l'ensemble des points j qui représentent chaque production. Nous pouvons alors illustrer plusieurs niveaux de représentation à partir de ces mesures. Notons que les valeurs acoustiques sont reconnues pour être sensibles au sexe du locuteur, ainsi les femmes présentent un espace vocalique généralement plus grand que les hommes. Nous considérerons donc ce facteur dans nos analyses descriptives, puisque les métriques calculées à partir des valeurs de formants sont susceptibles d'y être également sensibles.

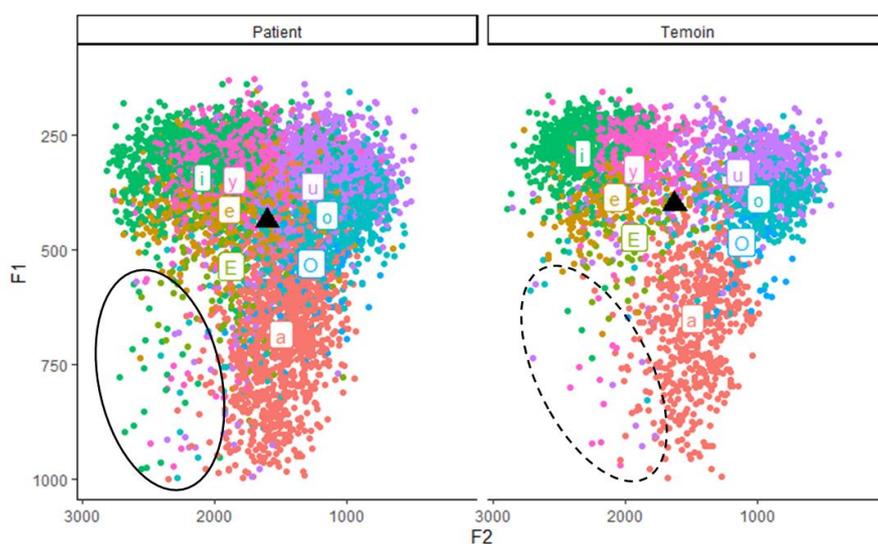


Figure 56 : Illustration de la visualisation de j

La Figure 57 illustre la représentation des valeurs de j , CG_k et CG_SYS_k dans l'espace bidimensionnel du plan F1/F2. Chaque couleur représente une catégorie de voyelle, chaque point est la conjonction des valeurs de F1 et F2 soit j , chaque étiquette de catégorie représente le centre de gravité de la catégorie et le triangle représente le centre de gravité du système. Ici les valeurs CG_k et CG_SYS_k sont moyennées par groupe de locuteur

(patient et contrôle). Les ellipses ajoutées sur la figure illustrent les valeurs hors du cadre de référence dont il a été question dans la première partie de cette section méthodologique. En effet ces représentations permettent de caractériser la répartition des mesures acoustiques de production et d'identifier visuellement celles qui s'éloignent anormalement de son nuage de point.

Nous mettrons ces représentations en lien avec les réponses des auditeurs aux tests de perception. Chaque point pourra être associé aux transcriptions proposées par les auditeurs en termes de transcription phonémique. Ainsi nous associons une mesure acoustique de production soit un point j dans le plan F1/F2 avec une transcription perceptive impliquée dans le calcul d'une distance locale entre la cible qui devait être produite et celle qui a été perçue.

Nous proposerons une analyse descriptive des systèmes vocaliques des locuteurs. Nous comparerons les deux groupes de locuteurs, patient et contrôle, en fonction de leur sexe. Les visualisations des groupes patients devraient montrer un espace vocalique plus compressé, par rapport à leur groupe contrôle respectif en termes de sexe. De plus, la dispersion des catégories devrait montrer plus de variabilité dans les groupes patients, entraînant davantage de chevauchement entre celles-ci par rapport aux groupes contrôles.

Concernant les visualisations associant les transcriptions des auditeurs avec j , les points éloignés du nuage de leur catégorie devraient montrer des transcriptions divergentes de leurs catégories cible. Il est également envisageable qu'un point se positionnant au centre du nuage d'une autre catégorie ait été transcrit comme celle-ci et non comme sa cible attendue. Néanmoins, au regard de l'étude des transcriptions effectives des auditeurs (5.3.2 Transcriptions des auditeurs), ces derniers transcrivent préférentiellement les segments proches du segment cible.

Les résultats des analyses conduites à partir des métriques statistiques et des métriques de visualisation, afin de mieux caractériser les altérations de la parole des patients par rapport aux locuteurs contrôles, sont présentées dans la partie suivante.

6.3. Résultats

6.3.1. Statistiques

Une analyse de variance (ANOVA) a été conduite, afin de tester les effets simples entre les scores perceptifs (PPDGlobal et PPDLocalVoy) et les métriques acoustiques (ID et VSA) avec les groupes de locuteurs patient et contrôle. Les mesures perceptives se sont révélées plus pertinentes que les mesures acoustiques pour discriminer deux groupes de locuteurs. Le PPDGlobal a montré un effet significatif ($F(1,122) = 59,35, p < 0.001$) sur les groupes de locuteurs (contrôles $\mu = 0.476$; patients $\mu = 1.276$) ce qui confirme sa capacité à les distinguer. Le PPDLocalVoy montre la même tendance que le PPDGlobal, il s'est révélé significatif ($F(1,122) = 53,81, p < 0.001$) pour discriminer les deux groupes de

locuteurs (contrôles $\mu = 0.278$; patients $\mu = 0.833$). La différence entre les scores moyens des groupes de locuteurs pour les scores globaux et locaux des voyelles peut s'expliquer par le nombre de segments considérés pour chacun des scores. Plus il y a de segments pris en compte plus ils sont susceptibles de contribuer à la distinction des groupes de locuteurs. Parmi les métriques acoustiques testées seul l'Indice de Distinctivité s'est révélé significatif ($F(1,122) = 17,14$, $p < 0.001$) dans sa capacité à discerner deux groupes de locuteurs. Le système des locuteurs contrôles présente plus de distinction que le système des patients (contrôles $\mu = 3.17$; patients $\mu = 2.28$). L'aire du triangle vocalique VSA n'a montré aucun effet sur les groupes de locuteurs ($p = 0.35$). Ceci indique que cette métrique n'est pas adaptée pour cette distinction (patients vs contrôles). Ces résultats nous invitent à nuancer nos prédictions, les corrélations entre les scores perceptifs et les mesures acoustiques devraient être positives pour l'ID mais l'aire du triangle vocalique VSA ne devrait pas montrer de relation avec celle-ci.

Afin de déterminer si les variables acoustiques revêtent une capacité à prédire les scores issus de l'évaluation perceptive, nous avons testé la corrélation entre les scores PPDGloiaux et PPDLocauxVoy et l'aire du triangle vocalique ainsi que l'Indice de Distinctivité, au moyen d'une méthode de régression linéaire. Elle permettra également d'évaluer le sens de l'interaction et dans quelle mesure ces évaluations perceptives et acoustiques sont complémentaires et congruentes ou concurrentes.

Comme attendu, l'aire du triangle vocaliques VSA n'a montré aucun effet de corrélation avec les scores perceptifs, que ce soit pour les PPDGloiaux ($p = 0.632$) ou pour les PPDLocauxVoy ($p = 0.988$). Cela signifie que la compression du système vocalique n'est pas une caractéristique spécifique de ce groupe de patients et n'entretient aucune relation avec les scores perceptifs.

En revanche, l'Indice de Distinctivité a montré un effet de corrélation significatif avec les scores perceptifs. Celui-ci corrèle avec les scores globaux ($F(1,122) = 20.93$, $r = 0.146$, $p < 0.001$) et locaux ($F(1,122) = 24.83$, $r = 0.169$, $p < 0.001$) de façon significative. Néanmoins, la considération du groupe de locuteur dans le modèle de régression linéaire révèle que cette corrélation n'est effective que pour le groupe Patient, ($F(3,120) = 26.11$, $r = 0.39$, $p < 0.016^*$) et qu'aucun effet n'est révélé pour le groupe Contrôle ($p = 0.96$), pour les scores perceptifs globaux. On observe les mêmes résultats pour les scores perceptifs locaux avec un effet de l'ID pour le groupe Patient ($F(3,120) = 26.45$, $r = 0.39$, $p < 0.006^{**}$) mais pas d'effet pour le groupe Contrôle ($p = 0.87$).

Cela signifie que l'Indice de Distinctivité représente une métrique complémentaire des scores perceptifs pour le groupe patient. Cela suggère que cet indice peut être approprié pour contribuer à l'évaluation de l'intelligibilité segmentale des patients. Toutefois, ces résultats doivent être nuancés puisque le corpus demeure hétérogène. Il est probable que cette métrique soit particulièrement adaptée pour certains locuteurs mais pas pour d'autres, au regard de la spécificité de leur trouble fonctionnel. Seule une étude de cas permettrait d'approfondir cette question.

Ces différences de corrélation entre les scores perceptifs et l'ID pour les groupes patients et contrôles nous ont poussée à définir de nouveaux groupes de locuteurs en fonction des scores perceptifs obtenus. Pour cela, nous avons divisé la cohorte de locuteurs en trois groupes, sur la base du seuil de la courbe ROC (Ghio, Lalain, et al., 2020) établit à 0,6 (traits moyen d'écart par phonème). Le premier groupe a un score PPDGlobal inférieur ou égal à 0,6 (« Low ») le second groupe à un sPPD compris entre 0,6 et 1,2 (« Medium ») et enfin le dernier groupe à un score PPDGlobal supérieur à 1,2 (« High »). Nous avons reproduit ce même filtrage pour les scores PPDLocaux des voyelles. Il est à noter que la division en fonction du score global produit des groupes équilibrés en nombre de locuteurs (40, 44 et 40 pour les PPD Low, Medium et High respectivement), alors que celle basée sur les scores locaux des voyelles produit des groupes déséquilibrés avec 63, 44 et 17 locuteurs dans les groupes de PPD locaux Low, Medium et High respectivement.

Les résultats ont montré que l'indice de distinctivité corrèle significativement avec les scores PPDGlobaux Low et High ($F(2,121) = 13.94$, $r = 0.187$, $p < 0.001$), mais cette relation est moins forte pour les scores PPDGlobaux Medium ($p = 0,059$). Néanmoins, dans le cas des scores perceptifs Locaux, les seuils Low Medium et High corrèlent avec l'ID de façon significative ($F(2,121) = 17.16$, $r = 0.221$, $p < 0.001$). Notre analyse montre donc que l'indice de distinctivité est corrélé avec les scores PPD pour les PPD bas et élevés, mais peu avec les sPPD moyens dans les scores perceptifs globaux, mais l'ID est corrélé avec les trois seuils préétablis dans les scores perceptifs Locaux. Cela peut s'expliquer par le fait que par essence, l'Indice de Distinctivité corrèle davantage avec un score perceptif calculé à partir des segments vocaliques uniquement plutôt qu'à partir de plusieurs types de segments car son calcul est basé sur ces mêmes segments vocaliques.

Ainsi, l'Indice de Distinctivité constitue une métrique complémentaire aux scores perceptifs, qui peut être congruente dépendamment du groupe de locuteurs et des besoins thérapeutiques du groupe patient. Les scores PPD globaux et locaux sont informatifs sur le degré d'altération, l'ID permet d'inférer le système vocalique mais la part de score et donc d'intelligibilité prédite par celui-ci est relativement faible (en moyenne 20%). Ainsi, la capacité de prédiction de l'indice de distinctivité est limitée au groupe patient, c'est donc un prédicteur à employer à bon escient, quand il est pertinent, mais qui ne peut être étendu à un ensemble de locuteurs.

Ces résultats suggèrent que la maximisation de la distinction entre chaque catégorie de voyelles permet probablement d'accroître l'intelligibilité segmentale dans un contexte pathologique mais pas dans le cas de parole saine. Néanmoins, cet indice se révèle pertinent au cas par cas dans le contexte pathologique, puisqu'il pourrait permettre d'adapter la prise en charge du patient pour lui permettre, si besoin, de tendre vers plus de distinctivité entre les catégories vocaliques. La capacité de prédiction des métriques acoustiques sur les scores perceptifs s'est révélée inexistante pour l'aire du triangle vocalique VSA et nuancée pour l'Indice de Distinctivité. Cependant, un panel d'autres métriques pourrait montrer des effets de prédictions sur les scores perceptifs, ces relations restent à déterminer.

Pour obtenir une image de la contribution des voyelles à l'intelligibilité segmentale globale nous avons testé la corrélation entre les scores PPDGlobaux et les scores PPDLocaux. Le modèle de régression linéaire a montré une corrélation significative ($F(1,122) = 307.1, r = 0.77, p < 0.001$) entre ces scores. Ce coefficient de corrélation indique que le score PPDLocalVoy permet d'expliquer 77% des scores PPDGlobaux. Ces résultats suggèrent que les scores PPDLocalVoy ont un pouvoir de prédiction des scores globaux conséquent. Dans un test de perception basé sur des pseudo-mots, les voyelles contribuent pour près de 80% à l'intelligibilité segmentale.

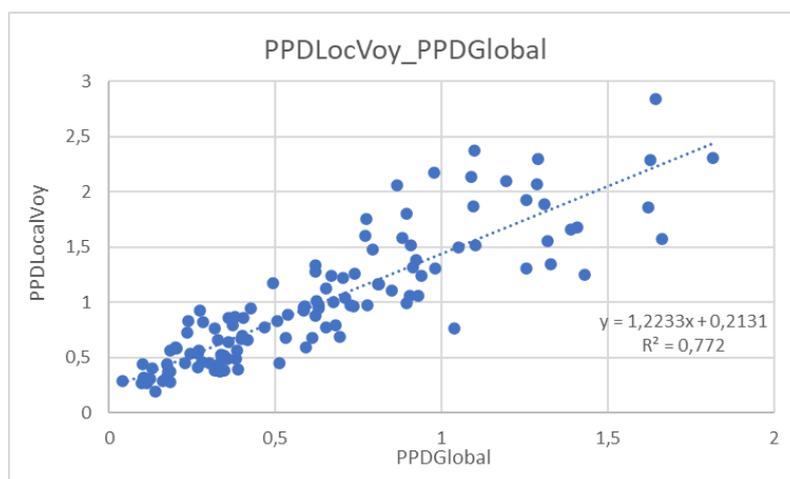


Figure 57 : Droite de régression linéaire illustrant la corrélation entre les Scores PPDGlobaux et les scores PPDLocaux des voyelles.

Ces résultats ont permis de caractériser les liens entre les évaluations perceptives et les mesures acoustiques. L'analyse descriptive présentée ci-après à partir des métriques de visualisation, également impliquées dans le calcul de l'ID, devrait faire écho aux analyses ci-dessus.

6.3.2. Visualisation

Dans cette section nous présenterons les résultats de la représentation des métriques j , CG_k , CG_SYS_k dans l'espace bidimensionnel du plan F1/F2 au moyen de graphiques.

Dans un premier temps nous comparerons l'allure générale des nuages de points, la position de Centres de Gravité de chaque catégorie et du centre de gravité du système par groupe de locuteurs (patient et contrôle) et par sexe (homme et femme). Dans ces représentations, chaque point j représente un segment vocalique produit, les labels (étiquettes de la catégorie) représentent les centres de gravité de chaque catégorie (CG_k), enfin le losange noir illustre le centre de gravité du système (CG_SYS_k). Nous avons également fait apparaître des ellipses qui illustrent la dispersion de chaque catégorie, et permettent d'estimer le chevauchement ou recouvrement des catégories les unes par rapport aux autres. Ensuite, nous proposerons des représentations similaires en remplaçant les points des nuages par les transcriptions phonémiques locales des auditeurs aux tests de perception, laissant transparaître deux niveaux d'informations supplémentaires : savoir si le segment cible qui correspond à un point dans le nuage a été

correctement identifié, et dans le cas contraire, par quel segment il a été transcrit et si la transcription est cohérente avec la place du point dans l'espace.

Tout d'abord, nous comparerons les deux groupes de locuteurs Patient et Contrôle au regard de leur sexe pour les hommes puis les femmes.

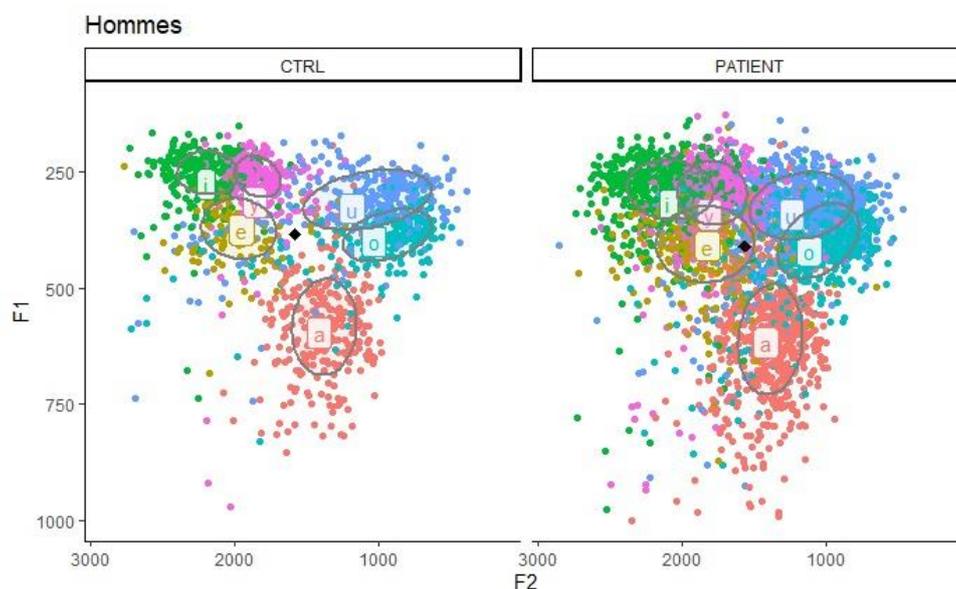


Figure 58 : Plan F1/F2 pour les locuteurs masculins, Contrôle vs. Patient

Cette figure 13 illustre la comparaison des groupes de locuteurs contrôles et patients pour les locuteurs de sexe masculin de ce corpus. L'allure des nuages de points montre plus de dispersion dans le groupe patient par rapport au groupe contrôle. On observe plus de recouvrement dans le groupe patient, les segments /i/, /y/ et /e/ se chevauchent, alors que cet effet n'est pas présent dans le groupe contrôle. En revanche, dans ce dernier les segments d'arrière /u/ et /o/ présentent un léger recouvrement, cet effet semble augmenté par la pathologie puisqu'il se retrouve dans le groupe patient et qu'il est plus important. Le recouvrement inter catégories observé pour les patients est également visible entre les voyelles hautes antérieures et postérieures. Ces observations résonnent non seulement avec les résultats des mesures acoustiques mais aussi avec les résultats de l'analyse des PPD locaux (5.3.1 PPD locaux). Dans les deux groupes de locuteurs, le segment bas /a/ ne présente aucun chevauchement avec un autre segment ce qui fait de lui le segment le plus distinctif sur le plan acoustique. Les centres de gravité de chaque catégorie, matérialisés par le label de la catégorie, sont plus proches du centre de gravité du système, en particulier pour le groupe patient (il va jusqu'à être inclus dans l'ellipse du segment /e/). Ceci suggère un système, en moyenne, plus compressé pour le groupe patient, par rapport au groupe contrôle. La taille du système est susceptible d'impacter la distinction entre les catégories de segments, comme relevé dans la première partie de ce commentaire. De plus, le centre de gravité du système est plus bas et plus postérieur dans le groupe patient par rapport au groupe contrôle. Ceci suggère que les segments vocaliques hauts sont moins distincts chez les patients par rapport aux sujets contrôles.

Nous proposerons un commentaire du même esprit pour la comparaison des groupes patient et contrôles des locutrices féminines.

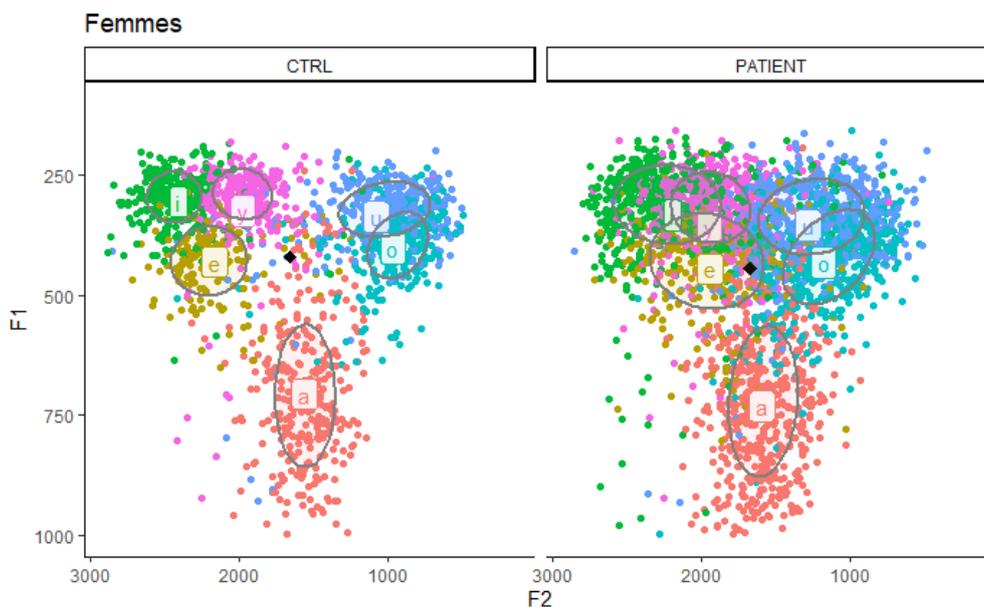


Figure 59 : Plan F1/F2 pour les locutrices féminines, Contrôle vs. Patient

Dans cette figure représentant les systèmes vocaliques des femmes pour les groupes patient et contrôle, on remarque qu'à l'image des hommes, les nuages de points des productions des patientes présentent plus de dispersion par rapport au groupe contrôle. De plus, dans le groupe patient, les segments /i/, /y/ et /e/ se chevauchent de façon importante alors qu'à l'inverse, ces segments sont bien dissociés dans le groupe contrôle. Les segments /u/ et /o/ se chevauchent chez les deux groupes de locutrices mais cet effet est plus important dans le groupe patient. Le segment /a/ demeure bien distingué des autres segments dans les deux groupes de locuteurs féminins. Les centres de gravité de chaque catégorie sont plus proches du centre de gravité du système dans le groupe patient, pour lequel il s'inclut dans l'ellipse du segment /e/.

Les différences entre les systèmes du groupe patient montrent des effets similaires entre les patients masculins et féminins, par rapport au groupe contrôle. En revanche, la comparaison des systèmes des hommes et des femmes met en lumière des différences. Comme attesté dans la littérature, les systèmes vocaliques des femmes sont plus grands que ceux des hommes. Cette caractéristique observée chez le groupe contrôle se retrouve dans le groupe patient ce qui suggère que la pathologie n'affecte pas toutes les propriétés du système acoustique.

Dans le groupe contrôle, les catégories /i/, /y/ et /e/ sont plus distinctes chez les femmes que chez les hommes. En revanche, dans le groupe patient, ces segments présentent davantage de chevauchement chez les femmes que chez les hommes. On observe donc un effet de chevauchement plus important chez les femmes patientes que chez les hommes patients, alors qu'on observe plus de distinction entre ces segments dans le groupe de femmes contrôles par rapport aux hommes. Cela suggère que les propriétés acoustiques des productions des femmes sont davantage impactées par la pathologie par rapport aux productions des hommes. De plus, les ellipses représentant un indice de dispersion des catégories, présentent plus d'amplitude chez les femmes patientes par rapport aux hommes du même groupe. Dans ce dernier les segments /i/, /y/, et /e/ se

chevauchent davantage dans le groupe de femmes par rapport aux hommes. Le chevauchement des segments /u/ et /o/ présente une dynamique similaire entre les groupes patients. Dans les groupes contrôles, les segments /i/, /y/ et /e/ se distinguent bien les uns des autres, chez les femmes la distinction est plus importante que chez les hommes. En revanche, dans les deux groupes de sexe on observe un chevauchement entre des segments /u/ et /o/ mais sa dynamique est différente, elle ne suit pas la même orientation.

Le segment /a/ plus distinct des autres segments, pour tous les groupes, toutefois la dispersion de ce segment est plus importante chez les femmes que chez les hommes, quel que soit le groupe de locuteurs.

Les résultats de cette étude ont montré que les scores PPD locaux permettent de distinguer deux groupes de locuteurs. Ils ont montré que la dimension de l'espace vocalique n'était pas corrélée avec les scores perceptifs locaux et globaux. La corrélation de ces scores perceptifs avec l'indice de distinctivité ne s'est révélée positive que dans le groupe patient. Cela suggère que l'indice de distinctivité représente un indice complémentaire aux scores perceptifs. Les métriques de visualisation ont confirmé les différences entre les dimensions des systèmes des hommes et des femmes. Ils ont également montré que les patients présentent davantage de chevauchement entre les catégories, par rapport au groupe contrôle. Cela suggère que la distinction entre les catégories impacte davantage l'intelligibilité segmentale chez les patients.

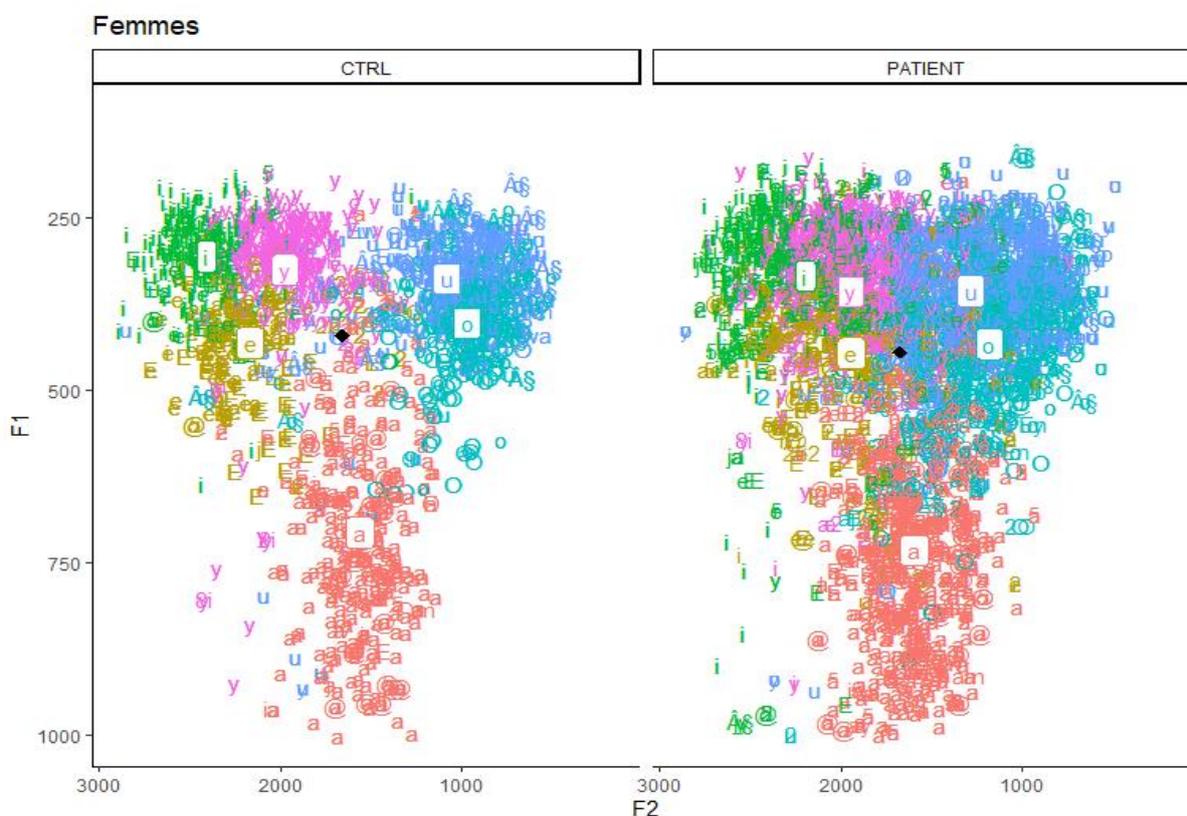
Aux termes de ces travaux, nous avons montré que l'approche segmentale est pertinente pour une évaluation de l'intelligibilité de ce niveau linguistique. L'approche par décodage acoustico-phonétique basée sur des pseudo-mots et sur le calcul d'un score analytique, complétée par l'approche par mesures acoustiques offrent un panorama riche des productions et perceptions des segments vocaliques. Ces résultats seront discutés dans la dernière section de ce document.

Toutefois, avant d'aborder cette discussion, nous nous permettons une parenthèse imaginative. Aux termes de ces analyses acoustiques, nous ne pouvons nous empêcher d'envisager une possible application clinique idéale. Dans cette partie, nous mettrons en avant un aspect de notre réflexion qui a émergé de ces analyses. En effet, nous l'avons évoqué, la conjonction des valeurs acoustiques de F1 et F2 avec les transcriptions effectives des auditeurs offre une interface visuelle entre production et perception. Les mesures acoustiques sur le signal de parole, mesures indirectes, sont associées aux réponses segmentales des auditeurs aux tests de perception, mesures directes.

Nous proposons donc une représentation des systèmes vocaliques où chaque point a été remplacé par la transcription phonétisée locale des auditeurs aux tests de perception. Cette représentation permet d'associer visuellement une mesure acoustique (F1/F2) avec une transcription perceptive.

La Figure 60 représente ainsi exactement la même répartition des données que la Figure 59. La différence réside dans les points de conjoncture F1 / F2. Elle fait apparaître

les transcriptions des auditeurs au test de perception. Sur cette figure sont représentés les réponses correctes, pour lesquelles la transcription correspond à la cible attendue et le score PPD local est de zéro, et les substitutions détaillées dans la partie 5.2, pour lesquelles la transcription proposée est différente de la cible attendue et le score PPD local supérieur ou égal à 1.



Nous précisons que la clarté de la représentation pourrait être améliorée par la réduction du nombre de locuteurs, néanmoins cette proposition constitue davantage une réflexion et un questionnement autour de l'application au contexte clinique, qu'un résultat en lui-même. C'est une réflexion qu'il nous tenait particulièrement à cœur de présenter ici. Dans le cadre des projets de recherche, nos collègues du LIA travaillent sur des mesures automatiques et sur l'apprentissage des réseaux de neurones (deep learning), pour contribuer également à l'amélioration de l'évaluation de l'intelligibilité. Leur approche et leurs outils automatiques ont participé à l'émergence de notre présente proposition.

Dans le cadre d'une possible application clinique, on pourrait imaginer qu'une telle représentation visuelle soit obtenue à partir de l'enregistrement des productions d'un patient, quasi immédiatement au cours du bilan. Ainsi, le locuteur produirait des pseudo-mots et serait enregistré. L'auditeur évaluateur pourrait écouter et transcrire les productions du patient en direct ou en différé, au moyen d'une application qui recueillerait ses transcriptions. Cet enregistrement pourrait être simultanément traité de façon automatique pour extraire les valeurs de F1 et F2. L'automatisation de l'ensemble

Discussion

Dans cette thèse, nous avons tenté d'expliciter la complexité de la notion d'intelligibilité, au travers des problématiques liées à son évaluation clinique. Nous avons vu qu'au cours des années, voire des siècles, cette notion a fait l'objet de réflexions, de propositions de définitions mais sans qu'un consensus ne puisse émerger. Selon nous, c'est cette absence de consensus et de définition claire qui entraîne les difficultés et les limites de son évaluation perceptive clinique. Ceci probablement en raison d'un manque de prise en considération de la complexité des mécanismes de communication parlée et de la complexité de l'organisation langagière. De plus, la notion d'intelligibilité, nous l'avons montré, est quasi nécessairement liée à la notion de mesure, d'évaluation, que ce soit dans un contexte clinique, ou d'apprentissage. Au terme de nos réflexions, il nous a semblé qu'une définition opérationnelle de l'intelligibilité devrait comprendre le locuteur, l'auditeur, le contexte communicationnel ainsi que le niveau linguistique évalué.

Ainsi, nous concevons l'intelligibilité comme « le degré de précision avec lequel les unités linguistiques d'un signal acoustique de parole, produit par un locuteur, sont décodées par un auditeur ». Cette définition reflète mieux selon nous la complexité qui réside dans la notion d'intelligibilité : l'évaluation de la qualité de production des unités linguistiques produites par un locuteur, au travers de la perception d'un auditeur et au moyen d'une méthode d'évaluation segmentale. Une telle définition pourrait être plus précise et surtout adaptable aux enjeux de l'évaluation. Par exemple : « le degré de précision avec lequel les unités [lexicales] [prosodiques] d'un signal acoustique de parole, produit par un locuteur, sont [identifiées] [décodées] par un auditeur ».

Ce concept ainsi défini est adapté à l'évaluation et peut être étendu à d'autres pathologies (dysarthries). L'enjeu de l'évaluation de cette composante dans le domaine clinique repose sur la production d'unités linguistiques de parole qui doivent être contrôlées pour permettre d'évaluer le niveau souhaité. Dans le cadre clinique de la prise en charge des séquelles fonctionnelles après un cancer de la cavité buccale ou de l'oropharynx, et telle que conceptualisée ici en termes d'unités ciblées, la notion d'intelligibilité segmentale concerne le niveau articulatoire de la parole et donc de la production des unités phonémiques de bas niveau.

L'efficacité des mécanismes de perception de la parole qui permet la réussite de l'échange communicationnel même dans des conditions défavorables, doit être réduite et contrôlée, dans le cadre d'une évaluation perceptive clinique. En particulier, le recours de l'auditeur aux informations de haut niveau nécessite dans notre cas, d'être neutralisé. Comme nous l'avons montré dans la partie consacrée, la sous-évaluation de l'intelligibilité provient souvent du fait que l'on mesure les performances du locuteur, bruitées par celles de l'auditeur. Pour neutraliser les biais perceptifs de l'auditeur, responsables par exemple des effets de restauration (2.1.3.6 Intégration perceptive), nous avons montré que le choix du matériel linguistique et la façon dont il est élaboré et généré puis choisi par l'évaluateur

est crucial en ce qu'il impacte directement l'évaluation. Une nouvelle tâche d'évaluation de l'intelligibilité basée sur des pseudo-mots a été élaborée dans le cadre du projet de recherche C2SI pour répondre à cette problématique. Le matériel linguistique utilisé présente l'avantage d'être parfaitement contrôlé en termes d'occurrence, de fréquence et de position des phonèmes dans les listes générées de manière à obtenir un équilibre. L'emploi de formes linguistiques non lexicales permet de contraindre l'auditeur à considérer majoritairement les unités de bas niveaux que constituent les sons de parole, et accomplir une tâche de décodage acousti-phonétique (DAP). Ce principe offre la possibilité de réaliser une évaluation perceptive qui soit pertinente en termes de déficit segmental, adaptée à une évaluation de l'intelligibilité telle que définie dans ce travail. L'idée selon laquelle les plus petites unités de sons, les phonèmes, constituent la base pyramidale de la parole (comme le suggèrent les modèles, cf parties 2.1.3.2 et 2.1.3.5) laisse envisager que la réparation de la production de ces unités puisse contribuer à la réparation d'unités plus larges et à la prise en charge adaptée du handicap communicationnel des patients traités pour un cancer ORL.

La production de la parole est un phénomène dynamique par essence. Nous estimons que son évaluation au moyen de représentations phonologiques statiques, reste opérationnelle si l'on considère comme non hermétique la « frontière » entre phonétique et phonologie. L'évaluation de l'intelligibilité segmentale, au moyen d'un score considérant les distances inter-phonémiques entre les unités produites et les unités perçues, constitue un outil pertinent pour le clinicien afin de mesurer les déficits fonctionnels de parole et d'orienter la rééducation de la parole.

Le premier axe de cette thèse s'est attaché à valider cette pertinence de manière expérimentale, au travers d'une approche perceptive par décodage acoustico-phonétique. Une série d'expérience de perception de la parole ont permis de questionner la nature du matériau linguistique pour neutraliser le recours des auditeurs aux informations de haut niveau.

Dans une première expérience nous avons testé la pertinence de l'emploi de pseudo-mots, par rapport à des mots, pour une évaluation segmentale de l'intelligibilité. Il s'agissait de comparer les évaluations perceptives basées sur des mots versus des pseudo-mots produits par les locuteurs Patient et Contrôles. Les résultats ont montré que les évaluations basées sur les mots (donc un matériau linguistique lexical qui permet le recours aux informations Top-down) donnent des scores instables, qui évoluent à la baisse au fur et à mesure de l'exposition des auditeurs aux mots. Plus précisément, cela signifie qu'un locuteur voit son score d'intelligibilité s'améliorer au cours de la tâche simplement parce que l'auditeur reconnaît un élément lexical de sa langue. Pour autant, la production du locuteur qui est présentée est toujours la même. Au contraire, les évaluations basées sur les pseudo-mots sont stables dans le temps ce qui signifie qu'il n'y a pas d'apprentissage. Évalué sur la base de productions de pseudo-mots, le locuteur présente toujours le même score d'intelligibilité segmentale. Ce résultat montre un avantage de l'évaluation par pseudo-mots par rapport aux mots. Notre interprétation est que le recours aux pseudo-mots permet de contrôler les effets de restauration lexicale, en limitant le recours au lexique et donne ainsi une mesure plus fiable. Le pseudo-mot est

produit comme une unité isolée, sans lien avec d'autres éléments linguistiques. L'auditeur ne peut s'appuyer sur ces informations pour décoder le signal acoustique. Il est contraint d'utiliser les seules informations dont il dispose, à savoir le signal acoustique du pseudo-mot produit par un locuteur, ainsi que ses connaissances sur les unités de bas niveau et leur agencement (une séquence /str/ existe en français dans le mot « estrade » par exemple. Cet enchaînement de phonèmes appartient au champ des possibles pour un auditeur, en revanche la séquence /sbt/ est inexistante dans la langue et inconnue de l'auditeur, elle ne constitue donc pas un agencement phonémique envisageable). Certaines réponses des auditeurs sont lexicalisées. Dans une précédente étude (Rebourg, 2018) nous avons examiné les réponses des auditeurs à un test de jugement perceptif de l'intelligibilité par pseudo-mots. Nous avons montré, sur un total de 19354 réponses que 6% étaient des réponses lexicalisées. Nous avons conclu à un phénomène d'attractivité lexicale inévitable, mais qui n'impacte pas l'évaluation. Restauration lexicale et attraction lexicale sont deux phénomènes différents qui n'impactent pas l'évaluation de la même manière. Ces deux phénomènes entraînent une réponse lexicale de l'auditeur. Mais, dans un cas il perçoit un mot et reconstruit une information sémantique à partir de ses connaissances lexicales, dans l'autre il perçoit un pseudo-mot et est perceptivement attiré par une cible connue et stockée dans son lexique mental. Même si une part d'attraction vers le lexique demeure celle-ci ne perturbe pas l'évaluation puisque ce qui a été perçu par l'auditeur-évaluateur diverge nécessairement de la cible initiale qui n'est pas lexicale. En revanche les effets de restauration lexicale sont susceptibles de perturber l'évaluation. Une méthode d'évaluation, qui permet de contourner ou neutraliser ce mécanisme, comme la méthode par DAP offre une évaluation perceptive plus robuste car moins bruitée. Considérant ce premier résultat qui confirme la stabilité des évaluations, et son pouvoir discriminant, cette évaluation basée sur des pseudo-mots présente un réel intérêt clinique : l'intelligibilité peut être plus précisément évaluée en bilan initial et peut s'envisager dans le cadre d'évaluations de suivi, à différents moments du parcours de soin.

Cependant, au terme de cette première expérience, un niveau questionnement a émergé du design expérimental : la stabilité des évaluations par pseudo-mots relève-t-elle de la qualité du matériel linguistique utilisé ou de sa quantité. En effet, dans notre design, les auditeurs écoutaient toujours les mêmes 50 mots de la BECD mais les listes de pseudo-mots étaient toujours différentes (n=1000). Dans une tâche classique d'évaluation clinique l'examen est composé de listes de mots courtes et fermées, la répétition du test par le clinicien constitue un entraînement qui lui permet de mémoriser les items. L'évaluation des déficits est perturbée par les effets d'apprentissage du matériel linguistique, relevant d'intégration d'informations de hauts niveaux linguistiques. Le principe du décodage acoustico-phonétique repose sur un très grand nombre de pseudo-mots qui permettent de générer un très grand nombre de listes variées et équilibrées. Nous avons donc conduit une seconde expérience visant déterminer si la qualité (mots vs. pseudo-mots) ou la quantité (nombre) avait influencé la neutralisation de l'effet d'apprentissage du matériel linguistique lexical relevée dans notre première expérience (le score moyen baisse au cours de l'exposition). Au moyen d'un nouveau design expérimental basé sur la production d'une seule et unique liste de pseudo-mots par des

locuteurs patients et contrôles, nous avons comparé l'évolution des scores moyens au cours de l'exposition à autant de mots que de pseudo-mots (n = 50). Nos résultats ont confirmé l'effet d'apprentissage induit par l'exposition aux items lexicaux qui entraînent le recours aux informations de haut niveau. Ils ont également mis en lumière un effet d'apprentissage des pseudo-mots lorsque les mêmes stimuli sont répétés. Autrement dit, ces résultats indiquent que c'est la quantité de matériel linguistique, en plus de sa qualité, qui permet de neutraliser les effets d'apprentissage du matériel linguistique au cours de la répétition d'une évaluation perceptive. Les mécanismes de perception de la parole montrent à nouveau leur efficacité, la répétition d'une tâche d'écoute entraîne l'auditeur à apprendre des items linguistiques non lexicaux qu'il est capable de stocker même en l'absence de signification. Il améliore ses performances de décodage en ayant recours à de nouvelles informations de haut niveau. Ces observations rejoignent les études qui suggèrent que les connaissances linguistiques affectent la perception de la parole.

Aux termes de ces deux expériences, nous avons montré que la quantité de matériel linguistique disponible pour une évaluation perceptive de l'intelligibilité segmentale permet de neutraliser les effets d'apprentissage de celui-ci. Et que les pseudo-mots constituent des unités linguistiques permettant de neutraliser les effets de restauration lexicale (observé avec des mots). En d'autres termes, les pseudo-mots constituent des unités pertinentes pour évaluer l'intelligibilité segmentale, mais ceux-ci doivent être disponible en grand nombre pour que la neutralisation des effets d'apprentissage soit également possible. Ces observations rejoignent celles de [Grataloup et al. \(2009\)](#) selon lesquelles les pseudo-mots sont plus difficiles à reconstruire que des mots puisqu'ils n'ont jamais été entendu auparavant et ne sont pas stockés en tant que représentation lexicale. Nos résultats suggèrent qu'ils peuvent être appris et s'ajouter aux connaissances de haut niveau linguistique (lexical) d'un auditeur. De plus, un effet d'attraction lexicale demeure comme évoqué précédemment.

L'ensemble des résultats disponibles montrent la pertinence des pseudo-mots pour réaliser une évaluation de l'intelligibilité segmentale par la méthode de décodage acoustico-phonétique. Ce test a été élaboré pour s'adapter aux contraintes de l'évaluation perceptive clinique. Il se destine donc aux professionnels de la prise en charge de la parole pathologique, les cliniciens (ORL, phoniâtres, orthophonistes). Il est donc apparu essentiel d'évaluer cette méthode au regard de la qualité de l'auditeur. Pour cela nous avons conduit une troisième étude perceptive auprès de deux groupes d'auditeurs, naïfs et experts. Celle-ci était destinée à évaluer la pertinence de l'utilisation de pseudo-mots pour neutraliser le recours aux informations de haut niveau et a concerné l'expertise auditive des auditeurs. En effet, nous avons abordé le recours aux informations de haut niveau contextuelles des auditeurs lorsqu'ils perçoivent de la parole. Parmi ces informations disponibles, certaines sont spécifiques au cadre clinique. Le clinicien évaluateur a connaissance du patient et de son dossier médical. Sa pathologie (cancers ORL), la localisation, le stade et la taille de la tumeur, ou encore du traitement dont il a bénéficié (chirurgies, radiothérapie, chimiothérapie) représentent des informations qu'il mobilise dans son évaluation. Ces informations de haut niveau clinique lui confèrent la capacité d'inférer les déficits articulatoires des patients. Par là même il se montre capable

de restaurer une information acoustiquement dégradée. Par exemple, dans le cas d'une tumeur du voile du palais traitée par chirurgie, le clinicien anticipe une perturbation de la fermeture vélo pharyngée qui va impacter la production de la parole en termes de nasalité. Les effets produits sur le lexique le sont également. Si le patient produit les mots « honneur » et « amuser » le clinicien se montre capable de restaurer les mots « odeur » et « abuser » présents dans le matériel linguistique de l'évaluation en tant que cible. Pour cela il mobilise deux types d'informations de hauts niveaux : linguistique (lexicale) et clinique (métadonnées).

Ainsi, les auditeurs experts, en confrontant leurs connaissances linguistiques (phonologiques, physiologiques ...) aux métadonnées cliniques du patient, se montrent capables de prédire l'altération du trouble de parole. Cet effet connu d'intégration des informations de haut niveau clinique n'est pas manipulé dans notre étude puisqu'aucune métadonnées cliniques n'a été communiquée aux auditeurs. Dans cette étude, il s'agissait d'évaluer la qualité d'expertise auditive. Les 18 auditeurs experts recrutés pour cette étude étaient tous orthophonistes. Ces spécialistes de la parole l'évaluent quotidiennement et ont développé des facultés de restauration phonémique de la parole altérée a priori plus importantes que les auditeurs naïfs. Ils bénéficient de leur expérience dans l'écoute et le décodage de parole pathologique. Afin de mesurer l'impact de cet effet d'expertise auditif sur l'évaluation perceptive nous avons comparé l'évolution des scores moyens au cours de l'exposition aux pseudo-mots de deux groupes d'auditeurs. Nos résultats ont montré qu'au regard de leur qualité, les auditeurs proposent des évaluations dont l'évolution des scores moyens entre les groupes d'auditeurs est comparable. Cependant, celles-ci ne sont pas totalement équivalentes puisqu'elles présentent une différence moyenne de 0,1 trait d'écart moyen par phonème. L'intelligibilité des locuteurs est jugée plus dégradée par les auditeurs naïfs que par les experts. La dimension de cet effet différentiel est faible. Les scores moyens restent stables au cours de l'exposition pour les deux groupes d'auditeurs. Cela signifie que les pseudo-mots permettent de neutraliser le recours aux informations de haut niveau linguistique quand ils sont présentés en très grand nombre. Toutefois, les différences de scores moyens entre les auditeurs naïfs et experts montrent que les auditeurs experts sont de meilleurs décodeurs des paramètres acoustico-phonétique de la parole altérée, que les auditeurs naïfs. Cela signifie que les auditeurs experts conservent leur qualité d'expertise auditive dans une évaluation perceptive. Cette différence ne perturbe pas l'évaluation, les effets d'apprentissage et de restauration lexicale demeurent neutralisés dans les deux groupes d'auditeurs.

Dans cette étude, les effets de restauration de haut niveau clinique sont en partie neutralisés puisqu'aucune donnée clinique n'a été communiquée aux auditeurs. L'effet d'expertise auditif ici testé n'est pas totalement neutralisé mais ne perturbe pas l'évaluation de l'intelligibilité segmentale par les cliniciens. Leurs capacités d'écoute et de décodage de la parole pathologique suggèrent qu'en leur qualité d'experts, ils retrouvent probablement des informations manquantes du signal. Grâce à leur expertise d'écoute, ils ont peut-être une capacité plus accrue à reconstruire des unités de bas niveau à partir d'indices très fins, peut-être de coarticulation. Cela peut être associé aux théories linguistiques selon lesquelles un auditeur se montre capable de restaurer de unités phonémiques.

Par ailleurs, la faible différence entre les deux groupes d'auditeurs questionne la capacité d'un auditeur à devenir expert. En effet, l'entraînement au décodage des unités phonémiques par l'exposition à la parole pathologique pourrait-il permettre à un auditeur d'accroître sa capacité à décoder des unités phonologiques de bas niveau ? Il serait ainsi, envisageable que des personnes impliquées dans un environnement linguistiquement pathologique (aidants, famille, infirmières, aides-soignantes) puissent présenter des facultés d'expertises auditives proches/équivalentes de celles des auditeurs experts.

Aux termes de ces trois expériences en perception de la parole nous avons confirmé que les pseudo-mots constituent un matériau linguistique pertinent pour neutraliser les effets de restauration lexicale. Disponibles en grande quantité, ils permettent de neutraliser l'impact des effets d'apprentissage du matériel linguistique au cours d'une évaluation. Enfin, les effets d'expertise auditive des cliniciens ne peuvent être neutralisés, au regard de leur compétences perceptives. Cependant, ils ne perturbent pas l'évaluation si les autres biais perceptifs sont contrôlés (restauration et apprentissage). De plus, dans chacune des expériences les pseudo-mots se sont montrés capables de distinguer deux groupes de locuteurs (patient vs contrôle) ce qui suggère qu'une évaluation perceptive par une méthode de décodage acoustico-phonétique basée sur l'utilisation de pseudo-mots et un calcul de scores analytiques permettant d'exprimer la dégradation de l'intelligibilité segmentale en nombre moyen de traits altérés par phonème est pertinente pour une évaluation en contexte clinique.

Ces résultats ont été obtenus par le calcul des scores de Déviation Phonologique Perçue représentant un score de distance inter-phonémique cumulée en nombre de traits moyen altérés par phonèmes. Celui-ci permet de calculer les scores d'intelligibilité segmentale. Au moyen d'une matrice de coût segmental basée sur la théorie des traits articulatoires et d'un algorithme de calcul de distance d'édition entre deux chaînes de caractères, la distance phonémique entre une cible phonétique qui devait être produite et une cible phonologique qui a été perçue a été établie. Ces précédents résultats portent sur les scores de PPD globaux, autrement dit sur un score de distance cumulée (entre les unités phonémiques et entre les différents stimuli), c'est un score perceptif global par locuteur issu des transcriptions d'auditeurs.

Afin de mesurer / d'évaluer encore plus directement le niveau segmental et d'affiner davantage l'analyse des segments phonémiques, la méthode de calcul a été décomposée. Elle permet désormais d'obtenir des scores dits locaux, soit une distance phonème à phonème. Deux informations ont fait l'objet d'une analyse, d'une part les scores PPD locaux (nombre) et les transcriptions locales des auditeurs (phonème). Le calcul de scores de distance cumulés en fonction des segments cibles représente la moyenne des scores locaux pour les différents segments. Ainsi, nous avons comparé les scores de distance cumulées locales des segments vocaliques, consonantiques et des clusters. Les scores locaux permettent d'identifier les segments les plus dégradés. Nos résultats ont montré que les clusters sont moins bien identifiés par les auditeurs par rapport aux consonnes et aux voyelles. Ils ont montré que les voyelles sont moins bien reconnues et identifiées que les consonnes, mais que cette différence de score moyen est imputable aux voyelles

nasales qui semblent plus difficiles à identifier que les voyelles orales. Ainsi, les scores moyens des voyelles orales et des consonnes sont équivalents. Ces résultats questionnent le choix/la pertinence d'une unité phonémique plus ou moins robuste face à l'altération, pour une évaluation segmentale de l'intelligibilité.

Il apparaît que chaque élément constitutif de la parole présente un intérêt pour une évaluation globale du déficit linguistique. La question se pose de savoir, dans notre cadre clinique, quelles sont les unités qui constituent les meilleurs candidats pour l'évaluation ? En d'autres termes, quel degré de robustesse d'un segment capture le mieux les dégradations fines de l'altération de la parole reflétant le plus fidèlement le handicap communicationnel. Nous pouvons facilement inférer que les unités les plus robustes chez le groupe contrôle, si elles conservent ces propriétés dans le groupe patient, constitueraient des unités pertinentes aptes à évaluer la parole pathologique en contexte clinique. Cependant, l'étude de Marczyk et al. (2020) sur la réduction du nombre d'items dans les listes montre que les clusters sont des unités particulièrement pertinentes et à conserver pour réduire la liste de pseudo-mots, de 50 à 16 items produits, tout en proposant des scores de déviation phonologique perçue globaux équivalents. Ces résultats suggèrent que les structures phonémiques complexes comme les clusters sont aptes à refléter des dégradations fines du signal acoustique. En effet, ils permettent probablement d'intégrer plus d'éléments à l'évaluation et de capturer/représenter des effets inhérents à la production de la parole comme la coarticulation. Cette étude sur les scores globaux suggère, au regard de nos résultats sur les scores locaux, que les unités moins robustes et plus complexes se montrent plus sensibles aux variations pathologiques et sont donc à même de capturer des altérations fines. Ces structures complexes ont montré leur pertinence quand elles sont considérées dans un score global de distances cumulées (A. K. Marczyk et al., 2020). Mais leur contribution dans une considération locale n'a pas été évaluée et demeure une question d'intérêt.

Ici, nous avons jugé pertinent d'évaluer d'abord les structures moins complexes et mieux décrites qui contribuent le plus à l'intelligibilité, au regard des éléments de littérature (Chen et al., 2013; Cole et al., 1996; Fogerty & Kewley-Port, 2009; Kewley-Port et al., 2007). Ceux-ci suggèrent que les voyelles contribuent davantage à l'intelligibilité (au sens large) que les consonnes et des mesures de la dimension du système vocalique sont régulièrement associées à des troubles de la production de la parole (Ishikawa et al., 2019; Kent & Kim, 2003; Weismer et al., 2001). L'approche perceptive par décodage acoustico-phonétique de pseudo-mots permet un examen approfondi des segments vocaliques à partir des scores locaux (exprimés en nombre moyen de traits altéré par voyelle) et des réponses (transcriptions) des auditeurs. A partir des scores locaux, nous avons pu établir une hiérarchie pour le degré d'altération des segments vocaliques. Les voyelles nasales sont nettement plus dégradées que les voyelles orales. Même si nous n'avons pas exploré en détails les voyelles nasales, leur dégradation significativement plus importante que leurs homologues orales nous oblige à évoquer les conceptions théoriques divergentes à propos de la nasalité. La conception classique de la nasalité considère qu'une voyelle orale ne se distingue de son homologue nasal qu'au travers du trait de nasalité associé au mouvement du voile du palais. Une nouvelle conception émerge d'études plus récentes (Delvaux, 2009) et conçoit que les voyelles nasales

revêtent finalement des caractéristiques articulatoires qui leurs sont propres, et ne peuvent être considérées comme un pur équivalent. Nous avons évoqué dans notre partie théorique que l'ouverture de la mandibule entraîne un recul sensible de l'axe d'antériorité vers l'arrière à mesure que l'aperture augmente, il se pourrait que la mobilisation du voile du palais impacte aussi la position d'autres articulateurs. Ceci pourrait expliquer l'importante différence entre les scores locaux observés pour les voyelles nasales et orales.

Parmi les voyelles orales, les voyelles postérieures sont plus dégradées que les antérieures et les voyelles hautes sont plus dégradées que les voyelles moyennes et basses. La voyelle haute et postérieure /u/ est la plus dégradée. Les mêmes effets sont observés dans les deux groupes de locuteurs mais ils sont accentués par la pathologie. Plus précisément, la hiérarchie observée pour chacun des deux groupes est la même mais les scores locaux pour les patients montrent une dégradation plus importante. Cela suggère que la production des segments vocaliques qui impliquent un faible degré d'aperture et une position linguale postérieure sont moins intelligibles de manière générale. Ceci fait écho aux travaux sur l'acquisition du langage qui font état de l'ordre d'acquisition (Buhr, 1980; Davis & MacNeilage, 1990; Gayraud & Kern, 2007; Lieberman, 1980) des phonèmes vocaliques et présentent les voyelles antérieures ouvertes d'âge d'acquisition plus précoce que les voyelles postérieures fermées ; l'altération pourrait suivre un ordre inverse. Chez les patients la dégradation est majorée par la pathologie. Une étude sur ce même corpus qui visait à mettre en lien les métadonnées cliniques (région et taille de tumeur) et les scores locaux pour les segments consonantiques et vocaliques avait conclu à ce même résultat non seulement pour les voyelles mais aussi pour les consonnes (en particulier la consonne occlusive [k]). L'interprétation qui avait été donnée était en lien avec la localisation de la tumeur et le type de traitement dont le patient avait bénéficié (Lalain et al., 2022).

Nous avons expliqué nos observations par un lien avec l'ordre d'acquisition des phonèmes d'une part et avec l'impact des traitements des cancers de la cavité buccale et de l'oropharynx. Nous l'avons souligné en première partie, l'intégrité du système physiologique de production constitue un fondamental pour la qualité de la production mais ce n'est pas la seule et unique condition. Chaque segment semble se comporter de manière spécifique, peut-être en fonction de sa place dans l'acquisition du langage. Il est possible que les segments les plus tardivement acquis soient les moins intelligibles et les plus gravement dégradés en cas de troubles de la parole séquellaires d'un cancer ORL. Les séquelles fonctionnelles résultant du cancer et de ses traitements atteignent l'intégrité du système physiologique et la mobilité des structures anatomiques impliquées dans production de la parole. La perte de substance dans le cas de chirurgies ou la rigidification des tissus dans le cas de traitement par radiothérapie impactent la dynamique du geste articulatoire. Au regard de ces résultats, il nous semble qu'une approche perceptive par DAP pour évaluer l'intelligibilité segmentale est intéressante puisqu'elle permet outre le contrôle des biais perceptifs, une analyse fine des segments de parole et une caractérisation des segments les plus ou les moins intelligibles. Il nous paraît cependant important de souligner que l'identification de segments plus ou moins robustes à la dégradation n'est pertinente qu'en considérant ces segments au sein d'un système

linguistique donné. En effet, nous avons montré que les voyelles hautes et plus particulièrement la voyelle haute postérieure /u/ sont les plus dégradées mais ceci est en contradiction avec les observations de Ishikawa et al. (2019) qui montre au contraire que ce sont les voyelles basses qui sont les moins intelligibles en anglais. Il nous semble donc que les observations doivent être mises en perspective avec le système vocalique de la langue considérée : la densité du système vocalique sur les différents axes d'aperture et d'antériorité impacte probablement la distinction entre les catégories (cf théorie dispersion/focalisation). Plus un axe comprend de segments et moins ces segments sont focalisés. Le fait que les voyelles hautes soient plus nombreuses en français qu'en anglais et que les basses soient plus nombreuses en anglais qu'en français explique probablement cette différence de résultat avec notre étude.

Dans notre analyse des scores locaux nous avons également examiné les transcriptions des auditeurs. La correspondance entre le phonème ciblé et les phonèmes transcrits a permis d'associer une production phonologique avec ses transcriptions phonétiques perçues au niveau segmental. Au travers de cette analyse, nous avons tenté d'établir des constantes de substitution telles que les voyelles nasales sont plus souvent perçues comme le segment oralisé le plus proche (perte de nasalité) ; les voyelles fermées seraient perçues préférentiellement comme plus ouvertes ou encore les voyelles postérieures seraient perçues principalement comme plus antérieures. Nous avons exploré ici l'idée selon laquelle il est possible de caractériser la dynamique perceptive sur les axes antéro-postérieur et/ou de hauteur ainsi qu'en termes d'arrondissement et/ou de nasalité.

Nos résultats ont montré que les auditeurs produisent sensiblement les mêmes erreurs entre les deux groupes de locuteurs. En moyenne, les voyelles cardinales (/a/, /i/, /u/) présentent plus d'écart avec leurs transcriptions par rapport aux autres voyelles. Les voyelles moyennes (/e/ et /o/) présentent le moins d'écart moyen avec leurs transcriptions. La majorité des segments transcrits sont peu éloignés, en termes de traits articulatoires, de leur cible initiale. Ces résultats ont été confirmés par les analyses des scores locaux et par les représentations visuelles. Cela signifie que les auditeurs transcrivent préférentiellement des segments peu éloignés, en moyenne un trait d'écart seulement par rapport à la transcriptions attendue. Cet effet s'observe chez les deux groupes de locuteurs mais la diversité des réponses dans le groupe patient est plus importante par rapport au groupe contrôle, cet effet est majoré pour les patients dont la production est la plus altérée. En d'autres termes, l'auditeur transcrit au mieux ce qu'il perçoit et en général cette cible perçue est proche de la cible attendue. Néanmoins, l'éventail de choix est plus large et la divergence de nombre de traits articulatoires plus importante relativement au degré d'altération de la production. Enfin, l'examen des transcriptions des auditeurs a montré que pour tous les segments substitués, l'archiphonème /E/ constitue une réponse récurrente. Les segments altérés tendent à être produits de manière plus centrale. D'un point de vue articulatoire, leur réalisation se rapproche d'une position neutre des articulateurs. Ceci, nous l'avons vu, fait écho aux travaux qui montrent une centralisation du système dans les troubles de la parole.

Ces résultats sur les segments substitués montrent par ailleurs que les auditeurs ne proposent pas de transcriptions de phonèmes étrangers à leur système vocalique. Au

même titre que le locuteur qui est contraint pas son système physiologique, l'auditeur est contraint par le système phonologique de sa langue. Il n'a à sa disposition qu'un nombre limité d'unité phonémique dans son répertoire pour représenter sa perception. Cela indique que les connaissances linguistiques influencent la perception de la parole et suggère que l'auditeur est probablement influencé par le haut niveau. De plus, un effet inhérent à la production de la parole n'est pas capturé dans ces analyses segmentales, la coarticulation. Il est pourtant envisageable que l'auditeur s'appuie sur ces éléments pour identifier la cible attendue même si celle-ci est mal produite, puisque les caractéristiques acoustiques d'un segment se répercutent dans les segments adjacents. L'absence de prise en compte des effets de coarticulations dans la méthode de calcul des scores constitue probablement une limite à cette approche.

D'autres part, Chomsky & Halle (1968) soulignent qu'une transcription phonétique « représente l'interprétation du locuteur-auditeur plutôt que des propriétés directement observables du signal » en ce sens, « l'existence de certains décalages entre la transcription et le signal peut être comprise ». L'évaluation perceptive au moyen du système de perception humaine est imparfaite mais constitue le *gold standard* (Balaguer, Boisguérin, et al., 2019). La méthode d'évaluation par décodage acoustico-phonétique basée sur des pseudo-mots a montré son efficacité et permet d'améliorer l'évaluation perceptive.

Le second axe de cette thèse s'est attaché à évaluer s'il était pertinent et envisageable de préciser cette méthode au travers d'une approche par mesures acoustiques. Celles-ci revêtent un caractère particulièrement intéressant puisqu'elles permettent d'extraire des mesures directes sur le signal acoustique, au moyen d'une méthode automatique standardisée. Si l'on considère les liens au moins supposés entre production et perception, la production de segments identifiés comme peu intelligibles s'accompagne probablement d'une signature acoustique qui ne permet pas d'identifier perceptivement le segment cible. Ces mesures spectrales participent au calcul de métriques représentatives des dimensions du système vocalique et de la dynamique des segments qui le constitue. Par ailleurs, elles possèdent un pouvoir de représentation de ce système dans un espace bidimensionnel le figurant.

Dans ces travaux nous avons choisi les métriques de l'aire du triangle vocalique (Vowel Space Area, VSA) qui renseigne sur la taille générale du système vocalique d'un locuteur, ainsi que l'indice de distinctivité (ID) qui propose un indice de distinction entre les catégories de segments vocaliques. Les éléments de la littérature indiquent que l'aire du triangle vocalique est comprimée dans les pathologies d'origine dysarthrique. Cette observation devrait pouvoir être étendue aux troubles de la parole séquellaires d'un cancer traité de la CB ou de l'OP. Selon nous, la distinction entre les catégories de segments du système devrait s'en trouver impactée. Nous avons exploré leur capacité à distinguer les deux groupes de locuteurs, ainsi que leurs corrélations avec les scores perceptifs locaux et globaux.

Nos résultats ont montré que l'aire du triangle vocalique ne permet pas de distinguer les deux groupes de locuteurs, et ne montre pas de lien avec les scores perceptifs. L'aire de l'espace vocalique s'était montrée une mesure pertinente pour distinguer des

locuteurs sains de locuteurs dysarthriques (Weismer et al., 1988, 2001). Plus encore, cette mesure permettait de distinguer 2 types de dysarthrie. Nous ne retrouvons pas cette différence dans notre corpus. Il est possible que la différence de langue (anglais vs français) et donc du système phonologique vocalique entre notre corpus et celui de Weismer et collègues explique cette différence de résultat. Il est également possible que la taille de la cohorte de locuteurs impacte les résultats. Dans l'étude de Weismer et al. (1988), les trois groupes comprennent 39 sujets (19 sujets sains, 10 SLA et 10 Parkinson). Notre corpus comprend les productions de 127 locuteurs (40 contrôles et 87 patients). Il s'agit d'une cohorte bien plus importante et il n'est pas exclu qu'un grand groupe de locuteurs se caractérise par une plus grande hétérogénéité. Une solution à envisager serait de subdiviser notre groupe de patients selon des critères cliniques par exemple (taille ou localisation de la tumeur, délai post traitement...) ou démographiques (âge, sexe).

En revanche, l'indice de distinctivité permet de distinguer les deux groupes de locuteurs. Le groupe de locuteurs contrôles présente un indice plus élevé que dans le groupe patient. Cela signifie que les catégories de voyelles présentent plus de distinction les unes envers les autres chez les contrôles par rapport aux patients. Ces résultats suggèrent qu'au regard des scores PPD, un ID élevé devrait être associé avec un sPPD bas reflétant peu d'altération, contrairement à un ID bas qui devrait davantage être associé avec un sPPD élevé. Néanmoins, les tests de corrélation ont montré que l'ID corrèle avec les scores perceptifs locaux et globaux dans le groupe patient mais pas dans le groupe contrôle. Cela suggère qu'un ID bas reflète d'avantage un score PPD élevé et se révèle pertinent pour caractériser les patients, mais pas les contrôles (au regard des scores perceptifs). Ainsi, le lien entre l'intelligibilité segmentale et l'ID est modéré puisque dépendant du groupe de locuteur. Cela suggère que les scores d'intelligibilité segmentale bas (groupe contrôle) n'impliquent pas forcément une forte distinction entre les catégories, en revanche, un score d'intelligibilité segmentale élevé (groupe patient) montre un ID plus élevé reflétant moins de distinction entre les catégories de voyelles. De plus, les premiers résultats concernant la distinction entre les deux groupes de locuteurs, ont été confirmés par les visualisations. En effet, on observe davantage de chevauchement entre les catégories dans le groupe patient par rapport au groupe contrôle dans lequel elles présentent plus de distinction.

Nos résultats montrent que ces mesures représentatives du système vocalique dans sa globalité ou plus spécifiquement des catégories de voyelles, ne permettent pas de prédire le score perceptif, en revanche, l'indice de distinctivité constitue une métrique complémentaire susceptible d'expliquer une part de la variabilité des scores perceptifs du groupe patient. Les mesures de VSA et d'ID demeurent indicatives et pertinentes pour caractériser le système vocalique des locuteurs. Il s'agissait ici d'une démarche expérimentale visant à évaluer leur relation avec les scores perceptifs. Celle-ci s'est révélée ténue.

La représentation de l'espace vocalique au moyen de métriques permettant leur visualisation s'est montrée pertinente pour caractériser les systèmes vocaliques des locuteurs et fait écho, dans une certaine mesure, aux résultats observés au cours de

l'analyse des transcriptions proposées par les auditeurs. La comparaison de ceux-ci entre chaque groupe de locuteur patient et contrôles, en considérant leur sexe, ont montré des caractéristiques qui semblent permettre de les distinguer. Ces visualisations constituent un outil pertinent pour obtenir un nouveau regard sur les productions des locuteurs au niveau segmental. Elles ont permis de confirmer que les systèmes vocaliques des hommes et des femmes présentent des dimensions sensiblement différentes, en moyenne, celui des femmes est plus grand que celui des hommes. Cet effet est maintenu dans le groupe patient. Ils ont également montré que la distinction entre les catégories de segments dans le système vocalique des femmes semble davantage impactée par rapport aux hommes dans le groupe de patient, les segments /i/, /y/ et /E/ présentent davantage de recouvrement. Alors que les femmes du groupe contrôle présentent davantage de distinction par rapport aux hommes. Cela suggère que les caractéristiques acoustiques de la parole des femmes sont plus sensibles aux séquelles fonctionnelles. Ces spécificités doivent être considérées en contexte clinique.

D'autre part, d'un point de vue clinique, il est envisageable de considérer qu'une mesure de l'intelligibilité au moyen d'une évaluation perceptive, complétée par une mesure acoustique automatique, sont pertinentes pour évaluer l'intelligibilité segmentale de la parole d'un patient.

Enfin, plusieurs limites ont émergé au fil de ces travaux, elles s'apparentent à des aspects techniques et méthodologiques. Les premières concernent le corpus C2SI, objet des analyses des scores locaux et des transcriptions des auditeurs, en lui-même. En effet, la cohorte de locuteurs patient est hétérogène au regard des métadonnées cliniques. Tous les stades T et N sont représentés, les combinaisons de traitements dont les patients ont bénéficié sont variées et les localisations sont nombreuses. De plus, il est délicat d'établir un lien franc entre la localisation de la tumeur et la dégradation. La localisation relevée pour être intégrées dans les métadonnées requiert un examen du site lésionnel, nécessairement consécutif au traitement. L'étendue de cette localisation est différente du site lésionnel initial et ne peut être précisée. Celles-ci pourraient pourtant présenter des spécificités qui leurs sont propres notamment en fonction de la taille de la tumeur réséquée. En effet, une tumeur de la langue ou une tumeur du voile (traitée par chirurgie, traitement localisé) n'a probablement pas les mêmes conséquences sur la production du locuteur, la première affecte d'abord la mobilité de la langue sur l'axe antéro-postérieur, alors que la seconde perturbe d'abord l'oralité en ne permettant pas de séparer totalement les cavités nasales et buccale et entraînant un phénomène de nasalisation. De plus, les mêmes tumeurs traitées par radiothérapie par exemple ne devraient présenter totalement les mêmes conséquences sur la production de la parole. Le maintien de la structure existante est probablement bénéfique. Néanmoins le traitement radiothérapique impacte l'ensemble des tissus mous qui constituent le système physiologique de production de la parole. La perte de souplesse des tissus a des conséquences importantes sur la mobilité et la vitesse de mouvement des articulateurs. Au regard de l'hétérogénéité du corpus, les conséquences des tailles, traitements et localisations sur la production de la parole ne peuvent être précisées individuellement.

De plus, le corpus tel que constitué ne permet pas non plus d'établir des sous-groupes homogènes en ces termes. Cela compromet les perspectives d'établir un index précis. Néanmoins ce corpus reflète la diversité des pathologies et l'individualisation du soin proposé au patient par le corps médical en charge. Il permet d'éclairer les tendances par groupe de locuteur.

Les secondes limites relevées concernent les mesures acoustiques au regard de la qualité des enregistrements et de l'alignement segmental, ainsi que de leurs conséquences sur la détection des formants. Tout d'abord, les enregistrements présentent des différences de gain d'entrée. Si celui-ci n'est pas contrôlé, le signal acoustique peut être saturé ou au contraire manquer d'énergie et d'amplitude spectrale. Ces phénomènes liés aux conditions des enregistrements (à l'hôpital au cours d'une consultation pour les patients) constituent des limites de terrain. De plus, les signaux acoustiques bruts des locuteurs employés pour les mesures acoustiques n'ont pas été normalisés, contrairement aux stimuli préparés dans le cadre des tests de perception. La normalisation permet, entre autres, de limiter les effets liés aux différences de qualité des enregistrements. Son absence est susceptible de perturber les mesures automatiques. En effet, l'analyse automatique de la parole saine constitue déjà un défi en elle-même. Sa grande variabilité explique pour parti ces difficultés. Par exemple, les différentes méthodes automatiques de détection et de calcul des formants font l'objet de nombreux travaux conséquents, elles sont comparées et évaluées (Flanagan, 1956; Jemaa, 2013; Nearey et al., 2002). De plus les portions de signal et les différents points retenus sur celui-ci pour calculer une valeur moyenne varient d'une étude à l'autre. Le manque de standardisation des procédures tient principalement au fait que la recherche expérimentale suit son cours. Les travaux sur les formants et leurs détections montrent qu'ils sont sensibles aux différences anatomiques du conduit vocal liées aux genres et à l'âge. Cela suggère que ces difficultés sont majorées lorsqu'il s'agit de productions pathologiques. Nos travaux ont montré que les productions altérées des patients représentent un niveau de complexité supérieur pour les systèmes automatiques puisqu'ils présentent davantage encore de variabilité. Par exemple, l'aligneur s'est révélé sensible aux pauses intra-pseudo-mots produites par certains patients. Ceci à entrainer un décalage dans l'alignement automatique qui délimite incorrectement les positions des frontières dans l'annotation segmentale. Les valeurs de formants relevées sur ces signaux sur la base des annotations ont présenté des valeurs hors du cadre de référence. Ce même cadre peut être questionné. En effet, les études qui s'appliquent à établir un référentiel pour les valeurs de formants se basent principalement sur de la parole saine. Si celui-ci est employé comme un filtre pour trier des données, ces limites sont susceptibles d'occulter certains effets d'intérêts liés aux caractéristiques de cette parole pathologique. L'absence de référentiels fiables pour celle-ci et les limites qu'elle impose aux systèmes automatiques constituent une limite importante. La mise à l'épreuve des systèmes automatiques sur différents types de parole contribue à leur perfectionnement et l'évaluation de leur capacité à distinguer des groupes de locuteurs (patients versus contrôle, natif versus langue seconde, ...) pourrait participer à leur amélioration.

Enfin, la dernière limite dont il est ici question concerne la méthode de calcul des scores de Déviation Phonologique Perçue, au moyen d'une matrice de coût basée sur la théorie des traits articulatoires et de l'algorithme de Wagner Fischer. Tout d'abord, nous l'avons évoqué, elle ne prend pas en considération les effets de coarticulation puisqu'elle considère les phonèmes individuellement. Les effets de restauration phonémique permis par la coarticulation ne sont pas capturés dans le calcul des scores perceptifs. De plus, la mesure des phénomènes de délétion, d'insertion et de substitution doit faire l'objet de nouvelles analyses pour être prises en compte. Des études de validation des listes à 16 mots devront également être envisagées. La composition des listes doit notamment être repensée pour que le passage à 16 items demeure équilibré. Une réflexion devra considérer les clusters en particulier, à conserver dans ces listes réduites. En effet, dans la version initiale, les clusters (CC) quand ils sont en position d'attaque de la seconde syllabe (CVCCV) peuvent refléter des structures homo (/fukli/) ou hétéro-syllabique (/kaspu/). La place des segments dans le pseudo-mots doit être mieux considérée pour affiner davantage la précision de ce score perceptif segmental.

Aux termes de cette discussion, les limites relatives aux effets de restauration lexicale et phonétique, des effets d'apprentissages, de la méthode de calcul des scores perceptifs, des aspects automatiques de l'alignement du signal acoustique avec sa transcription, de la sensibilité des systèmes de détection des formants, de la qualité des enregistrements, qui constituent des limites de terrain, méthodologiques et techniques ont été discutés au regard de nos résultats et des éléments de littérature.

Certaines de ces limites pourraient être repoussées : une analyse plus précise des similitudes entre les locuteurs contrôles et patients devraient être éclairantes sur les mécanismes de production de certains segments (par exemple les nasales), sur les mécanismes de production (examen des compensations) et sur les mécanismes de perception (examen des restaurations phonémiques). Dans le cadre clinique, nos analyses ici présentées pourraient être approfondies, par des études de cas par exemple, ou par l'examen des relations entre nos évaluations et les métadonnées cliniques des patients. Autant de nouvelles analyses que la richesse de ce corpus permet d'envisager.

Conclusion

« Evaluation de l'intelligibilité après un cancer ORL : approches par décodage acoustico-phonétique et par mesures acoustiques »

Aux termes de ces travaux, nous avons vu que l'imprécision de la définition du concept d'intelligibilité a des répercussions sur son évaluation. Nous avons souligné l'importance de la compréhension des mécanismes de production et de perception de la parole et du recours aux concepts théoriques de phonétique et phonologie pour envisager l'amélioration de l'évaluation de l'intelligibilité. Dans notre panorama théorique et clinique, nous avons proposé d'évaluer la pertinence d'une évaluation de l'intelligibilité segmentale, dans le cadre clinique de la cancérologie ORL, au moyen de deux approches : une approche perceptive par décodage acoustico-phonétique de pseudo-mots associée à une méthode de calcul de score de Déviation Phonologique Perçue et complétée par une approche par mesures acoustiques.

Nous nous sommes ainsi emparée d'une problématique ancienne : le concept de l'évaluation de l'intelligibilité. Nous avons envisagé, considéré et proposé différentes approches pour entreprendre et participer à l'amélioration de sa mesure. Nos résultats ont montré que les approches proposées permettent une évaluation plus fiable de l'intelligibilité au sens où nous la définissons. Parce qu'ils ont été considérés, les biais perceptifs de l'évaluation ont été contrôlés. L'expertise auditive clinique, identifiée comme un possible biais de l'évaluation, n'apparaît plus comme tel à la suite de ces travaux. L'approche perceptive par décodage acoustico-phonétique de pseudo mots s'est montrée robuste et a permis d'envisager non seulement une évaluation globale de l'intelligibilité mais aussi une identification précise des segments de parole les plus altérés. En complément de cette approche, les mesures acoustiques proposées ont fait partiellement écho aux évaluations perceptives. Les approches proposées dans cette thèse ont donc permis de résoudre au moins partiellement les problèmes identifiés comme potentielles sources des difficultés de l'évaluation clinique de l'intelligibilité et contribuent à son amélioration.

Ainsi, les approches proposées permettent d'envisager le développement d'un outil d'évaluation perceptive et acoustique, plus précis et plus fiable que les méthodes traditionnelles. Utilisé en bilan initial il pourrait permettre de mesurer l'évolution du déficit de parole au cours du suivi clinique du patient, et mesurer l'efficacité d'un traitement, d'une rééducation sur la parole par exemple. La méthode d'évaluation par décodage acoustico-phonétique de pseudo-mots, a montré sa pertinence et a été validée selon des critères scientifiques avec une méthodologie reproductible (basée sur des tests de perception). Les batteries de tests classiques n'ont pas fait l'objet d'autant d'études et d'analyses à propos du matériau et des niveaux linguistiques visés par l'évaluation,

pourtant la démarche expérimentale nous semble indispensable pour assurer la stabilité et la robustesse des tests proposés.

Même si les résultats sont encourageants l'approche perceptive que nous avons proposée mérite sans doute des améliorations, au regard de certains résultats obtenus et qui continuent de nous interroger. Par exemple, l'expertise auditive des cliniciens montre qu'ils sont de meilleurs décodeurs que les auditeurs naïfs, ce qui suggère qu'ils prennent en compte des indices acoustiques fins pour restaurer les segments dégradés et qu'ils sont capables de restauration phonémique de façon plus efficace. Ils pourraient par exemple se révéler plus sensibles aux indices de coarticulation que notre approche ne nous permet pas pour le moment de mesurer. La poursuite de ces travaux devra sans doute considérer l'évaluation de l'intelligibilité segmentale non pas comme l'évaluation de la production par la perception, mais plutôt comme l'évaluation d'un déficit articulatoire compensé par un système perceptif hautement adaptatif. En ce sens, ce n'est pas seulement une mesure de l'altération que vise l'évaluation de l'intelligibilité, mais également une mesure de l'adaptation du système de production. En effet, le patient lorsqu'il produit, s'adapte à la configuration et aux spécificités de son système de production puisque c'est le propre de la communication de concilier la variabilité de la production et l'adaptabilité de la perception de la parole, en vue de la réussite de l'échange communicationnel. L'approche par mesures acoustiques a quant à elle présenté des limites essentiellement liées aux difficultés méthodologiques rencontrées, mais permet, elle aussi, de projeter des applications cliniques.

Ces approches ne permettent pas d'apporter de solution à toutes les difficultés rencontrées. Certaines ont été identifiées et peuvent être contrôlées (biais perceptifs), d'autres n'ont pas été appréciées ici mais pourraient faire l'objet de recherches futures, puisque c'est le propre de la recherche que d'interroger. A date, nous pouvons tout de même envisager une application clinique de la méthode de DAP qui intègre les systèmes automatiques pour une évaluation en direct, c'est l'enjeu porté par d'autres projets de recherches en cours.

Enfin, les deux approches proposées se voulaient complémentaires mais nous n'avons pas pu observer de véritable correspondance entre les mesures proposées, les variables acoustiques (mesures indirectes) ici testées avec les scores perceptifs (mesures directes), ce qui révèle davantage encore la complexité de la relation non biunivoque entre production et perception de la parole. Probablement que la notion de mesure, en elle-même, constitue également une limite à l'évaluation de l'objet communicationnel complexe qu'est la parole, et n'est pas apte à refléter parfaitement la relation entre ce qui est produit *par le corps* et perçu *par les sens*.

Cette relation d'interdépendance forte entre production et perception s'est aussi imposée à nous au cours de la rédaction à travers le choix des termes à utiliser pour commenter certains résultats ; L'évaluation de la production se réalise au moyen de la perception, aussi, le score calculé est intrinsèquement dépendant du locuteur et de l'auditeur. Ils entretiennent une relation d'interdépendance, à l'image des mécanismes de perception et de production de la parole. Déterminer si c'est le segment qui est 'mal

produit' ou si celui-ci est 'mal perçu', renvoie aux fondements même de ce questionnement complexe.

Vous êtes sur le point d'achever la lecture de cette communication écrite. Nous espérons avoir su transmettre notre proposition argumentée d'une évaluation de l'intelligibilité en considération avec le niveau linguistique dont il est question dans celle-ci et applicable au contexte clinique, sans par ailleurs négliger ses limites. Bientôt, et si vous le permettez, nous défendrons ces travaux dans une communication orale dont l'enjeu de transmission, *par des sons articulés*, de notre pensée, sera tout aussi important que cette communication écrite. Dans l'attente des échanges communicationnels scientifiques qui animeront ce jour.

Bibliographie

Abitbol, J., & Vallancien, B. (s. d.). *VOIX, physiologie*. Encyclopædia Universalis. Consulté 18 octobre 2021, à l'adresse <http://www.universalis-edu.com.lama.univ-amu.fr/encyclopedie/voix-physiologie/>

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception : Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, 31(6), 265-272. <https://doi.org/10.1016/j.tins.2008.02.011>

André, C., Ghio, A., Cavé, C., & Teston, B. (2003). PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. *XVth ICPhS*, 1421-1424.

AP-HM - Hôpitaux Universitaires de Marseille. (2013, mars 25). *Comment soigner les cancers ORL ? - C la Santé*. <https://www.youtube.com/watch?v=45Enl2jK9Io>

Aristote. (1846). *Traité de l'âme*. Librairie Philosophique de Ladrance.

Astésano, C., Balaguer, M., Farinas, J., Fredouille, C., Gaillard, P., Ghio, A., Giusti, L., Laaridh, I., Lalain, M., Lepage, B., Nocaudie, O., Pinquier, J., Pont, O., Pouchoulin, G., Robert, D., Sicard, E., & Woisard, V. (2018). *Carcinologic Speech Severity Index Project : A Database of Speech Disorder Productions to Assess Quality of Life Related to Speech After Cancer*. 7.

Auzou, P., Ozsancak, C., Jan, M., Leonardon, S., Menard, J., Gaillard, M., Eustache, F., & Hannequin, D. (1998). Evaluation clinique de la dysarthrie : Présentation et validation d'une méthode. *Revue Neurologique*, 154(6-7), 523-530.

Auzou, P., & Rolland-Monnoury, V. (2006). *Batterie d'évaluation clinique de la dysarthrie* (Ortho Edition, Vol. 2). ORTHO.

Balaguer, M., Boisguérin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019). Assessment of impairment of intelligibility and of speech signal after oral cavity and oropharynx cancer. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 136(5), 355-359. <https://doi.org/10.1016/j.anorl.2019.05.012>

Balaguer, M., Boisguerin, A., Galtier, A., Puech, M., Farinas, J., Pinquier, J., & Woisard, V. (2019). *Facteurs influençant l'intelligibilité et la sévérité du trouble chronique de la parole des patients traités pour un cancer de la cavité buccale ou de l'oropharynx*. 23-24. <https://oatao.univ-toulouse.fr/24882/>

Barefoot, S. M., Bochner, J. H., Johnson, B. A., & Eigen, B. A. vom. (1993). Rating Deaf Speakers' Comprehensibility. *American Journal of Speech-Language Pathology*, 2(3), 31-35. <https://doi.org/10.1044/1058-0360.0203.31>

Bashford, J. A., Warren, R. M., & Brown, C. A. (1996). Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. *Perception & Psychophysics*, 58(3), 342-350. <https://doi.org/10.3758/BF03206810>

- Bechet, F. (2001). LIA PHON : Un système complet de phonétisation de textes. *Traitement Automatique des Langues, TAL*, 42(1), 47-67.
- Beck, J. M. (2005). Perceptual analysis of voice quality : The place of vocal profile analysis. In *A figure of speech : A festschrift for John Laver* (p. 286). Lawrence Erlbaum Associates.
- Blanc, E., Giusti, L., Ghio, A., & Serge, P. (2014). Adaptation en français du test d'intelligibilité de la version révisée du " Frenchay Dysarthria Assessment " (FDA-2). *Congrès de la Société Française de Phoniatrie*. <https://hal.archives-ouvertes.fr/hal-01615204>
- Boersma, P., & Weenink, D. (1992). *Praat : Doing phonetics by computer [Computer program]* (Version 6.1.16, 2020) [Audio]. University of Amsterdam. praat.org
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I : Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20(3-4), 255-272. [https://doi.org/10.1016/S0167-6393\(96\)00063-5](https://doi.org/10.1016/S0167-6393(96)00063-5)
- Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3, 219-252. <https://doi.org/10.1017/S0952675700000658>
- Buhr, R. D. (1980). The Emergence of Vowels in an Infant. *Journal of Speech, Language, and Hearing Research*, 23(1), 73-94. <https://doi.org/10.1044/jshr.2301.73>
- Calliope (Firm), T., J. P., & Fant, G. (1989). *La Parole et son traitement automatique*. Masson.
- Chen, F., Wong, L., & Wong, E. (2013). Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility. *The Journal of the Acoustical Society of America*, 134, EL178-84. <https://doi.org/10.1121/1.4812820>
- Chomsky, N. (2009). Syntactic Structures. In *Syntactic Structures*. De Gruyter Mouton. <https://doi.org/10.1515/9783110218329>
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.
- Clarke, P., Radford, K., Coffey, M., & Stewart, M. (2016). Speech and swallow rehabilitation in head and neck cancer : United Kingdom National Multidisciplinary Guidelines. *The Journal of Laryngology & Otology*, 130(S2), S176-S180. <https://doi.org/10.1017/S0022215116000608>
- Cole, R. A., Yonghong Yan, Mak, B., Fany, M., & Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, 2, 853-856. <https://doi.org/10.1109/ICASSP.1996.543255>
- Crochemore, E., & Vannier, F. (2001). Analyse phonétique de la parole dysarthrique. In *Les Dysarthries* (Auzou P., Özsancak C., Brun V. (Eds), p. 71-82). Masson.
- Cummins, N., Sethu, V., Epps, J., Schnieder, S., & Krajewski, J. (2015). Analysis of acoustic space variability in speech affected by depression. *Speech Communication*, 75, 27-49. <https://doi.org/10.1016/j.specom.2015.09.003>

Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of Correct Vowel Production. *Journal of Speech, Language, and Hearing Research*, 33(1), 16-27. <https://doi.org/10.1044/jshr.3301.16>

Décret n°2002-721 du 2 mai 2002 relatif aux actes professionnels et à l'exercice de la profession d'orthophoniste, 2002-721 (2002).

Définition classification TNM. (2019, février 21). <https://www.e-cancer.fr/Dictionnaire/C/classification-TNM>

Delvaux, V. (2009). Perception du contraste de nasalité vocalique en français. *Journal of French Language Studies*, 19(1), 25-59. <https://doi.org/10.1017/S0959269508003566>

Denes, P. B., & Pinson, E. N. (1993). The Speech Chain. In *The Physics and Biology of Spoken Language: Vol. 2nd edition* (p. 98-101). W.H. Freeman and Company. https://www.cambridge.org/core/product/identifier/S0025100300004904/type/journal_article

Données globales d'épidémiologie des cancers—Epidémiologie des cancers. (s. d.). Consulté 22 novembre 2021, à l'adresse <https://www.e-cancer.fr/Professionnels-de-sante/Les-chiffres-du-cancer-en-France/Epidemiologie-des-cancers/Donnees-globales>

Fang, C., Li, H., Ma, L., & Zhang, M. (2017). Intelligibility Evaluation of Pathological Speech through Multigranularity Feature Extraction and Optimization. *Computational and Mathematical Methods in Medicine*, 2017, 1-8. <https://doi.org/10.1155/2017/2431573>

Fant, G. (1970). *Acoustic Theory of Speech Production*. Walter de Gruyter.

Feng, Y., & Chen, F. (2022). Nonintrusive objective measurement of speech intelligibility : A review of methodology. *Biomedical Signal Processing and Control*, 71, 103204. <https://doi.org/10.1016/j.bspc.2021.103204>

Féry, C. (1999). Book review, Speech : A special code. *Journal of Pragmatics*, 31, 129-134.

Flanagan, J. L. (1956). Evaluation of Two Formant-Extracting Devices. *The Journal of the Acoustical Society of America*, 28(1), 118-125. <https://doi.org/10.1121/1.1908190>

Fletcher, H., & Steinberg, J. C. (1929). Articulation Testing Methods. *Bell System Technical Journal*, 8(4), 806-854. <https://doi.org/10.1002/j.1538-7305.1929.tb01246.x>

Fogerty, D., & Humes, L. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America*, 131, 1490-1501. <https://doi.org/10.1121/1.3676696>

Fogerty, D., & Kewley-Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *The Journal of the Acoustical Society of America*, 126, 847-857. <https://doi.org/10.1121/1.3159302>

Fontan, L. (2012). *De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication* [Thèse de doctorat - Linguistique, Université Toulouse 2 Le Mirail]. <https://tel.archives-ouvertes.fr/tel-00797883>

- Fu, Q. J., Zhu, M., & Wang, X. S. (2011). Development and validation of the Mandarin speech perception test. *Journal of the Acoustical Society of America*, 129, 267-273.
- Fujisaki, H., & Kawashima, T. (1970). On the Modes and Mechanisms of Perception of Speech Sounds. *The Journal of the Acoustical Society of America*, 47(1A), 57-57. <https://doi.org/10.1121/1.1974614>
- Ganong, W. F. (1980). Phonetic Categorization in Auditory Word Perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110-125.
- Garnier, M., Bailly, L., Dohen, M., Welby, P., & Loevenbruck, H. (2006). Étude acoustique et articulatoire de la parole Lombard. *XXVIemes Journees d'Etudes sur la Parole*, 077. <https://hal.archives-ouvertes.fr/hal-00214304>
- Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., & Dahlgren, N. L. (1993). *DARPA TIMIT: Acoustic-phonetic continuous speech corpus CD-ROM, NIST speech disc 1-1.1* (NIST IR 4930; p. NIST IR 4930). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.4930>
- Gayraud, F., & Kern, S. (2007). Phonological characteristics of target nouns as a function of age of acquisition. *Enfance*, 59(4), 324-338. <http://www.cairn.info/revue-enfance-2007-4-page-324.htm?contenu=auteurs>
- Gendrot, C., & Adda-Decker, M. (2005a). *Impact of Duration on F1/F2 Formant Values of Oral Vowels: An Automatic Analysis of Large Broadcast News Corpora in French and German*. 4.
- Gendrot, C., & Adda-Decker, M. (2005b). Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *Interspeech 2005*, 2453-2456. <https://halshs.archives-ouvertes.fr/halshs-00188096>
- Gentil, M. (1992). Phonetic intelligibility testing in dysarthria for the use of French language clinicians. *Clinical Linguistics & Phonetics*, 6(3), 179-189. <https://doi.org/10.3109/02699209208985529>
- Georgeton, L., Paillereau, N., Landron, S., Gao, J., & Kamiyama, T. (2012). *Analyse formantique des voyelles orales du français en contexte isolé: À la recherche d'une référence pour les apprenants de FLE*. 145-152. <https://halshs.archives-ouvertes.fr/halshs-00977591>
- Ghio, A. (1997). *Achile: Un dispositif de décodage acoustico-phonétique et d'identification lexicale indépendant du locuteur à partir de modules mixtes* [Thèse Science et Technologies de l'Information]. Université de droit, d'économie et des sciences d'Aix-Marseille III, Faculté de Sciences et Techniques de Saint-Jerome.
- Ghio, A., Giusti, L., Blanc, E., & Pinto, S. (2020). French adaptation of the "Frenchay Dysarthria Assessment 2" speech intelligibility test. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 137(2), 111-116. <https://doi.org/10.1016/j.anorl.2019.10.007>

- Ghio, A., Guisti, L., Blanc, E., Pinto, S., Lalain, M., Robert, D., Fredouille, C., & Woisard, V. (2016). Quels tests d'intelligibilité pour évaluer les troubles de la production de la parole ? *Actes de la conférence conjointe JEP-TALN-RECITAL*. JEP.
- Ghio, A., Lalain, M., Giusti, L., Pouchoulin, G., Robert, D., Rebourg, M., Fredouille, C., Laaridh, I., & Woisard, V. (2018a). Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique. *XXXIe Journées d'Études sur la Parole*, 285-293. <https://doi.org/10.21437/jep.2018-33>
- Ghio, A., Lalain, M., Giusti, L., Pouchoulin, G., Robert, D., Rebourg, M., Fredouille, C., Laaridh, I., & Woisard, V. (2018b). Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique. *XXXIe Journées d'Études sur la Parole*, 285-293. <https://doi.org/10.21437/JEP.2018-33>
- Ghio, A., Lalain, M., Rebourg, M., Fredouille, C., & Woisard, V. (2020). Evaluation de l'intelligibilité de patients avec traitement du cancer des cavités orales et pharyngales. *6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)*, 271-279. <https://hal.archives-ouvertes.fr/hal-02798549>
- Giraud, E., & Tessier, A. (2018). *Intérêt d'une prise en charge orthophonique des cancers ORL pendant la radio-chimiothérapie*. 25. <https://doi.org/dumas-02084337>
- Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, 52(3), 862-874. <https://doi.org/10.1016/j.neuroimage.2009.10.023>
- Grataloup, C., Hoen, M., Veillet, E., Collet, L., Pellegrino, F., & Meunier, F. (2009). Speech Restoration : An Interactive Process. *Journal of Speech, Language, and Hearing Research*, 52(4), 827-838. [https://doi.org/10.1044/1092-4388\(2008/06-0235\)](https://doi.org/10.1044/1092-4388(2008/06-0235))
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39(5), 350-365. <https://doi.org/10.1016/j.jcomdis.2006.06.013>
- Hardcastle, W. J., & Beck, J. M. (2014). *A Figure of Speech : A Festschrift for John Laver*. Routledge.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature*, 408(6815), 927-928. <https://doi.org/10.1038/35050160>
- Hirano, M. (1981). *Clinical Examination of Voice* (Springer Verlag).
- Huet, K., & Harmegnies, B. (2000). *Contribution à la quantification du degré d'organisation des systèmes vocaliques*. 5.
- Ishikawa, K., Nudelman, C., Park, S., & Ketring, C. (2019). Perception and Acoustic Studies of Vowel Intelligibility in Dysphonic Speech. *Journal of Voice*, 35(4), 659.e11-659.e24. <https://doi.org/10.1016/j.jvoice.2019.12.022>

Ishikawa, K., & Webster, J. (2020). The Formant Bandwidth as a Measure of Vowel Intelligibility in Dysphonic Speech. *Journal of Voice*, S0892199720304033. <https://doi.org/10.1016/j.jvoice.2020.10.012>

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI). *American Journal of Speech-Language Pathology*, 6(3), 66-70. <https://doi.org/10.1044/1058-0360.0603.66>

Jakobson, R. (1941). *Child Language, Aphasia, and Phonological Universals* (The Hague : Mouton).

Jakobson, R., Fant, C. G. M., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Acoustics Laboratory, Massachusetts Institute of Technology.

Jarzé, S., Smithson Barrière, D., Révis, J., & Ghio, A. (2017a). Analyse perceptive des voix dysphoniques. Influences et corrélations entre les dimensions G, R et B de l'échelle d'Hirano. *Journées de Phonétique Clinique*. <https://hal.archives-ouvertes.fr/hal-01615030>

Jarzé, S., Smithson Barrière, D., Révis, J., & Ghio, A. (2017b). Analyse perceptive des voix dysphoniques. Influences et corrélations entre les dimensions G, R et B de l'échelle d'Hirano. *Journées de Phonétique Clinique*. <https://hal.archives-ouvertes.fr/hal-01615030>

Jemaa, I. (2013). *Suivi de Formants par analyse en Multirésolution, interface gomme-machine* [Université de Lorraine; Faculté des Sciences de Tunis]. tel-00836717

Johnson, A., Jacobson, B., Grywalski, C., Silbergleit, A., Jacobson, G., & Benninger, M. (1997). The Voice Handicap Index (VHI) : Development and Validation. *American Journal of Speech-Language Pathology*, 6, 66-70.

Karl, J. R., & Pisoni, D. B. (1994). The role of talker-specific information in memory for spoken sentences. *The Journal of the Acoustical Society of America*, 95(5), 2873-2873. <https://doi.org/10.1121/1.409447>

Kent, R. D., Kent, J. F., Rosenbek, J. C., Vorperian, H. K., & Weismer, G. (1997). A speaking task analysis of the dysarthria in cerebellar disease. *Folia Phoniatica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, 49(2), 63-82. <https://doi.org/10.1159/000266440>

Kent, R. D., & Kim, Y. -J. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics*, 17(6), 427-445. <https://doi.org/10.1080/0269920031000086248>

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward Phonetic Intelligibility Testing in Dysarthria. *Journal of Speech and Hearing Disorders*, 54(4), 482-499. <https://doi.org/10.1044/jshd.5404.482>

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic Studies of Dysarthric Speech : Methods, Progress and Potential. *Journal of Communication Disorders*, 32, 141-186. [https://doi.org/S0021-9924\(99\)00004-0](https://doi.org/S0021-9924(99)00004-0)

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-

impaired listeners. *The Journal of the Acoustical Society of America*, 122(4), 2365-2375. <https://doi.org/10.1121/1.2773986>

Kuhl, P. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories. Monkeys do not. *Perception & psychophysics*, 50, 93-107. <https://doi.org/10.3758/BF03212211>

Laccourreye, L., Ettienne, V., Prang, I., Couloigner, V., Garabedian, E.-N., & Loundon, N. (2015). Perception, production et intelligibilité de la parole de l'enfant francophone sourd profond implanté cochléaire précoce après infection congénitale par le cytomégalovirus. *Annales françaises d'Oto-rhino-laryngologie et de Pathologie Cervico-faciale*, 132(6), 290-294. <https://doi.org/10.1016/j.aforl.2015.02.004>

Ladefoged, P. (1995). *Elements of Acoustic Phonetics* (2d edition, Éd.). University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/E/bo3683670.html>

Lagier, A., Assaiante, C., Legou, T., Amy de la Bretèque, Benoît, Meynadier, Y., & Giovanni, A. (2009). Étude de la voix chuchotée : Pression sous-glottique réelle et analyse posturale. *3è Journées de Phonétique Clinique*. <https://hal.archives-ouvertes.fr/hal-01212698>

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form : A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245-294. [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Papers in Laboratory Phonology VII*, 637-675.

Lalain, M., Ghio, A., Giusti, L., Robert, D., Fredouille, C., & Woisard, V. (2020). Design and Development of a Speech Intelligibility Test Based on Pseudowords in French : Why and How? *Journal of Speech, Language, and Hearing Research*, 63(7), 2070-2083. https://doi.org/10.1044/2020_JSLHR-19-00088

Lalain, M., Monestier, L., Ghio, A., Fredouille, C., Rebourg, M., Fakhry, N., & Woisard, V. (2022). Prédiction du degré d'altération de l'intelligibilité chez des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. *32ème édition des Journées d'Etudes sur la Parole*.

Lambourg, C. (2002). Évaluation de l'intelligibilité de la parole dans les établissements recevant du public sonorisés. Les méthodes normalisées sont-elles adaptées ? *Acoustique & Technique, La sonorisation des EPR*(n° 29), 9-16.

Larousse, É. (2019, octobre 29). *Définitions : Intelligibilité - Dictionnaire de français Larousse*. Larousse. <https://www.larousse.fr/dictionnaires/francais/intelligibilit%C3%A9/43559>

Levelt, W. J. F. (1990). Speaking : From Intention to Articulation. In M. F. Garrett, *Language and Speech* (Vol. 33, p. 273-291). <http://journals.sagepub.com/doi/10.1177/002383099003300303>

Levelt, W. J. F., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, 22(1), 1-38; discussion 38-75. <https://doi.org/10.1017/s0140525x99001776>

- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-461. <https://doi.org/10.1037/h0020279>
- Lieberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Lieberman, P. (1980). Chapter 7—On the development of vowel production in young children. In G. H. Yeni-komshian, J. F. Kavanagh, & C. A. Ferguson (Éds.), *Child Phonology* (p. 113-142). Academic Press. <https://doi.org/10.1016/B978-0-12-770601-6.50012-7>
- Liljencrants, J., & Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems : The Role of Perceptual Contrast. *Language*, 48(4), 839-862. <https://doi.org/10.2307/411991>
- Lindblom, B. (1990a). Explaining Phonetic Variation : A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Éds.), *Speech Production and Speech Modelling* (p. 403-439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16
- Lindblom, B. (1990b). On the communication process : Speaker listener interaction and the development of speech. In *Augmentative and Alternative Communication* (p. 220-230).
- Lombard, C. (2009). *Orthophonie et cancérologie ORL : Rôle de l'orthophoniste auprès des patients opérés d'une chirurgie buccopharyngolaryngée* (p. 169) [Other, UHP - Université Henri Poincaré]. <https://hal.univ-lorraine.fr/hal-01887335>
- Longeart, M. (2019a, octobre 21). *Aristote* [Site Académique]. AC Grenoble Philosophie. <http://www.ac-grenoble.fr/PhiloSophie/logphil/auteurs/aristote.htm>
- Longeart, M. (2019b, octobre 21). *Platon* [Site Académique]. AC Grenoble Philosophie. <http://www.ac-grenoble.fr/PhiloSophie/logphil/auteurs/platon.htm>
- Marczyk, A., Ghio, A., Lalain, M., Rebourg, M., Fredouille, C., & Woisard, V. (2021). Optimizing linguistic materials for feature-based intelligibility assessment in speech impairments. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01610-9>
- Marczyk, A. K., Ghio, A., Lalain, M., Rebourg, M., Fredouille, C., & Woisard, V. (2020). Have a Cake and Eat it Too : Assessing Discrimination Performance of an Intelligibility Index Obtained from a Reduced Sample Size. *12th Conference on Language Resources and Evaluation (LREC 2020)*, 1784-1788. <https://hal.archives-ouvertes.fr/hal-02611678>
- Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception : A new test. *Speech Communication*, 2(1), 15-35. [https://doi.org/10.1016/0167-6393\(83\)90061-4](https://doi.org/10.1016/0167-6393(83)90061-4)
- McClelland, J. L., Rumelhart, D. E., & Group, P. R. (1987). *Parallel Distributed Processing, Volume 2 : Explorations in the Microstructure of Cognition: Psychological and Biological Models*. MIT Press.
- Meunier, C. (2007). Phonétique acoustique. In *Les dysarthries* (Solal, p. 164-173).

Meunier, C., & Ghio, A. (2018). Caractériser la distinctivité du système vocalique des locuteurs. *XXXIIe Journées d'Études sur la Parole*, 469-477. <https://doi.org/10.21437/JEP.2018-54>

Meynadier, Y. (2010). *Éléments de phonétique acoustique*. 49.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41(5), 329-335. <https://doi.org/10.1037/h0062491>

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders*, 48(6), 601-612. <https://doi.org/10.1111/1460-6984.12061>

Monestier, L., Lalain, M., Ghio, A., Woisard, V., & Fakhry, N. (2021). Prédiction du déficit d'intelligibilité chez des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. *Journées de Phonétique Clinique*, 2.

Nearey, T. M., Assmann, P. F., & Hillenbrand, J. M. (2002). Evaluation of a strategy for automatic formant tracking. *The Journal of the Acoustical Society of America*, 112(5), 2323-2323. <https://doi.org/10.1121/1.4779372>

Nguyen, N. (2005). La perception de la parole. In N. Nguyen, S. Wauquier, & J. Durand (Éds.), *Phonologie et phonétique* (p. 425-447). Hermès. <https://hal.archives-ouvertes.fr/hal-00142953>

Ohala, J. J. (1970). Aspects of the control and production of speech. *UCLA Working papers in phonetics*, 15, 192.

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. [And : A response to Pierrehumbert's commentary]. In *Papers in Laboratory Phonology I: Between the grammar and the physics of speech* (J. Kingston & M. Beckman (eds.), p. 258-175; 280-282). Cambridge University Press.

Paternostro, R. (2008). Le dévoisement des voyelles finales. *Rassegna Italiana di Linguistica Applicata*, 3(40), 129. <https://halshs.archives-ouvertes.fr/halshs-00639082>

Peckels, J., & Rossi, M. (1973). Le Test de diagnostic par paires minimales : Adaptation au français du Diagnostic Rhyme Test de Voiers. *Revue d'acoustique*, 27, 245-262.

Perry, A., & Frowen, J. (2006). Speech and swallowing function in head and neck cancer patients : What do we know? *Cancer Forum*, 30.

Philippart, M., Delvaux, V., Kathy, H., Monnier, M., Piccaluga, M., & Harmegnies, B. (2020). Évolution de l'organisation du système vocalique du français en contexte de bilinguisme simultané. *SHS Web of Conferences*, 78, 09008. <https://doi.org/10.1051/shsconf/20207809008>

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics*, 18(3), 375-394. [https://doi.org/10.1016/S0095-4470\(19\)30380-8](https://doi.org/10.1016/S0095-4470(19)30380-8)

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816-847. <https://doi.org/10.1016/j.neuroimage.2012.04.062>

Rebourg, M. (2017). *Mesure de l'intelligibilité : La tâche de décodage acoustico-phonétique dans le développement d'outils de mesure* [Mémoire de Master 1 - sciences du langage - Théorie Linguistique : terrain et expérimentation - Linguistique expérimentale]. Aix-Marseille Université.

Rebourg, M. (2018). *Validation d'une tâche de Décodage Acoustico Phonétique : Lexicalisation, mémorisation, familiarisation* [Mémoire de Master 2 - Sciences du langage - Linguistique expérimentale]. Aix-Marseille Université.

Rossi, M. (1981). De la physiologie à la perception phonémique. In *Modèles linguistiques: Vol. Fascicule 2* (p. 5-23). Presses Universitaires de Lille.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125(1), 28-51. <https://doi.org/10.1037/0096-3445.125.1.28>

Samuel, A. G. (2001). KNOWING A WORD AFFECTS THE FUNDAMENTAL PERCEPTION OF THE SOUNDS WITHIN IT. *PSYCHOLOGICAL SCIENCE*, 12(4), 4.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In *Intelligibility in Speech Disorders: Theory, measurement and management* (Raymond D. Kent, p. 11-34). John Benjamins Publishing Company. <https://benjamins.com/catalog/sspl.1.02sch>

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336-354. <https://doi.org/10.1016/j.jneuroling.2009.12.004>

Steeneken, H. J. M. (1982). Ontwikkeling en toetsing van een Nederlandstalige Diagnostische Rijmtest voor het testen van spraakcommunicatiekanalen. *Rapport IZF*, 13.

Suarez-Cunqueiro, M.-M., Schramm, A., Schoen, R., Seoane-Lestón, J., Otero-Cepeda, X.-L., Bormann, K.-H., Kokemueller, H., Metzger, M., Diz-Dios, P., & Gellrich, N.-C. (2008). Speech and Swallowing Impairment After Treatment for Oral and Oropharyngeal Cancer. *Archives of Otolaryngology-Head & Neck Surgery*, 134(12), 1299. <https://doi.org/10.1001/archotol.134.12.1299>

Tikofsky, R. S. (1970). A Revised List for the Estimation of Dysarthric Single Word Intelligibility. *Journal of Speech and Hearing Research*, 13(1), 59-64. <https://doi.org/10.1044/jshr.1301.59>

Tikofsky, R. S., & Tikofsky, R. P. (1964). Intelligibility Measures of Dysarthric Speech. *Journal of Speech and Hearing Research*, 7(4), 325-333. <https://doi.org/10.1044/jshr.0704.325>

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie* (Prague). trad. fr. Cantineau J. 1949, Principes de phonologie, Paris, Klincksieck.

Vazquez, M.-P., Soupre, V., Bénateau, H., Seigneuric, J.-B., Martinez, H., Taieb, M.-F., Trichet-Zbinden, C., & Picard, A. (2007). Les séquelles vélopharyngées dans les fentes labioalvéolopalatovélaires. Véloplasties et pharyngoplasties par lambeau pharyngé à pédicule supérieur ou inférieur. *Revue de Stomatologie et de Chirurgie Maxillo-faciale*, 108(4), 334-342. <https://doi.org/10.1016/j.stomax.2007.06.012>

Voiers, W. D., Sharpley, A. D., & Hehmsoth, C. J. (1973). *Research on Diagnostic Evaluation of Speech Intelligibility*. Defense Technical Information Center. <https://doi.org/10.21236/AD0755918>

Warren, R. M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science*, 167(3917), 392-393. <https://doi.org/10.1126/science.167.3917.392>

Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatica et Logopaedica*, 53(1), 1-18. <https://doi.org/10.21236/AD0755918>

Weismer, G., Kent, R. D., Hodge, M., & Martin, R. (1988). The acoustic signature for intelligibility test words. *The Journal of the Acoustical Society of America*, 84(4), 1281-1291. <https://doi.org/10.1121/1.396627>

Woisard, V. (2011a). *Impact de l'intelligibilité dans les troubles de la production de la parole* [Thèse de doctorat - Sciences du langage]. Aix-Marseille Université.

Woisard, V. (2011b). *Impact de l'intelligibilité dans les troubles de la production de la parole* [Thèse de doctorat - Sciences du langage]. Aix-Marseille Université.

Woisard, V., Andrieux, M. P., & Puech, M. (2006). Validation d'un questionnaire d'auto-évaluation du handicap pour les troubles de la déglutition oropharyngée (Deglutition Handicap Index). *Revue de laryngologie - otologie - rhinologie*, 127(5), 315-325. <https://www.lissa.fr/rep/articles/17425006>

Woisard, V., Astésano, C., Balaguer, M., Farinas, J., Fredouille, C., Gaillard, P., Ghio, A., Giusti, L., Laaridh, I., Lalain, M., Lepage, B., Mauclair, J., Nocaudie, O., Piquier, J., Pouchoulin, G., Puech, M., Robert, D., & Roger, V. (2020). C2SI corpus : A database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-020-09496-3>

Yorkston, K. M. (1992). Intelligibility measurement as a tool in the clinical management of dysarthric speakers. In *Intelligibility in Speech Disorders: Theory, Measurement, and Management*. John Benjamins Publishing.

Zappa, A., Mestre, D., Pergandi, J.-M., Bolger, D., & Frenck-Mestre, C. (2022). Cross-linguistic gender congruency effects during lexical access in novice L2 learners : Evidence from ERPs. *Language, Cognition and Neuroscience*. <https://doi.org/10.1080/23273798.2022.2039726>

Zue, V. W., & Seneff, S. (1996). Transcription and Alignment of the TIMIT Database. In H. Fujisaki (Éd.), *Recent Research Towards Advanced Man-Machine Interface Through Spoken Language* (p. 515-525). Elsevier Science B.V. <https://doi.org/10.1016/B978-044481607-8/50088-8>

Table des figures

Figure 1 : Schéma de la chaîne de communication parlée (Denes & Pinson, 1993)	22
Figure 2 : Coupe sagittale des localisations cavité buccale et oropharynx	30
Figure 3 : Organes impliqués dans la production de parole	39
Figure 4 : Schéma du trajet de l'air à travers le conduit vocal et les résonateurs	39
Figure 5 : Anatomie et physiologie de l'oreille	43
Figure 6 : Illustration du système vocalique du français.....	60
Figure 7 : Schéma du processus de phonétisation des stimuli et des réponses de auditeurs.....	77
Figure 8 : Matrice de coût des segments vocaliques, en nombre de traits différents entre les segments. Ê, Û et Ô sont des archiphonèmes correspondant respectivement à {e, ε}, {oe, ø} et {o, ɔ} (Ghio et al., 2018a).....	78
Figure 9 : Exemple de comparaison entre les deux chaînes de caractères que constituent la cible et la transcription de l'auditeur (/μ/ = /ε̃/).....	79
Figure 10 : Localisation de l'atteinte physiologique, divisée en deux groupes, la cavité buccale (39% des patients) – plancher, langue, rétromolaire et mandibule - et l'oropharynx (61% des patients) – amygdale, base de langue, oropharynx et voile du palais -.....	93
Figure 11 : Répartition des patients en fonction de la taille de la tumeur (en %)	94
Figure 12 : Répartition des patients en fonction de l'envahissement des ganglions lymphatiques (en %).....	94
Figure 13 : Proportion de patients traités par chirurgie et type d'intervention chirurgicale pratiquée (en %)	94
Figure 14 : Type de traitement reçu par les patients, exprimé en pourcentage. « RT » Radiothérapie ; « Chimio » Chimiothérapie ; « ChirT » Chirurgie de la tumeur ; « ChirN » Chirurgie des ganglions ; lorsque le patient a reçu une combinaison de traitement, l'ordre est indiqué par l'ordre des abréviations (« ChirT_Rt » signifie que le patient a bénéficié d'une chirurgie de la tumeur, puis d'une radiothérapie) (en %).....	95
Figure 15 : Localisation de l'atteinte physiologique, divisée en deux groupes, la cavité buccale (60% des patients) – plancher, langue, rétromolaire et mandibule - et l'oropharynx (40% des patients) – amygdale, base de langue, oropharynx et voile du palais-.....	98
Figure 16 : Répartition de la population de patients en fonction de la taille de la tumeur (en %)	99
Figure 17 : Répartition de la population de patients en fonction de l'envahissement des ganglions lymphatiques (en %)	99
Figure 18 : Proportion de patients traités par chirurgie et type d'intervention chirurgicale pratiquée (en %)	99

Figure 19 : Type de traitement reçu par les patients, exprimé en pourcentage. « RT » Radiothérapie ; « Chimio » Chimiothérapie ; « Chir » Chirurgie de la tumeur et/ou des ganglions (en %).....	100
Figure 20 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item lexical BECD dans la passation.....	112
Figure 21 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item non lexical DAP dans la passation.....	113
Figure 22 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item non lexical DAP dans la passation, avec des auditeurs experts.....	114
Figure 23 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type d'auditeur, naïfs et experts, pour le matériel linguistique non lexical DAP.....	115
Figure 24 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type d'auditeur, naïfs et experts, pour le matériel linguistique lexical BECD.....	115
Figure 25 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item lexical BECD dans la passation.....	118
Figure 26 : Représentation en nuage de points du score d'intelligibilité moyen en fonction de la position de l'item non lexical DAP dans la passation.....	118
Figure 27 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type de matériel linguistique, mots de la BECD et pseudo-mots du DAP.....	119
Figure 28 : Répartition des locuteurs en fonction du groupe (patient vs contrôle) et du sexe (homme et femme), en nombre par catégorie et en pourcentage du total des locuteurs.....	122
Figure 29 : Répartition des patients (en %) selon la localisation estimée de la tumeur, en gris les organes de la cavité buccale et en bleu ceux de l'oropharynx.....	123
Figure 30 : Répartition de la population de patients en fonction de la taille de la tumeur (en %).....	124
Figure 31 : Type de traitement reçus par les patients, exprimé en pourcentage. « RT » Radiothérapie ; « Chimio » Chimiothérapie ; « ChirT » Chirurgie de la tumeur ; « ChirN » Chirurgie des ganglions ; lorsque le patient a reçu une combinaison de traitement, l'ordre est indiqué par l'ordre des abréviations (« ChirN_Rt » signifie que le patient a bénéficié d'une chirurgie des ganglions, puis d'une radiothérapie).....	125
Figure 32 : Scores PPD locaux moyens selon les macro-classes phonémiques pour les groupes de locuteurs patients et contrôles.....	131
Figure 33 : Scores PPD locaux moyens selon les macro-classes phonémiques consonnes et les sous classes voyelles nasales et voyelles orales, pour les groupes de locuteurs patients et contrôles.....	132
Figure 34 : Scores PPD locaux moyens selon la macro-classe phonémique consonnes et la sous classe voyelles orales, pour les groupes de locuteurs patients et contrôles.....	132
Figure 35 : Scores PPD locaux moyens par segment vocalique, pour les groupes de locuteurs patients et contrôles.....	134

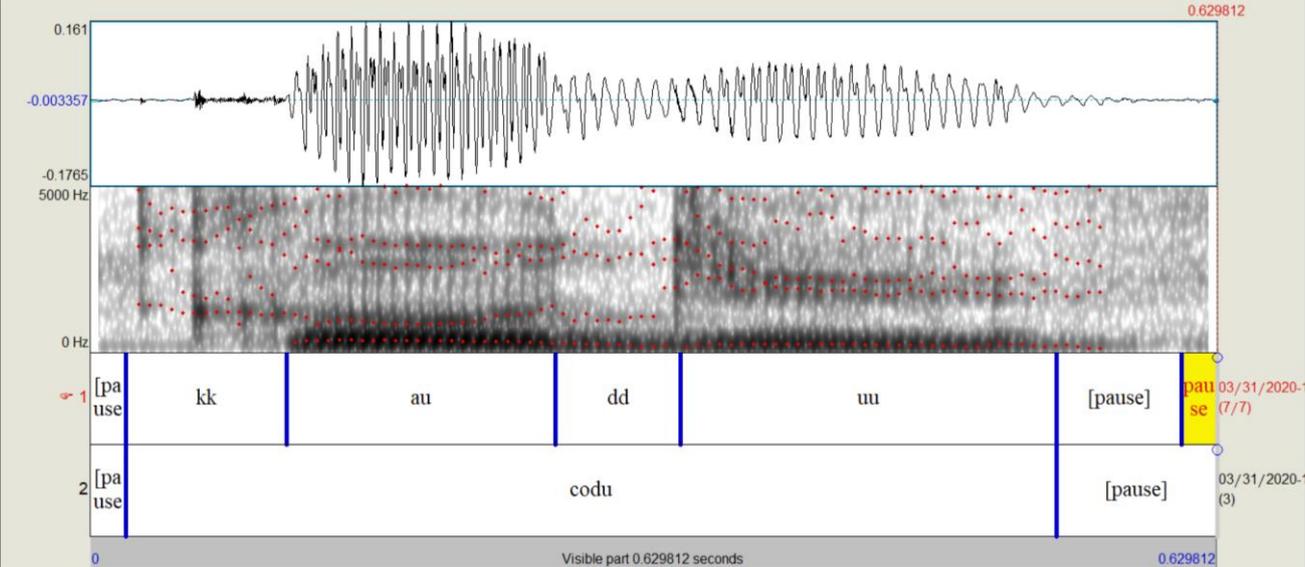
Figure 36 : Scores PPD locaux moyens selon les sous classes phonémiques nasales et orales, pour les groupes de locuteurs patients et contrôles	134
Figure 37 : Scores PPD locaux moyens selon la position avant (Front) vs arrière (Back) pour les groupes de locuteurs patients et contrôles	135
Figure 38 : Scores PPD locaux moyens en fonction de la hauteur (High, Medium, Low), pour les groupes de locuteurs patients et contrôles	136
Figure 39 : Scores PPD locaux par voyelle orale, pour les groupes de locuteurs patients et contrôles	137
Figure 40 : Taux de réponses correctes et de substitutions pour les segments vocaliques comportant le trait de nasalité /@/ et /μ/, dans les groupes contrôle et patient	139
Figure 41 : Taux de réponses correctes et de substitutions pour le segment /i/, dans les groupes contrôle et patient	141
Figure 42 : Taux de réponses correctes et de substitutions pour le segment /y/, dans les groupes contrôle et patient.....	141
Figure 43 : Taux de réponses correctes et de substitutions pour le segment /E/, dans les groupes contrôle et patient.....	143
Figure 44 : Taux de réponses correctes et de substitutions pour le segment /Ô/, dans les groupes contrôle et patient.....	143
Figure 45 : Taux de réponses correctes et de substitutions pour le segment /a/, dans les groupes contrôle et patient.....	145
Figure 46 : Taux de réponses correctes et de substitutions pour le segment /u/, dans les groupes contrôle et patient.....	146
Figure 47: Illustration du signal acoustique d'un locuteur patient prononçant le pseudo-mot « ganlou ». En premier l'oscillogramme du signal produit, suivit par le spectrogramme à bande large sur lequel apparaissent les formants (en rouge). L'annotation segmentale en phonème (convention LIAPhon) se retrouve dans la première tire d'annotation, la seconde représente le pseudo-mot qui devait être produit.	153
Figure 48 : Illustration du signal acoustique d'un patient prononçant le pseudo-mot « prega ».....	153
Figure 49 : Illustration du signal acoustique d'un patient produisant le pseudo-mot « chovu ».	154
Figure 50 : Signal acoustique d'un patient produisant le pseudo-mot « flifou ». En premier l'oscillogramme du signal produit, suivit par le spectrogramme à bande large sur lequel apparaissent les formants (en rouge). L'annotation segmentale en phonème (convention LIAPhon) se retrouve dans la première tire d'annotation, la seconde représente le pseudo-mot qui devait être produit. La troisième tire a été ajoutée par nos soins pour proposer une correction de l'alignement phonémique.....	155
Figure 51 : Signal acoustique d'un patient produisant le pseudo-mot « floulu ».....	156
Figure 52 : Signal acoustique de la production du pseudo-mot « touca » par une patiente.....	157
Figure 53 : Signal acoustique de la production du pseudo-mot « nianrtou » par un patient.....	157

Figure 54 : Signal acoustique de la production du pseudo-mot « zimi » par un patient	158
Figure 55 : Signal acoustique d'un patient qui aurait dû produire le pseudo-mot « fanju »	159
Figure 56 : Illustration de la visualisation de j	163
Figure 57 : Droite de régression linéaire illustrant la corrélation entre les Scores PPDGlobaux et les scores PPDLocaux des voyelles.....	167
Figure 58 : Plan F1/F2 pour les locuteurs masculins, Contrôle vs. Patient	168
Figure 59 : Plan F1/F2 pour les locutrices féminines, Contrôle vs. Patient.....	169
Figure 60 : Plan F1/F2 pour les locutrices féminines, Contrôles vs. Patients, chaque 'point' représente les transcriptions phonémiques des auditeurs	171
Figure 61 : Plan F1/F2 de 10 les locutrices féminines, 5 contrôles vs. 5 patientes, chaque 'point' représente les transcriptions phonémiques des auditeurs	172

ANNEXES

A. Exemple de spectrogramme

Un exemple de signal acoustique étiqueté et aligné avec sa transcription LIAPhon



B. Projet C2SI

Le projet de recherche Carcinologic Speech Severity Index (C2SI), financé par l'Institut National du Cancer (INCa) (Grant 2014-135) en 2014, « Sciences Humaines et Sociales, épidémiologie et Santé Publique » dirigé par le professeur Virginie Woisard du Centre Hospitalier Universitaire de Toulouse vise à développer une « Mesure de l'impact des traitements des cancers de la cavité buccale et du pharynx sur la production de la parole par l'Indice de sévérité des troubles de la production de la parole ».

Nos questions de recherches s'insèrent pour partie dans ce projet, et c'est de ce projet que provient l'un des corpus mobilisés dans nos travaux. Nous l'avons souligné, notre projet de recherche trouve sa genèse dans le cadre du projet C2SI et le dépasse.

L'objectif du projet de recherche C2Si est de développer une technique d'évaluation automatique qui permettrait de mesurer l'impact du trouble de la production de la parole sur les habilités communicatives en proposant un index de sévérité pour des patients traités pour un cancer VADS.

Evaluer l'intelligibilité de la parole est une tâche ardue, qui nécessite plusieurs tâches pour couvrir toutes les dimensions du handicap communicationnel. Ce projet le souligne en proposant plusieurs tâches qui mettent en jeu différents niveaux linguistiques en vue de leur évaluation. Les tâches orientées vers la compréhension permettent l'intégration d'informations de haut niveau alors que les tâches qui visent à évaluer la qualité de production des sons incrémentent la perception des unités de bas niveau.

Plusieurs objectifs soutiennent ce projet, notamment la création d'un Indice de Sévérité de la Parole Carcinologique (C2SI) au travers de traitements automatiques de la parole, et plus particulièrement de scores d'intelligibilité, dans les cancers de la cavité buccale et du pharynx. Cet indice devrait avoir valeur de prédictibilité du handicap de parole, des conséquences du cancer et des traitements sur la communication parlée. De plus, les caractéristiques des déficits fonctionnels et des troubles de la parole devraient permettre de décrire des profils de patients. Ces objectifs se concrétisent au travers de deux principaux livrables : « - un indice de sévérité du trouble de la parole (C2SI) utilisable pour quantifier les conséquences de l'altération de la parole et - un score d'intelligibilité automatique »

Cet ambitieux projet implique plusieurs équipes de recherches pluridisciplinaires, comprenant des médecins ORL, phoniâtres, des orthophonistes, des informaticiens, des ingénieurs de recherche et des linguistes. Ce projet collaboratif animé par des acteurs dynamiques est également multicentrique et mobilise le Centre Hospitalier Universitaire de Toulouse, le Laboratoire Octogone-Lordat et l'Institut de Recherche en Informatique de Toulouse (IRIT) des Universités de Toulouse, le Laboratoire d'Informatique d'Avignon (LIA) et le Laboratoire Parole et Langage (LPL-Aix-Marseille Université) à Aix-en-Provence. L'association de cliniciens, d'informaticiens et de chercheurs permet de satisfaire les enjeux cliniques et linguistiques de l'évaluation de la parole au travers de critères de validation scientifiques des outils développés. Chaque équipe est assignée à la

réalisation de tâches prédéfinies. Chacune bénéficie des compétences des autres au travers de collaboration inter-équipes.

L'absence de test d'évaluation de parole pathologique fiable, pertinent et représentant le handicap communicationnel du patient motive pour partie l'élaboration de ce projet de recherche. L'évaluation de la parole est affaire de complexité et celle-ci est augmentée par la pathologie.

Locuteurs et tâches de production

Pour mener à bien ce projet, 87 patients traités pour un cancer de la cavité buccale et/ou de l'oropharynx, et 42 sujets sains, ont été recrutés et enregistrés sur différentes tâches de production de parole. Les critères d'inclusion des patients sont : - la prise en charge pour un cancer de la cavité buccale et/ou de l'oropharynx classé T2, T3 ou T4 et traité par chirurgie et/ou radiothérapie et/ou chimiothérapie ; - un délai post traitement à plus de 6 mois après la fin du traitement ; - être en état de rémission clinique. La population constituant ce corpus est plus largement décrite dans la partie 5.1 Corpus C2Si.

Pour servir les objectifs de ce projet, visant à déterminer une mesure automatique de l'intelligibilité, plusieurs tâches de production de parole ont été élaborées, puis enregistrées auprès des locuteurs suscités, afin de constituer un corpus de travail. Ces différentes tâches servent différents objectifs au travers de différentes évaluations. Chaque équipe a été investie dans l'élaboration de tâches prédéfinies relatives à leur domaine de compétences et les nombreuses collaborations inter-équipes permettent de cumuler les compétences et qualités de chacun.

Dans la partie suivante nous décrirons les différentes tâches de production auxquelles les locuteurs ont participé, en vue de l'enregistrement d'un corpus.

Dans la tâche « AAA » le locuteur reçoit pour consigne de produire 3 fois le phonème vocalique /a/ tenu le plus longtemps possible. Cette tâche permet de réaliser des mesures acoustiques sur un phonème unique, en termes de stabilité, temps de phonation, harmoniques... et est destinée à évaluer la qualité de la voix.

Deux tâches de parole continue sont proposées au travers de tâches de lecture. Le premier paragraphe de la nouvelle « La chèvre de M. Seguin » par Alphonse Daudet, est lu à haute voix pour la tâche « LEC ». C'est un texte bien connu et utilisé dans l'évaluation phonétique clinique. Pour la tâche « SVT » un ensemble de 50 phrases (sélectionné parmi une liste de 300 phrases) est lu par le locuteur. Ce sont des phrases vraies (« Le chat miaule ») ou des phrases fausses (« Le chat aboie »), qui se distinguent toujours par le mot cible. Quand ces phrases sont présentées à des auditeurs, ces derniers doivent décoder et comprendre la totalité de la phrase afin de pouvoir déterminer si elle est vraie ou fausse (tâche de compréhension).

Une évaluation de plusieurs niveaux prosodiques est proposée au travers de trois tâches de production basées sur la lecture de phrases. L'expression de la modalité (tâche MOD) est évaluée par la production de dix phrases différentes selon trois modalités : interrogative, assertive, exclamative (« Tu viens à la plage ? / . / ! »). La portée de la

focalisation (tâche FOC) est évaluée par la production d'un ensemble de 20 phrases dans lesquelles le mot cible doit être mis en avant par le focus prosodique. Ces phrases sont produites en réponse à une question posée par une synthèse vocale selon deux modalités : congrue (« Tu as vu un canard ou un cochon dans le jardin ? ») et incongrue (« Tu as vu un canard dans le jardin ou dans la cour de l'école ? ») et le locuteur de répondre en lisant la phrases réponse « J'ai vu un CANARD dans le jardin ». Il doit mettre l'emphase prosodique sur le mot en capitales. La dernière tâche est une tâche de désambiguïsation syntaxique (SYN) dans laquelle les locuteurs produisent des phrases avec un contraste spécifique. Par exemple, dans la phrase « Les chevaux et les poneys blancs », l'adjectif « blancs » peut porter sur « les chevaux et les poneys » ou seulement sur « les poneys ». La désambiguïsation du sens de cette phrase peut être marquée prosodiquement « Les chevaux et les poneys _ blancs » ou « Les chevaux _ et les poneys blancs »

Les tâches de prosodie (MOD, FOC et SYN) et la tâche de SVT, ont fait l'objet de tests de perception auprès de 147 auditeurs naïfs. Chaque phrase a été évaluée 3 fois par 3 auditeurs différents. L'auditeur reçoit une consigne spécifique à chaque type de production. Dans la tâche MOD, il doit déterminer la modalité de la phrase perçue (interrogative, assertive ou exclamative). Pour la tâche FOC, l'auditeur doit juger de la congruité de la phrase perçue par rapport à la question posée (congrue vs incongrue). Dans la tâche SYN, l'auditeur doit choisir entre deux images laquelle représente la phrase perçue. Enfin, pour la SVT, l'auditeur doit juger si la phrase est vraie ou si elle est fausse. C'est la compréhensibilité qui est évaluée dans cette dernière.

Deux tâches de production de parole spontanée sont également enregistrées. La première est une description d'une image choisie par le locuteur parmi plusieurs (DES). Toutes les images ont pour thème la mer et les bateaux. Le locuteur doit décrire l'image à l'évaluateur afin qu'il puisse la dessiner sur la base d'explications orales. Dans la dernière tâche (SPO), le locuteur doit donner son opinion, pendant au moins 3 minutes, sur un questionnaire auquel il a répondu avant la session d'enregistrement.

Les enregistrements des tâches de parole continue et parole spontanée permettent d'entraîner des processus automatiques pour la reconnaissance et l'évaluation de la parole. Elles font également l'objet d'évaluations perceptives par des auditeurs humains. La comparaison des évaluations automatiques et humaines contribue à la recherche fondamentale en linguistique et en informatique et peut permettre de montrer qu'une évaluation automatique peut être équivalente à celle d'un auditeur humain, en termes de prédiction du handicap de parole.

Une tâche (DAP) d'évaluation de la qualité articulatoire au travers du décodage acoustico-phonétique de pseudomots est également proposée. C'est une tâche de production d'une liste de 50 pseudo-mots. Ces productions sont ensuite présentées à des auditeurs dans un tests de jugement perceptif dans lequel ils transcrivent ce qu'ils ont perçu. Une comparaison entre le pseudo-mot qui devait être produit et la transcription permet d'obtenir un score de distance à la cible en termes de traits phonétiques. Les détails concernant la tâche de DAP sont à retrouver dans la partie 3.3.1 Méthode de Décodage acoustico-phonétique.

C. Questionnaire auditeurs experts

1. Vous travaillez en Profession libérale ? (..... %)
Autre (CH, SSR, CAMSP, ...) ? (..... %)

2. Comment caractériseriez-vous votre patientèle générale ? (Ex : 50% langage écrit, 25% dysarthrie, 25% dyscalculie)
.....
.....

3. Etes-vous amené à évaluer l'intelligibilité de vos patients ? Oui – Non

Si oui, pour quelles pathologies ?
.....
.....

4. Pour évaluer l'intelligibilité utilisez-vous des bilans ? Oui – Non

Si oui, lesquels ?
.....
.....

5. Avez-vous déjà utilisé la BECD ? Oui – Non

L'utilisez-vous souvent ? Oui – Non
Environ fois par semaine – mois
Pour quelles pathologies ?
.....
.....

Pouvez-vous citer des items ? (Citez tous ceux dont vous vous souvenez)
.....
.....

6. Concernant l'évaluation de l'intelligibilité dans votre pratique, êtes-vous satisfait ?
Avez-vous des remarques
.....
.....
.....
.....
.....
.....

Cochez les items qui font partie de la liste d'items disponibles dans la Batterie d'évaluation clinique de la dysarthrie (BECD) :

- abri
- artiste
- bouche
- boxeur
- clenche
- cœur
- collet
- craie
- dent
- docteur

- froid
- groin
- grumeau
- harpe
- herbe
- impôt
- miel
- mouche
- noix
- paix

- paquet
- patrie
- peur
- plancher
- poil
- pont
- spasme
- sport
- taxi
- thym

A. Scripts Praat

This script opens each sound file in a directory, looks for a corresponding TextGrid and extracts f0, F1, F2 and intensity from the midpoint(s) of any labelled interval(s) in the specified TextGrid tier.

It also extracts the duration of the labelled interval(s) as well as labels from other intervals located at the same midpoint(s) on other tiers.

All these results are written to a tab-delimited text file.

The script is a modified version of the script "collect_formant_data_from_files.praat" by Mietta Lennes, available here: <http://www.helsinki.fi/~lennes/praat-scripts/>

This script was first modified by Dan McCloy (drmcclay@uw.edu) in December 2011 and later modified to add intensity and labels from other tiers by Esther Le Grézause (elg1@uw.edu) in May 2016.

This script was last modified by Anna Marczyk, June 2020

This script is distributed under the GNU General Public License.

Copyright 4.7.2003 Mietta Lennes

```
form Get pitch formants intensity and duration from labeled segments in files
  comment Directory of sound files. Be sure to include the final "/"
  text sound_directory C:/Users/Marie/Nextcloud/Manip2/C2SI_PseudoMots/
  sentence Sound_file_extension .wav
  comment Directory of TextGrid files. Be sure to include the final "/"
  text textGrid_directory C:/Users/Marie/Nextcloud/Manip2/C2SI_PseudoMots/
  sentence TextGrid_file_extension .TextGrid
  comment Full path of the resulting text file:
  text resultsfile C:/Users/Marie/Nextcloud/Manip2/output/output_Test.csv
  comment Which tier do you want to analyze?
  integer Tier 1
  comment Formant analysis parameters
  positive Time_step 0.01
  integer Maximum_number_of_formants 5
  positive Maximum_formant_(Hz) 5500
  positive Window_length_(s) 0.025
  real Preemphasis_from_(Hz) 50
  comment Pitch analysis parameters
  positive Pitch_time_step 0.01
  positive Minimum_pitch_(Hz) 75
  positive Maximum_pitch_(Hz) 300
endform
```

Make a listing of all the sound files in a directory:

```

Create Strings as file list... list 'sound_directory$'*'sound_file_extension$'
numberOfFiles = Get number of strings

# Assigning tier number
identification_tier = 1
syllable_tier = 2

# Check if the result file exists:
if fileReadable (resultsfile$)
    pause The file 'resultsfile$' already exists! Do you want to overwrite it?
    filedelete 'resultsfile$'
endif

# Create a header row for the result file: (remember to edit this if you add or change
the analyses!)
header$ = "Loc   voyelle       duration   identification syllable       F1       F2
'newline$"
fileappend "'resultsfile$" 'header$'

# Open each sound file in the directory:
for ifile to numberOfFiles
    filename$ = Get string... ifile
    Read from file... 'sound_directory$"filename$'

# get the name of the sound object:
soundname$ = selected$ ("Sound", 1)

# Look for a TextGrid by the same name:
gridfile$ = "textGrid_directory$"soundname$"textGrid_file_extension$"

# if a TextGrid exists, open it and do the analysis:
if fileReadable (gridfile$)
    Read from file... 'gridfile$'

    select Sound 'soundname$'
    To Formant (burg)... time_step maximum_number_of_formants
maximum_formant window_length preemphasis_from

    select Sound 'soundname$'
    To Pitch... pitch_time_step minimum_pitch maximum_pitch

    select Sound 'soundname$'
    To Intensity... minimum_pitch time_step

```

```

select TextGrid 'soundname$'
numberOfIntervals = Get number of intervals... tier

# Pass through all intervals in the designated tier, and if they have a label,
do the analysis:
for interval to numberOfIntervals
  label$ = Get label of interval... tier interval
  if label$ <> ""
    # duration:
    start = Get starting point... tier interval
    end = Get end point... tier interval
    midpoint = (start + end) / 2
    duration = end-start
    cursor33 = start + 0.33*duration
    cursor66 = start + 0.66*duration
    duration_msec = duration*1000

    # get the matching interval (at the midpoint) in the identification tier
    identification_interval = Get interval at time...
identification_tier midpoint
    # get label of interval in the identification tier
    label_identification$ = Get label of interval...
identification_tier identification_interval

    # get the matching interval (at the midpoint) in the syllable tier
    syllable_interval = Get interval at time... syllable_tier midpoint
    # get label of interval in the syllable tier
    label_syllable$ = Get label of interval... syllable_tier
syllable_interval

    # formants:
    select Formant 'soundname$'
    f1_mean = Get mean... 1 cursor33 cursor66 Hertz
    f2_mean = Get mean... 2 cursor33 cursor66 Hertz

    # Save result to text file:
    resultline$ = "'soundname$'      'label$'
'duration_msec'  'label_identification$'      'label_syllable$'      'f1_mean'
'f2_mean'      'newline$'"
    fileappend "'resultsfile$'" 'resultline$'

    # select the TextGrid so we can iterate to the next interval:
    select TextGrid 'soundname$'

```

```
        endif
    endfor
    # Remove the TextGrid, Formant, and Pitch objects
    select TextGrid 'soundname$'
    plus Formant 'soundname$'
    plus Pitch 'soundname$'
    Remove
endif
# Remove the Sound object
select Sound 'soundname$'
Remove
# and go on with the next sound file!
select Strings list
endif
endfor

# When everything is done, remove the list of sound file paths:
Remove
```

B. Script R

```
#           VOWEL QUALITY METRICS
#           Final script by A. Marczyk
#           July 2020

require(graphics)
library(phonR)
setwd("C:/Users/Marie/Nextcloud/Manip2/Output")
data <-
read.table("C:/Users/Marie/Nextcloud/Manip2/Output/Output_Final_MetaD.txt",
head=TRUE, fill=TRUE)
summary(data)

## Checker les variables

str(data)
data$DURATION=as.numeric(data$DURATION)
data$F2=as.numeric(data$F2)
data$F1=as.numeric(data$F1)
data$Sexe=as.factor(data$Sexe)
data$LexiPhon=as.factor(data$LexiPhon)

summary(data)# check

### subsetting

voyd<-subset(data, DURATION>= 30 & DURATION <=400)
summary(voyd$DURATION)
voyd<-subset(voyd, F1=1000)
data<-subset(voyd, LexiPhon %in% c("a", "i", "u", "e", "o", "y", "O", "E", "9", "@", "5"))
summary(data)
str(data)

# we compute the center of gravity k of the first formant F1
# per vowel category (voyelle) and speaker (Loc) and store it as CG_K_F1

data$CG_K_F1<-ave(data$F1, data$ID_Loc, data$LexiPhon, na.rm=TRUE)
summary(data)
head(data) # we check we got it right
```

```

# we compute the center of gravity k of the second formant F2
# per vowel category (voyelle) and speaker (Loc) and store it as CG_K_F2

data$CG_K_F2<-ave(data$F2, data$ID_Loc, data$LexiPhon, na.rm=TRUE)
summary(data)
head(data, n=15) # we check we got it right

# we compute the euclidean distance between every vowel category and its center of
gravity
# see the paper for the formula

data$a<-((data$F1 - data$CG_K_F1))^2 # we compute squared F1 of a particular vowel
# minus center of gravity of F1 for this vowel's category (CG_k_F1_new)
data$b<-((data$F2 - data$CG_K_F2))^2 # same for F2

data$d_intra <- sqrt(data$a + data$b)# we square root the sum of squared F1 and F2
indexes
# and store it as a variable d_intra

head(data, n=15) # we check we got it right

#=====
# COMPUTING THE SYSTEM'S CENTER OF GRAVITY
#=====

#=====
#           A. EXEMPLAR
#=====
## OK pas de pb particulier ##

# we compute the center of gravity for each formant
# i.e. mean of all the F1 and F2 per speaker (across vowel categories)
# we store it as new variables CG_SYST_J_F1 and CG_SYST_J_F2

data$CG_SYST_J_F1<-ave(data$F1, data$ID_Loc, na.rm=TRUE)
summary(data)
head(data, n=15) # we check we got it right

data$CG_SYST_J_F2<-ave(data$F2, data$ID_Loc, na.rm=TRUE)
summary(data)
head(data, n=15) # we check we got it right

```

```

# we compute the system center of gravity as a function of F1 and F2
# we can compute either EXEMPLAR index or PROTOTYPE index

c<-((data$CG_SYST_J_F1-data$F1))^2 # we compute the system's center of gravity for
F1
d<-((data$CG_SYST_J_F2-data$F2))^2 # same for F2

data$DE_SYST_J <- sqrt(c+d) # we square root the sum of squared F1 and F2 indexes
# and store it as a new variable CG_SYST_J, i.e. the system's center of gravity as a
function of F1 and F2
head(data, n=15) # we check we got it right

#=====
#           B. PROTOTYPE
#=====

####added for ID prototype

data$d_intra_voy<-ave(data$d_intra, data$LexiPhon, data$ID_Loc)
data$d_intra_exemplar<-ave(data$d_intra_voy, data$ID_Loc)
summary(data)
head(data, n=15) # we check we got it right

### boucle

Loclist<-as.list(levels(data$ID_Loc))

for (ID_Loc in Loclist)
{
  print (ID_Loc)
  dataLoc<- data[ which(data$ID_Loc==ID_Loc),] # I extract one mean of DE_SYST_K for
each category per Loc
  loc_d_intra_voy <-unique(dataLoc$d_intra_voy) # I select unique value per voyelle
  print (mean (loc_d_intra_voy)) # I check if there is just one mean per Loc
  data[ which(data$ID_Loc==ID_Loc),"d_intra_mean"] = mean (loc_d_intra_voy) # I
compute the mean and insert the value in the dataframe)
}

###

data$CG_SYST_K_F1<-ave(data$CG_K_F1, data$ID_Loc)
summary(data)
head(data, n=15) # we check we got it right

```

```

data$CG_SYST_K_F2<-ave(data$CG_K_F2, data$ID_Loc)
summary(data)
head(data, n=15) # we check we got it right

data$e<-((data$CG_K_F1-data$CG_SYST_K_F1))^2 # we compute the system's center
of gravity for F1
data$f<-((data$CG_K_F2-data$CG_SYST_K_F2))^2 # same for F2

data$d_inter <- sqrt(data$e +data$f) # we square root the sum of squared F1 and F2
indexes
# and store it as a new variable CG_SYST_K, i.e. the system's center of gravity as a
function of F1 and F2
head(data, n=15) # we check we got it right

# compute the final Distinctivity Index according to PROTOTYPE or EXEMPLAR option

#=====
# EXEMPLAR
# ID(J) = mean DE_SYST_J/mean DE_VOY
#=====

data$DE_SYST_J_mean<-ave(data$DE_SYST_J, data$ID_Loc)
summary(data)
head(data, n=15) # we check we got it right

data$ID_exemplar <- data$DE_SYST_J_mean/data$d_intra_exemplar

#=====
# PROTOTYPE
# ID(K) = mean DE_SYST_K_mean_new/mean d_intra
#=====

Loclist<-as.list(levels(data$ID_Loc))

for (ID_Loc in Loclist)
{
print (ID_Loc)
dataLoc<- data[ which(data$ID_Loc==ID_Loc),] # I extract one mean of DE_SYST_K for
each category per Loc
loc_d_inter <-unique(dataLoc$d_inter) # I select unique value per voyelle
print (mean (loc_d_inter)) # I check if there is just one mean per Loc
data[ which(data$ID_Loc==ID_Loc),"d_inter_mean"] = mean (loc_d_inter) # I compute
the mean and insert the value in the dataframe)
}

```

```

head(data, n=15) # we check we got it right

data$ID_prototype <- data$d_inter_mean/data$d_intra_mean
head(data, n=15) # we check we got it right

# compute Skodda Vowel Index
#  $(F2\ i + F1\ a) / (F1\ i + F1\ u + F2\ u + F2\ a)$ 

# create variables

f_list = c() # we create a list that contains the ID_Skodda for each non duplicated
speaker

locuteurs = as.character(data$ID_Loc)
locuteurs = locuteurs[!duplicated(locuteurs)]
for (l in locuteurs) {

F1a <- mean(data$CG_K_F1[data$LexiPhon == "a" & data$ID_Loc == l])
F1i <- mean(data$CG_K_F1[data$LexiPhon == "i" & data$ID_Loc == l])
F1u <- mean(data$CG_K_F1[data$LexiPhon == "u" & data$ID_Loc == l])
F2a <- mean(data$CG_K_F2[data$LexiPhon == "a" & data$ID_Loc == l])
F2i <- mean(data$CG_K_F2[data$LexiPhon == "i" & data$ID_Loc == l])
F2u <- mean(data$CG_K_F2[data$LexiPhon == "u" & data$ID_Loc == l])

f_list = c(f_list, ((F2i+F1a)/(F1i+F1u+F2u+F2a)))
}

# compute Formant Centralization Ratio (FCR, Sappir 2010)
#  $(F2u + F2a + F1i + F1u) / (F2i + F1a)$ 

# create variables

g_list = c() # we create a list that contains the FCR for each non duplicated speaker
locuteurs = as.character(data$ID_Loc)

locuteurs = locuteurs[!duplicated(locuteurs)]
for (l in locuteurs) {

F1a <- mean(data$CG_K_F1[data$LexiPhon == "a" & data$ID_Loc == l])
F1i <- mean(data$CG_K_F1[data$LexiPhon == "i" & data$ID_Loc == l])
F1u <- mean(data$CG_K_F1[data$LexiPhon == "u" & data$ID_Loc == l])
F2a <- mean(data$CG_K_F2[data$LexiPhon == "a" & data$ID_Loc == l])

```

```

F2i <- mean(data$CG_K_F2[data$LexiPhon == "i"& data$ID_Loc == l])
F2u <- mean(data$CG_K_F2[data$LexiPhon == "u"& data$ID_Loc == l])

g_list = c(g_list, ((F2u + F2a + F1i + F1u)/(F2i + F1a))) }

# compute French adapted Formant Centralization Ratio (aFCR, Sappir 2010)
# aFCR=(F2u+F1i+F1u+F2o + F2O)/(F2i+F1a+F2e+F2E)

# create variables

l_list = c() # we create a list that contains the FCR for each non duplicated speaker
locuteurs = as.character(data$ID_Loc)
locuteurs = locuteurs[!duplicated(locuteurs)]
for (l in locuteurs) {

F1a <- mean(data$CG_K_F1[data$LexiPhon == "a" & data$ID_Loc == l])
F1i <- mean(data$CG_K_F1[data$LexiPhon == "i"& data$ID_Loc == l])
F1u <- mean(data$CG_K_F1[data$LexiPhon == "u"& data$ID_Loc == l])
F2a <- mean(data$CG_K_F2[data$LexiPhon == "a"& data$ID_Loc == l])
F2i <- mean(data$CG_K_F2[data$LexiPhon == "i"& data$ID_Loc == l])
F2u <- mean(data$CG_K_F2[data$LexiPhon == "u"& data$ID_Loc == l])
F2o <- mean(data$CG_K_F2[data$LexiPhon == "o"& data$ID_Loc == l])
F2O <- mean(data$CG_K_F2[data$LexiPhon == "O"& data$ID_Loc == l])
F2e <- mean(data$CG_K_F2[data$LexiPhon == "e"& data$ID_Loc == l])
F2E <- mean(data$CG_K_F2[data$LexiPhon == "E"& data$ID_Loc == l])

l_list = c(l_list, ((F2u+F1i+F1u+F2o+F2O)/(F2i+F1a+F2e+F2E))) }

# compute F2RR (Ratio d'étendue de F2 (cf. Sapir et al., 2010))
# F2RR= F2i/F2u

# create variables

n_list = c() # we create a list that contains the FCR for each non duplicated speaker
locuteurs = as.character(data$ID_Loc)
locuteurs = locuteurs[!duplicated(locuteurs)]
for (l in locuteurs) {

F2i <- mean(data$CG_K_F2[data$LexiPhon == "i"& data$ID_Loc == l])
F2u <- mean(data$CG_K_F2[data$LexiPhon == "u"& data$ID_Loc == l])

n_list = c(n_list, (F2i/F2u)) }

```

```

# compute F1RR (Ratio d'étendue de F1 (cf. Sapir et al., 2010))
# F1RR= 2 F1a/(F1i+F1u)

# create variables

m_list = c() # we create a list that contains the FCR for each non duplicated speaker
locuteurs = as.character(data$ID_Loc)
locuteurs = locuteurs[!duplicated(locuteurs)]
for (l in locuteurs) {

  F1a <- mean(data$CG_K_F1[data$LexiPhon == "a" & data$ID_Loc == l])
  F1i <- mean(data$CG_K_F1[data$LexiPhon == "i"& data$ID_Loc == l])
  F1u <- mean(data$CG_K_F1[data$LexiPhon == "u"& data$ID_Loc == l])

  m_list = c(m_list, (2*F1a/(F1i+F1u))) }

#=====
#      Formant Centralization Ratio
#=====

# compute tendency towards prototypes /iua/ Ratio
# (F1E+F1O+F2O)/(F2E+F1e+F1o)
# F2E+F1e+F1o/F1E+F1O+F2O

# create variables

z_list = c() # we create a list that contains the FCR for each non duplicated speaker
locuteurs = as.character(data$ID_Loc)

locuteurs = locuteurs[!duplicated(locuteurs)]
for (l in locuteurs) {

  F1E <- mean(data$CG_K_F1[data$LexiPhon == "E" & data$ID_Loc == l])
  F1O <- mean(data$CG_K_F1[data$LexiPhon == "O"& data$ID_Loc == l])
  F2O <- mean(data$CG_K_F2[data$LexiPhon == "O"& data$ID_Loc == l])
  F2E <- mean(data$CG_K_F2[data$LexiPhon == "E"& data$ID_Loc == l])
  F1e <- mean(data$CG_K_F1[data$LexiPhon == "e"& data$ID_Loc == l])
  F1o <- mean(data$CG_K_F1[data$LexiPhon == "o"& data$ID_Loc == l])

  z_list = c(z_list, (F1E+F1O+F2O)/(F2E+F1e+F1o)) }

### compute pVSA

```

```

locuteurs = as.character(data$ID_Loc)

hull.area <- with(data, convexHullArea(F1, F2, group=locuteurs))
poly.area <- with(data, vowelMeansPolygonArea(F1, F2, LexiPhon,
      poly.order=c("i", "e", "a", "o", "u"), group=locuteurs))

a_list = c()
a_list = with(data, vowelMeansPolygonArea(F1, F2, LexiPhon,
      poly.order=c("i", "e", "a", "o", "u"), group=locuteurs))
locuteurs = locuteurs[!duplicated(locuteurs)]

#=====

# save raw data in a csv format
# this data contains all the computations
summary(data)
write.table(data, "OutPut_1.csv", sep = ";", row.names=FALSE)

# save data with different ID per speaker (simple dataset)
locut_data=data[, c("ID_Loc", "ID_exemplar", "ID_prototype")]
print(locut_data) #just checking

# eliminate the duplicates
locut_data = locut_data[!duplicated(locut_data$ID_Loc),]

# add the ID_Skodda to the locut_data set
# first, create an empty column in this dataset
locut_data ["ID_Skodda"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {
  if (locut_data[i,"ID_Loc"] == locuteurs[i])
    locut_data[i,"ID_Skodda"] = f_list[i] }

# add the FCR to the locut_data set
# first, create an empty column in this dataset
locut_data ["FCR"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {
  if (locut_data[i,"ID_Loc"] == locuteurs[i])
    locut_data[i,"FCR"]=g_list[i] }

```

```

# add the aFCR to the locut_data set
# first, create an empty column in this dataset
locut_data ["aFCR"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {
  if (locut_data[i,"ID_Loc"] == locuteurs[i])
    locut_data[i,"aFCR"]=l_list[i] }

# add the F2RR to the locut_data set
# first, create an empty column in this dataset
locut_data ["F2RR"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {
  if (locut_data[i,"ID_Loc"] == locuteurs[i])
    locut_data[i,"F2RR"]=n_list[i] }

# add the F1RR to the locut_data set
# first, create an empty column in this dataset
locut_data ["F1RR"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {
  if (locut_data[i,"ID_Loc"] == locuteurs[i])
    locut_data[i,"F1RR"]=m_list[i]      }

# add the ProVI to the locut_data set
# first, create an empty column in this dataset
locut_data ["ProVI"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {
  if (locut_data[i,"ID_Loc"] == locuteurs[i])
    locut_data[i,"ProVI"]=z_list[i]      }

# add the pVSA to the locut_data set
# first, create an empty column in this dataset
locut_data ["pVSA"] = 0

# then, we fill with with ordered values from the f_list and locuteurs
for (i in 1:length(locuteurs)) {

```

```

if (locut_data[i,"ID_Loc"] == locuteurs[i])
  locut_data[i,"pVSA"]=a_list[i] }

# save
write.table(locut_data, "OutPut_2.csv", sep = ";", row.names=FALSE)

# correlation between IDs

m<-cor(locut_data$ID_exemplar, locut_data$ID_prototype)
summary(m)
cor(locut_data$ID_exemplar, locut_data$ID_Skodda)
cor(locut_data$ID_prototype, locut_data$ID_Skodda)
cor(locut_data$ID_prototype, locut_data$ID_exemplar)
plot(locut_data$ID_exemplar, locut_data$ID_prototype)
plot(locut_data$ID_exemplar, locut_data$ID_Skodda)
plot(locut_data$ID_prototype, locut_data$ID_Skodda)
plot(locut_data$ID_prototype, locut_data$aFCR)
plot(locut_data$ID_prototype, locut_data$FCR)
plot(locut_data$ID_prototype, locut_data$F2RR)

cor(locut_data$ID_prototype, locut_data$F2RR)
cor(locut_data$ID_prototype, locut_data$aFCR)
cor(locut_data$ID_prototype, locut_data$F1RR)

```