



HAL
open science

Les mesures d'intelligibilité: état de l'art, considérations pratiques pour l'applicabilité clinique et explorations acoustiques

Timothy Pommée

► To cite this version:

Timothy Pommée. Les mesures d'intelligibilité: état de l'art, considérations pratiques pour l'applicabilité clinique et explorations acoustiques. Sciences de l'information et de la communication. Université Paul Sabatier - Toulouse III, 2021. Français. NNT: 2021TOU30141 . tel-04136563

HAL Id: tel-04136563

<https://theses.hal.science/tel-04136563>

Submitted on 21 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Timothy POMMEE

Le 30 novembre 2021

**Les mesures d'intelligibilité : état de l'art, considérations
pratiques pour l'applicabilité clinique et explorations acoustiques**

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et
Télécommunications de Toulouse**

Spécialité : **Informatique et Télécommunications**

Unité de recherche :

IRIT : Institut de Recherche en Informatique de Toulouse

Thèse dirigée par

Julien PINQUIER et Virginie WOISARD

Jury

Mme Nathalie HENRICH BERNARDONI, Rapporteure

M. Marc DE BODT, Rapporteur

Mme Julie MAUCLAIR, Examinatrice

Mme Dominique MORSOMME, Examinatrice

M. Julien PINQUIER, Directeur de thèse

Mme Virginie WOISARD, Co-directrice de thèse

Remerciements

J'adresse mes remerciements à tous ceux qui ont rendu possible la réalisation de ce travail.

En premier lieu, je remercie mon directeur de thèse, Julien Pinquier, ma co-directrice de thèse, Virginie Woisard, ainsi que ma co-encadrante, Julie Mauclair, pour leur confiance et le temps qu'ils m'ont accordé, pour leurs conseils avisés et surtout pour leur patience. Je leur suis également très reconnaissant de parvenir, malgré la charge de travail qui leur incombe, à toujours garder la banane (☺) et à offrir à leurs étudiants un cadre motivant et encourageant.

Malgré les conditions particulières liées à la pandémie, vous êtes parvenus à maintenir un suivi hebdomadaire de qualité, toujours plein d'humour mais aussi d'opinions individuelles – subjectives donc ! – qui ont permis d'alimenter de riches débats et de contribuer à ma maturation professionnelle et scientifique.

Je voudrais également remercier Nathalie Henrich-Bernardoni et Marc de Bodt d'avoir accepté de lire cette thèse et d'en être rapporteurs. Je remercie Dominique Morsomme d'avoir accepté d'être présidente du jury. Enfin, merci à tous les membres du jury d'avoir accepté d'assister à la présentation de ce travail.

La réalisation de cette thèse ainsi que les voyages et échanges enrichissants qui en ont fait partie ont été rendus possibles grâce au financement octroyé par le programme de recherche et d'innovation Horizon 2020 de l'Union européenne dans le cadre de la convention de subvention Marie Skłodowska-Curie n° 766287. Merci aussi à tous les étudiants doctorants et partenaires de ce projet pour les bons moments passés ensemble.

Mes remerciements sont aussi adressés à Renée Speyer de l'Université d'Oslo, qui m'a formé à la revue systématique de la littérature et à la méthodologie Delphi.

Un grand merci aussi à tous les membres de l'équipe SAMoVA de l'IRIT de m'avoir accueilli à bras ouverts et d'avoir partagé avec moi de nombreuses aventures au fil de ces trois années. Plus que des collègues, vous êtes rapidement devenus de vrais amis. Merci notamment à Dupond pour les superbes moments passés en sa compagnie aux quatre coins du monde ainsi que pour les nombreuses collaborations et l'entraide qui m'ont permis d'en arriver à ce résultat final.

Je tiens aussi à exprimer ma gratitude envers les membres du CHU et de l'oncopole de Toulouse qui ont accepté de me consacrer de leur temps pour les enregistrements de locuteurs.

Mes remerciements s'adressent bien sûr aux membres du Laboratoire d'Informatique d'Avignon, du Laboratoire Parole Langage d'Aix-Marseille Université et du Laboratoire de Neuropsycholinguistique de l'Université Toulouse II Jean-Jaurès pour les nombreuses collaborations et l'aide qu'ils m'ont apportée dans mes différents projets.

Merci également à tous les participants à mes diverses études, cliniciens, experts de la parole, locuteurs, en France et dans de nombreux autres pays, pour le temps qu'ils ont accordé à mes études.

Enfin, je tiens également à exprimer mes remerciements à tous mes proches, mes amis et ma famille ainsi que plus particulièrement à Adrien, qui m'ont aidé de près ou de loin à mener à bien ce travail, en m'encourageant jusqu'au bout ainsi qu'en me permettant de décompresser au besoin.

Table des matières

LISTE DES FIGURES	3
LISTE DES TABLEAUX	3
INTRODUCTION.....	8
CHAPITRE 1. INTELLIGIBILITE	15
1.1. ÉTAT DE L'ART	15
1.2. INTELLIGIBILITE ET COMPREHENSIBILITE : DEFINITION ET MESURES.....	17
1.2.1. <i>Choix de la méthode</i>	17
1.2.2. <i>La méthode Delphi</i>	18
1.2.3. <i>Notre étude Delphi</i>	19
CHAPITRE 2. ANALYSE DES BESOINS	39
2.1. L'ÉVIDENCE-BASED PRACTICE	39
2.2. PREMIER PILIER DE L'EBP : LA RECHERCHE SCIENTIFIQUE.....	40
2.2.1. <i>Les revues systématiques de la littérature</i>	40
2.2.2. <i>Besoins en termes de mesure d'intelligibilité identifiés dans la littérature</i>	42
2.3. DEUXIEME PILIER DE L'EBP : L'EXPERIENCE CLINIQUE ET PRATIQUE	55
2.3.1. <i>Évaluation des troubles de la parole chez l'adulte – situation actuelle et besoins en pratique clinique francophone</i>	55
2.3.2. <i>Évaluation de la parole chez l'adulte – opinions d'experts</i>	75
2.4. TROISIEME PILIER DE L'EBP : L'AVIS DES PATIENTS.....	86
2.5. CONCLUSION DU CHAPITRE : BESOINS EN TERMES DE MESURE DE LA PAROLE	91
CHAPITRE 3. SUPPORTS D'ÉVALUATION ET ÉCHANTILLONS DE PAROLE	94
3.1. SUPPORTS ET ÉCHANTILLONS – ÉTAT DE L'ART EN PRATIQUE FRANCOPHONE.....	94
3.1.1. <i>Outils et supports disponibles en langue française</i>	94
3.1.2. <i>Unités phonémiques et textes de référence</i>	99
3.2. NOUVEAU TEXTE DE RÉFÉRENCE	102
3.2.1. <i>Identification des besoins et définition de l'objectif</i>	103
3.2.2. <i>Recensement, sélection et hiérarchisation des critères de construction</i>	105
3.2.3. <i>Conception du texte</i>	113
3.3. CONCLUSION DU CHAPITRE.....	116
CHAPITRE 4. TRAITEMENT DU SIGNAL	117
4.1. ÉTAT DE L'ART – MESURES ACOUSTIQUES DE L'INTELLIGIBILITE DE LA PAROLE	117
4.2. EXPLORATION ACOUSTIQUE PRELIMINAIRE SUR LES CONSONNES : LA « BANANE DE LA PAROLE » ...	121
4.2.1. <i>Motivation</i>	121
4.2.2. <i>Méthodologie</i>	122
4.2.3. <i>Résultats</i>	125
4.2.4. <i>Discussion</i>	126
4.3. LES MOMENTS SPECTRAUX SUR LES CONSONNES NON-SONANTES EN FRANÇAIS.....	127
4.3.1. <i>Moments spectraux sur segmentation manuelle</i>	130
4.3.2. <i>Moments spectraux sur alignement forcé</i>	147
4.3.3. <i>Moments spectraux en parole pathologique</i>	158
4.4. CONCLUSION DU CHAPITRE.....	171
CHAPITRE 5. ESSAI PILOTE DE MESURES ACOUSTIQUES SUR NOTRE NOUVEAU TEXTE... 172	172
5.1. METHODOLOGIE	172
5.2. RESULTATS ET DISCUSSION	175
5.2.1. <i>Données descriptives</i>	175
5.2.2. <i>Objectif 1 : Reproductibilité</i>	179
5.2.3. <i>Objectif 2 : Vitesse de lecture</i>	181
5.2.4. <i>Objectifs secondaires</i>	182
5.3. CONCLUSION DU CHAPITRE.....	184
CONCLUSION GENERALE	185
CONCLUSION.....	185
PERSPECTIVES.....	189
BIBLIOGRAPHIE.....	192
ANNEXES	221

Liste des figures

Figure 1 – Organigramme du déroulement de notre étude Delphi	20
Figure 2 – Mesures perceptives qui caractérisent au mieux l’intelligibilité de la parole (gris foncé) et la compréhension (gris clair) ; pour faciliter la visualisation, les résultats ont été classés par ordre décroissant pour les mesures d’intelligibilité ; p.-m. : pseudomots, phr. non-pred. : phrases non prédictibles	29
Figure 3 – Mesures « objectives » qui caractérisent au mieux l’intelligibilité de la parole (gris foncé) et la compréhension (gris clair).....	30
Figure 4 – L’intelligibilité et la compréhension dans la production de la parole.....	34
Figure 5 – Modèle interactionnel de la communication de [SCHRAMM 54].....	34
Figure 6 – Boucle de communication orale ; en gris : étapes relatives à l’intelligibilité.....	36
Figure 7 – Les trois principaux piliers de l’EBP	39
Figure 8 – Prise en compte du premier pilier de l’EBP : revue systématique de la littérature	40
Figure 9 – Les étapes de la revue systématique de la littérature ; les chiffres verts indiquent les cinq étapes clés. Adapté et traduit de Kaufman, 2011	41
Figure 10 – Diagramme de flux illustrant le processus de sélection selon les lignes directrices PRISMA	46
Figure 11 – Prise en compte du deuxième pilier de l’EBP : enquête adressée aux cliniciens.....	55
Figure 12 – Composantes de la CIF (Organisation mondiale de la santé 2001)	56
Figure 13 – Distribution des participants dans le monde (gauche) et en France métropolitaine, en Belgique et en Suisse (droite). Les points en dehors de ces pays se trouvent en Algérie, en Tunisie et dans les DOM-TOM.	60
Figure 14 – Types de microphones utilisés pour les bilans des TPP ; « autres » : microphones sur pied, dictaphones et micros-cravates.....	63
Figure 15 – Scores recueillis lors de l’évaluation des troubles de la parole.....	64
Figure 16 – Distribution des niveaux de satisfaction attribués aux outils d’évaluation de la parole.....	65
Figure 17 – Manques rapportés et solutions souhaitées par les cliniciens concernant l’évaluation des troubles de la parole.....	67
Figure 18 – Synthèse des besoins cliniques en matière d’évaluation de la parole identifiés dans notre enquête	73
Figure 19 – Prise en compte des premier et deuxième piliers de l’EBP : interviews semi-dirigées d’experts de la parole.....	75
Figure 20 – Satisfaction avec les outils d’évaluation perceptive (N=8).....	80
Figure 21 – Satisfaction avec les outils d’évaluation « objective » (N=9)	81
Figure 22 – Prise en compte du troisième pilier de l’EBP : questionnaire adressé aux patients	86
Figure 23 – Synthèse des études menées pour identifier les besoins en termes de mesure de l’intelligibilité selon le modèle EBP	91
Figure 24 – Échelle de granularité des unités de production de la parole	99
Figure 25 – Processus de création du nouveau texte pour l’évaluation de la parole et de la voix.....	103
Figure 26 – Classement moyen (RM) de chacun des critères de création de texte (incluant les non-classements) ; en noir : le plus et le mieux classé (RMN <7,5 et noté par au moins 50 % des participants) ; en gris foncé : le moins et le moins bien classé (RMN >10 et noté par moins de 25 % des participants) ; en hachuré : résultat clivé entre le nombre d’occurrences et le RMN	109
Figure 27 – Exemple de représentations de la trajectoire des formants depuis laquelle est calculé l’AAVS (tiré de [WHITFIELD 17]. Ces trajectoires ont été extraites d’une phrase prononcée par un homme en parole suraccentuée (haut) et en parole conversationnelle (bas). Les cercles noirs représentent les trajectoires F1-F2, et la couleur de fond représente la distribution de la densité (bleu = faible/aucune densité ; rouge = forte densité)	119
Figure 28 – Représentation de l’analyse de différence de la pente F2 aux frontières consonantiques (tiré de Van Son and Pols 1999) ; la pente de la frontière VC est soustraite de la pente de la frontière CV	120
Figure 29 – La « banane de la parole » utilisée en audiométrie pour évaluer l’impact du déficit auditif sur la perception de la parole ; récupéré sur https://hearinghealthcare.org/blog/the-audiogram-and-what-it-means/	122
Figure 30 – Exemple d’analyse préliminaire de la stabilité des spectres et de l’influence du contexte vocalique sur la consonne /ʃ/ prononcée par une locutrice	124

Figure 31 – Spectres LPC de la voyelle /u/ pour un ordre de prédiction de 8, 10, 12, 14, 16 et 17 [DISSEN 19]...	125
Figure 32 – « Bananes de la parole » calculées pour les six locuteurs (en haut, en orange) et pour les neuf locutrices (en bas, en bleu), montrant la moyenne des principaux pics spectraux des consonnes sur un graphique à deux axes (Intensité x Fréquence).....	126
Figure 33 – Représentation schématique des quatre moments spectraux	128
Figure 34 – Représentation schématique de courbes avec un SKEW négatif (gauche) et positif (droite), en comparaison à une courbe symétrique (SKEW de 0) ; traduit et adapté de https://analystprep.com/cfa-level-1-exam/quantitative-methods/unimodal-distribution-mean-mode-median/	128
Figure 35 – Graphiques en boîtes à moustaches pour les quatre moments spectraux au point médian des consonnes plosives et fricatives ; les boîtes bleues représentent l'écart interquartile, la ligne bleue à l'intérieur des boîtes la médiane, le cercle jaune la moyenne, les moustaches incluent tous les points situés dans un intervalle de 1,5 EIQ du quartile supérieur/inférieur, les points bleus représentent les valeurs qui se trouvent en dehors de ces limites.....	132
Figure 36 – Exemple de comparaison de valeurs mesurées en chambre anéchoïque (« anech ») et en pièce calme (« calm ») pour le CoG des plosives.....	133
Figure 37 – Évolution par fenêtre temporelle des moments spectraux dans les fricatives non voisées et voisées ; les étoiles indiquent les différences « horizontales » statistiquement significatives entre les fenêtres d'analyse ($\alpha = 0,0167$) ; les flèches verticales indiquent les différences « verticales » statistiquement significatives entre les lieux d'articulation	134
Figure 38 – Évolution par fenêtre temporelle des moments spectraux dans les plosives non voisées et voisées....	135
Figure 39 – Résumé des distinctions entre les lieux d'articulation basées sur les moments spectraux dans les fricatives (2 diagrammes supérieurs) et dans les plosives (2 diagrammes inférieurs) ; en vert : CoG et SKEW, en bleu : SD et KURT, paires respectivement liées théoriquement.....	136
Figure 40 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les fricatives ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres.....	139
Figure 41 – Représentation schématique du plateau (bleu clair) et des composantes spectrales principales à l'intérieur de celui-ci (bleu foncé) ainsi que du CoG (rouge), pour le /s/ (haut) et pour le /ʃ/ (bas)	141
Figure 42 – Spectre plat de la fricative labiodentale /f/	142
Figure 43 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les plosives ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres	143
Figure 44 – Composantes spectrales principales pour /t/ (gauche) et pour /k/ (droite).....	143
Figure 45 – Spectre plat de la plosive bilabiale /p/	144
Figure 46 – Représentation schématique de l'alignement forcé	147
Figure 47 – Exemples de résultat d'alignement forcé avec inclusion de la transition vocalique dans le segment consonantique	148
Figure 48 – Résultat de la segmentation subphonémique effectuée à l'aide de la méthode FBDS.....	149
Figure 49 – Fusion de l'alignement forcé et de la méthode FBDS	149
Figure 50 – Exemple d'un burst détecté par notre algorithme pour la plosive /t/ (en rouge, la rupture correspondant au début du burst, ainsi que la fenêtre d'analyse de 5 ms qui la suit).....	150
Figure 51 – Exemple d'un burst détecté par notre algorithme pour la plosive /d/ (en rouge, la rupture correspondant au début du burst, ainsi que la fenêtre d'analyse de 5 ms qui la suit)	150
Figure 52 – Exemple d'une fenêtre d'analyse de 10 ms (en rouge) identifiée par notre algorithme pour la fricative /f/, à l'intérieur de la phase stable de la consonne (en jaune).....	151
Figure 53 – Exemple d'une fenêtre d'analyse de 10 ms (en rouge) identifiée par notre algorithme pour la fricative /ʒ/ à l'intérieur du segment le plus faible en intensité.....	151
Figure 54 – Graphiques en boîtes à moustaches pour les quatre moments spectraux pour les consonnes plosives et fricatives, issus de la méthode FBDS	153
Figure 55 – Exemple d'inclusion de transition vocalique dans l'alignement forcé d'une fricative voisée, corrigée par notre algorithme	154
Figure 56 – Résumé des distinctions entre les lieux d'articulation basées sur les moments spectraux issus de notre algorithme, dans les fricatives (2 diagrammes supérieurs) et dans les plosives (2 diagrammes inférieurs) ; en vert : CoG et SKEW, en bleu : SD et KURT, paires respectivement liées théoriquement	155
Figure 57 – Représentation schématique des valeurs des moments spectraux en parole pathologique par rapport aux valeurs des sujets sains	160

Figure 58 – Données descriptives des moments spectraux issus des enregistrements de sujets sains et de patients traités pour un cancer ORL, pour les fricatives (gauche) et pour les plosives (droite) ; les points situés à plus de deux fois le quartile 99 ont été exclus pour éviter une échelle trop large pour SKEW (1 point exclu >8,52) et KURT (6 points exclus >44,27).....	162
Figure 59 – Sévérité moyenne par patient.....	163
Figure 60 – Fréquences de réponse pour les trois niveaux de difficulté de transcription sur les quatre sessions d'écoute successives (order 1 = premier groupe d'écoute, order 4 = dernier groupe d'écoute)	164
Figure 61 – Pourcentages de transcriptions correctes par patient	165
Figure 62 – Exemple d'un mauvais alignement dû à une très faible distinctivité des phonèmes du pseudomot /ø g ø/	165
Figure 63 – Valeurs des moments spectraux du /p/ de CMB11 (ligne rouge) par rapport aux fricatives non voisées des sujets sains.....	166
Figure 64 – Représentation schématique de la centralisation des valeurs de SD et de Kurt pour le /f, v/ et /s/ du patient CMH25	167
Figure 65 – Représentation schématique des valeurs de moments spectraux du /s/ pour le patient CMH25 (en rouge), au regard des valeurs de /s, v, ʒ/ chez les sujets sains (en pointillés)	167
Figure 66 – Pour PFG13, distinctions phonémiques par les moments spectraux préservées (en vert) et non-préservées (en noir)	167
Figure 67 – Représentation des plosives des sujets sains dans un espace à trois dimensions, grâce à la technique t-SNE (créée par Wessel Stoop de l'Université de Radboud) ; disponible à l'adresse : https://cls.ru.nl/staff/wstoop/moments/interactive.html	170
Figure 68 – Exemple fictif d'une courbe standard du phonème /s/ (en pointillé : CoG ; en orange pâle : SD et KURT ; en bleu et mauve : SKEW) avec la courbe de production d'un patient superposée (en rouge) ainsi que des probabilités d'identité phonémique ; dans la partie supérieure droite, représentation en curseur des quatre moments spectraux (en vert, la zone des valeurs de référence pour le phonème cible, en rouge, la valeur de la production du patient) avec des pistes d'interprétation (en bleu).....	170
Figure 69 – Exemple de triangle vocalique obtenu de la lecture du texte entier par une locutrice belge. F1 en ordonnées (haut-bas), F2 en abscisses (droite-gauche), en demi-tons ; ligne discontinue : triangle de référence ; pointillés : triangle employé par le locuteur (1 E.T.) ; ligne continue : triangle employé par le locuteur (2 E.T.) ; « x » : voyelles cardinales et centroïde (moyenne géométrique des fréquences des voyelles cardinales) ; en gris, pour chaque segment vocalique, le point le plus proche du centroïde (non employé dans cette étude) ; 1 = surface relative du triangle 1 E.T. ; 2 = surface relative du triangle 2 E.T. ; N = nombre total de segments vocaliques	173
Figure 70 – Représentation schématique des valeurs des moments spectraux pour les fricatives dans le texte par rapport à celles extraites des phrases porteuses	177
Figure 71 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les fricatives dans le texte ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres	177
Figure 72 – Représentation schématique des valeurs des moments spectraux pour les plosives dans le texte par rapport à celles extraites des phrases porteuses	178
Figure 73 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les plosives dans le texte ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres.....	178
Figure 74 – Boxplots des valeurs du CPPs (en dB) au fil des cinq lectures.....	180
Figure 75 – Boxplots des valeurs de l'aire relative du triangle vocalique (en %) au fil des cinq lectures	180
Figure 76 – Boxplots des valeurs du CoG (en Hz) pour la consonne /b/ au fil des cinq lectures	181
Figure 77 – Boxplots des valeurs de la pente de F2 pour la semi-consonne /j/ au fil des cinq lectures	181
Figure 78 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux (c.-à-d. du burst) pour les plosives non voisées ; AligDroite =frontière droite du segment consonantique de l'alignement forcé ...	292
Figure 79 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux (c.-à-d. du burst) pour les plosives voisées ; AligDroite =frontière droite du segment consonantique de l'alignement forcé	293
Figure 80 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux pour les fricatives non voisées	293
Figure 81 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux pour les fricatives voisées	294

Liste des tableaux

Tableau 1 – Énoncés présentés au tour 2 et pourcentages d'accord.....	27
Tableau 2 – Stratégie de recherche pour les deux bases de données.....	44
Tableau 3 – Évaluation de la qualité méthodologique des 22 articles inclus à l'aide de l'outil d'évaluation critique Qualsyst de [KMET 04] et du niveau de preuve selon la hiérarchie du National Health and Medical Research Council [NHMRC 99].....	46
Tableau 4 Répartition des genres et des âges dans les études retenues dans notre revue systématique	47
Tableau 5 – Corrélations significatives et non significatives entre mesures acoustiques et évaluations perceptives de la parole	50
Tableau 6 – Description des orthophonistes ayant participé à l'enquête.....	61
Tableau 7 – Description des patients atteints de troubles de la parole	62
Tableau 8 – Description du panel d'experts ayant participé aux interviews semi-dirigées.....	76
Tableau 9 – Activités professionnelles et niveaux d'éducation des participants aux interviews semi-dirigées	77
Tableau 10 – Description des patients ayant répondu au questionnaire concernant l'évaluation de leur trouble de la parole	87
Tableau 11 – Données concernant les bilans de parole	89
Tableau 12 – Besoins et manques identifiés à travers les quatre études menées concernant les mesures de la parole, solutions envisagées et contributions concrètes dans le cadre de cette thèse.....	92
Tableau 13 – Unités ciblées dans les outils d'évaluation francophones.....	97
Tableau 14 – Énoncés concernant la création de textes présentés au deuxième tour de l'enquête Delphi et pourcentages d'accord	109
Tableau 15 – Critères de construction contrôlés dans la première partie du texte	114
Tableau 16 – Médiane [intervalle interquartile] des moments spectraux pour chaque phonème.....	136
Tableau 17 – Coefficients de corrélation de Spearman entre les moments spectraux et probabilités	137
Tableau 18 – Comparaison des spectres des bursts des plosives dans notre étude (rangée supérieure) avec les patterns identifiés par [BLUMSTEIN 79] (rangée du bas).....	145
Tableau 19 – Médiane [intervalle interquartile] des moments spectraux pour chaque phonème ; en gras : valeurs significativement différentes des autres sur la même demi-rangée	155
Tableau 20 – Coefficients de corrélation de Spearman entre les moments spectraux issus de la méthode FBDS et probabilités	156
Tableau 21 – Description des patients traités pour un cancer ORL inclus dans notre étude	158
Tableau 22 – Données descriptives des mesures CPPs, slope et tilt pour les six lectures de texte par les 17 locuteurs	175
Tableau 23 – Données descriptives de l'aire relative du triangle vocalique pour les six lectures par les 17 locuteurs	176
Tableau 24 – Médiane [intervalle interquartile] des moments spectraux pour chaque phonème de la lecture du nouveau texte ; en gras : valeurs significativement différentes des autres sur la même demi-rangée	176
Tableau 25 – Médiane [intervalle interquartile] des pentes formantiques des semi-consonnes dans le nouveau texte	179
Tableau 26 – Données descriptives des mesures CPPs, slope et tilt pour les deux occurrences de la phrase entièrement voisée lors de la lecture du texte entier par les 17 locuteurs	182
Tableau 27 – Données descriptives des mesures pour lesquelles une différence significative de genre a été mesurée.....	183
Tableau 28 – Données descriptives des mesures CPPs, slope et tilt pour les locuteurs belges, canadiens et français	183

*« Instrumental assessment of pathological speech is of limited use unless normal variability
can be quantitatively defined. »*

[BUTCHER 89]

*« Measurements can be no better than the knowledge and skills of the clinician who chooses
and obtains them. »*

« Never trust a computer completely. »

[BAKEN 00]

Introduction

Problématique

La parole est une fonction essentielle de la vie quotidienne qui nécessite des interactions complexes entre la génération de la pression de l'air, la vibration des plis vocaux et la modulation par les cavités de résonance de l'appareil phonatoire [FITCH 00; HONDA 08]. Ne pas parvenir à se faire comprendre peut avoir un impact sur l'éducation et la formation, limiter la participation professionnelle et sociale, et ainsi réduire la qualité de vie [HUSTAD 08].

Les troubles de la parole peuvent se manifester dans de nombreux contextes et provenir de mécanismes physiopathologiques variés, fonctionnels ou organiques [DUFFY 16]. Ils peuvent apparaître au cours du développement (p. ex. trouble articulaire, fente vélopalatine, surdité), mais également être de nature acquise, chez l'enfant ou chez l'adulte [KENT 00]. Un trouble de la parole acquis peut par exemple survenir à la suite d'une atteinte neurologique (p. ex. accident vasculaire cérébral, traumatisme crânien ou pathologie neurodégénérative) ou de modifications structurelles (p. ex. à la suite d'une intervention chirurgicale). Au regard de la diversité des étiologies des troubles de la parole, ceux-ci touchent une partie conséquente de la population, d'autant plus grandissante vu le vieillissement de cette dernière. En effet, le vieillissement est accompagné de modifications physiologiques de la parole [KURUVILLA-DUGDALE 20], mais aussi d'une apparition plus fréquente de pathologies telle la maladie de Parkinson [DORSEY 07]. La prévalence des troubles de la parole ainsi que les conséquences importantes que ceux-ci peuvent avoir sur la qualité de vie des individus qui en souffrent [PIACENTINI 14; WALSH 11] en font une problématique d'intérêt croissant, autant en recherche scientifique que dans les domaines applicatifs. Cet intérêt se retrouve en première ligne dans le domaine de la santé, avec un besoin clinique grandissant concernant l'évaluation et la prise en charge des troubles de la parole. Il ne se limite cependant pas aux domaines médical et paramédical, mais touche également toutes les disciplines associées, dont le domaine émergent de l'informatique.

Contexte TAPAS

Dans ce contexte est né le projet européen Training Network on Automatic Processing of Pathological Speech (TAPAS)¹. Ce réseau pluridisciplinaire Horizon 2020 est financé par le programme Actions Marie Skłodowska-Curie de l'Union européenne. Il rassemble 15 jeunes chercheurs d'horizons variés (ingénieurs, linguistes, informaticiens et cliniciens) ainsi que de nombreux partenaires académiques et industriels, dans l'objectif général d'une amélioration de la qualité de vie des patients atteints de troubles de la parole. Ce réseau facilite, notamment pour les doctorants qui en font partie, les collaborations interdisciplinaires et internationales avec d'autres chercheurs dans le domaine de la parole. Ces échanges sont d'autant plus sollicités que chaque doctorant est tenu d'effectuer une mobilité de minimum deux mois auprès d'un des partenaires du projet. Ce projet européen offre de plus de nombreuses opportunités de formation, notamment par les séminaires (« training events ») organisés tous les 6 mois par l'un de ses partenaires. Ceux-ci visent non seulement à développer les connaissances théoriques et appliquées des jeunes chercheurs (p. ex. en s'assurant que chacun d'eux a une connaissance de base des troubles de la parole et de la prise en charge orthophonique, ainsi que de l'informatique

¹ <https://www.tapas-etn-eu.org/>

et de l'éthique de la recherche), mais aussi à les préparer pour l'après-thèse (p. ex. grâce à un séminaire portant sur l'entrepreneuriat). Enfin, le réseau TAPAS encourage fortement la participation de ses membres à des conférences internationales pour la diffusion des savoirs et le développement de réseaux professionnels.

Concernant les travaux menés, le projet se décline en trois volets : (a) la « détection », visant le développement de nouvelles techniques de traitement du signal audio pour la détection précoce et non invasive des troubles de la parole ; (b) la « thérapie », dont l'objectif est le développement de nouveaux outils de prise en charge automatisés ; (c) l'« assistance à l'autonomie à domicile », visant à améliorer les outils de domotique basés sur le traitement du signal audio pour améliorer l'autonomie des patients atteints de troubles de la parole.

Objectif

C'est dans le deuxième volet de ce programme international, celui de la « thérapie », que s'inscrit la présente thèse. Avec un regard pluridisciplinaire, alliant les domaines du traitement du signal et de l'orthophonie, elle vise à étudier et contribuer à l'**amélioration des outils de mesure acoustique de la parole**. En effet, une évaluation exhaustive et une quantification du déficit sont nécessaires afin de déterminer le degré général de l'altération et de cibler les objectifs du plan thérapeutique pour une prise en charge individualisée et efficace. De plus, une évaluation fiable est requise pour pouvoir évaluer l'efficacité thérapeutique, ainsi que l'évolution pathologique éventuelle [KENT 92; MILLER 13; STIPANCIC 16; SUSSMAN 12]. D'une part, ce travail de recherche sur la mesure acoustique de la parole répond à cette nécessité d'une évaluation exhaustive et fiable de la parole en pratique clinique. D'autre part, il répond également au besoin en recherche scientifique d'une bonne connaissance des mesures acoustiques et de leur comportement en parole saine avant leur application à la parole pathologique.

Dans le vaste domaine des troubles de la parole et de leurs manifestations variées, ce travail ciblera plus spécifiquement la **mesure de l'intelligibilité**, que nous définirons dans le premier chapitre. Au vu de la grande hétérogénéité des étiologies, des symptômes et des degrés de sévérité des atteintes pouvant provoquer des troubles de la parole, ce travail concernera plus particulièrement l'évaluation des troubles de la parole chez l'**adulte**. Bien que toujours très vaste, ce domaine n'englobe donc que les troubles survenant alors que les systèmes cognitif, moteur et sensoriel sont arrivés à leur état final de maturation. Ceci ne sous-entend bien évidemment pas un caractère forcément stable des troubles de la parole chez l'adulte (p. ex. détérioration dans les pathologies neurodégénératives) et n'empêche pas une grande variabilité dans les manifestations rencontrées. En effet, nous observons déjà une importante variabilité inter- et intra-individuelle de la production de la parole, même dans un contexte « sain » [KENT 83; LINDBLOM 90]. Cette variabilité peut se retrouver ensuite accentuée par la pathologie. C'est pourquoi les analyses expérimentales de cette thèse sont, dans un premier temps, appliquées à la **parole saine**. Elles sont ensuite investiguées en parole pathologique, en ciblant plus précisément les troubles de la parole chez les adultes traités pour un **cancer de la sphère oropharyngée**.

Plan

Le premier chapitre traitera de l'intelligibilité de la parole, qui est l'un des concepts phares de ce travail. Nous nous pencherons d'abord, dans un bref état de l'art, sur l'importante diversité terminologique observée dans le domaine de l'évaluation de la parole, associée à une certaine ambiguïté dans l'utilisation des différents termes [WALSH 05, 06; WOOD 71]. Nous définirons ensuite plus en détail l'intelligibilité et la compréhensibilité, deux termes fréquemment employés dans l'évaluation de la parole, souvent de manière non équivoque. Pour cela, nous avons mené une **étude de consensus international Delphi** visant la rédaction d'une définition désambiguïsée de ces termes, basée sur la consultation d'un panel d'experts de la parole. Nous décrirons ici également les types de mesures de ces deux concepts, perceptives et acoustiques, étroitement liées à leur définition.

Une fois la terminologie précisée, le deuxième chapitre de cette thèse abordera plus en détail l'**analyse des besoins** en termes de mesure de l'intelligibilité. Nous introduirons tout d'abord le modèle de l'**Evidence-Based Practice (EBP)**, qui servira de fil conducteur dans cette exploration des besoins. L'EBP est un courant visant à intégrer, pour les prises de décisions cliniques, trois sources d'informations : la recherche scientifique systématique, l'expérience clinique et les valeurs et préférences du patient [DOLLAGHAN 07]. À cela peut s'ajouter, selon les auteurs, le contexte local [FOSTER 13; RYCROFT-MALONE 04]. Notre analyse des besoins débutera ainsi par une investigation des données de la recherche scientifique concernant les mesures acoustiques de l'intelligibilité. Avant d'aborder les mesures de la parole en contexte pathologique, il est primordial d'en connaître les comportements en parole saine. Nous soulignerons l'importance de cette connaissance préalable en abordant la **variabilité de la parole naturelle** chez l'adulte sain, autant entre individus qu'entre différentes productions d'un même locuteur. Ensuite, nous exposerons les résultats d'une **revue systématique** menée selon les lignes directrices PRISMA [LIBERATI 09; MOHER 09], afin de répondre à la question des manques et besoins identifiés concernant les études scientifiques sur les mesures acoustiques de l'intelligibilité.

La suite de ce chapitre d'analyse des besoins traitera du second pilier de l'EBP, celui de l'expérience des professionnels concernés. Les cliniciens sont en première ligne pour l'amélioration de la qualité de vie des patients atteints de troubles de la parole. Nous nous pencherons sur la réalité clinique de l'évaluation de la parole sur le terrain. Pour cela, nous décrirons une **enquête adressée aux cliniciens** francophones en Belgique, en France, au Luxembourg, en Suisse et au Maghreb. Cette enquête, répondant également dans une certaine mesure à la prise en compte du « contexte local » de l'EBP, a permis de relever les manques et besoins exprimés par les cliniciens en termes de mesures de la parole. Ensuite, toujours dans le volet de l'expérience professionnelle, nous avons souhaité directement entendre le point de vue de chercheurs de différents horizons, concernant les pratiques actuelles et les manques en termes de mesure de la parole. En complément des données récoltées dans la littérature scientifique et auprès des cliniciens, nous avons pour cela réalisé des **interviews semi-dirigées** d'experts de la parole.

Enfin, nous aborderons brièvement le troisième pilier de l'EBP, qui concerne les valeurs et préférences du patient. En effet, l'opinion du patient est de plus en plus prise en compte dans la planification de la prise en charge [VRANCEANU 09], suivant la tendance croissante du « soin

centré sur le patient » (« patient-centered care » dans la littérature anglophone) [STEWART 01]. L'objectif, dans le présent travail, était de prendre en compte l'avis du patient dans le processus d'évaluation de la parole, avant même la planification thérapeutique ultérieure. Nous décrivons pour cela les résultats d'un **questionnaire destiné à des patients** adultes francophones souffrant d'un trouble de la parole.

Après avoir fait le point sur les besoins en termes de mesures de la parole en recherche scientifique et en pratique clinique, nous aborderons, dans le troisième chapitre, les **échantillons de parole** sur lesquels ces mesures peuvent être employées. Ce chapitre débutera par un bref état de l'art des supports d'évaluation de la parole existants et des différentes unités de parole ciblées. En effet, la parole peut être évaluée sur des échantillons de parole contrôlés de différents niveaux de granularité (p. ex. listes de mots ou de phrases), ou sur des échantillons plus naturels (p. ex. parole conversationnelle). Nous commencerons par rappeler les pratiques cliniques actuelles en ce qui concerne les supports et cibles des bilans de parole, à l'aide des données récoltées dans l'enquête adressée aux cliniciens francophones. Des données de l'enquête de consensus Delphi seront également reprises pour souligner les unités à cibler, selon le panel d'experts consulté, pour évaluer respectivement l'intelligibilité et la compréhensibilité. Cet état de l'art nous amènera à porter une attention particulière aux deux extrêmes de l'échelle de granularité des unités d'analyse : les phonèmes d'un côté et les textes de l'autre. Nous soulignerons l'importance de l'évaluation **phonémique** en abordant la contribution respective des consonnes et des voyelles à l'intelligibilité. Nous nous reposerons aussi sur la variabilité des consonnes et des voyelles, qui sera ici décrite en rapport avec l'influence du contexte phonémique et en soulignant l'importance du contrôle de ce contexte pour la validité de certaines mesures. Nous décrivons ensuite l'importance de l'évaluation de la parole sur base de **textes de référence**. Ceux-ci offrent une évaluation plus fonctionnelle que les unités d'analyse de plus bas niveau, tout en permettant le contrôle de divers paramètres linguistiques. Nous recenserons les textes de référence les plus fréquemment employés pour l'évaluation de la parole et de la voix, ainsi que leurs critères de construction.

Ceci nous permettra d'introduire la nécessité de **nouveaux textes** de référence plus spécifiquement créés à des fins d'évaluation de la parole et de la voix. Ce sera là le sujet de la deuxième partie de ce chapitre, où sera présenté un **groupe de travail francophone** mis en place à cet effet, intégrant des experts français, belges et canadiens de différents horizons. Le fruit du travail de ce groupe sera décrit en lien avec une partie des données de l'enquête de consensus Delphi, qui a permis de recenser et de hiérarchiser les critères à prendre en compte pour la création de textes.

Dans le chapitre quatre sera ensuite abordé le **traitement du signal** de parole à proprement parler, par l'investigation concrète de mesures acoustiques. Nous commencerons par un bref état de l'art en décrivant les résultats de la revue systématique menée pour répondre à la question de recherche suivante : « Quelles sont les mesures acoustiques spectrales employées dans la littérature scientifique pour quantifier l'intelligibilité en parole saine ? » Nous décrivons ensuite les expériences pratiques effectuées dans l'objectif de proposer de nouvelles pistes pour l'évaluation acoustique de la parole.

L'état de l'art a notamment permis de constater une prépondérance de mesures spectrales sur les voyelles, au détriment des consonnes [KAY 12]. L'aire du triangle vocalique, construit en mesurant les valeurs fréquentielles des deux premiers formants (F1 et F2) des voyelles, est une mesure courante dans l'évaluation de l'intelligibilité [TURNER 95; VILAIN 15; WEISMER 00]. Par le placement des voyelles sur un graphique à deux axes, F1 et F2, le triangle vocalique obtenu permet d'interpréter les valeurs mesurées par rapport au triangle « standard ». Cette mesure est employée autant en pratique clinique qu'en recherche scientifique. À notre connaissance, aucune alternative similaire n'existe cependant à ce jour pour les consonnes, si ce n'est en perception de la parole. En effet, en audiométrie, la « **banane de la parole** » est une représentation des consonnes par un point sur un graphique à deux axes, fréquence et intensité [KLANGPORNKUN 13; NORTHERN 84; SKINNER 78]. Les points, correspondant à la composante fréquentielle principale « standard » de chaque consonne et à son intensité, permettent ainsi la délimitation d'une « aire des consonnes » sur un audiogramme. C'est le rapport de la courbe d'audition de chaque patient à cette aire qui permet ensuite de prédire la perception des différentes consonnes, qui jouent un rôle non négligeable dans l'intelligibilité de la parole [BONATTI 05; OWREN 06]. Notre première étude acoustique a ainsi comme objectif de transposer cette « banane de la parole » du domaine de la perception vers celui de la production de la parole. Cette étude vise à reconstruire une telle « aire des consonnes » sur base d'enregistrements de parole d'adultes sains.

Après cette première exploration, nous décrivons la suite de notre investigation des mesures spectrales consonantiques en nous penchant sur les **moments spectraux**. Au contraire de la méthode précédente qui se basait sur la détection de pics spectraux, la technique des moments spectraux considère le spectre consonantique dans son entièreté. Ce spectre est analysé comme une distribution statistique et est décrit par quatre mesures [JONGMAN 00] : le centre de gravité (« center of gravity »), l'écart-type (« standard deviation »), l'asymétrie (« skewness ») et l'aplatissement (« kurtosis »). Ces mesures suscitent un intérêt en application clinique [ACHER 14; HERNANDEZ 19; KAY 12; KENT 99; TJADEN 17]. Elles seraient liées aux mesures articulatoires physiologiques [CATTELAINE 19; ISKAROUS 11], permettraient d'« objectiver » les différences entre parole naturelle et « atypique » et d'évaluer l'efficacité thérapeutique ou la progression pathologique [TJADEN 04]. De plus, elles sont non invasives et nécessitent peu d'équipement. C'est pourquoi nous avons tout d'abord souhaité analyser le comportement de ces mesures en parole adulte saine, sur les consonnes non-sonantes (les plosives et les fricatives) du français. Nos analyses seront décrites en détail dans ce chapitre, en abordant l'influence du genre, la reproductibilité des mesures en fonction des conditions d'enregistrement, l'emplacement de la fenêtre d'analyse et l'influence du contexte phonémique. Pour permettre une applicabilité en contexte clinique, nous nous attarderons ensuite sur l'utilisation des mesures spectrales en combinaison avec l'alignement forcé automatique. Ce dernier permet en effet aux cliniciens, souvent sujets à des contraintes temporelles importantes, d'éviter la tâche chronophage de segmentation manuelle. Enfin, nous aborderons les analyses de moments spectraux appliqués à la parole pathologique. Pour cela, nous décrivons les comportements de ces mesures sur des enregistrements de patients traités pour un cancer de la sphère oropharyngée. Les résultats seront ensuite comparés à ceux obtenus en parole saine, ainsi qu'à deux scores de référence perceptifs obtenus sur ces mêmes sujets.

Dans le chapitre cinq sera décrite la **mise en application de mesures acoustiques sur notre texte de référence** créé par le groupe de travail présenté en chapitre trois. En effet, les mesures décrites dans les chapitres précédents ont été effectuées dans des contextes phonémiques très contrôlés et peu naturels. Leur utilisation sur un texte permettrait une évaluation plus fonctionnelle de la parole. Comme l'emploi de mesures acoustiques sur un texte de référence en maximise la pertinence clinique, nous présenterons ici un protocole d'analyse acoustique sur notre nouveau texte facile d'utilisation par des experts non-informaticiens. Nous analyserons la faisabilité de quatre types de mesures acoustiques : qualité vocale, voyelles, consonnes et semi-consonnes. Nous investiguerons leur reproductibilité à travers des lectures répétées de notre texte ainsi que l'influence de la vitesse de lecture, mais aussi celle du genre et de la variante régionale du français. Nous aborderons également brièvement les différences entre les mesures de moments spectraux effectuées sur des phrases porteuses et sur notre nouveau texte de référence.

Nous concluons ensuite ce travail par une synthèse des contributions apportées tout au long des études que nous avons menées. Enfin, nous décrivons les perspectives pour de futures études visant l'exploration acoustique des sons de la parole et la création d'outils d'évaluation de l'intelligibilité exhaustifs, fiables, mais aussi et surtout cliniquement pertinents et employables.

Résumé des Contributions

Les principales contributions concrètes de ce travail dans le cadre de la problématique ciblée sont résumées dans la liste suivante :

- définition de l'intelligibilité et de la compréhensibilité et de leur mesure par une étude de consensus international (cf. [Chapitre 1](#)) ;
- revue systématique de la littérature concernant les mesures acoustiques de l'intelligibilité en parole saine (cf. « 2.2. [Premier pilier de l'EBP : la recherche scientifique](#) ») ;
- état des lieux des bilans de la parole adulte et des manques et besoins en clinique francophone (cf. « 2.3.1. [Évaluation des troubles de la parole chez l'adulte – situation actuelle et besoins en pratique clinique francophone](#) ») ;
- proposition d'un ensemble minimal de tâches pour l'évaluation de la parole (cf. Chapitre 2 - [Annexe G](#)) ;
- suggestions pour l'enregistrement de la parole en vue d'analyses acoustiques (cf. Chapitre 2 - [Annexe G](#)) ;
- interviews semi-dirigées de chercheurs dans le domaine de la parole concernant la mesure de celle-ci (cf. « 2.3.2. [Évaluation de la parole chez l'adulte – opinions d'experts](#) »)
- prise en compte de l'opinion de patients concernant l'évaluation de leur trouble de la parole (cf. « 2.5. [Troisième pilier de l'EBP : l'avis des patients](#) ») ;
- rédaction d'un texte de référence en français spécifiquement créé pour l'évaluation de la parole et de la voix (cf. « 3.2. [Nouveau texte de référence](#) ») ;
- analyses acoustiques au niveau segmental pour l'évaluation de l'intelligibilité : « banane de la parole » et moments spectraux (cf. [Chapitre 4](#)) ;
- implémentation d'un algorithme de détection de burst des plosives et du segment stable à l'intérieur des fricatives (cf. « 4.3.2. [Moments spectraux sur alignement forcé](#) ») ;

- mise en place d'un protocole permettant l'extraction aisée et rapide de mesures acoustiques du nouveau texte de référence (cf. [Chapitre 5](#)).

Ces contributions ont été mises en valeur à travers des présentations dans des conférences internationales [BALAGUER, POMMEE 20b; POMMEE 20; POMMEE 19a; POMMEE 21d; POMMEE 19b] ainsi que dans des articles scientifiques revus par des pairs [BALAGUER, POMMEE 21; BALAGUER, POMMEE 20c; POMMEE 21a, 21b, 21c] et dans des chapitres d'ouvrage [BALAGUER, POMMEE n.d.; POMMEE n.d.]. Deux articles scientifiques concernant respectivement le nouveau texte de référence et les analyses de moments spectraux sont en cours de rédaction.

Chapitre 1. Intelligibilité

1.1. État de l'art

Le bilan clinique de parole vise à évaluer plusieurs dimensions afin d'obtenir une vision complète et individualisée de la parole de chaque patient. Généralement, ce bilan comprend une exploration de la sensibilité et de la motricité orofaciales, ainsi qu'une évaluation fonctionnelle, à l'aide de diverses mesures perceptives et instrumentales, des multiples dimensions de la parole : la respiration, la phonation, la résonance, l'articulation (planification, programmation et exécution motrice), l'intelligibilité, la compréhensibilité, ainsi que l'impact psychosocial du trouble de la parole [ALTAHER 19; DYKSTRA 07; POMMEE 21a; RUMBACH 19].

La structure globale du bilan de parole ainsi que les tâches et outils employés pour en évaluer les différentes dimensions sont variables selon les pays, les contextes et les cliniciens [GUREVICH 17; POMMEE 21a]. Ce manque de standardisation, que nous aborderons plus en détail dans le chapitre suivant, ne concerne pas uniquement les tâches et outils, mais également la terminologie employée. En effet, dans l'évaluation des multiples dimensions de la parole, une grande diversité terminologique peut être observée, empreinte d'une certaine ambiguïté. De fait, il existe un manque de consensus concernant les définitions de certains termes liés à la parole, et, à l'inverse, concernant les termes attribués aux différents concepts évalués. Ce manque de consensus concerne non seulement la désignation des concepts ciblés, mais aussi, inéluctablement, la manière dont ils seront évalués.

Une ambiguïté particulière se présente pour les termes «intelligibilité» et «compréhensibilité», deux termes fréquemment employés et dont le premier sera le focus principal de cette thèse. Cette ambiguïté est observée à la fois dans l'usage terminologique par les cliniciens, dans les batteries d'évaluation existantes et dans la littérature scientifique.

Concernant l'usage terminologique en pratique clinique, le manque de consensus est notamment apparu dans une enquête que nous avons menée auprès de cliniciens francophones et dont la méthodologie sera détaillée dans le chapitre suivant. Parmi les 119 orthophonistes ayant participé, 63 % disent évaluer l'intelligibilité de la parole de manière perceptive, contre 45 % pour la compréhensibilité. Si nous reprenons les définitions les plus couramment citées de l'intelligibilité et de la compréhensibilité, la première correspond à un concept plus analytique de décodage acoustico-phonétique du signal de parole (p. ex. [GHIO 18; HUSTAD 08; LALAIN 20; YORKSTON 96]), tandis que la seconde réfère à un concept plus global de déduction du sens du message véhiculé, grâce à des informations syntaxiques, sémantiques, pragmatiques et paralinguistiques [BAREFOOT 93]. Cependant, au vu des tâches utilisées par ces cliniciens, le concept évalué semble en réalité être majoritairement celui de la compréhensibilité : à titre d'exemple, 95 % utilisent la parole spontanée, 77 % la lecture de texte et 76 % la répétition de phrases. La répétition de pseudomots et la lecture de pseudomots quant à elles, sont utilisées par 61 % et 28 % des orthophonistes, respectivement. Pourtant, ces dernières tâches court-circuitent les processus cognitifs de compensation sémantique et syntaxique (dits «top-down») permettant à l'auditeur de suppléer une parole peu intelligible au niveau phonémique ou syllabique.

Cette confusion terminologique se retrouve également dans les manuels des batteries d'évaluation disponibles à ces cliniciens. Dans la Batterie d'Évaluation Clinique de la Dysarthrie [AUZOU 06], par exemple, l'intelligibilité est décrite comme « le degré de précision avec lequel le message est compris par l'auditeur ». Le « score d'intelligibilité » y est basé sur une lecture de mots et de phrases, ainsi que sur de la parole spontanée (*sic*). De plus, les sous-épreuves sont intitulées « compréhension de mots » et « compréhension de phrases », ajoutant à la confusion.

Enfin, l'ambiguïté terminologique est également observée dans la littérature scientifique [DENMAN 19; WALSH 05, 06; WOISARD 10]. Cette problématique était déjà abordée en 1971 par Wood, qui expliquait (traduit de l'anglais) : « De nombreux termes et leurs significations ne sont pas bien cristallisés, car le sujet est en constante évolution ; les concepts eux-mêmes sont souvent incertains et fluctuants [...]. Cette croissance de l'orthophonie [...] a généré des centaines de termes, dont certains sont interchangeables, et dont certains ont des significations différentes selon les personnes » [WOOD 71]. Plus récemment, l'Australian Institute of Health and Welfare affirmait que cette problématique était toujours d'actualité (traduit de l'anglais) : « La classification et la terminologie utilisées pour décrire les troubles de la parole sont particulièrement entachées d'incohérences, notamment l'utilisation d'interprétations différentes pour une même terminologie ou de terminologies différentes pour une même signification » [AUSTRALIAN INSTITUTE OF HEALTH AND WELFARE 03]. La revue systématique que nous avons menée dans le cadre de cette thèse et qui sera décrite plus en détail dans les chapitres deux et quatre a également permis de mettre en évidence cette ambiguïté terminologique dans la littérature scientifique [POMMEE 21c]. En effet, dans les articles retenus, les termes employés pour référer à la production de la parole, parfois même de manière interchangeable, sont nombreux. Ils incluent l'intelligibilité, la précision articulatoire, la clarté de parole, la compréhensibilité et la sévérité. Parmi les 22 articles retenus, seuls cinq abordent la définition des concepts intéressés, dont quatre définissent l'intelligibilité [BUNTON 01; A. R. FLETCHER 17; MARTEL-SAUVAGEAU 17; YUNUSOVA 05], qui était ciblée dans cette revue. Cette dernière est définie — et par conséquent évaluée — de manières différentes selon les auteurs, leurs domaines de recherches et leurs objectifs. [MARTEL-SAUVAGEAU 17], par exemple, définissent l'intelligibilité comme le degré de compréhension d'un énoncé oral par un auditeur. [BUNTON 01], quant à eux, définissent l'intelligibilité comme (traduit de l'anglais) « le nombre d'unités de parole discrètes correctement reconnues par un auditeur ». Si nous considérons les tâches employées pour évaluer la production de la parole, seuls huit articles emploient des tâches concordant avec l'intelligibilité telle que nous l'avons définie ci-avant, en utilisant des tâches d'identification phonémique ou des listes de paires minimales de mots. Les 14 autres articles emploient des tâches d'identification de mots ou des jugements globaux sur des mots ou sur des phrases, à l'aide par exemple d'échelles visuelles analogiques ou d'échelles ordinales. À nouveau, ces tâches impliquent des processus cognitifs de compensation (« top-down ») de haut niveau et évaluent donc plutôt la compréhensibilité que l'intelligibilité.

Les définitions des termes référant à la parole sont donc variables, déterminées notamment par le domaine de pratique ou de recherche de celui qui les emploie. En effet, les phonéticiens, audiologistes, orthophonistes et autres professionnels travaillant dans le domaine de la production ou de la perception de la parole référeront potentiellement à des concepts

différents lorsqu'ils parlent de l'«intelligibilité». Ceci peut amener à des problèmes de communication entre professionnels et avoir un impact non seulement en pratique clinique, mais également en recherche scientifique [DENMAN 19; WALSH 05, 06].

Dans un contexte clinique, cela peut impacter l'efficacité de la prise en charge du patient dans une équipe pluridisciplinaire. Des malentendus peuvent également apparaître lorsqu'un patient est réorienté d'un professionnel à un autre. Dans un contexte scientifique, en plus d'entraver les débats scientifiques entre chercheurs, l'ambiguïté terminologique peut entraîner des difficultés à comparer et à combiner les résultats de recherches. De ce fait, elle réduit l'efficacité de la recherche scientifique, entrave les collaborations interdisciplinaires et favorise des avancées plutôt ségréguées (p. ex. à l'intérieur des différents domaines de recherche) au lieu d'un progrès plus homogène et transdisciplinaire. En plus de son impact dans les domaines cliniques et scientifiques, le manque de terminologie consensuelle entrave également le lien entre ces deux domaines, en affectant la transposition des découvertes fondamentales en applications cliniques [DENMAN 19; ROULSTONE 15].

Au vu du manque de consensus concernant la terminologie liée à l'évaluation de la parole et de son impact dans les domaines clinique et scientifique, nous avons mené une enquête internationale selon la méthode Delphi [CHALMERS 19; VON DER GRACHT 12]. L'objectif de cette étude était de proposer une définition plus exhaustive et consensuelle de l'intelligibilité et de la compréhensibilité. En objectif secondaire, cette étude visait à définir les mesures perceptives et instrumentales d'évaluation de ces deux concepts, étroitement liées à leur définition. Nous allons maintenant décrire cette étude Delphi, après avoir abordé différentes méthodes de recherche de consensus et développé les avantages de la méthode choisie.

1.2. Intelligibilité et compréhensibilité : définition et mesures

1.2.1. Choix de la méthode

Afin d'obtenir un consensus sur une question donnée, trois méthodes sont couramment employées en recherche dans le domaine de la santé : la conférence de consensus, le groupe nominal (ou panel d'experts) et la méthode Delphi [JONES 95; MCMILLAN 16]. La conférence de consensus a initialement été développée par le National Institutes of Health (NIH) pour la validation de la sécurité et de l'efficacité de nouvelles technologies dans le domaine de la santé et leur adaptation à la pratique clinique [PERRY 80]. Lors d'une telle conférence, un groupe d'experts se rassemble pendant quelques jours afin de rédiger des recommandations consensuelles. Il s'agit là d'un événement fastidieux à organiser, épuisant pour les participants (potentielles sessions nocturnes) et pour lequel les problèmes inhérents au contexte de groupe, telle la pression sociale vers la conformité, ne sont pas contrôlés. Enfin, aucun critère formel de décision ni processus de vote ne sont utilisés, et seuls des résultats qualitatifs peuvent en être tirés [LETRILLIART 11].

Au contraire des conférences de consensus, les méthodes de groupe nominal et de Delphi sont des approches structurées et systématiques, permettant l'obtention de résultats quantitatifs. Tout comme pour la conférence de consensus, le groupe nominal est une rencontre d'experts en face-à-face [MCMILLAN 14]. Cependant, il utilise une série très structurée de tours consécutifs afin

d'évaluer et de discuter différentes questions (voir [JONES 95] pour une description plus détaillée des différentes étapes).

Enfin, la méthode d'enquête groupée Delphi consiste elle aussi en un processus en plusieurs étapes (généralement trois tours successifs). Son objectif est d'obtenir un consensus parmi un groupe de participants ayant une expertise — professionnelle ou basée sur l'expérience — dans un domaine cible [BIRKO 15; LINSTONE 02b]. Au contraire des deux méthodes précédentes, elle emploie des questionnaires remplis indépendamment par chaque expert au lieu de recourir au face-à-face. Les sondages sont effectués soit par courrier, soit en ligne. La méthode Delphi permet ainsi de faire intervenir des groupes d'experts plus importants, par exemple en comparaison à la taille recommandée de sept experts pour le groupe nominal [MCMILLAN 16]. Elle peut plus facilement s'utiliser en ligne, avec les avantages que cela apporte [HARTMAN 95; LINSTONE 02a; TUROFF 96]. Outre son coût peu élevé et l'absence de limites géographiques, son caractère quasi anonyme est un argument primordial en faveur de cette méthode [SINHA 11; VON DER GRACHT 12]. L'identité des participants n'est connue que par le modérateur, tandis qu'ils restent inconnus les uns des autres. Ceci permet une liberté d'expression exempte de pression sociale ou professionnelle de la part des pairs. En outre, la nature quasi anonyme combinée à l'utilisation de plusieurs tours consécutifs et aux feed-back structurés fournis aux participants après chaque tour permet de réduire les biais dans le processus de recherche de consensus [CHALMERS 19]. C'est au vu de ses nombreux avantages que nous avons choisi d'utiliser la méthode Delphi pour élaborer une définition plus consensuelle de l'intelligibilité et de la compréhensibilité et de leur évaluation.

1.2.2. La méthode Delphi

La méthode Delphi a initialement été développée par la RAND Corporation à des fins militaires [DALKEY 63]. Elle a ensuite, à partir des années 80, été appliquée à divers domaines de recherche [VON DER GRACHT 12] dans l'objectif de faire des prévisions ou de prendre des décisions concernant des problématiques contemporaines [CHALMERS 19], de générer de nouvelles idées ou de déterminer des priorités [MCMILLAN 16]. Dans le domaine de la santé, cette méthode est aujourd'hui couramment utilisée notamment pour l'élaboration de lignes directrices [MCMILLAN 16], l'évaluation de l'adéquation de traitements [BEERS 91], les prévisions de prévalence de pathologies [CHIN 90] ou encore l'amélioration de l'enseignement et de la formation dans les professions de santé [FASSER 92].

Dans une enquête Delphi traditionnelle, lors du premier tour, les participants se voient généralement poser des questions ouvertes, afin d'éviter que les modérateurs n'imposent leur point de vue [CUNNINGHAM 19; HSU 07; MCMILLAN 16; SINHA 11]. Dans une enquête Delphi « modifiée », les énoncés à évaluer, construits au préalable sur base d'une recherche documentaire, sont directement fournis au premier tour [CUNNINGHAM 19; DENMAN 19]. Dans les deux cas, lors des tours suivants, les experts sont invités à donner leur avis concernant divers énoncés dérivés des réponses du tour précédent, à l'aide d'échelles d'évaluation ou de classements. Après chaque tour, les réponses du groupe sont synthétisées par le modérateur et les points d'accord et de désaccord sont identifiés. Un feed-back contrôlé agrégé [VON DER GRACHT 12] est alors fourni aux participants au tour suivant pour expliquer les modifications

apportées aux énoncés dans l'objectif de tendre vers un consensus. Les participants donnent ensuite à nouveau leur avis sur ces énoncés [DENMAN 19; DIAMOND 14]. Si l'objectif initial de la méthode Delphi était la recherche de consensus, cette technique est aujourd'hui également utilisée pour mettre en évidence et étudier les points de désaccord (parfois appelés points de « dissensus »), ou encore des « clusters de consensus ». En effet, l'identification de groupements d'opinions est intéressante, car elle permet de révéler les éventuels points de vue spécifiques aux différentes communautés ou domaines de recherche [BIRKO 15]. Ainsi, les itérations sont généralement effectuées jusqu'à ce qu'une stabilité des réponses soit observée, qu'un consensus soit atteint ou non (voir [CHALMERS 19; VON DER GRACHT 12] pour des exemples de mesures de stabilité).

1.2.3. Notre étude Delphi

A. Méthodologie

Design de l'étude

Les étapes de l'étude Delphi modifiée que nous avons menée sont résumées en [Figure 1](#). Nous allons maintenant décrire les étapes de la définition du problème, la sélection du panel, la recherche documentaire et la construction des trois tours Delphi consécutifs.

1. Définition du problème

Comme nous l'avons développé dans l'état de l'art de ce chapitre, il existe une ambiguïté dans la terminologie liée à la parole, notamment concernant l'intelligibilité et la compréhensibilité. Cette ambiguïté peut conduire à un problème de communication entre professionnels et entraver l'efficacité de la prise en charge des patients ainsi que l'avancement de la recherche. Elle conduit également à un manque de consensus sur les tâches et les mesures d'évaluation de la parole à inclure dans un bilan standard. L'objectif de notre étude Delphi était donc d'élaborer une définition plus consensuelle de l'intelligibilité et de la compréhensibilité et de leur évaluation.

2. Sélection d'experts

Notre étude s'adressait aux professionnels (cliniciens, chercheurs et enseignants) impliqués dans des activités liées aux troubles de la parole et/ou de la fluence (bégaiement/bredouillement). Les « activités » étaient définies comme des activités cliniques, de recherche, universitaires ou industrielles, ou une combinaison de celles-ci, si au moins approximativement 20 % étaient liées à la parole (selon l'auto-estimation par les participants). Ces professionnels devaient avoir une maîtrise intermédiaire de l'anglais à l'écrit.

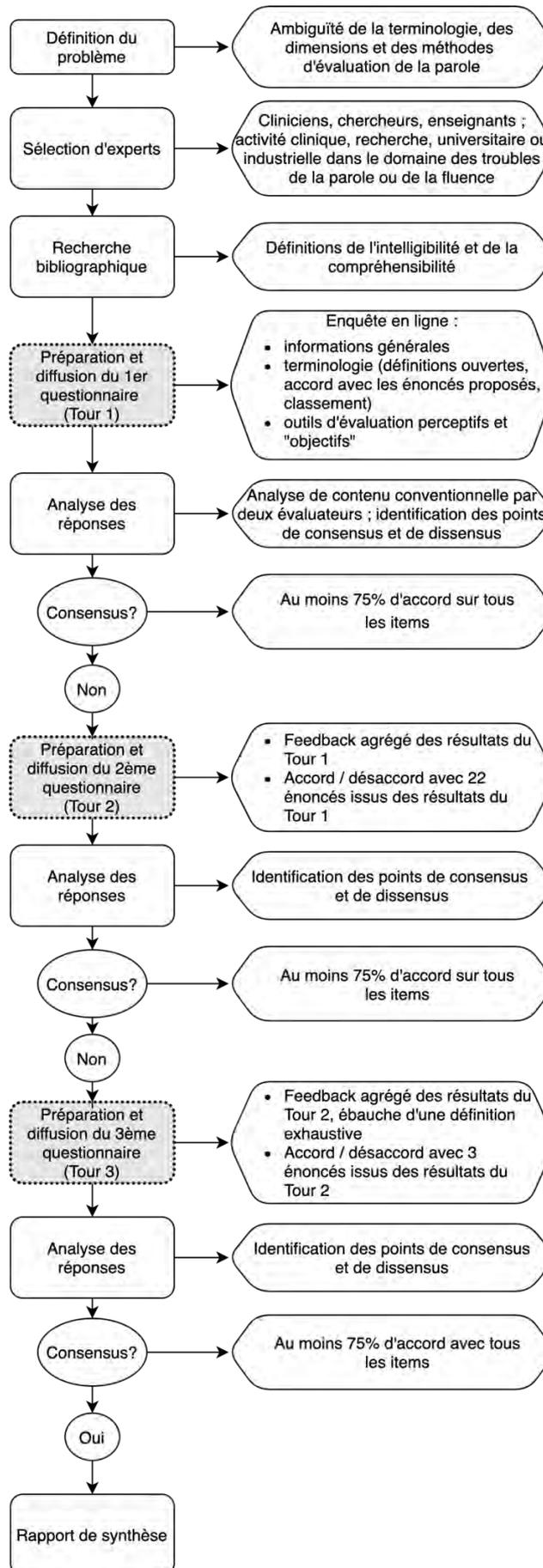


Figure 1 – Organigramme du déroulement de notre étude Delphi

Le recrutement a été effectué via :

- les associations professionnelles nationales (orthophonie, phoniatrie, voix, acoustique, informatique/traitement du signal, linguistique/phonétique) ; plus de 200 organisations dans le monde entier ont été contactées par courrier électronique ;
- les réseaux sociaux (Twitter, Facebook), où les professions et associations ciblées ont également été sollicitées dans des groupes privés ;
- courrier électronique à plus de 50 experts de la parole identifiés par des recherches bibliographiques sur PubMed, qui avaient au moins trois publications dans le domaine de la parole, étaient auteurs d'un livre de référence ou participaient à des projets de recherche liés à la parole pathologique.

À chaque tour, les non-répondants ont été exclus des tours suivants. Une analyse *a posteriori* utilisant des statistiques descriptives a été effectuée pour vérifier que les caractéristiques du groupe d'experts ne s'en retrouvaient pas modifiées. Nous avons également vérifié que le consensus croissant tout au long des tours n'était pas biaisé par les éventuels abandons d'experts majoritairement en désaccord. À cette fin, les données quantitatives des trois tours, avec et sans les participants qui se sont retirés, ont été comparées à l'aide de statistiques descriptives.

3. Recherche bibliographique

En étape préliminaire, une recherche bibliographique a été effectuée pour identifier les définitions de l'intelligibilité et de la compréhensibilité dans la littérature scientifique. Cinq définitions, considérées comme reflétant le mieux les différentes interprétations des deux termes, ont été retenues pour être présentées dans le premier tour de cette étude Delphi (après définition libre des deux concepts par les participants pour éviter de les biaiser) :

- [GHIO 18] :

« La perception de la parole est un processus complexe qui intègre à la fois un flux ascendant d'informations provenant du signal vocal mais aussi un flux descendant fondé sur les informations de haut niveau détenues par l'auditeur. Le flux ascendant ("bottom-up") est principalement une opération de décodage acoustico-phonétique qui consiste à identifier les phonèmes à partir du signal de parole. Les phonèmes, pouvant être considérés comme les plus petites unités permettant d'opposer du sens, sont les éléments de base de l'**intelligibilité** du discours. [...] Le décodage acoustico-phonétique est donc le processus fondamental pour mesurer perceptivement l'**intelligibilité** d'un locuteur. »

- [HODGE 10] (traduit de l'anglais²) :

« L'**intelligibilité**, ou la mesure dans laquelle la parole d'une personne est compréhensible pour une autre, est un indicateur fonctionnel de la compétence communicationnelle orale. Elle reflète la capacité d'un locuteur à convertir le langage en un signal physique (la parole) et la capacité d'un auditeur à percevoir et à décoder ce signal pour recouvrer le sens du message du locuteur. »

- [HUSTAD 08] (traduit de l'anglais²) :

« L'**intelligibilité** fait référence à la précision avec laquelle le signal acoustique d'un locuteur peut être recouvert par un auditeur. »

² Cf. définition originale en [Annexe A](#).

- [YORKSTON 96] (traduit de l'anglais²) :

« Le terme **intelligibilité** réfère au degré avec lequel le signal acoustique (l'énoncé produit par le locuteur dysarthrique) est compris par un auditeur. [...] Les concepts de compréhensibilité et d'intelligibilité peuvent être distingués par le fait que la **compréhensibilité** intègre des informations indépendantes du signal telles que la syntaxe, la sémantique et le contexte physique. »

- [BAREFOOT 93] (traduit de l'anglais²) :

« La compréhensibilité est définie comme la mesure dans laquelle un auditeur comprend les énoncés produits par un locuteur dans un contexte de communication. Selon nous, la compréhensibilité concerne à la fois les domaines de la parole et du langage, tandis que l'intelligibilité concerne principalement le domaine de la parole. La principale distinction entre la compréhensibilité et l'intelligibilité est que la compréhensibilité est destinée à rendre compte de caractéristiques communicationnelles des énoncés qui vont au-delà du domaine auditif-acoustique. La **compréhensibilité**, dans notre utilisation du terme, incorpore explicitement des caractéristiques contextuelles telles que la syntaxe, la sémantique et la pragmatique, et implique une activité de communication en face à face dans laquelle des énoncés signifiants sont produits par des locuteurs et traités par des auditeurs. »

L'énoncé suivant a été ajouté afin d'inciter la réflexion des participants :

« L'**intelligibilité** et la **compréhensibilité** peuvent être employées comme des synonymes. »

4. *Tours successifs et analyse des données*

Cette enquête de consensus Delphi a été menée en trois tours consécutifs (voir [Figure 1](#)), entre juillet et décembre 2020. Le tour 1 était disponible pendant deux mois et demi ; les tours 2 et 3 étaient disponibles pendant 1 mois, respectivement. Les questionnaires en ligne sont toujours disponibles sur la plateforme LimeSurvey³. Le premier questionnaire était en accès libre, avec des procédures de vérification de sécurité et de doublons. Les questionnaires suivants étaient réservés aux participants des tours précédents et nécessitaient un code individuel fourni par le modérateur. Tous les questionnaires ont été soumis à un essai pilote par cinq chercheurs afin d'obtenir une estimation du temps de réponse et de détecter et corriger les problèmes d'exécution éventuels (bogues et problèmes de structure logique).

À chaque tour, les participants indiquant un désaccord avec un énoncé étaient encouragés à expliquer la raison de celui-ci, mais les commentaires n'étaient pas obligatoires afin de ne pas biaiser les réponses en faveur de réponses positives.

Les données qualitatives et quantitatives ont été analysées à l'aide du logiciel Stata/MP (version 14, StataCorp, College Station, TX).

Critères de consensus et d'arrêt

Le seuil de consensus a été défini avant l'analyse des données comme étant l'accord d'au moins 75 % du panel d'experts [DENMAN 19; DIAMOND 14].

Le nombre maximal de tours prévu était de quatre, et le critère d'arrêt était soit l'obtention d'un consensus sur tous les items de chacune des deux principales parties investiguées (terminologie et mesures d'évaluation de la parole), soit une stabilité des réponses.

³ Tour 1 : <https://enquetes.univ-tlse3.fr/index.php/623792?lang=en> – Tour 2 : <https://enquetes.univ-tlse3.fr/index.php/372685?lang=en> – Tour 3 : <https://enquetes.univ-tlse3.fr/index.php/526489?lang=en>

Tour 1

Le premier tour était construit en trois parties (31 questions) :

1. Informations générales sur le participant ;
2. Terminologie (définitions ouvertes de l'intelligibilité et de la compréhensibilité⁴, degré d'accord avec et classement de préférence des définitions proposées et énumération d'autres termes utilisés) ;
3. Mesures d'évaluation perceptive et « objective »⁵ de l'intelligibilité et de la compréhensibilité (choix multiples).

Des statistiques descriptives ont été réalisées sur les données démographiques quantitatives et qualitatives afin de dresser un tableau global du groupe d'experts.

Deux évaluateurs (Mathieu Balaguer et moi-même, orthophonistes et étudiants doctorants), en aveugle par rapport à l'identité des participants, ont effectué individuellement une analyse de contenu conventionnelle [DENMAN 19; HSIEH 05] des définitions spontanées de l'intelligibilité et de la compréhensibilité ainsi que des définitions de la littérature, afin d'identifier les principaux thèmes et concepts récurrents. Une analyse de fréquence a ensuite été utilisée pour identifier les tendances, en quantifiant le nombre d'experts évoquant chacun de ces concepts dans leurs définitions spontanées. Les degrés d'accord et les rangs de préférence concernant les énoncés tirés de la littérature ont également été analysés à l'aide d'une analyse de fréquence, en tenant compte des concepts inclus dans chacune des définitions.

Les réponses ouvertes concernant les éventuels autres termes utilisés pour décrire et évaluer les troubles de la parole ont été regroupées sémantiquement en termes génériques et spécifiques par les deux évaluateurs, et une analyse de fréquence a été appliquée.

Enfin, une analyse de fréquence a aussi été utilisée pour identifier les principales tendances concernant les méthodes d'évaluation perceptive et « objective » de l'intelligibilité et de la compréhensibilité.

L'ensemble de ces résultats a ensuite servi à rédiger 22 énoncés pour le deuxième tour, ciblant chacun des principaux concepts identifiés.

Tour 2

Pour la mise en contexte du deuxième tour, les réponses et commentaires du premier tour ont été synthétisés et présentés aux participants sous forme groupée et anonyme, conjointement avec les 22 nouveaux énoncés basés sur ces résultats. Le deuxième tour était structuré en deux parties (« Terminologie » et « Mesures d'évaluation de la parole »). Tous les énoncés étaient regroupés selon les six concepts identifiés dans l'analyse de contenu du premier tour et ont été évalués à l'aide de réponses binaires (Accord/Désaccord), avec commentaires facultatifs.

⁴ Les participants avaient la possibilité, à plusieurs reprises, de modifier leurs définitions libres tout au long du questionnaire, par exemple après avoir été invités à évaluer les définitions existantes.

⁵ Le terme « objectif » a été utilisé dans l'enquête en ligne et sera donc également utilisé dans la description des résultats de cette étude. Cependant, la notion d'objectivité est sujette à discussion, certains experts soutenant que même les mesures acoustiques restent relativement subjectives, car elles sont effectuées par l'humain, avec des biais subjectifs subsistant dans la procédure d'enregistrement, les paramètres d'analyse, le choix des stimuli, de la fenêtre d'analyse... Par conséquent, ce terme sera ici employé entre guillemets. Le sens initialement envisagé était celui de mesures « reproductibles, instrumentales », par opposition aux méthodes perceptives, plus subjectives.

Tour 3

Un troisième tour a été nécessaire pour clarifier trois énoncés qui se sont avérés ambigus lors du deuxième tour. Après ce troisième tour, le critère d'arrêt (consensus sur tous les énoncés) a été atteint. Une ébauche de définition de l'intelligibilité et de la compréhensibilité, intégrant tous les éléments consensuels des tours précédents, a ainsi pu être rédigée.

B. Résultats

Groupe d'experts

Quarante experts ont complété le tour 1 ; trente-quatre experts ont complété le tour 2 (85 %) ; trente-trois experts ont complété le tour 3 (97 % du tour 2, 83 % du tour 1). Un taux d'abandon total de 17 % entre le tour 1 et le tour 3 a donc été observé.

Des données détaillées sur les participants à chaque tour sont disponibles en [Annexe B](#). Les tendances décrites ci-après sont constantes tout au long des trois tours, malgré les quelques abandons. La majorité des membres du groupe d'experts sont des orthophonistes travaillant dans le domaine des troubles de la parole, de la fluence et de la voix. Les autres professions majoritairement observées sont les linguistes, les ORL/phoniatres et les informaticiens. Plus de la moitié des experts ont au moins 10 ans d'expérience dans le domaine de la parole et de la voix. Leur activité principale est la recherche pour environ 40 %, la pratique clinique et l'activité académique pour environ un tiers des experts, respectivement ; seuls deux participants initiaux étaient engagés dans une activité industrielle. La grande majorité des experts exercent au moins deux activités principales ; l'activité clinique et la recherche sont combinées dans la moitié des cas. Plus de la moitié des participants ont un diplôme de troisième cycle (doctorat) obtenu en moyenne en 2009-2010 (± 8 ans).

La France, le Royaume-Uni et l'Allemagne sont les pays les plus représentés, tandis que la langue principale la plus fréquemment parlée au travail est l'anglais, suivi du français et de l'allemand.

Les populations de patients sont plutôt équilibrées en ce qui concerne les groupes d'âge, avec une légère prévalence pour la population âgée. De même, les pathologies neurologiques acquises et dégénératives sont les plus rencontrées.

Remarque préliminaire sur les abandons

La vérification *a posteriori* a révélé que le fait de retirer ou non des analyses les sept participants s'étant désistés en cours de route n'a pas d'impact significatif sur les conclusions et les valeurs de consensus concernant la terminologie et les mesures d'intelligibilité et de compréhensibilité. Ainsi, les données de tous les participants ont été retenues.

Terminologie – Intelligibilité et Compréhensibilité

Pour rappel, lors du **premier tour**, les participants devaient d'abord fournir des définitions spontanées de l'intelligibilité et de la compréhensibilité et énumérer d'éventuels autres termes employés. Ils devaient ensuite indiquer leur accord avec six énoncés tirés de la littérature (cf. « 1.2.3. A.3. [Recherche bibliographique](#) ») sur une échelle à 6 points (1 : Pas du tout d'accord - 6 : Tout à fait d'accord ; commentaire facultatif) et les classer par ordre décroissant de préférence.

L'analyse de contenu conventionnelle des définitions spontanées de l'intelligibilité et de la compréhensibilité a révélé six concepts principaux. Ces six concepts figurent à la fois dans les définitions spontanées et dans les définitions de la littérature :

- synonymie : mentions de l'intelligibilité et de la compréhensibilité comme étant des synonymes ;
- reconstruction du contenu de parole : définitions de l'intelligibilité ou de la compréhensibilité en rapport avec la précision de la reconstruction par l'auditeur du contenu transmis, soit au niveau du signal acoustique, soit au niveau sémantique ;
- production phonétique-acoustique : pour l'intelligibilité, mention de la contribution des capacités de production de bas niveau du locuteur à la reconstruction du contenu ;
- décodage acoustico-phonétique : pour l'intelligibilité, mention de la contribution des capacités de décodage de bas niveau du côté de l'auditeur à la reconstruction du contenu ;
- communication fonctionnelle : mise en évidence de la contribution de la compréhensibilité à la communication fonctionnelle ;
- éléments contextuels : mention des éléments linguistiques, extralinguistiques et paralinguistiques comme contributeurs à la compréhensibilité.

Les résultats suivants sont présentés pour chacun de ces principaux concepts identifiés.

1. *Synonymie*

Dans les définitions spontanées, aucun participant n'a mentionné que l'intelligibilité et la compréhensibilité sont des synonymes.

Soixante-dix-huit pour cent (31/40) des experts n'étaient pas d'accord avec l'affirmation « L'intelligibilité et la compréhensibilité peuvent être employées comme des synonymes » (degré d'accord moyen : 2,18/6 ; mode : 1/6).

Soixante-treize pour cent (29/40) ont classé cet énoncé au dernier rang de préférence par rapport aux cinq autres. Seuls cinq pour cent (2/40) l'ont classé en première position. Quinze pour cent (6/40) des participants ont souligné dans leurs commentaires que l'intelligibilité fait référence au locuteur plutôt qu'à l'auditeur.

2. *Reconstruction du contenu*

La majorité des participants, dans leurs définitions spontanées, ont indiqué que l'intelligibilité et la compréhensibilité permettent la reconstruction par un auditeur du contenu de parole transmis (intelligibilité : 63 %, 25/40 ; compréhensibilité : 85 %, 34/40). Concernant l'intelligibilité, 90 % (36/40) ont précisé que le contenu est véhiculé par le signal sonore.

Quatre-vingt-dix-huit pour cent (39/40) étaient d'accord avec la définition de [YORKSTON 96] (degré d'accord moyen : 5,48/6 ; mode : 6/6), qui précise, en ce qui concerne l'intelligibilité, que l'information est véhiculée par le signal acoustique.

3. *Communication fonctionnelle*

Concernant la compréhensibilité, dans leurs définitions spontanées, les participants ont mis l'accent sur l'aspect fonctionnel de la communication.

Quatre-vingt-treize pour cent (37/40) d'entre eux étaient d'accord avec la définition de [BAREFOOT 93], qui définit la compréhensibilité comme un indicateur de la communication fonctionnelle, par opposition à l'intelligibilité. Cette dernière n'est pas un indicateur direct de la communication fonctionnelle selon les commentaires des répondants. Treize pour cent (5/40)

des participants qui étaient d'accord avec la définition de [BAREFOOT 93] ont cependant jugé trop restrictif de se limiter à la communication en face à face.

Soixante-dix-huit pour cent (31/40) ont classé cette définition dans les trois premières et aucun participant ne l'a classée en dernière position.

4. *Production phonétique-acoustique*

Dans leurs définitions spontanées, en ce qui concerne l'intelligibilité, certains participants ont indiqué que la reconstruction du contenu est permise par la capacité de production phonétique-acoustique du locuteur (10 %, 4/40), afin d'obtenir un message clair (clarté : 15 %, 6/40) et facilement compris par le locuteur (facilité de compréhension : 8 %, 3/40).

Quatre-vingt-dix pour cent (36/40) sont d'accord avec la définition de [HODGE 10], qui est la seule à prendre en compte le concept de production phonétique-acoustique en lien avec l'intelligibilité (à côté des concepts de décodage acoustico-phonétique, de communication et de récupération du sens du message).

5. *Décodage acoustico-phonétique*

Spontanément, 35 % (14/40) des participants ont indiqué que pour l'intelligibilité, la reconstruction du contenu est basée sur les capacités de décodage acoustico-phonétique de l'auditeur, et qu'elle est liée aux capacités sensorielles de l'auditeur (5 %, 2/40).

Quatre-vingt-quinze pour cent (38/40) étaient d'accord avec la définition de [GHIO 18] et 100 % avec celle de [HUSTAD 08], qui lient toutes deux exclusivement le concept de décodage acoustico-phonétique à l'intelligibilité. Cependant, certains participants (8 %, 3/40) ont émis des doutes sur la limitation au niveau phonémique et l'exclusion d'éléments de niveau supérieur.

6. *Éléments contextuels*

Dans leur définition spontanée de la compréhensibilité, certains participants ont indiqué que la reconstruction du message par l'auditeur combine le décodage acoustico-phonétique (5 %, 2/40) avec des éléments contextuels (48 %, 19/40), en s'appuyant par exemple sur des capacités linguistiques (10 %, 4/40) ou sur le langage non verbal (15 %, 6/40).

Quatre-vingt-dix-huit pour cent (39/40) étaient d'accord avec la définition de [YORKSTON 96], qui assimile l'intelligibilité au décodage acoustico-phonétique et la compréhensibilité à la reconstruction du sens du message à l'aide d'informations syntaxiques, sémantiques et contextuelles. Quinze pour cent (6/40), tout en étant d'accord avec cette définition, ont estimé que l'intelligibilité incorpore également des informations indépendantes du signal. Des participants qui étaient d'accord ainsi que des participants qui n'étaient pas d'accord avec la définition de [HODGE 10] ont souligné qu'elle décrit davantage la compréhensibilité que l'intelligibilité.

L'ensemble de ces résultats a été utilisé pour rédiger 16 énoncés binaires (Accord/Désaccord) sur la terminologie pour le **deuxième tour**, ciblant chacun des six concepts principaux décrits précédemment. Ces énoncés ainsi que les pourcentages d'accord au tour 2 sont présentés dans le [Tableau 1](#). Les énoncés originaux (en anglais) sont disponibles en [Annexe A](#).

Tableau 1 – Énoncés présentés au tour 2 et pourcentages d'accord

A. Terminologie		
Concept	Énoncé (traduit de l'anglais)	Accord
<i>Synonymie</i>	- « L'intelligibilité et la compréhensibilité sont deux termes liés à la parole mais ne sont pas synonymes. »	97 % (33/34)
	- « L'intelligibilité et la compréhensibilité réfèrent aux capacités de production de la parole. »	76 % (26/34)
	- « L'évaluation de l'intelligibilité et de la compréhensibilité ne doit pas tenir compte des capacités perceptives de l'auditeur. »	41 % (14/34)
<i>Reconstruction du contenu</i>	- « L'intelligibilité et la compréhensibilité permettent toutes deux la reconstruction d'un message par l'auditeur. »	97 % (33/34)
	- « Plus précisément, l'intelligibilité d'un message est spécifiquement portée par le signal acoustique. »	97 % (33/34)
<i>Communication fonctionnelle</i>	- « L'intelligibilité et la compréhensibilité contribuent à la communication. »	100 % (34/34)
	- « La compréhensibilité renvoie davantage à la dimension fonctionnelle de la communication que l'intelligibilité. »	94 % (32/34)
	- « La compréhensibilité fait référence à la communication fonctionnelle et ne se limite donc pas au contexte de face-à-face (p. ex. conversations téléphoniques audio). »	94 % (32/34)
<i>Production phonétique-acoustique</i>	- « Pour l'intelligibilité, la reconstruction du message est rendue possible par les capacités de production phonétique-acoustique du locuteur. »	97 % (33/34)
<i>Décodage acoustico-phonétique</i>	- « Pour l'intelligibilité, la reconstruction du message est rendue possible par les capacités de décodage acoustico-phonétique de l'auditeur. »	97 % (33/34)
	- « L'unité d'analyse la plus pertinente pour l'intelligibilité est le phonème. »	85 % (29/34)
<i>Éléments contextuels</i>	- « La compréhensibilité combine, en plus du décodage acoustico-phonétique, des éléments contextuels, tels que le contexte linguistique ou non verbal. »	94 % (32/34)
	- « L'intelligibilité, qui réfère au décodage acoustico-phonétique, est donc une composante de la compréhensibilité. »	94 % (32/34)
	- « L'intelligibilité inclut également des éléments indépendants du signal. »	76 % (26/34)

B. Mesures de la parole		
Concept	Énoncé (traduit de l'anglais)	Accord
<i>Mesures perceptives</i>	- « Les évaluations au niveau des mots sont également liées à l'intelligibilité. »	100 % (34/34)
	- « Perceptivement, la compréhensibilité est évaluée au mieux au moyen de mesures liées au sens. »	94 % (32/34)
	- « Les phrases non prédictibles permettent une évaluation perceptive de l'intelligibilité. »	91 % (31/34)

	- « Perceptivement, l'intelligibilité est évaluée au mieux en utilisant des scores de % correct sur des éléments de bas niveau (phonèmes, syllabes, pseudomots). »	76 % (26/34)
<i>Mesures</i> « objectives »	- « Les mesures objectives suprasegmentales (débit, intonation, accentuation...) concernent à la fois l'intelligibilité et la compréhensibilité. »	91 % (31/34)
	- « La qualité vocale réfère également à l'intelligibilité, car elle contribue au décodage acoustico-phonétique. »	85 % (29/34)
	- « Aucune mesure objective ne permet d'évaluer directement la compréhensibilité. »	76 % (26/34)
	- « L'intelligibilité est évaluée au mieux à l'aide de mesures acoustiques au niveau du phonème. »	74 % (25/34)

Note : Les énoncés sont triés par concept cible (pour la terminologie) et par type de mesure (pour les mesures de la parole), puis par ordre décroissant d'accord moyen. En gras, les énoncés pour lesquels le seuil de consensus n'a pas été atteint ou qui se sont avérés contradictoires avec d'autres résultats. Ces énoncés ont été reformulés dans le tour suivant.

Un seul énoncé n'a pas atteint le seuil d'accord lors de ce **deuxième tour** : seuls 41 % (14/34) des participants étaient d'accord avec l'affirmation « L'évaluation de l'intelligibilité et de la compréhensibilité ne doit pas tenir compte des capacités perceptives de l'auditeur ». Dans leurs commentaires, les participants en désaccord ont souligné que la perception fait partie de la communication et doit être prise en compte (95 %, 19/20), bien que dans une plus grande mesure pour la compréhensibilité (30 %, 6/20). De plus, l'analyse des commentaires des participants a révélé que la perception pouvait être interprétée à différents niveaux de la boucle de communication : au niveau auditif périphérique (dépistage auditif), mais aussi au niveau des compétences linguistiques réceptives de l'auditeur, ainsi qu'en rapport avec l'environnement sonore (p. ex. bruit de fond).

C'est pourquoi, au **tour 3**, l'énoncé a été précisé comme suit : « Dans le contexte de l'évaluation perceptive, bien que les facteurs de perception de la parole du côté de l'auditeur doivent être contrôlés au préalable (c.-à-d. son audition, mais aussi les compétences linguistiques réceptives et l'environnement sonore), l'intelligibilité et la compréhensibilité sont utilisées pour évaluer la production de la parole du côté du locuteur ». Reformulé ainsi, 85 % (28/33) des participants étaient d'accord avec cet énoncé. Trois de ceux qui n'étaient pas d'accord (60 %) ont à nouveau indiqué que les deux concepts, mais plus spécifiquement la compréhensibilité, incluent également la capacité de l'auditeur à reconstruire le contenu transmis.

Terminologie – Autres Termes

Outre l'intelligibilité et la compréhensibilité, 70 % (28/40) des experts utilisent également d'autres termes, qui peuvent être regroupés en trois catégories principales (avec un total de 72 mentions) :

- termes génériques (43 %, 31/72) : p. ex. articulation/précision articulatoire (8), caractère naturel (6), sévérité (5) ;
- termes spécifiques (36 %, 26/72) : p. ex., liés à la voix (phonation, qualité vocale, intensité... ; 7), prosodie (6) ;
- autres (21 %, 15/72) : p. ex. relatifs au contexte pathologique/taxonomie (constance des erreurs, typicité... ; 10) ou à l'impact fonctionnel/la qualité de vie (2).

Mesures de la parole

Nous allons maintenant décrire les résultats des trois tours concernant les mesures perceptives et « objectives » de l'intelligibilité et de la compréhension préconisées par les experts.

Au premier tour, les participants devaient indiquer à l'aide de choix multiples (avec options « Autre » et « Aucune ») les mesures perceptives et « objectives » pour les deux concepts.

1. Mesures perceptives

Selon les réponses des participants au **premier tour** concernant les mesures perceptives qui caractérisent le mieux l'intelligibilité et la compréhension de la parole (voir [Figure 2](#)) :

- l'intelligibilité est mesurée au mieux en utilisant des scores de transcription orthographique (p. ex. le % d'items corrects) :
 - o en utilisant des mots réels (50 %, 20/40) ;
 - o au niveau des phonèmes (48 %, 19/40) ;
 - o en utilisant des pseudomots/non-mots (38 %, 15/40) ;
 - o en utilisant des phrases non prédictibles (38 %, 15/40).
- la compréhension est évaluée au mieux à l'aide de mesures liées à la sémantique :
 - o questions sur le contenu sémantique (60 %, 24/40) ;
 - o jugement sémantique sur des phrases (vrai-faux) (50 %, 20/40) ;
 - o sélection d'images basée sur des phrases (43 %, 17/40) ;
 - o évaluation subjective globale sur des échelles de Likert ou des échelles visuelles analogiques (38 %, 15/40).

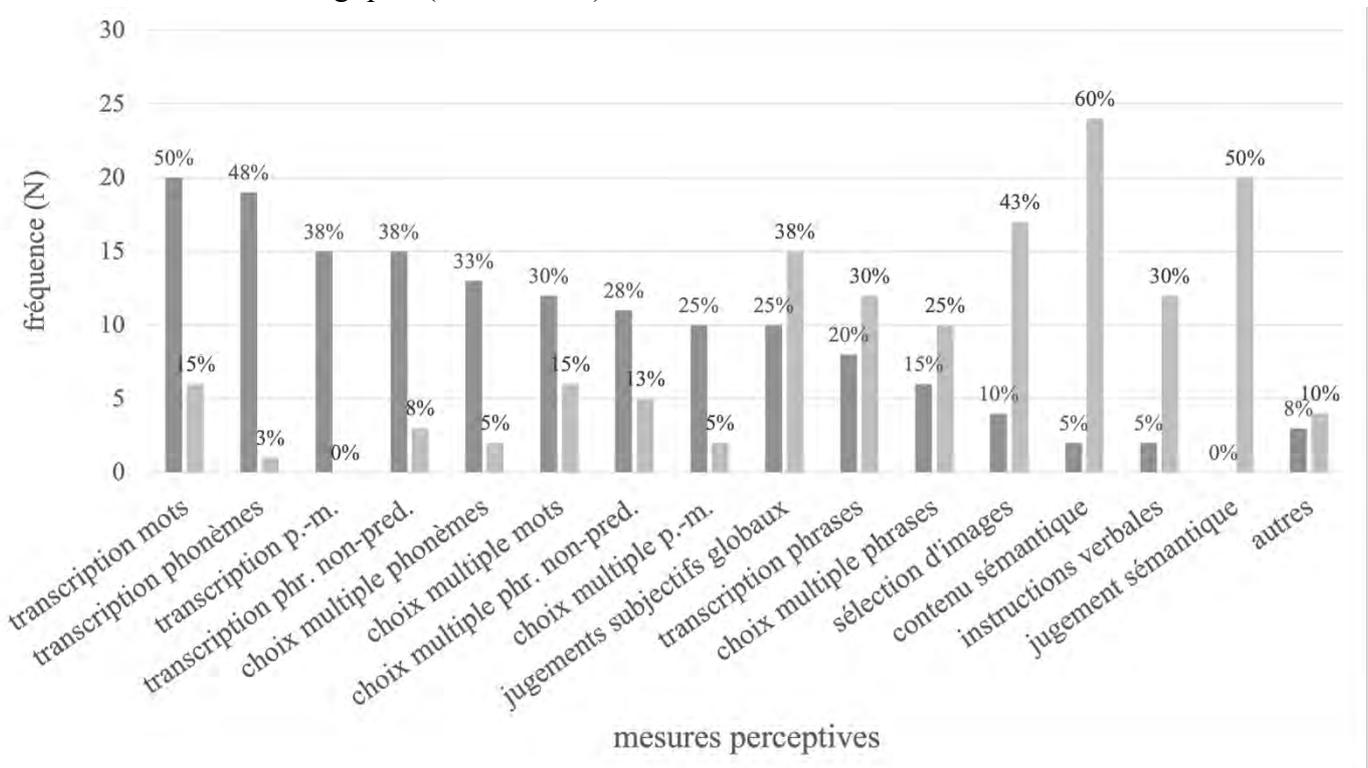


Figure 2 – Mesures perceptives qui caractérisent au mieux l'intelligibilité de la parole (gris foncé) et la compréhension (gris clair) ; pour faciliter la visualisation, les résultats ont été classés par ordre décroissant pour les mesures d'intelligibilité ; p.-m. : pseudomots, phr. non-pred. : phrases non prédictibles

Les mesures sur les transcriptions d'unités linguistiques de bas niveau (phonèmes, pseudomots, mots) sont le plus souvent utilisées pour l'intelligibilité (38 % pour les pseudomots, 48 % pour les phonèmes, 50 % pour les mots).

Des mesures de plus haut niveau sont préférées pour l'évaluation de la compréhensibilité (p. ex. 60 % pour les questions de contenu sémantique). Les scores au niveau du mot sont associés à l'intelligibilité plus qu'à la compréhensibilité, ce qui est cohérent avec la réserve exprimée concernant la réduction de l'intelligibilité au seul niveau des phonèmes. Trois participants (7 %) soulignent que les mesures au niveau du mot sont plus fonctionnelles et permettent de prendre en compte la coarticulation. Deux autres (5 %) soulignent que l'utilisation de mesures au niveau du mot reste un défi en raison de la mémorisation par l'auditeur et des processus de compensation basés sur ses connaissances linguistiques. En outre, deux autres experts (5 %), tout en convenant que les unités de bas niveau sont d'un intérêt majeur pour évaluer l'intelligibilité de la parole, soulignent que les symptômes au niveau de la phrase, comme la respiration dans la dysarthrie, s'en retrouvent négligés.

2. Mesures « objectives »

Selon les réponses des participants au **premier tour** concernant les mesures « objectives » de l'intelligibilité et la compréhensibilité de la parole (voir [Figure 3](#)) :

- l'intelligibilité est évaluée au mieux par des mesures acoustiques :
 - o sur les consonnes (63 %, 25/40) et les voyelles (53 %, 21/40) ;
 - o au niveau suprasegmental (40 %, 16/40) ;
 - o de qualité vocale (35 %, 14/40).
- la compréhensibilité n'est pas évaluée objectivement (68 %, 27/40), ou est évaluée à l'aide de mesures suprasegmentales (33 %, 13/40).

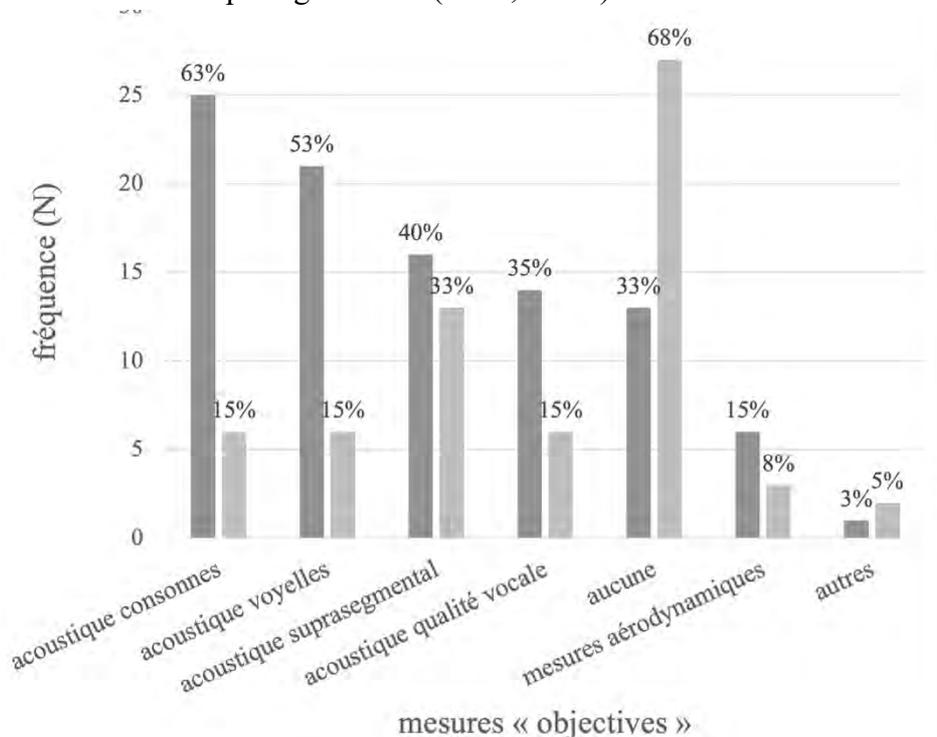


Figure 3 – Mesures « objectives » qui caractérisent au mieux l'intelligibilité de la parole (gris foncé) et la compréhensibilité (gris clair)

Sur base de ces résultats, six énoncés binaires (Accord/Désaccord) concernant l'évaluation de l'intelligibilité et de la compréhensibilité ont été élaborés pour le **deuxième tour**. Ces énoncés, ainsi que les pourcentages d'accord qui en résultent au tour 2, sont présentés dans le [Tableau 1](#).

Un seul énoncé, relatif à la granularité/au niveau d'analyse pour l'évaluation acoustique de l'intelligibilité, n'a pas atteint le seuil d'accord lors du tour 2 : « L'intelligibilité est évaluée au mieux à l'aide de mesures acoustiques au niveau du phonème ». Soixante-quatorze pour cent (25/34) des experts étaient d'accord avec cette affirmation. Ceux qui n'étaient pas d'accord ont précisé que ces mesures ne sont pas exclusives et qu'une combinaison de mesures au niveau des phonèmes, des mots et des phrases est recommandée, en tenant compte des différents contextes phonémiques. Cependant, d'après les commentaires des participants, il est apparu que certains d'entre eux avaient interprété l'énoncé comme faisant référence à des mesures sur des phonèmes isolés uniquement, ne tenant donc pas compte du contexte phonémique. Par conséquent, cet énoncé a été reformulé lors du **troisième tour** : « L'intelligibilité est évaluée au mieux à l'aide de mesures acoustiques sur les consonnes, les voyelles et les glides (y compris les transitions de formants entre phonèmes), qu'il s'agisse de phonèmes isolés ou intégrés dans des syllabes, des (pseudo)mots ou des phrases ». Pour cette nouvelle affirmation plus spécifique, le seuil de consensus a été atteint, avec 79 % (26/33) des experts en accord. Trois de ceux qui restaient en désaccord (43 %) ont précisé que ce qui les dérangeait était la formulation « au mieux ». En effet, selon eux, il n'existe pas de « meilleure » façon de faire, mais plutôt une nécessité de considérer plusieurs concepts et dimensions. Les termes « au mieux »/« meilleur » ont donc été évités lors de la rédaction de notre définition (voir « Résultat final » ci-après).

Toujours en ce qui concerne l'évaluation de l'intelligibilité, un second énoncé du **deuxième tour** a retenu notre attention : « L'intelligibilité inclut également des éléments indépendants du signal. » Bien qu'il ait atteint le seuil de consensus (76 % [26/34] étaient d'accord), ce résultat s'est révélé incohérent avec d'autres réponses (par exemple, 97 % d'accord avec l'affirmation « L'intelligibilité d'un message est spécifiquement portée par le signal acoustique »). Il semblait donc y avoir une ambiguïté dans l'expression « éléments indépendants du signal », ambiguïté qui a été explicitement évoquée dans les commentaires de certains participants. Par « éléments indépendants du signal », nous entendions toutes les informations qui ne sont pas transmises par le signal acoustique, y compris la connaissance du sujet de la conversation, les connaissances générales, l'utilisation du contexte linguistique et de la communication non verbale... Par conséquent, ces « éléments indépendants du signal » faisaient référence aux processus cognitifs descendants (« top-down »), indépendamment des processus de décodage acoustico-phonétique ascendants (« bottom-up »). L'énoncé a donc été reformulé lors du **troisième tour**, comme suit : « L'intelligibilité, par opposition à la compréhensibilité, n'inclut pas les éléments indépendants du signal (c.-à-d. les informations provenant des processus cognitifs descendants : connaissance du sujet de la conversation, connaissances générales, utilisation du contexte linguistique et de la communication non verbale...). ». Cette nouvelle formulation a donné lieu à 91 % (30/33) d'accord. Les éléments indépendants du signal ont donc été associés à la compréhensibilité plutôt qu'à l'intelligibilité dans la définition ci-après.

Résultat final

L'objectif de cette étude était d'élaborer une définition plus consensuelle de l'intelligibilité et de la compréhensibilité. Tout au long du processus Delphi, il est rapidement apparu que les orthophonistes/phoniatries, les informaticiens, les linguistes, les audiologistes, etc. ont des visions légèrement différentes mais complémentaires des concepts en question. Le paragraphe suivant reprend tous les éléments consensuels recueillis au fil du processus Delphi tout en conciliant les points de vue des différents domaines d'expertise (version originale en anglais en [Annexe A](#)) :

Intelligibilité et compréhensibilité

L'intelligibilité et la compréhensibilité sont deux termes relatifs à la parole, mais ils ne sont pas synonymes. Ils font tous deux référence à l'évaluation des capacités de production du locuteur et contribuent tous deux à la communication. Par conséquent, bien que la production de la parole soit ciblée, les facteurs de perception de la parole par l'auditeur ne peuvent être négligés (c.-à-d. les troubles auditifs du côté de l'auditeur doivent au moins être exclus).

L'**intelligibilité** fait référence à la reconstruction d'un *énoncé* au niveau acoustico-phonétique, l'information liée à l'intelligibilité est donc portée par le signal acoustique (c.-à-d. l'intelligibilité se concentre sur l'information dépendante du signal). Cette reconstruction est rendue possible à la fois par les capacités de production phonétique-acoustique du locuteur et par les capacités de décodage acoustico-phonétique de l'auditeur.

Perceptivement, l'intelligibilité est évaluée au mieux sur des stimuli à faible prédictibilité : les phonèmes, syllabes et pseudomots, mais aussi les mots (en paires minimales) et les phrases non prédictibles pour une évaluation plus fonctionnelle prenant en compte la coarticulation et les symptômes au niveau de la phrase (p. ex. la respiration et la prosodie), tant que les processus de compensation cognitive descendante de l'auditeur sont court-circuités (c'est-à-dire, sans aide du contexte sémantique ou linguistique).

Instrumentalement, l'intelligibilité peut être évaluée à l'aide de mesures acoustiques sur les consonnes, les voyelles et les glides (y compris les transitions de formants entre phonèmes), qu'il s'agisse de phonèmes isolés ou intégrés dans des syllabes, des (pseudo-)mots ou des phrases. En outre, dans certains cas, la qualité de la voix contribue également à l'intelligibilité, car elle joue un rôle dans certains contrastes phonémiques. Les paramètres suprasegmentaux (évalués objectivement par exemple par le débit de parole ou l'accentuation) contribuent également à l'intelligibilité.

La **compréhensibilité** fait référence à la reconstruction d'un *message* au niveau sémantique-discursif, ultérieure à la reconstruction acoustico-phonétique. L'intelligibilité est donc une composante de la compréhensibilité. Outre le décodage acoustico-phonétique, cette dernière comprend également des éléments contextuels indépendants du signal, tels que le contexte linguistique ou non verbal. Cependant, l'on peut être compréhensible sans que toutes les unités de bas niveau soient nécessairement décodées avec précision ; par conséquent, si l'intelligibilité affecte la compréhensibilité, cette dernière n'en dépend toutefois pas entièrement.

La compréhensibilité renvoie à la dimension plus fonctionnelle de la communication et, perceptivement, elle est appréciée au mieux à l'aide d'évaluations liées au sens (c'est-à-dire en tenant compte des processus cognitifs descendants qui peuvent compenser la dégradation des informations acoustiques et phonétiques). Aujourd'hui, aucune mesure instrumentale « objective » n'est encore adaptée pour évaluer la compréhensibilité à proprement parler (c'est-à-dire la transmission du sens global du message). Cependant, certains paramètres suprasegmentaux contribuent à la compréhensibilité et peuvent être évalués objectivement (p. ex. les mesures de débit et d'intonation).

C. Discussion

Groupe d'experts

S'il a été démontré que la taille de l'échantillon dans les études Delphi a un impact significatif sur les indices de consensus [BIRKO 15], il n'existe pas de directives claires quant à la taille d'échantillon recommandée. Les groupes d'experts varient d'un nombre aussi restreint que six à des tailles d'échantillon supérieures à 1000 experts, en fonction du sujet traité et des ressources disponibles [C. POWELL 03] ; une taille de groupe de 12 à 15 experts a été rapportée dans de nombreuses études Delphi [MCPHERSON 18]. Dans la nôtre, 40 experts ont participé au premier tour. En plus de la taille initiale de l'échantillon, il est important de minimiser l'attrition du nombre de participants tout au long du processus Delphi, car les abandons peuvent conduire à une surestimation du consensus final et entraver la fiabilité des résultats [SINHA 11]. Par conséquent, dans la lettre d'invitation, les participants à cette étude ont été invités à prendre part uniquement si la participation à l'ensemble des quatre tours initialement prévus était envisagée. Cela a potentiellement limité le nombre de participants initial, mais a finalement contribué à un faible taux d'abandon de 17 % (7/40) entre le premier et le troisième tour, alors que des taux de 20 à 30 % sont généralement attendus [CHALMERS 19]. En plus de la taille d'échantillon satisfaisante de notre étude, les caractéristiques du groupe d'experts sont également intéressantes, puisqu'il comprend des participants issus de divers domaines liés à la parole, avec des parcours et des contextes culturels et linguistiques variés. Il est également à noter que les profils du groupe sont restés constants tout au long du processus Delphi, sans changement majeur dans les professions représentées, les pays, le niveau d'éducation, les langues, l'ancienneté et les populations de patients. Par conséquent, notre groupe Delphi a été considéré comme satisfaisant à la fois en termes de taille et de composition pour répondre à l'objectif principal de cette étude, que nous allons maintenant aborder.

Terminologie

Cette étude Delphi a abouti à une définition étayée de l'intelligibilité et de la compréhensibilité et de leur évaluation, intégrant l'ensemble des éléments consensuels identifiés tout au long du processus. La [Figure 4](#) résume cette définition et illustre la relation entre les deux concepts.

Une façon de différencier l'intelligibilité et la compréhensibilité dans la définition proposée est le terme utilisé pour désigner le processus de (re)construction respectif : le terme « énoncé » (« *utterance* ») est employé pour référer au contenu acoustique et phonétique de la parole (et, par conséquent, à l'intelligibilité), tandis que le terme « message » est utilisé comme un terme plus général référant à la (re)construction au niveau sémantique (c.-à-d. à la compréhensibilité, qui comprend également les éléments d'intelligibilité). En effet, comme deux participants l'ont suggéré dans leurs commentaires, alors que le terme « message » peut être défini comme faisant référence au « thème ou à l'idée sous-jacente » (donc, avec une idée de « signification »), le terme « énoncé » se détache du contenu sémantique communiqué et se rapporte plutôt au signal acoustique transmis.

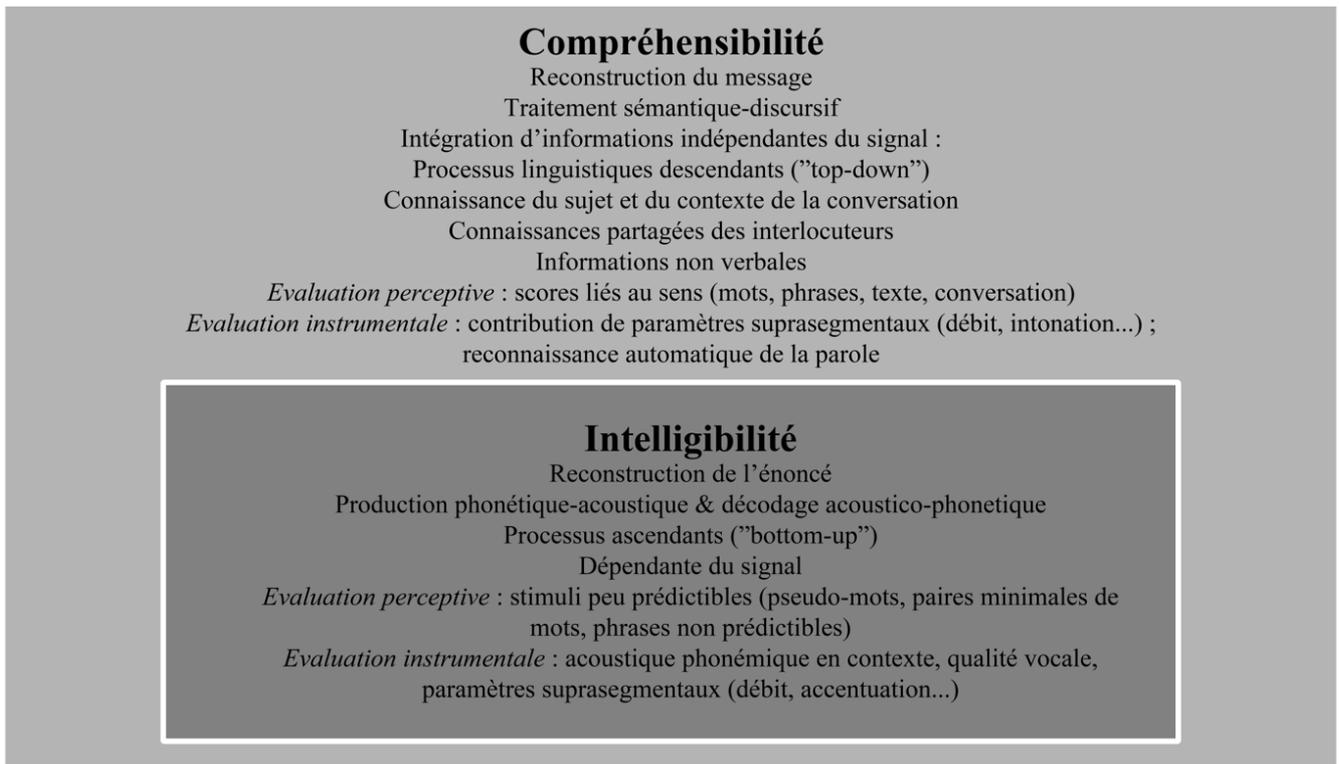


Figure 4 – L'intelligibilité et la compréhensibilité dans la production de la parole

Si l'intelligibilité et la compréhensibilité sont principalement destinées à décrire la production de la parole d'un individu, le côté de l'auditeur, depuis lequel les deux concepts sont généralement évalués et définis, joue également un rôle important. En effet, il s'est avéré sans équivoque dans notre étude Delphi que l'intelligibilité et la compréhensibilité contribuent toutes deux à la communication, et que la communication humaine fonctionnelle (« échange verbal entre un locuteur et un interlocuteur dont il sollicite une réponse »⁶) nécessite par définition un locuteur et un auditeur. Schramm décrivait déjà en 1954 la communication comme un échange bidirectionnel de messages entre deux « interprètes » [SCHRAMM 54]. Dans son modèle interactionnel (voir Figure 5), un message est encodé et transmis via un canal, puis décodé par le récepteur qui, à son tour, fournit un retour à l'émetteur. Bien que restrictif, ce modèle soulignait déjà l'importance du locuteur et de l'auditeur et de leurs domaines d'expérience respectifs dans le processus de communication.

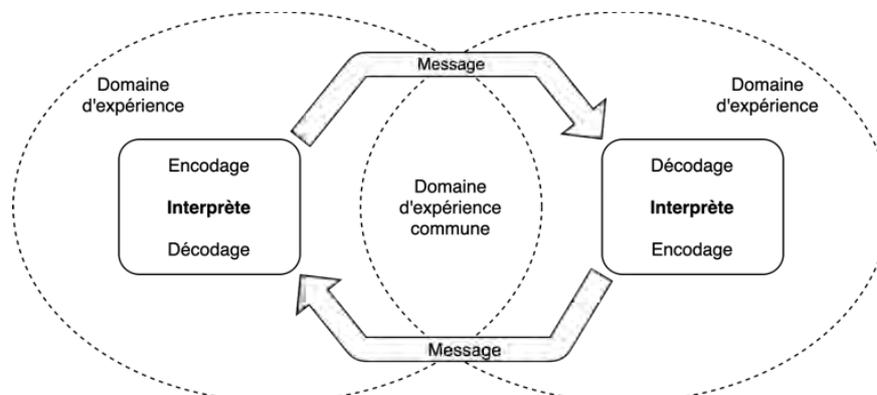


Figure 5 – Modèle interactionnel de la communication de [SCHRAMM 54]

⁶ <https://www.larousse.fr/dictionnaires/francais/communication/17561>

L'intelligibilité intègre donc non seulement la précision de la production phonétique-acoustique par le locuteur, mais aussi le décodage acoustico-phonétique par l'auditeur. La compréhensibilité implique en outre de nombreux facteurs de plus haut niveau, qui sont également liés à la fois au locuteur (p. ex. des indices non verbaux et l'intonation pour pallier une faible intelligibilité) et à l'auditeur (p. ex. la connaissance que l'auditeur a du locuteur, de ses intentions et de ses émotions).

La [Figure 6](#), que nous avons construite sur base du modèle de communication orale de Levelt [LEVELT 93], illustre plus en détail l'intelligibilité et la compréhensibilité dans la boucle de communication. À chacun des niveaux de cette dernière, des variations peuvent apparaître en fonction des caractéristiques naturelles ou pathologiques du locuteur et de l'auditeur (p. ex. genre, âge, accent régional, trouble de la parole). Les étapes de cette boucle qui constituent le concept d'intelligibilité sont surlignées en gris (cf. aussi l'[Annexe C](#) pour des exemples de facteurs physiologiques ou pathologiques pouvant affecter spécifiquement l'intelligibilité). Les autres niveaux, qui se rapportent à la (re)construction sémantique et pragmatique du message véhiculé, constituent le concept de compréhensibilité. Ce dernier, en plus du contenu linguistique, comprend également des indices paralinguistiques⁷ (p. ex. soupirs, grognements) et extralinguistiques (p. ex. langage corporel et expressions faciales), ainsi que des éléments contextuels (p. ex. connaissance préalable du sujet de conversation et de l'interlocuteur et domaine d'expérience commune), qui peuvent faciliter la reconstruction du message transmis.

Outre les caractéristiques du locuteur et de l'auditeur, l'intelligibilité et la compréhensibilité peuvent être affectées par les caractéristiques du canal de communication (p. ex. dans le cas d'une restitution incomplète du contenu fréquentiel du signal lors d'une conversation téléphonique) et du contexte de communication (p. ex. en raison du bruit de fond ou de l'acoustique environnementale, de la réverbération).

Mesures de la parole

Comme nous l'avons développé ci-avant, l'intelligibilité et la compréhensibilité sont le plus souvent définies du point de vue de l'auditeur, en termes de reconstruction du contenu de parole transmis plutôt que par sa construction initiale par le locuteur. Les définitions de ces deux concepts intègrent très souvent des éléments relatifs aux mesures perceptives et acoustiques.

Sur le plan perceptif, l'intelligibilité peut être évaluée à l'aide de stimuli peu prédictibles (phonèmes, syllabes, pseudomots, paires minimales de mots ou phrases non prédictibles)⁸.

⁷ Définis comme « tout ce que l'on peut trouver dans le signal de parole [...] et qui ne peut être décrit uniquement en termes strictement phonétiques et/ou linguistiques » [SCHULLER 13]. Notons que les indices paralinguistiques peuvent être attribués à la fois à l'intelligibilité et à la compréhensibilité. En effet, les facteurs tels que le débit de parole et la qualité vocale peuvent avoir un impact direct sur l'intelligibilité [EVITTS 16; TJADEN 14]. Cependant, les caractéristiques paralinguistiques du signal de parole peuvent également fournir des informations supplémentaires au contenu linguistique et ainsi augmenter plus directement la compréhensibilité, sans nécessairement impacter l'intelligibilité (p. ex. si le locuteur augmente le volume vocal pour indiquer l'insatisfaction).

⁸ En ce qui concerne les unités d'analyse de l'intelligibilité, il est nécessaire de souligner que dans certaines langues, comme les langues tonales, les mesures suprasegmentales (par exemple la courbe intonative) contribuent vraisemblablement à un degré plus élevé à l'intelligibilité que dans les langues occidentales, par rapport aux mesures au niveau du phonème [CHEN 11].

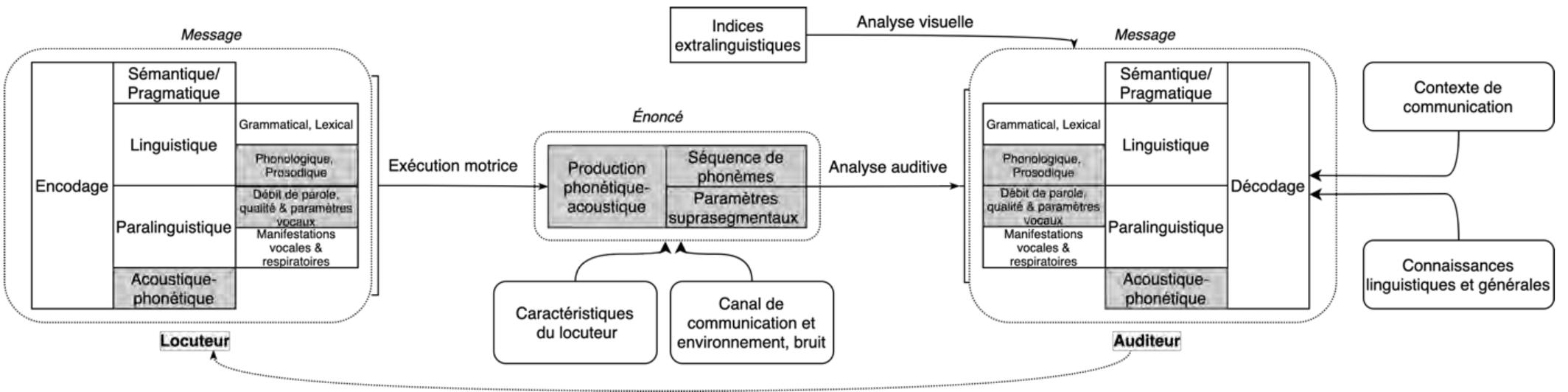


Figure 6 – Boucle de communication orale ; en gris : étapes relatives à l'intelligibilité

L'utilisation de mots réels et de phrases standard, bien que plus fonctionnelle, est sujette à la mémorisation par l'auditeur et à des processus de compensation basés sur ses connaissances linguistiques et se prête donc davantage à l'évaluation de la compréhensibilité. Toutefois, en parole naturelle, les unités de bas niveau (séquences de phonèmes/syllabes), ciblées dans l'évaluation de l'intelligibilité, sont malgré tout intégrées dans un énoncé avec ses caractéristiques prosodiques (suprasegmentales). Les caractéristiques segmentales et suprasegmentales contribuent donc toutes deux à l'intelligibilité. Pour cette raison, les paires minimales de mots et les phrases non prédictibles ont été incluses dans notre définition, en plus des phonèmes, des syllabes et des pseudomots. Elles minimisent l'influence des informations linguistiques tout en permettant de prendre en compte les phénomènes coarticulatoires et ceux du niveau phrastique.

Afin d'évaluer l'intelligibilité avec moins de subjectivité, il peut être envisagé de se passer de la dimension « auditeur ». Une solution consiste à recourir à des mesures assistées par ordinateur. Pour l'évaluation acoustique de l'intelligibilité — tout comme pour nous venons de le décrire pour l'évaluation perceptive — bien que les mesures acoustiques au niveau du phonème soient un indicateur instrumental primordial, les contextes phonémiques et suprasegmentaux dans lesquels les unités ciblées sont produites doivent être pris en compte. Idéalement, les mesures d'intelligibilité devraient donc être effectuées sur de la parole continue afin de permettre une évaluation plus fonctionnelle, reflétant les conditions naturelles de la parole. Ainsi, l'évaluation automatique de l'intelligibilité, par exemple, est généralement réalisée sur la parole continue, mais en excluant des algorithmes les indices contextuels pour se concentrer sur l'évaluation de la production phonétique-acoustique [FREDOUILLE 19]. L'évaluation automatique de la compréhensibilité, elle, reste cependant toujours problématique. Le *modus operandi* habituel des systèmes de reconnaissance automatique de la parole (RAP), par exemple, ne peut être que partiellement associé au concept de compréhensibilité. Certains indices contextuels sont en effet utilisés par les algorithmes pour reconstruire les messages prononcés. La compréhensibilité est alors évaluée en examinant la précision des résultats de ces systèmes de RAP. Une analyse plus approfondie des aspects lexicaux, syntaxiques et sémantiques de la production du locuteur est ensuite possible. Cependant, la communication humaine est bien plus complexe et comporte de nombreuses dimensions para- et extralinguistiques que les systèmes de RAP actuels ne prennent pas encore en compte. Néanmoins, il est à noter que la prise en compte des informations extra- et paralinguistiques et contextuelles fait l'objet de plus en plus d'investigations, notamment dans le domaine des interactions personne-machine, et que des résultats prometteurs ont été observés [KENNINGTON 15; PORZEL 11; SCHULLER 13, 19].

En conclusion, il convient donc de prendre en compte à la fois les mesures acoustiques et perceptives, à différents niveaux de granularité (segmental et suprasegmental), et sur divers échantillons de parole (phonèmes et syllabes isolés, mots, pseudomots et phrases). Seule leur combinaison permet une évaluation exhaustive de l'intelligibilité et de la compréhensibilité et offre ainsi des informations sur la parole du patient tant au niveau segmental que fonctionnel.

D. Conclusion du chapitre et perspectives

Notre enquête de consensus Delphi a permis de rédiger une définition étayée de l'intelligibilité et de la compréhensibilité de la parole, incluant tous les éléments consensuels recueillis au fil du processus et prenant ainsi en compte les points de vue de différents domaines d'expertise. Ce processus a permis de mieux comprendre ces deux termes couramment utilisés dans le domaine de la parole et ainsi de spécifier leur évaluation en décrivant les tâches et mesures qui correspondent le mieux à leur définition. Bien que l'intelligibilité et la compréhensibilité soient liées et contribuent toutes deux à la communication fonctionnelle, elles concernent deux niveaux différents de reconstruction du contenu de parole transmis. L'intelligibilité fait référence au décodage acoustico-phonétique de l'énoncé, tandis que la compréhensibilité désigne la reconstruction du sens du message. Par conséquent, l'évaluation perceptive de l'intelligibilité nécessite l'utilisation d'échantillons de parole non prédictibles (pseudomots, paires minimales de mots, phrases non prédictibles), tandis que l'évaluation de la compréhensibilité repose sur le sens et le contexte et implique des stimuli et des tâches de production de parole plus fonctionnels.

De nombreuses ambiguïtés persistent dans la terminologie liée à l'évaluation de la parole, par exemple en ce qui concerne les termes « objectif », « subjectif », « instrumental », « perceptif », « outil », « moyen » et « mesure », pour lesquels nous trouvons diverses interprétations dans la littérature. Comme l'a souligné un expert du panel, « ... il semblerait que ce soit la sémantique interlinguistique qui contribue à ce problème épineux plutôt que les principes scientifiques sur lesquels nous sommes probablement d'accord ». Par conséquent, des études supplémentaires sont nécessaires pour clarifier les questions liées à la terminologie, générer des définitions moins équivoques et faciliter les avancées de la recherche sur la parole et la voix grâce à une meilleure communication entre les experts. La méthodologie Delphi semble être un moyen approprié à cette fin, au vu de ses caractéristiques que nous avons décrites ci-avant. Par le recours à de multiples itérations, le processus Delphi favorise une approche plus réfléchie pour la résolution de problèmes [HSU 07]. Comme l'ont souligné les participants, cette méthode est ainsi « propice à la réflexion et intellectuellement stimulante », et remet parfois en question les termes quotidiennement utilisés et ancrés dans les usages. Enfin, dans l'attente d'une terminologie plus consensuelle dans le domaine de l'évaluation de la parole, nous recommandons fortement aux auteurs de définir clairement les concepts visés lorsqu'ils introduisent leur recherche, afin d'éviter toute ambiguïté et de permettre à des personnes de différents horizons de comprendre sans équivoque le sens attribué.

Maintenant que nous avons défini l'intelligibilité, nous pouvons nous pencher plus en détail sur la question de sa mesure.

Chapitre 2. Analyse des besoins

Dans ce chapitre, nous allons aborder les besoins en termes de mesure de l'intelligibilité, autant au niveau clinique qu'en recherche scientifique. À cette fin, nous nous baserons sur le modèle de l'Evidence-Based Practice (EBP) [SACKETT 96], qui servira de fil conducteur dans cet état de l'art.

2.1. L'Evidence-Based Practice

L'EBP est le plus souvent définie comme « l'utilisation consciencieuse, explicite et judicieuse des meilleures preuves actuelles dans la prise de décisions concernant les soins de chaque patient » [SACKETT 96].

C'est dans les années 90 qu'apparaît pour la première fois le terme « Evidence-Based Medicine » (EBM). Cette approche, dans les sciences de la santé, est née de la nécessité de désaccentuer l'intuition, les observations non systématiques dans l'expérience clinique et le raisonnement physiopathologique pur comme motifs suffisants pour la prise de décisions cliniques [EVIDENCE-BASED MEDICINE WORKING GROUP 92]. En lieu de cela, l'EBP promeut l'importance de la systématisme des observations ainsi que la prise en compte des données issues de la recherche clinique, à une époque où une grande variabilité de pratiques cliniques ainsi qu'un essor des connaissances sont observés depuis plusieurs décennies [MCCURTIN 12]. Ainsi, pour guider la prise de décisions cliniques, le courant de l'EBP vise à intégrer trois sources d'informations (voir [Figure 7](#)) : les données récentes de la recherche scientifique, l'expérience clinique et les valeurs et préférences du patient dûment informé [DODD 07; DOLLAGHAN 07].

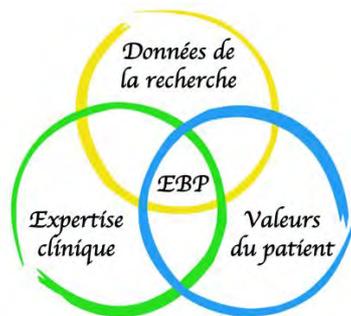


Figure 7 – Les trois principaux piliers de l'EBP

En orthophonie, ce courant est de plus en plus enseigné [DURIEUX 18] et adopté dans la pratique clinique courante [BUTLER 19]. Les étapes de base de l'EBP sont les suivantes [SWANSON 10] :

- formuler la question de recherche ;
- trouver les meilleures données de la recherche scientifique ;
- évaluer ces données de manière critique ;
- intégrer les données de la recherche, l'expertise clinique et la situation individuelle du patient ;
- évaluer l'efficacité suite aux étapes précédentes.

Ce modèle de l'EBP structurera notre analyse des besoins développée dans la suite de ce chapitre, pour un état de l'art prenant en compte ses trois principaux piliers.

2.2. Premier pilier de l'EBP : la recherche scientifique

2.2.1. Les revues systématiques de la littérature

Le premier pilier de l'EBP consiste en la prise en compte des données de la recherche systématique (voir [Figure 8](#)). Ainsi, les revues systématiques et les méta-analyses, suivies des études randomisées contrôlées, sont considérées comme les plus hauts niveaux de « preuve » de la recherche [DE MIRANDA 19; MCCURTIN 12].



Figure 8 – Prise en compte du premier pilier de l'EBP : revue systématique de la littérature

Les revues systématiques permettent d'interroger les bases de données scientifiques de manière rigoureuse pour répondre à une question spécifique, tout en limitant au maximum les biais de sélection [MCCURTIN 12]. Les cinq étapes principales d'une revue systématique de la littérature (voir [Figure 9](#)) sont les suivantes [KHAN 03] :

1. *Formuler la question de recherche :*

La question de recherche est formulée de préférence sous une forme structurée et explicite, p. ex. suivant le modèle P(ES)ICO [FINEOUT-OVERHOLT 05; SCHARDT 07; SCHLOSSER 06] : Population/Patient/Problème, Environnement, Stakeholders/Parties prenantes, Intervention, Comparaison, Outcome/Résultat ;

2. *Identifier les publications pertinentes :*

Une recherche exhaustive doit être menée dans les différentes ressources existantes, dont les bases de données en ligne, en adaptant les critères d'inclusion et d'exclusion en fonction de la question de recherche ;

3. *Évaluer la qualité des études :*

La qualité des études est évaluée sur base du design d'étude, dont la hiérarchie est communément décrite sous forme d'une pyramide des « niveaux de preuve » [BURNS 11]. Ensuite, la méthodologie et la qualité de rédaction des articles retenus sont analysées de manière détaillée en utilisant une grille d'évaluation critique (« critical appraisal tool ») adaptée aux designs retenus, telle la grille QualSyst [KMET 04] ;

4. *Résumer les données :*

Les données sont synthétisées sous forme d'un tableau reprenant les caractéristiques de chaque étude retenue (score de qualité méthodologique, design d'étude, population, mesures d'efficacité employées, conclusions principales...). Dans une méta-analyse,

des analyses statistiques sont ensuite appliquées pour comparer et combiner les résultats des études entre eux ;

5. *Interpréter les résultats :*

Enfin, l'ensemble des résultats est interprété et discuté afin de répondre à la question de recherche initiale, tout en prenant en compte d'éventuels biais de publication ou limites des études, par exemple.

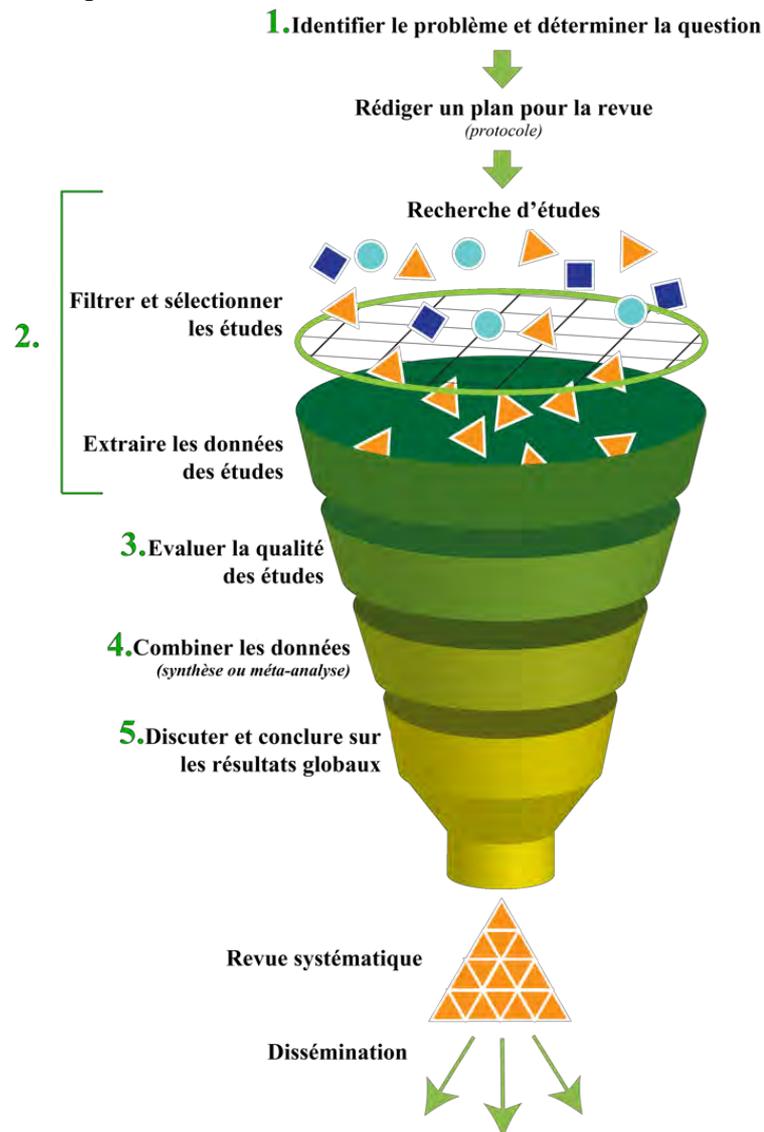


Figure 9 – Les étapes de la revue systématique de la littérature ; les chiffres verts indiquent les cinq étapes clés. Adapté et traduit de Kaufman, 2011⁹

Différentes lignes directrices ont été formulées afin de structurer les étapes de recherche et de sélection des études, ainsi que la rédaction de la revue. Ces lignes directrices ont pour objectif de maximiser la rigueur et la fiabilité méthodologiques, mais aussi la transparence du processus. L'un des outils les plus employés est le Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [LIBERATI 09; MOHER 09; PAGE 21]. C'est cette méthodologie que nous avons utilisée pour mener notre revue systématique, que nous allons maintenant décrire.

⁹ <https://ccrg.cochrane.org/sites/ccrg.cochrane.org/files/public/uploads/What%20authors%20do%204%2012%2015.pdf>

2.2.2. Besoins en termes de mesure d'intelligibilité identifiés dans la littérature

A. Variabilité en parole naturelle « saine »

Nous avons précédemment introduit l'importance d'une évaluation exhaustive de la parole pathologique pour une prise en charge individualisée et efficace de chaque patient en pratique clinique. Cependant, avant de se pencher sur la parole pathologique, il est primordial de s'attarder sur la parole dite « saine » ou « typique ». Toute mesure qui se veut applicable en pratique clinique doit en effet d'abord être considérée en parole saine, afin d'en comprendre les comportements. Encore trop souvent, la parole saine est considérée comme « parfaite » et son intelligibilité présumée à 100 % (p. ex. « Le groupe contrôle avait une intelligibilité normale » [VAN LIERDE 12]).

L'un des principaux problèmes que pose la mesure de la parole est son caractère hautement variable, en contexte pathologique comme en parole « saine » [BENZEGUIBA 07; MILLER 13]. Certains locuteurs « sains » sont en effet plus intelligibles que d'autres, et il a été montré que cela est lié à la production acoustico-phonétique du locuteur plutôt qu'à la perception de l'auditeur [BOND 94; R. M. COX 87; HAZAN 04; HOOD 80]. De nombreuses études décrivent la variabilité inter- et intralocuteur en parole saine, qui est influencée par divers facteurs tels que la coarticulation, les probabilités phonotactiques, l'état émotionnel du locuteur, le débit de parole, le style d'élocution, la physiologie, l'accent régional, le contexte de communication et bien d'autres [BENZEGUIBA 07; LINDBLOM 90; MCCLOY 15; MEUNIER 07; K. J. REILLY 13].

L'analyse des « erreurs » ou accidents de parole conduit souvent à la théorie du compromis entre vitesse et précision (« speed-accuracy trade-off ») [GUENTHER 95; MEUNIER 07; TREMBLAY 17]. Pour expliquer ce compromis, Lindblom a proposé la théorie de l'« hyper/hypo-parole » dite « H & H » [BOND 94; LINDBLOM 90], selon laquelle une intelligibilité élevée peut être atteinte par différentes stratégies acoustiques et phonétiques [R. M. COX 87; GUENTHER 95; HAZAN 04; LAVOIE 02]. Il est donc primordial de bien comprendre ces stratégies, de distinguer quelles variations peuvent être attribuées aux contraintes de la parole spontanée en contexte de communication naturelle, et quelles déviations indiquent une parole altérée à proprement parler, afin de permettre une évaluation plus précise.

Une autre question importante est de savoir si ces « imperfections » peuvent être capturées par des mesures acoustico-phonétiques. Plusieurs études, notamment sur le vieillissement physiologique [HAZAN 17; KURUVILLA-DUGDALE 20], indiquent qu'une grande partie de la variabilité de la parole saine est imputable à des caractéristiques acoustiques-phonétiques spécifiques du locuteur [BRADLOW 96; METZ 90].

Le caractère « imparfait » de la parole saine souligne la nécessité de bien comprendre les comportements des mesures qu'on y applique, avant de se pencher sur la parole pathologique. C'est ainsi que nous avons mené une revue systématique de la littérature scientifique selon les lignes directrices PRISMA, en nous intéressant aux mesures acoustiques employées dans l'investigation de l'intelligibilité en parole saine. Nous discuterons ici dans un premier temps uniquement les résultats en termes de besoins relatifs à la mesure de l'intelligibilité en parole saine. Les résultats concernant les mesures acoustiques à proprement parler seront discutés dans le [Chapitre 4](#).

B. Notre revue systématique PRISMA

Méthodologie

1. Critères d'éligibilité

Pour être inclus dans cette revue, les articles devaient :

- traiter à la fois des notions d'intelligibilité¹⁰ et d'acoustique spectrale de la parole (exclusion notamment des articles traitant de l'acoustique ambiante) ;
- étudier la parole naturelle de locuteurs sains adultes âgés de plus de 18 ans (exclusion donc également des articles étudiant la parole modifiée ou vocodée lorsqu'aucune donnée sur la parole non modifiée n'était fournie) ;
- utiliser des mesures acoustiques segmentales (et non seulement des mesures acoustiques globales, telles que le spectre moyen à long terme sur une phrase entière) ;
- être rédigés en anglais ;
- être des articles originaux (exclusion des présentations orales, des études de cas, des lettres d'auteurs, des actes de conférences et des revues) ;
- inclure au moins six locuteurs sains.

Les critères d'exclusion étaient les suivants :

- l'étude exclusive de la voix/phonation (dysphonie, mesures de qualité vocale) ;
- l'étude des langues tonales, pour lesquelles les analyses de l'intelligibilité reposent en outre sur le ton lexical et la prosodie [DING 03; YIU 94] ;
- l'utilisation exclusive de mesures temporelles (telles que la durée des voyelles ou le débit de parole) ;
- l'étude de la perception de la parole par des auditeurs malentendants ;
- l'application de techniques de traitement automatique de la parole, telles que les réseaux neuronaux profonds.

Aucun critère d'exclusion lié à la date de publication n'a été utilisé, car certaines sources pertinentes connues des auteurs remontent au milieu des années 1950.

Tous les critères d'éligibilité devaient être respectés pour que les articles soient inclus dans cette revue.

2. Sources d'information et stratégie de recherche

La recherche bibliographique a été effectuée le 4 décembre 2018 dans deux bases de données biomédicales : Embase et PubMed. Les listes de références des articles inclus ont été consultées pour identifier d'éventuels autres articles pertinents. Les termes de recherche et la syntaxe sont présentés dans le [Tableau 2](#).

¹⁰Les études utilisant des méthodes d'évaluation perceptive évaluant l'« intelligibilité » selon les auteurs mais ne correspondant pas à notre définition axée sur les unités segmentales de bas niveau n'ont pas été exclues *a priori* mais ont été différenciées dans la discussion.

Tableau 2 – Stratégie de recherche pour les deux bases de données

Base de données	Termes de recherche (descripteurs MeSH et termes en texte libre)	Nombre d'entrées
Embase	((speech intelligibility/) OR (Intelligibil*.ab. OR Intelligibil*.ti. OR comprehensibil*.ab. OR comprehensibil*.ti. OR understandabil*.ab. OR understandabil*.ti.)) AND (acoustics/ OR speech analysis/ OR acoustic analysis/ OR sound analysis/ OR phonetics/ OR signal processing/ OR fourier analysis/ OR sound detection/ OR sound/ OR frequency/ OR frequency analysis/ OR pitch/ OR noise/ OR signal noise ratio/)	3326
PubMed	((("Speech Intelligibility"[Mesh]) OR (intelligibil*[Title/Abstract] OR comprehensibil*[Title/Abstract] OR understandabil*[Title/Abstract])) AND ("Acoustics"[Mesh] OR "Speech Acoustics"[Mesh] OR "Speech Production Measurement"[Mesh] OR "Phonetics"[Mesh] OR "Signal Processing, Computer-Assisted"[Mesh] OR "Fourier Analysis"[Mesh] OR "Sound Spectrography"[Mesh] OR "Sound"[Mesh] OR "Signal-To-Noise Ratio"[Mesh] OR "Noise"[Mesh]))	3393
		Total : 6719
		Total après exclusion des doublons : 4818

Les titres et les résumés ont été récupérés via EndNote X9 et examinés par deux évaluateurs indépendants (les mêmes que précédemment), en appliquant les critères de sélection susmentionnés. En raison du grand nombre de résumés, l'ensemble a été divisé en deux. Chaque évaluateur a donc examiné la moitié de l'ensemble, plus un lot de 20 % des résumés choisis aléatoirement dans l'autre moitié. Ainsi, 40 % des résumés ont été lus par les deux évaluateurs, ce qui a permis de mesurer un Kappa pondéré pour évaluer la concordance interévaluateurs. Les différences dans les évaluations d'éligibilité ont été résolues par consensus. Les textes intégraux des articles sélectionnés ont ensuite été récupérés et examinés par chaque évaluateur. Un organigramme illustrant le processus de sélection des articles conformément aux directives PRISMA [LIBERATI 09] est présenté en [Figure 10](#) dans la section Résultats.

3. Évaluation critique de la qualité méthodologique et du niveau de preuve

La qualité méthodologique des articles sélectionnés a été évaluée à l'aide de l'outil QualSyst [KMET 04]. Cet outil a été développé sous forme d'une grille de cotation afin d'évaluer méthodologiquement les articles de recherche quantitative et qualitative. Il permet d'analyser, entre autres, le design d'étude, la question de recherche, la sélection et la description du groupe d'étude, et le contrôle des facteurs de confusion.

La hiérarchie élaborée par le National Health and Medical Research Council (NHMRC, 1999) a été utilisée pour évaluer le « niveau de preuve ». Six niveaux sont décrits :

- niveau I : niveau le plus élevé, revues systématiques d'essais contrôlés randomisés ;
- niveau II : essais contrôlés randomisés ;
- niveau III-1 : essais contrôlés pseudorandomisés ;
- niveau III-2 : études comparatives avec groupe contrôle parallèle et allocation non randomisée (études de cohorte, études cas-témoins, ou séries chronologiques interrompues avec un groupe contrôle) ;

- niveau III-3 : études comparatives sans groupe contrôle parallèle, avec groupe contrôle historique, deux ou plusieurs études à un seul bras, ou séries chronologiques interrompues sans groupe contrôle parallèle ;
- niveau IV : niveau le plus bas, séries de cas.

Les études de niveau IV et les avis d'experts n'ont pas été retenus, de même que les revues systématiques.

4. Données extraites

Après la sélection des articles basée sur les critères d'éligibilité et l'évaluation de la qualité méthodologique, les informations suivantes ont été extraites pour chaque article : la population d'étude (effectif, âge, genre, langue), l'échantillon de parole utilisé (phonèmes ciblés), les mesures acoustiques, les mesures perceptives d'intelligibilité, la principale conclusion concernant le lien entre l'acoustique et l'intelligibilité le cas échéant, et les données descriptives si disponibles.

Résultats

1. Sélection d'études

Au total, 4818 titres et résumés ont été extraits des bases de données (après élimination automatique de la plupart des doublons). Avant de se réunir pour un consensus, les évaluateurs étaient en accord sur les critères d'éligibilité pour 1792 (93 %) des 1928 résumés évalués en commun, avec un Kappa pondéré de 0,89 — correspondant à un accord « presque parfait » [LANDIS 77].

Deux cent soixante-sept textes intégraux d'articles ont été examinés, dont 22 ont été retenus. Pour des raisons de lisibilité, un code d'identification a été attribué à chacun des 22 articles (voir [Tableau 3](#)), qui sera utilisé pour les citations *in texto* tout au long de ce chapitre.

Neuf de ces études portaient sur le lien entre les mesures acoustiques spectrales et perceptives (A01-A09). Les 13 articles restants, bien que n'évaluant pas le lien en soi, ont été retenus parce qu'ils fournissaient des données quantitatives à la fois pour les mesures acoustiques et pour les évaluations perceptives chez des locuteurs sains.

Le processus de sélection des études est illustré dans la [Figure 10](#). Une synthèse détaillée des 22 études incluses est disponible dans l'[Annexe D](#).

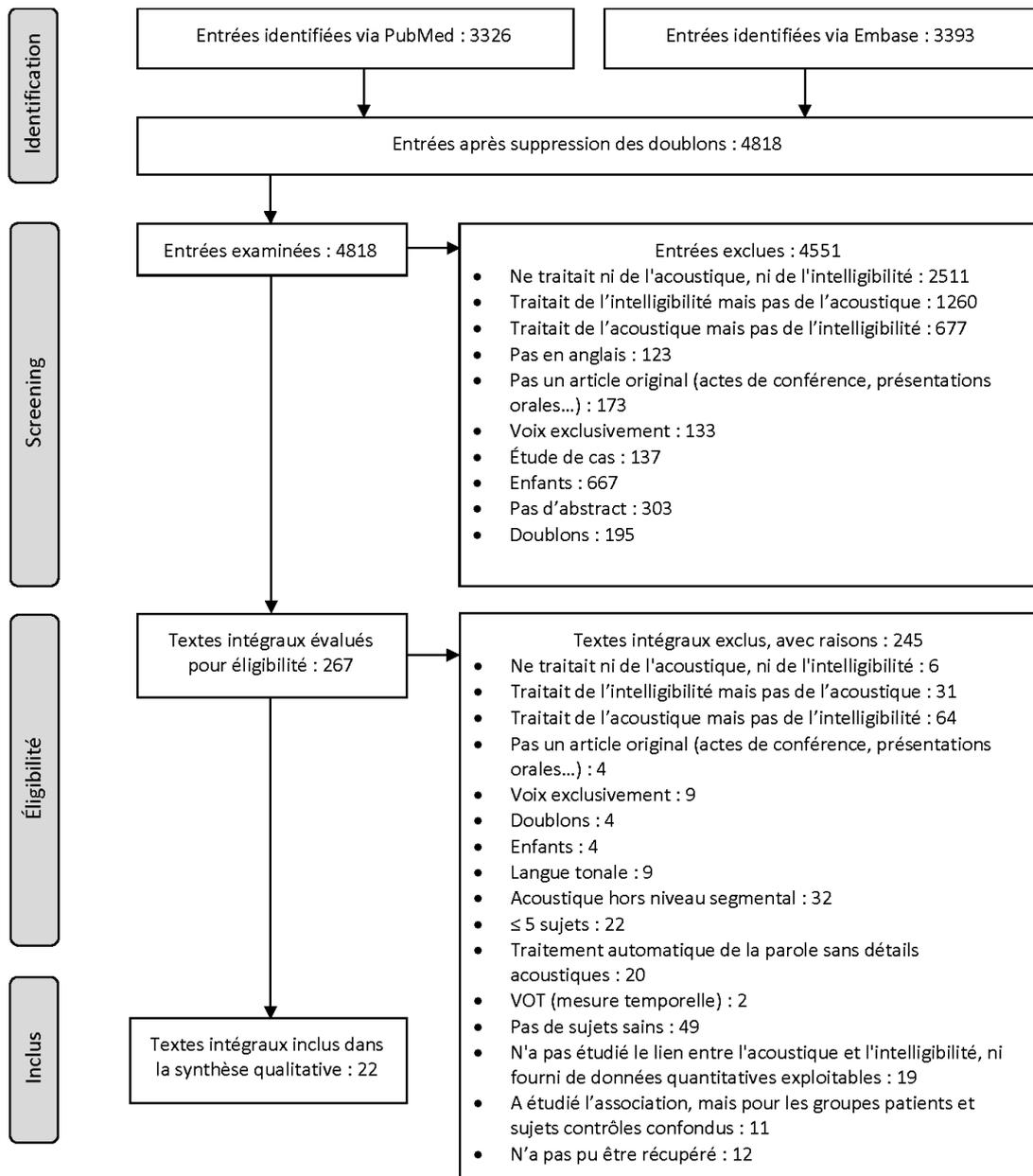


Figure 10 – Diagramme de flux illustrant le processus de sélection selon les lignes directrices PRISMA

2. Évaluation de la qualité méthodologique

Les scores QualSyst des 22 articles variaient entre 71 % (bonne qualité méthodologique) et 100 % (haute qualité). La qualité méthodologique d'un seul article a été évaluée comme étant « bonne », celle des 21 autres a été évaluée comme étant « haute ». Selon la hiérarchie du NHMRC pour l'évaluation du niveau de preuve, 14 articles ont été classés comme sources de niveau III-2 (étude comparative avec groupe contrôle parallèle), les huit autres articles ont été classés comme sources de niveau III-3 (étude comparative sans groupe contrôle parallèle). L'évaluation pour chaque article est indiquée dans le [Tableau 3](#).

Tableau 3 – Évaluation de la qualité méthodologique des 22 articles inclus à l'aide de l'outil d'évaluation critique Qualsyst de [KMET 04] et du niveau de preuve selon la hiérarchie du National Health and Medical Research Council [NHMRC 99]

Référence	Score Qualsyst ¹ (%)	Qualité méthodologique	Niveau de preuve NHMRC
A01. McRAE 02	20/24 (83)	Haute	III-2
A02. HAZAN 04	21/24 (88)	Haute	III-3

A03. NEEL 08	18/22 (82)	Haute	III-3
A04. FERGUSON 14	20/22 (91)	Haute	III-3
A05. WHITFIELD 17	21/24 (88)	Haute	III-3
A06. KATZ 91	20/24 (83)	Haute	III-3
A07. FLEGE 92	21/24 (88)	Haute	III-3
A08. BUNTON 01	21/24 (88)	Haute	III-2
A09. FERGUSON 07	20/24 (83)	Haute	III-3
A10. WEISMER 92	17/24 (71)	Bonne	III-2
A11. HOHOFF 03	20/24 (83)	Haute	III-3
A12. YUNUSOVA 05	20/24 (83)	Haute	III-2
A13. DE BRUIJN 09	21/24 (88)	Haute	III-2
A14. VAN LIERDE 12	20/24 (83)	Haute	III-2
A15. SKODDA 13	21/24 (88)	Haute	III-2
A16. WHITFIELD 14	23/24 (96)	Haute	III-2
A17. NEEL 15	22/22 (100)	Haute	III-2
A18. DWIVEDI 16	24/24 (100)	Haute	III-2
A19. CONNAGHAN 17	21/24 (88)	Haute	III-2
A20. FLETCHER 17	22/24 (92)	Haute	III-2
A21. KIM 17	22/24 (92)	Haute	III-2
A22. MARTEL-SAUVAGEAU 17	23/24 (96)	Haute	III-2

¹ Qualité méthodologique : haute >80 % ; bonne 60–79 % ; appropriée 50–59 % ; mauvaise <50 %.

Note : Les études ont été classées selon 1) le type de résultat : A01-A05 = corrélation directe entre les mesures acoustiques et perceptives ; A06-A09 = étude indirecte du lien entre les mesures acoustiques et perceptives ; A10-A22 : données quantitatives fournies pour les mesures acoustiques et perceptives, sans étude du lien ; 2) l'ordre chronologique puis alphabétique.

3. Caractéristiques des études

Populations d'étude

Parmi les 22 études, 14 incluaient à la fois un groupe de sujets et un groupe de témoins sains, dont seul le groupe témoin a été conservé pour la présente analyse. Les huit études restantes n'incluaient que des locuteurs sains comme groupe d'étude. En considérant que seules les études comprenant plus de cinq sujets ont été retenues, la taille médiane de l'échantillon de l'étude était de 15 (min. : 8, max. : 93), avec un écart interquartile de 18,5.

Les répartitions des facteurs genre et âge sont détaillées dans le [Tableau 4](#). En ce qui concerne les langues étudiées, 17 des 22 études (73 %) ont été menées sur anglais. Onze d'entre elles investigaient l'anglais américain (dont trois spécifiaient un dialecte de l'Upper Midwest), une l'anglais britannique, une l'anglais néo-zélandais, et les quatre autres ne spécifiaient pas la variante anglaise. Deux études ont été réalisées en néerlandais, deux en français (dont une en français québécois), une en allemand et une à la fois en coréen et en anglais.

Tableau 4 Répartition des genres et des âges dans les études retenues dans notre revue systématique

Facteur	Répartition	N
Genre	Hommes et femmes	20/22
	<i>Ratio 1:1</i>	13/20
	<i>Léger déséquilibre (différence <20 %)</i>	4/20
	<i>Prépondérance hommes (différence >20 %)</i>	3/20
	Hommes uniquement (A10)	1/20
	Genre non renseigné (A14)	1/20
Âge	>50 ans	11/22
	<50 ans	10/22
	Âge non renseigné (A03)	1/22

Échantillons de parole et mesures spectrales

Les différents phonèmes analysés dans les études ont majoritairement été extraits de mots isolés ou de mots dans des phrases. Deux études ont analysé des phonèmes isolés (voyelle tenue /i/ en A14, et /s/ en A18).

Voyelles. Dix-huit des 22 articles (82 %) ont étudié l'acoustique des voyelles. Les voyelles cardinales /i, u, a, æ/ sont les plus étudiées (8/18 études). Un article (A20) a étudié les voyelles cardinales de l'anglais néo-zélandais /a:, i:, o:/. Seules trois études ont analysé le panel plus large de voyelles /i, ɪ, e, ε, æ, ɑ, ʌ, o, ʊ, u/. Trois autres études ne mentionnaient pas explicitement les voyelles utilisées pour les analyses. Aucune des études ne s'est intéressée aux voyelles nasales.

En ce qui concerne l'analyse spectrale des voyelles, dix-sept des 18 études ont utilisé des mesures stationnaires de formants, quatre études ont examiné des mesures dynamiques.

Semi-voyelles. Un article (A22) a étudié les deux glides /w, j/ en plus des voyelles, en utilisant les pentes spectrales de F2 comme mesure du taux de modification du tractus phonatoire.

Consonnes. Huit articles (36 %) ont étudié les consonnes. Les consonnes les plus investiguées sont les fricatives /s, ʃ/ (6 études). Les deux autres articles traitaient des plosives /t, d/ (contraste de voisement, A07) et de la vélaire /x/ (en néerlandais, A13), respectivement. Aucune de ces études ne portait sur les consonnes nasales ni sur les liquides.

Parmi ces huit articles, cinq ont utilisé l'analyse des moments spectraux. Quatre d'entre eux ont utilisé le premier moment, tandis que le cinquième a employé le deuxième moment. Les autres mesures acoustiques ont été utilisées dans des études uniques et sont rapportées dans le tableau des résultats ([Annexe D](#)).

Une liste des définitions et des formules des mesures acoustiques utilisées dans les études retenues et rapportées dans le tableau des résultats ([Annexe D](#)) est disponible en [Annexe E](#).

Mesures perceptives

Pourcentage d'identification correcte. Dix études ont utilisé le pourcentage de stimuli correctement identifiés. Un article n'a pas décrit la tâche d'identification (A02). Les neuf autres ont employé une tâche à choix multiple, six dans lesquelles l'auditeur devait choisir la cible parmi une liste de mots, deux dans lesquelles l'auditeur devait choisir entre deux cibles (A06 et A07), et une (A19) dans laquelle l'auditeur devait choisir parmi 12 voyelles (monophthongues ou diphtongues). Aucune de ces études n'a utilisé une tâche de transcription.

Échelles ordinales. Sept études ont utilisé des échelles à intervalles réguliers de type Likert [LIKERT 32], dont cinq demandaient aux auditeurs d'évaluer l'« intelligibilité globale », trois l'« articulation », une la « clarté de parole », une la « précision de parole » et une la « sévérité de parole ». Deux études ont utilisé des échelles d'évaluation où un score élevé indiquait une évaluation positive ; quatre études ont utilisé des échelles où un score élevé signifiait une évaluation négative. Une étude a utilisé les deux types d'échelles (A13).

Échelles visuelles analogiques (EVA). Cinq études ont employé des échelles visuelles analogiques, dont deux demandaient aux évaluateurs de juger la « clarté de parole » (A05, A16)

et les autres respectivement l'« intelligibilité globale » et la « précision de parole » (A17), la « précision articulatoire » (A20), la « facilité de compréhension » (A20) et « ce que [l'auditeur] a compris de ce que la personne a dit » (A22). Trois des études ont utilisé des EVA positives, les deux autres ont utilisé des EVA négatives.

Estimation directe de la magnitude (EDM). Deux études ont utilisé l'estimation directe de la magnitude avec un modulus de 100. Dans l'une des études, il était demandé aux auditeurs d'évaluer la « sévérité globale » (échelle négative) (A01), dans l'autre, l'« intelligibilité globale » (instruction : « facilité de compréhension ») sur une échelle positive (A12).

Mesure du résultat

Neuf des 22 articles retenus ont analysé le lien entre l'acoustique spectrale et les mesures perceptives. Deux méthodologies différentes peuvent être identifiées. Cinq articles (A01-A05) ont directement abordé la corrélation entre les mesures acoustiques et perceptives. Quatre autres (A06-A09) ont indirectement étudié ce lien, en analysant les différences acoustiques entre des groupes créés sur base de leur intelligibilité perçue (A09), ou en analysant les différences acoustiques entre deux phonèmes/syllabes correctement perçus : /si/ vs /su/ (A06) ; /t/ vs /d/ (A07) ; /ε/ vs /æ/ et /i/ vs /ε/ (A08). Les 13 autres articles (A10-A22) ont analysé les mesures spectrales ainsi que les mesures perceptives, sans toutefois aborder directement le lien entre les deux.

4. Résumé des conclusions

Les conclusions des différentes études sont présentées dans le tableau des résultats ([Annexe D](#)), classées en trois catégories : les études traitant directement du lien entre les mesures spectrales et perceptives ; les études étudiant indirectement ce lien ; et les études fournissant uniquement des données descriptives sans analyser le lien.

En ce qui concerne la première catégorie, les corrélations significatives et non significatives sont présentées dans le [Tableau 5](#). Des corrélations significatives entre les mesures spectrales et les évaluations perceptives ont été mesurées pour les voyelles uniquement, pour les mesures stationnaires de F1 et F2 (A04), l'intervalle de F1 chez les hommes (A03), la différence de F2 entre /i/ et /u/ (A02), l'aire de l'espace vocalique (VSA, A03), le changement relatif de l'aire de l'espace vocalique acoustico-articulatoire (A05), la valeur moyenne de mouvement des formants chez les femmes (A03) et la mesure dynamique de longueur de vecteur (A04).

Parmi les études qui ont abordé indirectement le lien entre l'acoustique spectrale et les évaluations perceptives (c.-à-d. sans corrélation), A06 et A07 ont ciblé les mesures consonantiques, tandis que A08 et A09 se sont concentrées sur les voyelles. Dans A06, l'énergie centroïde et le pic spectral de la fricative dans le son /s/ dans /si/ et /su/ se sont avérés être des marqueurs acoustiques de l'effet de coarticulation, les valeurs étant significativement plus élevées pour le /s/ dans les syllabes identifiées comme /si/. Dans A07, les auteurs ont observé des fréquences d'attaque stationnaires de F1 significativement plus élevées pour les voyelles précédant /t/ que pour /d/, chez des locuteurs anglais natifs. Les auteurs ont conclu que cette mesure acoustique est un bon indicateur de la perception du contraste de voisement dans les consonnes plosives apico-alvéolaires.

Tableau 5 – Corrélations significatives et non significatives entre mesures acoustiques et évaluations perceptives de la parole

	Voyelles											Consonnes			
	F1	F2	Intervalle		Distance euclidienne F1-F2	F1≠ /i-a/	F2≠ /i-u/	Distance vocalique	VSA	AAVS	Mouvement formants	Rapport dynamique	Longueur vecteur	Longueur trajectoire	1 ^{er} moment spectral
		F1	F2												
EDM									∅ (A01)						∅ (A01)
Likert	∅ (A02)	∅ (A02)			∅ (A02)	∅ (A02)	✓ (A02)								
EVA									✓ (A05)						
%corr	∅ (A02, A03)	∅ (A03)	✓ (M, A03)	∅ (A03)	∅ (A02)	∅ (A02)	✓ (A02)	∅ (A03)	✓ (A03)		✓ (F, A03)	∅ (A03)	✓ (A04)	∅ (A04)	
	✓ (A04)	✓ (A04)	∅ (F, A03)								∅ (M, A03)				

Note : ✓ = corrélation significative ; ∅ = corrélation non significative ; F1/F2 = premier et second formant ; F1≠ /x,y/ = différence de F1 entre les voyelles x et y ; VSA = « vowel space area » (aire de l'espace vocalique) ; AAVS = « articulatory-acoustic vowel space » (espace vocalique acoustico-articulatoire) ; EDM = estimation directe de la magnitude ; Likert : échelle à intervalle régulier de type Likert ; EVA = échelle visuelle analogique ; %corr = score de pourcentage correct d'identification

En ce qui concerne les mesures visant les voyelles, des différences significatives entre F0-F1 et F1-F2 ont été observées dans A08, pour les voyelles correctement identifiées dans les paires /ε-æ/ et /I-ε/. Les auteurs ont ainsi conclu que ces mesures sont liées à l'intelligibilité de la parole, puisqu'elles semblent liées à la perception du contraste de hauteur de langue. La différence F1-F2 a été considérée comme l'indice primaire, tandis que la différence F0-F1 a été interprétée comme un indice secondaire, lié à la différence F2-F1. En A09, la mesure du « changement spectral » s'est avérée significativement plus élevée chez les locuteurs montrant un bénéfice élevé en parole claire (« clear speech ») sur l'identification des mots.

Discussion – Les besoins identifiés dans la recherche scientifique

Les données de cette revue confirment la nature variable et « imprécise » de la parole chez les locuteurs adultes sains. En raison des différentes tâches d'évaluation et instructions (p. ex. évaluation sur des EVA de l'intelligibilité vs de la précision articulatoire) et des unités de parole ciblées (p. ex. pourcentage d'identification correcte de phonèmes vs de mots), aucune mesure groupée de variabilité n'a pu être calculée sur l'ensemble des études de cette revue. Cependant, nous pouvons observer que parmi les études utilisant le pourcentage d'identification correcte, par exemple, alors que quatre ont trouvé des valeurs supérieures à 90 % (sur des mots, des voyelles isolées et des voyelles dans des syllabes consonne-voyelle-consonne [CVC]), quatre autres ont trouvé des scores moyens entre 60,6 % et 71 % (sur des phonèmes dans des syllabes CVC et sur des syllabes). La variabilité de la parole chez les locuteurs sains est également observée entre les sujets à l'intérieur des différentes études. Par exemple, alors que trois des études utilisant des pourcentages d'identification correcte rapportent un écart-type relativement faible (variant de 1,12 % à 4 %), les études utilisant des échelles ordinales montrent une plus grande variabilité : si tous les résultats sont normalisés en pourcentages, les écarts-types varient de 6,25 % à 12 %. Bien qu'une part de cette plus grande variabilité puisse être liée au type d'échelle employée, ces résultats tendent tout de même à illustrer que même chez les locuteurs sains, les limites physiologiques ne permettent pas toujours au système de production de la parole de répondre aux nombreuses exigences de la parole spontanée. Les

« imprécisions » qui en résultent se situent principalement au niveau des phonèmes [ROSSI 98; SCHILLER 06], entraînant un certain recouvrement des catégories phonémiques, c.-à-d. des réductions vocaliques et consonantiques, ainsi que des omissions de phonèmes [BENZEGUIBA 07; GUENTHER 95; MEUNIER 07; VAN SON 96, 99].

L'objectif de cette revue était d'étudier de plus près la manière dont ces variations en parole saine peuvent être mesurées afin d'être prises en compte lors de l'analyse de la parole pathologique. Les dates de publication des articles retenus — dont trois seulement datent des années 1990 — montrent que l'essor de la technologie a suscité un intérêt croissant pour l'étude acoustique de la parole. Ceci est principalement dû au fait que les mesures acoustiques ne doivent plus être entièrement effectuées manuellement et sont donc plus rapides à obtenir, mais aussi plus fiables. Nous allons donc maintenant nous pencher sur les tendances et les besoins identifiés dans l'analyse des études acoustiques de l'intelligibilité en parole saine.

Tout d'abord, dans cette revue, nous remarquons que la plupart des études utilisant des mesures spectrales se sont concentrées sur les voyelles. La réduction vocalique en parole spontanée informelle est un phénomène universel et bien décrit [VAN SON 96]. Il existe deux types de réduction [MAUROVÁ PAILLEREAU 16] : la centralisation vocalique et l'assimilation contextuelle. La centralisation vocalique est observée lorsque les fréquences des formants tendent vers celles d'une voyelle neutre, tandis que l'assimilation contextuelle se produit lorsque les fréquences des formants d'une voyelle évoluent vers le locus acoustique des consonnes voisines. Les voyelles jouent un rôle important dans l'intelligibilité de la parole [CHEN 13; COLE 96; KEWLEY-PORT 07] et sont également plus pratiques à analyser spectralement, car elles sont par définition voisées et composées d'ondes périodiques et peuvent être tenues (contrairement aux consonnes plosives). Les études ciblant les caractéristiques spectrales des consonnes sont plus rares dans notre revue, bien que les consonnes soient réduites autant que les voyelles dans la parole spontanée et que cette réduction articulaire affecte également l'intelligibilité [VAN SON 99]. Alors que les voyelles se sont avérées avoir un effet plus important sur la discrimination de l'identité du locuteur, les consonnes semblent effectivement essentielles pour l'identification des mots [BONATTI 05; OWREN 06]. De plus, [LINDBLOM 90] postulait déjà que, malgré les effets de coarticulation, une combinaison de caractéristiques spectrales pouvait permettre une bonne distinction entre les consonnes plosives. L'intelligibilité des consonnes, leur variabilité et leur réduction en parole saine, ainsi que les indices spectraux associés (en plus des indices temporels, déjà mieux documentés), devraient ainsi être davantage explorés. Le **besoin d'investiguer davantage les mesures acoustiques spectrales sur les consonnes** est donc un premier constat souligné dans notre revue. Nous nous pencherons plus en détail sur l'importance des consonnes et des voyelles dans l'intelligibilité dans le [Chapitre 3](#).

Dans notre revue, diverses mesures acoustiques spectrales, principalement sur les voyelles, se sont avérées liées à l'intelligibilité perçue de la parole chez les locuteurs sains (cf. [Chapitre 4](#) pour une discussion plus en détail de ces mesures). Cependant, les résultats montrent qu'aucune de ces mesures n'explique un pourcentage élevé de la variance des scores d'intelligibilité perceptive. Bien que les mesures acoustiques permettent une étude plus objective de la parole, elles ne représentent pas de manière exhaustive le signal de la parole,

mais ciblent plutôt des indices spécifiques qui sont considérés comme théoriquement pertinents dans chaque étude. Il convient également de garder à l'esprit que la bonne perception de phonèmes repose sur plusieurs caractéristiques phonémiques [JAKOBSON 51] et que ce n'est pas un seul trait, mais l'ensemble des unités discrètes de parole qui constitue la notion d'intelligibilité [FLANAGAN 72]. Par conséquent, une combinaison de mesures acoustiques, prenant en compte divers traits phonémiques et aspects spectraux, pourrait être une première voie vers une évaluation plus exhaustive de l'intelligibilité de la parole (p. ex. [BRADLOW 96; KENT 89b; J. LEE 14; LINDBLOM 90; WEISMER 08]). Avant d'utiliser des mesures acoustiques segmentales sur des populations spécifiques de patients, cette revue montre donc que des recherches approfondies sont encore nécessaires afin de **mieux comprendre leur comportement chez les locuteurs sains**. Ceci aidera à **identifier les combinaisons acoustiques exhaustives et pertinentes pour rendre compte de l'intelligibilité** et à fournir des données normatives à partir d'un large ensemble de locuteurs sains.

En outre, il existe une intrication complexe des caractéristiques acoustiques segmentales avec des facteurs à d'autres niveaux de granularité tels que l'intonation, l'accentuation (p. ex. différences acoustiques entre voyelles accentuées et non accentuées en A19), la qualité vocale et le débit de parole [KURUVILLA-DUGDALE 20; METZ 90; SMILJANIĆ 09; WHITFIELD 17]. Un troisième besoin identifié dans notre revue consiste donc en la **considération des aspects suprasegmentaux lors d'analyses de phénomènes segmentaux**.

D'autres besoins ressortent de l'analyse des méthodologies des articles retenus. La diversité de ces méthodologies démontre tout d'abord que la parole peut être étudiée de différentes manières, tant au niveau perceptif qu'acoustique. Sur les 22 articles retenus dans notre revue, seuls cinq ont abordé la définition du ou des concepts liés à la parole, dont quatre (A08, A12, A20, A22) ont fourni une définition de l'intelligibilité. Compte tenu des différents termes utilisés pour désigner la production de la parole — chacun d'entre eux faisant référence à un concept spécifique — il existe un besoin primordial de fournir des **définitions non ambiguës des principaux concepts abordés** dans les articles de recherche scientifique. De même, les **tâches d'évaluation et les mesures acoustiques doivent être décrites en détail**, afin de permettre une interprétation des résultats en conséquence ainsi qu'une reproductibilité de la méthodologie. Nous constatons en effet que même si plusieurs études utilisent la même mesure, la population étudiée, l'échantillon phonémique, la méthode d'extraction et de calcul et la présentation des résultats sont très différents (ce qui, selon l'objectif de chaque étude, est tout à fait légitime), mais surtout parfois non rapportés, ce qui rend difficile la mise en relation des valeurs obtenues. Pour illustrer ce point, une tentative de comparaison des résultats de mesures acoustiques similaires utilisées dans les différentes études est présentée en [Annexe F](#). Une description détaillée de la méthodologie semble donc être une nécessité, surtout dans le domaine de l'étude acoustique de la parole où de nombreux facteurs peuvent contribuer à des variations importantes.

Des considérations concernant la méthodologie peuvent également être soulignées en ce qui concerne les caractéristiques des populations étudiées. La majorité des études ont inclus des hommes et des femmes dans une proportion équilibrée. Cependant, très peu d'entre elles ont réellement différencié les résultats par genre, en particulier dans les groupes témoins, pour

lesquels les résultats sont très souvent agrégés. Il est pourtant établi que les valeurs des formants vocaliques, par exemple, varient entre les hommes et les femmes [BRADLOW 96; COLEMAN 71; HILLENBRAND 95; YANG 96], ainsi que les composantes de bruit dans les consonnes [PÉPIOT 15]. D'une manière générale, le **facteur genre doit être davantage pris en compte**, et l'information sur le genre du groupe d'étude devrait être systématiquement précisée a minima. Un moyen de contourner le problème des comparaisons des valeurs entre genres est d'utiliser des échelles Bark [A. R. FLETCHER 17], ce qui a pu être observé dans certaines des études de cette revue. Toujours concernant les populations étudiées, alors que la moitié des études ont été réalisées sur des groupes d'étude âgés de plus de 50 ans, aucune étude n'a spécifiquement examiné l'impact de l'âge sur les mesures spectrales ou sur l'intelligibilité perçue de la parole. Il serait intéressant de **prendre en compte le facteur âge** afin d'analyser l'évolution de l'acoustique de la parole et de l'intelligibilité perçue dans le vieillissement normal [KURUVILLA-DUGDALE 20]. En effet, il a été démontré que la parole varie tout au long de la vie en raison de modifications physiologiques et neuromusculaires [BENJAMIN 97; BILODEAU-MERCURE 16; HAZAN 17, 18; HOOPER 09; TREMBLAY 17]. Ces modifications articulatoires induites par l'âge diminuent l'intelligibilité de la parole [HAZAN 17; KURUVILLA-DUGDALE 20; PARNELL 87; SHUEY 89], principalement par une réduction de la précision de la production phonémique et un ralentissement du débit de parole. **L'étude des modifications de la parole dans le vieillissement « normal »** par rapport au vieillissement pathologique pourrait permettre de mieux comprendre les stratégies de production dans la parole saine.

Conclusion

Nous avons mené une revue systématique de la littérature selon la méthodologie PRISMA afin d'identifier les besoins en termes de mesure de la parole saine. Nos résultats mettent en évidence que la parole saine présente des imprécisions inhérentes et n'est donc pas, comme on le présume souvent, intelligible à 100 %. Une meilleure compréhension de ces imprécisions fournira une ligne de base plus réaliste pour l'étude de la parole pathologique. À cet effet, des recherches approfondies doivent encore être menées pour mieux comprendre les comportements des mesures acoustiques chez le locuteur sain et identifier les combinaisons acoustiques pertinentes qui pourraient expliquer les variations perçues de la parole (p. ex. les réductions vocaliques et consonantiques). Il serait notamment intéressant d'investiguer davantage les mesures acoustiques spectrales sur les consonnes, tout en considérant les aspects suprasegmentaux. Notre étude souligne également la nécessité de fournir dans les études des définitions non ambiguës des principaux concepts abordés. En plus de cela, décrire de manière suffisamment détaillée les tâches d'évaluation et les mesures acoustiques employées est primordial pour permettre la réplication, les comparaisons croisées ou méta-analyses, et la mise en commun des données. Les facteurs genre et âge doivent en outre être davantage pris en compte, en précisant a minima les informations sur leur répartition dans les groupes d'étude.

Perspectives

Les mesures du domaine temporel n'ont pas été prises en compte dans cette revue afin de limiter le bruit lors de la recherche initiale dans les bases de données (p. ex. les nombreuses études sur le débit de parole et les pauses dans les troubles de la fluence). Seules les mesures du domaine fréquentiel ont donc été incluses. À l'avenir, il serait toutefois intéressant d'examiner le lien entre les mesures du domaine temporel (comme le délai de voisement ou « voice onset time ») et l'intelligibilité perçue, car les mesures du domaine temporel et du domaine fréquentiel fournissent des données complémentaires [FLOEGEL 20; X. LI 08]. Le nombre plus élevé d'études portant sur les voyelles par rapport aux consonnes pourrait également découler de cette décision méthodologique. D'autres études sur les mesures temporelles pourraient clarifier s'il s'agit d'une tendance générale parmi les mesures au niveau du phonème, ou si elle est limitée aux mesures spectrales.

Enfin et surtout, bien que notre analyse se soit concentrée sur les études écrites en anglais, il serait également utile d'examiner les études écrites dans — et donc axées sur — d'autres langues. L'exemple le plus contrasté pour illustrer l'intérêt d'étudier d'autres langues est celui des langues tonales. Dans ces dernières, les constituants acoustiques et perceptifs de l'intelligibilité de la parole peuvent être très différents de ceux des langues occidentales. Les mesures suprasegmentales (p. ex. le contour intonatif), notamment, peuvent y contribuer à un degré plus élevé à l'intelligibilité, par rapport aux mesures au niveau du phonème [CHEN 11].

2.3. Deuxième pilier de l'EBP : l'expérience clinique et pratique

La difficulté de transposition des données de la recherche vers la pratique clinique est l'une des barrières à l'utilisation de l'EBP et provient entre autres de l'individualité de chaque patient et de l'hétérogénéité des symptômes en contexte pathologique [BUTLER 19]. La prise en compte de l'expérience clinique est donc un élément primordial afin d'adapter au mieux les outils d'évaluation et de prise en charge de patients atteints de troubles de la parole. C'est pourquoi nous avons dans ce travail souhaité dresser un tableau des pratiques cliniques actuelles en matière d'évaluation clinique de la parole et obtenir l'avis de cliniciens quant aux besoins en termes de mesure de celle-ci (voir [Figure 11](#)). Nous avons pour cela mené une **enquête en ligne adressée aux cliniciens francophones** en Belgique, en France, au Luxembourg, en Suisse et au Maghreb. En nous limitant au contexte clinique francophone, nous avons également souhaité répondre dans une certaine mesure à la prise en compte du contexte « local » parfois inclus dans le schéma de l'EBP [FOSTER 13; RYCROFT-MALONE 04], en gardant bien sûr à l'esprit l'hétérogénéité des pratiques même au sein d'une communauté linguistique donnée.

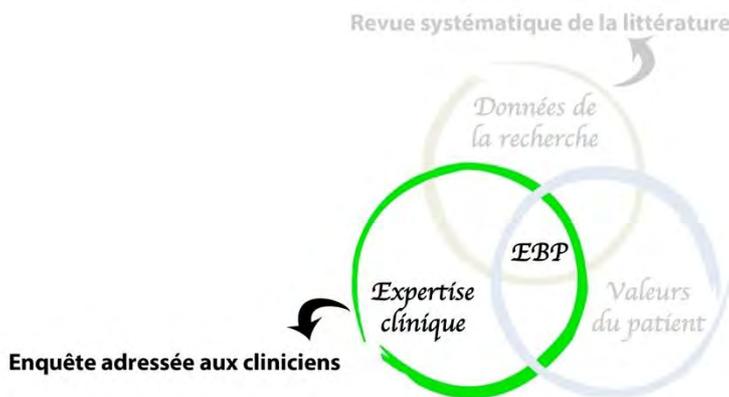


Figure 11 – Prise en compte du deuxième pilier de l'EBP : enquête adressée aux cliniciens

En plus de l'expérience clinique, comme nous traitons d'une thématique qui touche de nombreux domaines, nous avons également souhaité élargir notre recherche des besoins à d'autres domaines applicatifs. C'est pourquoi nous avons mené des **entrevues semi-dirigées avec des chercheurs de différents horizons** traitant de la parole dans des thématiques applicatives, afin d'obtenir leurs opinions sur les pratiques actuelles et les manques et besoins ressentis concernant les mesures de la parole.

2.3.1. Évaluation des troubles de la parole chez l'adulte – situation actuelle et besoins en pratique clinique francophone

Le bilan de parole vise à évaluer plusieurs dimensions du trouble, qui peuvent être regroupées — sur base du modèle de la Classification Internationale du Fonctionnement, du Handicap et de la Santé (CIF [ORGANISATION MONDIALE DE LA SANTÉ 01]) — en trois niveaux (voir [Figure 12](#)).

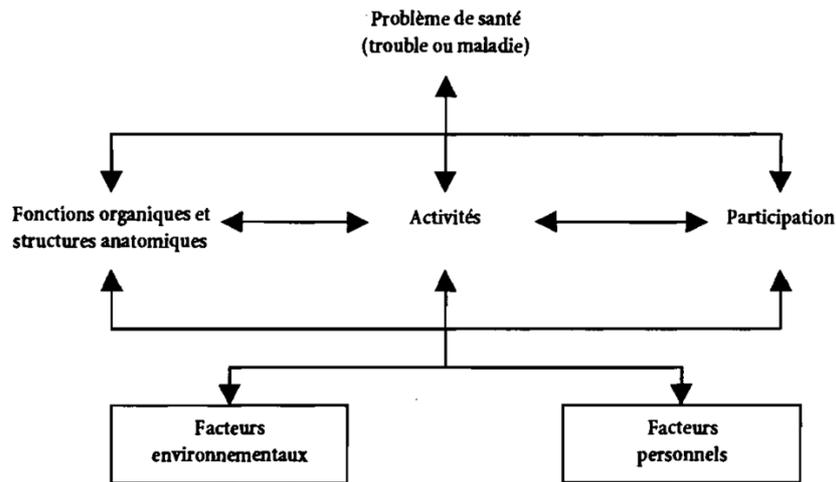


Figure 12 – Composantes de la CIF (Organisation mondiale de la santé 2001)

Le **premier niveau** d'analyse est celui **des structures et des fonctions corporelles**, où le clinicien intègre des informations sur le contexte pathologique, sur les déficits/modifications structurelles et leur caractère stable ou évolutif. Ce niveau comprend l'examen des trois principaux sous-systèmes de la production de la parole : la respiration, la phonation et l'articulation. Le clinicien évalue la posture du patient, sa respiration au repos et pendant la parole, la coordination pneumophonique, la fonction phonatoire, la résonance, la motricité et la sensibilité des articulateurs, ainsi que la coordination, la programmation et la planification des mouvements articulaires [BALLARD 15; DYKSTRA 07].

Le **deuxième niveau**, celui des **activités**, concerne les limitations que le déficit induit au niveau de la communication du patient, sur sa capacité à partager des informations avec autrui. À ce niveau, deux dimensions sont distinguées : l'intelligibilité et la capacité de communication fonctionnelle. Ce niveau comprend donc le concept plus analytique de l'intelligibilité et le niveau plus fonctionnel de la compréhensibilité, deux concepts complémentaires que nous avons définis au [Chapitre 1](#). Il est à noter que si les éléments sous-jacents de l'intelligibilité de la parole (respiration, phonation, articulation) peuvent être analysés au niveau des « structures et fonctions corporelles », le concept d'intelligibilité en lui-même est considéré comme relevant de l'utilisation intégrée de ces structures et fonctions pour produire une parole intelligible, et est donc inclus dans le niveau « activités » de la CIF [DYKSTRA 07; YORKSTON 96].

Enfin, au **troisième niveau** de la CIF, le clinicien doit évaluer le niveau de **participation**, c.-à-d. la capacité du patient à interagir avec son environnement et les restrictions que le trouble de la parole induit dans son contexte psychosocial.

Pour chacun de ces niveaux, les méthodes d'évaluation de la parole utilisées aujourd'hui sont variées [RUMBACH 19]. Elles sont principalement perceptives [ALTAHER 19] et subjectives, par exemple : les échelles visuelles analogiques, les échelles de Likert et l'estimation directe de la magnitude [CARMICHAEL 07; KENT 89b; SCHIAVETTI 92; WHITEHILL 02], la transcription orthographique de stimuli [GHIO 16; HUSTAD 06, 08], l'évaluation de la compréhensibilité de mots/de phrases sur des listes fermées d'items [HUSTAD 06; KENT 89b] et l'évaluation de la compréhensibilité de récits via des questions de contenu [HUSTAD 08]. Toutes ces méthodes permettent d'évaluer différents concepts globaux (p. ex. la sévérité du trouble, le caractère naturel, la clarté de la parole) ou plus spécifiques (p. ex. la prosodie, la qualité vocale, le débit

de parole). Elles peuvent s'appliquer à différents niveaux de granularité : sur des phonèmes isolés, des syllabes, des (pseudo-)mots, des phrases, de la lecture à voix haute d'un texte, ou une conversation.

Enfin, des questionnaires d'auto-évaluation sont parfois utilisés pour évaluer la perception du trouble de la parole par le patient, ainsi que l'impact fonctionnel et la qualité de vie.

Les outils d'évaluation disponibles en clinique francophone seront présentés plus en détail dans le [Chapitre 3](#) de ce travail.

Compte tenu de la diversité des méthodes d'évaluation de la parole, mais aussi de la variabilité internationale de la formation des orthophonistes, les pratiques d'évaluation doivent être explorées dans les différents pays et communautés linguistiques. Une meilleure compréhension de ce sujet permet de développer des lignes directrices cliniques, de nouveaux outils cliniques et l'information des programmes de formation de futurs orthophonistes [RUMBACH 19]. Alors que des enquêtes ont été menées dans différents pays et communautés linguistiques [COLLIS 12; CONWAY 15; GUREVICH 17; MILLER 17; RUMBACH 19], la littérature actuelle ne fournit pas de données sur les pratiques d'évaluation de la parole dans les pays francophones. Par conséquent, au moyen d'une enquête destinée aux orthophonistes et phoniâtres francophones, le premier objectif de cette étude était de fournir un aperçu des pratiques cliniques actuelles concernant l'évaluation des troubles de la parole chez les patients adultes. Cette vue d'ensemble a ensuite permis d'aborder le second objectif de cette étude, à savoir l'identification des manques et des besoins rapportés par ces cliniciens concernant l'évaluation de la parole chez l'adulte.

A. Méthodologie

Recrutement

Notre questionnaire en ligne s'adressait aux cliniciens francophones (orthophonistes et phoniâtres) traitant des troubles de la parole chez l'adulte¹¹, en Belgique, en France, au Luxembourg, en Suisse et au Maghreb.

Le recrutement a été effectué via :

- les associations professionnelles nationales (Fédération Nationale des Orthophonistes, Union Professionnelle des Logopèdes Francophones, Association Romande des Logopédistes Diplômés...);
- des groupes privés dédiés aux professions ciblées sur les réseaux sociaux ;
- d'autres moyens, comme le bouche-à-oreille et les listes de diffusion d'autres organisations.

Le nombre global d'orthophonistes actifs était approximativement connu (25 607 en France au 1er janvier 2019 [DIRECTION DE LA RECHERCHE 19], 6 392 dans les régions francophones de Belgique au 31 décembre 2018 [DIRECTION GÉNÉRALE SOINS DE SANTÉ 18], 105 au Luxembourg au 31 décembre 2017 [MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA

¹¹ Définition présentée en début de questionnaire : « Le terme 'troubles de production de la parole' (TPP) est un terme générique qui englobe un grand nombre de pathologies d'origines variées, impactant la production de sons articulés. Ces difficultés peuvent être regroupées en troubles de la parole d'origine neurologique (dysarthries et apraxies) et d'origine organique/anatomique (malformations de l'appareil bucco-pharyngé, séquelles de traitement carcinologique ORL radiothérapique et/ou chirurgical, insuffisance vélaire, surdité...). Nous ne traiterons pas dans cette étude d'autres troubles de l'élocution tels que les troubles de la fluence ou les troubles articulatoires développementaux. »

RECHERCHE 20], 76 dans les hôpitaux de Suisse latine en 2014 [OBSERVATOIRE SUISSE DE LA SANTÉ 16]). Cependant, le nombre hypothétique de cliniciens répondant aux critères d'inclusion ainsi que la taille de l'échantillon des orthophonistes finalement contactés par les associations professionnelles n'étaient pas connus des auteurs. Aucune hypothèse n'a donc pu être émise sur le nombre attendu de réponses. À titre de comparaison, les tailles d'échantillon d'autres enquêtes ciblant les orthophonistes dans le domaine de la dysarthrie étaient les suivantes : 296 aux États-Unis [GUREVICH 17], 80 en République d'Irlande [CONWAY 15], 119 et 146 au Royaume-Uni [COLLIS 12; MILLER 11] et 56 en Australie [RUMBACH 19]. La seule enquête concernant les orthophonistes travaillant dans le domaine de la dysarthrie en France ciblait le nord-ouest du pays ainsi que Paris, avec un échantillon limité à 20 participants [JAN 07].

Le questionnaire

1. Conception et construction du questionnaire

L'enquête en ligne a été réalisée à l'aide de la plateforme LimeSurvey. Le questionnaire était en accès libre, avec des procédures intégrées de vérification des doublons et de sécurité. Deux questions « filtres » ont été utilisées pour vérifier la conformité aux critères d'inclusion, mettant fin à la participation en cas de réponse négative. La première faisait référence à la profession (orthophoniste ou phoniatre), la seconde à la prise en charge de patients adultes souffrant de troubles de la parole tels que définis (Oui/Non).

Le questionnaire comprenait 49 questions : 18 questions ouvertes (entrées numériques, textes courts et textes longs) et 31 questions fermées (choix unique et multiple, réponses binaires, classements, échelles de Likert), regroupées en six catégories principales :

1. Informations sur le participant ;
2. Données sur le parcours académique et professionnel du participant et sur sa pratique actuelle, afin de pouvoir analyser les pratiques d'évaluation en fonction des différents profils de cliniciens ;
3. Informations concernant l'expérience professionnelle dans le domaine des troubles de la parole, afin d'estimer le niveau d'expertise du répondant ;
4. Des questions sur la patientèle, pour décrire le groupe cible des évaluations de la parole ;
5. Données sur l'équipement matériel/logiciel et l'évaluation de la parole effectuée pour les troubles de la parole, afin d'examiner les pratiques actuelles ;
6. Informations sur les éventuelles lacunes de l'évaluation de la parole.

Certaines questions étaient facultatives, afin d'éviter l'abandon du questionnaire en cas de difficultés de réponse. Une case « commentaire » était prévue pour permettre aux participants de compléter leurs réponses si nécessaire.

Le questionnaire a été testé au préalable afin d'obtenir une estimation du temps de réponse et de détecter d'éventuels problèmes d'exécution. Un panel de huit cliniciens a été sélectionné sur la base de leur expérience dans les troubles de la parole chez l'adulte. Le temps de réponse a été estimé à environ 35 minutes et certains bogues informatiques et problèmes de structure logique ont été corrigés.

Le questionnaire final a été lancé le 25 janvier 2019 et était disponible jusqu'au 31 mai 2019. Il est toujours disponible en ligne pour le lecteur¹².

¹² <https://enquetes.univ-tlse3.fr/index.php/622914?lang=fr>

2. *Analyse statistique des données*

Les données ont été analysées à l'aide du logiciel Stata/MP (version 14, StataCorp, College Station, Texas).

Le test de Shapiro-Wilk, recommandé pour les petits échantillons, a d'abord été utilisé pour tester la normalité des distributions des données dans les variables quantitatives, et a conduit au choix d'analyses non paramétriques.

Des statistiques descriptives ont été réalisées sur toutes les données quantitatives et qualitatives et ont servi de base pour décrire les pratiques des cliniciens francophones, ainsi que les besoins et les lacunes signalés.

Le test exact de Fisher, adapté aux petits échantillons et aux données qualitatives catégorielles, a été utilisé pour étudier les différences éventuelles parmi les cliniciens quant aux auto-évaluations du niveau d'expertise en matière d'évaluation de la parole et de la suffisance des outils d'évaluation. Deux questions de recherche ont ainsi été explorées :

- « Les cliniciens ayant une proportion plus élevée de troubles de la parole dans leur patientèle rapportent-ils un plus haut niveau d'expertise dans l'évaluation des troubles de la parole ? » ;
- « Les cliniciens qui effectuent un nombre élevé d'évaluations de troubles de la parole ont-ils une perception différente de la suffisance des outils d'évaluation de la parole ? ».

Des régressions logistiques ont été utilisées afin d'évaluer si le niveau auto-estimé en évaluation des troubles de la parole est lié au nombre d'années de prise en charge de ceux-ci et au nombre de formations continues sur la parole, en ajustant respectivement pour le nombre total de formations sur les troubles de la parole et pour le nombre d'années de prise en charge de ceux-ci. Les odd-ratios (OR) calculés permettent d'étudier le lien entre la variable quantitative d'intérêt (ici le niveau autoévalué) et une ou plusieurs autres variables explicatives.

Le test U de Mann-Whitney a été employé pour étudier la différence de satisfaction relative aux outils d'évaluation quantitatifs et qualitatifs.

Enfin, le test de Kruskal-Wallis a été utilisé pour analyser l'influence du nombre d'années d'expérience dans la prise en charge des troubles de la parole ainsi que du nombre de formations continues liées à ces troubles sur la perception de la suffisance des outils d'évaluation.

Afin d'illustrer les manques rapportés et les solutions envisagées par les cliniciens en ce qui concerne les outils d'évaluation de la parole, leurs réponses ouvertes ont d'abord été analysées par deux évaluateurs indépendants et aveugles à l'identité des participants (les mêmes que précédemment). Cette analyse de contenu visait à identifier les tendances et thématiques récurrentes, afin de définir des regroupements sémantiques. Après une réunion de consensus entre les deux évaluateurs, les fréquences auxquelles ces catégories ont été mentionnées par les cliniciens ont ensuite été utilisées pour construire un graphique sur diagrammes.net (voir [Figure 17](#)), comme suit : pour les manques (cercles), le diamètre de départ (correspondant à une seule mention) était de 80 pt, augmenté de 15 pt pour chaque mention supplémentaire. Les solutions (nuages) commençaient à une largeur de 120 pt, avec un pas d'incrémentation de 22,5 pt pour chaque nouvelle mention.

B. Résultats

Avant d’aborder la question des besoins des cliniciens francophones en matière d’évaluation des troubles de la parole, nous allons brièvement exposer les résultats concernant les participants (données démographiques, formation initiale et continue, et caractéristiques de la patientèle) et décrire les pratiques d’évaluation des troubles de la parole.

Deux cent soixante-dix-sept réponses ont été soumises. Après un nettoyage de la base de données (suppression des doublons, des entrées en cas d’abandon avant les premières questions concernant l’activité professionnelle, et de celles ayant une réponse négative à l’une des deux questions filtres), 119 entrées ont été retenues.

Le temps médian pour remplir le questionnaire était de 20 min 47 s (écart interquartile [EIQ] : 14 min 23 s).

Description de la population

1. Participants orthophonistes

Les données détaillées des 119 cliniciens dont les réponses ont été incluses dans cette étude sont disponibles dans le [Tableau 6](#).

L’âge médian des participants, tous orthophonistes, était de 39 ans (EIQ : 20). Quarante-cinq pour cent d’entre eux exerçaient en France¹³, avec une répartition assez équilibrée (voir [Figure 13](#)). La plupart d’entre eux exerçaient en libéral (74 %), avec une médiane de 13,5 ans d’exercice (EIQ : 18). Soixante-douze pour cent d’entre eux ont déclaré n’avoir suivi aucune formation continue sur les troubles de la parole.



Figure 13 – Distribution des participants dans le monde (gauche) et en France métropolitaine, en Belgique et en Suisse (droite). Les points en dehors de ces pays se trouvent en Algérie, en Tunisie et dans les DOM-TOM.

Les orthophonistes considèrent généralement leur niveau d’expertise en évaluation des troubles de la parole comme « intermédiaire » (60 %).

Ce niveau d’expertise autoévalué est lié à la proportion de troubles de la parole dans la patientèle (tests exacts de Fisher : $p=0,003$). Il est, en outre, davantage influencé par le nombre d’années d’expérience que par le nombre de formations liées aux troubles de la parole. La probabilité de se considérer comme expert en évaluation de la parole n’est pas plus grande en fonction du nombre total de formations, compte tenu des années d’expérience dans la prise en charge de ces troubles ; ceci est également vrai pour la probabilité de se considérer comme

¹³Les caractéristiques de notre échantillon suggèrent une bonne représentativité de la population des orthophonistes français : l’âge moyen (40,4 ans dans l’échantillon vs 43,4 ans en France), le nombre d’orthophonistes exerçant en cabinet libéral ou mixte (74 % vs 81 % en France) et le nombre de cliniciens salariés (26 % vs 19 % en France).

novice. En revanche, la probabilité de se considérer comme novice est significativement plus faible pour les orthophonistes ayant plus de 10 ans d'expérience de prise en charge des troubles de la parole, par rapport à ceux ayant moins de 5 ans d'expérience (régression logistique : OR=0,06, c.-à-d. 94 % de probabilité en moins ; 95 % CI=[0,01 ; 0,32]), en tenant compte du nombre total de formations liées aux troubles de la parole.

Tableau 6 – Description des orthophonistes ayant participé à l'enquête

Âge		
	Médiane	39
	EIQ	20
	Min.-max.	24–72
Lieu d'exercice		N (%)
Pays	France	101 (85 %)
	Belgique	12 (10 %)
	Suisse	4 (3 %)
	Algérie	1 (1 %)
	Tunisie	1 (1 %)
Milieu ¹	Urbain	64 (54 %)
	Suburbain	29 (24 %)
	Rural	32 (27 %)
Type d'exercice		N (%)
	Libéral uniquement	76 (65 %)
	Salarié uniquement	30 (26 %), dont 21 (70 %) en public
	Mixte	11 (9 %)
Formation initiale		
Diplomation	Médiane (année de dipl.)	2004
	EIQ (années)	21
	Min.-max.	1975–2018
	2011–2017 (N, %)	35 (30 %)
Années de pratique	Médiane	13,5
	EIQ	18
	Min.-max.	1–44
	>20 années (N, %)	38 (32 %)
Années de prise en charge des TP	Médiane	10
	EIQ	14
	Min.-max.	1–42
Formation continue		
	Formations récentes non liées aux TP (nombre moyen de formations distinctes, ET) ²	5 ± 3
	Formations récentes liées aux TP (nombre moyen, ET)	2 ± 1
	Aucune formation liée aux troubles de la parole (N, %)	69 (72 %)
	Parmi les formations liées aux TP (N, %)	
	Formations aux logiciels d'analyse de la voix/parole ≤3 ans	5 (11 %)
	Formations aux logiciels d'analyse de la voix/parole >3 ans	9 (22 %)
Niveau d'expertise en TP autoévalué		N (%)
Évaluation	Novice	25 (26 %)
	Intermédiaire	57 (60 %)
	Expert	13 (14 %)
Prise en charge	Novice	22 (23 %)
	Intermédiaire	63 (66 %)
	Expert	10 (11 %)

¹ Une combinaison était possible en cas de multiples lieux d'exercice.

² Au cours des trois années précédant la complétion du questionnaire.

Note : TP = troubles de la parole ; EIQ = écart interquartile ; ET = écart-type

2. Patients atteints de troubles de la parole

Les données détaillées concernant les patients atteints de troubles de la parole sont présentées dans le [Tableau 7](#).

La proportion de patients atteints de troubles de la parole dans la patientèle globale ne dépasse pas un quart dans 62 % des cas. La majorité de ces patients sont âgés de plus de 50 ans (87 %), et présentent le plus souvent des troubles neurologiques (99 % : pathologies neurodégénératives, neurologiques acquises et tumeurs neurologiques). Les troubles de la parole ont un impact moteur et fonctionnel variable, et induisent dans la grande majorité des cas (95 %) des conséquences psychoémotionnelles et une restriction de la participation aux activités sociales (92 %).

Tableau 7 – Description des patients atteints de troubles de la parole

Pourcentage de patients atteints de TP dans la patientèle	N (%)
0-25 %	55 (62 %)
25-50 %	21 (24 %)
50-75 %	6 (7 %)
75-100 %	6 (7 %)
Âge	
Médiane	66,75
EIQ	12,5
Patients TP >50 ans (N, %)	76 (87 %)
Pathologies¹	
N (%)	
Troubles neurologiques	85 (99 %)
Troubles de la fonction motrice orofaciale	37 (43 %)
Troubles de la communication et du langage dans le cadre d'un handicap moteur, sensoriel ou mental	17 (20 %)
Déficits moteurs	
N (%)	
Imprécision des mouvements articulatoires	80 (93 %)
Difficultés à enchaîner rapidement des mouvements alternatifs - diadococinésies	69 (80 %)
Ralentissement des mouvements	68 (79 %)
Troubles du tonus musculaire des articulateurs	62 (72 %)
Réduction de l'amplitude des mouvements simples	58 (67 %)
Difficultés de coordination des mouvements simples	47 (55 %)
Fatigabilité	43 (50 %)
Tremblements	32 (37 %)
Modification de la sensibilité des articulateurs	31 (36 %)
Syncinésies	24 (28 %)
Impact sur la parole	
N (%)	
Imprécision articulatoire des phonèmes	85 (99 %)
Trouble du débit de la parole (lent, saccadé...)	77 (90 %)
Difficultés dans le séquençage phonémique - dans la coarticulation	75 (87 %)
Rhinolalie ouverte	48 (56 %)
Substitutions phonémiques	40 (47 %)
Omissions phonémiques	40 (47 %)
Inversions phonémiques	12 (14 %)
Ajouts phonémiques	12 (14 %)
Rhinolalie fermée	7 (8 %)
Impact psychosocial	
N (%)	
Psychoémotionnel (faible estime de soi, frustration...)	82 (95 %)
Participation à des activités sociales	79 (92 %)
Exécution de tâches	34 (40 %)
Répercussions financières	16 (19 %)

¹Selon la Nomenclature Générale des Actes en Orthophonie, 1^{er} janvier 2019

Évaluation clinique des troubles de la parole : situation actuelle

1. Fréquence des sessions d'évaluation

Le nombre moyen d'évaluations par trimestre est de 6,32 par orthophoniste. Près de la moitié des cliniciens effectuent moins de trois évaluations de troubles de la parole par trimestre.

2. Outils d'évaluation

Batteries d'évaluation. Environ un quart des praticiens (26 %) n'utilisent aucune batterie d'évaluation existante. Les autres utilisent les batteries suivantes, soit entièrement (34 %), soit partiellement (40 %) : Batterie d'Évaluation Clinique de la Dysarthrie (BECD) [AUZOU 06], Évaluation Clinique de la Dysarthrie (ECD, ancienne version de la BECD) [AUZOU 98], Test quantitatif d'intelligibilité [DITTNER 10], Frenchay Dysarthria Assessment-Second Edition (FDA-2) [ENDERBY 08] ou protocole MonPaGe [LAGANARO 21; LEVEQUE 16; PERNON 20; TROUVILLE 21].

Enregistrement de la parole. Seuls deux orthophonistes déclarent ne pas faire d'enregistrements. Trois des 76 cliniciens ayant répondu à la question effectuent des enregistrements dans une pièce insonorisée (deux salarié[e]s et un[e] en libéral). Vingt-neuf pour cent des orthophonistes ont répondu à la question sur le type de microphone utilisé (voir [Figure 14](#)).

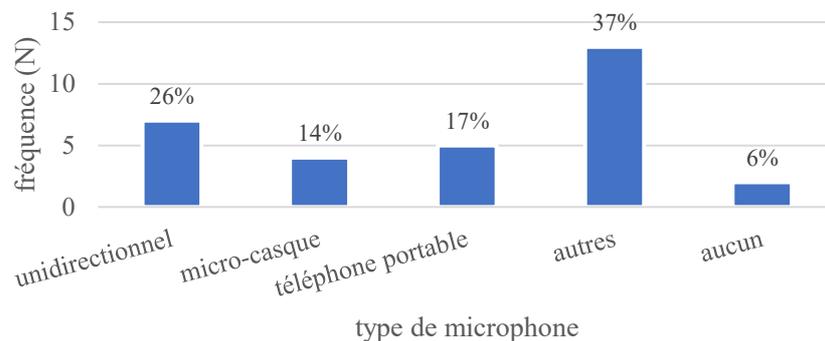


Figure 14 – Types de microphones utilisés pour les bilans des TPP ; « autres » : microphones sur pied, dictaphones et micros-cravates

Seuls 14 % des participants ont répondu à la question du mode d'enregistrement (mono : 53 %, stéréo : 29 %, « je ne sais pas » : 18 %) et 7 % ont répondu à la question de la fréquence d'échantillonnage des enregistrements.

Vingt-sept pour cent des orthophonistes ont répondu à la question sur le logiciel d'enregistrement/d'analyse utilisé : Praat (n=17) [BOERSMA 01], Vocalab (n=16) [MENIN-SICARD 01], Audacity (n=5) [AUDACITY TEAM 99], Diadolab [MENIN-SICARD 06] et Sound Studio (n=1) [FELT TIP INC. 08]. Les deux orthophonistes qui n'utilisent aucun logiciel sont les mêmes qui n'utilisent pas de microphone.

Tâches et mesures. Les tâches les plus fréquemment utilisées sont la répétition de mots (95 %), la parole spontanée (95 %), les tâches de motricité et de sensibilité orofaciales (85 %), la lecture de textes (77 %) et la répétition de phrases (76 %). Les autres tâches moins fréquemment utilisées (moins de 69 %) sont, par ordre décroissant : la parole semi-dirigée, la

lecture de mots, la répétition de pseudomots, la lecture de phrases, la lecture de pseudomots, les tâches de phonation (temps maximal phonatoire, projection vocale, coordination pneumophonique), les diadococinésies et les tâches de prosodie.

Les principaux scores obtenus lors de ces tâches (voir [Figure 15](#)) sont les évaluations perceptives globales d'« intelligibilité » (63 %), de « sévérité » (59 %) et de « compréhensibilité » (45 %), suivies du pourcentage d'erreurs entre phonèmes attendus et perçus (41 %). Quatorze pour cent indiquent qu'ils ne calculent aucun score sur la base des tâches de parole.

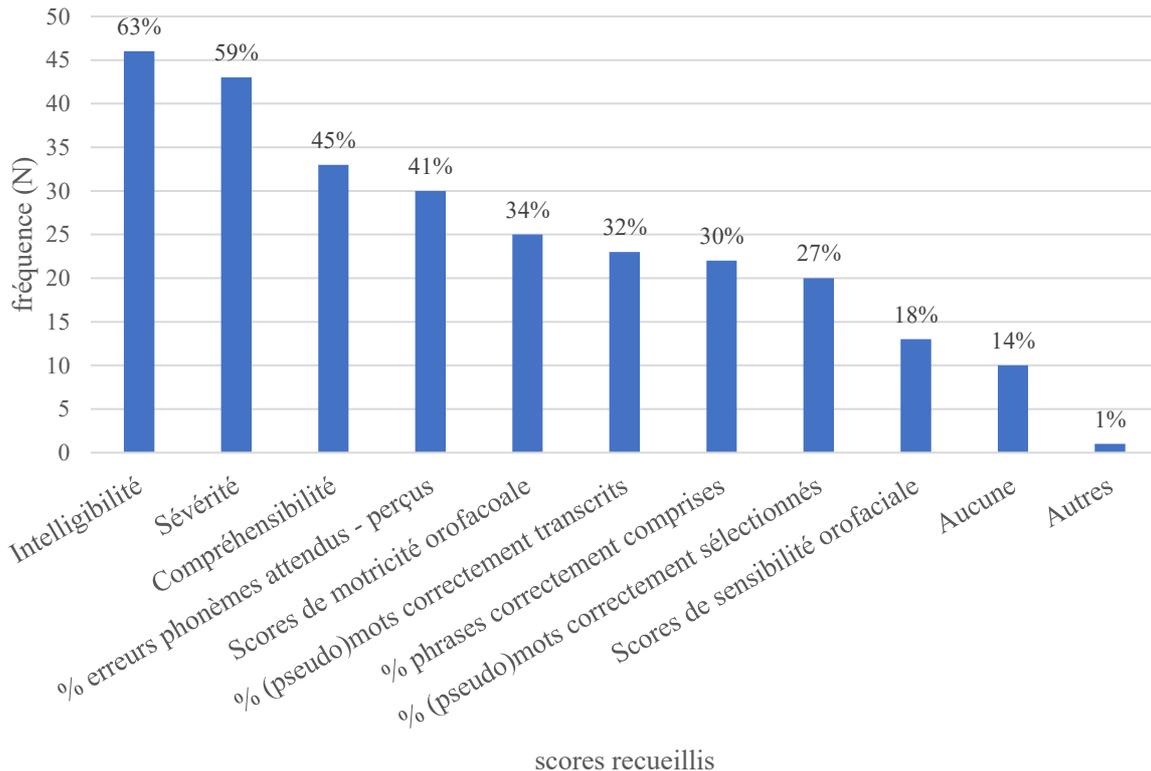


Figure 15 – Scores recueillis lors de l'évaluation des troubles de la parole

Dans la majorité des cas (82 %), le clinicien se fie uniquement à son propre jugement. Dix-sept pour cent des orthophonistes combinent leur évaluation avec celle d'un second auditeur.

Impact fonctionnel et qualité de vie. Parmi les questionnaires de handicap, d'impact fonctionnel et de qualité de vie liés à la parole, le plus couramment utilisé est le Voice Handicap Index (70 %) [JACOBSON 97]. Le Speech Handicap Index (15 %) [DEGROOTE 11] et le Phonation Handicap Index (11 %) [FICHAUX-BOURIN 09] sont rarement employés. Près de 25 % des orthophonistes n'utilisent aucun questionnaire.

Durée de l'évaluation et fatigabilité des patients. La durée médiane du bilan de parole est de 78 minutes, avec une anamnèse de 30 minutes (EIQ=20) et une durée d'épreuves de 48 minutes (EIQ=20). La plupart des patients atteints de troubles de la parole ont un niveau de fatigabilité modéré à élevé (niveau de fatigabilité >3/5 : 93 %).

Manques et besoins en matière d'évaluation de la parole

1. Suffisance des outils d'évaluation de la parole

Dans l'ensemble, les outils disponibles ne sont jugés que moyennement satisfaisants. Sur une échelle allant de 1 (totalement insuffisant) à 5 (totalement suffisant), les outils d'évaluation quantitative¹⁴ sont notés significativement plus bas que les outils d'évaluation qualitative ($2,57 \pm 1$ et $3,14 \pm 0,99$, respectivement ; test U de Mann-Whitney : $p=0,001$). La distribution des niveaux de satisfaction attribués est présentée dans la [Figure 16](#).

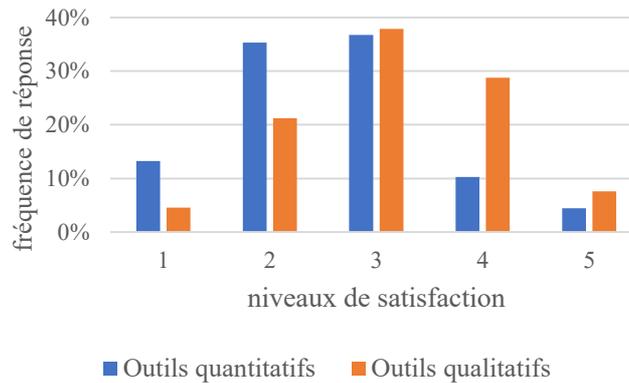


Figure 16 – Distribution des niveaux de satisfaction attribués aux outils d'évaluation de la parole

Les commentaires des orthophonistes soulignent le manque de fiabilité des outils d'évaluation actuellement disponibles, qui sont largement basés sur des évaluations subjectives. En voici quelques exemples :

- « J'ai l'impression de faire un bilan davantage qualitatif que quantitatif, malgré l'utilisation de la batterie ECD » ;
- « La cotation, même si elle se veut objective¹⁵, demeure fonction de l'examineur, de sa sensibilité et de ce qu'il considère comme étant pathologique (ortho expérimentée vs stagiaire ou ortho débutante par exemple) » ;
- « Fiabilité longitudinale si on garde les mêmes méthodes d'enregistrement et le même juge (or biais dans l'évaluation), fiabilité interjuges délicate. » ;
- « Il y a selon moi une part assez importante à la subjectivité du praticien dans la notation des résultats du patient concernant la réussite de certains items parfois » ;
- « Bilans absolument pas reproductibles donc fiables ! On s'est basés sur la BECD pour une étude clinique, c'est une catastrophe en termes de reproductibilité et tous nos résultats sont biaisés ! » ;
- « Manque d'outils objectifs notamment sur la sévérité globale du trouble de la parole et sur la quantification des altérations segmentaires ».

¹⁴ Aucune définition des termes « outils quantitatifs » et « qualitatifs » n'a été fournie aux participants. La distinction entre outils qualitatifs, quantitatifs, subjectifs, objectifs, perceptifs et instrumentaux reste clairement ouverte au débat. Si l'absence de définitions précises doit être gardée à l'esprit lors de l'interprétation du résultat de moindre satisfaction à l'égard des outils « quantitatifs », d'autres éléments non ambigus des résultats de l'enquête contribuent bien à la conclusion d'une utilisation importante de mesures perceptives subjectives, ainsi que d'un manque et d'un besoin de méthodes instrumentales plus objectives et fiables.

¹⁵ Il convient de noter que les notions de « quantification » et d'« objectivité » sont parfois confondues. Une évaluation quantitative n'est pas nécessairement objective.

La perception de la suffisance des outils d'évaluation n'est influencée ni par les années d'expérience dans la prise en charge des troubles de la parole (Kruskal-Wallis : $p > 0,25$), ni par le nombre de formations liées à ces troubles (Kruskal-Wallis : $p > 0,46$), ni par le nombre d'évaluations réalisées par trimestre (test exact de Fisher : $p > 0,08$).

2. Manques rapportés et solutions souhaitées par les cliniciens

Le graphique en [Figure 17](#) montre les manques rapportés par les orthophonistes qui ont participé à cette enquête, concernant les outils d'évaluation de la parole chez les adultes, ainsi que les solutions souhaitées.

Trois grandes catégories de manques ressortent des réponses :

- le manque de validité et de fiabilité des outils (en orange dans la [Figure 17](#)) : il s'agit principalement de la subjectivité de ces outils, entraînant un manque de fiabilité et de reproductibilité, ainsi qu'un manque de repères normatifs ;
- les difficultés liées aux aspects pratiques de l'utilisation des outils (en rouge dans la [Figure 17](#)) : de nombreux cliniciens se plaignent, entre autres, de la nature chronophage et onéreuse de ces outils et de la complexité de leur accès et de leur utilisation ;
- des lacunes dans l'applicabilité clinique (en jaune dans la [Figure 17](#)) : les outils actuellement disponibles manquent d'exhaustivité, sont insuffisants pour la pose de diagnostic et la planification de la prise en charge, et ne permettent pas l'évaluation de paramètres spécifiques tels que la prosodie, la nasalité et le débit de parole ; les outils d'évaluation quantitative fiables semblent rares à ce jour, et les thérapeutes y sont insuffisamment familiarisés.

En ce qui concerne le manque de validité et de fiabilité, d'après les solutions souhaitées rapportées par les cliniciens, ces derniers profiteraient de nouvelles mesures reproductibles, ainsi que de critères de cotation clairement définis et de bases de données audio de référence pour améliorer les évaluations subjectives, et de données normatives pour comparer les performances des patients.

Toujours selon les commentaires des répondants, les aspects pratiques de l'utilisation d'un tel outil d'évaluation seraient améliorés par le développement d'une solution abordable, facile d'accès comme d'utilisation, avec des conditions simplifiées d'enregistrement et de stockage des données. La portabilité serait également un grand avantage dans les différents contextes cliniques.

Afin de pallier le manque d'exhaustivité des outils actuellement existants, nos résultats montrent qu'une attention particulière devrait également être accordée à certains paramètres tels que la prosodie, la nasalité et le débit de parole. De plus, des tâches spécifiques, telles que l'utilisation de pseudomots pour cibler les paramètres de bas niveau de la parole, devraient être combinées avec des tâches d'évaluation plus écologiques. En outre, des questionnaires adressés à l'entourage du patient pourraient permettre d'obtenir une vision plus complète de la communication du patient.

Enfin, les cliniciens bénéficieraient également d'une meilleure information sur les outils d'évaluation de la parole existants et des formations continues, notamment en ce qui concerne l'évaluation acoustique des troubles de la parole.

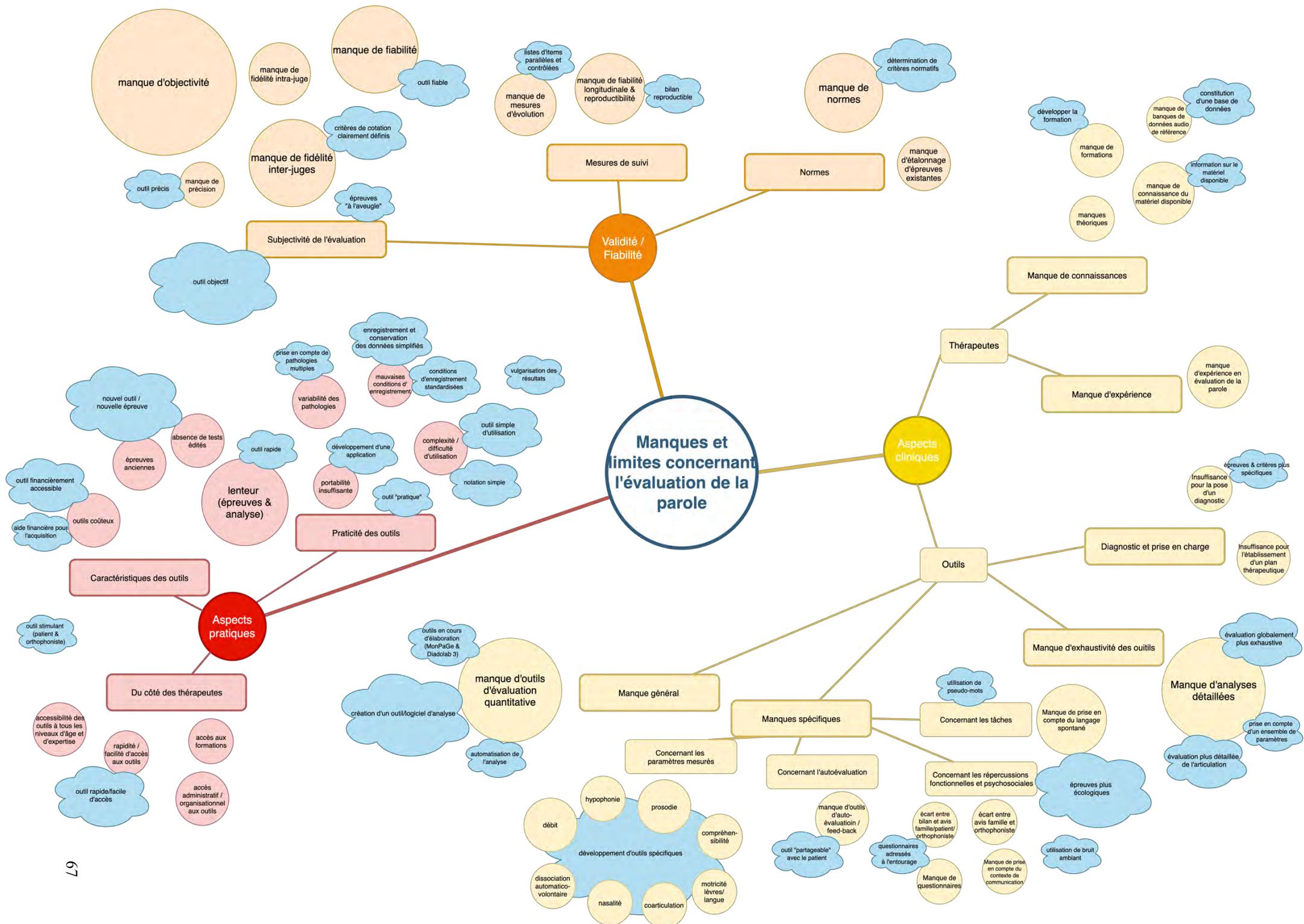


Figure 17 – Manques rapportés et solutions souhaitées par les cliniciens concernant l'évaluation des troubles de la parole

C. Discussion

Comme développé dans l'introduction, plusieurs dimensions doivent être abordées dans l'évaluation de la parole afin de fournir un plan thérapeutique adapté à la situation spécifique de chaque patient. Deux besoins principaux ont été identifiés dans cette étude. Tout d'abord, un **besoin de standardisation** de l'évaluation de la parole : nos résultats montrent que non seulement la structure globale de l'évaluation de la parole (c.-à-d. les dimensions évaluées) diffère d'un clinicien à l'autre, mais aussi les tâches d'évaluation utilisées pour chaque dimension, ainsi que la procédure d'enregistrement audio. Une deuxième problématique importante qui ressort de nos résultats et qui sera discutée ici est la subjectivité de la plupart des tâches d'évaluation, ainsi que le **besoin de mesures plus fiables**.

Standardisation

Le manque de standardisation de l'évaluation est observé principalement à deux niveaux : les tâches de production de la parole dans chaque dimension et les conditions d'enregistrement.

En ce qui concerne les **tâches de production de la parole** (parole spontanée, répétition ou lecture de mots, de phrases, de textes...), le plus souvent, les orthophonistes créent des évaluations « à la carte ». Les batteries d'évaluation ne sont utilisées dans leur intégralité que par un tiers des orthophonistes, tandis que les autres n'en tirent que quelques épreuves (négligeant ainsi parfois certaines dimensions de l'évaluation de la parole). La préférence pour cette évaluation « informelle » par rapport aux outils standardisés a été rapportée au niveau international dans plusieurs autres enquêtes [CONWAY 15; GUREVICH 17; PALMER 05; RUMBACH 19]. Les causes possibles sont l'absence de critères standardisés pour choisir entre les outils d'évaluation [PALMER 05], le manque de temps et de financement, l'insatisfaction vis-à-vis des outils existants et la difficulté d'accès [GUREVICH 17; RUMBACH 19]. Bien sûr, des différences existent entre les contextes de soins de santé des pays francophones et peuvent expliquer certains des résultats. En France par exemple, où exerçait la majorité des répondants, le clinicien dispose d'une autonomie totale dans le choix des outils d'évaluation, alors qu'en Belgique, il existe des directives plus strictes (listes limitatives) qui conditionnent l'intervention de l'assurance maladie nationale.

De plus, nos résultats montrent que la satisfaction limitée quant aux outils d'évaluation actuellement disponibles est indépendante de l'ancienneté des cliniciens, de leur patientèle et de leur formation continue. Le fait que cette insatisfaction soit ainsi généralisée souligne que la cause première de ce ressenti provient très probablement des outils d'évaluation eux-mêmes et met en évidence la nécessité de remédier directement à ce problème en proposant de nouvelles alternatives standardisées.

Il convient de noter que les batteries d'évaluation de la parole standardisées françaises n'étaient pas disponibles au moment de la réalisation de l'enquête : La BECD/ECD n'était plus disponible sur le marché, et la traduction française du FDA-2 [BLANC 14] a été présentée dans le cadre d'un mémoire de maîtrise en orthophonie, mais n'est toujours pas publiquement disponible. Bien que cela souligne un manque inhérent d'outils d'évaluation de la parole standardisés, il faut également en tenir compte quant au constat de préférence des méthodes informelles d'évaluation de la parole. En effet, la disponibilité des outils d'évaluation ainsi que

leurs limites contribuent toutes deux à l'insatisfaction des cliniciens ; leurs contributions respectives à cette dernière restent cependant inconnues.

Le **protocole d'enregistrement** utilisé par les cliniciens varie également beaucoup, ce qui peut compromettre la fiabilité des mesures acoustiques. Tout d'abord, les instructions données aux cliniciens restent très souples, ce qui entraîne une grande variabilité des mesures (par exemple, dans le BECD, l'intensité peut être mesurée sur une phrase simple et courte ou sur une voyelle soutenue). De plus, alors que la procédure d'enregistrement du signal de parole peut affecter les valeurs mesurées, les modalités d'enregistrement dans les évaluations de la parole sont très disparates. Les appareils utilisés ne sont pas standardisés, certains cliniciens utilisant des microphones de haute qualité alors que d'autres préfèrent utiliser des appareils plus accessibles et ne nécessitant pas d'investissement supplémentaire (par exemple leur téléphone portable). Enfin, les nombreuses non-réponses aux questions sur les paramètres d'enregistrement indiquent que peu d'orthophonistes semblent à l'aise avec ces paramètres acoustiques, vraisemblablement à cause d'une connaissance limitée de ceux-ci.

Pour pallier le manque général de standardisation, les répondants suggèrent une meilleure information des cliniciens quant aux outils d'évaluation de la parole disponibles — en particulier quant à l'utilisation des outils d'évaluation acoustique —, des directives plus claires concernant les tâches d'évaluation de la parole, ainsi que des formations continues en évaluation de la parole.

Une meilleure standardisation de l'évaluation de la parole, au moins en ce qui concerne les dimensions à évaluer, aurait deux avantages principaux. Premièrement, elle permettrait une procédure d'évaluation reproductible, et donc un recueil des données plus fiable pour les évaluations de suivi (c.-à-d. une meilleure fidélité intrajuge). Les comparaisons seraient non seulement plus pertinentes entre l'évaluation initiale et les évaluations de suivi pour chaque patient, mais aussi entre des patients présentant des profils cliniques similaires. De plus, dans la philosophie de l'EBP, des procédures d'évaluation homogènes entre professionnels permettraient de faciliter la mise en commun et la comparaison des résultats pour promouvoir des interventions fondées sur les preuves, grâce à une meilleure fiabilité interjuges.

À la lumière des présents résultats, des recommandations générales pour les enregistrements audio et une proposition indicative d'un ensemble minimal de tâches pour l'évaluation de la parole chez l'adulte sont disponibles en [Annexe G](#). Ces suggestions générales sont destinées à proposer une structure globale de l'évaluation de la parole, mais peuvent nécessiter des adaptations en cas de comorbidités (p. ex. déficits cognitifs ou sensoriels associés) et des tâches supplémentaires liées à des contextes pathologiques spécifiques. Une discussion exhaustive de cet ensemble minimal dépasse le cadre de ce travail. Nous limiterons donc notre discussion à deux points particuliers qui ont été mis en évidence dans nos résultats.

Premièrement, il ressort que les cliniciens utilisent peu les pseudomots, probablement en raison de leur absence dans les batteries standardisées. Certains cliniciens estiment qu'il s'agit là d'une lacune. En effet, l'utilisation de pseudomots permet d'éviter les processus de compensation sémantique et syntaxique «top-down» [GHIO 17] et ainsi d'évaluer l'**intelligibilité**, l'une des cibles principales dans la prise en charge des troubles de la parole [WALSHE 08]. Pourtant, l'utilisation prédominante de mots et de phrases indique que

l'évaluation porte plutôt sur la **compréhensibilité**. Le fait de se concentrer davantage sur les tâches fonctionnelles de la parole est tout à fait cohérent dans un contexte clinique, car cela répond au mieux à la plainte du patient. Cependant, comme nous l'avons déjà souligné au [Chapitre 1](#), la terminologie utilisée par les orthophonistes semble révéler un manque de clarté quant à ces concepts : la plupart disent évaluer l'intelligibilité, alors qu'au vu des tâches employées, le concept ciblé correspond davantage à la compréhensibilité. Par conséquent, en premier lieu, la standardisation de l'évaluation de la parole bénéficierait d'une définition consensuelle des termes liés à la parole, tels que l'intelligibilité et la compréhensibilité, afin de permettre des instructions plus cohérentes pour le clinicien et d'éviter toute ambiguïté. Cet objectif a été poursuivi dans le cadre de cette thèse et est abordé dans le [Chapitre 1](#). Plus important encore que la terminologie, un consensus sur les dimensions conceptuelles qui doivent être évaluées pour répondre aux objectifs de l'évaluation de la parole est nécessaire. Nous estimons, par exemple, qu'en plus d'une tâche de compréhensibilité et d'une évaluation de la capacité du patient à agir sur ses partenaires de communication et son environnement, une tâche ciblant l'intelligibilité semble également pertinente. En effet, cette dernière permet des mesures de performance plus fiables et une définition d'objectifs plus spécifiques lors de la planification thérapeutique. Par conséquent, notre ensemble minimal comprend à la fois des tâches d'intelligibilité et des tâches d'évaluation de la compréhensibilité plus fonctionnelles.

Deuxièmement, au-delà du déficit symptomatique, la plainte du patient concerne également le plus souvent les **répercussions fonctionnelles et psychosociales** du trouble de la parole. Ainsi, les modèles pathologiques strictement biomédicaux ne suffisent pas à décrire et à comprendre l'impact des troubles de la parole [WADE 04; WRAY 17], dans des environnements de vie et de travail spécifiques à chaque patient. Cet impact est le plus souvent évalué à l'aide de questionnaires. Cependant, il apparaît que le questionnaire le plus utilisé est le Voice Handicap Index, qui a été validé pour les troubles de la voix. Son utilisation dans le contexte des troubles de la parole manque donc de spécificité. Le Speech Handicap Index (SHI) et le Phonation Handicap Index (PHI), validés pour les troubles de la parole, sont quant à eux rarement utilisés. Ce constat, associé au fait que près d'un quart des orthophonistes n'utilisent aucun questionnaire, souligne que persiste un manque de prise en compte de l'impact des déficits sur la vie quotidienne des patients, malgré un intérêt croissant pour les notions de qualité de vie dans les domaines médicaux et paramédicaux [ORGANISATION MONDIALE DE LA SANTÉ 01; RUMBACH 19]. Néanmoins, le SHI a été inclus dans la nouvelle version de la BECD (octobre 2019), ce qui démontre l'intérêt grandissant pour l'évaluation de l'impact fonctionnel des déficits de la parole. Le PHI, qui évalue spécifiquement les troubles de la parole perçus par le patient, a été validé pour le français en 2009 dans une population de patients majoritairement atteints de troubles neurologiques [FICHAUX-BOURIN 09] ; une étude de validation complémentaire a ensuite été réalisée en 2019, dans une population traitée pour un cancer de la bouche ou de l'oropharynx [BALAGUER 19]. De même, la validation de la version française du Dysarthria Impact Profile [ATKINSON-CLEMENT 19] montre une forte corrélation entre les scores d'impact psychosocial et d'« intelligibilité » (*sic*) (pourcentage de mots et de phrases correctement transcrits).

Enfin, nous pensons qu'un texte standardisé créé spécifiquement pour l'évaluation de la parole et de la voix est nécessaire. En effet, il existe une pléthore de textes dans différentes

langues, dont certains ont été créés dans un but spécifique (p. ex. le Zoo Passage en anglais [S. G. FLETCHER 72], qui ne comprend que des consonnes orales pour l'évaluation de la fermeture vélopharyngée, ou les textes de Kuo et Weismer [KUO 16], spécifiquement créés pour évaluer la réduction vocalique). Cependant, aucun de ces textes ne semble vraiment répondre aux besoins des cliniciens et des chercheurs en matière d'évaluation courante de la parole et de la voix. Un texte unifié répondant à la plupart des critères requis pour les tâches standard d'évaluation de la parole et de la voix (p. ex. inventaire phonémique complet, utilisation de divers contours prosodiques, présence de phrases à prédominance nasale, contenu émotionnellement neutre...) serait donc une option intéressante. À cette fin, j'ai participé à la mise en place d'un groupe de travail multidisciplinaire, et une partie de l'enquête internationale de consensus Delphi a également abordé cette problématique. Les résultats de ce travail seront exposés au [Chapitre 3](#).

Pour synthétiser notre discussion concernant la standardisation de l'évaluation clinique de la parole, voici les différentes solutions envisagées :

- instructions et directives claires pour le clinicien ;
- consensus sur les objectifs et les dimensions de l'évaluation de la parole ;
- ensemble standardisé minimal de tâches (cf. [Annexe G](#)) ;
- définitions consensuelles des termes relatifs à la parole (cf. [Chapitre 1](#)) ;
- recommandations pour les enregistrements audio (cf. [Annexe G](#)) ;
- meilleure information sur les outils d'évaluation de la parole disponibles ;
- formations continues spécifiques en évaluation (acoustique) de la parole ;
- texte standardisé spécialement conçu pour l'évaluation de la parole et de la voix (cf. [Chapitre 3](#)).

Fiabilité et reproductibilité

Outre le manque de standardisation de l'évaluation de la parole, une autre caractéristique des tests en pratique clinique actuelle est leur **subjectivité**, avec une forte prévalence de l'**évaluation perceptive**, même dans les batteries de tests « formelles ». La BECD, par exemple, est basée exclusivement sur l'évaluation perceptive (à l'exception du niveau de pression acoustique et de la fréquence fondamentale). Ceci implique un biais important et largement discuté lié à la subjectivité de l'auditeur, d'autant plus que les définitions et les critères d'évaluation sont sujets à un haut degré de variabilité interprétative. Les notions de fiabilité et de reproductibilité de l'évaluation sont des points cruciaux évoqués par les orthophonistes ayant participé à cette enquête. Ils rapportent également une plus grande satisfaction vis-à-vis de l'évaluation qualitative par rapport aux outils quantitatifs, indépendamment du nombre d'années d'expérience dans les troubles de la parole, du nombre de formations continues suivies et du nombre d'évaluations réalisées par trimestre. L'utilisation majeure de méthodes subjectives et qualitatives mise en évidence dans nos données corrobore les résultats d'études antérieures. Celles-ci démontraient également un emploi limité d'outils d'évaluation non perceptifs, malgré le souhait des cliniciens de disposer de moyens plus fiables, étant donné la subjectivité actuelle de l'évaluation tant informelle que formelle [CONWAY 15; MILLER 17; PALMER 05].

Afin de réduire la subjectivité de l'évaluation perceptive, plusieurs cliniciens ont exprimé le souhait de disposer d'une **base de données de référence** avec des enregistrements de parole pathologique de différents degrés, pour mieux calibrer leurs jugements de niveau de sévérité des troubles. En effet, de nombreuses études ont montré l'amélioration de la fiabilité interjuges par l'utilisation d'ancrages auditifs et textuels dans l'évaluation des échantillons de voix et de parole, notamment par une meilleure distinction perceptive entre les enregistrements « sains » et les niveaux de sévérité légers [AWAN 09; BRINCA 15].

L'utilisation de **mesures instrumentales non perceptives** est une autre solution pertinente pour compenser le manque d'objectivité. À ce jour, ces outils restent peu disponibles, et les paramètres acoustiques utilisés dans l'évaluation de la parole sont assez pauvres. De plus, ils ne portent en grande partie pas sur l'articulation des phonèmes. Une étude décrivant l'évaluation instrumentale de la dysarthrie en France [JAN 07] indique que celle-ci consiste principalement en des mesures acoustiques de la phonation (intensité, fréquence fondamentale, jitter, shimmer, étendue vocale) ainsi qu'en des mesures aérodynamiques (p. ex. volume pulmonaire courant, temps maximum phonatoire, pression intraorale). Le manque d'exhaustivité de l'évaluation et la non-prise en compte de paramètres spécifiques de la parole (tels que la prosodie, le débit de parole et la nasalité) sont effectivement des manques soulignés par nos participants. Pour les futurs outils d'évaluation, il serait donc intéressant d'introduire des mesures acoustiques plus spécifiques au domaine de l'articulation de la parole/de phonèmes, permettant au clinicien d'obtenir une analyse rendant plus directement compte de la performance de production de la parole de ses patients. Il est à noter que les outils d'évaluation acoustique de la parole sont de plus en plus développés. Le logiciel Diadolab version 3, par exemple, est sorti en avril 2019, peu avant la fin de la période de réalisation de l'enquête en ligne. Ce logiciel comprend un module d'évaluation et un module de prise en charge, et utilise diverses mesures acoustiques de la parole (prosodie, phonologie et diadococinésies). Un seul participant a déclaré avoir utilisé ce logiciel. Le Carcinologic Speech Severity Index (C2SI) [ASTÉSANO 18; WOISARD 21] est un indice automatisé d'évaluation de la sévérité des troubles de la parole validé sur une population oncologique, dont le score est calculé sur trois tâches (répétition de pseudomots, lecture de texte et voyelle tenue). Enfin, le protocole informatisé MonPaGe, publié en avril 2021, est une étape prometteuse vers la résolution des deux principaux problèmes discutés dans cette étude. Il utilise un ensemble standardisé de mesures perceptives et acoustiques et combine l'investigation de modules analytiques tels que des pseudomots pour évaluer la précision articulatoire, avec des tâches écologiques telles que des situations d'interaction communicative [LAGANARO 21; LEVEQUE 16; PERNON 20; TROUVILLE 21]. Cependant, les indices acoustiques sont encore limités aux mesures de la phonation et des voyelles. Les consonnes contribuent également de manière significative à l'intelligibilité de la parole [BONATTI 05; CHANG 17; OWREN 06] et devraient donc être étudiées de manière plus approfondie pour les futurs outils d'évaluation. Ceci sera l'objet de notre quatrième chapitre.

Dans l'ensemble, nos résultats montrent que les outils d'évaluation de la parole actuellement disponibles dans les pays francophones ne semblent pas répondre aux attentes et aux besoins des cliniciens. Tout comme les enquêtes menées dans d'autres communautés linguistiques [COLLIS 12; CONWAY 15; GUREVICH 17; MILLER 17; RUMBACH 19], notre étude souligne donc le besoin de nouveaux outils sensibles, standardisés, fiables et exhaustifs pour

évaluer les troubles de la parole à la fois au niveau analytique et à des niveaux plus fonctionnels. La [Figure 18](#) synthétise les besoins exprimés par les cliniciens dans notre enquête en termes d'évaluation de la parole. Il est intéressant de noter que les perspectives identifiées sont similaires à celles rapportées par Gurevich et Scamihorn aux États-Unis (traduit de l'anglais) : « La nécessité de développer de nouveaux outils formels rentables, utiles, efficaces et simples à utiliser pour évaluer objectivement (*sic*) l'intelligibilité (*sic*) est avérée. Une des méthodes pour éliminer la subjectivité de l'évaluation de l'intelligibilité (*sic*) est de se tourner vers la technologie. » [GUREVICH 17]

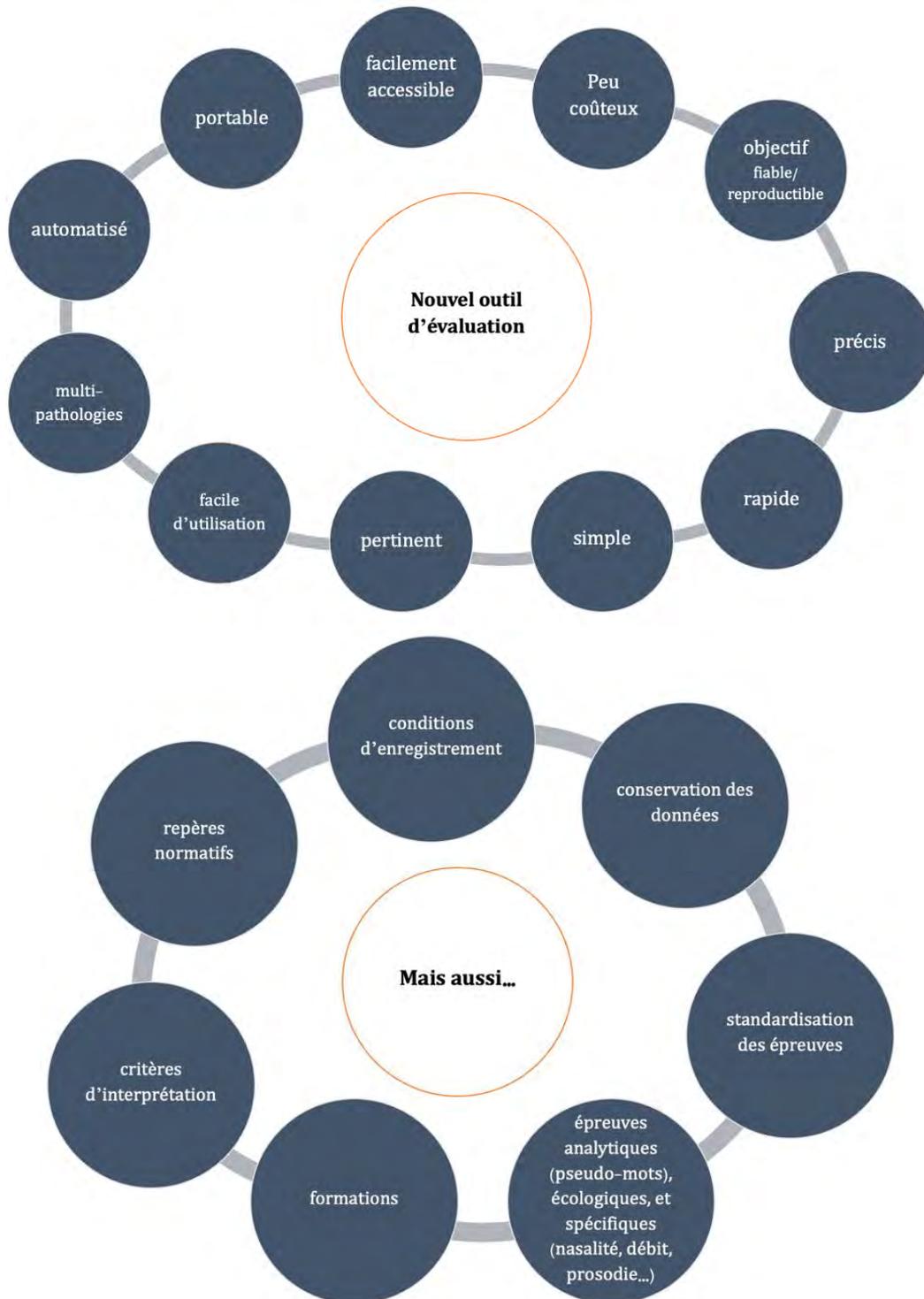


Figure 18 – Synthèse des besoins cliniques en matière d'évaluation de la parole identifiés dans notre enquête

D. Conclusion

Ce n'est qu'en prenant en compte les opinions des cliniciens que le transfert en pratique clinique d'un outil créé par des experts scientifiques peut se faire de manière idéale. Aussi, afin de prendre en compte le deuxième pilier de l'EBP, nous avons mené une enquête auprès des cliniciens francophones concernant les pratiques actuelles et les manques et besoins en termes d'évaluation de la parole.

Les résultats de cette étude soulignent la nécessité de proposer de nouveaux outils d'évaluation des troubles de la parole qui sont fiables/reproductibles, faciles à utiliser et financièrement accessibles. L'évaluation s'en trouverait plus simple, précise, rapide, écologique et pertinente dans le contexte des multiples pathologies pouvant affecter la parole, avec des critères d'interprétation et des repères normatifs clairement définis. La création d'un nouvel outil standardisé d'évaluation de la parole répondant à ces critères, incluant notamment des mesures acoustiques, permettrait une adéquation optimale avec les besoins exprimés par les thérapeutes.

Pour qu'un tel nouvel outil soit fiable et valide, et compte tenu de la familiarité limitée de nombreux cliniciens avec les paramètres d'enregistrement détaillés, une formation plus exhaustive des cliniciens à l'évaluation acoustique de la parole, ainsi qu'une standardisation et une simplification de la procédure d'enregistrement et des conditions de stockage des données semblent nécessaires. En outre, l'évaluation bénéficierait également d'une standardisation des dimensions à évaluer, des tâches (incluant des pseudomots, des épreuves plus « écologiques », et des épreuves plus spécifiques pour le débit de parole, la nasalité, la prosodie...), ainsi que des instructions fournies. Sur la base de ces constats, un ensemble de tâches minimales ainsi que des suggestions pour les conditions d'enregistrement sont fournis à titre de recommandation pour les cliniciens dans l'Annexe G.

2.3.2. Évaluation de la parole chez l'adulte – opinions d'experts

Pour compléter les données récoltées dans la littérature scientifique et auprès des cliniciens, nous avons mené des interviews semi-dirigées auprès de chercheurs dans différents domaines applicatifs liés à la parole (voir [Figure 19](#)). L'objectif était d'obtenir leur avis concernant la définition de l'intelligibilité et son évaluation, ainsi que concernant les manques et lacunes des outils d'évaluation existants.

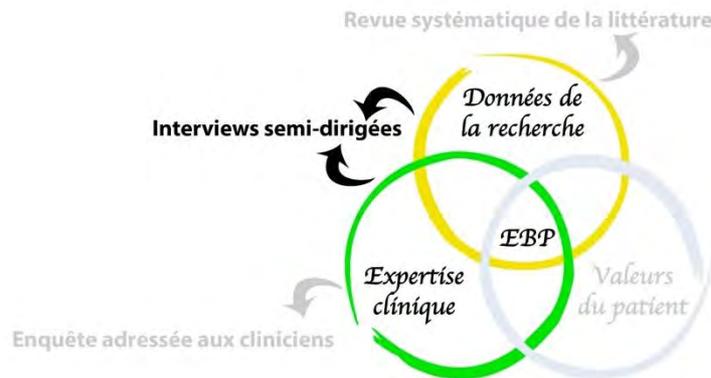


Figure 19 – Prise en compte des premier et deuxième piliers de l'EBP : interviews semi-dirigées d'experts de la parole

A. Méthodologie

Recrutement

Les participants ont été recrutés sur la base de leur activité dans le domaine de la recherche sur la parole, c.-à-d. qu'ils répondaient à au moins un des critères suivants : avoir publié au moins trois articles scientifiques liés à la recherche sur la parole ; être l'auteur d'un livre de référence ; participer à des projets de recherche liés à la parole pathologique.

Interviews

Les interviews étaient menés en vidéoconférence ou en face à face. Il s'agissait d'interviews semi-dirigées, sous forme d'une conversation guidée par des questions prédéterminées, laissant libre cours à l'apparition d'autres questions éventuelles au fil de l'échange [BARRIBALL 94; IMBERT 10; WHITING 08]. Chaque interview était précédée d'une brève présentation de ma part (parcours d'études), du contexte TAPAS, du sujet de ma thèse et des principaux axes de travail, d'une définition des troubles de la parole, ainsi que d'une introduction des objectifs et du déroulement de l'interview. Aucune définition de concepts liés à la parole n'était fournie (notamment de l'intelligibilité lors de la présentation du sujet de ma thèse) afin de ne pas induire de biais dans les réponses des participants. Le guide d'entretien comprenait des questions de 4 grandes catégories :

1. Présentation de l'expert (formation et activité professionnelle, type d'exercice...);
2. Caractéristiques de la population atteinte de troubles de la parole (principales pathologies, limitations physiques, fonctionnelles, répercussions sociales...);
3. Terminologie (définition de l'intelligibilité; le cas échéant, différence entre intelligibilité, clarté, compréhensibilité, sévérité du trouble);
4. Évaluation de la parole (éléments indispensables lors d'un bilan des troubles de la parole, outils perceptifs et automatiques connus, mesures employées, sur quel support,

degré de satisfaction, prise en compte de la dimension fonctionnelle et de la qualité de vie, manques et points problématiques, solutions envisageables...).

Chaque interview était enregistrée avec l'accord du participant, puis retranscrite orthographiquement. La transcription orthographique a été effectuée à l'aide d'un outil de reconnaissance automatique de la parole de l'équipe SAMoVA¹⁶ (fondée sur la boîte à outils Kaldi¹⁷) [HEBA 21], puis corrigée manuellement. Une analyse qualitative du contenu de ces interviews a ensuite été menée indépendamment par deux lecteurs (les mêmes que précédemment). Après une première lecture de l'interview, ces lecteurs ont effectué un étiquetage des fragments de texte pour indiquer les thèmes principaux et identifier les réponses aux questions prédéterminées. Un tableur Excel a ensuite permis de mettre en parallèle les éléments de réponse identifiés par chacun des deux lecteurs, afin de s'assurer que ceux-ci étaient identiques et de limiter les biais de sélection de réponses. Un tableau reprenant les éléments de réponse identifiés conjointement par les deux lecteurs a ensuite été établi, permettant l'analyse finale des données.

B. Résultats

Participants

Trente experts ont été contactés par courriel. Un nombre total de 14 entretiens semi-dirigés a été réalisé, dont huit en ligne et six en face à face, en France, en Angleterre et à Taiwan. Le [Tableau 8](#) reprend une description plus détaillée du panel de participants.

Tableau 8 – Description du panel d'experts ayant participé aux interviews semi-dirigées

Pays	N (%)
États-Unis (Arizona, Wisconsin, Utah)	3 (21,4 %)
France	3 (21,4 %)
Angleterre	2 (14,3 %)
Australie	2 (14,3 %)
Allemagne	1 (7,1 %)
Belgique	1 (7,1 %)
Suède	1 (7,1 %)
Pays-Bas	1 (7,1 %)
Domaine professionnel	
Recherche en orthophonie	6 (42,8 %)
Traitement du signal	4 (28,6 %)
Sciences de la parole, phonétique, psycholinguistique, neurosciences, audiologie	4 (28,6 %)
Niveau d'éducation	
MCF/HDR, professeur agrégé	7 (50 %)
Professeur	3 (21,4 %)
Professeur émérite	2 (14,3 %)
Docteur	1 (7,1 %)
Master	1 (7,1 %)

¹⁶ <https://www.irit.fr/departement/signaux-images/samova/>

¹⁷ <https://kaldi-asr.org/>

Ancienneté (année d'obtention du titre de docteur)	
2001-2010	7 (50 %)
1970-2000	5 (36 %)
2012	1 (7,1 %)
En cours	1 (7,1 %)
Population d'étude	
TP d'origine neurologique	11 (79 %)
TP développemental chez l'enfant	4 (28,6 %)
TP chez la personne malentendante	3 (21,4 %)
TP d'origine neurologique (enfant)	2 (14,3 %)
TP dans un contexte oncologique	1 (7,1 %)
Parole dans l'apprentissage des langues	1 (7,1 %)
Parole saine (auparavant pathologique)	1 (7,1 %)

Dans le [Tableau 9](#), un croisement des activités professionnelles et des niveaux d'éducation des participants donne un aperçu des situations professionnelles du panel de participants au moment de l'interview.

Tableau 9 – Activités professionnelles et niveaux d'éducation des participants aux interviews semi-dirigées

			Activité professionnelle, N (%)				
			Recherche	Académique	Industrie	Clinique	Informatique
			14 (100 %)	12 (86 %)	2 (14,3 %)	6 (42,8 %)	4 (28,6 %)
Niveau d'éducati on, N (%)	Master	1 (7,1 %)	1 (7,1 %)			1 (7,1 %)	
	Docteur	1 (7,1 %)	1 (7,1 %)		1 (7,1 %)		1 (7,1 %)
	MCF/HDR	7 (50%)	7 (50%)	7 (50%)		3 (21,4%)	3 (21,4%)
	Professeur	3 (21,4 %)	3 (21,4 %)	3 (21,4 %)	1 (7,1 %)	1 (7,1 %)	
	Prof. émérite	2 (14,3 %)	2 (14,3 %)	2 (14,3 %)		1 (7,1 %)	

Pathologies et symptômes de parole identifiés dans les populations d'étude

Les populations d'étude des experts interrogés sont principalement les patients atteints de dysarthrie et d'apraxie, de dysphonie et de dysphagie. Les participants rapportent une grande variabilité des symptômes physiques et fonctionnels entre les différentes pathologies, mais également au sein d'un groupe pathologique. Ainsi, même si une lésion peut être localisée, il sera difficile de prédire les symptômes, notamment à cause des phénomènes de compensation et d'adaptation.

Terminologie des concepts liés à la parole : intelligibilité, compréhension et autres termes

Les stratégies utilisées pour définir les termes diffèrent selon les participants. Alors que certains utilisent des définitions théoriques, d'autres décrivent les concepts au moyen de tâches d'évaluation (p. ex. le pourcentage de phonèmes corrects pour définir l'intelligibilité, la réalisation d'une tâche par l'auditeur pour la compréhension).

L'**intelligibilité** est définie comme (N=14) :

- un concept analytique/concernant l'information phonémique/acoustique (N=6) ;
- la capacité de comprendre un interlocuteur au niveau des mots (N=4) ;
- la capacité d'être compris par l'auditeur (N=3) ;
- la capacité de « recevoir un message » sans information contextuelle (N=1).

Le concept de « micro-intelligibilité » a également été mentionné dans l'un des entretiens comme un concept segmental de bas niveau, par rapport à l'intelligibilité au niveau du mot (« indépendante du sens » [*sic*]).

La **compréhensibilité** est définie comme (N=13) :

- un concept de « plus haut niveau » que l'intelligibilité, utilisant des indices lexicaux, contextuels, visuels pour saisir le contenu sémantique global du message : 9 ;
- un synonyme de l'intelligibilité : 4.

Contrairement à ces deux termes, qui sont largement utilisés (malgré les disparités dans les définitions) et sont quantifiables à l'aide de tâches d'évaluation spécifiques, d'autres termes sont utilisés de manière plus sporadique et qualitative. Ces termes sont purement subjectifs et présentent donc une variabilité encore plus grande. Ils sont considérés par certains comme étant corrélés aux concepts d'intelligibilité ou de compréhensibilité, alors que d'autres soulignent l'indépendance de ces concepts (p. ex., traduit de l'anglais : « *une personne peut être intelligible tout en présentant un mode de production de parole qui n'est pas complètement naturel* »).

Par exemple, le « **caractère naturel** » (« naturalness ») de la parole (également « acceptabilité » ou « étrangeté » [« bizarreness »]), qui est défini comme un terme qualitatif, qui (traduit de l'anglais) « intègre [...] de nombreux symptômes de la parole comme le débit de la parole, ou [...] la qualité de la voix [...] la façon dont le patient module [...] c'est une évaluation très spontanée de la parole ». Il est défini par rapport aux patterns de production de parole habituels (N=8) :

- en termes négatifs : parole qui ne ressemble pas à la parole saine (« à quel point quelqu'un peut sonner étrange, inhabituel » ; « degré auquel la parole sonne déviante ») ;
- en termes positifs : « sonne-t-elle normale, est-ce un plaisir de l'écouter ? » ; « la parole est naturelle si elle n'attire pas l'attention sur elle-même d'une manière inhabituelle ».

La « **sévérité** » (N=6) est également considérée comme un terme qualitatif général, qui inclut l'intelligibilité avec d'autres paramètres perçus de la parole, tels que le caractère naturel (la parole peut « sonner sévèrement détériorée », tout en restant intelligible et compréhensible), ainsi que des aspects plus fonctionnels relatifs à l'impact du trouble de la parole dans la vie quotidienne du patient. Notons qu'un expert distingue deux « types » de sévérité : la sévérité globale, qui peut être évaluée par des auditeurs naïfs, et la sévérité de paramètres spécifiques de la parole (phonation, articulation, prosodie...), qui doit être évaluée par des évaluateurs experts. Une fois encore, ceci démontre l'utilisation non consensuelle des termes liés à la parole. Enfin, la « **clarté** » (« clarity », N=5) fait référence à la précision de la production de la parole. À nouveau, celle-ci peut être considérée en une certaine mesure indépendante de l'intelligibilité de la parole : « vous comprenez le segment phonémique [...] mais peut-être n'est-il pas parfaitement précis [...] la dysarthrie est un bon exemple où [...] vous pouvez comprendre les segments [...] mais ce n'est pas parfaitement clair et précis ».

Les autres termes mentionnés par les participants sont les suivants :

- la précision (« accuracy ») : selon l'un des experts, « le nombre de phonèmes que vous percevez correctement », un concept de niveau inférieur par rapport à l'intelligibilité (définie comme la précision au niveau des mots par ce même expert) ;

- l'efficacité de la communication (« communication efficiency ») : « le nombre de mots intelligibles produits dans une période de temps [...] le temps qu'il faut pour transférer ce message » ;
- l'effort communicatif (« communicative effort ») : « l'effort cognitif ou physiologique investi dans la production de la parole » ;
- l'effort de l'auditeur perçu (« perceived listener effort ») : « le degré d'effort nécessaire pour comprendre un patient » ;
- mais aussi la pertinence des pauses, la nasalité, la qualité vocale, l'imprécision consonantique et les allongements phonémiques.

Évaluation de la parole : stimuli, mesures perceptives et « objectives » et impact fonctionnel

1. Stimuli

Selon les réponses des participants, les mesures de la parole peuvent être effectuées à différents niveaux de granularité : au niveau du phonème (2/14), du mot (11/14), de la phrase (10/14, dont non prédictibles/asémantiques : 4/10), sur des lectures à voix haute d'un texte (4/14), de la parole semi-spontanée ou conversationnelle (6/14). Sept experts jugent l'utilisation de pseudomots intéressante pour se débarrasser du contenu lexico-sémantique et des indices syntaxiques pour l'évaluation de l'intelligibilité.

2. Mesures perceptives

À ces différents niveaux de granularité, de nombreux concepts peuvent être évalués (intelligibilité, sévérité, nasalité, pertinence des pauses...).

Pour évaluer ces concepts, il existe plusieurs méthodes d'évaluation perceptive :

- évaluations subjectives à l'aide d'échelles visuelles analogiques, échelles de Likert, ou estimation directe de la magnitude ;
- transcription de mots ou de phrases simples/non-prévisibles, avec cotation ultérieure du pourcentage de consonnes/voyelles/mots corrects ;
- choix multiple de mots avec voisins phonémiques/paires minimales ;
- inventaire phonémique ;
- analyse des distorsions.

Des outils spécifiques ont également été mentionnés par les participants, dont le Mayo Clinic System [DARLEY 69a, 69b], l'Assessment of Intelligibility of Dysarthric Speech (AIDS) [YORKSTON 84], le Frenchay Dysarthria Assessment [ENDERBY 08].

Huit participants ont renseigné un **niveau de satisfaction** précis avec les mesures perceptives sur une échelle de 1 (pas du tout satisfait) à 5 (tout à fait satisfait). La distribution des résultats est illustrée en [Figure 20](#). Parmi les autres experts, l'un a expliqué qu'il attribuerait un niveau de satisfaction de 2/5 lorsque les outils sont utilisés par un seul évaluateur et un niveau de 4/5 lorsqu'un jury d'experts est impliqué. Selon lui, l'implication de plusieurs évaluateurs permet d'identifier et de traiter les valeurs aberrantes. Un autre participant n'a pas pu se décider entre 3 et 4/5. Enfin, quatre participants n'ont pas donné de note.

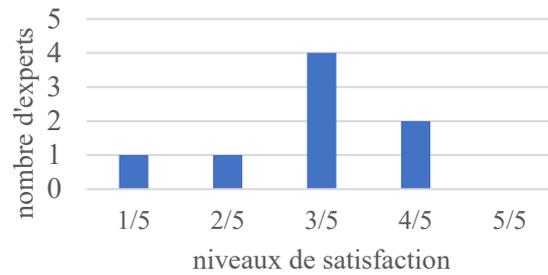


Figure 20 – Satisfaction avec les outils d'évaluation perceptive (N=8)

Bien que les outils perceptifs restent le gold standard nécessaire dans l'évaluation de la parole, la plupart des participants (11/14) se plaignent de la **subjectivité** de ces outils et de la **variabilité** inter- et intraévaluateur. Ils expliquent que les outils perceptifs ne sont pas assez rigoureux et ne sont pas fiables, car ils sont influencés par la familiarité de l'évaluateur avec le patient, avec le type de parole et de pathologie (les référents internes étant idiosyncrasiques et en constante évolution), par le contexte d'évaluation, le sujet de la conversation ainsi que le fait que le locuteur soit visible ou non. La nécessité d'une formation plus approfondie à l'utilisation des mesures perceptives ainsi que la création de bases de données de référence sont mentionnées comme des moyens possibles pour réduire la variabilité.

Cinq experts mentionnent également le caractère **chronophage** de certains outils (par exemple, la transcription orthographique, pourtant considérée comme la plus fiable et la plus précise).

D'autres commentaires concernant les outils d'évaluation perceptive ont été évoqués. Parmi eux, les commentaires positifs sont les suivants :

- les évaluations sont faciles à comprendre ;
- ils sont utiles pour prédire l'intelligibilité lors de la création de mesures automatiques ;
- ils servent à fixer des objectifs thérapeutiques ;
- ils sont utiles pour structurer la pensée des cliniciens (p. ex. la liste de descripteurs de Darley, Aronson et Brown, qui aide le clinicien à se concentrer sur les différents paramètres de la parole sans se perdre dans les multiples dimensions de celle-ci).

Les commentaires négatifs sont :

- les évaluations sont parfois difficiles à coter ;
- elles sont dépendantes de la tâche, ce qui pose le problème de leur généralisation ;
- elles ne sont pas assez sensibles, en particulier dans les parties intermédiaires du continuum d'évaluation, et ne sont donc pas idéales comme mesures d'évolution.

3. Mesures « objectives »

Au même titre que les mesures perceptives, les mesures dites « objectives » peuvent être employées à différents niveaux de granularité. Il existe des mesures au niveau segmental (p. ex., mesures de formants sur les voyelles), ainsi que des mesures qui s'appliquent sur des unités plus larges de parole (p. ex. débit de parole, prosodie).

Les mesures « objectives » les plus couramment utilisées sont les mesures de formants des voyelles, mentionnées par 9 experts : l'aire de l'espace vocalique, l'indice d'articulation vocalique (VAI), le ratio de centralisation des formants (FCR), les mesures formantiques isolées, les transitions de formants, la distinctivité des formants, la mesure de l'enveloppe convexe [MCCLOY 12; STORY 17] et l'espace vocalique acoustico-articulatoire.

Pour les consonnes, les experts mentionnent le délai d'établissement du voisement (« voice onset time », VOT), les moments spectraux, le bruit dans les fricatives, les mesures de nasalité (à l'aide d'antiformants ou d'appareils dédiés tels que le Nasometer), les équations de locus/la pente du second formant, le creux du troisième formant (« F3 dip »), et les mesures d'intensité sur le burst des plosives en position finale de mots.

Les mesures de voix/phonation sont également évoquées (fréquence fondamentale, niveau de pression sonore, résonance, temps maximum phonatoire), ainsi que les mesures de timing (de bas niveau, telle la durée des phonèmes, mais aussi de plus haut niveau tel le débit de parole). D'autres mesures incluent : la prosodie, la cinématique, les vecteurs spectraux et cepstraux (pour la reconnaissance automatique de la parole) ainsi que des mesures continues prises sur l'ensemble d'un enregistrement de parole (rapport hautes fréquences/basses fréquences, analyse du pattern syllabique, jalons acoustiques [« acoustic landmarks »]).

Neuf participants ont fourni un **niveau de satisfaction** précis avec ces mesures « objectives » sur une échelle de 1 (pas du tout satisfait) à 5 (tout à fait satisfait). La distribution des résultats est illustrée en [Figure 21](#). Parmi les autres experts, l'un a estimé que la satisfaction dépend de l'objectif : si, par exemple, l'on veut mesurer le triangle vocalique d'un individu, la satisfaction sera de 5/5, mais si l'on veut généraliser cette mesure, on sera confronté à une grande variabilité, qui constitue un défaut majeur des mesures objectives. Un autre participant a expliqué que pour la parole saine, la satisfaction avec les mesures objectives serait de 4/5 (reconnaissance automatique de la parole) ; cependant, lorsque l'on considère la « parole non standard », ce niveau tombe à 2/5 en raison d'un manque de corpus et d'intérêt général. Un dernier expert n'a pas pu trancher entre 3 et 4/5. Enfin, deux experts n'ont pas donné de note.

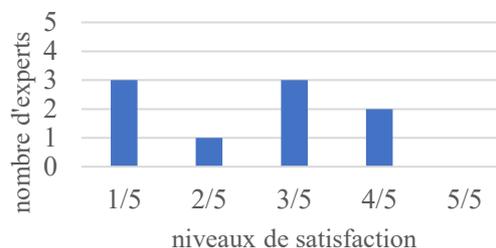


Figure 21 – Satisfaction avec les outils d'évaluation « objective » (N=9)

Parmi les principaux manques évoqués concernant les mesures « objectives » de la parole, nous retrouvons en premier lieu leur **manque de corrélation avec les scores perceptifs** (5/14). Ainsi, selon un participant, le manque de lien direct entre les mesures acoustiques, la perception et les dimensions articulatoires ne permet pas au clinicien de savoir, à partir de la mesure acoustique, comment modifier l'articulation du patient pour parvenir à la cible ou améliorer la qualité perçue de la parole. Cependant, ces mesures permettent des regroupements de phonèmes problématiques ; il revient ensuite au clinicien d'interpréter les résultats sur un plan articulatoire, sur base de ses connaissances acoustico-phonétiques et physiologiques.

Un autre expert explique que bien que les moyens de mesure de la parole s'améliorent et que nous avons peu à peu une meilleure compréhension des corrélats perceptifs et acoustiques de l'intelligibilité, l'éventail des mesures est aujourd'hui trop large. Il reste ainsi selon lui à identifier les indicateurs les plus sensibles de l'intelligibilité. Cette tâche reste ardue, car il en existe une multitude de corrélats acoustiques, qui dépendent du contexte pathologique. Quatre

participants précisent qu'une mesure unique de l'intelligibilité ne serait pas pertinente, mais qu'il faut plutôt envisager un *ensemble de mesures* qui rendraient compte des différents traits distinctifs des phonèmes, par exemple, et de la *multidimensionnalité de la parole*.

Un deuxième manque des mesures objectives, évoqué par quatre participants, est le *manque de généralisabilité et de reproductibilité*, qui est principalement lié à la grande variabilité de la parole, autant saine qu'intra- et interpathologique (p. ex. due aux phénomènes de compensation et d'adaptation). Deux experts évoquent la variabilité liée à la coarticulation, qui reste difficile à maîtriser.

Une troisième lacune, également évoquée par quatre participants, concerne la *difficulté d'utilisation* de ces outils par les cliniciens, en lien avec un *manque de formation* à leur utilisation. Conjointement, des problèmes plutôt liés aux aspects pratiques sont également évoqués par d'autres experts, tels le coût des logiciels et de l'équipement nécessaire (2/14) et la nature chronophage de certaines mesures (2/14) concomitante à une difficulté persistante à automatiser certaines d'entre elles (2/14).

En plus de cela, et similairement à ce qui a été observé pour les mesures perceptives, trois experts déplorent un *manque de standardisation des procédures* pour la mesure de la parole. Ce manque de standardisation concerne notamment la procédure d'enregistrement des stimuli et les paramètres pour l'extraction des mesures. Concernant les conditions d'enregistrement, deux experts mentionnent qu'en recherche, elles doivent être adaptées à l'objectif : pour une applicabilité clinique ultérieure du résultat, les enregistrements doivent être effectués dans des conditions « naturelles/réelles » (p. ex. utilisation du téléphone portable, enregistrement dans un bureau classique) ; si l'objectif est l'investigation de paramètres fins, des conditions plus contrôlées doivent être employées (p. ex. pièce insonorisée, utilisation d'un microphone de haute qualité). Concernant les paramètres d'enregistrement et d'extraction, l'un des experts précise que dans les logiciels commercialisés (p. ex. Nasomètre, Visipitch), bien souvent, ces paramètres ne sont pas contrôlables, ce qui rend parfois difficile l'interprétation des résultats. Un autre ajoute que les changements forcés de logiciels liés à la commercialisation (p. ex. arrêt de commercialisation par une compagnie) posent un problème pour la récolte de données normatives. En effet, dans la plupart des cas, des données acquises avec un logiciel donné ne peuvent pas être combinées avec celles récoltées avec d'autres logiciels, au vu des paramètres d'extraction et d'analyse pouvant différer et n'étant pas toujours personnalisables.

Trois experts soulignent aussi que de nombreuses mesures ne fonctionnent pas bien dans la parole pathologique et ne sont donc pas utilisables en clinique, notamment par *manque de spécificité et de sensibilité*. Les experts illustrent cela à l'aide d'exemples tels que l'impossibilité de calculer un VOT en cas d'occlusion incomplète dans l'articulation de plosives, ou l'impossibilité d'utiliser les mesures formantiques en cas d'hypernasalité. Un expert informaticien souligne que les difficultés de développement de mesures automatiques en parole pathologique sont entre autres dues à un manque important de corpus. Un autre participant, au vu du manque de sensibilité et de spécificité des mesures automatiques, regrette l'absence de scores de confiance. Selon lui, il faudrait intégrer notamment dans les outils de classification automatique, en plus des catégories spécifiques, une catégorie « on ne sait pas », une « zone floue/ambiguë ». Il argumente cela en avançant que les experts, même en évaluation

perceptive, ont également des difficultés à classer certains stimuli, et que cette catégorie doit en conséquence être reflétée dans les mesures automatiques.

Enfin, trois experts déplorent le *manque de mesures sur les consonnes*, deux autres le *manque de mesures spécifiques* pour les différentes dimensions de la parole (articulation, prosodie, nasalité...).

4. Mesures d'impact fonctionnel et des répercussions psychosociales

L'impact fonctionnel et les répercussions psychosociales sont de plus en plus pris en compte dans les domaines médicaux et paramédicaux. Pour l'évaluation de l'*impact fonctionnel* du trouble de la parole, un expert souligne l'importance de l'évaluation de la parole par des partenaires de communication ou par des auditeurs naïfs, par exemple en utilisant le *crowdsourcing*, au lieu de l'évaluation par les experts. En effet, cette dernière n'est pas représentative des partenaires de communication du patient au quotidien. Un autre expert mentionne l'évaluation de la parole dans le bruit, également plus représentative de l'utilisation fonctionnelle de la parole que les évaluations réalisées dans des conditions d'écoute « optimales ».

Pour l'évaluation de l'impact fonctionnel et celle des *répercussions psychosociales* (la frontière entre ces deux concepts semblant assez floue), la plupart des experts mentionnent les questionnaires (10/14). Ceux-ci peuvent être des questionnaires soit d'auto-évaluation, soit d'évaluation par les proches et les partenaires de communication du patient. Ils abordent principalement l'efficacité et les difficultés communicationnelles dans différents contextes et le degré de satisfaction et de frustration dans l'accomplissement d'objectifs liés à la communication. L'avis des proches et des partenaires de communication est un élément important dans l'évaluation fonctionnelle et écologique de la parole ; elle est mentionnée par cinq experts. L'un d'entre eux précise que chez l'enfant, l'évaluation de l'impact fonctionnel est très difficile et que les questionnaires complétés par les parents (p. ex. le *Intelligibility in Context Scale* [MCLEOD 12], également validé en français) sont alors indispensables. Parmi ces questionnaires, nous retrouvons d'un côté ceux liés à la voix (p. ex. le *Voice Handicap Index* [JACOBSON 97], le *Voice-Related Quality of Life* [HOGIKYAN 99] et le *Voice Activity and Participation Profile* [MA 01]), de l'autre les questionnaires spécifiques à la parole (p. ex. le *Dysarthria Impact Profile* [WALSHE 09], le *Communication Participation Item Bank* [BAYLOR 13], le *Speech Handicap Index* [RINKEL 08] et le *Phonation Handicap Index* [FICHAUX-BOURIN 09], tous deux validés en français [BALAGUER 19]).

Cependant, certains experts soulignent que ces questionnaires montrent des faiblesses importantes. L'un d'eux déplore le manque de constance des scores dans les études longitudinales, malgré une évolution linéaire de la sévérité de la dysarthrie. Le lien entre la sévérité de la dysarthrie et l'impact fonctionnel/psychosocial serait donc très « souple », influencé par de très nombreux facteurs (fatigue, humeur, interaction sociale positive...). De plus, un expert souligne que beaucoup de questionnaires employés sont plutôt informels, la plupart d'entre eux n'étant pas validés, parfois même non publiés. D'autres faiblesses concernent des outils spécifiques. Par exemple, l'un des participants estime la passation du *Dysarthria Impact Profile* difficile et son interprétation peu aisée, au vu notamment de nombreuses doubles négations. Le *Speech Handicap Index*, lui, est jugé peu pertinent, car il

s'agit principalement d'une translation mot-à-mot du Voice Handicap Index en remplaçant le terme « voix » par « parole », sans adapter les questions au contexte des troubles de la parole.

En plus des questionnaires, trois experts soulignent que les conversations informelles avec le patient ou ses pairs quant à l'impact du trouble de la parole sont également très informatives. Elles permettent d'informer le clinicien quant aux comportements communicationnels dans des situations spécifiques, à la satisfaction par rapport à l'efficacité communicationnelle, mais aussi quant aux adaptations éventuelles mises en place.

Enfin, un participant soulève une question qui persiste dans l'évaluation de l'impact psychosocial du trouble de la parole : quels sont finalement les aspects de la parole qui rendent une personne attractive/la parole naturelle aux oreilles du grand public ?

C. Discussion et Conclusion

Afin de compléter les données récoltées dans la littérature scientifique (cf. « 2.2.2. B. [Notre revue systématique PRISMA](#) ») et auprès des cliniciens (cf. « 2.3.1. [Évaluation des troubles de la parole chez l'adulte – situation actuelle et besoins en pratique clinique francophone](#) »), nous avons mené des interviews semi-dirigées auprès de chercheurs dans différents domaines applicatifs liés à la parole.

Encore une fois — tout comme dans la revue systématique et l'enquête aux cliniciens — les données de cette étude montrent une grande variabilité de la terminologie liée à la parole. Définir ces concepts s'est avéré être une tâche difficile — difficulté parfois explicitement exprimée par les experts interrogés — avec un manque de consensus apparaissant dans le panel.

Comme relevé dans l'enquête aux cliniciens, face à la subjectivité des mesures perceptives et au caractère chronophage de certaines d'entre elles, le besoin de mesures plus « objectives » et fiables est incontestable. Cependant, il semble aujourd'hui encore difficile d'identifier les mesures les plus pertinentes pour rendre compte de l'intelligibilité de la parole et des autres concepts. En effet, aucune d'entre elles ne permet d'expliquer une part importante de la variance de l'intelligibilité perçue, ce qui avait également été observé dans notre revue systématique de la littérature. Il est intéressant de noter que le manque de mesures acoustiques sur les consonnes, identifié dans notre revue, est aussi relevé par certains chercheurs interviewés. L'une des difficultés principales dans la recherche des indices acoustiques pertinents est la variabilité de la parole, qui se retrouve non seulement en parole saine, mais également à un plus haut degré encore dans la parole pathologique — autant à l'intérieur même d'un groupe diagnostique qu'entre les différentes pathologies. La variabilité existe et pose également problème au niveau des procédures d'enregistrement et d'extraction des mesures. En effet, les mesures acoustiques sont sensibles non seulement à l'échantillon de parole utilisé, mais aussi aux conditions matérielles et logicielles (qualité du micro, bruit ambiant, paramètres d'enregistrement, formats de fichiers, logiciel et paramètres d'analyse...). Cette sensibilité rend les mesures encore trop peu fiables à ce jour, au vu de l'absence de standardisation des procédures. Une difficulté dans la comparaison et la mise en commun des résultats des études scientifiques résulte de ces variabilités méthodologiques, ce que nous avons effectivement observé dans notre revue systématique.

En plus de ces aspects fondamentaux, qui rendent difficile la création même d'outils d'évaluation « objective » pertinents, valides et fiables, des freins existent aussi quant à leur applicabilité et leur utilisation en pratique clinique courante. Bien qu'un besoin soit clairement exprimé par les cliniciens, les aspects pratiques (coût, manque de temps, manque de maîtrise des outils informatiques/de connaissances en acoustique), déjà relevés dans notre enquête précédente, limitent encore le transfert des avancées scientifiques en outils cliniques. Pourtant, comme le conclut l'un des experts du panel, une utilisation des outils d'évaluation plus « objectifs » complémentaire aux outils d'évaluation perceptive subjectifs — sans prétendre à les remplacer — permettrait une évaluation plus exhaustive et fiable des troubles de la parole.

Enfin, un intérêt grandissant existe pour les mesures d'impact fonctionnel et psychosocial ou de qualité de vie liée à la parole. Un grand nombre de questionnaires existe à ce jour, dont certains ont été spécifiquement créés ou adaptés puis validés pour les troubles de la parole.

Comme cela avait aussi été relevé dans l'enquête aux cliniciens, un grand intérêt est également porté aux questionnaires remplis par l'entourage du patient, permettant de recueillir l'avis de ses partenaires de communication dans ses milieux de vie quotidiens. Une évaluation des aspects fonctionnels et psychosociaux en plus des mesures de la parole semble indispensable, au vu du manque de lien direct entre ces mesures. En effet, l'impact du trouble de la parole sur le quotidien du patient ne peut être prédit par les mesures strictement liées à la production de la parole ni par un raisonnement purement physiopathologique, car il est influencé par de nombreux facteurs externes. Il reste encore à ce jour à analyser ces facteurs et à mieux comprendre les interactions complexes qui existent entre le trouble de la parole et les facteurs psychosociaux et contextuels individuels à chaque patient.

2.4. Troisième pilier de l'EBP : l'avis des patients

Le troisième pilier de l'EBP concerne les valeurs et préférences du patient. Pour aborder celui-ci, nous avons souhaité connaître l'avis de patients à propos du processus d'évaluation clinique de leur parole (voir [Figure 22](#)) : son déroulement, les manques et besoins éventuels, ainsi que la prise en compte des répercussions du trouble dans la détermination de la stratégie thérapeutique. De fait, la diversité des pratiques cliniques et des outils utilisés rend nécessaire un état des lieux des méthodes actuelles, pour cibler au mieux les manques et besoins à adresser. Bien que l'avis et les valeurs du patient soient de plus en plus considérés dans la prise en charge, dans une visée de « soins centrés sur le patient » [STEWART 01; VRANCEANU 09], nous avons voulu mettre en avant ses valeurs et préférences du patient dans l'étape préliminaire à cette prise en charge. En effet, cette étape de bilan de parole est tout aussi importante, car elle déterminera la mise en place du plan thérapeutique subséquent ainsi que son efficacité. Nous avons donc développé et diffusé un questionnaire destiné à des patients adultes francophones souffrant d'un trouble de la parole, dans le cadre d'un stage d'initiation à la recherche proposé aux étudiants en orthophonie.



Figure 22 – Prise en compte du troisième pilier de l'EBP : questionnaire adressé aux patients

A. Méthodologie

Diffusion

Le questionnaire a été diffusé par deux étudiantes du Centre de Formation Universitaire en Orthophonie de Toulouse, entre janvier et octobre 2020 :

- directement à des orthophonistes de leur entourage (précédents stages, connaissances) ;
- en demandant à d'autres étudiantes de remettre des exemplaires à leurs maîtres de stage ;
- via le réseau social Facebook, où des groupes spécialisés en orthophonie ont été ciblés ;
- en passant par les syndicats régionaux de la Fédération Nationale des Orthophonistes et par des associations d'orthophonie ;
- en envoyant des demandes à différentes structures cliniques (centres hospitaliers, centres de rééducation et de réadaptation fonctionnelle).

Les patients ciblés étaient des patients atteints de troubles de la parole d'origine neurologique (dysarthrie, apraxie) ou organique (malformations de l'appareil bucco-pharyngé, séquelles de traitement carcinologique ORL radiothérapeutique et/ou chirurgical, insuffisance

vélaire...), ayant effectué un bilan orthophonique. L'effectif escompté était d'une trentaine de patients, dont 15 présentant un trouble d'origine neurologique et 15 un trouble d'origine organique. Les situations géographiques des patients devaient être multiples, avec des suivis effectués par des orthophonistes différents.

Dans un premier temps, les questionnaires ont été diffusés en fournissant aux orthophonistes des exemplaires papier ainsi que des enveloppes préimprimées. Par souci de confidentialité et pour éviter tout biais de réponse, il était demandé aux patients d'envoyer directement eux-mêmes le questionnaire complété par voie postale, ou de le rendre à leur orthophoniste dans l'enveloppe préalablement scellée.

À partir du confinement dû à la crise sanitaire provoquée par la COVID-19 et au vu de la patientèle à risque dans l'épidémie, la possibilité d'une complétion directe du questionnaire dans une version PDF adaptée a été mise en place. Les patients pouvaient alors directement téléverser le document dans un dossier sécurisé, sans accès aux autres fichiers téléversés dans ce même dossier par d'autres patients.

Questionnaire

Le questionnaire, anonyme, était composé de 27 questions réparties sur 4 parties :

1. Informations personnelles du patient : genre, âge, département, lieu de vie actuel (ville, banlieue, campagne) ;
2. Pathologie : étiologie/diagnostic, début ;
3. Trouble de parole : apparition premiers signes ;
4. Évaluation de sa parole (premier suivi et suivi actuel) : évaluateur, date du premier/dernier bilan, durée du bilan, niveau de fatigue, explications de l'orthophoniste, adaptation à la plainte, enregistrement de parole, auto-évaluation, prise en compte des conséquences du trouble...

Le questionnaire complet est disponible en [Annexe H](#). La durée de complétion estimée était de 15 minutes.

B. Résultats

Population

Onze participants ont répondu au questionnaire, dont 6 hommes et 5 femmes. La médiane d'âge était de 74 ans (EIQ : 19 ans ; min. : 59, max. 88). Les données descriptives de l'échantillon de patients sont disponibles dans le [Tableau 10](#).

Seul un participant a utilisé le questionnaire en sa version informatique.

Tableau 10 – Description des patients ayant répondu au questionnaire concernant l'évaluation de leur trouble de la parole

Département	N (%)
Haute-Garonne	4 (36 %)
Mayenne	3 (27 %)
Tarn-et-Garonne	1 (9 %)
Dordogne	1 (9 %)
Rhône	1 (9 %)
Alpes-Maritimes	1 (9 %)
<hr/>	
Pathologie	

Accident vasculaire cérébral	4 (36 %)
Cancer ORL	4 (36 %)
Parkinson	2 (18 %)
Atrophie multisystématisée	1 (9 %)
<hr/>	
Type d'exercice orthophoniste	
Libéral	8 (73 %)
Libéral et hôpital	2 (18 %)
Maison médicale	1 (9 %)
<hr/>	
Lieu d'exercice orthophoniste	
Ville	7 (64 %)
Banlieue	2 (18 %)
Campagne	2 (18 %)

Bilan de parole

Les données détaillées concernant le bilan de parole sont disponibles dans le [Tableau 11](#). Pour 9 patients, le dernier bilan de parole date de moins d'un an avant la complétion du questionnaire. La plupart d'entre eux (7/11) rapportent une durée de bilan entre 30 et 60 minutes. La fatigabilité est très variable à travers les pathologies. Par exemple, parmi les 4 patients évalués à la suite d'un AVC, deux rapportent une fatigue quasi nulle après le bilan, malgré une durée de plus d'une heure pour l'un d'entre eux, alors qu'un autre rapporte une fatigue très importante, pour une durée de bilan entre 30 et 60 minutes. Les quatre patients évalués à la suite d'un cancer de la sphère ORL rapportent tous une fatigue entre 1 et 2 sur une échelle de 0 à 3, malgré une durée de bilan inférieure à 30 minutes pour deux d'entre eux.

Concernant la prise en compte des conséquences du trouble de la parole par l'orthophoniste, le patient indiquant que celles-ci n'ont « pas du tout » été prises en compte a également indiqué une absence d'auto-évaluation de sa parole, d'enregistrement et de questions sur l'impact fonctionnel et psychosocial. Ce patient indique également qu'il aimerait une meilleure prise en compte des conséquences de son trouble. Le patient indiquant que les conséquences de son trouble avaient été « peu » prises en compte indique lui aussi une absence de questions sur l'impact fonctionnel et psychosocial et le souhait d'une meilleure prise en compte des conséquences de son trouble. Enfin, plus de la moitié (6/11) aimeraient que les conséquences de leur trouble de la parole soient davantage prises en compte lors du bilan.

Tableau 11 – Données concernant les bilans de parole

Date dernier bilan		N (%)
<6 mois		5 (45 %)
6 mois – 1 an		4 (36 %)
1 – 2 ans		2 (18 %)
Durée du bilan		
15-30 minutes		2 (18 %)
30-60 minutes		7 (64 %)
>1 heure		2 (18 %)
Fatigue après bilan (0 : pas du tout – 3 : beaucoup)		
0		4 (36 %)
1		4 (36 %)
2		2 (18 %)
3		1 (9 %)
Explications exercices réalisés (0 : pas du tout – 3 : très claires)		
2		2 (18 %)
3		9 (82 %)
Adaptation à la plainte		
Oui		11 (100 %)
Enregistrement de parole		
Oui		4 (36 %)
Non		7 (64 %)
Auto-évaluation parole		
Oui		5 (45 %)
Non		6 (55 %)
Prise en compte des conséquences fonctionnelles¹ du trouble (0 : pas du tout – 3 complètement)		
0		1 (9 %)
1		1 (9 %)
2		1 (9 %)
3		8 (73 %)
Questionnaire conséquences		
Oui		3 (27 %)
Non		8 (73 %)
Souhait meilleure prise en compte des conséquences du trouble (0 : pas du tout – 3 tout à fait)		
0		2 (18 %)
1		2 (18 %)
2		5 (45 %)
3		1 (9%)
NR		1 (9%)
Communication compréhensible des résultats		
Oui		9 (82 %)
Non		2 (18 %)

¹ « pour demander des choses aux gens, pour discuter avec eux, pour faire des activités de groupe... »

Note : NR = non renseigné

C. Discussion et Conclusion

Afin de prendre en compte l'opinion des patients concernant l'évaluation de la parole, nous les avons interrogés via un questionnaire créé à cet effet.

L'échantillon de patients ayant répondu à notre questionnaire est très réduit. En effet, seuls onze patients ont participé, ce qui est largement insuffisant pour tirer des conclusions générales, et encore plus pour des analyses différenciées (par exemple par diagnostic) et des croisements de données. Nous allons cependant ici retenir deux points qui ressortent majoritairement de ces données, puis aborder les raisons potentielles de l'échec du recrutement.

Tout d'abord, il s'avère que moins de la moitié des patients disent avoir été **enregistrés** lors de leur bilan de parole. Les données de l'enquête aux cliniciens, elles, avaient montré que seuls deux des 119 cliniciens n'effectuaient pas d'enregistrement de la parole lors du bilan. Cette différence peut être expliquée par le fait que l'enquête aux cliniciens était plus spécifiquement adressée aux professionnels prenant en charge les troubles de la parole, alors qu'aucun contrôle n'a été effectué quant à l'expertise dans ce domaine des orthophonistes ayant réalisé le bilan des patients dans la présente enquête. Ainsi, une hypothèse émise est que les cliniciens davantage familiarisés avec les troubles de la parole recourent plus facilement aux enregistrements de la parole (que ce soit pour des analyses acoustiques ou pour une réécoute ultérieure).

Le deuxième point observé est que peu de patients rapportent une **auto-évaluation** de leur propre parole et l'utilisation de **questionnaires sur les conséquences fonctionnelles et psychosociales** de leur trouble de la parole. Dans l'enquête aux cliniciens, près de 25 % d'entre eux disaient ne pas utiliser de questionnaire, contre presque 75 % dans les données ci-présentes. Encore une fois, l'hypothèse concernant le caractère plus spécialisé des cliniciens interrogés par rapport aux orthophonistes des patients dans cette deuxième étude est d'application.

Ainsi, cette étude nous permet de souligner l'importance de l'enregistrement d'échantillons de parole lors des bilans cliniques initiaux et de suivi, que ce soit pour l'évaluation perceptive ou acoustique. De plus, l'intégration dans le bilan de l'opinion du patient concernant les conséquences fonctionnelles et psychosociales de son trouble est un point fondamental de la démarche EBP. Le désir d'une meilleure prise en compte de ces conséquences exprimée par les patients souligne l'importance de l'utilisation de questionnaires d'impact fonctionnel et de qualité de vie, en plein essor dans la recherche scientifique actuelle.

2.5. Conclusion du chapitre : besoins en termes de mesure de la parole

Pour prendre en compte les trois principaux piliers de l'EBP dans l'investigation des besoins actuels en matière de mesure de la parole, nous avons mené quatre études, tel qu'illustré dans la Figure 23.

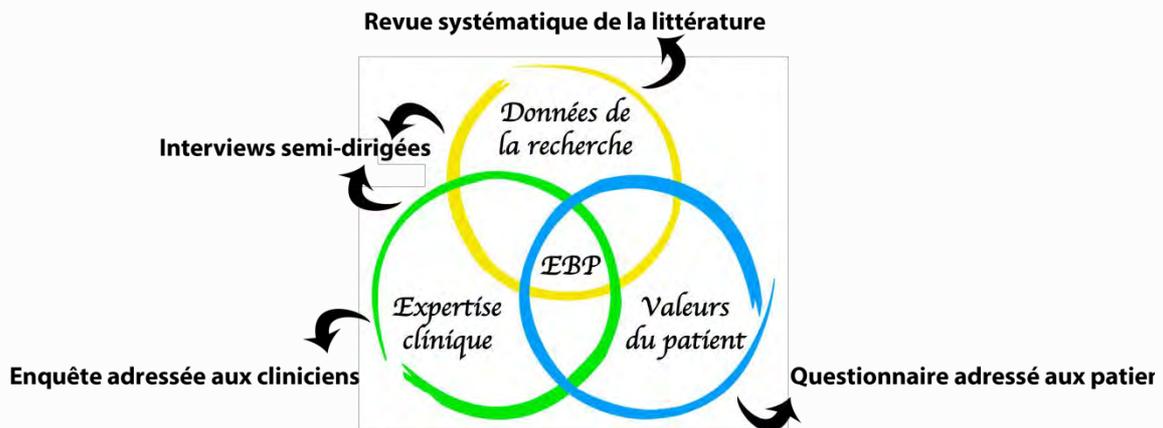


Figure 23 – Synthèse des études menées pour identifier les besoins en termes de mesure de l'intelligibilité selon le modèle EBP

Ces quatre études nous ont permis d'identifier diverses lacunes et besoins liés à la problématique de cette thèse. Le Tableau 12 rassemble tous ces éléments identifiés ainsi que les solutions suggérées et les contributions de cette thèse pour y répondre.

Tableau 12 – Besoins et manques identifiés à travers les quatre études menées concernant les mesures de la parole, solutions envisagées et contributions concrètes dans le cadre de cette thèse

Besoin ou manque	Source ¹	Solutions suggérées	Contribution
<ul style="list-style-type: none"> • Variabilité de la parole, saine et pathologique • Besoin de données normatives 	<ul style="list-style-type: none"> • RS • EC • IE 	<ul style="list-style-type: none"> • Mieux connaître le comportement des mesures de la parole saine (notamment données normatives), puis seulement en parole pathologique • Prendre en compte les contextes phonémiques et les facteurs suprasegmentaux 	<ul style="list-style-type: none"> • Investigation de mesures acoustiques en parole saine, avec analyse de variation en fonction du contexte phonémique (cf. Chapitre 4)
<ul style="list-style-type: none"> • Hétérogénéité et ambiguïté terminologique 	<ul style="list-style-type: none"> • RS • EC • IE 	<ul style="list-style-type: none"> • Définitions plus consensuelles • Définitions claires dans les rapports d'études scientifiques 	<ul style="list-style-type: none"> • Définition de l'intelligibilité et de la compréhensibilité (cf. Chapitre 1)
<ul style="list-style-type: none"> • Multitude de mesures perceptives et « objectives », manque de standardisation du bilan 	<ul style="list-style-type: none"> • RS • EC • IE 	<ul style="list-style-type: none"> • Standardisation du bilan concernant les tâches de production/consensus sur les objectifs et dimensions de l'évaluation de la parole • Instructions et directives claires et non ambiguës 	<ul style="list-style-type: none"> • Proposition d'un ensemble minimal de tâches d'évaluation des différentes dimensions de la parole (cf. Annexe G)
<ul style="list-style-type: none"> • Aucune mesure acoustique n'explique un % élevé de la variance de l'intelligibilité perçue • Manque de corrélation entre mesures acoustiques et perceptives 	<ul style="list-style-type: none"> • RS • IE 	<ul style="list-style-type: none"> • Identifier les indices les plus pertinents et les combinaisons les plus à même de rendre compte de l'intelligibilité perçue 	<ul style="list-style-type: none"> • Analyse du lien entre les mesures acoustiques investiguées et une évaluation perceptive (cf. Chapitre 4)
Variabilité méthodologique : <ul style="list-style-type: none"> • Variabilité des paramètres d'enregistrement et d'extraction des mesures acoustiques • Manque de prise en compte des facteurs âge et genre 	<ul style="list-style-type: none"> • RS • EC • IE 	<ul style="list-style-type: none"> • Standardiser les procédures d'enregistrement et d'extraction des mesures • Décrire plus en détail la méthodologie dans les rapports d'études • Investiguer les effets de l'âge et du genre sur l'intelligibilité de la parole 	<ul style="list-style-type: none"> • Suggestions pour l'enregistrement audio de la parole (cf. Annexe G)
<ul style="list-style-type: none"> • Manque global d'outils « objectifs » (applicabilité clinique limitée, manque d'exhaustivité des mesures proposées et de prise en compte de paramètres spécifiques de la parole, fatigabilité des patients) 	<ul style="list-style-type: none"> • EC • IE • QP 	<ul style="list-style-type: none"> • Cibler les paramètres spécifiques de la parole (prosodie, débit, nasalité, articulation...) • Combiner tâches analytiques (dont pseudomots) et tâches plus écologiques (dont lecture de texte) 	<ul style="list-style-type: none"> • Investigation de mesures au niveau segmental pour l'évaluation de l'intelligibilité (cf. Chapitre 4) • Proposition d'un nouveau texte de référence spécifiquement créé pour l'évaluation de la parole et de la voix (cf. Chapitre 3)

<ul style="list-style-type: none"> • Manque de sensibilité et de fiabilité des outils perceptifs 	<ul style="list-style-type: none"> • EC • IE 	<ul style="list-style-type: none"> • Mesures reproductibles et fiables (« objectives ») • Critères de cotation clairement définis • Base de données audio de référence • Données normatives • Formation plus approfondie à l'utilisation d'outils d'évaluation perceptive 	<ul style="list-style-type: none"> • Investigation de mesures acoustiques « objectives » en parole saine (données normatives) (cf. Chapitre 4)
<ul style="list-style-type: none"> • Aspect chronophage des outils perceptifs 	<ul style="list-style-type: none"> • EC • IE 	<ul style="list-style-type: none"> • Automatisation 	<ul style="list-style-type: none"> • Investigation de mesures acoustiques semi-automatiques (cf. Chapitre 4)
<ul style="list-style-type: none"> • Mesures acoustiques spectrales principalement limitées aux voyelles 	<ul style="list-style-type: none"> • RS • IE 	<ul style="list-style-type: none"> • Investiguer davantage les mesures sur les consonnes et les glides 	<ul style="list-style-type: none"> • Investigation de mesures consonantiques (cf. Chapitre 4)
<ul style="list-style-type: none"> • Manque de spécificité et de sensibilité des mesures « objectives », surtout en parole pathologique 	<ul style="list-style-type: none"> • EC • IE 	<ul style="list-style-type: none"> • Utilisation de corpus plus importants • Meilleur ciblage des mesures en fonction des symptômes spécifiques visés 	<ul style="list-style-type: none"> • Analyse du comportement des mesures acoustiques en parole pathologique (cancers ORL) (cf. Chapitre 4)
<ul style="list-style-type: none"> • Difficultés liées aux aspects pratiques des mesures « objectives » (chronophage, onéreux, difficile d'accès et d'utilisation) 	<ul style="list-style-type: none"> • EC • IE 	<ul style="list-style-type: none"> • Outil abordable, facile d'accès et d'utilisation, portable • Meilleure information quant aux outils existants • Formation à la mesure (acoustique) de la parole 	<ul style="list-style-type: none"> • Utilisation d'un alignement forcé automatique et de mesures acoustiques semi-automatiques, obtenues à l'aide d'un matériel abordable et dans des conditions similaires à celles observées en clinique courante (cf. Chapitre 4)
<ul style="list-style-type: none"> • Manque de prise en compte de l'impact fonctionnel et psychosocial spécifique à la parole 	<ul style="list-style-type: none"> • EC • IE • QP 	<ul style="list-style-type: none"> • Auto-évaluation par le patient • Utilisation de questionnaires adressés à l'entourage • Améliorer l'information et la diffusion d'outils spécifiques existants • Valider les outils dans des pathologies spécifiques • Mieux comprendre les facteurs médiateurs de l'impact du trouble de la parole 	<ul style="list-style-type: none"> • Collaboration avec Mathieu Balaguer, dont la thèse porte sur la mesure automatique de l'impact fonctionnel et psychosocial des troubles de la parole (cf. https://hal.archives-ouvertes.fr/hal-02505380)

¹RS = revue systématique, EC = enquête cliniciens, IE = interviews d'experts, QP = questionnaire patients

Chapitre 3. Supports d'évaluation et échantillons de parole

Jusqu'ici, nous avons abordé la définition de l'intelligibilité et de la compréhensibilité ainsi que leur mesure, et avons fait le point sur les besoins en termes de mesures de la parole en recherche scientifique et en pratique clinique. Nous allons maintenant nous pencher dans ce troisième chapitre sur les supports et échantillons de parole sur lesquels ces concepts peuvent être évalués.

3.1. Supports et échantillons – état de l'art en pratique francophone

Comme nous l'avons vu au [Chapitre 2](#), la parole peut être évaluée perceptivement ou instrumentalement sur des échantillons de différents niveaux de granularité (p. ex. mots isolés vs phrases), plus ou moins fonctionnels (p. ex. lecture à voix haute d'un texte vs échantillon de parole conversationnelle). Au vu de la grande diversité de tâches et d'outils d'évaluation évoquée dans les chapitres précédents, nous nous attarderons ici plus spécifiquement sur le contexte francophone.

3.1.1. Outils et supports disponibles en langue française

Il existe, pour l'évaluation de la parole, des tâches isolées, ainsi que des batteries d'évaluation standardisées composées de multiples tâches. Nous allons tout d'abord décrire les outils d'évaluation disponibles en langue française, avant de nous pencher sur les unités ciblées dans ceux-ci pour l'évaluation de l'intelligibilité.

A. Outils d'évaluation standardisés

Trois batteries sont actuellement disponibles en français : la BECD [AUZOU 06], la FDA-2 [GHIO 20] et le protocole MonPaGe [LAGANARO 21; LÉVÊQUE 16; TROUVILLE 21]. Rappelons avant de les décrire que, parmi les cliniciens que nous avons interrogés, environ un quart n'utilisent aucune batterie d'évaluation, et 40 % ne les emploient que partiellement. Or, l'utilisation partielle des batteries d'évaluation pose un problème de validité, étant donné que celles-ci sont validées non pas par tâche, mais sur la passation complète (prenant notamment en compte un ordre précis des tâches). L'utilisation « à la carte » de tâches extraites de ces batteries induit donc, entre autres, des problèmes dans l'interprétation des résultats vis-à-vis des normes établies.

BECD

La batterie d'évaluation standardisée la plus utilisée est la Batterie d'Évaluation Clinique de la Dysarthrie (BECD), ou sa version plus ancienne, l'Évaluation Clinique de la Dysarthrie (ECD) [AUZOU 98]. Cette batterie comporte six modules principaux :

1. Évaluation de la sévérité de la dysarthrie :
 - jugements perceptifs globaux (qualité vocale, prosodie, intelligibilité...);
 - score d'intelligibilité : compréhension de 10 mots et de 10 phrases porteuses et jugement de sévérité sur de la parole spontanée ;
 - Test Phonétique d'Intelligibilité : adaptation française du « Single Word Intelligibility test » de Kent [CROCHEMORE 01; GENTIL 92; KENT 89b], destiné

à identifier les altérations phonétiques. Liste de 52 mots bisyllabiques en choix multiple avec 13 séries de contrastes phonétiques ;

2. Analyse perceptive : jugements perceptifs sur une échelle de Likert de 35 critères et échelle d'évaluation perceptive de la qualité vocale GRBAS-I, sur conversation libre ou sur un extrait de « Pierrot » [DE MAUPASSANT 82] ;
3. Analyse phonétique :
 - répétition de 33 phonèmes isolés ;
 - répétition de 37 mots simples ;
 - répétition de 25 mots complexes ;
4. Examen moteur : grille d'évaluation motrice des organes impliqués dans la production de la parole ;
5. Auto-évaluation : Voice Handicap Index, remplacé par le Speech Handicap Index dans la nouvelle version de 2019 ;
6. Mesures acoustiques instrumentales : mesures d'intensité et de hauteur vocale.

La BECD, batterie pionnière dans l'évaluation des troubles de la parole, se base en grande partie sur des jugements perceptifs.

FDA-2

La version française du « Frenchay Dysarthria Assessment 2 » consiste en la lecture de 10 mots et de 10 phrases tirés aléatoirement parmi une liste de 101 mots et 51 phrases. Les auteurs ont pris en compte les améliorations du FDA-2 par rapport à sa première version en anglais, mais aussi les spécificités liées à la langue française, pour proposer un nouveau matériel d'évaluation. Ainsi, les 101 mots sont contrôlés au niveau de la fréquence lexicale, de la structure phonétique, des types grammaticaux et des voisins phonétiques. Les 51 phrases sont contrôlées au niveau de la structure morphosyntaxique, des modes et des temps, des distributions phonétiques, des fréquences lexicales des mots cibles et de la prédictibilité des phrases.

L'évaluation du FDA-2 se fait donc uniquement de manière perceptive.

MonPaGe

Le protocole MonPaGe est le protocole standardisé d'évaluation de la parole le plus récent. Publié en 2021, il comprend huit modules :

1. Intelligibilité : 15 mots cibles (avec voisins phonologiques), tirés aléatoirement parmi 437, à placer selon la consigne énoncée par le patient sur une grille de 25 formes colorées (p. ex. « mettez le mot “désert” dans le triangle rouge ») ;
2. Pneumophonatoire : /a/ tenu, temps maximum phonatoire et modulation de l'intensité vocale ;
3. Pseudomots : lecture/répétition de 50 pseudomots, évaluation perceptive de la précision articulatoire des consonnes et des voyelles du français ainsi que des clusters consonantiques ;
4. Diadococinésies : répétition de six séquences syllabiques aussi rapidement et aussi précisément que possible, avec mesure du nombre de syllabes sur 4 secondes ;
5. Phrases : lecture à voix haute de six phrases déclaratives et interrogatives, avec évaluation de la prosodie démarcative et distinctive par mesure de la fréquence fondamentale, du débit de parole et du contraste mélodique

6. Semaine : évaluation de la parole « automatique », récitation en boucle des jours de la semaine avec mesure du nombre de mots produits sur 20 secondes
7. Texte : lecture d'un texte de 188 mots permettant l'évaluation de la précision articulatoire, de la coarticulation, de la prosodie dans sa fonction expressive, du débit de parole et de l'effet de fatigue ;
8. Description ou Spontané : description d'une image ou réponse à deux questions.

Ce protocole utilise donc un ensemble standardisé de mesures perceptives et acoustiques. Cependant, les indices acoustiques sont encore limités aux mesures de la phonation et des voyelles.

Autres outils standardisés

D'autres épreuves standardisées existent, bien qu'elles soient moins employées.

Le « Test de diagnostic par paires minimales » [PECKELS 73], version française du « Diagnostic Rhyme Test » [VOIERS 83], est principalement utilisé en recherche en raison de sa longue durée [GHIO 16]. En effet, il comprend 432 mots présentés par paires, dont les consonnes initiales diffèrent par leurs traits phonétiques.

Le Test quantitatif d'intelligibilité [DITTNER 10] comprend quatre épreuves : une lecture des 12 mois de l'année (cotation phonétique), une lecture de 34 mots triphonémiques (cotation sémantique ; parmi 10 listes phonétiquement équilibrées), une lecture de 10 phrases et une description d'image.

Notons aussi qu'une épreuve dite de « décodage acoustico-phonétique » est en cours de validation : il s'agit d'un test d'intelligibilité constitué de pseudo-mots afin d'éviter le biais des processus de compensation « top-down » [GHIO 17, 18].

Enfin, pour évaluer l'impact fonctionnel du trouble de la parole, deux outils spécifiques existent en français, mais sont encore peu utilisés (par 11 % des cliniciens dans notre enquête, respectivement) : les versions françaises du Speech Handicap Index [DEGROOTE 11; RINKEL 08] et du Phonation Handicap Index (PHI) [BALAGUER 19; FICHAUX-BOURIN 09]. De plus, le Carcinologic Handicap Index (CHI) a été créé spécifiquement pour les patients atteints de cancer de la tête et du cou [BALAGUER 17; BALAGUER 20a]. Malgré l'existence de ces outils spécifiques, le questionnaire le plus couramment utilisé reste le Voice Handicap Index (70 %) [JACOBSON 97], dont le contenu cible les troubles de la voix. Enfin, près de 25 % des orthophonistes n'utilisent aucun questionnaire.

B. Unités ciblées pour l'évaluation de l'intelligibilité

Selon le panel d'experts interrogés dans le cadre de notre enquête Delphi, perceptivement, l'évaluation de l'intelligibilité s'effectue au mieux sur des stimuli à faible prédictibilité : les phonèmes, syllabes et pseudomots, mais aussi les mots (en paires minimales) et les phrases non prédictibles. Instrumentalement, l'intelligibilité peut être évaluée à l'aide de mesures acoustiques sur les consonnes, les voyelles et les glides, qu'il s'agisse de phonèmes isolés ou intégrés dans des syllabes, des (pseudo-)mots ou des phrases. Les paramètres suprasegmentaux (évalués objectivement par exemple par le débit de parole ou l'accentuation) contribuent également à l'intelligibilité, nécessitant donc des échantillons du niveau de la phrase.

Nous allons maintenant nous pencher sur les unités ciblées dans les batteries d'évaluation citées ci-avant, particulièrement pour l'évaluation de l'intelligibilité. Le [Tableau 13](#) synthétise les différentes unités employées dans les outils disponibles.

Tableau 13 – Unités ciblées dans les outils d'évaluation francophones

Batterie	Phonèmes/ syllabes isolé(e)s	Pseudo- mots	Mots	Phrases	Lecture de texte	Discours induit ou semi- spontané
BECD	X		X	X	X	X
FDA-2			X	X		
MonPaGe		X	X	X	X	X
Test de diagnostic par paires minimales			X			
Test quantitatif d'intelligibilité			X	X		X

Dans la BECD, trois épreuves visent l'évaluation de l'intelligibilité. Premièrement, l'épreuve « Score d'intelligibilité »¹⁸ se base sur une lecture de mots et de phrases et sur de la parole « spontanée » (*sic*). Les sous-épreuves sont intitulées « compréhension de mots » et « de phrases ». Cependant, elles emploient des mots avec des voisins phonologiques (p. ex. paquet/parquet,abri/appris), limitant leur prédictibilité et permettant ainsi une évaluation plus proche de l'intelligibilité que de la compréhensibilité. Deuxièmement, le « Test Phonétique d'Intelligibilité », destiné à identifier les altérations des contrastes phonétiques, emploie lui aussi des mots avec des paires minimales (p. ex. donna/donnant/tonna/tonnant). Cependant, chaque groupe de quatre mots voisins phonologiques est présenté quatre fois — une fois pour chaque mot. L'examineur peut donc identifier certains mots par déduction, sur base des réponses aux présentations précédentes du même groupe (p. ex. s'il entend « donna » au lieu de « tonna », mais avait déjà identifié « donna » auparavant). De plus, cinq versions de ces listes existent pour éviter les phénomènes d'habituation, mais celles-ci comprennent les mêmes items, dans un ordre différent. Comme le précisent les auteurs de la batterie, cette épreuve présente un effet plafond, les patients atteints de dysarthrie légère atteignant facilement la note maximale de 52/52. Troisièmement, avec son module « Analyse phonétique », la BECD est la seule batterie à utiliser une répétition de phonèmes isolés, permettant l'appréciation de l'articulation « pure » de ces derniers, sans effets de coarticulation. Il ne s'agit pas ici d'une épreuve d'identification de phonèmes (l'évaluateur sachant quel phonème doit être prononcé), mais plutôt d'une évaluation de la précision articulatoire. Certaines distinctions phonémiques ne sont cependant pas testées (p. ex. /œ̃/ vs /ɛ̃/ ; /œ/ vs /ø/), limitant l'utilisation de l'épreuve dans certaines

¹⁸ Notons à nouveau l'ambiguïté terminologique, observée également dans la recherche scientifique et dans l'usage clinique : l'intelligibilité est décrite par les auteurs comme « le degré de précision avec lequel le message est *compris* par l'auditeur » (en citant Yorkston et Beukelman, 1980), ou par « le nombre d'unités de parole reconnu par l'auditeur » ; la compréhensibilité est ensuite définie comme « le degré avec lequel un auditeur comprend la parole à partir du signal acoustique (intelligibilité) et des autres informations qui contribuent à la compréhension de ce qui vient d'être produit [...] telles que des connaissances sur le sujet traité, le contexte sémantique ou syntaxique, les gestes et d'autres indices ».)

régions francophones où ces contrastes sont utilisés — ce qui peut expliquer en partie pourquoi de nombreux orthophonistes n'utilisent pas la batterie dans son intégralité. De plus, dans la répétition de mots complexes, la complexité des items est croissante, ce qui peut amener un biais au vu de la fatigabilité des patients dysarthriques et de la longueur du bilan.

La version française de la FDA-2 emploie des mots et des phrases lus par le patient et transcrits par l'évaluateur. Bien que les auteurs aient diminué la prédictibilité des phrases en proposant chaque début de phrase deux fois avec une fin différente (p. ex. « Je ne veux pas changer » et « Je ne veux pas de thé »), les processus de compensation top-down ne sont pas entièrement court-circuités dans ces deux épreuves. En effet, la sémantique des phrases reste toujours plausible ; l'évaluation se rapproche ainsi ici davantage de la compréhensibilité.

Dans le protocole MonPaGe, deux modules ciblent l'évaluation de l'intelligibilité. Premièrement, le module « Intelligibilité » emploie des mots avec voisins phonologiques — tout comme la BECD et la FDA-2. Cependant, dans ce protocole, la prédictibilité est davantage réduite, car ces mots sont tirés parmi 437 mots qui ont tous entre un et six distracteurs dans la langue française. De plus, l'intelligibilité est évaluée dans une tâche communicationnelle, étant donné que les mots sont intégrés dans des consignes que le patient donne à l'évaluateur. Deuxièmement, le protocole MonPaGe est à ce jour le seul outil disponible qui propose également une tâche de pseudomots. Celle-ci permet l'analyse de la précision articulatoire des phonèmes du français — à l'image de l'épreuve « Analyse phonétique » de la BECD — mais dans un contexte plus « naturel » que la production isolée. De fait, ces phonèmes sont ici intégrés dans des séquences phonémiques respectant les règles phonotactiques du français. En outre, l'évaluateur, ne connaissant pas à l'avance le pseudomot prononcé, doit décoder celui-ci ; cette tâche permet donc véritablement une évaluation de l'intelligibilité.

C. Autres supports utilisés en clinique et en recherche francophone

D'autres supports non intégrés dans des outils standardisés sont fréquemment employés pour obtenir des échantillons de parole, en pratique clinique tout comme en recherche dans le domaine. Nous allons en lister quelques-uns, sans toutefois prétendre à dresser un inventaire exhaustif :

- listes de mots :
 - listes cochléaires de Lafon [LAFON 64] : 20 listes de 17 mots triphonémiques limitant l'intervention de la suppléance mentale, initialement créées pour l'audiométrie ;
 - listes de Fournier [FOURNIER 51] : 40 listes de 10 mots dissyllabiques, également initialement destinées à l'audiométrie ;
- listes de phrases :
 - les phrases de Combescure [COMBESCURE 81] : 20 listes de 10 phrases, phonétiquement équilibrées par liste ;
 - phrases de Lonchamp et de Vaissière : phrases qui incluent tous les phonèmes français : « *Au loin un gosse trouve, dans la belle nuit complice, une merveilleuse et fraîche jeune campagne.* » et « *Il faut déjà que vous sachiez que les bords de telles rues ne sont qu'un peu glissants le matin à Zermatt.* »

- textes :
 - La Bise et le Soleil : traduction du texte anglais « The North Wind and the Sun » [INTERNATIONAL PHONETIC ASSOCIATION 99] ;
 - La chèvre de Monsieur Seguin [DAUDET 69] ;
 - trois textes de Harmegnies [HARMEGNIES 88].

3.1.2. Unités phonémiques et textes de référence

Les outils d'évaluation disponibles utilisent donc des unités de production de la parole variées, allant du phonème au texte ou à la parole semi-spontanée sur l'échelle de granularité (voir [Figure 24](#)).

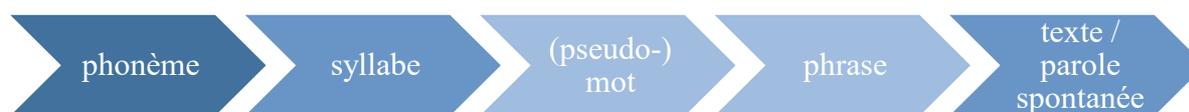


Figure 24 – Échelle de granularité des unités de production de la parole

Chaque type d'unité permet une évaluation répondant à un objectif spécifique, qui peut différer selon la nature de l'évaluation — perceptive ou instrumentale. Pour l'intelligibilité, bien que la cible principale soit le niveau phonémique (voire syllabique), les paramètres suprasegmentaux (p. ex. prosodiques) y contribuent également, nécessitant donc des échantillons du niveau de la phrase, voire du texte. Nous allons maintenant nous pencher plus spécifiquement sur les deux extrémités de l'échelle de granularité des unités d'analyse, toutes deux importantes dans l'évaluation de l'intelligibilité. En effet, l'une représente l'unité principale de l'évaluation analytique de l'intelligibilité, tandis que l'autre permet d'intégrer cette analyse dans un contexte plus naturel de production de parole.

A. Phonèmes

Les voyelles et les consonnes sont les unités principales ciblées lors de l'évaluation segmentale de l'intelligibilité. De nombreuses études montrent un traitement différentiel de ces deux catégories phonémiques au niveau neurolinguistique [CARAMAZZA 00; CARREIRAS 08a, 08b]. Se pose alors inévitablement la question de leur contribution respective à l'intelligibilité de la parole, autant en production qu'en perception de la parole.

Nous avons vu, au [Chapitre 2](#), que l'évaluation acoustique spectrale est aujourd'hui principalement investiguée sur les **voyelles** [HONDA 19; KAY 12]. Celles-ci jouent effectivement un rôle important dans l'intelligibilité de la parole [CHEN 13; COLE 96; KEWLEY-PORT 07]. L'analyse acoustique y est également plus aisée, étant donné leur nature voisée, leur composition d'ondes périodiques et la possibilité d'en effectuer une production tenue — par opposition aux consonnes plosives. Pourtant, plusieurs études montrent que les voyelles portent davantage d'informations suprasegmentales [FOGERTY 12; OWREN 06]. Elles contribueraient ainsi plutôt à la compréhensibilité — en permettant par exemple l'identification de structures syntaxiques [NESPOR 03] — ainsi qu'à la discrimination de l'identité du locuteur [OWREN 06].

À l'inverse, des études perceptives ont montré que chez le sujet atteint d'une surdité, la perte auditive des **consonnes** affecte davantage la perception de la parole que celle des voyelles

[SCHOW 18]. Les consonnes contribueraient davantage à l'intelligibilité au niveau lexical, autant chez le sujet normo-entendant que chez le sujet atteint d'une surdité [BONATTI 05; FOGERTY 10, 12; NESPOR 03; OWREN 06]. Le rôle distinctif plus important des consonnes au niveau lexical semble de plus indépendant du répertoire phonémique des langues étudiées [CUTLER 00]. Plusieurs études ont ainsi observé une préférence « universelle » pour les modifications vocaliques, lorsque des participants devaient transformer des pseudo-mots en mots en changeant un seul phonème (p. ex. « cobra » pour « kebra », bien plus rarement « zebra »). Cette tendance a été observée tant dans les langues avec un répertoire phonémique équilibré (p. ex. le néerlandais) que dans les langues moins équilibrées (p. ex. l'espagnol, plus pauvre en voyelles) [CUTLER 00, 02; MARKS 02; VAN OOIJEN 96]. Les auteurs avancent l'hypothèse que cette préférence de modification vocalique « universelle » est due au fait que, dans toutes les langues (*sic*), changer une consonne change plus facilement un mot en un autre mot, par rapport aux voyelles [CUTLER 00].

Les consonnes jouent donc elles aussi un rôle non négligeable dans l'intelligibilité de la parole. De plus, les réductions consonantiques en parole spontanée sont aussi fréquentes que les réductions vocaliques [VAN SON 99]. C'est pourquoi, dans ce travail, nous avons souhaité nous concentrer davantage sur l'étude acoustique spectrale des consonnes afin de répondre au manque de données qui se fait d'autant plus ressentir dans les études francophones.

Au-delà du phonème ciblé, il convient aussi de prendre en compte son contexte phonémique. En effet, il est largement établi que la production (ergo, l'acoustique) et la perception des phonèmes sont influencées par leur contexte, du fait des phénomènes de coarticulation [LIBERMAN 67; SUOMI 85]. La **coarticulation** est définie comme le chevauchement sur l'axe temporel des mouvements accomplis par les articulateurs et donc des caractéristiques phonétiques et acoustiques des sons de la parole [NGUYEN 01].

L'influence du contexte vocalique sur l'acoustique des consonnes a par exemple été montrée dans une étude retenue dans notre revue systématique de la littérature. [KATZ 91] ont en effet observé que l'énergie spectrale se situait plus haut dans les fréquences pour le /s/ dans /si/ que dans /su/. Ceci reflète l'influence de la coarticulation labiale (effet de l'arrondissement de la voyelle) et linguale (effet de l'antériorité de la voyelle), également observée dans de nombreuses autres études (p. ex. [BLUMSTEIN 79; LIBERMAN 52; MANN 80; SUOMI 85; WHALEN 81]).

Inversement, le contexte consonantique a lui aussi une influence importante sur la perception, la production et l'acoustique des voyelles [FLEMMING 09; HILLENBRAND 01], notamment sur le second formant [MAUROVA PAILLEREAU 16; STEVENS 63]. Ainsi, à titre d'exemple, un contexte consonantique alvéolaire a généralement un effet centralisant sur les voyelles [GENDROT 05], tout comme les contextes labiaux qui abaissent le second formant des voyelles antérieures non-arrondies /i, e, ε/, tandis qu'ils élèvent le second formant des voyelles postérieures et labialisées /u, o, ɔ/ [MAUROVA PAILLEREAU 16].

Cette variabilité des voyelles et des consonnes due à leur contexte phonémique s'ajoute ainsi aux nombreuses autres sources de variabilité intralocuteur (telle que la fatigue) et à la variabilité inter-locuteurs de production de la parole. Elle souligne l'importance du contrôle de ce contexte pour la validité des mesures acoustiques. C'est pourquoi nos investigations

acoustiques décrites dans le [Chapitre 4](#) se sont, dans un premier temps, déroulées dans des contextes phonémiques contrôlés. Cependant, pour une évaluation plus fonctionnelle, ces mesures doivent également être appliquées à des échantillons de parole plus naturels. À cet effet, les textes de référence présentent un intérêt majeur, car ils permettent d'étudier les fonctions des composants de la parole et de la voix intégrés dans un contexte de parole continue (plus proche de la parole spontanée que les mots ou les phrases isolés), avec des stimuli cibles dans un environnement contrôlé [R. PATEL 13]. Nous allons maintenant nous pencher sur les principaux textes de référence employés pour l'évaluation de la voix et de la parole et allons décrire leurs critères de construction.

B. Textes de référence

Tant en pratique clinique qu'en recherche, les textes de référence permettent d'obtenir un aperçu rapide des caractéristiques de la parole du patient [AUZOU 06], avec moins d'hésitations que dans la parole conversationnelle [VASILESCU 04] et une meilleure prédictibilité par l'évaluateur. Cependant, bien qu'il existe une pléthore de textes dans différentes langues, aucun d'entre eux ne semble vraiment répondre aux attentes et aux besoins des cliniciens et des chercheurs pour l'évaluation courante de la parole et de la voix. Certains ont été créés dans un but spécifique, tel le Zoo Passage en anglais [S. G. FLETCHER 72] pour l'évaluation de la fermeture vélopharyngée, ou les textes de Kuo et Weismer [KUO 16] pour l'évaluation de la réduction vocalique. Pour la plupart des autres, peu d'informations sont disponibles sur la façon dont ils ont été créés. Certains sont de simples extraits d'œuvres littéraires (par exemple « La chèvre de Monsieur Seguin » d'Alphonse Daudet [DAUDET 69] en français), ou des traductions de textes sans adaptation à la langue cible [JESUS 15]. Pour l'évaluation de la voix et de la parole en français, nous avons évoqué cinq textes employés en pratique clinique et en recherche :

- « La chèvre de Monsieur Seguin » [DAUDET 69] est un texte souvent considéré à tort comme phonétiquement équilibré. En effet, il s'agit là d'une nouvelle écrite par l'écrivain Alphonse Daudet, sans aucune visée d'utilisation clinique ou scientifique.
- C'est aussi le cas pour l'extrait du conte « Pierrot » de Guy de Maupassant, employé dans la BECD.
- « La Bise et le Soleil » est également un texte souvent dit phonétiquement équilibré. Ce texte a été traduit de la fable d'Ésope, d'abord en anglais par l'International Phonetic Association (« The North Wind and the Sun » [INTERNATIONAL PHONETIC ASSOCIATION 99]). Cependant, déjà dans cette première traduction, certains phonèmes de l'anglais étaient manquants et des défauts d'équilibrage phonétique étaient notés [DETERDING 06]. On ne peut donc pas s'attendre à un équilibre phonétique dans les nombreuses traductions subséquentes.
- Les trois textes de Harmegnies [HARMEGNIES 88] ont, pour leur part, été créés dans le cadre d'un travail sur l'évaluation de la parole. Ils comprennent 277 phonèmes et sont phonétiquement équilibrés. Cependant, l'emploi de mots rares et de structures syntaxiques peu communes induit des perturbations dans la fluence de la lecture et dans la prosodie.
- Enfin, le texte de 188 mots du protocole MonPaGe [LAGANARO 21] (cf. [Annexe K](#)) a été spécifiquement créé pour l'évaluation de la parole dysarthrique ou apraxique et validé sur

des locuteurs francophones de Belgique, France, Québec et Suisse. C'est le seul texte francophone à ce jour pour lequel les critères de construction ont été détaillés par les auteurs ; il a été publié en 2021. Ceux-ci seront décrits plus en détail dans le sous-chapitre suivant (cf. « 3.2.2. A. [Recensement](#) »).

Nous avons ainsi constaté qu'il n'existait aucun texte en français créé spécifiquement pour répondre aux besoins des cliniciens et chercheurs en matière d'évaluation de la parole et dont les critères de construction sont documentés (le texte du protocole MonPaGe n'a été publié qu'en 2021). C'est pourquoi, dans le cadre de ce travail, nous avons souhaité répondre à ce manque en proposant un nouveau texte de référence, spécifiquement conçu à cet effet et adapté aux spécificités culturelles et linguistiques. À cette fin, un groupe de travail francophone international et pluridisciplinaire a été mis en place en octobre 2019. Ce groupe intégrait dix experts, dont cinq français, trois belges et deux canadiens. Deux membres étaient médecins ORL, quatre linguistes et quatre orthophonistes/logopèdes (cf. [Annexe I](#) pour la liste des membres de ce groupe).

Lors de la création de ce groupe, les auteurs du texte du protocole MonPaGe qui n'était alors pas encore publié — dont Véronique Delvaux, et Cécile Fougeron, membres de notre groupe — nous ont alors informés de l'existence de ce dernier. Ses critères de construction nous ont été communiqués, et l'autorisation nous a été octroyée de l'utiliser comme base pour la création d'un nouveau texte. En effet, bien que de nombreux critères avaient été pris en compte dans sa construction (cf. « 3.2.2. A. [Recensement](#) »), ceux-ci n'étaient pas basés sur un état des lieux des besoins des cliniciens et des chercheurs. Nous allons maintenant décrire le travail effectué par notre groupe de travail pour identifier ces besoins et construire un nouveau texte compte tenu de ces derniers.

3.2. Nouveau texte de référence

La construction du nouveau texte s'est faite en trois grandes étapes résumées en [Figure 25](#), que nous allons maintenant décrire :

1. Recensement des besoins des cliniciens et chercheurs (mesures à appliquer sur le futur texte, population cible) et définition de l'objectif du texte ;
2. Recensement de critères de construction, sélection et hiérarchisation ;
3. Construction du nouveau texte.

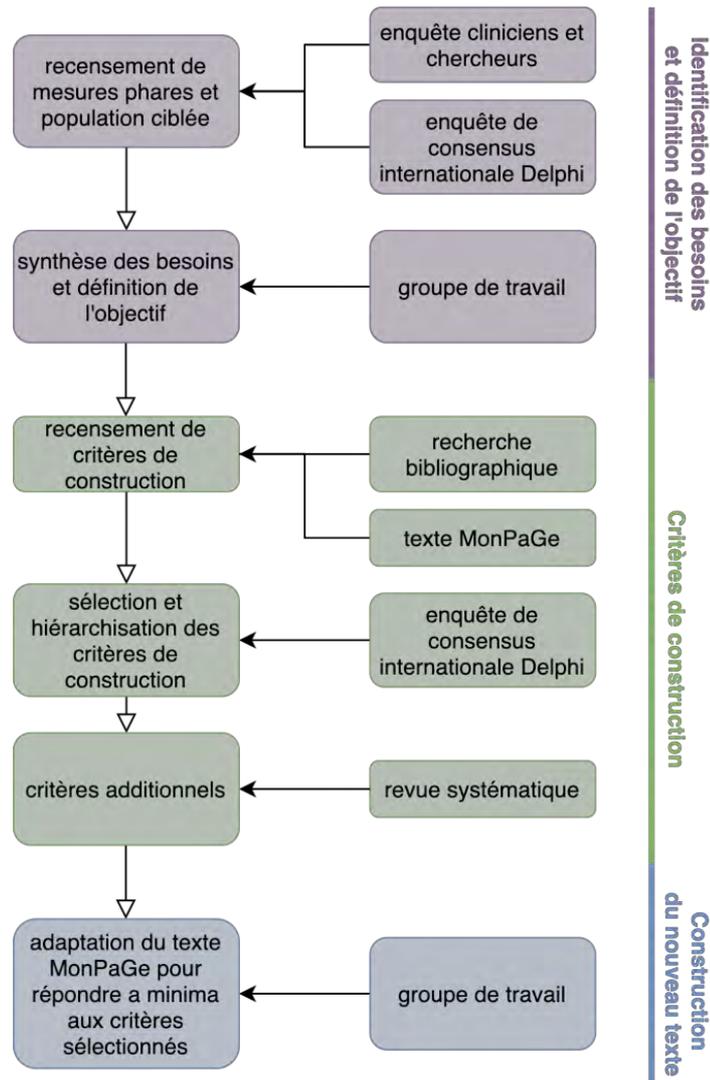


Figure 25 – Processus de création du nouveau texte pour l'évaluation de la parole et de la voix

3.2.1. Identification des besoins et définition de l'objectif

Premièrement, une brève enquête (via LimeSurvey) a été envoyée à 40 cliniciens ayant répondu à notre enquête décrite au [Chapitre 2](#) et ayant accepté d'être recontactés. Cette enquête a également été diffusée, via les membres du groupe, à des chercheurs concernés (phonéticiens, linguistes, spécialistes de la prosodie...). L'objectif était de recenser les trois mesures phares que ces professionnels souhaiteraient pouvoir appliquer à un texte dans le cadre de l'évaluation de la parole et de la voix, ainsi que les domaines d'intervention/d'étude ciblés (p. ex. voix, parole, bégaiement) et la population concernée (enfants, adolescents, adultes).

Treize personnes ont répondu à l'enquête (12 orthophonistes, 1 orthophoniste enseignante et chercheuse). Tout d'abord, les résultats de ce recensement ont montré que le terme « mesure » peut être interprété de manière très large. En effet, les réponses ne correspondaient que dans quelques cas à une « mesure » quantifiable et concrète (p. ex. aire de l'espace vocalique) ; plus souvent, il s'agissait des concepts à évaluer (p. ex. intelligibilité, variations du débit...), ou même parfois de caractéristiques du texte à proprement parler (p. ex. voyelles et groupes consonantiques du français dans différents contextes ; fréquence, régularité et longueur des mots...). Les réponses, classées en 6 catégories non exclusives, étaient :

- articulation (15/39¹⁹) : intelligibilité globale et segmentale, précision phonétique ; structures consonantiques simples à complexes ; erreurs articulatoires ; analyse des consonnes, voyelles et groupes consonantiques du français dans différents contextes ; qualité de l'articulation (comparaison avec le spontané) ; degré de nasalité ; résonance ; coarticulation ;
- prosodie (11/39) : variations du débit ; rythme ; débit/vitesse ; modalité, focus ; contours intonatifs/variations de la hauteur ; vitesse articulatoire, accentuations primaires et secondaires, pauses ;
- phonation (5/39) : intensité ; F0 ; raucité ; tenue phonatoire ; qualité vocale ;
- fluence (4/39) : % de blocages ; % d'allongements ; % de répétitions ;
- lecture (3/39) : temps de lecture, qualité de la transposition graphophonétique, rapidité/précision ;
- langage (1/39) : contrôle des mots (fréquence, régularité, longueur).

Nous observons donc un intérêt des cliniciens pour des mesures principalement du domaine de l'articulation et de la prosodie, mais aussi de la phonation et de la fluence.

Concernant les domaines d'application de ces différentes catégories de « mesures », les résultats montrent une prépondérance de la dysarthrie (25/39²⁰) et du bégaiement/bredouillement (16/39), ainsi que de la dysphonie (10/39). Les autres domaines incluent les troubles développementaux (5/39), le transgendérisme (3/39), l'aphasie/les troubles du langage chez l'adulte (3/39) et la dyslexie/dysorthographe (1/39).

Enfin, en ce qui concerne la population cible, un intérêt majeur est accordé aux adultes (29/39), mais la population des enfants et adolescents n'est pas négligeable (17/39).

Deuxièmement, notre enquête de consensus Delphi a également permis le recensement de mesures de parole et de voix souhaitées sur un nouveau texte, selon 40 experts internationaux (cf. [Chapitre 1](#)). Les participants ont énuméré les mesures ou évaluations qu'ils souhaitaient pouvoir utiliser sur un texte standardisé sous forme de réponses ouvertes (minimum 3). Deux évaluateurs (les mêmes que précédemment), en aveugle par rapport à l'identité des participants, ont procédé à une analyse de contenu de ces réponses ouvertes afin d'identifier des groupements sémantiques :

- vingt-deux des 40 experts souhaitaient appliquer des mesures suprasegmentales ;
- vingt-et-un désiraient appliquer des mesures de qualité vocale ;
- dix-sept souhaitaient pouvoir appliquer des scores de pourcentage d'identification correcte, dont 11 des mesures au niveau du phonème ;
- dix-sept étaient intéressés par l'application de mesures d'articulation/de précision phonétique, dont dix en utilisant des mesures acoustiques ;
- neuf voulaient mesurer l'intelligibilité (sans préciser comment/à quel niveau de granularité) ;
- sept souhaitaient appliquer des évaluations subjectives globales liées à la parole.

¹⁹ Chacun des 13 participants a renseigné 3 mesures.

²⁰ Certaines mesures peuvent s'appliquer à plus d'un domaine ; la somme totale de ces résultats n'égale donc pas 39.

Le groupe s'est alors réuni pour synthétiser les résultats en termes de besoins et pour définir ensemble un objectif. Nous avons décidé de cibler les personnes âgées de minimum 12 ans, pour éviter les thématiques trop infantilisantes pour l'adulte, mais aussi et surtout pour cibler uniquement des individus dont la fluidité de lecture est suffisamment développée.

Ainsi, le texte résultant de ce travail est destiné à fournir un support standardisé pour l'évaluation, chez les patients âgés d'au moins 12 ans :

- de l'articulation des sons de la parole (dysarthrie, apraxie) ;
- des variations prosodiques et du comportement phonatoire (dysphonie, féminisation vocale) ;
- de la fluence/des disfluences (bégaiement/bredouillement).

Il n'est pas destiné à l'évaluation du niveau de lecture (rapidité/précision du décodage graphophonémique ; français langue étrangère), ni à l'évaluation langagière (p. ex. aphasie).

3.2.2. Recensement, sélection et hiérarchisation des critères de construction

Une fois l'objectif du texte fixé, quatre sources de données ont permis la sélection et la hiérarchisation des critères de construction de celui-ci :

- une recherche bibliographique d'articles décrivant explicitement la création de textes pour l'évaluation de la parole ;
- la collaboration avec les auteurs du protocole MonPaGe, qui a été publié un an et demi après la création de ce groupe de travail et qui inclue un texte à lire à voix haute spécifiquement créé pour l'évaluation de la parole ;
- notre enquête de consensus internationale Delphi, où les participants ont sélectionné et classé les critères à prendre en compte pour la création d'un nouveau texte ;
- notre revue systématique de la littérature, pour l'identification des mesures acoustiques corrélées à l'évaluation perceptive de l'intelligibilité.

A. Recensement

Une recherche bibliographique a permis de dresser une liste non exhaustive de listes de phrases et de textes communément utilisés en anglais, en français et en néerlandais²¹ (cf. [Annexe J](#)). Nous avons identifié quatre articles décrivant plus en détail la création de textes. Ces articles ont été utilisés comme source principale pour établir une première liste de critères :

- [R. PATEL 13] : leur objectif était de fournir un nouveau texte anglais mieux adapté à l'évaluation des troubles de la parole et aux diagnostics différentiels par rapport au texte « My Grandfather » [VAN RIPER 63] ;
- [MARTENS 10a] : ces auteurs ont élaboré un texte flamand dans le cadre d'un nouveau protocole d'évaluation des compétences prosodiques de patients dysarthriques ;
- [T. POWELL 06] : dans cette étude, les caractéristiques phonétiques, lexicales et structurelles de 15 textes en anglais ont été analysées et comparées ;
- [KUO 16] : leur objectif était d'étudier les variations intra- et interindividuelles de l'articulation des voyelles chez des locuteurs sains de l'anglais américain.

²¹J'étais à ce moment-là en déplacement à l'Université de Radboud (Nimègues, Pays-Bas).

Les critères identifiés dans ces articles ont ensuite été enrichis par un échange avec les auteurs du texte du protocole MonPaGe [LAGANARO 21], qui avaient rédigé un texte [FOUGERON 19; PERNON 20] :

- court (moins de 200 mots) ;
- incluant des mots également présentés de manière isolée dans le protocole (p. ex. les jours de la semaine) ;
- comprenant des glides ;
- avec des répétitions de mots ;
- permettant l'évaluation de la coarticulation voyelle-à-voyelle (« Papy » vs « Papa ») ;
- qui inclut les voyelles cardinales /a,i,u/ dans des mots monosyllabiques consonne-voyelle (CV), CVC et des mots bisyllabiques CVCV ;
- avec une même séquence en début et en fin de texte (fatigabilité) ;
- incluant des dialogues et des variations mélodiques ;
- permettant l'évaluation de la nasalité sur des segments à prédominance nasale ;
- comprenant des structures syllabiques complexes enchaînées.

Sur base de cette collaboration, ainsi que sur base de la recherche documentaire, la liste suivante de 25 critères potentiels à prendre en compte lors de la création d'un nouveau texte a été établie. Les critères qui seront sélectionnés parmi ceux-ci seront décrits plus en détail ci-après (cf. 3.2.2. C. « [Critères principaux et pertinence](#) »).

Critères au niveau du phonème

1. Inventaire phonémique complet (le texte comprend tous les phonèmes de la langue cible au moins une fois) ;
2. Équilibre phonétique (les phonèmes inclus ont approximativement la même fréquence d'occurrence que dans une conversation ordinaire dans la langue cible) ;
3. Présence de clusters (bi- et triconsonantiques) ;
4. Prise en compte des positions dans le mot (début/milieu/fin) et du contexte phonémique ;
5. Inclusion des voyelles cardinales à plusieurs reprises ;
6. Prise en compte de la coarticulation voyelle-à-voyelle (p. ex. papy – papa) ;
7. Présence d'occlusives sourdes initiales et de clusters, surtout de séquences phonème non voisé/phonème voisé (bégaiement) ;
8. Présence de mots avec des glides (dynamique articulatoire linguale, dysarthrie) et de mots induisant des mouvements linguaux d'un bout à l'autre du triangle vocalique (p. ex. kayak, caillou, kiwi) ;

Critères au niveau lexical

9. Ratio de mots de fonction/de contenu (troubles de la fluence) ;
10. Contrôle de la complexité articulatoire/phonétique des mots ;
11. Répétition de mots à plusieurs reprises (cohérence des erreurs, apraxie vs dysarthrie) ;
12. Présence de paires de mots de longueur croissante (p. ex. amuse/amusement) ;
13. Contrôle de la fréquence lexicale ;

14. Couplage avec items de tâches d'évaluation isolées pour comparer la production de parole en condition isolée vs continue (p. ex. inclure des items de listes de mots isolés des batteries d'évaluation existantes) ;
15. Contrôle du voisinage phonologique des mots ;

Critères au niveau phrastique

16. Présence de divers patterns intonatifs (discours rapporté, interrogatives, déclaratives, accentuation contrastive...);
17. Inclusion d'une phrase à prédominance nasale vs orale (insuffisance vélaire, fente palatine...);
18. Présence d'une phrase 100 % voisée ;
19. Répétition d'une séquence identique en début et en fin de texte (effets de fatigue) ;
20. Présence de phrases à durée variable (soutien respiratoire, groupes de respiration, fatigabilité) ;
21. Contrôle de la longueur moyenne des énoncés ;
22. Contrôle de la complexité syntaxique ;

Critères généraux

23. Calcul de l'index de lisibilité (p. ex. Flesch-Kincaid) ;
24. Contrôle de la thématique (p. ex. thématique contemporaine, contenu émotionnellement aussi neutre que possible, « banal », pour éviter la charge émotionnelle, mais aussi la mémorisation) ;
25. Contrôle de la longueur globale du texte.

B. Sélection et hiérarchisation

Afin de sélectionner les critères les plus importants de cette liste et de les hiérarchiser, nous avons demandé aux participants de l'étude internationale Delphi (cf. [Chapitre 1](#)), dans le **premier tour**, de sélectionner et de classer, parmi la liste de 25 critères décrits ci-avant, ceux qui leur paraissaient essentiels à contrôler pour la création d'un nouveau texte. Tout critère non sélectionné était considéré comme non essentiel ; des critères additionnels pouvaient être proposés par les participants, en spécifiant la langue cible. Les réponses et les commentaires de ce premier tour ont été analysés indépendamment et de manière anonymisée par deux évaluateurs (les mêmes que précédemment).

Trois mesures ont été effectuées pour chaque critère : le rang moyen (RM), la fréquence à laquelle un critère a été noté (et a donc été considéré par le participant comme essentiel) et le rang moyen lorsqu'il a été noté (RMN, excluant donc les non-notations).

Pour les RM, les critères non classés se sont vu attribuer la valeur la plus basse de 26 (pour 25 critères disponibles). Cette mesure est un indicateur composite, qui prend en compte non seulement le nombre de fois que chaque critère a été évalué, mais aussi le rang qui lui a été attribué à chaque évaluation.

Les fréquences de notation et les RMN permettent, pour leur part, des nuances d'informations différentes : alors que le RMN nous informe sur l'importance relative de chaque critère pour le groupe d'experts qui l'a sélectionné, la fréquence de classement nous informe quant à elle sur la proportion d'experts qui ont jugé un critère comme pertinent, sans toutefois

prendre en compte l'importance relative qui lui est attribuée. Les seuils pour qu'un critère soit considéré comme essentiel pour le groupe d'experts ont été fixés *a priori* : un critère devait être classé par au moins la moitié des participants (c.-à-d. que la majorité des experts le trouvent important), et son RMN devait être inférieur à 7,5/25 (33 %) (c.-à-d. qu'il devait être classé important par ceux qui l'ont sélectionné).

La [Figure 26](#) reprend l'ensemble des critères et leur classement moyen. Les huit critères les plus et les mieux classés sont repris dans la liste suivante. Les résultats sont présentés sous forme de rangs moyens (RM, sur 25) avec, entre parenthèses, la proportion de participants ayant classé le critère et le rang moyen excluant les non-notations (RMN) :

- inventaire phonémique complet : 9,5 (68 % ; RMN : 1,6) ;
- équilibre phonétique : 11,6 (63 % ; RMN : 3) ;
- inclusion de divers patterns d'intonation : 11,9 (73 % ; RMN : 6,6) ;
- complexité articulatoire/phonétique des mots : 12,6 (68 % ; RMN : 6,1) ;
- prise en compte de la position des phonèmes dans les mots, ainsi que de leur contexte phonémique : 15,6 (50 % ; RMN : 5,3) ;
- inclusion de clusters consonantiques : 15,9 (50 % ; RMN : 5,8) ;
- présence répétée de certains mots : 16,4 (50 % ; RMN : 6,7) ;
- inclusion de différentes longueurs de phrases : 16,6 (50 % ; RMN : 7,3).

Les deux critères les moins bien classés présentent également les plus faibles occurrences de sélection :

- complexité syntaxique : 23,5 (20 % ; RMN : 13,3) ;
- présence d'une phrase 100 % voisée : 23,3 (18 % ; RMN : 10,6).

Deux autres critères ont été mis en évidence, à la lumière des résultats clivants entre le nombre d'occurrences et le RMN :

- La « longueur globale du texte » a été classée par 53 % des participants, mais avec un RMN de 9,14. Ce critère semble donc important à considérer par la majorité des participants, sans toutefois être jugé prioritaire.
- La prise en compte de la coarticulation voyelle-à-voyelle a été classée par seulement 20 % des participants, mais avec un RMN de 6,5. Par conséquent, bien qu'elle n'ait pas été choisie par la majorité, elle semble être un critère assez important pour un sous-groupe d'experts.

En outre, 8 % des experts ajoutent dans leurs commentaires que le texte devrait également contenir des pseudo-mots (qui n'étaient pas initialement inclus dans les critères disponibles). Cet élément a été ajouté au deuxième tour pour validation par l'ensemble du panel.

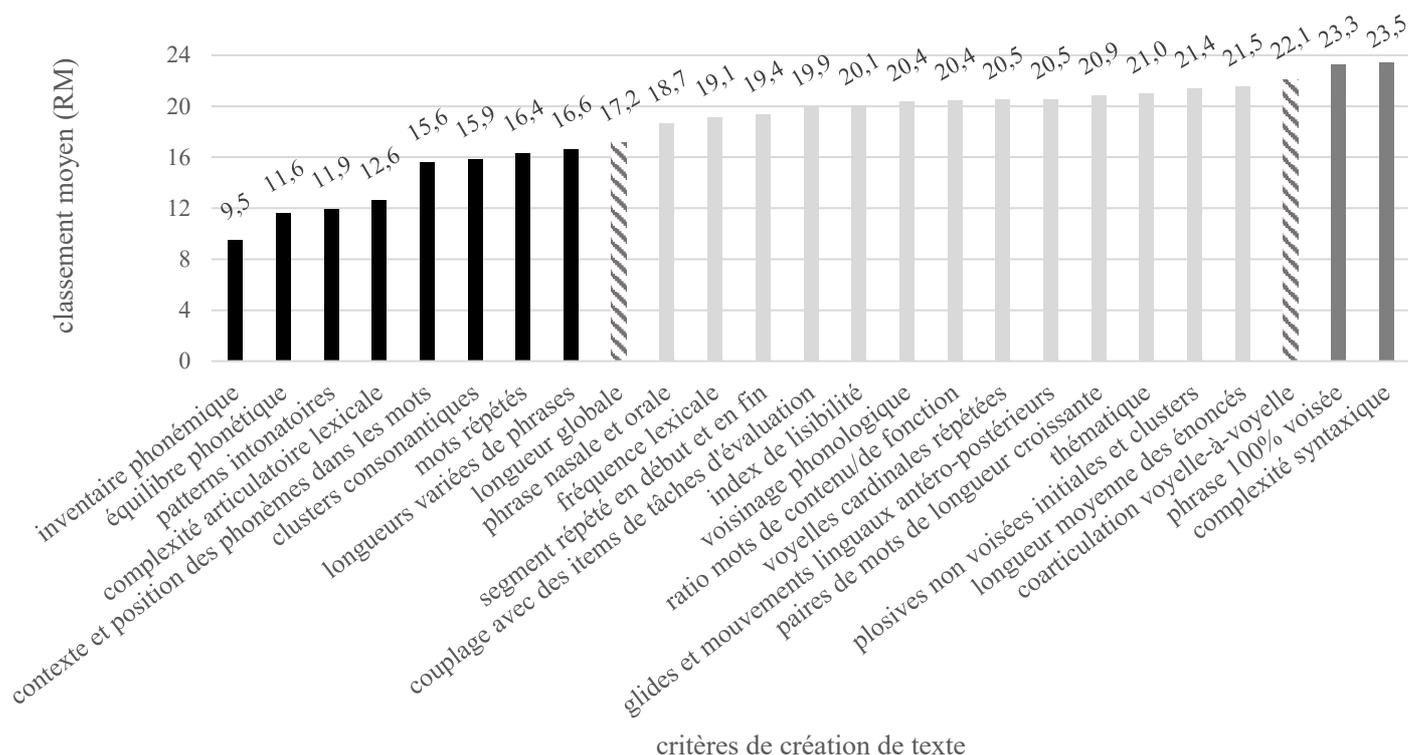


Figure 26 – Classement moyen (RM) de chacun des critères de création de texte (incluant les non-classements) ; en noir : le plus et le mieux classé (RMN < 7,5 et noté par au moins 50 % des participants) ; en gris foncé : le moins et le moins bien classé (RMN > 10 et noté par moins de 25 % des participants) ; en hachuré : résultat clivé entre le nombre d'occurrences et le RMN

Au **deuxième tour**, les participants ont ensuite évalué (Accord/Désaccord et commentaire facultatif) un set des critères les plus et les mieux classés au premier tour, ainsi que deux énoncés supplémentaires concernant les critères « clivants » et un concernant les pseudo-mots.

Les résultats de ce tour sont présentés dans le [Tableau 14](#) sous forme de pourcentage d'accord pour chacune des affirmations regroupées par concepts cibles (les astérisques indiquent que le seuil de consensus n'a pas été atteint). Un consensus a été atteint pour 3 des 4 énoncés (accord > 75 % entre les évaluateurs), avec un accord élevé (> 90 %) pour deux d'entre elles. Pour le set de huit critères essentiels, seul un expert n'a pas indiqué son accord, en précisant que les critères 7 et 8 n'étaient pour lui pas importants car ils peuvent être testés d'autres manières.

Tableau 14 – Énoncés concernant la création de textes présentés au deuxième tour de l'enquête Delphi et pourcentages d'accord

Concept	Énoncé	Accord
<i>Pseudo-mots</i>	- « Le texte devrait contenir des pseudo-mots afin d'évaluer l'intelligibilité de la parole. »	50 % (17/34)*
<i>Set de critères essentiels</i>	- « Les principaux critères à prendre en compte lors de la création d'un texte pour l'évaluation de la parole et de la voix sont :	97 % (33/34)
	1. Inventaire phonémique complet (c.-à-d. le texte inclue tous les phonèmes de la langue cible au moins une fois).	
	2. Équilibre phonétique (c.-à-d. les phonèmes inclus sont présents à une fréquence approximativement identique à celle à	

laquelle ils apparaissent dans une conversation ordinaire dans la langue cible).

3. Inclusion de divers patterns d'intonation (questions, déclarations, accentuation contrastive...).
4. Contrôle de la complexité articulatoire/phonétique des mots.
5. Prendre en compte la position des phonèmes dans les mots (début/milieu/fin), ainsi que leur contexte vocalique/consonantique.
6. Inclusion de clusters consonantiques.
7. Présence répétée de certains mots (cohérence des erreurs, apraxie vs dysarthrie).
8. Inclusion de différentes durées de phrases (soutien respiratoire, fatigabilité...).

<i>Critères avec opinions clivés</i>	- « La longueur totale du texte doit être prise en compte lors de la création d'un texte pour l'évaluation de la parole et de la voix. »	97 % (33/34)
	- « La coarticulation voyelle-à-voyelle doit être prise en compte lors de la création d'un texte pour l'évaluation de la parole et de la voix. »	79 % (27/34)

*seuil de consensus non atteint

Seuls 50 % des experts étaient d'accord avec l'énoncé concernant les pseudo-mots. Parmi ceux-ci, 41 % étaient strictement contre l'inclusion de pseudo-mots dans un texte (principalement parce qu'ils ne sont pas naturels et pourraient donc avoir un impact sur la prosodie), tandis que 41 % autres ont précisé qu'ils étaient permis mais non nécessaires (en fonction de l'objectif du texte).

Un **troisième tour** a été nécessaire pour clarifier cet énoncé. En effet, à la lecture des commentaires, des ambiguïtés semblaient exister dans la formulation. Par conséquent, l'affirmation a été reformulée et précisée au troisième tour : « Le texte devrait contenir des pseudo-mots afin d'évaluer le décodage acoustico-phonétique, à condition que l'intégration se fasse de manière à respecter au mieux le caractère naturel de la lecture (par exemple, pseudo-mots comme noms propres ou incantations). »

Lors de ce dernier tour, avec un accord à 73 % (24/33), les répondants ne sont toujours pas parvenus à un consensus. Ceux qui maintiennent leur désaccord restent principalement sur leur opinion que les pseudo-mots ne sont pas assez naturels ; l'un d'entre eux recommande à la place de recourir à l'utilisation d'énoncés à faible prédictibilité. Au vu de cette absence de consensus, il a été conclu que les pseudo-mots ne devraient pas être systématiquement inclus dans un texte pour l'évaluation de la parole et de la voix. Celui-ci doit donc se concentrer (en tout cas en ce qui concerne l'évaluation perceptive) sur la compréhensibilité, d'une manière plus fonctionnelle et écologique. L'évaluation de l'intelligibilité, en revanche, nécessite alors une tâche spécifique et distincte, utilisant des stimuli peu prévisibles (pseudo-mots, paires de mots minimaux ou phrases imprévisibles).

C. Critères principaux et pertinence

Ainsi, bien que le choix des critères puisse évidemment varier en fonction d'objectifs spécifiques, cette étude Delphi nous a permis d'identifier **dix critères principaux** parmi les 25 d'origine, dont nous allons maintenant brièvement discuter l'importance :

Critères au niveau du phonème

1. **Inventaire phonémique.** Comme nous l'avons déjà décrit, les voyelles et les consonnes jouent un rôle important dans l'intelligibilité de la parole. De plus, il a été démontré que les phonèmes peuvent être affectés de manière différentielle dans les troubles de la parole [H. KIM 10]. Des effets phonémiques différentiels ont également été mis en évidence sur la qualité de la voix (p. ex. le caractère soufflé sur les phonèmes voisés), à la fois dans l'évaluation perceptive [CAMARGO 19] et dans l'évaluation automatique [POUCHOULIN 07; REVIS 04]. Il est donc important que le texte contienne tous les phonèmes de la langue cible, au moins une fois.
2. **Équilibre phonétique.** Pour maximiser la pertinence fonctionnelle, le texte doit refléter au mieux la parole naturelle dans la langue cible. Par conséquent, les phonèmes inclus doivent se produire à peu près à la même fréquence que dans l'utilisation parlée de la langue au quotidien, et les règles phonotactiques telles que la distribution des voyelles et des consonnes doivent être prises en compte [JESUS 15].
3. **Position et contexte des phonèmes.** Le contexte phonémique influence la perception et l'acoustique des consonnes et des voyelles, comme nous l'avons déjà évoqué. De plus, dans les troubles de la fluence, la position des phonèmes est importante, car les études montrent une augmentation des disfluences sur les mots avec des consonnes initiales non voisées [ADAMS 74; WELLS 83], sur les mots avec des phonèmes non voisés suivis de phonèmes voisés [WELLS 83] et sur les clusters initiaux [WOLK 00].
4. **Clusters consonantiques.** Parmi les différents troubles de la parole rencontrés, les personnes apraxiques ont particulièrement tendance à simplifier les clusters de consonnes [GALLUZZI 15; ROSENBEK 84], notamment les clusters intrasyllabiques [AICHERT 04]. De plus, les clusters en position initiale de mots [WOLK 00], ainsi que les clusters homorganiques intersyllabiques — c.-à-d. deux consonnes ayant le même lieu d'articulation, de part et d'autre d'une frontière syllabique —, semblent induire davantage de disfluences chez les personnes qui bégaient [HUINCK 04]. Enfin, les clusters augmentent la complexité phonétique des mots (voir point 5) et peuvent donc également être affectés dans la parole dysarthrique.
5. **Coarticulation voyelle-à-voyelle.** Alors que les études sur la coarticulation dans la dysarthrie et l'apraxie ont donné des résultats équivoques [HARDCASTLE 09; KATZ 00], des résultats récents ont montré que les patterns de coarticulation voyelle-à-voyelle peuvent être liés à des difficultés de planification et de coordination dans la dysarthrie et l'apraxie de la parole [D'ALESSANDRO 19]. Il a également été constaté que la coarticulation voyelle-à-voyelle diminue avec l'âge [D'ALESSANDRO 18]. En plus d'inclure toutes les voyelles dans divers contextes phonémiques, la coarticulation voyelle-à-voyelle devrait donc également être prise en compte, ne serait-ce que pour permettre une étude plus approfondie de celle-ci dans les troubles de la parole.

Critères au niveau lexical

6. **Complexité articulatoire/phonétique des mots.** La complexité articulatoire des mots peut aujourd'hui être mesurée à l'aide de différents modèles, tels que l'« Index of Phonetic Complexity » (IPC) en anglais [JAKIELSKI 98, 00] et en français [H. LEE 14]¹. Le contrôle de la complexité articulatoire des mots présente notamment un intérêt pour l'évaluation des patients atteints d'apraxie, qui ont des difficultés de planification phonétique et sont donc sensibles à la structure phonologique lexicale [ZIEGLER 15]. La complexité phonétique est également intéressante pour l'évaluation des locuteurs dysarthriques, car leur intelligibilité peut varier en fonction de la complexité articulatoire des mots [H. KIM 10; KURUVILLA-DUGDALE 18]. De plus, il a également été démontré que les disfluences dans les troubles de la fluence augmentent pour les mots longs et complexes [BROWN 45; SASISEKARAN 14].
7. **Répétitions de mots.** Alors que la constance des erreurs comme critère de diagnostic de l'apraxie de la parole a été débattue [STAIGER 12], [HALEY 18] ont récemment montré que cette constance dépend des unités d'analyse. Ils ont ainsi confirmé l'inconstance au niveau des mots dans l'apraxie. Par conséquent, certains mots devraient être répétés tout au long du texte afin de faciliter le diagnostic différentiel entre l'apraxie et la dysarthrie.

Critères au niveau phrastique

8. **Prosodie.** La prosodie a des fonctions à la fois linguistiques, paralinguistiques et extralinguistiques [DI CRISTO 12]. Elle contribue à l'intelligibilité de la parole au niveau phonémique et à la compréhensibilité aux niveaux lexical et supra-lexical, mais elle fournit également des informations sur le sens profond du message (p. ex. intentions et attitude du locuteur) ainsi que sur le locuteur lui-même (p. ex. genre, âge, personnalité, accent étranger). Les troubles de la parole et de la voix peuvent affecter les différentes fonctions de la prosodie, par exemple en empêchant la production de certains patterns d'intonation ou en perturbant le débit et le rythme de la parole. Par conséquent, le texte doit contenir divers patterns intonatifs (phrases déclaratives/interrogatives/exclamatives/impératives) et tenir compte des fonctions de découpage syntaxique et d'accentuation contrastive (« focus »), ainsi que d'autres fonctions spécifiques à la langue (p. ex. accentuation lexicale ou ton dans les langues tonales).
9. **Longueurs de phrases.** La longueur des phrases a un impact sur l'intelligibilité de la parole, sur le débit de parole et sur les pauses de parole chez les locuteurs dysarthriques [ALLISON 14, 19]. En outre, tant pour les troubles de la parole que pour les troubles de la voix, des phrases plus longues permettent d'étudier le soutien respiratoire et phonatoire, les patterns respiratoires, la fatigabilité et la pertinence des groupes de souffle [R. PATEL 13]. Par exemple, en cas de paralysie des plis vocaux, un essoufflement (dyspnée) peut être noté à la fin de phrases plus longues [TESTON 04].

Critère général

10. **Longueur du texte.** La longueur totale du texte sera évidemment déterminée par tous les autres critères pris en compte lors de sa conception. Un compromis doit être trouvé entre l'exhaustivité du texte et le temps généralement limité disponible pour les évaluations de la parole et de la voix, ainsi que la fatigabilité des patients.

Ces lignes directrices peuvent s'appliquer aux langues indo-européennes romanes et germaniques telles que l'anglais, le français et l'allemand. D'autres groupes linguistiques, tels que les langues indo-européennes balto-slaves ou sino-tibétaines (y compris le chinois), ne sont pas ou peu représentés dans le groupe d'experts. Des études spécifiques sont nécessaires pour identifier les critères les plus importants qui s'appliquent aux spécificités de ces langues. En outre, cet ensemble doit être complété en cas d'objectifs de recherche spécifiques, pour lesquels des critères particuliers sont souvent nécessaires (voir par exemple [KUO 16; VAN ZUNDERT 98]).

3.2.3. Conception du texte

Sur base du travail d'identification et de hiérarchisation des critères de construction effectué en amont, notre groupe de travail s'est alors lancé dans la conception concrète du nouveau texte. Nous sommes partis du texte du protocole MonPaGe, en en gardant les principaux critères de construction et en le rendant conforme *a minima* aux dix critères principaux sélectionnés dans l'étude Delphi.

En plus de ces critères, nous avons également souhaité prendre en compte les mesures acoustiques corrélées avec l'évaluation perceptive de l'intelligibilité, identifiées dans notre revue systématique :

- pour l'articulation des voyelles, les mesures formantiques stationnaires (F1, F2 et intervalles ; différence de F2 entre /i/ et /u/, différence F1-F2 pour les paires /i-ε/ et /ε-æ/, aire de l'espace vocalique) et dynamiques (mouvements des formants, longueur de vecteur) : pour permettre ces mesures, nous avons souhaité intégrer dans le texte les voyelles cardinales à de multiples reprises, ainsi que les voyelles sur l'axe /i-e-ε-a/ du triangle vocalique (correspondant au degré d'aperture de la cavité orale) ;
- pour l'articulation des consonnes, l'énergie centroïde et le pic spectral dans la fricative /s/ précédant /i/ et /u/ (mesure de coarticulation anticipatoire), ainsi que la fréquence de F1 dans les voyelles précédant /t/ et /d/ (indice de voisement) : pour permettre ces mesures, nous avons souhaité intégrer les syllabes /si/ et /su/ ainsi que des syllabes /Vt/ et /Vd/.

Enfin, nous avons voulu intégrer toutes les consonnes non-sonantes (plosives et fricatives) dans un même contexte vocalique, afin de pouvoir effectuer des mesures consonantiques dans un contexte vocalique plus contrôlé, diminuant ainsi les différences acoustiques liées à la coarticulation (cf. mesures investiguées au Chapitre 4).

Notre volonté était de contrôler un maximum de critères dans une première partie du texte afin de rendre possible une passation rapide.

Voici le texte résultant ; sa transcription phonétique en français standard (avec les variations de prononciation françaises et belges et liaisons possibles) et en français québécois est disponible en [Annexe L](#). La première partie de ce texte, pour une éventuelle passation rapide, est présentée en gras. Les critères contrôlés dans cette première partie sont décrits dans le [Tableau 15](#) ; les critères pour l'ensemble du texte sont listés en [Annexe M](#). Seule la répétition de mots n'a pas pu être intégrée dans la première partie.

Transcription orthographique du nouveau texte

Lundi matin, Alice et son Papa vont à Malibou.

Là-bas, ils rejoignent Papy après un voyage sans soucis.

Il fait chaud, mais la brise légère et l'air iodé de la mer les ravivent.

Vers midi, Alice s'exclame : « J'ai vraiment très très faim ! ».

Papy les guide alors vite vers un café luxueux au bord de l'eau : Le Bigorneau Salé.

Mardi, ils vont à la plage.

Il n'y a pas un nuage dans le ciel. Papa s'interroge : « Avons-nous pris la crème solaire ? »

« Bien sûr ! », répond Alice.

Mercredi, Papa et Papy se baladent en bavardant.

Pendant ce temps, Alice se détend en lisant un roman et mange un bonbon à l'ananas.

Jeudi, elle va faire un jogging.

Papa lui crie : « Nous partons faire quelques achats ! »

Au magasin, Papy achète des noix de macadamia.

Vendredi, ils visitent un musée d'art abstrait.

Papa s'extasie devant un splendide tableau et demande : « Qui a donc créé cette œuvre ? ».

Samedi matin, Alice s'entraîne pour la soirée karaoké en répétant rapidement : « pataka pataka pataka ».

Samedi soir, ils fêtent leur départ en dansant la java sous le lilas.

Comme à l'arrivée, il fait chaud, mais la brise légère et l'air iodé de la mer les ravivent.

Dimanche, Alice, Papa et Papy quittent Malibou.

Ils rentrent affamés.

À table, il y a de la pizza garnie et des lasagnes aux champignons.

Rassasiés, ils s'exclament : « Quel séjour extraordinaire ! »

Tableau 15 – Critères de construction contrôlés dans la première partie du texte

Critère à contrôler	Mise en application	Source
Niveau phonémique		
Inventaire phonémique complet	Inclusion de tous les phonèmes du français au moins une fois (sauf /ɲ/ qui apparaît dans la suite du texte)	Delphi
Équilibre phonétique ²²	Corrélation Spearman de 0,75 entre les fréquences dans le texte et dans la langue française selon [TUBACH 90], p<0,0001 ; test U de Mann-Whitney : U = 555, p = 0,50 → distribution des fréquences identique sur le texte et en langue française	
Inclusion d'occlusives sourdes initiales et de clusters consonantiques (tr. de la fluence)	« Papa », « Papy », « brise », « exclame », « vraiment », « très très », « café »	
Inclusion de clusters consonantiques	« après », « brise », « exclame », « vraiment », « très très », « luxueux », « Bigorneau »	
Inclusion d'éléments permettant de mesurer la coarticulation voyelle-à-voyelle	« Papa » vs « Papy »	

²² Certains phonèmes peuvent être prononcés de deux manières selon les variantes régionales (/e/-/ɛ/, /o/-/ɔ/, /œ/-/ø/, /œ/-/ɛ/). Dans la plupart des cas, cependant, chacun des phonèmes de ces paires peut être élicité (p. ex. « mer » sera toujours prononcé /mɛr/, même dans les régions françaises où « après » est prononcé /a pr e/). Seules les occurrences « non ambiguës » ont été comptabilisées dans le calcul de l'équilibre phonétique, en excluant les occurrences où deux prononciations étaient possibles.

Répétitions multiples des voyelles cardinales	/a/ : 22 /i/ : 16 /u/ : 2 Inclusion des voyelles cardinales précédées d'une plosive identique pour la mesure de l'aire du triangle avec facilité de segmentation (/b a/ : « là-bas » ; /b i/ : « Bigorneau » ; /b u/ : « Malibou ») ; /i/ en syllabe ouverte pour éviter le relâchement en français québécois	Revue systématique
Inclusion de /i – e – ε – a/ pour la mesure de la différence F1-F2 (contraste hauteur de langue)	Prise en compte des phénomènes d'affrication en français québécois, des variations régionales de prononciation vocalique en France et en Belgique ainsi que de la facilité de segmentation en fonction du contexte phonémique : - /i/ de « midi » ou de « Bigorneau » (pas d'affrication, syllabe ouverte en français québécois) ou « Papy » (occlusive sourde, facilité de segmentation) - /e/ de « Alice et son » ou « iodé de » (/ses/ ou /ded/, facilité de segmentation) - /ε/ de « l'air » ou « vers » (prononciation /ε/ même en France, « mer » aussi mais précédé d'une nasale risquant d'avoir une plus grande influence coarticulatoire sur l'acoustique de la voyelle) - /a/ de « Papa » ou « café » (occlusives sourdes, facilité de segmentation)	
Inclusion de la consonne /s/ suivie des voyelles cardinales /s a, s i, s u/ pour la mesure du 1 ^{er} moment spectral sur la fricative	/s a/ : « salé » /s i/ : « soucis » /s u/ : « soucis »	
Inclusion de /t,d/ après /i/ (mesure de F1 en offset de /i/ devant /t,d/, indice de voisement)	« midi », « guide » et « vite »	
Inclusion d'éléments permettant d'analyser le contraste nasal/oral	Consonnes : /p a, b a/ (« Papa », « Papy » / « là-bas ») vs /m a/ (« matin », « Malibou ») Voyelles : /s a/ (« salé ») vs /s ǎ/ (« sans ») /f ε/ (« fait ») vs /f ẽ/ (« faim »)	
Enchaînement de voyelles et de fricatives sourdes (flux d'air continu avec voisement intermittent, poser la voix sur le souffle)	« sans soucis », « il fait chaud »	Groupe de travail
Inclusions de mots avec voyelle initiale, en position initiale de phrase ou après un signe de ponctuation induisant une pause : incitent d'éventuels coups de glotte (forçage vocal)	Début de phrase : « il fait chaud » Après un signe de ponctuation (liaison possible ou coup de glotte probable chez hypertoniques) : « Alice » (x2), « ils rejoignent »	
Niveau lexical		

Contrôle de la complexité articulatoire/phonétique des mots	Inclusion de mots simples et de mots complexes ; complexité calculée à l'aide de l'Indice de Complexité Phonétique [H. LEE 14], intégrant huit paramètres dans son calcul : - Min : 0 ; Max : 12 - Moyenne : 2,6 ; E.T. : 2,5 - Médiane : 2 ; EIQ : 3 19 % des mots ne présentent aucun paramètre de complexité ; 48 % présentent entre 1 et 3 paramètres de complexité, 33 % présentent plus de 4 paramètres de complexité.	Delphi
Niveau phrastique		
Inclusion de diverses longueurs de phrases	Phrases de 9 à 16 mots	Delphi
Inclusion de variations prosodiques (focus et modalité)	Alice s'exclame : « J'ai vraiment très très faim ! » (focus et modalité) Fonction structurale de la prosodie (rythme linguistique, accentuation)	
Inclusion d'un segment entièrement voisé d'au moins une dizaine de syllabes pour les mesures vocales	« mais la brise légère et l'air iodé de la mer les ravivent. »	Groupe de travail
Niveau global		
Longueur du texte	5 phrases, 58 mots, 185 phonèmes. Partie courte pour une passation rapide éventuelle avec un maximum de critères contrôlés.	Delphi

3.3. Conclusion du chapitre

Après avoir fait un bref état des lieux des outils et supports d'évaluation utilisés en langue française, nous avons décrit plus en détail la contribution des phonèmes à l'intelligibilité, ainsi que l'importance des textes de référence pour l'évaluation de la parole.

Au vu des limites des textes existants, nous avons mis en place un groupe de travail international d'experts francophones belges, français et québécois afin de construire un nouveau texte de référence pour l'évaluation de la parole et de la voix. Notre étude internationale de consensus Delphi nous a permis de sélectionner les critères les plus importants pour la création de ce texte, qui a ensuite été validé par l'ensemble des membres de notre groupe. Malgré les nombreux critères contrôlés, il présente une facilité de lecture moyenne (score de facilité de lecture de Flesch : 55,5 / 100). Inspiré du texte du protocole MonPaGe, il a été pensé spécifiquement pour l'évaluation de la parole et de la voix, et plus particulièrement pour rendre possibles des mesures acoustiques afin de compléter l'évaluation perceptive. La faisabilité et la reproductibilité de ces mesures acoustiques seront investiguées dans le Chapitre 5.

Chapitre 4. Traitement du signal

Tout au long de ce travail, nous avons identifié un besoin de mesures moins subjectives et plus reproductibles. Ce besoin a été identifié non seulement dans la population des cliniciens, mais également dans le domaine de la recherche scientifique sur la parole. Pour répondre à ce besoin, nous allons, dans ce chapitre, décrire les analyses acoustiques menées dans l'objectif de proposer de nouvelles pistes de mesures plus objectives de l'intelligibilité, utilisables en clinique et en recherche.

4.1. État de l'art – mesures acoustiques de l'intelligibilité de la parole

Le plus souvent, l'acoustique consonantique est investiguée dans le domaine temporel²³. Pourtant, un intérêt pour les mesures fréquentielles est également observé, tant pour les mesures stationnaires (p. ex. localisation du pic spectral [NISSEN 03]) que pour les mesures dynamiques (p. ex. pentes de F2, transitions formantiques, équations du locus [BRANCAZIO 98; DELATTRE 55; FANT 69; JONGMAN 00; Y. KIM 11; LIBERMAN 58]).

[HALLE 57], soutenus par [STEVENS 78] ainsi que par d'autres études plus récentes (p. ex. [FRANCIS 00; MOOSHAMMER 07]), avançaient déjà l'idée d'indices invariants dans les spectres consonantiques permettant de différencier les lieux d'articulation. Lindblom postulait lui aussi que, malgré les effets de coarticulation, une combinaison de caractéristiques spectrales pouvait permettre une bonne distinction entre les consonnes plosives [LINDBLOM 90]. Enfin, [SHANNON 02], passant en revue la recherche sur la perception de la parole sur les 25 années passées, conclut que l'information spectrale est primordiale pour la compréhension (*sic*) de la parole, en comparaison avec les indices des domaines temporel et d'intensité.

Notre revue systématique, dont la méthodologie et les résultats ont été détaillés dans le chapitre deux (cf. « 2.2.2. B. [Notre revue systématique PRISMA](#) »), nous a permis de répondre à la question de recherche suivante : « Quelles sont les mesures acoustiques spectrales employées dans la littérature scientifique pour quantifier l'intelligibilité en parole saine ? ».

Les mesures spectrales recensées dans notre revue sont variées et peuvent être regroupées en mesures vocaliques (majoritaires), consonantiques et semi-vocaliques. Dans les deux premières catégories, nous pouvons trouver des mesures stationnaires — prises à un instant t — et des mesures dynamiques — rendant compte de l'évolution temporelle du contenu spectral. Les mesures sur les semi-voyelles sont, par la nature non stationnaire de leur production articulatoire, inévitablement dynamiques. Le lecteur trouvera en [Annexe E](#) une liste complète des mesures identifiées dans notre revue systématique. Nous allons ici nous concentrer sur les mesures qui se sont avérées liées aux mesures perceptives dans les études retenues.

Premièrement, pour les **voyelles**, les **mesures stationnaires** de formants sont liées aux scores d'identification vocalique [BUNTON 01; FERGUSON 14; HAZAN 04; NEEL 08] : premier

²³ Par exemple : durée du bruit de friction [BEHRENS 88; JONGMAN 00; NIRGIANAKI 14], durée de la voyelle précédant une occlusion [ABDELLI-BERUH 04], délai de voisement [ABRAMSON 17; CHO 19; WINN 20] ; amplitude du burst [MAUCLAIR 13; NISSEN 03], amplitude normalisée des fricatives [JONGMAN 00]

formant [F1], deuxième formant [F2], intervalle de F1, différence de F2 entre /i/ et /u/, différence F1-F2 dans les paires /i-ε/ et /ε-æ/, aire de l'espace vocalique (VSA). La mesure VSA est couramment utilisée pour rendre compte de la centralisation des voyelles, souvent en parole pathologique [LIU 05; SAPIR 09; WEISMER 01]. Il a également été démontré qu'elle est sensible aux différences d'intelligibilité dans la parole saine [BOND 94] et aux changements articulatoires dans la parole surarticulée [LAM 12; SMILJANIĆ 09]. L'aire de l'espace vocalique est dans une certaine mesure liée à la taille et à la forme des cavités de résonance modulées par les positions de la mâchoire et de la langue [SANDOVAL 13]. Elle fournit ainsi un aperçu global de l'espace de travail articulatoire. Cependant, cette mesure a montré des résultats inconsistants [LANSFORD 14; SAPIR 09] et pourrait ne pas être assez sensible aux changements subtils d'articulation des voyelles, tant en parole saine [FERGUSON 07] que dans les troubles moteurs de la parole [WHITFIELD 14]. [SAPIR 09] ont ainsi expliqué que toutes les distances euclidiennes de l'espace vocalique ne contribuent pas de manière équivalente à la différenciation entre les locuteurs sains et pathologiques. À la lumière de cette asymétrie de sensibilité des formants vocaliques aux changements articulatoires, ils ont ainsi suggéré d'utiliser la distance euclidienne entre /i/ et /u/, qui s'est avérée être le marqueur le plus sensible. La différence de F2 entre /i/ et /u/ s'est également avérée liée à l'intelligibilité des voyelles dans l'étude de [HAZAN 04]. En outre, [LAM 12] ont constaté que, dans la parole surarticulée, les voyelles fermées tendues et laxes (/i, ɪ, u, ʊ/) contribuaient le plus à l'expansion de l'espace vocalique. Ces observations indiquent que les mesures de formants dans ces voyelles devraient être priorisées à des fins de diagnostic.

Plusieurs alternatives à la mesure VSA ont été suggérées, comme l'index d'articulation des voyelles (vowel articulation index, VAI) [SKODDA 13] et son opposé, le ratio de centralisation des formants (formant centralisation ratio, FCR) [A. R. FLETCHER 17]. Celles-ci ont été conçues pour minimiser la variabilité entre locuteurs et maximiser la sensibilité à la réduction vocalique [SAPIR 10, 11]. Cependant, toutes les mesures ci-dessus n'utilisent que le point médian de trois à quatre voyelles aux extrêmes de l'espace vocalique. [WHITFIELD 14, 17] ont quant à eux proposé une autre mesure, l'espace vocalique acoustico-articulatoire (AAVS), qui utilise des mesures de formants sur les parties voisées d'un énoncé entier en parole continue (voir [Figure 27](#)) et fournit ainsi une mesure plus globale, également supposée plus sensible. Il a notamment été montré dans notre revue systématique que l'AAVS est significativement plus large en parole surarticulée [WHITFIELD 17]. Il serait ainsi intéressant d'étudier plus en détail la manière dont la mesure de l'AAVS est corrélée aux évaluations perceptives de l'intelligibilité et rend compte des variations en parole saine.

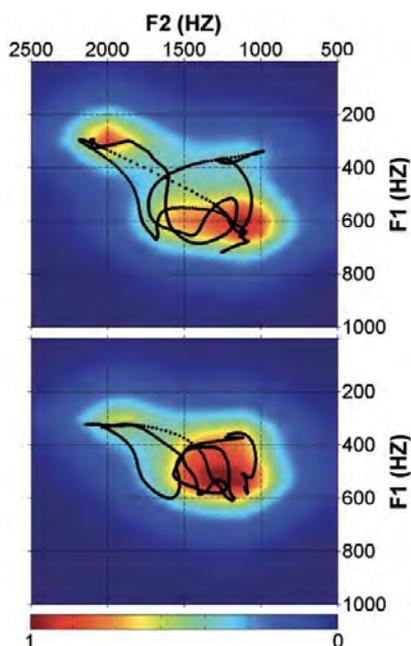


Figure 27 – Exemple de représentations de la trajectoire des formants depuis laquelle est calculé l’AAVS (tiré de [WHITFIELD 17]). Ces trajectoires ont été extraites d’une phrase prononcée par un homme en parole suraccentuée (haut) et en parole conversationnelle (bas). Les cercles noirs représentent les trajectoires F1-F2, et la couleur de fond représente la distribution de la densité (bleu = faible/aucune densité ; rouge = forte densité)

En ce qui concerne les **mesures dynamiques** des formants, les résultats des mesures « mouvement des formants » [NEEL 08], « longueur du vecteur » [FERGUSON 14] et « changement spectral » [FERGUSON 07] révèlent que les voyelles présentant des changements plus importants dans l’espace F1xF2 sont significativement mieux identifiées. [LAM 12] ont montré que les mesures dynamiques des formants des voyelles présentaient également des valeurs plus élevées en parole surarticulée. Ces mesures sont liées aux mouvements linguaux antéro-postérieurs intravocaliques et aux changements de hauteur de langue. Elles pourraient donc également être utiles dans l’étude des imprécisions dues aux contraintes motrices dans la parole conversationnelle et, par la suite, dans la parole pathologique.

Les études ciblant les caractéristiques spectrales des **consonnes** sont plus rares dans notre revue, bien que les consonnes soient réduites autant que les voyelles dans la parole courante et que cette réduction articulatoire affecte leur intelligibilité [VAN SON 99]. L’énergie centroïde de la fricative (ou « centre de gravité » [CoG]) et le pic spectral de la fricative dans le son /s/ de /si/ et /su/ se sont avérés être des indicateurs acoustiques de l’effet de coarticulation des voyelles sur la consonne [KATZ 91]. Le CoG est la première des quatre mesures de moments spectraux [JONGMAN 00] et correspond à la « fréquence qui divise le spectre en deux moitiés » [YOON 15]. Il a été constaté qu’elle est abaissée dans les non-plosives en parole spontanée de locuteurs sains par rapport à la parole lue [VAN SON 96, 99], ce qui en fait une mesure acoustique pertinente de la réduction consonantique. Les mesures de moments spectraux décrivent l’ensemble du spectre phonémique comme une distribution statistique. [EVERS 98] avançaient déjà qu’il est plus judicieux de considérer l’aspect global des spectres de consonnes sibilantes plutôt que des régions fréquentielles spécifiques. En effet, les sibilantes sont caractérisées par deux sources sonores, l’une au niveau de la constriction linguale et l’autre au niveau des dents [FANT 60], ce qui rend les emplacements des pics spectraux difficiles à prédire.

De même, la forme spectrale des consonnes est moins définie que la structure claire des formants des voyelles. Par conséquent, la description de la forme spectrale globale des consonnes devrait être préférée à l'utilisation de régions de fréquences spécifiques [FANT 60; STEVENS 78]. Un autre argument en faveur de l'utilisation des moments spectraux est qu'ils sont considérés comme étant corrélés avec la longueur et la forme de la cavité antérieure à la constriction articuloire [BEHRENS 88; KAY 12; STEVENS 98; YOON 15]. Par conséquent, ils peuvent *a priori* conduire à une interprétation articuloire. Nous aborderons ces mesures dans nos analyses acoustiques ci-après.

Tout comme pour les voyelles, un autre type de mesure a été utilisé dans les articles retenus pour analyser les consonnes et les glides : les **transitions de formants dynamiques**, parmi lesquelles la pente du second formant (F2). La mesure de la pente F2, utilisée dans les glides dans l'étude de [MARTEL-SAUVAGEAU 17], est « une mesure dynamique qui reflète la vitesse à laquelle les mouvements de la parole peuvent être effectués » [KENT 89a] et est donc liée à la vitesse de la parole. [VAN SON 99], étudiant les corrélats acoustiques de la réduction consonantique en parole saine, ont constaté que la « différence de pente de F2 » (c.-à-d. la différence entre la pente de F2 dans les frontières consonantiques en syllabe voyelle-consonne-voyelle [VCV], voir [Figure 28](#)) est plus faible en parole spontanée qu'en parole lue. Cette réduction de différence des pentes de F2 indique une plus faible coarticulation induite par la consonne en syllabe VCV, donc une articulation réduite de cette consonne. L'utilisation de mesures de transition de formants est d'autant plus intéressante qu'il a été démontré que dans le vieillissement sain, une diminution de l'intelligibilité peut être en partie attribuée à un ralentissement des mouvements de la langue [KURUVILLA-DUGDALE 20]. En parole pathologique, la pente de F2 est considérée comme un indicateur de la lenteur articuloire (à différencier du débit de parole) et donc de la sévérité globale du trouble de la parole [KENT 89a; Y. KIM 09; WEISMER 92, 95].

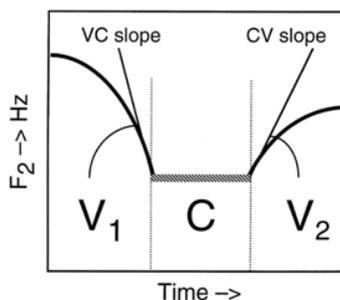


Figure 28 – Représentation de l'analyse de différence de la pente F2 aux frontières consonantiques (tiré de Van Son and Pols 1999) ; la pente de la frontière VC est soustraite de la pente de la frontière CV

Diverses mesures acoustiques spectrales, principalement sur les voyelles, se sont donc avérées liées à l'intelligibilité perçue de la parole chez les locuteurs sains. Cependant, aucune de ces mesures n'explique un pourcentage élevé de la variance des scores d'intelligibilité perceptive. Par conséquent, une combinaison de mesures acoustiques, prenant en compte divers traits phonémiques et aspects spectraux, pourrait être une première voie vers une évaluation plus exhaustive de l'intelligibilité de la parole (p. ex. [BRADLOW 96; KENT 89b; J. LEE 14; LINDBLOM 90; WEISMER 08]). En outre, les caractéristiques acoustiques segmentales sont influencées par les facteurs à d'autres niveaux de granularité tels que l'intonation, l'accentuation (p. ex. les différences acoustiques entre les voyelles accentuées et non accentuées

dans l'étude de [CONNAGHAN 17]), la qualité vocale et le débit de parole. Aussi, avant d'utiliser des mesures acoustiques segmentales sur des populations spécifiques de patients, des recherches approfondies sont encore nécessaires pour mieux comprendre leur comportement chez les locuteurs sains, pour identifier les combinaisons acoustiques pertinentes qui pourraient expliquer les variations de parole perçues et pour fournir des données normatives à partir d'un large ensemble de locuteurs sains.

L'intérêt des mesures acoustiques consonantiques pour l'étude de la parole a été maintes fois souligné, notamment sur les consonnes non-sonantes [KENT 99]. Aussi, dans cette thèse, nous avons souhaité contribuer à l'investigation de **mesures acoustiques segmentales** sur les **consonnes en parole saine francophone**, en prenant en compte l'influence de la coarticulation et des facteurs suprasegmentaux.

4.2. Exploration acoustique préliminaire sur les consonnes : la « banane de la parole »

Notre première exploration acoustique a fait suite au constat d'une prépondérance de mesures spectrales sur les voyelles, au détriment des consonnes. Inspirés de l'aire du triangle vocalique, nous souhaitons proposer une mesure équivalente pour les consonnes. Notre objectif était ici de tendre vers une représentation globale visuelle de l'articulation des consonnes, facilement interprétable notamment par un clinicien lors du bilan de parole.

4.2.1. Motivation

L'aire du triangle vocalique, construit en mesurant les valeurs fréquentielles des deux premiers formants (F1 et F2) des voyelles, est en effet une mesure courante dans l'évaluation de l'intelligibilité [TURNER 95; VILAIN 15; WEISMER 00]. Par le placement des voyelles sur un graphique à deux axes, F1 et F2, le triangle vocalique obtenu permet d'interpréter les valeurs mesurées par rapport au triangle « standard ». Cette mesure est employée autant en pratique clinique qu'en recherche scientifique. Elle offre un aperçu global de l'espace articulo-vocalique. Aucune alternative similaire n'existe cependant à ce jour pour les consonnes, si ce n'est en perception de la parole. En effet, en audiométrie, la « banane de la parole » (voir [Figure 29](#)) est une représentation des consonnes par un point sur un graphique à deux axes, fréquence et niveau de pression sonore [KLANGPORNKUN 13; NORTHERN 84; SKINNER 78]. Ces points, correspondant à la composante fréquentielle principale « standard » de chaque consonne et à son intensité, permettent la délimitation d'une « aire des consonnes » sur un audiogramme. C'est le rapport de la courbe d'audition de chaque patient à cette aire qui permet ensuite de prédire la perception des différentes consonnes, primordiales pour l'intelligibilité de la parole [BONATTI 05; OWREN 06]. Par cette méthode, l'audioprothésiste peut estimer l'impact du déficit auditif sur la perception de la parole, et ajuster l'amplification des bandes de fréquences de l'appareil auditif en conséquence.

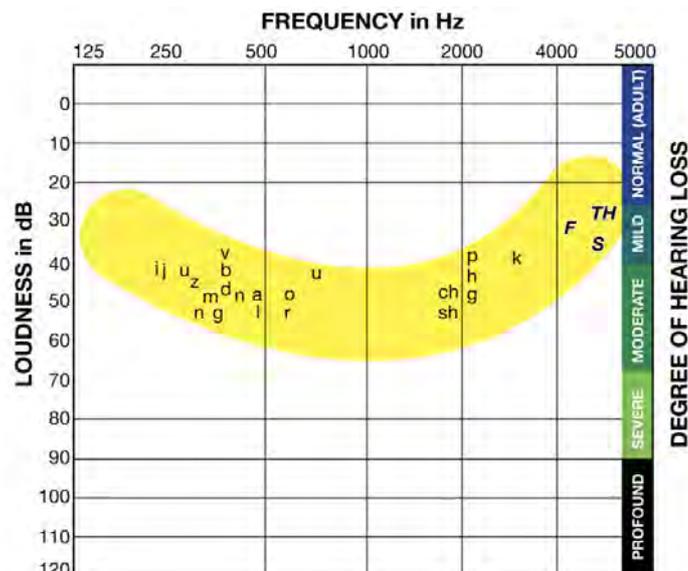


Figure 29 – La « banane de la parole » utilisée en audiométrie pour évaluer l'impact du déficit auditif sur la perception de la parole ; récupéré sur <https://hearinghealthcare.org/blog/the-audiogram-and-what-it-means/>

Notre première étude acoustique a ainsi comme objectif de transposer cette « banane de la parole » du domaine de la perception vers celui de la production de la parole. Nous visons à reconstruire une telle « aire des consonnes » sur base d'enregistrements de parole d'adultes sains, pour les 16 consonnes principales du français. Une démarche similaire a été réalisée pour les consonnes de l'anglais [NORTHERN 84; SKINNER 78] et du thaï [KLANGPORNKUN 13]. Pour le français, [BECHET 12] ont calculé à l'aide des F2 et F3 des aires consonantiques sur les occlusives sonores /b,d,g/, représentant les trois lieux d'articulation les plus communs à l'échelle universelle [ABRY 03].

Nous souhaitons déterminer si un pic d'énergie spectrale récurrent existe pour les différentes consonnes, pour pouvoir localiser celles-ci sur les axes de fréquence et de niveau de pression sonore. Le clinicien pourrait ainsi comparer la représentation graphique obtenue pour le patient, à la banane « normale » construite sur base d'enregistrements de sujets sains. Nous pourrions notamment imaginer le calcul de l'aire de cette banane (telle l'aire du triangle vocalique déjà employée actuellement), ou des mesures selon chaque axe (p. ex. diminution de la banane sur l'axe fréquentiel et/ou du niveau de pression sonore). Par l'automatisation de la construction d'une telle représentation des consonnes produites par le sujet, nous espérons pouvoir proposer un outil d'évaluation objective de l'intelligibilité de la parole.

4.2.2. Méthodologie

A. Stimuli et enregistrements

Dans cette étude pilote, 15 sujets sains francophones natifs et sans antécédents neurologiques/ORL (6 hommes, 9 femmes, âge médian : F=47, EIQ=18, min=23, max=76 ; M=41, EIQ=17 ; min=24 ; max=76) ont produit la phrase porteuse « Le sac /øCø/ convient », où C représente chacune des 16 consonnes : /b, f, d, f, g, ʒ, k, l, m, n, p, r, s, t, v, z/.

Les enregistrements standardisés ont été effectués dans une pièce anéchoïque avec un micro serre-tête super cardioïde à condensateur (Thomann t.bone HC-444 TWS) à environ 5 cm de la bouche du locuteur, et un enregistreur numérique Zoom H4n Pro avec des paramètres

d'enregistrement constants (fréquence d'échantillonnage de 48 kHz, 16 bits par échantillon, niveau d'entrée de 30 dB, enregistrement par un microphone mono). Le post-traitement du signal audio comprenait le sous-échantillonnage à 16 kHz²⁴. Les consonnes ont été manuellement segmentées via Praat. L'oscillogramme et le spectrogramme ont été utilisés pour déterminer les frontières des phonèmes. Pour les plosives, seule la première phase («transient») de l'explosion («burst») a été incluse — caractérisée par une bande verticale courte et sombre qui couvre la plupart des fréquences du spectrogramme —, excluant ainsi le silence ainsi que les segments fricatifs et aspiratifs subséquents [FANT 69]. En effet, [CATTELAÏN 19] a montré dans son étude sur les moments spectraux que la distinction entre les lieux d'articulation est la plus nette lors des toutes premières phases du burst (5 ms) et que le bruit de friction qui suit influence fortement les indices acoustiques. [LIU 00] avaient également constaté que le pic fréquentiel proéminent dans le burst était l'indice acoustique principal pour différencier le lieu d'articulation.

B. Analyse préliminaire

Une analyse préliminaire sur deux locuteurs supplémentaires (un homme, une femme) répétant trois fois les stimuli de parole dans quatre contextes vocaliques (/a, i, u, e/) nous a permis de vérifier visuellement que les spectres obtenus étaient reproductibles et stables et d'observer si ces contextes induisaient des spectres consonantiques différents. Les trois répétitions de chaque séquence VCV produisaient effectivement des spectres visuellement similaires, mais le contexte vocalique semblait influencer sur le spectre consonantique. Ce résultat est similaire chez l'homme et chez la femme. En [Figure 30](#), par exemple, nous pouvons observer que le /u/ induit un premier pic autour de 1800 Hz dans le spectre de la consonne /ʃ/ (contre un pic plus élevé pour les autres voyelles), et ce pour les trois répétitions. Ceci peut être mis en lien avec l'arrondissement de la voyelle qui induit une labialisation de la consonne, et de par ce fait un abaissement des fréquences du bruit de friction [MEYNADIER 13].

Ainsi, nous nous sommes orientés vers les voyelles centrales (« neutres », [œ]/[ø]). En effet, au vu de l'influence du contexte vocalique sur le spectre consonantique, les voyelles extrêmes du triangle vocalique n'étaient pas de bons candidats. De plus, selon [FLEMMING 09], le « schwa » est une voyelle « faible », qui ne présente en elle-même pas une structure stable, mais aurait de ce fait également peu d'influence sur les consonnes qui l'entourent. Enfin, dans une perspective clinique, les voyelles « neutres » semblaient également les plus pertinentes, en raison de la difficulté de certains patients à produire les voyelles extrêmes du triangle vocalique. La voyelle neutre /œ/ étant le plus souvent prononcée /ø/ dans les régions du sud de la France, nous avons choisi cette dernière pour nous assurer que tous les sujets produisent le même contexte vocalique.

²⁴ Pour le codage prédictif linéaire, employé dans cette étude, une fréquence d'échantillonnage de deux fois la fréquence formantique maximale attendue (soit deux fois 8000 Hz, correspondant à la limite supérieure des fréquences des sons de la parole testés en audiométrie classique) est préconisée.

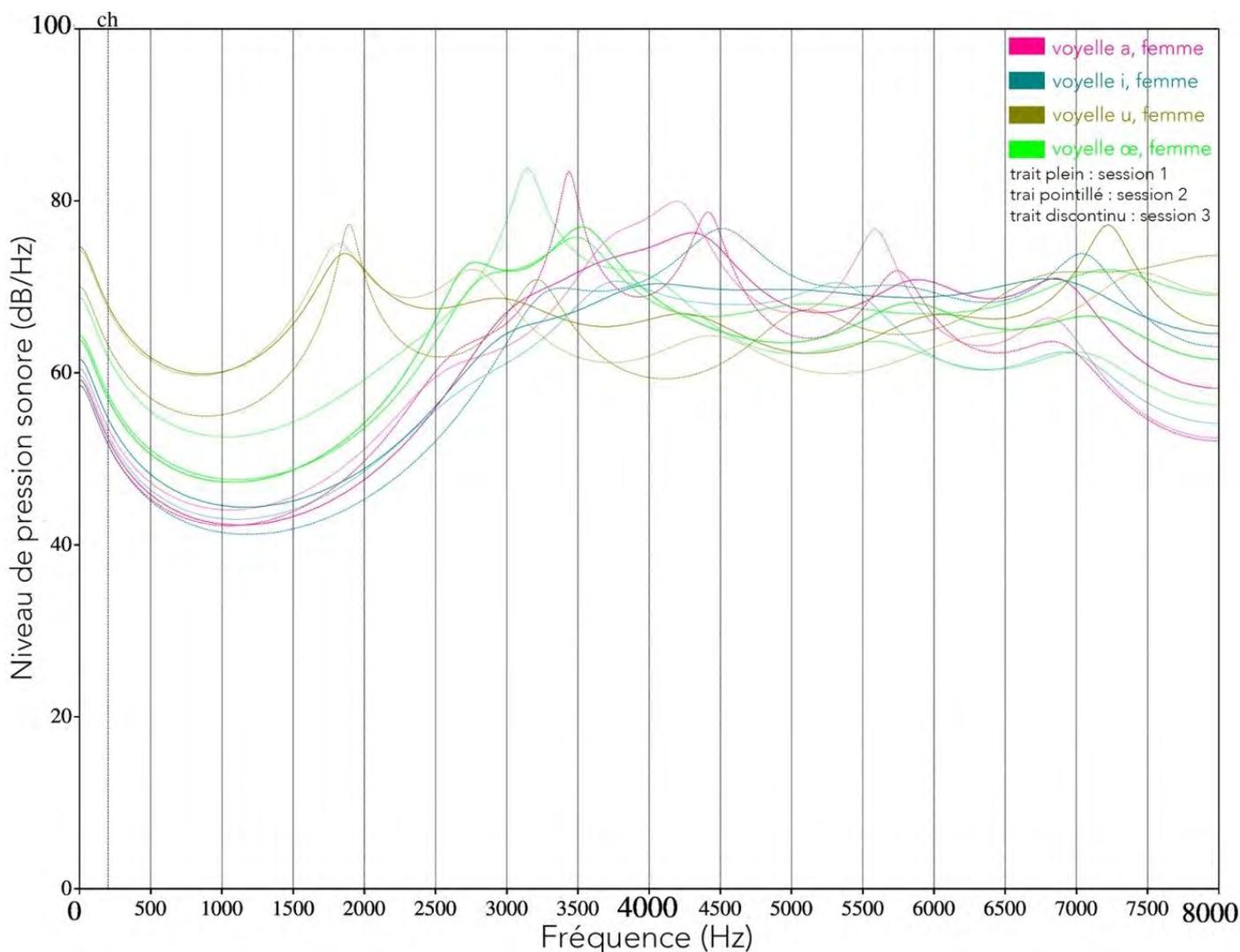


Figure 30 – Exemple d'analyse préliminaire de la stabilité des spectres et de l'influence du contexte vocalique sur la consonne /f/ prononcée par une locutrice

C. Analyse acoustique

L'analyse semi-automatique a tout d'abord consisté à tracer un spectre lissé grâce au codage prédictif linéaire (LPC) [ATAL 71; ZUE 76]. La méthode LPC part du postulat que la parole est produite selon la théorie source-filtre [FANT 60]. Elle se base sur la corrélation entre les échantillons consécutifs du signal de parole et décompose le signal sonore en ses deux composantes : source et filtre. La composante « filtre », correspondant au tractus vocal, est modélisée par un nombre de coefficients (ou « pôles »). Le choix de ce nombre de coefficients, ou « ordre de prédiction », va déterminer l'aspect plus ou moins lissé du spectre (voir [Figure 31](#)). La règle de base est de choisir un ordre deux fois plus élevé que le nombre de pics spectraux attendus [LUDLOW 19; VALLABHA 02]. La composante « source » est modélisée par deux coefficients additionnels représentant la source glottique.

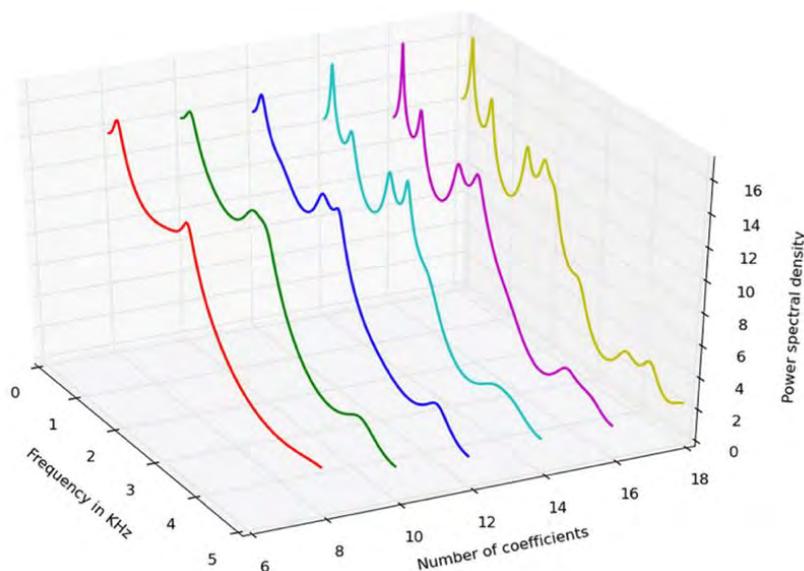


Figure 31 – Spectres LPC de la voyelle /u/ pour un ordre de prédiction de 8, 10, 12, 14, 16 et 17 [DISSEN 19]

Comme nous nous attendions à identifier au maximum 3 à 4 pics par consonne (sur base d’essais préliminaires avec différents ordres de prédiction), nous avons donc effectué un codage prédictif linéaire selon la méthode de Burg [ANDERSEN 74] en ordre 10. La largeur des fenêtres d’analyse était de 25 ms, avec un pas de 5 ms. Nous avons ensuite extrait un spectre au centre de chaque segment consonantique.

Ces spectres lissés nous ont permis d’identifier manuellement les pics spectraux proéminents pour chaque consonne, pour chaque locuteur. Les valeurs médianes pour les pics récurrents ont ensuite été transposées sur un graphique à deux axes (fréquence en Hz et niveau de pression sonore en dB). Pour être retenu, un pic devait :

- être observé sur les spectres des hommes et des femmes, dans une zone de fréquences similaire (différence <1000 Hz) ;
- être mesuré chez la majorité ($\geq 50\%$) des hommes et des femmes ;
- ne pas présenter une étendue fréquentielle trop large (<1500 Hz) ni un niveau de pression sonore trop faible (<5 dB/Hz).

Aucun pic récurrent n’a pu être identifié pour /p, b, v/.

Nous avons utilisé des tests U de Mann-Whitney pour analyser les différences entre les valeurs chez les hommes et chez les femmes, pour les fréquences et les intensités de chaque pic de chaque phonème.

4.2.3. Résultats

Sept pics se différençaient significativement ($p < 0,05$) au niveau de la fréquence (/ \int_{pic1} , \int_{pic2} , \int_{pic3} , s, l_{pic2}, d, t/), un seul au niveau de l’intensité (/t/). Ce résultat nous a amenés à réaliser deux « bananes » séparées en fonction du genre (voir [Figure 32](#)).

Une représentation basée sur les pics fréquentiels récurrents est globalement faisable, certains pics apparaissant effectivement de manière assez constante, autant chez les hommes que chez les femmes. Seules les plosives bilabiales et le /v/ ne présentaient pas de pic récurrent et n’ont donc pas pu être incluses dans les « bananes ».

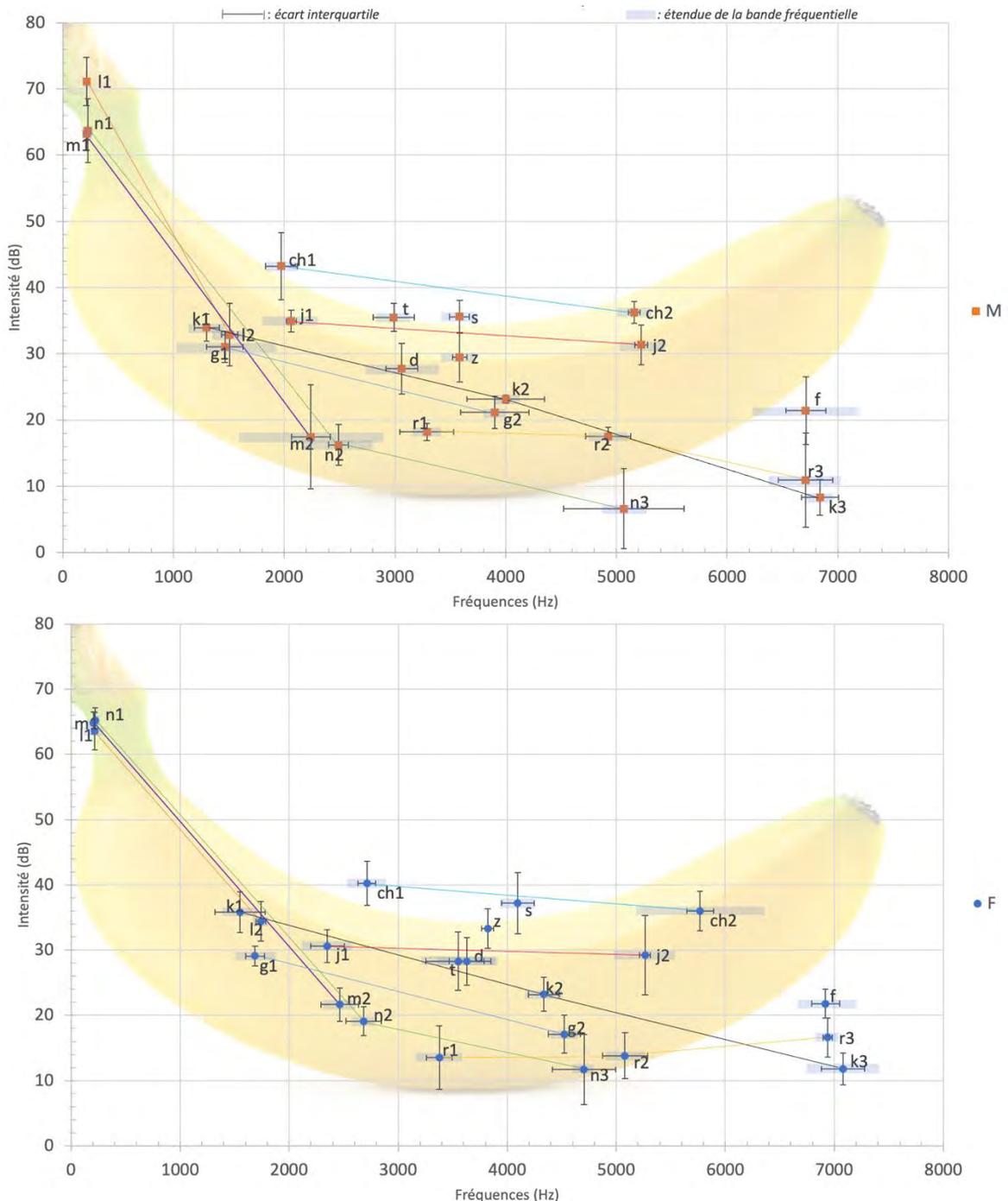


Figure 32 – « Bananes de la parole » calculées pour les six locuteurs (en haut, en orange) et pour les neuf locutrices (en bas, en bleu), montrant la moyenne des principaux pics spectraux des consonnes sur un graphique à deux axes (Intensité x Fréquence)

4.2.4. Discussion

Bien qu’il s’agissait ici d’une étude pilote — avec un échantillon limité au vu de la variabilité interindividuelle des données acoustiques [FANT 73; JOHNSON 93; LAVOIE 02; PETERSON 52] — notre étude pilote soulève différentes questions.

Tout d’abord, bien qu’une représentation basée sur les pics fréquentiels récurrents semble globalement faisable, la méthode de détection des pics ne semble pas pertinente pour tous les phonèmes. Les fricatives, notamment, sont plutôt caractérisées par des bandes ou des plateaux de fréquences renforcées, et non par des pics. Pour les plosives bilabiales, aucun pic

récurrent n'a été observé. D'autres indices acoustiques doivent ainsi être recherchés pour caractériser ces consonnes.

En outre, il reste à déterminer plus précisément comment analyser ces mesures : par exemple, devons-nous nous concentrer sur un pic principal ou prendre en compte plusieurs pics ? Si plusieurs pics sont pris en compte, l'absence de l'un d'eux a-t-elle nécessairement un impact sur l'intelligibilité ? Quelles seraient les possibilités d'analyse à effectuer sur des enregistrements pathologiques, par rapport aux valeurs « standard » (distances entre pics, plages consonantiques « standard » dans lesquelles au moins un pic doit se situer pour permettre la bonne perception du phonème...) ? Il serait également intéressant d'analyser l'intensité relative des pics via leur distance par rapport à une droite de régression linéaire à travers le spectre, afin de prendre en compte l'intensité globale de chaque production. Enfin et surtout, des études perceptives approfondies devraient être menées afin d'étudier la corrélation entre les pics spectraux et la perception des phonèmes (p. ex. en utilisant des échantillons de parole synthétiquement modifiée).

Alors que l'utilisation de pics spectraux pour différencier les phonèmes a été investiguée par d'autres auteurs, principalement sur les plosives (p. ex. [ALI 01; BONNEAU 08]), l'originalité de cette étude portait sur l'investigation de l'ensemble des consonnes. Notre objectif initial était de pouvoir proposer une représentation visuelle globale de l'articulation des consonnes basée sur les mesures acoustiques. Au vu des résultats obtenus, il s'est avéré que la détection de pics fréquentiels ne convenait pas entièrement à cet effet, l'interprétation clinique d'une telle représentation restant incertaine. Ce résultat met en question le processus utilisé pour la création de l'outil couramment employé par les audioprothésistes. Aussi, au lieu d'approfondir cette analyse par pics, nous nous sommes interrogés quant à l'existence d'autres mesures prenant en compte l'aspect plus global du spectre consonantique, afin par exemple de pouvoir représenter une enveloppe ou un profil spectral moyen pour chaque consonne. C'est ainsi que nous sommes venus à effectuer notre seconde exploration acoustique, celle des moments spectraux.

4.3. Les moments spectraux sur les consonnes non-sonantes en français

Au contraire de la méthode précédente qui se basait sur la détection de pics spectraux, les moments spectraux considèrent le spectre consonantique dans son entièreté. L'étude que nous allons décrire dans ce sous-chapitre répond ainsi à l'intérêt d'une analyse considérant l'aspect global du spectre des consonnes, qui avait déjà été souligné il y a plus de soixante ans [FANT 60; STEVENS 78]. Les moments spectraux ont montré des résultats très prometteurs pour l'analyse des consonnes. [JONGMAN 00] notamment, constata que les mesures de transition des formants et de durée du bruit n'étaient pas fiables pour distinguer les fricatives anglaises, alors que les moments spectraux permettaient de différencier les quatre points d'articulation.

A. Définition des moments spectraux

Lors de l'analyse en moments spectraux, le spectre consonantique est considéré comme une distribution statistique et est décrit par quatre mesures [JONGMAN 00] (voir [Figure 33](#)) : le centre de gravité (« center of gravity » ou CoG), l'écart-type (« standard deviation » ou SD), l'asymétrie (« skewness », SKEW) et l'aplatissement (« kurtosis », KURT). Le **premier**

moment (CoG, en Hertz), que nous avons déjà évoqué dans notre revue systématique, est la « fréquence qui divise le spectre en deux moitiés de telle sorte que la quantité d'énergie dans les régions de haute fréquence est égale à celle dans la région de basse fréquence » [YOON 15]. Par conséquent, un CoG élevé indique que le noyau spectral est situé dans les hautes fréquences, par exemple dans les sibilantes [HERNANDEZ 19]. Le **second moment (SD, en Hertz)**, mesure la dispersion du noyau spectral autour de ce centre de gravité. Un petit SD indique une concentration étroite autour du CoG. Le **troisième moment (SKEW)** est généralement considéré comme étant lié au CoG [CATTELAÏN 19; NEWMAN 03; TJADEN 10]. Alors qu'une asymétrie de 0 correspond à une distribution symétrique, une valeur positive indique que la distribution est asymétrique vers la droite, c.-à-d. que la queue droite de la distribution (valeurs au-dessus du mode et de la moyenne) s'étend plus loin que la queue gauche [JONGMAN 00] (voir [Figure 34](#)). SKEW peut également être interprété comme faisant référence à l'inclinaison spectrale [JONGMAN 00], une valeur positive indiquant une inclinaison négative (une forte concentration spectrale dans les basses fréquences), et inversement. Le **quatrième moment (KURT)** est quant à lui lié au SD [CATTELAÏN 19; NEWMAN 03; TJADEN 10] et constitue une mesure de l'acuité du pic. Plus le KURT est élevé, plus le pic est défini. Notons que SKEW et KURT ont tous deux des unités sans dimension [GOTTSMANN 11].

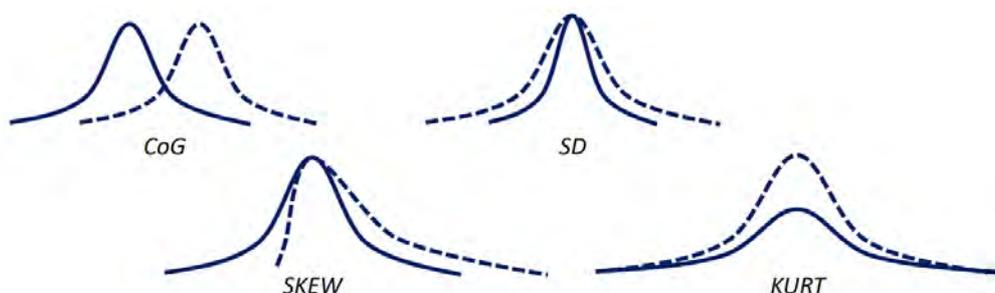


Figure 33 – Représentation schématique des quatre moments spectraux

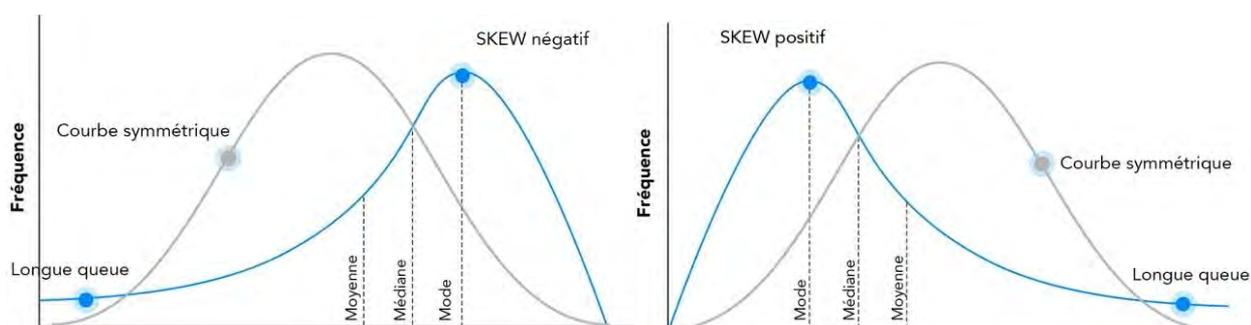


Figure 34 – Représentation schématique de courbes avec un SKEW négatif (gauche) et positif (droite), en comparaison à une courbe symétrique (SKEW de 0) ; traduit et adapté de <https://analystprep.com/cfa-level-1-exam/quantitative-methods/unimodal-distribution-mean-mode-median/>

Il est communément admis que la forme spectrale est caractéristique de la longueur et de la forme de la cavité antérieure à la constriction articulaire [BEHRENS 88; KAY 12; STEVENS 98; YOON 15]. Ainsi, pour les consonnes plus antérieures, le noyau spectral est concentré dans les fréquences plus élevées. De même, plus la cavité antérieure est longue, meilleure sera la définition de la forme spectrale [STEVENS 98]. De ce fait, les fricatives labiodentales et plosives labiales présentent en théorie des spectres plutôt plats, alors que les consonnes (palato-)alvéolaires présentent une forme spectrale bien définie et pointue [JONGMAN 00; REPP 89].

B. Intérêt clinique

Des liens directs entre les moments spectraux et les mesures articulatoires physiologiques ont été suggérés, par exemple entre le CoG et le SKEW dans les plosives et le mouvement de la langue [CATTELAÏN 19], ainsi qu'entre le CoG dans les fricatives et les mouvements de la mâchoire [ISKAROUS 11]. De ce fait, ces mesures suscitent entre autres un intérêt en application clinique [ACHER 14; HERNANDEZ 19; KAY 12; KENT 99; TJADEN 17].

De nombreuses études montrent que les moments spectraux peuvent être utilisés pour « objectiver » les différences entre la parole naturelle et « atypique ». En outre, leur mesure est non invasive et peut être effectuée avec peu d'équipement [MEYER 11]. Ainsi, les moments spectraux ont pu être utilisés pour évaluer la dysarthrie (p. ex. [DROMEY 03; HERNANDEZ 19; KAY 12; KENT 99; LAM 16; TJADEN 97, 17]). Par exemple, ils se sont révélés sensibles aux modifications de la parole, notamment dans la parole forte par rapport à la parole conversationnelle chez les personnes atteintes de la maladie de Parkinson [TJADEN 04]. Ils ont également été appliqués pour des patients glossectomisés (p. ex. [ACHER 14]), des personnes malentendantes (p. ex. [ABRAHAM 20; BLACKLOCK 04; SCARBEL 12]), mais aussi le développement de la parole chez les enfants (p. ex. [KÖRKKÖ 15; MUNSON 04; NISSEN 03, 05]). Ils seraient donc *a priori* pertinents pour l'évaluation de l'efficacité thérapeutique ou de la progression pathologique [TJADEN 10].

C. Le problème des langues

Puisque les mesures spectrales dépendent de paramètres articulatoires, nous nous attendons à ce qu'elles varient d'une langue à l'autre. En effet, alors que certains patterns semblent constants à travers les langues (p. ex. pour le CoG du /s/ [GORDON 02]), [SUNDARA 05] a montré des différences significatives entre les moments spectraux des plosives /d/ et /t/ en anglais canadien et en français canadien.

Or, très peu d'études sont disponibles pour le français. Quelques données ont été rapportées sur un échantillon de deux patients francophones glossectomisés [ACHER 14], sur deux plosives pour le français canadien [SUNDARA 05] et sur les fricatives chez des enfants francophones normo-entendants et malentendants [SCARBEL 12]. Seules deux thèses de doctorat rapportent des données de moments spectraux chez des adultes francophones sains. L'une ne traite que le CoG, le SKEW et le KURT, et ce uniquement sur les plosives [CATTELAÏN 19] ; l'autre ne rapporte que les données de CoG, sans préciser les paramètres d'extraction [KAHN 11].

D. Données disponibles en parole saine

Avant d'utiliser les moments spectraux sur la parole « atypique », il est essentiel d'en connaître le comportement en parole saine [BUTCHER 89; FORREST 88]. La plupart des études sur les moments spectraux se concentrent sur les fricatives [BLACKLOCK 04; JESUS 02] et rapportent rarement les quatre moments pour toutes ces fricatives [JONGMAN 00]. Pourtant, certains auteurs soutiennent que les moments spectraux devraient être utilisés pour la caractérisation des plosives, celles-ci représentant un tiers de toutes les consonnes en anglais [KENT 09] et en français [TUBACH 90]. De plus, une performance supérieure des moments spectraux a été démontrée sur les plosives [FORREST 88].

Pour ces différentes raisons, nous avons souhaité analyser le comportement de ces mesures en parole adulte saine, sur les consonnes non-sonantes²⁵ (plosives et fricatives) [KENT 99] du français. Nous allons maintenant décrire nos premières analyses sur base d'une segmentation manuelle, en abordant également l'influence du genre, la reproductibilité de ces mesures selon les conditions d'enregistrement, et le choix de la localisation de la fenêtre d'analyse. Ensuite, étant donné le caractère chronophage de la segmentation manuelle de phonèmes et pour permettre une applicabilité en contexte clinique au vu des contraintes temporelles, nous nous attarderons sur l'utilisation des mesures spectrales en combinaison avec l'alignement forcé automatique. Enfin, nous décrirons brièvement les comportements des moments spectraux sur des enregistrements de patients traités pour un cancer de la sphère oropharyngée.

4.3.1. Moments spectraux sur segmentation manuelle

A. Méthodologie

Locuteurs et enregistrements

Pour être inclus, les locuteurs devaient être adultes, de langue maternelle française, sans trouble de la parole passé ou présent, sans accent étranger perçu et sans plainte auditive.

Afin de vérifier la reproductibilité des mesures acoustiques ainsi que l'utilisabilité dans un contexte clinique, les enregistrements ont été réalisés dans une salle anéchoïque et dans un environnement plus naturel, à savoir un bureau calme.

Tous les enregistrements ont été effectués avec un microphone à condensateur supercardioïde Thomann t.bone HC 444 TWS de type serre-tête, à environ 5 cm des lèvres du locuteur, à l'aide d'un enregistreur ZOOM H4n Pro, avec des paramètres d'enregistrement constants : une fréquence d'échantillonnage de 48 kHz, 16 bits par échantillon, un niveau d'entrée de 30 dB avec enregistrement par un microphone mono. Le post-traitement du signal audio consistait en un sous-échantillonnage à 16 kHz.

Tâche de lecture

Chaque participant a lu à haute voix une liste de 12 phrases porteuses, avec la structure « Le sac /øCø/ convient », où C représente chacune des consonnes non-sonantes du français : /b, d, g, p, t, k, v, z, ʒ, f, s, ʃ/. La voyelle /ø/, comme pour l'étude précédente, a été choisie pour sa facilité de prononciation chez les sujets sains et chez les sujets atteints de troubles de la parole, ainsi que pour sa relative neutralité sur les plans articulatoire et acoustique permettant de limiter les effets de coarticulation [BLUMSTEIN 79; FLEMMING 09].

Segmentation et analyse acoustique

Les consonnes ont été annotées et extraites manuellement à l'aide de Praat. Les quatre moments spectraux (CoG, SD, SKEW, KURT) ont été calculés pour chaque consonne extraite, à trois points temporels [NISSEN 05] : à 15 % (début), à 50 % (milieu) et à 85 % (fin) de la consonne. Pour extraire les moments spectraux, un script Praat a été adapté de Mayer [MAYER 11]. La longueur de la fenêtre de Hamming a été fixée à 40 ms pour les

²⁵ Dans un premier temps, nous avons souhaité exclure les consonnes nasales de nos analyses car elles se sont plus sensibles aux caractéristiques du locuteur [KAHN 11]. En effet, leur articulation implique le passage de l'air par les cavités nasales, dont la forme dépend fortement de chaque individu.

fricatives et à 20 ms pour les plosives [ABRAHAM 20; BLUMSTEIN 79; GIBSON 19; JONGMAN 00; MAYER 11; NISSEN 03, 05]. Un filtre de préaccentuation de 6 dB/octave au-dessus de 100 Hz a été appliqué, les mesures des moments spectraux étant d'autant plus précises que la distribution spectrale est proche d'une courbe gaussienne [BLACKLOCK 04; FORREST 88; SUNDARA 05]. En outre, un filtre passe-haut en forme de Hann de 1000 Hz avec un lissage de 100 Hz a été utilisé pour réduire les éventuels effets de voisement dans les fréquences graves [ABRAHAM 20; ACHER 14; KÖRKKÖ 15; MARTEL-SAUVAGEAU 20; MUNSON 04; NISSEN 03; SUNDARA 05].

Des spectres LPC ont également été obtenus en utilisant l'algorithme de Burg avec un ordre de prédiction de 10, une fenêtre d'analyse de 25 ms et un pas de 5 ms. Ces spectres ont été obtenus pour le point médian des consonnes, pour chaque locuteur, après avoir appliqué les mêmes filtres de préaccentuation et passe-haut que pour l'analyse du moment spectral, ainsi qu'une normalisation de l'intensité à 70 dB pour faciliter la comparaison de ces spectres.

B. Résultats

Locuteurs et enregistrements

L'échantillon final comprend 37 enregistrements (9 en chambre anéchoïque, 28 en pièce calme²⁶) et est composé de 18 locuteurs féminins et 19 masculins, avec un âge médian de 29 ans (EIQ : 23, min : 19, max : 79). Tous les locuteurs résidaient dans la région française Occitanie (95 % en Haute-Garonne), 49 % (18/37) étaient natifs de cette région.

Données descriptives

Les résultats du test de normalité de Shapiro-Wilk²⁷ ont conduit au rejet ($p < 0,05$) de l'hypothèse nulle de distribution gaussienne pour les valeurs de CoG dans /p, t, k, d, g, s/, pour les valeurs de SD dans /t, k, d, g, ʒ, z/, pour les valeurs de SKEW dans /p, b, d, g, ʃ, ʒ, z/ et pour les valeurs KURT dans tous les phonèmes. Par conséquent, toutes les analyses statistiques ultérieures sur les moments spectraux ont été effectuées en utilisant des tests non paramétriques.

Deux enregistrements de /s/ ont été exclus des analyses statistiques ultérieures, car les valeurs KURT étaient de 57 et 129, respectivement (c.-à-d. plus de cinq fois le 99e quantile) ; à l'écoute de ces enregistrements, des sifflements aigus (sigmatismes) ont été détectés.

Les graphiques descriptifs en boîtes à moustaches sont présentés pour les quatre moments au point médian de chaque phonème en [Figure 35](#). Ces graphiques ont été créés en utilisant la procédure décrite dans Cox [N. J. COX 09].

Nous notons un nombre plus élevé de valeurs aberrantes pour la mesure KURT, tant pour les fricatives que pour les plosives.

Les spectres LPC, regroupés par consonne, sont présentés en [Annexe N](#).

²⁶ Six sujets ont été enregistrés dans les deux conditions ; leur enregistrement en chambre anéchoïque n'a été utilisé que pour les comparaisons d'environnement d'enregistrement, puis a été exclu.

²⁷ Les analyses statistiques ont été réalisées à l'aide du logiciel Stata/MP (v. 14, StataCorp, College Station, TX).

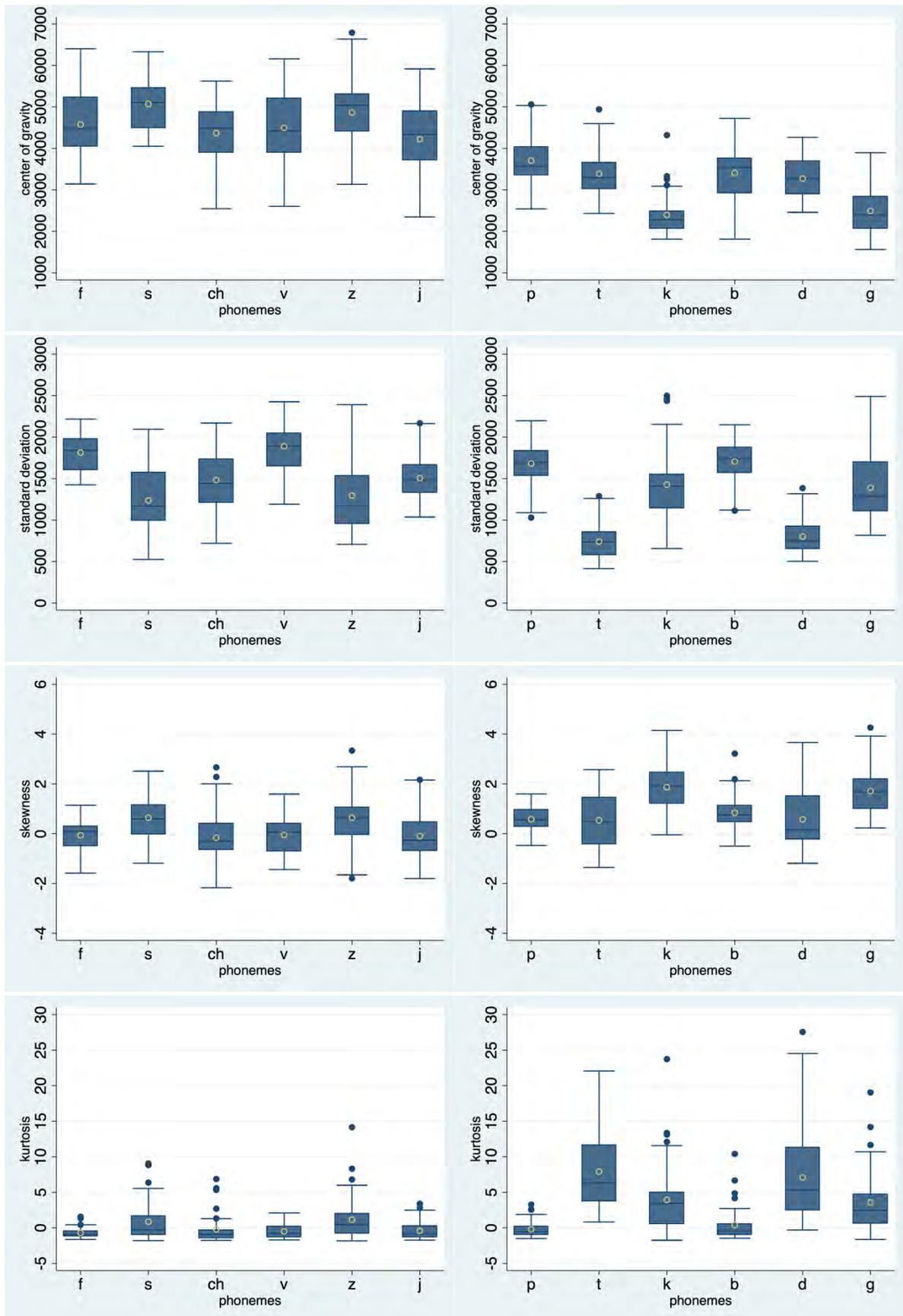


Figure 35 – Graphiques en boîtes à moustaches pour les quatre moments spectraux au point médian des consonnes plosives et fricatives ; les boîtes bleues représentent l'écart interquartile, la ligne bleue à l'intérieur des boîtes la médiane, le cercle jaune la moyenne, les moustaches incluent tous les points situés dans un intervalle de 1,5 EIQ du quartile supérieur/inférieur, les points bleus représentent les valeurs qui se trouvent en dehors de ces limites

Reproductibilité

Le test de Kruskal-Wallis, utilisé pour comparer les valeurs calculées sur des enregistrements en cabine anéchoïque (9 enregistrements²⁸) et en bureau calme (28), a donné des valeurs de probabilité allant de 0,0001 à 0,0013 pour les moments spectraux des fricatives et de 0,0001 pour tous les moments spectraux des plosives, indiquant que les différences entre certaines des médianes sont statistiquement significatives.

La comparaison post-hoc par paires (test de Dunn avec correction de Bonferroni) a cependant révélé qu'aucune différence n'existe pour un même phonème entre les valeurs des enregistrements en chambre anéchoïque et en chambre calme (voir [Figure 36](#)). La comparaison manuelle des six enregistrements effectués dans les deux environnements a conforté ce résultat.

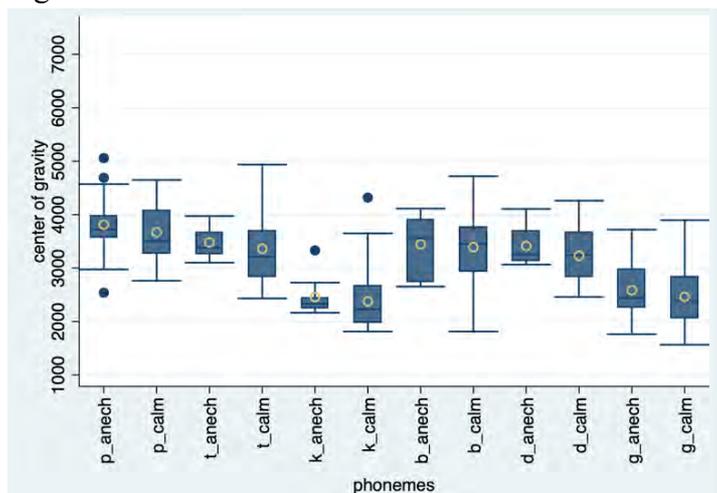


Figure 36 – Exemple de comparaison de valeurs en pièce anéchoïque (« anech ») et calme (« calm ») pour CoG des plosives

Effet du genre

Puisque les mesures spectrales reflètent la taille et la forme du tractus vocal, qui peuvent différer en fonction du genre du locuteur [SUNDARA 05], nous avons également voulu vérifier la différence entre les moments spectraux des locuteurs masculins et féminins. Le test de Dunn avec correction de Bonferroni, employé pour comparer les phonèmes par paires, a montré qu'aucune différence n'existe entre les genres pour un même phonème (p.ex. entre les moments spectraux du /s/ chez l'homme et du /s/ chez la femme).

Emplacement de la fenêtre d'analyse

Les résultats du test des rangs signés de Wilcoxon utilisé pour étudier les différences entre les emplacements de la fenêtre d'analyse (début, milieu, fin) pour chaque moment spectral, pour chaque phonème, sont illustrés dans la [Figure 37](#) et la [Figure 38](#).

Dans les fricatives, le SD et le KURT donnent le moins de différences statistiquement significatives à travers le temps, alors que le CoG et le SKEW montrent plus de variations.

Dans les plosives, tous les moments spectraux produisent un plus grand nombre de différences statistiquement significatives dans le temps que les fricatives.

Les patterns d'évolution temporelle pour les consonnes voisées et leurs homologues non voisées sont comparables (p. ex. patterns concaves et convexes pour le CoG et le SKEW dans les fricatives, respectivement).

²⁸ Exclusion des 6 enregistrements effectués dans les deux environnements, étant donné que le test de Kruskal-Wallis est destiné à comparer des échantillons indépendants.

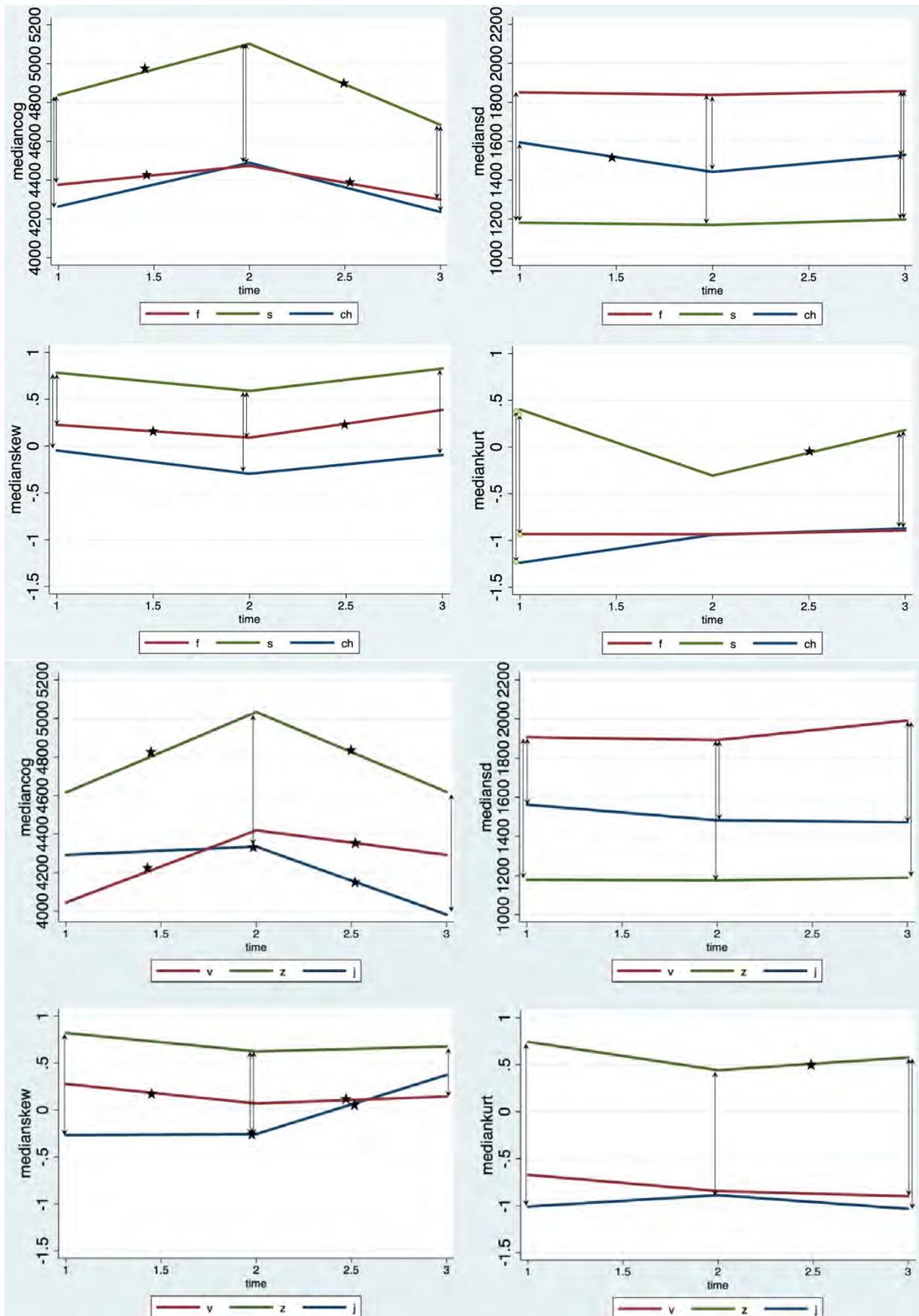


Figure 37 – Évolution par fenêtre temporelle des moments spectraux dans les fricatives non voisées et voisées ; les étoiles indiquent les différences « horizontales » statistiquement significatives entre les fenêtres d'analyse ($\alpha = 0,0167$) ; les flèches verticales indiquent les différences « verticales » statistiquement significatives entre les lieux d'articulation

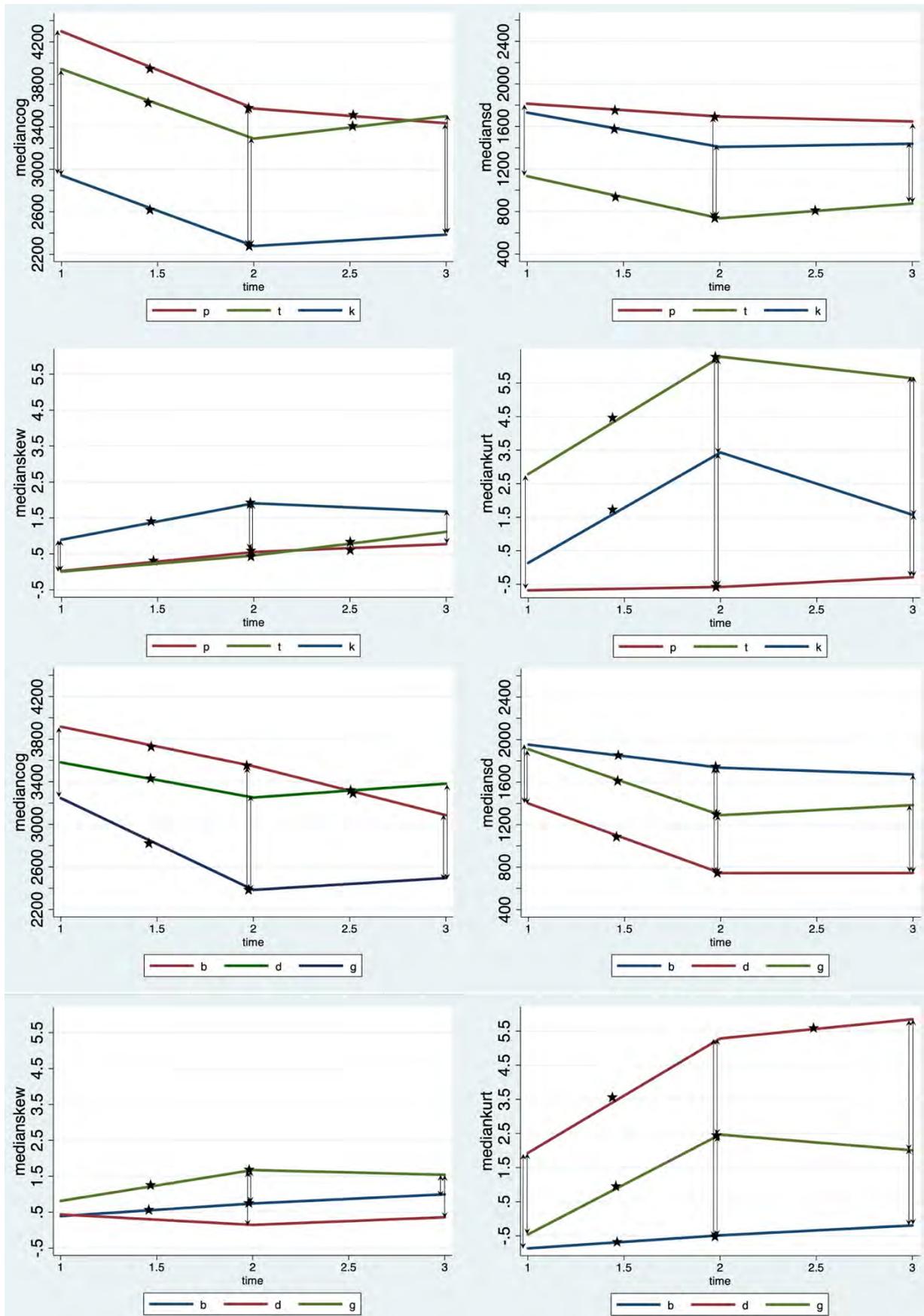


Figure 38 – Évolution par fenêtre temporelle des moments spectraux dans les plosives non voisées et voisées

Lieu d’articulation

Tous les tests de Kruskal-Wallis pour chaque moment spectral entre les phonèmes, au début, au milieu et à la fin des consonnes ont donné lieu à une différence statistiquement significative (avec des probabilités allant de 0,0001 à 0,003).

Les résultats des comparaisons posthoc par paires (test de Dunn avec correction de Bonferroni) sont également illustrés dans la [Figure 37](#) et la [Figure 38](#). Les différences statistiquement significatives sont indiquées par des flèches noires verticales, montrant entre quels lieux d’articulation chaque moment spectral diffère à un emplacement de fenêtre d’analyse donné.

La [Figure 39](#) résume le(s) moment(s) spectral(aux) qui diffère(nt) entre les lieux d’articulation des fricatives et des plosives voisées et non voisées, respectivement, au point temporel médian. Aucun moment spectral ne différencie les paires de phonèmes voisés et non voisés.

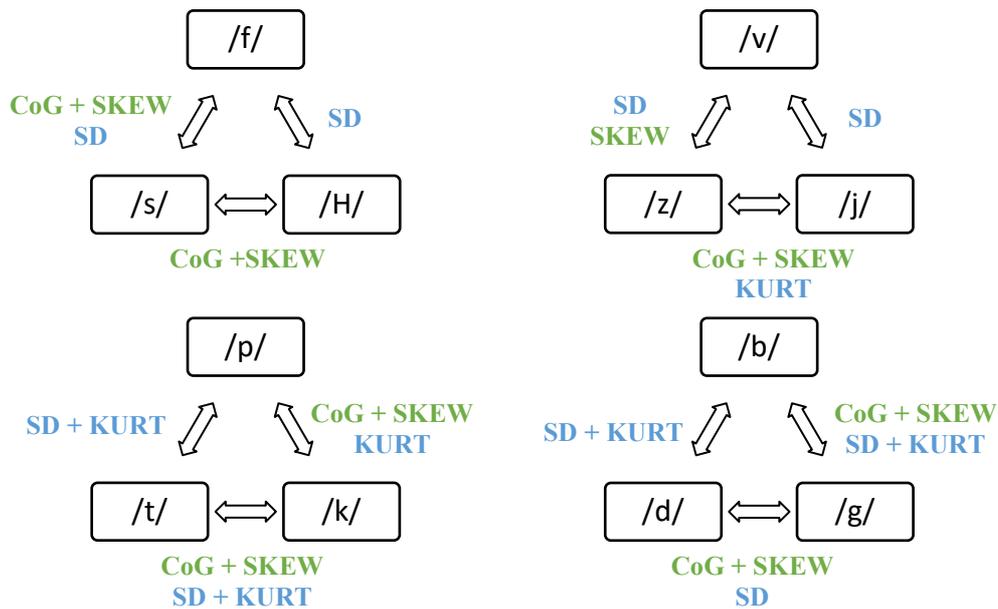


Figure 39 – Résumé des distinctions entre les lieux d’articulation basées sur les moments spectraux dans les fricatives (2 diagrammes supérieurs) et dans les plosives (2 diagrammes inférieurs) ; en vert : CoG et SKEW, en bleu : SD et KURT, paires respectivement liées théoriquement

Le [Tableau 16](#) illustre les moments spectraux qui différencient les fricatives et plosives au point médian consonantiques (voir aussi la [Figure 35](#) pour une comparaison visuelle).

Tableau 16 – Médiane [intervalle interquartile] des moments spectraux pour chaque phonème

	s	ʃ	f	z	ʒ	v
CoG	5103 [4498; 5459]	4490 [3903; 4869]	4474 [4051; 5230]	5033 [4417; 5304]	4334 [3719; 4900]	4419 [3904; 5208]
SKEW	0,59 [-0,01; 1,15]	-0,30 [-0,65; 0,41]	0,09 [-0,49; 0,29]	0,62 [-0,03; 1,06]	-0,26 [-0,67; 0,46]	0,07 [-0,70; 0,41]
SD	1170 [996; 1575]	1442 [1215; 1732]	1838 [1608; 1976]	1174 [954; 1532]	1483 [1331; 1664]	1892 [1649; 2047]
KURT	-0,31 [-0,92; 1,67]	-0,94 [-1,39; -0,32]	-0,93 [-1,13; -0,50]	0,44 [-0,76; 1,95]	-0,89 [-1,28; 0,23]	-0,85 [-1,27; 0,21]
	k	p	t	g	b	d
CoG	2277 [2075; 2479]	3573 [3356; 4028]	3288 [3026; 3654]	2384 [2071; 2836]	3546 [2925; 3761]	3252 [2903; 3690]

SKEW	1,91 [1,22; 2,46]	0,55 [0,30; 0,96]	0,45 [-0,41; 1,45]	1,67 [1,02; 2,18]	0,74 [0,47; 1,13]	0,14 [-0,22; 1,51]
SD	1408 [1147; 1550]	1693 [1537; 1835]	737 [579; 852]	1289 [1109; 1696]	1735 [1574; 1877]	744 [657; 920]
KURT	3,43 [0,55; 4,96]	-0,58 [-0,97; 0,17]	6,28 [3,75; 11,64]	2,47 [0,65; 4,67]	-0,49 [-0,91; 0,55]	5,28 [2,49; 11,30]

Note : en gras, les valeurs significativement différentes des autres sur la même demi-rangée

Corrélations entre moments

Les coefficients de corrélation de Spearman²⁹ entre les moments spectraux mesurés au point médian des consonnes et les probabilités sont présentés dans le [Tableau 17](#).

Dans les fricatives, CoG et SKEW sont modérément corrélés négativement ; SD et KURT sont fortement corrélés négativement. SKEW n'est que faiblement corrélé avec SD (négativement) et KURT (positivement). Il n'y a pas de corrélation significative entre CoG et SD ni entre CoG et KURT.

Dans les plosives, CoG et SKEW sont fortement corrélés négativement, de même que SD et KURT. KURT est modérément corrélé avec CoG (négativement) et avec SKEW (positivement). Une faible corrélation positive existe également entre CoG et SD. Il n'y a pas de corrélation significative entre SD et SKEW.

Tableau 17 – Coefficients de corrélation de Spearman entre les moments spectraux et probabilités

Fricatives	CoG	SD	SKEW
SD	-0,0148 1,0000		
SKEW	-0,6017 0,0000*	<i>-0,2436</i> 0,0016*	
KURT	-0,0299 1,0000	-0,7558 0,0000*	<i>0,2841</i> 0,0001*
Plosives	CoG	SD	SKEW
SD	<i>0,2713</i> 0,0003*		
SKEW	-0,8566 0,0000*	<i>-0,1216</i> 0,4236	
KURT	-0,5690 0,0000*	-0,8555 0,0000*	0,5272 0,0000*

Note : les astérisques indiquent les corrélations statistiquement significatives ; en gras, les corrélations supérieures à 0,5 ; en italique, les corrélations significatives mais faibles [HINKLE 03]

C. Discussion

Reproductibilité

Les moments spectraux calculés sur chaque phonème ne diffèrent pas entre les enregistrements réalisés dans une pièce insonorisée et ceux effectués dans un bureau calme avec le même micro. Ce résultat est important, car il indique que les mesures des moments spectraux sont robustes et ne sont pas significativement influencées par l'environnement

²⁹ Les lignes directrices de [HINKLE 03] ont été utilisées pour interpréter les coefficients de corrélation : 0,00 à 0,30 = peu ou pas de corrélation ; 0,31 à 0,50 = faible corrélation ; 0,51 à 0,70 = corrélation modérée ; 0,71 à 0,90 = forte corrélation ; 0,91-1,00 = très forte corrélation.

d'enregistrement. Elles pourraient donc être utilisées dans un contexte clinique, à condition toutefois que les enregistrements soient effectués avec un microphone de bonne qualité, dans une pièce relativement silencieuse.

Effet du genre

Malheureusement, la plupart des études récentes sur les moments spectraux ne font pas état du facteur genre. Seules quelques études en tiennent compte, et l'effet du genre mesuré est généralement de petite taille [JONGMAN 00; KAY 12; NISSEN 03]. Cependant, ces études utilisent des statistiques paramétriques sur des échantillons limités et ne rapportent pas de comparaisons par phonème, mais plutôt des différences de genre regroupées qui englobent toutes les comparaisons entre les différents phonèmes (p.ex. entre /p/ chez l'homme et /f/ chez la femme). Ceci ne permet pas de conclure si les moments spectraux diffèrent pour chaque phonème entre les locuteurs masculins et féminins. Seul [YOON 15] a analysé les différences de genre par phonème sur un corpus comprenant 192 locuteurs féminins et 493 masculins (TIMIT, anglais américain). L'auteur a rapporté des différences significatives entre les genres pour le CoG et le SD, uniquement pour les sibilantes non-voisées.

Dans notre étude, en considérant chaque phonème séparément, nous n'avons pas identifié de différences significatives entre les moments spectraux chez les locuteurs masculins et féminins. Il est important de souligner que le filtre passe-haut de 1000 Hz utilisé dans notre étude peut atténuer les différences entre les genres. En effet, les différences à basse fréquence (y compris celles liées à la fréquence fondamentale) sont effacées, ce qui ne laisse que des différences éventuelles entre les genres dans les régions des hautes fréquences.

Emplacement de la fenêtre d'analyse

Les différences mesurées entre les emplacements de la fenêtre d'analyse au début, au milieu et à la fin des consonnes ne donnent pas lieu à un pattern spécifique. Dans les fricatives, SD et KURT semblent les plus stables dans le temps, alors que CoG et SKEW montrent le plus de variations. Dans les plosives, tous les moments spectraux sont caractérisés par une plus grande variabilité dans le temps.

À la lumière des nombreuses différences, malgré le fait que la segmentation manuelle visait à extraire une partie stable de la consonne, nous avons décidé de poursuivre les analyses ultérieures sur le point médian de la consonne. Cette décision est basée sur deux faits.

Premièrement, à l'exception de KURT dans les fricatives non voisées, les mesures du point médian permettent des distinctions entre phonèmes qui sont satisfaisantes à la fois par leur nombre (par rapport au début des consonnes) et par leur concordance avec les données des résultats précédents dans la littérature.

Deuxièmement, alors qu'il n'existe pas de consensus [MEYER 11], de nombreux auteurs conseillent l'utilisation de fenêtres d'analyse centrales pour éviter les effets de « contamination » coarticulatoire (p. ex. [FLIPSEN 99; GIBSON 19; GORDON 02; HERNANDEZ 19; JANNEDY 17; KOENIG 13; MEYER 11; NIRGIANAKI 14; NISSEN 05; TJADEN 17]). Dans les fricatives, notamment, les articulateurs atteignent leur point le plus proche au milieu, puis restent relativement constants jusqu'aux 20 % finaux de la fricative, avant de se déplacer vers la position articulaire de la voyelle suivante [ISKAROUS 11; TJADEN 17]. Cette tendance est effectivement observée dans nos données, le pattern convexe du CoG dans les fricatives étant

comparable aux résultats de [ISKAROUS 11], qui observait une augmentation dans la première moitié du /s/ suivie d’une baisse, due aux mouvements de la mâchoire et de la langue. D’autres auteurs ont montré qu’alors que les mesures spectrales sur les fricatives sont plutôt constantes tout au long de leur durée, les pics de haute fréquence sont plus susceptibles d’être détectés au milieu de la consonne [BEHRENS 88; JONGMAN 00].

Contraste de voisement

Sur les graphiques d’évolution temporelle (voir la [Figure 37](#) et la [Figure 38](#)), les patterns pour les consonnes voisées sont comparables à leurs homologues non voisées. Les composantes à haute fréquence (au-dessus de 1000 Hz) semblent donc évoluer de manière similaire à travers les consonnes, indépendamment de leur voisement. Les différences de moments spectraux entre les lieux d’articulation sont également similaires dans les consonnes voisées et dans leurs homologues non voisées (p. ex. le CoG de /f/ diffère de /ʃ/ et /s/, celui de /v/ diffère de /ʒ/ et /z/). Dans les paires voisée/non-voisée, aucune différence statistiquement significative n’a été mesurée (voir [Tableau 16](#)). Nous pouvons donc conclure que l’effet du voisement dans les basses fréquences a été modéré avec succès par le filtre passe-haut, et que les différences mesurées sont plus susceptibles d’être attribuées aux lieux d’articulation. Notons que pour la distinction de voisement, des mesures efficaces existent déjà, dont le célèbre VOT dans les plosives [ABRAMSON 17; CHO 19; WINN 20]. D’autres mesures, telles que la fréquence fondamentale en début de la voyelle suivant la consonne, peuvent être utilisées pour compléter le VOT dans certaines langues (par exemple, le japonais [KONG 12]). Les études portant sur les distinctions de voisement dans les fricatives sont plus rares [MANIWA 09], mais la durée et l’amplitude du bruit se sont avérées être des indices efficaces [BEHRENS 88; JONGMAN 00; NIRGIANAKI 14; VAN DEN BERG 89].

Lieu d’articulation

Nous examinerons les résultats de CoG et de SKEW ainsi que de SD et de KURT conjointement, étant donné que ces moments sont théoriquement liés [CATTELAÏN 19; NEWMAN 03; TJADEN 10]. Ceci a été confirmé par nos mesures de corrélation, détaillées ci-après.

1. Dans les fricatives

Le lecteur trouvera en [Figure 40](#) une représentation schématique des moments spectraux selon le lieu d’articulation des fricatives.

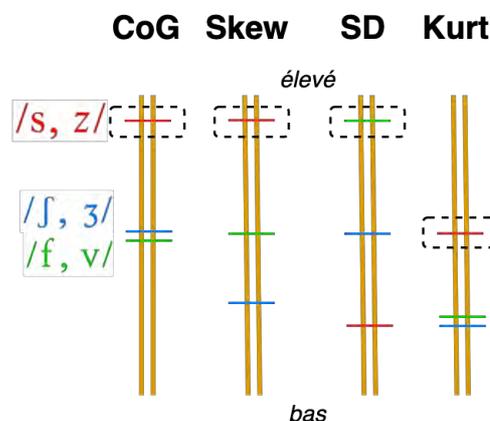


Figure 40 – Représentation schématique des moments spectraux en fonction du lieu d’articulation pour les fricatives ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres

Le **centre de gravité (CoG)** donne des valeurs plus élevées dans /s, z/ (comme dans l'étude française de [KAHN 11]). Ceci conforte l'hypothèse selon laquelle le CoG reflète le point de constriction plus antérieur des fricatives alvéolaires /s, z/ par rapport aux palato-alvéolaires /ʃ, ʒ/ [HERNANDEZ 19; JONGMAN 00; MEYER 11; NISSEN 03; REIDY 15; SCARBEL 12; TJADEN 17].

L'**asymétrie (SKEW)** est également significativement plus élevée dans /s, z/. Alors que les fricatives alvéolaires /s, z/ ont un SKEW positif, les fricatives palato-alvéolaires /ʃ, ʒ/ ont un SKEW négatif, et les labiodentales un SKEW proche de 0.

Si nous considérons les valeurs CoG et SKEW ensemble, le noyau spectral n'est donc pas situé dans les basses fréquences (puisque le CoG est élevé), mais la queue sur sa droite (les valeurs au-dessus de la moyenne) s'étend plus largement que pour les autres fricatives. Nos résultats sont en accord avec les résultats d'autres études, parmi lesquelles [TOMIAK 90], [AVERY 96] et [SCARBEL 12], qui ont également trouvé des valeurs de SKEW positives plus élevées pour /s/ combinées à un CoG plus élevé.

En ce qui concerne les **corrélations** entre le premier et le troisième moment spectral dans les fricatives, la corrélation négative entre le CoG et le SKEW indique *a priori* qu'un centre de gravité plus élevé est associé à une asymétrie plus faible ou négative, et inversement. Bien que ce résultat soit en accord avec les données de la littérature [TJADEN 10], il ne coïncide pas avec les résultats mentionnés ci-dessus (CoG plus élevé pour /s, z/ ainsi qu'une asymétrie positive). Suite à ces résultats statistiques quelque peu contradictoires, nous avons examiné les spectres lissés (voir tous les spectres en [Annexe N](#), et un exemple de spectre en [Figure 41](#)) afin de mieux comprendre le phénomène.

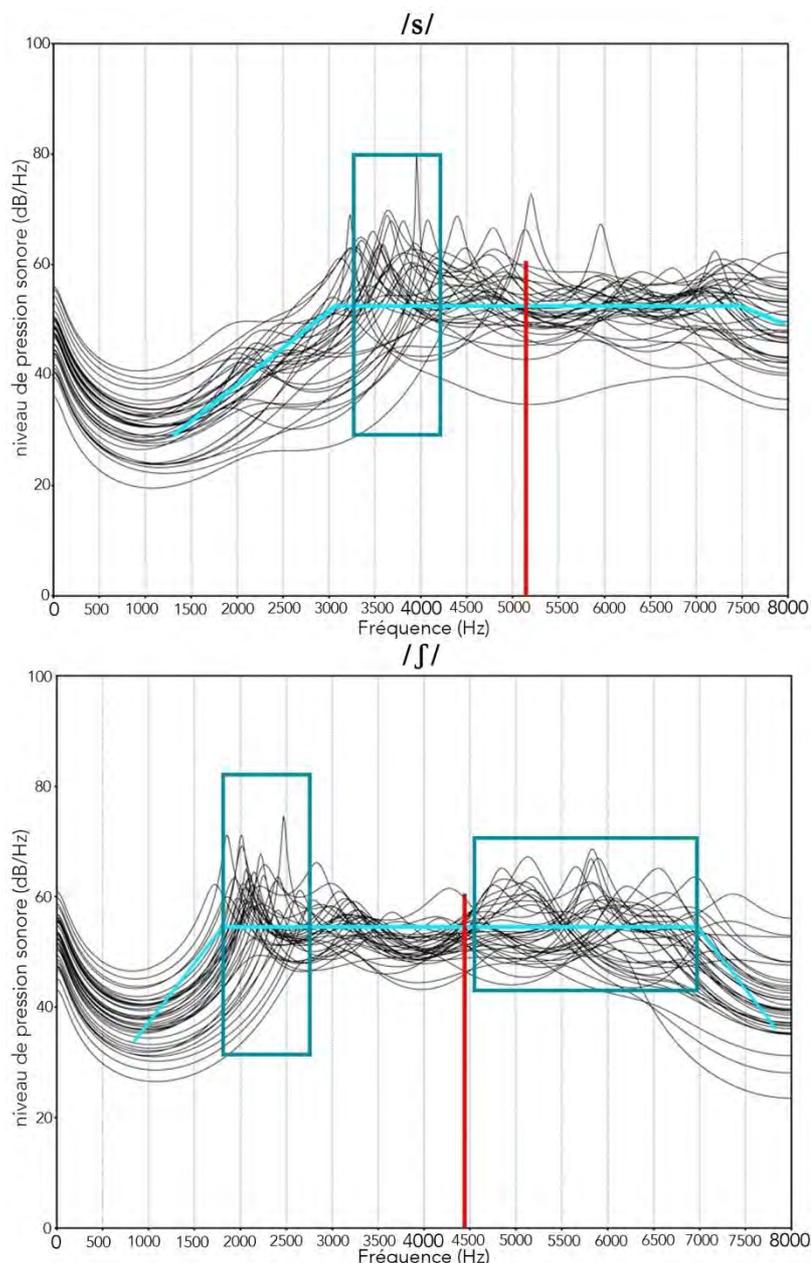


Figure 41 – Représentation schématique du plateau (bleu clair) et des composantes spectrales principales à l'intérieur de celui-ci (bleu foncé) ainsi que du CoG (rouge), pour le /s/ (haut) et pour le /ʃ/ (bas)

Tout d'abord, nous remarquons que les spectres des fricatives sont caractérisés par un plateau dans les hautes fréquences. Les deux plateaux montrent cependant des pics accentués en leur début. Pour /ʃ, ʒ/, nous observons également un léger regain d'énergie après 5000 Hz, avant que la « queue droite » ne soit interrompue autour de 7000 Hz.

Par conséquent, dans l'ensemble, le noyau spectral de /s, z/ (et donc le CoG) est plus élevé (c.-à-d. entre 3000 et 7500 Hz contre 2000 et 7000 Hz pour /ʃ, ʒ/), mais la majeure partie de l'énergie (mode) est située à gauche de ce plateau, avec des valeurs au-dessus du mode s'étendant plus largement que pour /ʃ, ʒ/. Pour /ʃ, ʒ/, non seulement la queue à droite du mode est plus soudainement interrompue dans les fréquences aiguës, mais le léger regain d'énergie dans sa partie droite amenuise la différence avec le pic énergétique de gauche, pouvant expliquer l'asymétrie plus négative.

La configuration particulière des spectres des fricatives, et plus particulièrement celle des palato-alvéolaires, peut être mise en parallèle avec la théorie d'une double source sonore pour les sibilantes, l'une au niveau de la constriction linguale et l'autre au niveau des dents [FANT 60]. De par ce fait, SKEW doit être interprété avec prudence et ne semble pas être une mesure discriminante fiable à elle seule pour les fricatives.

La **déviati**on standard (SD) présente des valeurs plus élevées pour /f,v/ par rapport à /ʃ,ʒ/ et /s,z/. Les fricatives labiodentales montrent donc une plus grande dispersion spectrale autour du centre de gravité (un spectre relativement plat, voir [Figure 42](#)), tandis que les fricatives alvéolaires et palato-alvéolaires sont caractérisées par des formes spectrales plus marquées. Cette observation corrobore les conclusions de [JONGMAN 00; F. LI 10; NISSEN 03, 05; REPP 89; SCARBEL 12], entre autres.

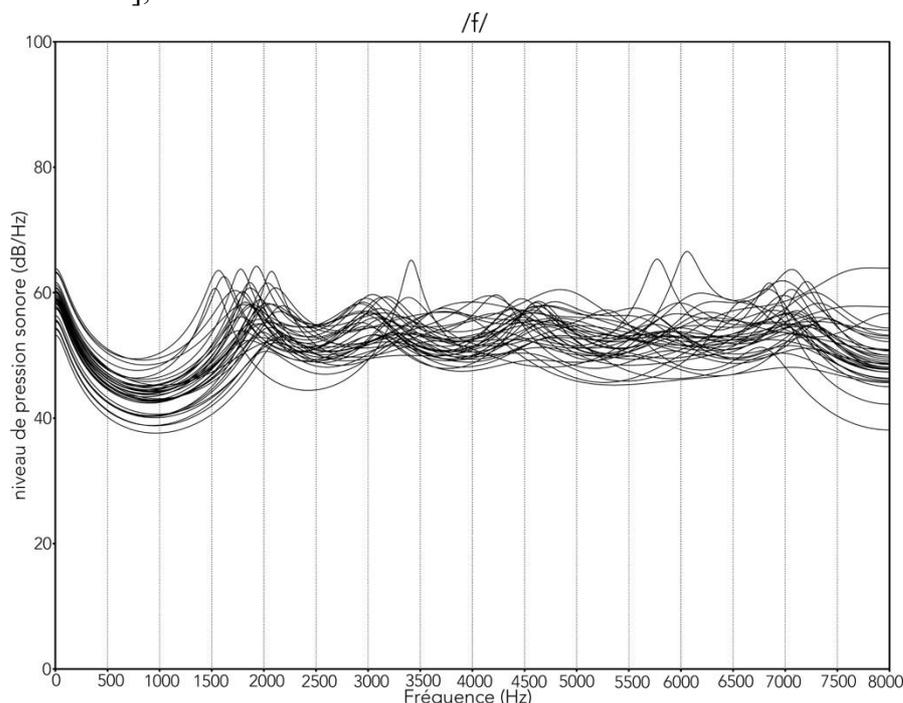


Figure 42 – Spectre plat de la fricative labiodentale /f/

Pour le quatrième moment spectral, le **kurtosis (KURT)**, les alvéolaires /s,z/ montrent une valeur positive suggérant un spectre avec un pic relativement distinct. Cette valeur diffère de la valeur négative des palato-alvéolaires /ʃ,ʒ/, qui présentent un spectre plus plat avec un pic moins clairement défini. Alors que cette mesure semble moins sensible ([CATTELAINE 19] et [ACHER 14] ont également conclu que KURT n'était pas une mesure suffisante pour distinguer les lieux d'articulation), les valeurs sont en accord avec les résultats de SD.

La **corrél**ation négative entre SD et KURT indique qu'une dispersion élevée du noyau spectral autour du CoG est associée à un spectre plus plat.

2. Dans les plosives

Le lecteur trouvera en [Figure 43](#) une représentation schématique des moments spectraux selon le lieu d'articulation des fricatives.

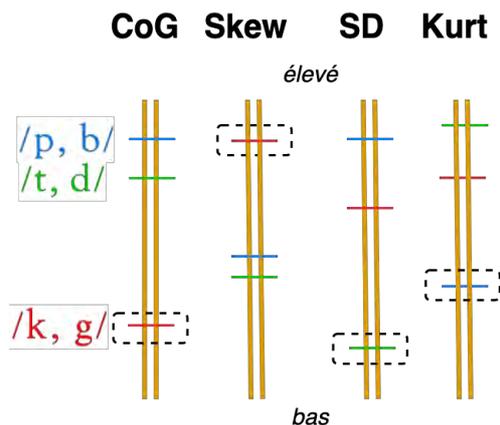


Figure 43 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les plosives ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres

Le **CoG** est significativement plus bas pour les consonnes palatales /k, g/ par rapport aux consonnes alvéolaires /t, d/ et bilabiales /p, b/. Ce résultat est en accord avec les données de la littérature [HALLE 57; TJADEN 17], dont également les études en français de [BONNEAU 08; CATTELAÏN 19] et [BONNEAU 08; CATTELAÏN 19].

Les résultats pour **SKEW** montrent que toutes les plosives sont positivement asymétriques, mais que les consonnes palatales /k, g/ ont une asymétrie positive plus élevée par rapport aux consonnes bilabiales /p, b/ et aux alvéolaires /t, d/.

Pris ensemble, les premier et troisième moments spectraux indiquent donc une concentration du noyau spectral dans les basses fréquences pour les consonnes palatales. Nous notons également que CoG et SKEW sont à nouveau fortement **corrélés** négativement, un centre de gravité plus élevé étant associé à une asymétrie plus négative.

Sur les spectres lissés des plosives, en comparaison à ceux des fricatives, nous observons des zones spectrales de haute énergie plus nettes pour /t, d/ et /k, g/ (autour de 3500 Hz et 1500 Hz, respectivement), sans vrais plateaux (voir [Figure 44](#)).

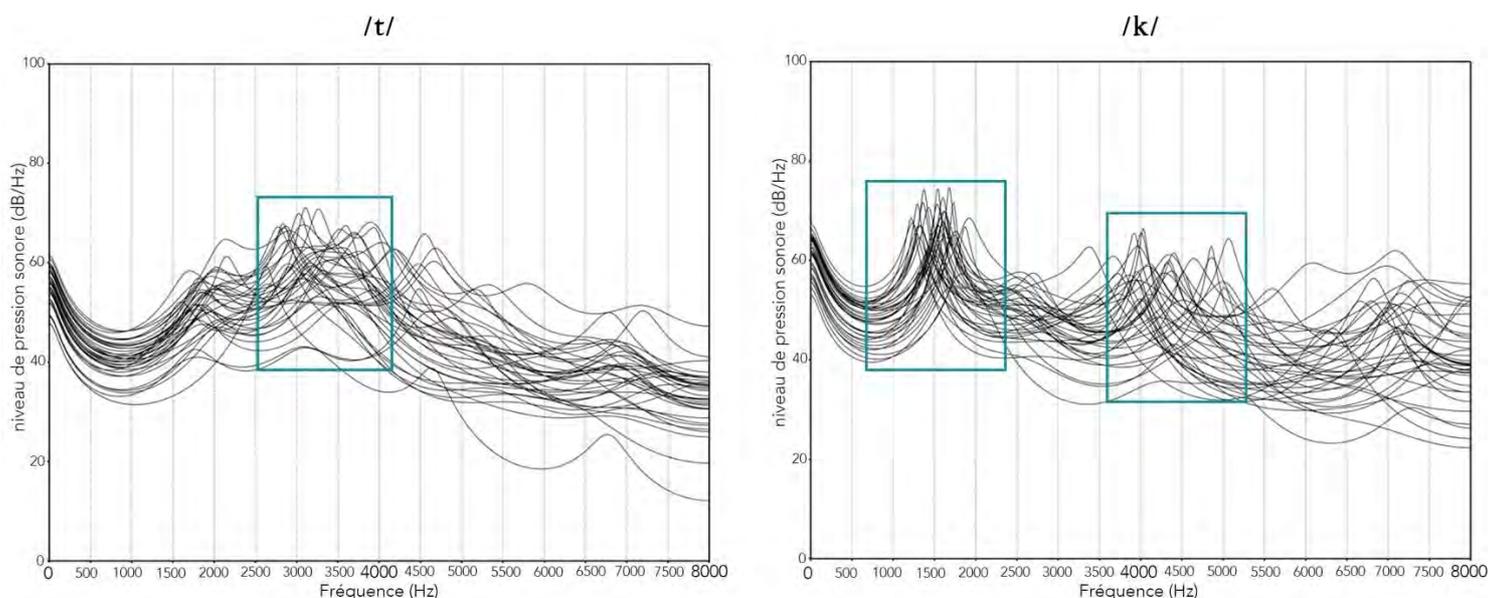


Figure 44 – Composantes spectrales principales pour /t/ (gauche) et pour /k/ (droite)

Au vu de leurs spectres, pour les plosives, un CoG plus élevé lié à une asymétrie faible ou négative (et inversement) est donc plus facilement interprétable ; tant le CoG que le SKEW sont individuellement informatifs dans les plosives, contrairement aux fricatives.

En ce qui concerne **SD**, alors que les consonnes alvéolaires /t,d/ sont caractérisées par des formes spectrales plus distinctes, les plosives palatales et bilabiales, avec une déviation standard plus importante, présentent un spectre plus plat — d'autant plus pour ces dernières (voir [Figure 45](#)). Pour les plosives palatales, une deuxième zone d'énergie spectrale autour de 4500 Hz peut expliquer le SD élevé malgré un pic net autour de 1500 Hz (voir [Figure 44](#)).

Enfin, les consonnes bilabiales /p,b/ ont un **KURT** négatif, indiquant un spectre plat, tandis que les autres consonnes ont des valeurs positives indiquant un pic plus net, d'autant plus pour les alvéolaires. Ces observations sont également faites dans d'autres études (p. ex. [ALI 01; BLUMSTEIN 79; FANT 73]), dont celle de [BONNEAU 08] investiguant les pics spectraux dans les plosives en français.

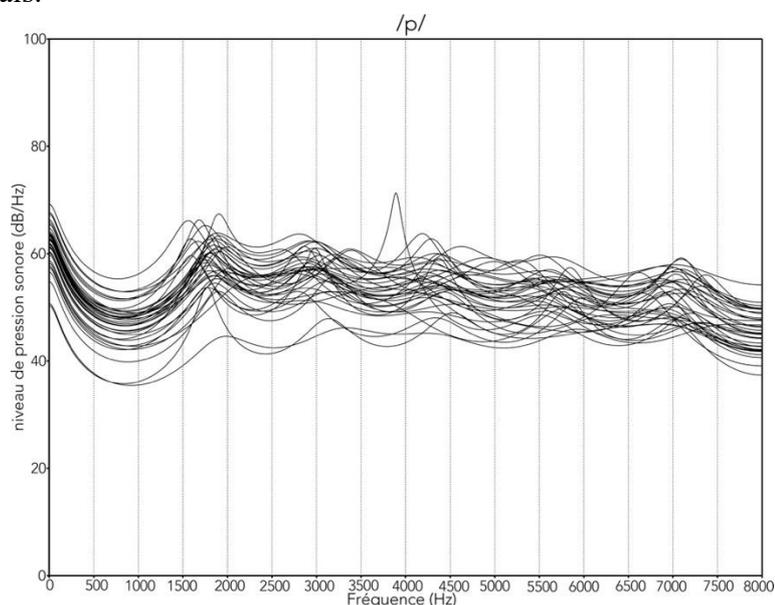
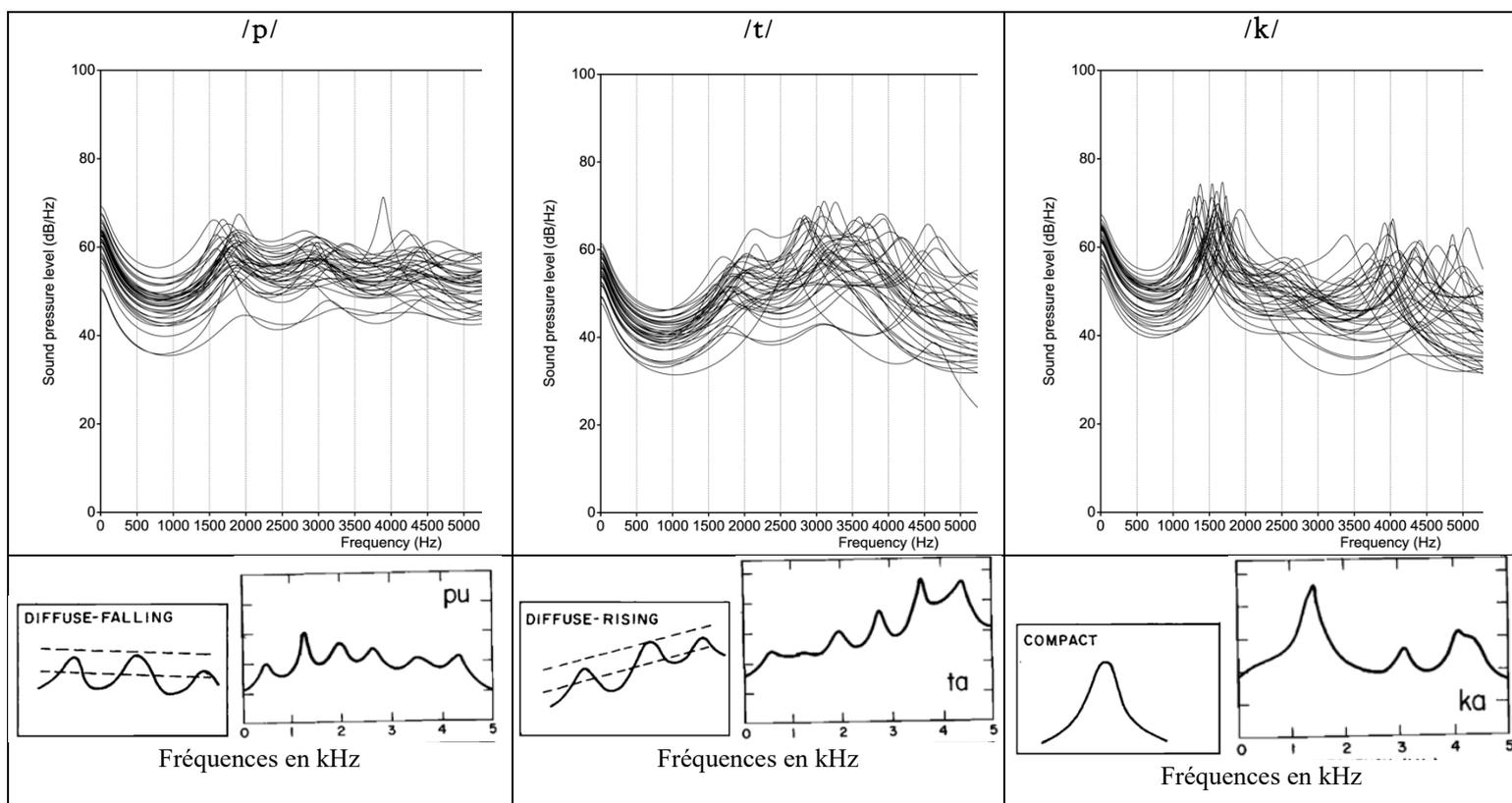


Figure 45 – Spectre plat de la plosive bilabiale /p/

SD et KURT sont à nouveau négativement **corrélés**, ce qui indique qu'une plus grande dispersion du noyau spectral autour du CoG est associée à un spectre plus plat. [SUNDARA 05] a également trouvé cette forte corrélation négative et a émis l'hypothèse que SD et KURT pourraient tous deux refléter un mouvement articulaire sous-jacent commun. La corrélation négative modérée que nous mesurons entre KURT et CoG, quant à elle, indique que plus la concentration spectrale est élevée dans les fréquences, moins la distribution est pointue. Cela peut s'expliquer en partie par le « roll-off » spectral (les hautes fréquences ont généralement moins d'énergie que les basses fréquences [BEHRMAN 18]), dont l'influence sur la perception des consonnes plosives a été montrée [ALEXANDER 09].

Notons que nos observations sur les plosives peuvent être mises en parallèle avec les trois patterns de bursts identifiés par [BLUMSTEIN 79] (voir [Tableau 18](#) pour une comparaison visuelle) : les bilabiales présentent un spectre plat, alors que les palatales présentent un pic plus net dans les fréquences moyennes. Les alvéolaires, elles, montrent des pics au-delà de 2000 Hz, avec une énergie spectrale concentrée dans les fréquences hautes, résultant en un spectre globalement « croissant ».

Tableau 18 – Comparaison des spectres des bursts des plosives dans notre étude (rangée supérieure) avec les patterns identifiés par [BLUMSTEIN 79] (rangée du bas)



Note : nos spectres ont été coupés pour correspondre à la même plage fréquentielle que les représentations de Blumstein et Stevens ; spectres des plosives non voisées de notre étude, ceux des voisées étant similaires

D. Conclusion

Globalement, le CoG et le SD semblent être les mesures les plus précises et fiables des moments spectraux. Le CoG est significativement plus élevé dans les fricatives alvéolaires et plus bas dans les plosives palatales. SD, d'autre part, est significativement plus élevé dans les fricatives labiodentales et les plosives bilabiales, alors que les plosives alvéolaires montrent une dispersion moins importante du noyau spectral.

SKEW doit être interprété avec prudence, surtout dans les fricatives, à la lumière de leurs formes spectrales particulières. KURT, en plus de montrer de nombreuses valeurs aberrantes dans les fricatives et les plosives, semble également moins sensible que SD dans les fricatives. Dans l'ensemble, nos résultats corroborent donc la conclusion de [FORREST 88] sur les plosives anglaises : afin de distinguer de manière consistante les lieux d'articulation, une mesure simultanée des quatre moments est nécessaire. Ainsi, une combinaison des deux paires CoG-SKEW et SD-KURT semble être la meilleure recommandation. Ces résultats nous amènent à conclure que les moments spectraux, tout en reflétant la variabilité de la parole saine, permettent malgré tout par leur combinaison la distinction des phonèmes prononcés. Nos données sont pertinentes pour des études ultérieures, notamment pour investiguer l'utilisation des moments spectraux en parole pathologique en français.

E. Perspectives

Les résultats concernant les moments spectraux dans la littérature sont variables, ce qui peut s'expliquer à la fois par des différences interlinguistiques, ainsi que par des disparités méthodologiques [LOUSADA 12]. Au vu du peu d'études disponibles pour les consonnes françaises, il convient donc de souligner le besoin d'études supplémentaires sur de grands ensembles de données, avec des descriptions méthodologiques détaillées.

Comme un effet de l'âge a été observé sur les moments spectraux [TAYLOR 20], des études ultérieures avec des échantillons plus importants pour différents groupes d'âge devraient également prendre ce facteur en compte. De plus, il serait intéressant d'investiguer plus en détail le facteur genre. En effet, nous n'avons pas observé dans notre étude de différences significatives entre les genres pour chaque phonème. Les distinctions entre phonèmes pourraient néanmoins être caractéristiques du genre du locuteur.

À travers la coarticulation, l'acoustique des consonnes — dont les moments spectraux [KATZ 91] — est influencée par le contexte phonémique. Ceci a été confirmé dans une étude pilote menée dans le cadre du stage d'initiation à la recherche d'Anthéa Frize — étudiante au Centre de Formation Universitaire en Orthophonie de Toulouse — que j'ai encadrée. Les résultats de cette étude sont résumés en Annexe O. Au vu de l'influence de la coarticulation [VAN SON 96, 99], mais aussi du type de parole (lue ou spontanée), les résultats rapportés dans cette étude ne doivent pas être généralisés à la parole spontanée. D'autres études sont notamment nécessaires pour explorer l'influence des voyelles et des consonnes voisines sur les moments spectraux d'une consonne cible. Cette question sera brièvement abordée dans le Chapitre 5. À des fins cliniques, pour l'instant, un ensemble de phrases avec un contexte vocalique contrôlé semble donc le plus pertinent pour maximiser la fiabilité des mesures. De plus, à la lumière de l'importante variabilité inter- et intralocuteur des mesures acoustiques, pour l'utilisation dans un contexte clinique, il est crucial d'utiliser les mêmes conditions d'enregistrement et paramètres d'analyse pour les mesures pré et post-thérapeutiques.

Enfin, nous tenons à souligner que les fenêtres d'analyse choisies sur base de la littérature scientifique (40 ms pour les fricatives, 20 ms pour les plosives) nous semblent *a posteriori* assez larges, surtout pour les plosives. En effet, alors que le burst d'une plosive peut durer jusqu'à 20 ms, sa première partie (« transient ») est un événement très court — d'environ 5 à 10 ms. L'utilisation d'une fenêtre de 20 ms peut notamment expliquer les valeurs plus variables dans le temps des moments spectraux que nous avons mesurées sur les plosives par rapport aux fricatives. C'est pourquoi, dans notre étude subséquente utilisant l'alignement forcé, nous avons tenu à réduire cette fenêtre d'analyse.

Toutes nos consonnes ont été segmentées manuellement, ce qui pose non seulement le problème de la subjectivité, mais aussi de l'aspect chronophage en vue d'une éventuelle utilisation en pratique clinique. Les outils d'alignement phonémique automatique, bien qu'ils doivent encore être utilisés « comme des outils, et non comme le remplacement complet d'un chercheur spécialisé » [MACKENZIE 20], montrent aujourd'hui des résultats prometteurs, y compris en français [GOLDMAN 11].

Ainsi, nous avons souhaité appliquer un alignement forcé sur nos stimuli de parole, afin d'investiguer si une extraction automatisée des moments spectraux est faisable, et si les résultats diffèrent de ceux obtenus jusque-là en segmentation manuelle.

4.3.2. Moments spectraux sur alignement forcé

Nous allons tout d’abord brièvement décrire l’alignement forcé, utilisé pour identifier de manière automatique les segments phonémiques du signal audio. Ensuite, nous décrirons l’algorithme qui nous a permis de détecter les parties stables à l’intérieur de ces segments pour les fricatives, ainsi que les bursts pour les plosives. Nous analyserons ensuite les moments spectraux obtenus à l’aide de ces outils et les comparerons à ceux tirés de la segmentation manuelle.

A. L’alignement forcé

Nous avons tout d’abord collaboré avec Imed Laaridh et Corinne Fredouille du Laboratoire d’Informatique d’Avignon (LIA), qui nous ont fourni des alignements forcés [BÜRKI 08] des phrases porteuses.

L’alignement forcé consiste à mettre en face du signal acoustique une séquence de phonèmes connue (voir [Figure 46](#)). Nous fournissons à l’outil d’alignement une transcription orthographique ou phonémique du contenu de parole, contrairement à la reconnaissance automatique où la séquence n’est pas connue *a priori*.

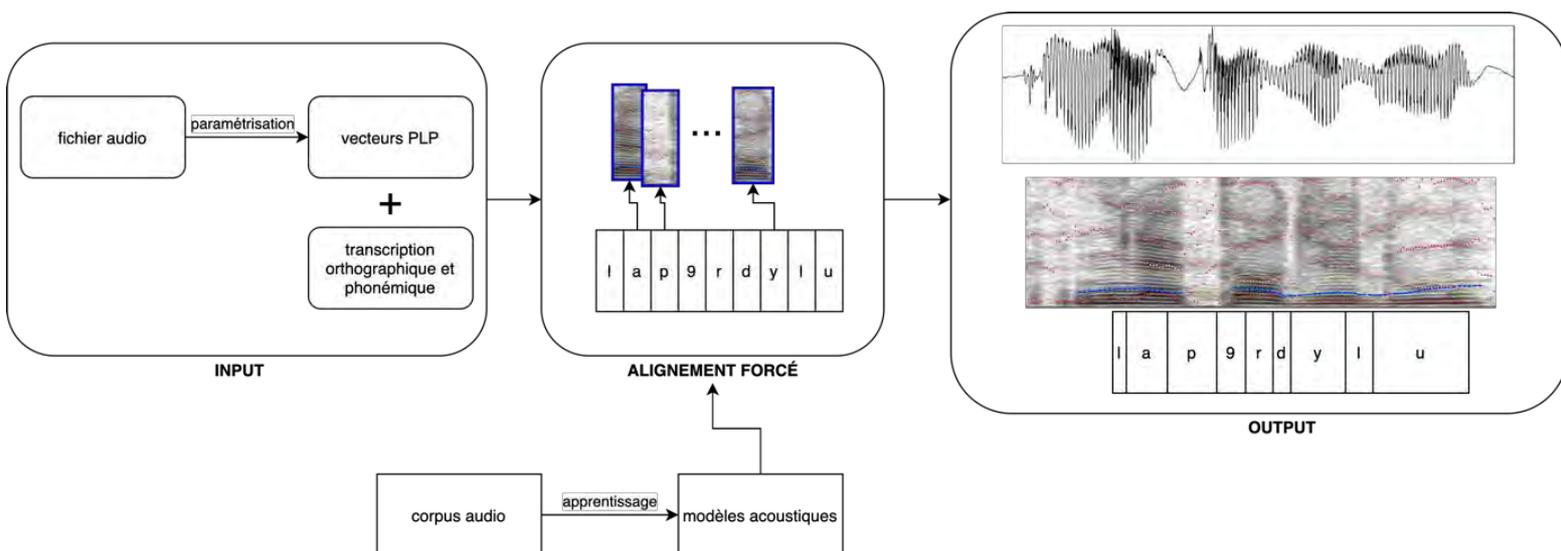


Figure 46 – Représentation schématique de l’alignement forcé

Les outils d’alignement forcé fonctionnent ensuite à l’aide d’un ou plusieurs modèles acoustiques. Ces modèles sont entraînés sur de nombreuses heures d’enregistrements de parole afin de modéliser chaque phonème et d’apprendre à en reconnaître des séquences. Le système va ainsi identifier les segments temporels dans le signal acoustique qui correspondent au mieux avec les segments phonémiques fournis.

Nous avons observé de plus près les résultats de l’alignement forcé, et avons constaté à plusieurs reprises une inclusion de la transition vocalique précédant la consonne dans le segment d’alignement forcé (voir [Figure 47](#)).

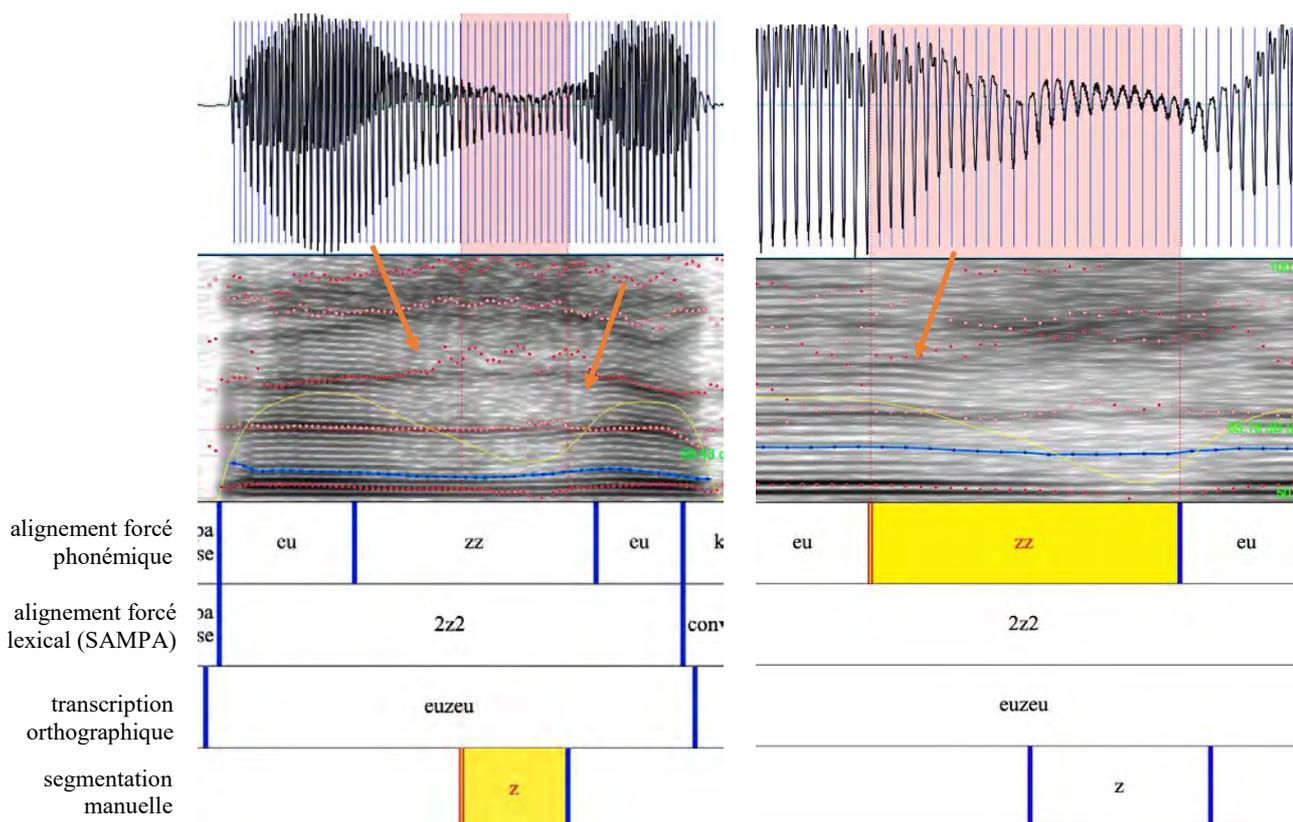


Figure 47 – Exemples de résultat d’alignement forcé avec inclusion de la transition vocalique dans le segment consonantique.

De plus, pour les plosives, l’alignement forcé fournit le segment entier de la plosive, incluant donc la phase de silence, le relâchement (« burst »), suivi du bruit de friction.

Nous avons donc souhaité combiner l’alignement forcé avec une seconde méthode pour affiner la recherche de la fenêtre d’analyse optimale pour les moments spectraux. Il s’agit de la méthode de segmentation « forward-backward divergence », que nous allons maintenant décrire.

B. La méthode « diverg »

L’algorithme « forward-backward divergence segmentation » (FBDS) [ANDRE-OBRECHT 88]³⁰ est destiné à détecter les événements acoustiques (« ruptures ») dans un signal de parole continue. Le signal est considéré comme une suite d’unités stationnaires, chacune caractérisée par un modèle autorégressif gaussien. Des analyses statistiques sont employées pour détecter les changements significatifs de paramètres dans ces unités. Deux modèles sont utilisés. Le modèle M1, correspondant à une courte fenêtre glissante qui démarre à la dernière rupture détectée, est comparé au modèle M0, correspondant au segment « grandissant » après la dernière rupture détectée. Lorsqu’un seuil de divergence entre les deux modèles est atteint, une nouvelle rupture est détectée ; ce seuil peut être adapté pour une détection plus ou moins sensible (un ordre de 2 est utilisé dans la méthode originale ; un ordre plus élevé amène une détection de ruptures plus nombreuses). Cette méthode est appliquée sur le signal en avant (« forward ») et en arrière (« backward ») et le résultat des deux passages est fusionné. Le résultat final est une segmentation subphonémique, tel qu’illustré en [Figure 48](#).

³⁰ <https://www.irit.fr/SAMOVA/site/research/analysis/audio-segmentation/>

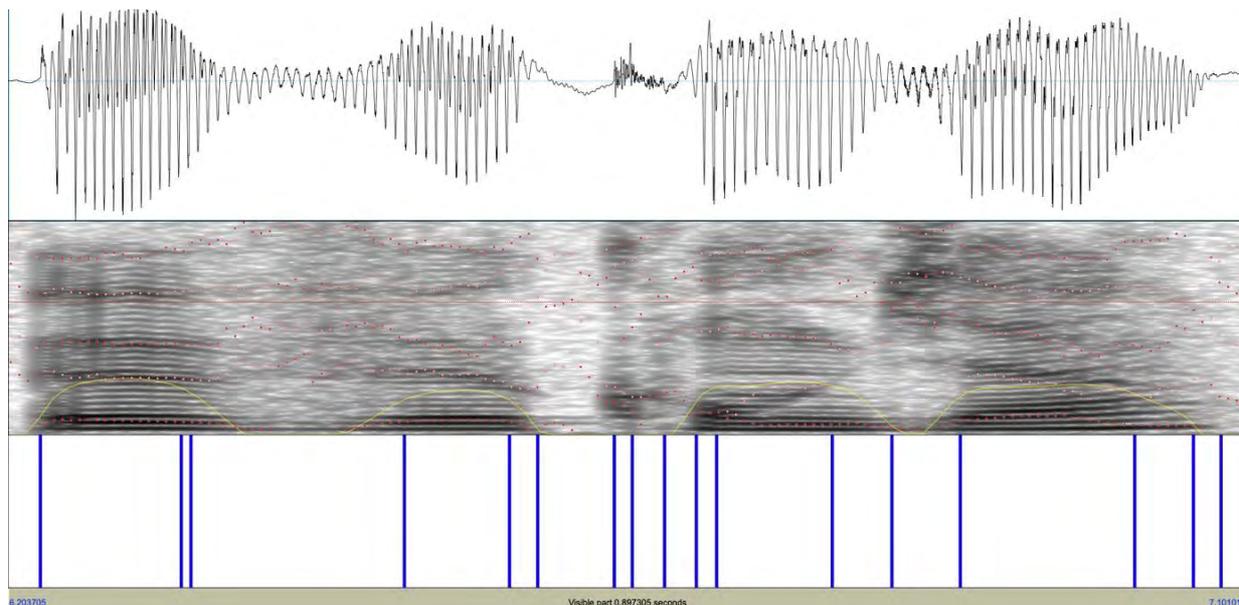


Figure 48 – Résultat de la segmentation subphonémique effectuée à l’aide de la méthode FBDS

Dans ce travail, en collaboration avec Régine André-Obrecht (auteur de la méthode FBDS), nous avons appliqué la détection d’évènements acoustiques afin d’identifier les segments stables du signal de parole. L’objectif était d’affiner le résultat obtenu à l’aide de l’alignement forcé. La première étape consiste en la fusion des résultats de l’alignement forcé et de la méthode FBDS. Cette étape permet de ne conserver que les ruptures acoustiques d’intérêt, c.-à-d. celles qui se situent à l’intérieur des segments consonantiques de l’alignement forcé (voir Figure 49). Pour la méthode FBDS, nous avons opté pour un ordre de sensibilité de 16, qui permet une grande précision de détection d’évènements acoustiques notamment pour les bursts. Étant donné que nous nous limitons aux ruptures à l’intérieur des segments de l’alignement forcé, le nombre important de ruptures détecté ne pose pas de problème.

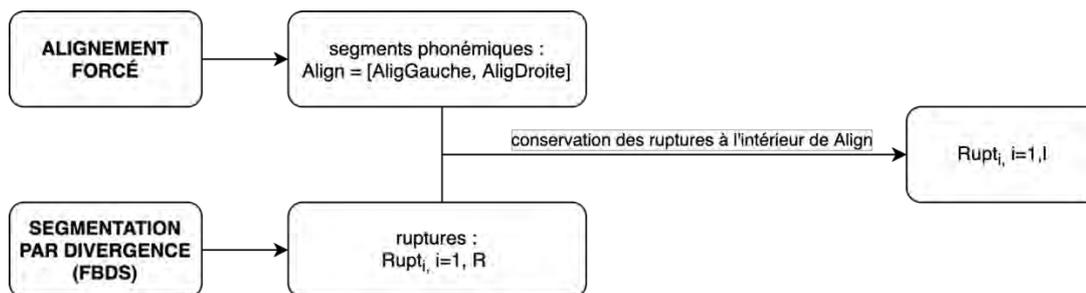


Figure 49 – Fusion de l’alignement forcé et de la méthode FBDS

La seconde étape est déclinée différemment selon le type de consonne. Pour les plosives, nous recherchons l’instant du burst ; pour les fricatives, nous recherchons la partie la plus stable.

Pour les plosives non voisées /p, t, k/ : Pour les plosives non voisées, nous utilisons un seuil d’intensité pour détecter la rupture correspondant au début du burst. Nous calculons tout d’abord l’intensité du point médian correspondant en théorie au segment silencieux de la plosive. Ensuite, le cas échéant, nous calculons l’intensité en dB après chaque rupture suivant ce point médian et la comparons à l’énergie du point médian et aux ruptures voisines. Un exemple de burst détecté par notre algorithme est présenté en Figure 50.

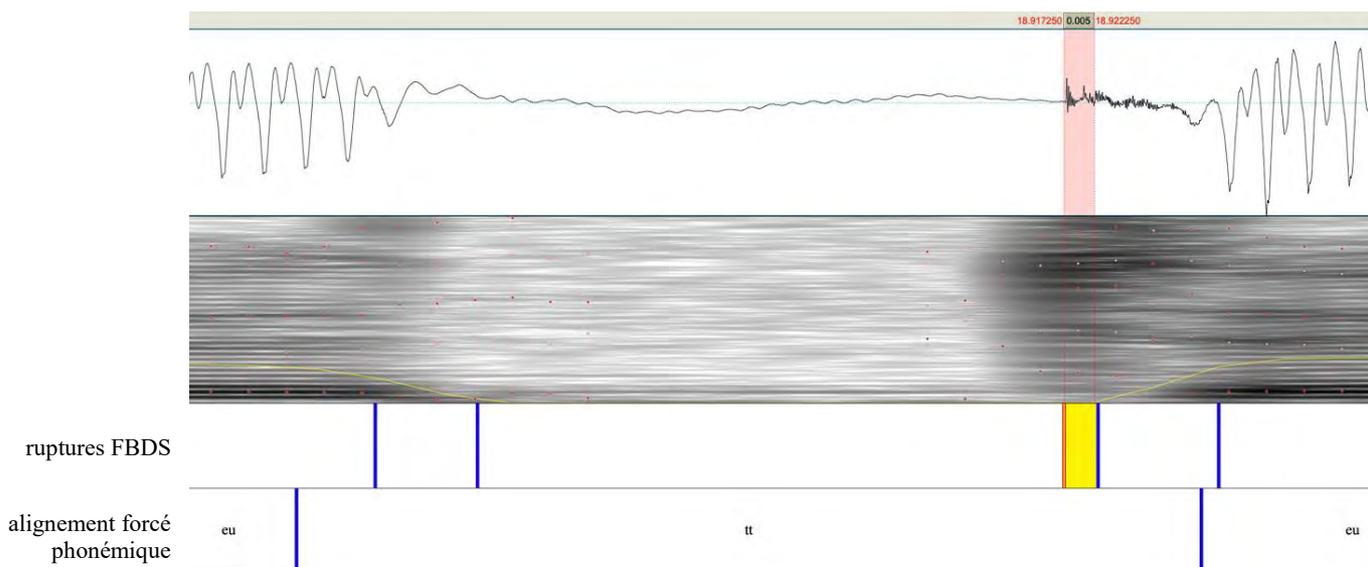


Figure 50 – Exemple d'un burst détecté par notre algorithme pour la plosive /t/ (en rouge, la rupture correspondant au début du burst, ainsi que la fenêtre d'analyse de 5 ms qui la suit)

Pour les plosives voisées /b, d, g/ : Pour trouver le burst dans les plosives voisées, le test d'intensité est peu adapté, non seulement car la phase précédant le burst n'est alors pas un silence, mais aussi parce que l'explosion est généralement moins intense. Nous cherchons donc dans ce cas les ruptures proches de la frontière droite du segment consonantique de l'alignement forcé. En priorité, nous sélectionnons la rupture la plus proche qui précède cette frontière ; en second lieu, nous optons pour la rupture qui suit cette frontière. Un exemple de burst détecté par notre algorithme est présenté en [Figure 51](#).

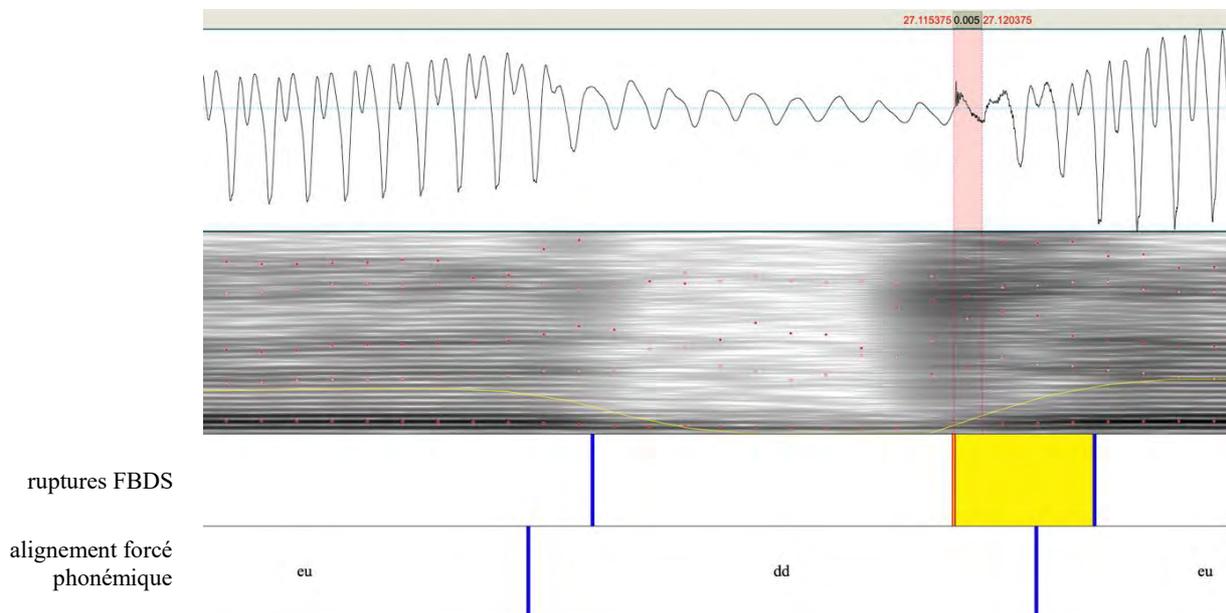


Figure 51 – Exemple d'un burst détecté par notre algorithme pour la plosive /d/ (en rouge, la rupture correspondant au début du burst, ainsi que la fenêtre d'analyse de 5 ms qui la suit)

Pour les fricatives non voisées /f, s, ʃ/ : La partie la plus stable des fricatives non voisées correspond généralement à la distance la plus longue entre deux ruptures à l'intérieur du segment consonantique de l'alignement forcé. Nous évitons ainsi les zones transitoires

vocaliques parfois incluses dans ce segment. Un exemple de fenêtre d'analyse identifié par notre algorithme est présenté en [Figure 52](#).

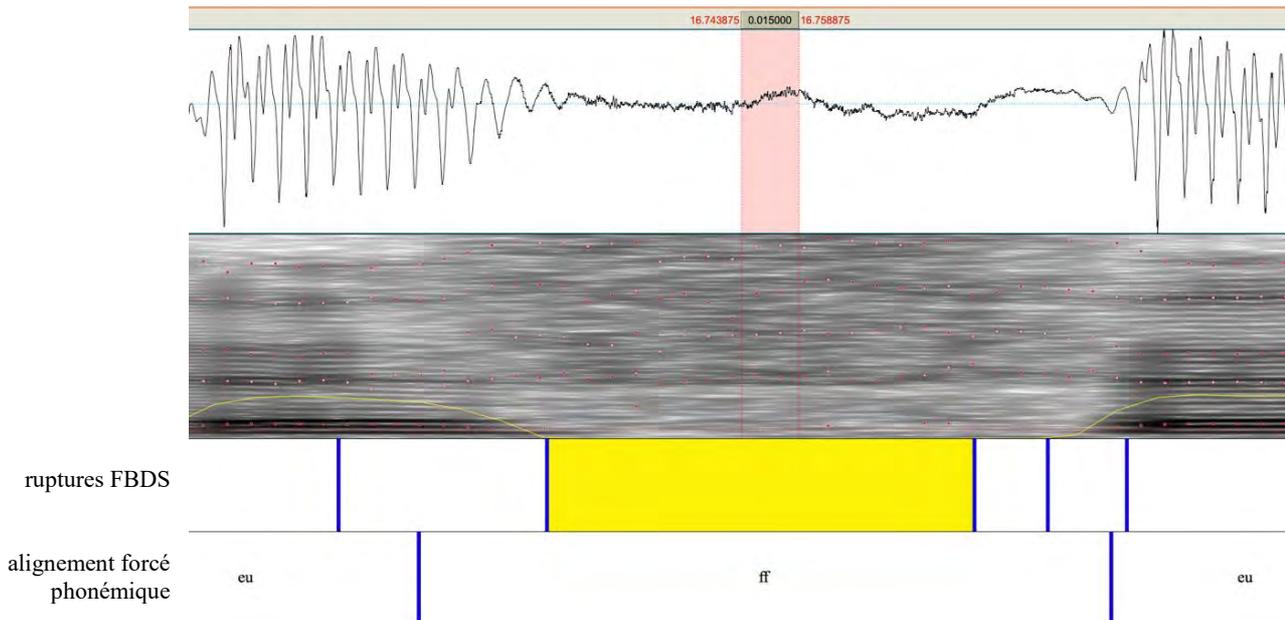


Figure 52 – Exemple d'une fenêtre d'analyse de 10 ms (en rouge) identifiée par notre algorithme pour la fricative /f/, à l'intérieur de la phase stable de la consonne (en jaune)

Pour les fricatives voisées /v, z, ʒ/ : Pour les fricatives voisées, la partie principale de la consonne présente une énergie moindre par rapport aux voyelles voisines. C'est cette baisse d'intensité qui est ciblée dans notre algorithme. Nous effectuons ainsi un test d'intensité sur chaque segment délimité par deux ruptures à l'intérieur du segment consonantique d'alignement forcé, afin d'identifier celui qui présente le minimum d'énergie. Un exemple de fenêtre d'analyse identifié par notre algorithme est présenté en [Figure 53](#).

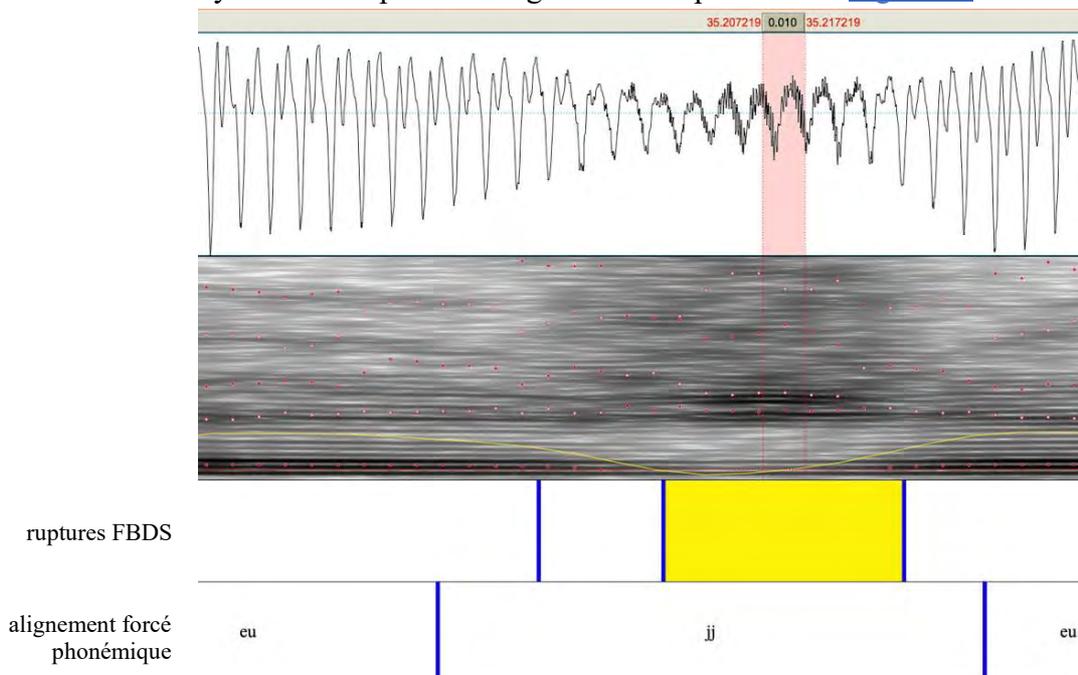


Figure 53 – Exemple d'une fenêtre d'analyse de 10 ms (en rouge) identifiée par notre algorithme pour la fricative /ʒ/ à l'intérieur du segment le plus faible en intensité

L'algorithme pour les quatre types de consonnes est détaillé en [Annexe P](#).

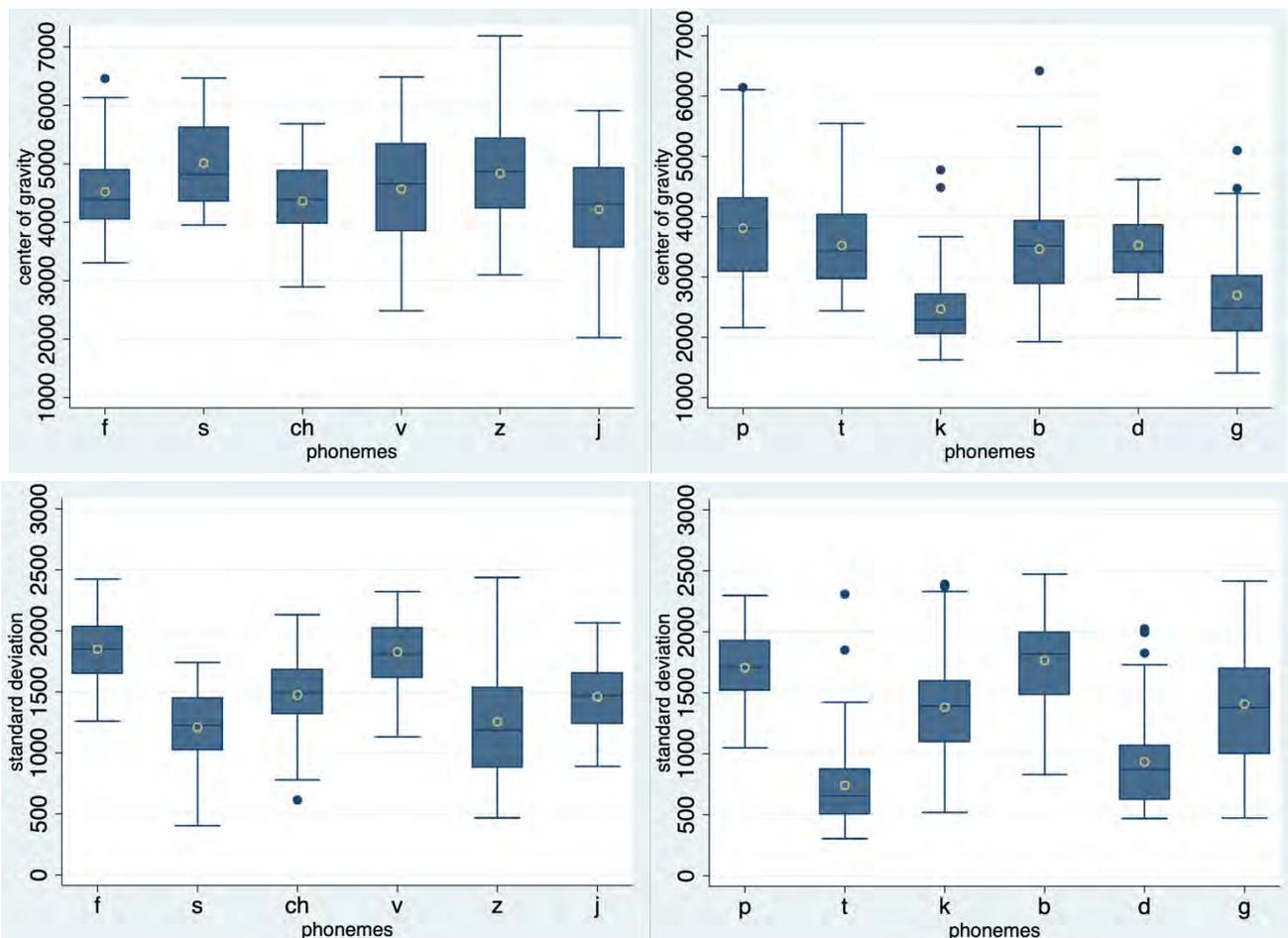
Les moments spectraux ont ensuite été calculés sur les fenêtres identifiées à l'aide de notre nouvel algorithme. Nous allons maintenant analyser les résultats ainsi obtenus et les comparer à ceux issus de la segmentation manuelle³¹. Par souci de lisibilité, la combinaison de l'alignement forcé avec la méthode de segmentation FBDS sera nommée « méthode FBDS » dans la suite de ce manuscrit.

C. Moments spectraux issus de la méthode FBDS

Statistiques descriptives

Six occurrences de phonèmes ont été exclues des analyses à venir à cause de valeurs aberrantes des moments spectraux, dont trois /s/ « sifflants », deux /ʒ/ à cause d'un décalage de l'alignement forcé, et un /d/ pour lequel ni l'écoute, ni l'observation de l'oscillogramme, du spectrogramme et de la détection de fenêtre d'analyse n'ont pu expliquer les valeurs aberrantes.

Les graphiques descriptifs en boîtes à moustaches pour les valeurs obtenues avec la méthode FBDS sont présentés pour les quatre moments pour chaque phonème en [Figure 54](#). Les spectres LPC, tracés avec les mêmes paramètres que pour la segmentation manuelle, sont disponibles en [Annexe R](#).



³¹ Les paramètres d'analyse étaient les mêmes que pour l'analyse sur la segmentation manuelle, excepté la taille de la fenêtre d'analyse qui a été adaptée (10ms pour les fricatives, 5ms pour les bursts des plosives).

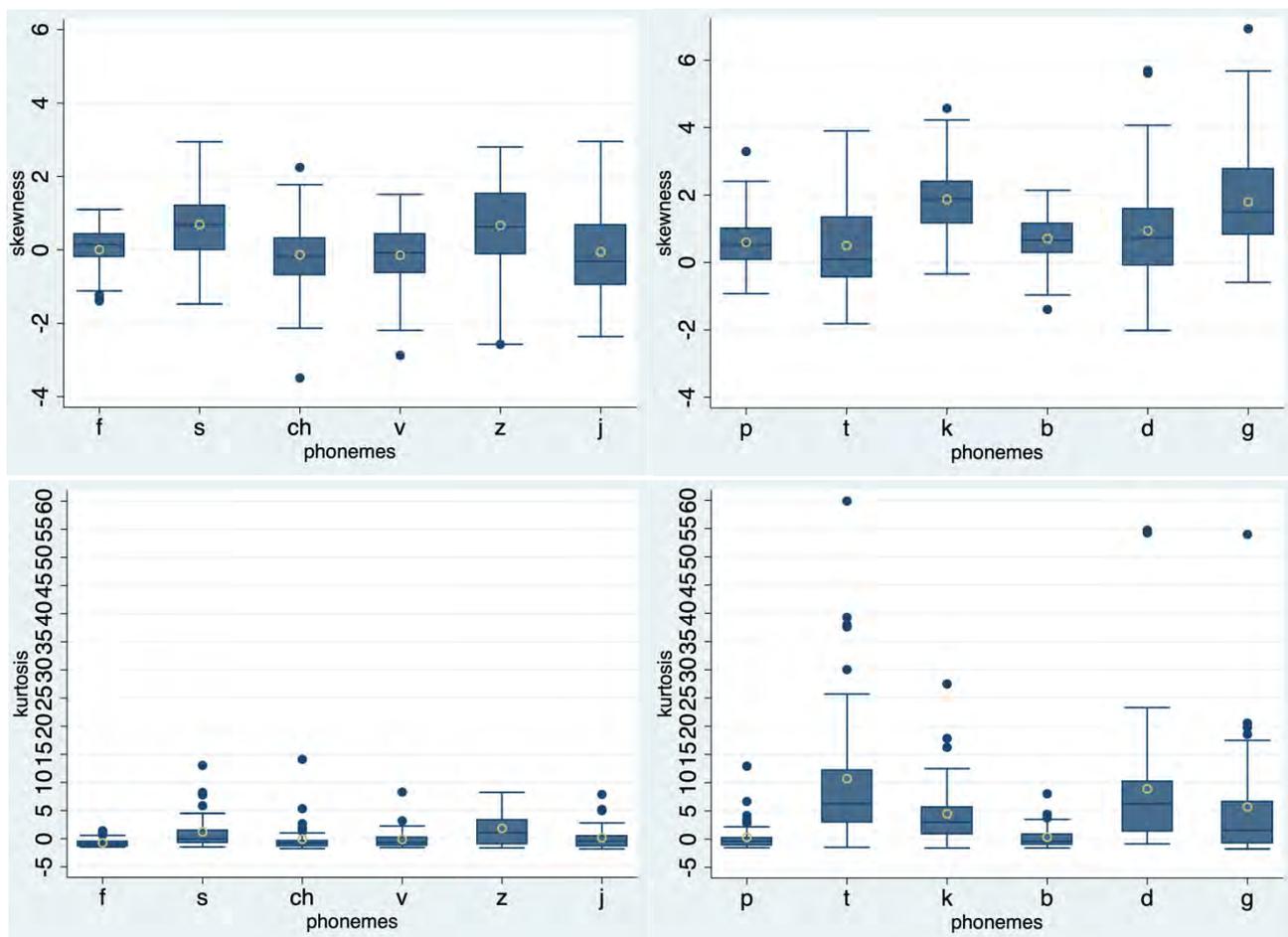


Figure 54 – Graphiques en boîtes à moustaches pour les quatre moments spectraux pour les consonnes plosives et fricatives, issus de la méthode FBDS

Segmentation manuelle vs méthode FBDS

Nous avons souhaité comparer les résultats issus de la segmentation manuelle avec ceux issus de notre nouvel algorithme. Pour cela, nous avons employé le test des rangs signés de Wilcoxon pour chaque moment spectral sur chaque phonème. Des différences significatives ont été trouvées entre les deux méthodes pour le CoG de /d/ ($p=0,0001$), pour le SD de /t/ ($p=0,007$) et de /z/ ($p=0,042$), et pour le KURT de /s/ ($p=0,016$) et de /z/ ($p=0,039$).

Le lecteur trouvera pour une comparaison facilitée en [Annexe Q](#) côte à côte les boîtes à moustaches des valeurs issues de la segmentation manuelle et de la méthode FBDS.

La méthode FBDS donne donc un résultat assez fidèle à la segmentation manuelle. Pour les fricatives, elle permet une détection de la fenêtre d'analyse dans un segment stable de la consonne, en minimisant l'inclusion de transitions vocaliques. Dans l'exemple en [Figure 55](#), alors que l'utilisation de l'alignement forcé seul aurait conduit à calculer les moments spectraux plus proche de la transition vocalique (au centre du segment en jaune), la méthode FBDS permet de considérablement réduire l'inclusion vocalique et calculer les moments spectraux dans les 10 ms centrales de la fricative (en rouge). Pour les plosives, la méthode FBDS donne des résultats concluants pour la détection du burst. Nous allons maintenant analyser de plus près les résultats des moments spectraux issus de la méthode FBDS.

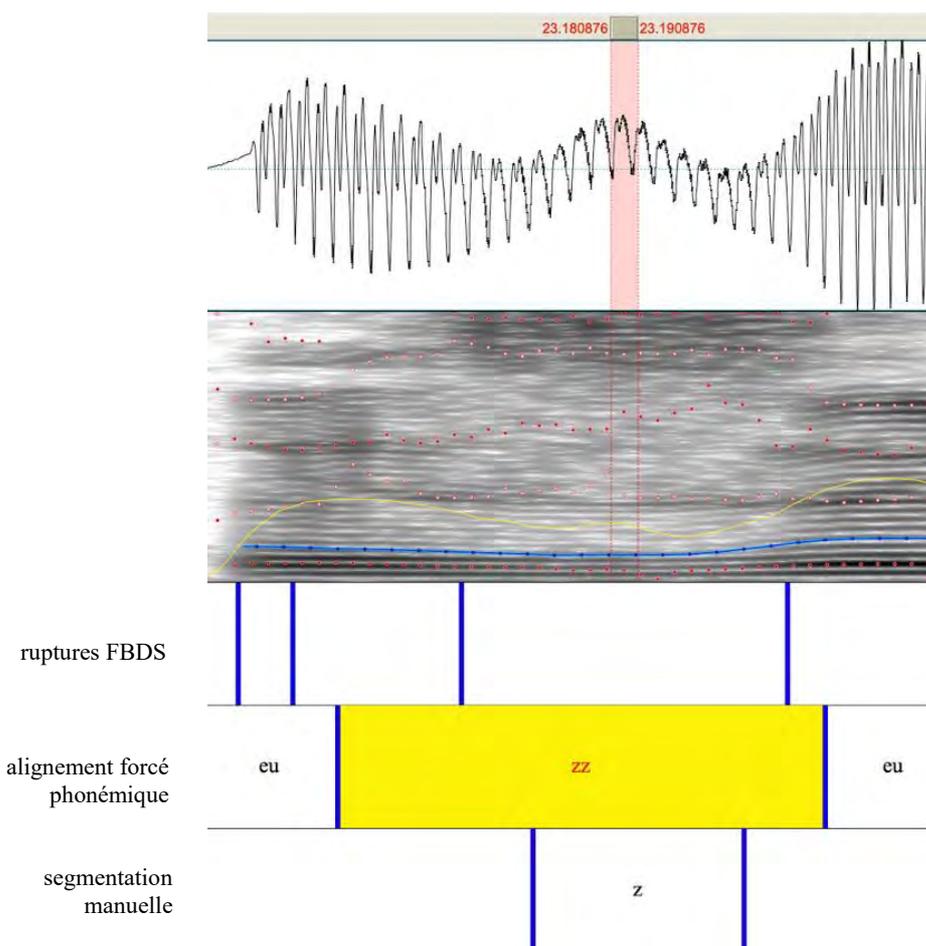


Figure 55 – Exemple d’inclusion de transition vocalique dans l’alignement forcé d’une fricative voisée, corrigée par notre algorithme

Moments spectraux et méthode FBDS - Effet du genre

Nous avons réitéré l’analyse d’effet du genre à l’aide du test de Dunn pour les comparaisons par paires, phonème par phonème. Sur les valeurs issues de notre algorithme, aucune différence statistiquement significative n’a été trouvée entre les genres des locuteurs pour les quatre moments spectraux. Comme pour la segmentation manuelle, les moments spectraux issus pour chaque phonème ne diffèrent donc pas entre les locuteurs masculins et féminins.

Moments spectraux et méthode FBDS - Lieu d’articulation

Nous avons étudié les différences de valeurs des moments spectraux issus de la méthode FBDS selon le lieu d’articulation (fricatives et plosives analysées séparément).

Tous les tests de Kruskal-Wallis pour chaque moment spectral ont donné lieu à une différence statistiquement significative (avec des probabilités de 0,0001 sauf 0,0035 pour le CoG des fricatives). La [Figure 56](#) résume les résultats du test de Dunn, c.-à-d. le(s) moment(s) spectral(aux) qui diffère(nt) entre les lieux d’articulation des fricatives et des plosives voisées et non voisées, respectivement. Aucun moment spectral ne différencie les paires de phonèmes voisé/non voisé.

Le [Tableau 19](#) illustre les moments spectraux qui différencient les fricatives et plosives au point médian consonantiques (voir aussi la [Figure 54](#) pour une comparaison visuelle).

Tableau 19 – Médiane [intervalle interquartile] des moments spectraux pour chaque phonème ; en gras : valeurs significativement différentes des autres sur la même demi-rangée

	s	ʃ	f	z	ʒ	v
CoG	4816 [4357; 5613]	4378 [3984; 4873]	4385 [4056; 4887]	4860 [4241; 5429]	4311 [3569; 4921]	4658 [3856; 5337]
SKEW	0,69 [0,01; 1,21]	-0,21 [-0,67; 0,31]	0,14 [-0,19; 0,43]	0,62 [-0,11; 1,52]	-0,32 [-0,94; 0,67]	-0,09 [-0,62; 0,42]
SD	1222 [1028; 1446]	1489 [1319; 1680]	1852 [1651; 2033]	1187 [885; 1530]	1471 [1240; 1654]	1813 [1691; 2069]
KURT	-0,09 [-0,72; 1,37]	-0,74 [-1,32; 0,39]	-0,96 [-1,25; -0,51]	1,09 [-0,88; 3,31]	-0,47 [-1,24; 0,39]	-0,67 [-1,21; 0,18]

	k	p	t	g	b	d
CoG	2288 [2064; 2705]	3803 [3095; 4301]	3437 [2965; 4025]	2484 [2110; 3022]	3514 [2888; 3930]	3413 [3074; 3863]
SKEW	1,88 [1,17; 2,39]	0,53 [0,08; 1,01]	0,08 [-0,43; 1,33]	1,5 [0,84; 2,77]	0,64 [0,30; 1,14]	0,73 [-0,08; 1,58]
SD	1391 [1101; 1593]	1708 [1523; 1926]	652 [508; 874]	1381 [1001; 1698]	1816 [1485; 1992]	870 [625; 1067]
KURT	2,96 [0,93; 5,56]	-0,48 [-1,1; 0,21]	6,28 [3,06; 12,11]	1,55 [-0,65; 6,6]	-0,42 [-0,97; 0,82]	6,29 [1,44; 10,19]

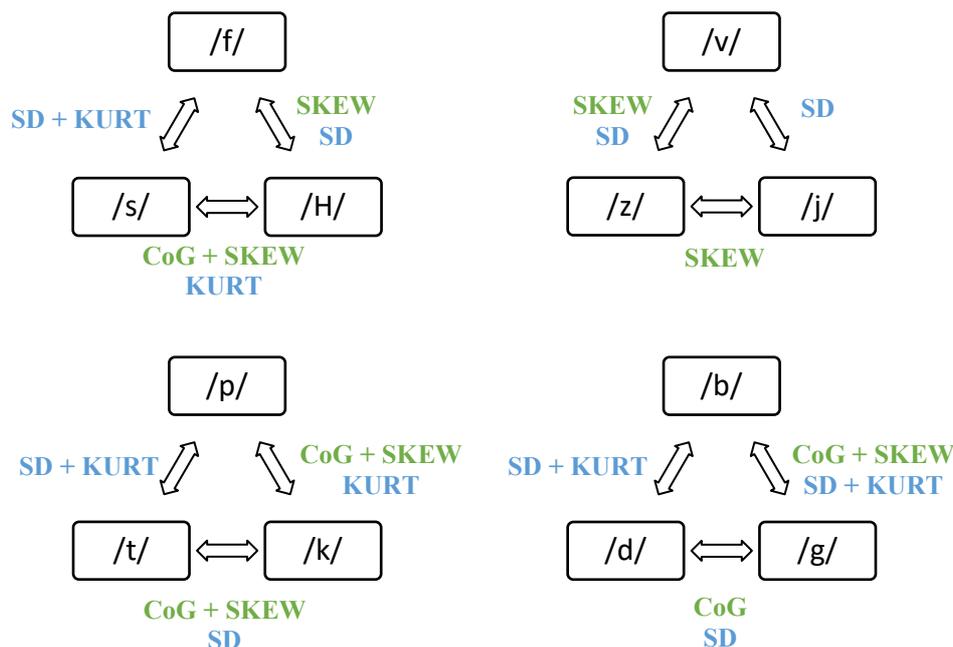


Figure 56 – Résumé des distinctions entre les lieux d’articulation basées sur les moments spectraux issus de notre algorithme, dans les fricatives (2 diagrammes supérieurs) et dans les plosives (2 diagrammes inférieurs) ; en vert : CoG et SKEW, en bleu : SD et KURT, paires respectivement liées théoriquement

Les distinctions entre les lieux d’articulation à l’aide des moments spectraux issus de notre algorithme sont similaires à ceux issus de la segmentation manuelle.

Pour la paire **COG-SKEW**, la significativité statistique de la différenciation consonantique par les deux moments spectraux disparaît entre /f/ et /s/ (malgré une valeur médiane plus élevée pour ce dernier), ainsi que la différenciation par le CoG entre /z/ et /ʒ/. Ceci peut être expliqué par un écart interquartile légèrement plus élevé par rapport à la

segmentation manuelle. La valeur de SKEW différencie maintenant le /ʃ/ du /f/ en plus du /s/. Pour les plosives, la distinction par SKEW disparaît entre /g/-/d/, mais demeure entre /g/-/b/.

Pour la paire **SD-KURT**, les distinctions par SD observées en segmentation manuelle restent toutes présentes. Dans les plosives non voisées, KURT ne différencie plus que le /p/ de /t/ et de /k/ (qui ne se différencient plus entre eux). Dans les fricatives, KURT ne différencie plus significativement le /z/ du /ʒ/ (probablement au vu d'une augmentation de l'EIQ du /z/, la médiane restant cependant élevée par rapport au /ʒ/ et au /v/), mais il différencie le /s/ du /ʃ/ et du /f/, ce qui n'était pas le cas dans notre première étude.

Moments spectraux et méthode FBDS - Corrélations entre moments spectraux

Les coefficients de corrélation de Spearman entre les moments spectraux issus de la méthode FBDS et les probabilités sont présentés dans le [Tableau 20](#).

Dans les fricatives, CoG et SKEW sont modérément corrélés négativement ; SD et KURT sont fortement corrélés négativement. SKEW n'est que faiblement corrélé avec KURT (positivement). Il n'y a pas de corrélation significative entre CoG et SD ni entre CoG et KURT.

Dans les plosives, CoG et SKEW sont fortement corrélés négativement, de même que SD et KURT. KURT est modérément corrélé avec CoG (négativement) et avec SKEW (positivement). Une faible corrélation positive existe également SD et CoG, ainsi qu'entre SD et SKEW. Il n'y a pas de corrélation significative entre SD et SKEW.

Tableau 20 – Coefficients de corrélation de Spearman entre les moments spectraux issus de la méthode FBDS et probabilités

<i>Fricatives</i>	CoG	SD	SKEW
SD	-0,0270 1,0000		
SKEW	-0,6565 0,0000*	-0,1767 0,0546	
KURT	-0,0810 1,0000	-0,7237 0,0000*	<i>0,2631</i> 0,0005*
<i>Plosives</i>	CoG	SD	SKEW
SD	<i>0,3217</i> 0,0000*		
SKEW	-0,8192 0,0000*	-0,2526 0,0009*	
KURT	-0,5580 0,0000*	-0,8810 0,0000*	0,5713 0,0000*

Note : Les astérisques indiquent les corrélations statistiquement significatives ; en gras, les corrélations supérieures à 0,50 ; en italique, les corrélations statistiquement significatives, mais faibles [HINKLE 03]

Les corrélations sont donc très similaires à celles mesurées pour les valeurs issues de la segmentation manuelle, mis à part l'apparition d'une faible corrélation négative entre SD et SKEW dans les plosives et sa disparition dans les fricatives.

Conclusion

Les valeurs issues de la méthode FBDS sont très similaires à celles obtenues en segmentation manuelle.

Les distinctions par la paire CoG-SKEW sont comparables avec celles identifiées en segmentation manuelle pour les plosives. Dans les fricatives, nous observons un pouvoir discriminatoire statistiquement diminué pour le CoG, au contraire de SKEW.

Les distinctions par la paire SD-KURT restent elles aussi similaires par rapport aux résultats en segmentation manuelle, autant pour les fricatives que pour les plosives.

Ainsi, comme dans notre première étude, le CoG est plus élevé dans les fricatives alvéolaires et plus bas dans les plosives palatales. SD, d'autre part, est plus élevé dans les fricatives labiodentales et les plosives bilabiales, alors que les plosives alvéolaires montrent une dispersion moins importante du noyau spectral. KURT, en plus de montrer de nombreuses valeurs aberrantes, aussi bien dans les fricatives que dans les plosives, semble moins sensible que SD dans les fricatives voisées.

Au vu de ces résultats, une combinaison des deux paires CoG-SKEW et SD-KURT semble rester la meilleure recommandation pour distinguer de manière consistante les lieux d'articulation consonantiques.

La mesure des moments spectraux est donc réalisable de manière automatisée à l'aide de l'algorithme que nous avons développé, sans nécessité de segmenter manuellement les consonnes. Cet algorithme se montre particulièrement performant pour la détection des bursts. À notre connaissance, bien que des études s'y soient intéressées, il n'existe à ce jour aucun algorithme de détection de bursts librement accessible. Ainsi, nous espérons que cette contribution pourra également servir à d'autres études acoustiques, notamment sur les bursts des plosives.

En plus de son utilité en recherche scientifique, d'un point de vue pratique, la mesure des moments spectraux semble donc également envisageable en contexte clinique pour l'analyse de l'articulation des consonnes. Afin d'aborder la question de la pertinence clinique des moments spectraux, nous allons dans le sous-chapitre suivant analyser les valeurs obtenues sur des enregistrements de parole pathologique.

4.3.3. Moments spectraux en parole pathologique

Afin d'investiguer les moments spectraux en parole pathologique, nous avons extrait ces mesures à l'aide de la méthode FBDS. Nous avons ensuite comparé les valeurs des patients à celles des sujets sains. Ensuite, pour mieux interpréter les mesures effectuées chez les patients, nous les avons comparées à deux scores perceptifs (intelligibilité et sévérité). Enfin, nous avons mené une étude de cas sur quatre patients pour analyser plus en détail le comportement des moments spectraux en parole pathologique. Nous allons maintenant décrire les résultats de ces analyses.

A. Enregistrements et extraction

Nous avons enregistré une lecture des phrases porteuses par 27 patients traités pour un cancer de la sphère ORL. Les données descriptives de cet échantillon sont reprises dans le [Tableau 21](#).

Tableau 21 – Description des patients traités pour un cancer ORL inclus dans notre étude

Âge		
Médiane		67
EIQ		12
Min.-max.		37-83
Sexe		
		N (%)
Homme		17 (63 %)
Femme		10 (37 %)
Localisation tumorale		
		N (%)
Cavité buccale		15 (56 %)
Oropharynx		11 (41 %)
Mixte		1 (3 %)
Intervention (combinaisons possibles)		
		N (%)
Chirurgie		24 (89 %)
Radiothérapie		26 (96 %)
Chimiothérapie		16 (59 %)
Délagi depuis la fin du traitement		
Médiane		40 mois
EIQ		131 mois
Min.-max.		6-564 mois
Critères de la classification TNM		
		N (%)
T (taille de la tumeur primitive et fixation aux organes voisins ; 4=tumeur étendue)	4	10 (37 %)
	3	3 (11 %)
	2	6 (22 %)
	1	2 (8 %)
	Non renseigné	6 (22 %)
N (nodes ; envahissement des ganglions lymphatiques ; dépend du nombre, de la taille et de la localisation des ganglions atteints)	3	1 (4 %)
	2	7 (26 %)
	1	4 (15 %)
	0	8 (29 %)
	Non renseigné	7 (26 %)
M (métastases ; 1=oui)	1	1 (4 %)
	0	19 (70 %)
	Non renseigné	7 (26 %)

Les moments spectraux ont été extraits en utilisant la méthode FBDS. Les données descriptives des moments spectraux chez les patients, côte à côte avec celles mesurées sur les sujets sains, sont illustrées en [Figure 58](#). Les spectres LPC obtenus avec les mêmes paramètres que précédemment sont disponibles en [Annexe S](#). Dix-neuf occurrences ont été exclues à cause d'un décalage de l'alignement forcé dû à des répétitions de phrases porteuses en fins d'enregistrements ou à des hésitations. Cinq occurrences supplémentaires ont été exclues à cause d'une mauvaise détection de la fenêtre d'analyse par l'algorithme FBDS, détectées lors des études de cas décrites ci-après.

A. *Patients vs sujets sains*

Nous avons utilisé le test U de Mann-Whitney pour mesurer les éventuelles différences entre les moments spectraux chez les sujets sains et chez les patients, pour chaque phonème. Des différences statistiquement significatives ($p < 0,05$) ont été mesurées pour :

- le CoG de /ʃ, s, z/ et /g, k, t/ ;
- le SD de /d, t/ ;
- le SKEW de /ʃ, ʒ/ et /k/ ;
- le KURT de /d, k, t/ ;

Comme illustré en [Figure 57](#), les valeurs de CoG indiquent chez les patients une concentration spectrale plus basse dans les fréquences pour les fricatives apico-dentales /s, z/ et pour les alvéolaires /ʃ, ʒ/, ce qui est cohérent avec les valeurs relevées dans d'autres études sur les moments spectraux en parole pathologique [BUDER 96; HERNANDEZ 19]. Les valeurs de CoG indiquent au contraire une concentration spectrale plus élevée dans les fréquences pour les plosives palatales /k, g/ et les apico-alvéolaires /t, d/, ce qui est également retrouvé dans d'autres études et peut être interprété comme le résultat soit d'un avancement du point d'articulation, soit d'une occlusion incomplète entraînant un bruit de friction [KAY 12]. Les valeurs de SD et de KURT indiquent des spectres plus étalés pour /t, d/.

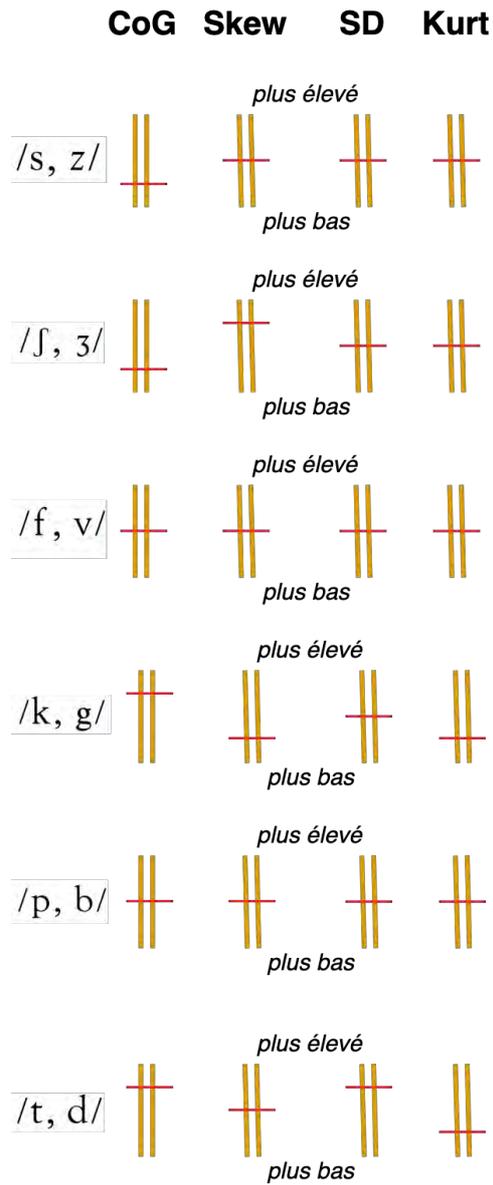
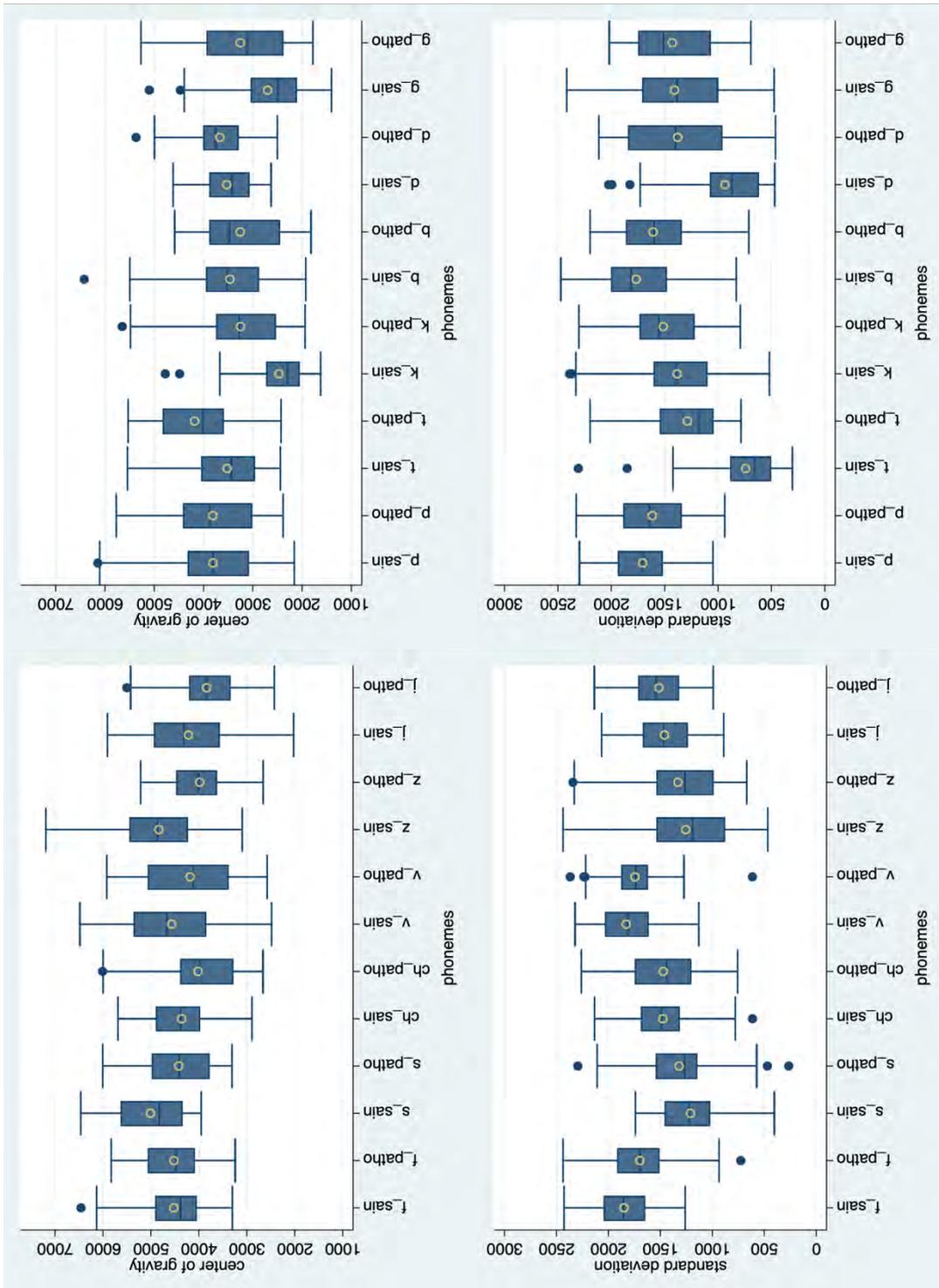


Figure 57 – Représentation schématique des valeurs des moments spectraux en parole pathologique par rapport aux valeurs des sujets sains



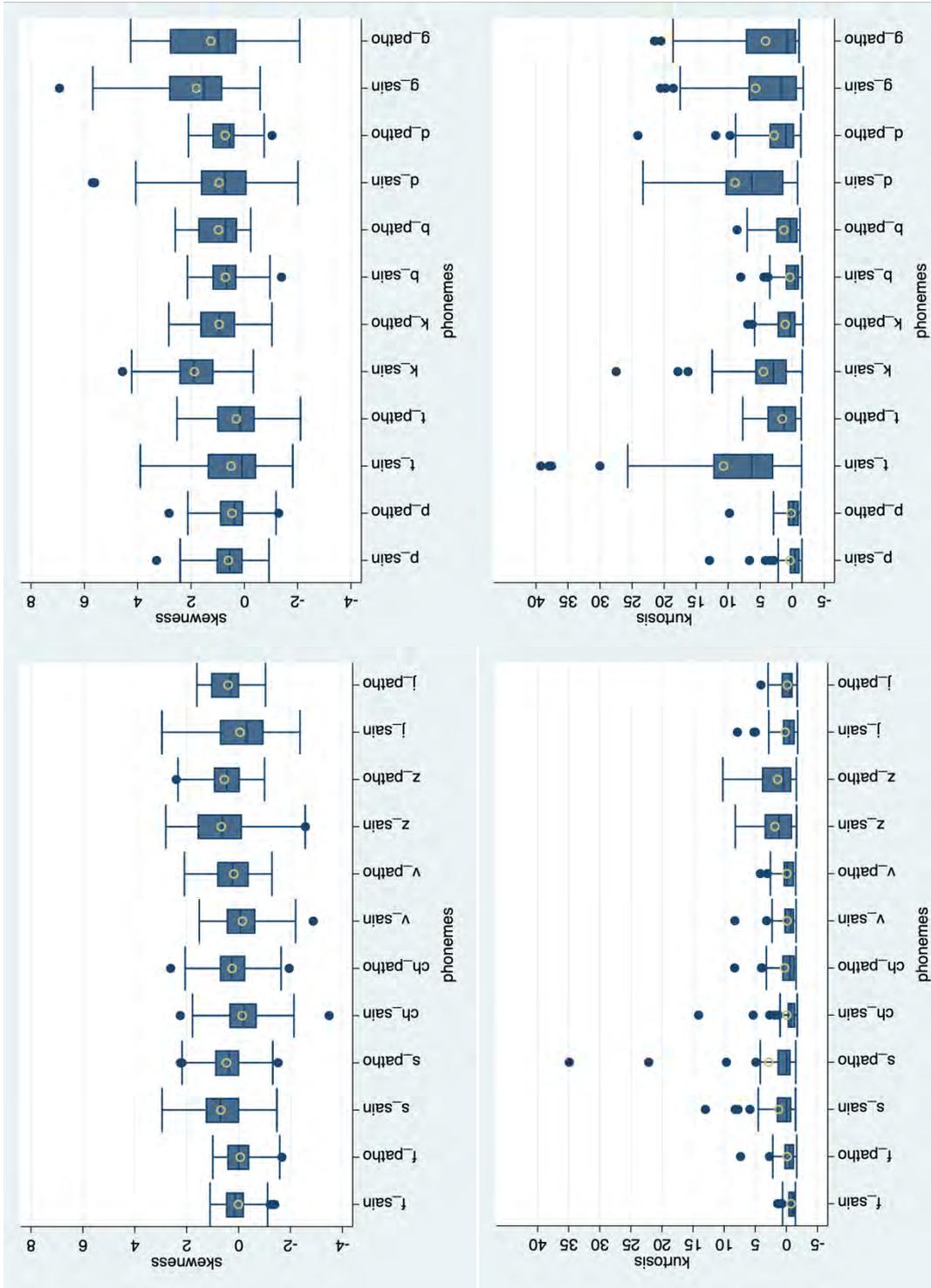


Figure 58 – Données descriptives des moments spectraux issus des enregistrements de sujets sains et de patients traités pour un cancer ORL, pour les fricatives (gauche) et pour les plosives (droite) : les points situés à plus de deux fois le quartile 99 ont été exclus pour SKEW (1 point exclu >8.52) et KURT (6 points exclus >44.27)

B. Scores perceptifs

Afin d'interpréter ces résultats, nous avons récolté deux scores perceptifs : la sévérité globale du trouble de la parole et l'intelligibilité. Nous allons maintenant décrire ces scores de référence, puis analyser les moments spectraux en parole pathologique au regard de ceux-ci.

Sévérité

L'évaluation perceptive de la sévérité du trouble a été effectuée par un jury d'experts de la parole pathologique, constitué d'un médecin phoniatre et de cinq orthophonistes exerçant en service ORL/cancérologie. Chacun de ces experts a évalué la sévérité globale sur une échelle de 0 (trouble très sévère) à 10 (absence de trouble) sur base de l'enregistrement d'un entretien semi-dirigé des patients inclus dans cette étude. Les scores de sévérité de nos 27 patients moyennés sur les 6 experts varient entre 0,5 et 10, avec une médiane de 6,33 (EIQ=4,5) (voir [Figure 59](#)). Notre corpus de patients est donc hétérogène quant à la sévérité du trouble de la parole perçu par le jury d'experts, certains ayant des degrés d'altération très sévères, d'autres ne semblant présenter aucun impact perceptible.

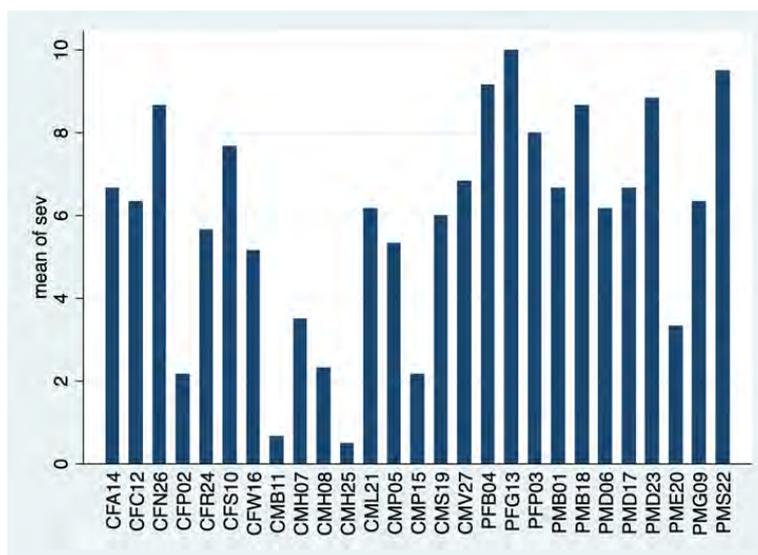


Figure 59 – Sévérité moyenne par patient

Intelligibilité

Pour le deuxième score perceptif, celui de l'intelligibilité, nous avons recruté des évaluateurs de langue maternelle française sans plainte auditive et dont la fréquence d'écoute de parole pathologique était inférieure à une fois par semaine. Notre panel d'évaluateurs était ainsi composé de neuf hommes et six femmes, avec un âge médian de 25 ans (EIQ=3,5 ; min.= 23, max. = 56). La tâche d'évaluation perceptive consistait en l'écoute des pseudomots /øCø/ extraits des 17 phrases porteuses prononcées par les 27 patients et la transcription orthographique de la consonne centrale (soit 459 stimuli). Pour chaque transcription, l'évaluateur devait également indiquer la facilité de transcription (facile, moyen, difficile). Afin de vérifier la fidélité test-retest, 10 % des pseudomots ont été présentés à deux reprises. Au total, 505 stimuli ont ainsi été présentés à chaque évaluateur, aléatoirement répartis sur 4 sessions d'écoute. Entre chaque session, une pause de minimum 5 minutes devait être respectée. Un essai pilote a permis d'estimer la durée totale de l'évaluation à une heure.

Un test Kappa de Cohen a permis de confirmer une fidélité test-retest substantielle [LANDIS 77] pour chaque juge concernant les transcriptions (77 % d'accord en moyenne ; Kappa moyen de 0,75). Pour le niveau de difficulté de transcription, la fidélité test-retest est modérée (77 % d'accord en moyenne ; Kappa moyen de 0,54).

Le test Kappa de Fleiss (pour des données catégorielles de plus de deux évaluateurs) a été utilisé pour mesurer la fidélité interjuges. Pour les transcriptions orthographiques, un accord substantiel est mesuré (coefficient Kappa de 0,67). Pour le niveau de difficulté de transcription estimé, l'accord est acceptable (« fair agreement » ; coefficient Kappa de 0,34).

Nous avons également vérifié qu'il n'y avait pas d'effet de fatigue se répercutant sur les réponses. Nous avons ainsi qualitativement comparé les scores de difficulté de transcription sur les quatre sessions d'écoute, et avons constaté que les proportions de réponses « facile », « moyen », « difficile » restaient similaires tout au long de l'évaluation (voir [Figure 60](#)).

-> order = 1			
difficulty	Freq.	Percent	Cum.
difficile	252	13.47	13.47
facile	1,265	67.61	81.08
moyen	354	18.92	100.00
Total	1,871	100.00	

-> order = 2			
difficulty	Freq.	Percent	Cum.
difficile	211	11.83	11.83
facile	1,281	71.80	83.63
moyen	292	16.37	100.00
Total	1,784	100.00	

-> order = 3			
difficulty	Freq.	Percent	Cum.
difficile	178	10.76	10.76
facile	1,207	72.97	83.74
moyen	269	16.26	100.00
Total	1,654	100.00	

-> order = 4			
difficulty	Freq.	Percent	Cum.
difficile	182	11.57	11.57
facile	1,150	73.11	84.68
moyen	241	15.32	100.00
Total	1,573	100.00	

Figure 60 – Fréquences de réponse pour les trois niveaux de difficulté de transcription sur les quatre sessions d'écoute successives (order 1 = premier groupe d'écoute, order 4 = dernier groupe d'écoute)

Nous avons également vérifié la cohérence entre le niveau de difficulté de transcription estimé et l'exactitude de la transcription fournie. Ces deux données sont effectivement concordantes, le nombre de transcriptions correctes étant plus élevé pour les stimuli jugés faciles à transcrire.

Les pourcentages de transcription correcte par sujet sont présentés en [Figure 61](#). Comme pour l'évaluation de la sévérité globale du trouble de la parole, nous observons une variation importante des scores. En effet, la transcription des phonèmes est majoritairement correcte pour certains patients, alors que pour d'autres, de nombreuses erreurs sont observées.

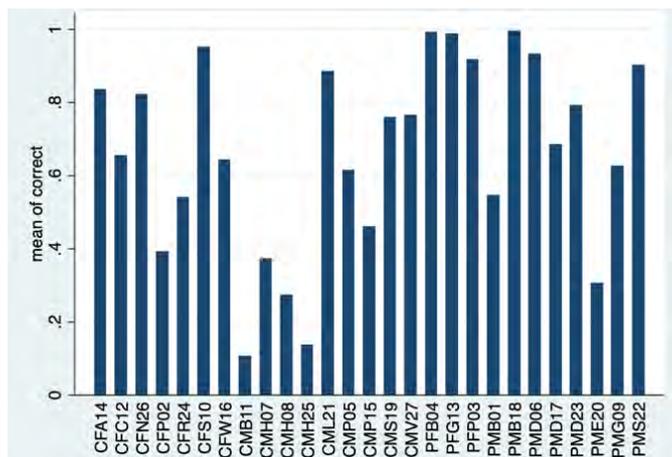


Figure 61 – Pourcentages de transcriptions correctes par patient

C. Étude de cas – moments spectraux en parole pathologique

Comme nous l’avons vu ci-avant, l’hétérogénéité de notre groupe de patients rend les statistiques groupées peu adaptées. C’est pourquoi nous avons voulu nous pencher plus spécifiquement sur l’étude de quelques cas identifiés à l’aide des évaluations perceptives. Afin d’analyser plus en détail les performances des moments spectraux en parole pathologique, nous avons donc sélectionné les deux patients avec les scores perceptifs les plus faibles (CMB11 et CMH25) et les deux patients avec les scores perceptifs les plus élevés (PFG13, PFB04).

1. Alignement et segmentation

Tout d’abord, cette analyse nous a permis de nous rendre compte que l’alignement forcé montre des difficultés en cas de parole fortement altérée. En effet, à de nombreuses reprises, l’articulation des trois phonèmes du pseudomot /øCø/ n’est pas suffisamment différenciée et l’alignement forcé identifie la quasi-totalité du pseudomot comme étant le premier /ø/ (voir [Figure 62](#)).

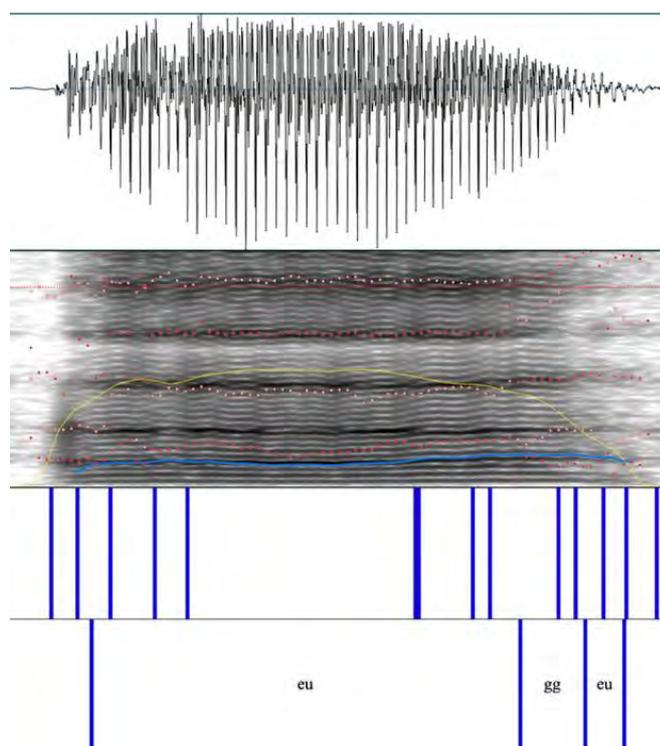


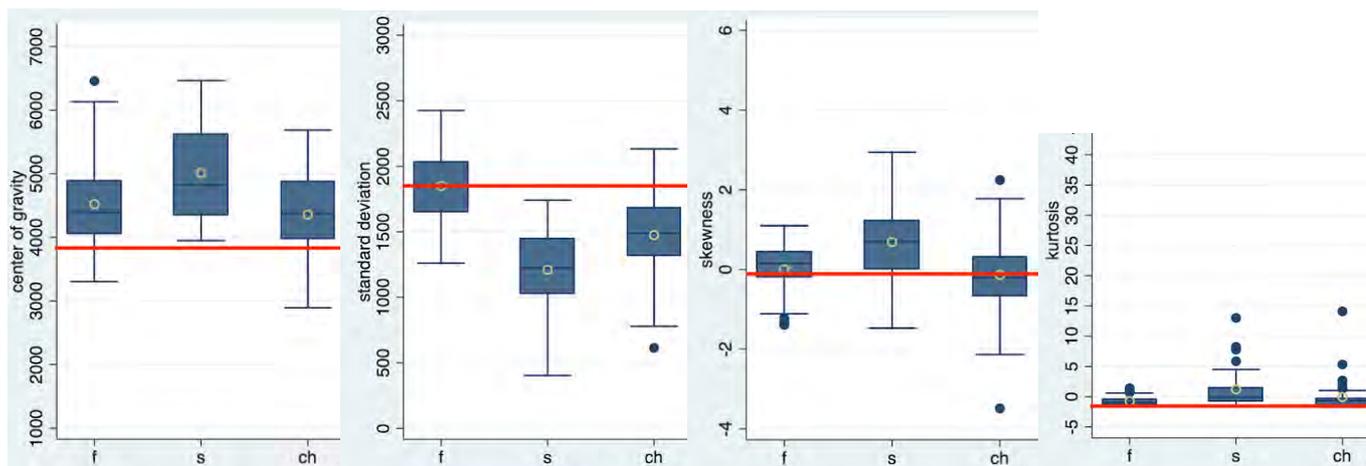
Figure 62 – Exemple d’un mauvais alignement dû à une très faible distinctivité des phonèmes du pseudomot /ø g ø/

Pour les patients avec des scores perceptifs élevés, l'alignement forcé performe de manière bien plus satisfaisante, les décalages étant rares. Cependant, la détection du burst par l'algorithme FBDS est peu satisfaisante chez la patiente PFB04, malgré une parole peu altérée. Ceci est potentiellement dû à la dysphonie modérée de la patiente renseignée dans son dossier médical (dysphonie modérée (1,5/3 ; 3 = dysphonie sévère), induisant une instabilité du signal vocal.

2. Moments spectraux et évaluation perceptive

L'analyse des données des patients à faibles scores perceptifs nous a permis de constater que les confusions perceptives se reflètent globalement dans les moments spectraux.

Par exemple, pour le patient CMB11, le /p/ a été correctement transcrit par un seul des 15 évaluateurs. Parmi les autres, les transcriptions les plus fréquentes sont /ʃ/ (5/14), /f/ (3/14) et /r/ (3/14). Comme illustré en [Figure 63](#), les CoG de /ʃ/ et /f/ — transcriptions les plus fréquentes du /p/ de notre patient — sont les plus bas des fricatives chez le sujet sain, et leur SD le plus élevé ; ce sont ces deux phonèmes qui se rapprochent donc le plus des valeurs mesurées pour notre patient. Le SKEW proche de 0 et légèrement négatif est lui aussi caractéristique du /f/ et du /ʃ/ chez le sujet sain, ainsi que le KURT négatif. Ainsi, les quatre valeurs mesurées pour le /p/ de CMB11, proches de celles du /f/ et du /ʃ/ chez le sujet sain, peuvent être mises en lien avec la transcription majoritaire en ces deux phonèmes.



Dans l'ensemble, nous observons que les plosives sont majoritairement perçues comme des fricatives ; or, nous y mesurons des CoG très élevés. L'occlusion semble donc insuffisante, conduisant non pas à un relâchement net de l'air phonatoire, mais plutôt à un bruit de friction caractérisé acoustiquement par une concentration spectrale dans les fréquences aiguës.

Les confusions parmi les fricatives peuvent elles aussi être mises en lien avec les moments spectraux. Par exemple, pour le patient CMH25, /f,v/ et /s/ montrent des spectres d'aplatissement (SD et KURT) centralisé et « inversé » par rapport aux tendances chez les sujets sains, celui de /f,v/ étant trop pointu et celui du /s/ étant aplati — se rapprochant ainsi l'un de l'autre et induisant des confusions perceptives dans les deux sens (voir [Figure 64](#)).

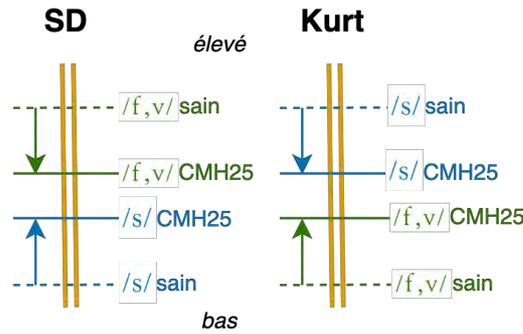


Figure 64 – Représentation schématique de la centralisation des valeurs de SD et de Kurt pour le /f, v/ et /s/ du patient CMH25

Pour ce même patient, l’apico-dentale /s/ est majoritairement assimilées au /v/ et au /ʒ/ lors de l’évaluation perceptive. Acoustiquement, pour le /s/, le SD élevé et le SKEW et KURT bas l’assimilent effectivement à /v/ et /ʒ/ (voir Figure 65).

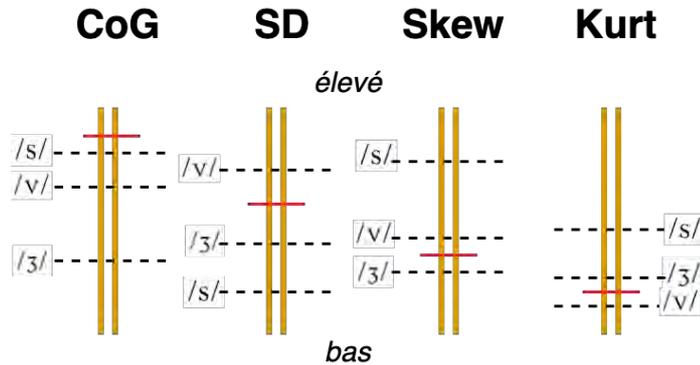


Figure 65 – Représentation schématique des valeurs de moments spectraux du /s/ pour le patient CMH25 (en rouge), au regard des valeurs de /s, v, ʒ/ chez les sujets sains (en pointillés)

L’analyse des données des patients à scores perceptifs élevés, pour sa part, révèle que c’est bien l’ensemble des quatre moments spectraux qu’il faut prendre en compte dans l’analyse de la distinctivité consonantique. En effet, nous observons des phonèmes correctement perçus dehors de l’intervalle interquartile des sujets sains et vont dans le sens opposé à la distinction consonantique. Par exemple, pour la patiente PFG13 (voir Figure 66), le SD élevé de /f, v/ qui les différencie normalement des autres fricatives est effectivement présent et même accentué ; le CoG de /s, z/ n’est pas comme attendu plus élevé que les autres, en revanche leur SKEW reste supérieur et permet de maintenir la distinction. De même, le /v/ montre certes un CoG trop élevé, mais son SD dans la norme permet de le différencier du /z/, et le SKEW élevé de ce dernier le différencie des deux autres fricatives.

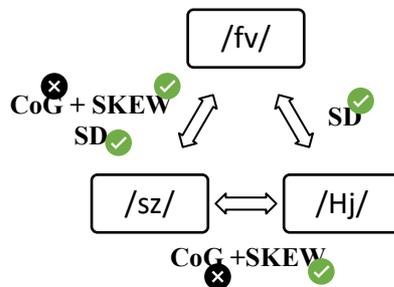


Figure 66 – Pour PFG13, distinctions phonémiques par les moments spectraux préservées (en vert) et non-préservées (en noir)

Ainsi, une déviation de la valeur absolue d’un moment spectral n’induit pas nécessairement une baisse de l’intelligibilité. C’est l’interprétation combinée et relative des valeurs qui est informative, au regard des distinctions chez les sujets sains (voir Figure 56).

Conclusion et Perspectives

L'objectif de cette étude était d'examiner les caractéristiques spectrales de la parole produite par des patients traités pour un cancer de la sphère ORL, à l'aide des moments spectraux. Les résultats suggèrent que l'analyse des moments spectraux fournit des informations utiles pour décrire les singularités acoustiques et perceptives de ces locuteurs. Cependant, comme cela est souvent le cas, notre groupe d'étude est caractérisé par une grande hétérogénéité de l'atteinte anatomophysiologique, de la sévérité du trouble de la parole, de l'intelligibilité consonantique et des stratégies articulatoires. L'étude conjointe des résultats perceptifs et acoustiques de quatre patients a néanmoins montré que les moments spectraux peuvent effectivement être interprétés en lien avec l'intelligibilité perceptive. Pour chaque consonne, l'interprétation requiert la prise en compte des quatre moments spectraux en regard de ceux des autres consonnes. Nos résultats montrent notamment qu'un CoG élevé pour une plosive est indicateur d'une production se rapprochant davantage d'une fricative, avec une occlusion insuffisante.

Néanmoins, dans notre étude de cas, nous avons arbitrairement choisi l'intervalle interquartile des sujets sains comme référence ; il reste encore à déterminer sur des échantillons plus larges et plus homogènes de patients (d'un point de vue perceptif) quel degré de déviation des moments spectraux a un impact sur la perception des phonèmes.

La distinctivité entre les différentes consonnes pourra aussi être analysée en calculant des ratios de moments spectraux entre chaque paire consonantique. Ceci permettra ensuite de comparer plus directement la distinctivité chez les sujets sains et celle dans diverses populations pathologiques pour lesquelles nous nous attendons à une réduction de cette distinctivité consonantique.

De plus, pour permettre une interprétation articulatoire des déviations des moments spectraux, de futures études devront prendre en compte un examen moteur de la sphère orofaciale des patients, ainsi qu'idéalement des mesures articulatoires physiologiques.

Enfin, les troubles vocaux et de nasalité, fréquemment associés aux troubles de la parole dans de nombreuses pathologies, n'ont pas pu être analysés dans ce travail. Pour les confusions consonantiques basées sur le voisement, d'autres mesures acoustiques complémentaires aux moments spectraux doivent être utilisées. Pour l'hypernasalité, il reste encore à analyser le comportement des moments spectraux dans les consonnes nasales, dont la structure spectrale se démarque des autres consonnes par la présence de formants et d'anti-formants. Il serait ainsi intéressant de déterminer en quelle mesure les moments spectraux s'appliquent à la distinction des phonèmes nasaux entre eux (/m, n, ŋ, ɲ/) et entre les phonèmes des paires nasales/non-nasales (p. ex. /b/ et /m/).

À ce jour, bien que les performances soient satisfaisantes, l'alignement forcé nécessite encore d'être amélioré en parole pathologique pour éviter des décalages phonémiques en cas de parole fortement altérée. De plus, l'algorithme FBDS, qui se montre performant pour les fricatives, doit lui aussi être adapté afin de permettre une détection plus robuste des bursts des plosives. Pour cela, il reste à déterminer quels facteurs influencent sa performance ; nos résultats préliminaires tendent par exemple à indiquer un impact de la dysphonie sur la détection des ruptures.

Enfin, à des fins d'utilisation clinique, il serait également intéressant de développer des moyens de visualisation des moments spectraux afin d'en faciliter l'interprétation. Pour cela, nous proposons deux pistes. La première consiste en la visualisation, à l'aide des moments spectraux, de toutes les consonnes prononcées par les sujets sains sur un graphique à trois dimensions. Nous avons collaboré avec Wessel Stoop de l'Université de Radboud (Nimègue) (<http://wesselstoop.ruhosting.nl/>), qui a pour cela utilisé un algorithme t-SNE (« t-distributed stochastic neighbor embedding »), une méthode de réduction de dimensions pour la visualisation de données [VAN DER MAATEN 08]. Cette technique permet donc de réduire les quatre moments spectraux en des coordonnées d'un espace à deux ou trois dimensions pour chaque consonne. Un exemple des nuages de points représentant chaque plosive selon ses quatre moments spectraux chez les sujets sains est visible en Figure 67. Sur cet exemple, nous pouvons observer que les trois lieux d'articulation des plosives forment des groupements de points. Nous pourrions alors afficher les productions d'un patient par rapport à ces zones consonantiques afin de déterminer elles s'en éloignent ou non, et, le cas échéant, effectuer des mesures de distance par rapport au point central de la zone cible. Cependant, cette méthode n'aurait d'intérêt que pour juger si un phonème est correctement prononcé ou non. En effet, d'un point de vue clinique, l'amalgamation des quatre moments spectraux en un point dans un espace 2D ou 3D ne permet aucune interprétation articulatoire ni piste pour le plan thérapeutique.

Aussi, une deuxième piste consiste en la réalisation de représentations spécifiques à chaque consonne. Cette technique consisterait à tracer des distributions « standard » basées sur les données des sujets sains pour chaque consonne. La courbe de la production du phonème par un patient pourrait ensuite y être superposée en combinaison avec des mesures de chevauchement et de probabilités d'identité phonémique. Un exemple fictif d'une telle représentation est illustré en Figure 68. Cette piste semble plus pertinente d'un point de vue clinique, étant donné qu'elle permettrait de cibler davantage la stratégie de rééducation articulatoire.

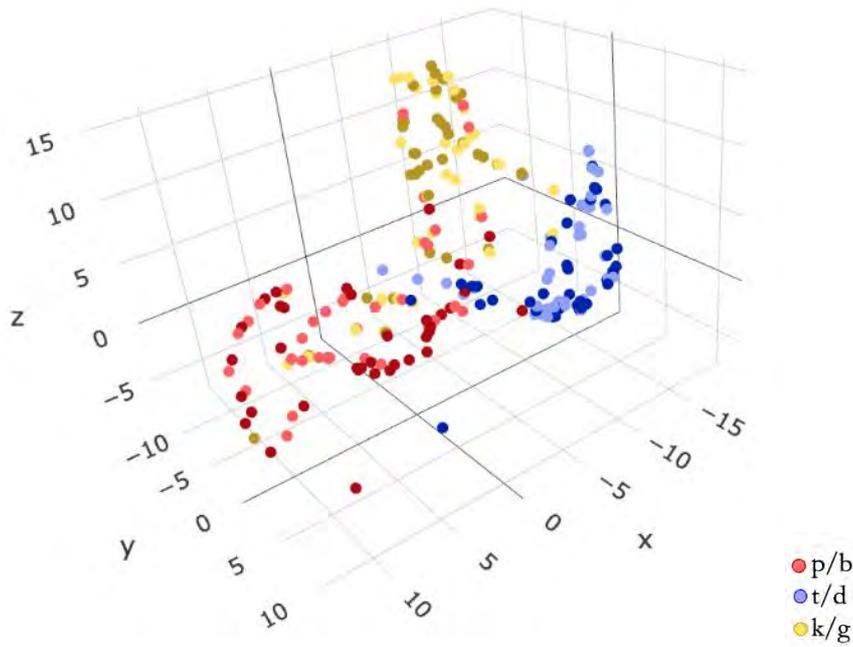


Figure 67 – Représentation des plosives des sujets sains dans un espace à trois dimensions, grâce à la technique t-SNE (créée par Wessel Stoop de l'Université de Radboud) ; disponible à l'adresse : <https://cls.ru.nl/staff/wstoop/moments/interactive.html>

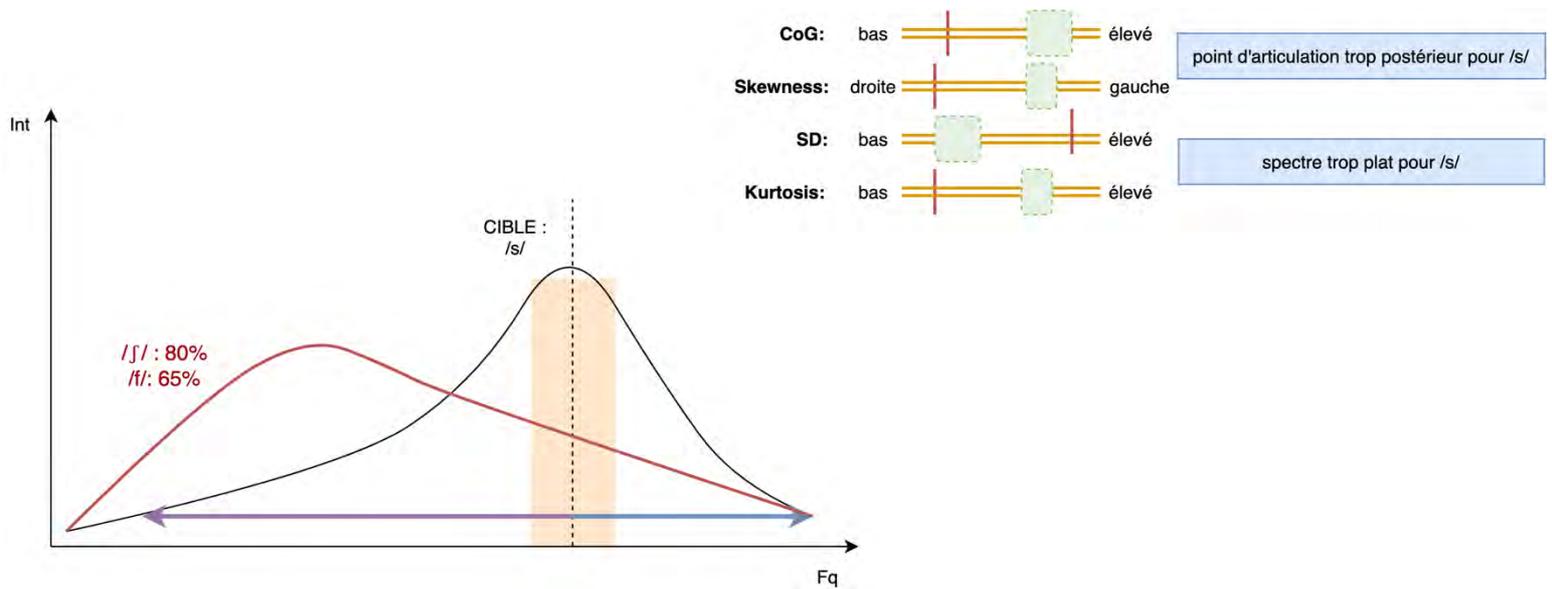


Figure 68 – Exemple fictif d'une courbe standard du phonème /s/ (en pointillé : CoG ; en orange pâle : SD et KURT ; en bleu et mauve : SKEW) avec la courbe de production d'un patient superposée (en rouge) ainsi que des probabilités d'identité phonémique ; dans la partie supérieure droite, représentation en curseur des quatre moments spectraux (en vert, la zone des valeurs de référence pour le phonème cible, en rouge, la valeur de la production du patient) avec des pistes d'interprétation (en bleu)

4.4. Conclusion du chapitre

Pour répondre au besoin de mesures moins subjectives et plus reproductibles de la parole, nous avons mené des analyses spectrales consonantiques, dans l'objectif de proposer de nouvelles pistes de mesures plus objectives de l'intelligibilité, utilisables en clinique et en recherche.

Après une étude pilote tentant de reproduire une « banane de la parole » [POMMEE 19a], nous avons investigué en détail le comportement des moments spectraux en parole saine, qui considèrent l'aspect global du spectre consonantique.

Dans un premier temps, nous avons investigué ces mesures sur base d'une segmentation consonantique manuelle. Une combinaison des deux paires CoG-SKEW et SD-KURT semble être la meilleure recommandation pour distinguer acoustiquement les différentes consonnes.

Ensuite, nous avons abordé l'utilisation des moments spectraux sur base de l'alignement forcé de nos stimuli de parole. L'alignement forcé employé seul ne permettant pas une détection suffisamment précise des bursts dans les plosives, nous avons développé un nouvel algorithme combinant cet alignement forcé avec la méthode de segmentation « forward-backward divergence ». Les résultats obtenus avec cette méthode sont fidèles à ceux issus de la segmentation manuelle. Notre algorithme permet la détection de bursts des plosives et de la fenêtre d'analyse dans un segment stable pour les fricatives. D'un point de vue pratique, la mesure des moments spectraux semble donc également envisageable en contexte clinique pour l'analyse de l'articulation des consonnes.

Afin d'aborder la question de la pertinence clinique, nous avons ensuite exploré l'utilisation des moments spectraux sur des enregistrements de patients traités pour un cancer de la sphère ORL. Les résultats suggèrent que l'analyse des moments spectraux fournit des informations utiles pour décrire les singularités acoustiques et perceptives de ces locuteurs. En effet, l'étude conjointe des résultats perceptifs et acoustiques de quatre patients a montré que les moments spectraux peuvent être analysés en lien avec l'intelligibilité perceptive. C'est l'interprétation combinée et relative des valeurs qui est informative, au regard des distinctions chez les sujets sains.

Chapitre 5. Essai pilote de mesures acoustiques sur notre nouveau texte

Nos études acoustiques précédentes ont été effectuées sur des stimuli très contrôlés. Or, l'emploi de mesures acoustiques sur un texte de référence en maximiserait la pertinence clinique. Aussi, nous avons souhaité clôturer ce travail par un exemple de protocole permettant un traitement acoustique sur notre nouveau texte de référence présenté en [Chapitre 3](#). Ce protocole emploie des mesures identifiées dans le Chapitre 4 (cf. « 4.1. [État de l'art – mesures acoustiques de l'intelligibilité de la parole](#) »).

Cette étude de faisabilité poursuit deux objectifs principaux : analyser la reproductibilité des mesures sur des lectures répétées du texte et investiguer l'influence de la vitesse de lecture à l'aide d'une lecture rapide. En plus de ces objectifs principaux, nous avons également souhaité observer si la lecture à voix haute de notre nouveau texte est faisable en termes de facilité de lecture et de longueur, et analyser l'effet du genre et de la variété régionale du français (français de France, de Belgique et du Québec).

5.1. Méthodologie

A. Protocole et mesures acoustiques

Notre texte étant destiné à l'utilisation par des chercheurs et des cliniciens d'horizons variés, nous avons rédigé un protocole facile d'utilisation par des professionnels non experts en informatique. Celui-ci permet le prétraitement des fichiers audio, un alignement forcé (p. ex. en utilisant le service gratuit en ligne « WebMAUS Basic »³²), puis l'extraction de quatre types de mesures acoustiques. L'objectif du texte de référence étant de permettre l'évaluation de la voix et de la parole, nous avons intégré des mesures de qualité vocale et des mesures liées à l'intelligibilité de la parole (cf. « 4.1. [État de l'art – mesures acoustiques de l'intelligibilité de la parole](#) ») pour chacune des trois classes de phonèmes (voyelles, consonnes, semi-consonnes) :

1. Mesures de qualité vocale³³ sur la phrase voisée (lue à deux reprises) :
 - la proéminence du pic cepstral lissé (« CPPs ») : distance entre l'amplitude du pic cepstral et l'amplitude du point ayant la même quéfrence sur la droite de régression à travers le cepstre lissé³⁴ (donc l'amplitude prédite à ladite quéfrence par la droite de régression) [HILLENBRAND 94]. L'utilisation de données du domaine cepstral permet des mesures plus fiables dans les dysphonies sévères [HEMAN-ACKAH 03]. En effet, ces dernières sont caractérisées par un signal peu périodique et se prêtent plus difficilement aux analyses du domaine fréquentiel tels le shimmer, le jitter et le rapport signal/bruit. De plus, les mesures cepstrales permettent l'évaluation du degré de la dysphonie sur un échantillon de parole continue [DELGADO-HERNANDEZ 18; LOWELL 11; SAUDER 17] ;

³² <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

³³ Adapté du script de l'Acoustic Voice Quality Index version 03.01 [BARSTIES 16]

³⁴ Le CPPs est mesuré après un « lissage », c'est à dire un moyennage d'un certain nombre de fenêtres cepstrales avant et après la fenêtre d'intérêt [HEMAN-ACKAH 03; HILLENBRAND 96].

- la pente du spectre moyen à long terme (« LTAS slope ») : division de l'énergie des fréquences basses (0-1 kHz) par celle des fréquences hautes (1-10 kHz) à travers le spectre moyen à long terme [VOLIN 11]. Les mesures du spectre moyen à long terme (tout comme les mesures cepstrales) ont gagné de l'intérêt dans l'évaluation vocale acoustique, car elles permettent tout comme les mesures cepstrales l'évaluation de la sévérité de la dysphonie sur des échantillons de parole continue [LOWELL 11] ;
 - l'inclinaison de la courbe de tendance à travers le spectre moyen à long terme (« LTAS tilt ») : il s'agit de la différence d'énergie de la droite de régression (prédiction de l'énergie en fonction de la fréquence) entre 0-1 kHz et 1-10 kHz à travers le spectre moyen à long terme [BARSTIES 15; SMITH 14] ;
2. Mesures vocaliques³⁵ : extraction des deux premiers formants des voyelles cardinales, traçage du triangle vocalique et calcul de l'aire du triangle relative à un triangle de référence. Un exemple de triangle vocalique obtenu à l'aide du script employé est illustré en [Figure 69](#)³⁶. Pour plus de détails quant à sa construction, voir [VAN SON 18].
- La mesure VSA est sensible aux différences d'intelligibilité dans la parole saine [BOND 94] et est couramment utilisée pour rendre compte de la centralisation des voyelles [LIU 05; SAPIR 09; WEISMER 01].

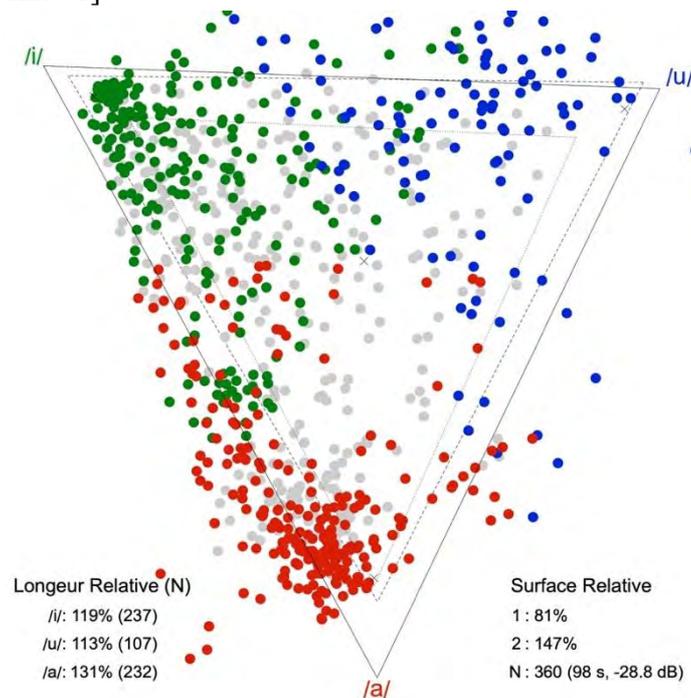


Figure 69 – Exemple de triangle vocalique obtenu de la lecture du texte entier par une locutrice belge. F1 en ordonnées (haut-bas), F2 en abscisses (droite-gauche), en demi-tons ; ligne discontinue : triangle de référence ; pointillés : triangle employé par le locuteur (1 E.T.) ; ligne continue : triangle employé par le locuteur (2 E.T.) ; « x » : voyelles cardinales et centroïde (moyenne géométrique des fréquences des voyelles cardinales) ; en gris, pour chaque segment vocalique, le point

³⁵ Adapté du script VowelTriangle (<https://github.com/robvanson/VowelTriangle>) [VAN SON 18]

³⁶ Pour chaque enregistrement, trois axes sont définis, entre le centroïde et chaque coin du triangle. Ensuite, chaque production vocalique est projetée sur l'axe correspondant, et la moyenne et l'E.T. des positions projetées sont calculés. Pour chaque axe, une longueur est définie comme la distance entre le centroïde et la moyenne, plus 1 ou 2 E.T. Le triangle construit par ces trois axes est le triangle effectivement utilisé par le locuteur. Ainsi, deux triangles sont tracés : l'un correspond à un écart-type, l'autre à deux écarts-types. Le pourcentage de l'aire du triangle utilisée par le locuteur est calculé, ainsi que la taille des axes individuels (valeurs à deux écarts-types).

le plus proche du centroïde (non employé dans cette étude) ; 1 = surface relative du triangle 1 E.T. ; 2 = surface relative du triangle 2 E.T. ; N = nombre total de segments vocaliques

3. Mesures consonantiques : calcul des moments spectraux sur les consonnes en contexte /aCa/³⁷ à l'aide de l'algorithme FBDS ;
4. Mesure semi-consonantique : mesure de la pente des trois premiers formants dans les glides. Le script que nous avons écrit permet d'extraire les formants selon la méthode de Burg sur chaque fenêtre de 5 ms du début à la fin du segment semi-consonantique. Le calcul des pentes se fait ensuite en soustrayant la moyenne des 5 dernières fenêtres de la moyenne des 5 premières fenêtres, et en divisant le tout par la durée du segment (en ms). Dans le vieillissement sain, une diminution de l'intelligibilité peut être en partie attribuée à un ralentissement des mouvements de la langue [KURUVILLA-DUGDALE 20]. En parole pathologique, la pente de F2 est considérée comme un indicateur de la lenteur articulatoire [KENT 89a; Y. KIM 09; WEISMER 92, 95].

Le dossier contenant les scripts et les consignes pour l'utilisation de ce protocole est disponible en ligne à l'adresse <https://cloud.irit.fr/index.php/s/mp6sNM7n4FUqpKX>. Nous y avons également joint les recommandations pour l'enregistrement d'échantillons de parole disponibles en [Annexe G](#).

B. Locuteurs et enregistrements

Afin de vérifier que ce protocole est employable, nous avons enregistré des locuteurs sains francophones en France (Toulouse), en Belgique (Liège et Eupen) et au Québec (Toronto). L'échantillon final est composé de 17 locuteurs (cinq femmes et trois hommes en France, trois femmes et trois hommes en Belgique, un homme et deux femmes au Québec), avec un âge médian de 29 ans (EIQ = 7).

Chaque locuteur a effectué une lecture à voix haute de notre texte en entier, puis quatre lectures de la première partie (en incluant la phrase « Mardi, ils vont à la plage » pour éviter un relâchement prosodique sur la dernière phrase de la première partie), et enfin une cinquième lecture de ce passage en lecture rapide.

Les enregistrements en France et en Belgique ont été effectués à l'aide d'un microphone serre-tête t.bone HC-444 TWS et un enregistreur numérique Zoom H4n Pro (fréquence d'échantillonnage de 48 kHz, 16 bits par échantillon, niveau d'entrée de 30 dB, monocanal) ; ceux au Québec à l'aide d'un microphone sur pied Blue Yeti Nano Premium (fréquence d'échantillonnage de 44,1 kHz, 16 bits par échantillon, monocanal). Le post-traitement du signal audio consistait en un sous-échantillonnage à 16 kHz.

³⁷ L'intégration dans un texte nous a amenés à choisir un autre contexte vocalique que le contexte /øCø/ précédemment utilisé. En effet, il est très difficile de trouver des mots comprenant cette séquence pour chacune des consonnes du français, tout en préservant la lisibilité du texte. De plus, le contexte vocalique en /a/ permet d'éviter les phénomènes d'affrication qui peuvent apparaître en français québécois lorsque les consonnes /t/ et /d/ sont suivies des voyelles /i/ ou /y/.

5.2. Résultats et Discussion

5.2.1. Données descriptives

Avant d’aborder les deux objectifs principaux (reproductibilité et vitesse de lecture), nous allons exposer les données descriptives des mesures acoustiques effectuées sur les enregistrements du texte, afin de mettre en perspective les analyses subséquentes.

A. Normalité des données

Le test de Shapiro-Wilk a indiqué une distribution non gaussienne des mesures suivantes, pour lesquelles des statistiques non paramétriques ont donc été employées :

- qualité vocale (CPPs, $p=0,03$) ; slope, $p<0,001$) ;
- moments spectraux sur les consonnes ;
- pentes formantiques des trois premiers formants ($p<0,001$).

Pour la mesure de surface relative du triangle vocalique, le test de Shapiro-Wilk ne nous a pas permis de rejeter l’hypothèse nulle de normalité des données ($p=0,06$) ; des statistiques paramétriques ont ainsi été employées.

B. Données descriptives

Qualité vocale

Les données descriptives pour les trois mesures sur les six lectures par les 17 locuteurs sont reprises dans le [Tableau 22](#). Il n’existe actuellement à notre connaissance pas de données normatives pour le CPPs, le *slope* et le *tilt* mesurés sur des échantillons francophones. Pour comparaison, [MURTON 20] ont mesuré un score seuil de 9,33 dB pour le CPPs sur la lecture du « Rainbow passage » en anglais (un score inférieur indiquant un trouble vocal). [PHADKE 20] ont mesuré un CPPs moyen de 10,5 dB ($\pm 1,2$) sur une lecture de texte par 40 locuteurs sains finnois. Enfin, [BATTHYANY 19] ont mesuré un CPPs moyen de 7,45 dB ($\pm 0,77$) sur une lecture de texte en néerlandais par 123 locuteurs sains. Pour le *slope*, [ANTONETTI 20] ont mesuré une valeur moyenne de -23,69 dB ($\pm 4,10$) sur une tâche de comptage par 53 locuteurs brésiliens. Aucune valeur de comparaison n’a pu être trouvée pour le *tilt* tel que mesuré dans le script original de l’AVQI 03.01.

Tableau 22 – Données descriptives des mesures CPPs, slope et tilt pour les six lectures de texte par les 17 locuteurs

	CPPs		Slope		Tilt	
	médiane	[IIQ]	médiane	[EIQ]	médiane	[IIQ]
Lecture 1 (texte entier)	10,71 dB	[10,07 ; 12,12]	-27,43 dB	[-30,03 ; -24,55]	-11,66 dB	[-12,17 ; -11,06]
Lecture 2 (1 ^e partie)	11,4 dB	[11,02 ; 13,18]	-26,45 dB	[-29,85 ; -23,48]	-11,48 dB	[-12,48 ; -11,07]
Lecture 3 (1 ^e partie)	11,48 dB	[10,74 ; 12,64]	-25,38 dB	[-29,63 ; -22,97]	-11,65 dB	[-12,31 ; -11,12]
Lecture 4 (1 ^e partie)	11,64 dB	[10,81 ; 12,97]	-26,52 dB	[-29,16 ; -22,97]	-11,61 dB	[-12,36 ; -11,09]
Lecture 5 (1 ^e partie)	11,17 dB	[10,7 ; 13,03]	-27,12 dB	[-28,71 ; -24,49]	-11,77 dB	[-12,48 ; -11,1]
Lecture 6 (1 ^e p. rapide)	11,72 dB	[10,66 ; 12,49]	-25,34 dB	[-28,98 ; -22,69]	-11,52 dB	[-12,47 ; -11,14]

Note : Pour la lecture 1, seule la 1^e occurrence de la phrase visée a été retenue ; IIQ = intervalle interquartile

Mesure vocalique

Nous allons ici nous concentrer sur la surface relative du triangle vocalique utilisé par nos locuteurs sains, en utilisant le triangle à 2 E.T. (comme dans [VAN SON 18]). Les données descriptives pour l’aire relative sur les six lectures par les 17 locuteurs sont reprises dans le

Tableau 23. Nous pouvons observer que nos locuteurs sains utilisent une aire vocalique étendue (dépassant l'aire de référence).

Tableau 23 – Données descriptives de l'aire relative du triangle vocalique pour les six lectures par les 17 locuteurs

	Aire relative du triangle vocalique	
	moyenne	[E.T.]
Lecture 1 (texte entier)	127,53 %	[21,81]
Lecture 2 (1 ^e partie)	132,41 %	[20,30]
Lecture 3 (1 ^e partie)	130,18 %	[22,83]
Lecture 4 (1 ^e partie)	130,29 %	[25,50]
Lecture 5 (1 ^e partie)	123,88 %	[18,73]
Lecture 6 (1 ^e partie, lecture rapide)	119,82 %	[23,91]

Mesures consonantiques

Vingt-cinq occurrences phonémiques ont été exclues à cause de valeurs aberrantes (supérieures au 99^e percentile ou inférieures au 1^e) d'au moins un des moments spectraux mesurés.

Les données descriptives pour les quatre moments spectraux pour chaque phonème sur la lecture entière du texte par les 17 locuteurs sont reprises dans le **Tableau 24**. Les phonèmes qui se différencient significativement des autres phonèmes de la même demi-rangée selon les résultats d'un test de Dunn y sont mis en évidence.

Tableau 24 – Médiane [intervalle interquartile] des moments spectraux pour chaque phonème de la lecture du nouveau texte ; en gras : valeurs significativement différentes des autres sur la même demi-rangée

	s	ʃ	f	z	ʒ	v
	N=16	N=15	N=17	N=16	N=17	N=31
CoG	6576 [6305; 6804]	4308 [3809; 5099]	4730 [4334; 5184]	6447 [5469; 6808]	3944 [3550; 4281]	3933 [2287; 4339]
SKEW	-1,70 [-2,07; -1,53]	0,92 [0,21; 1,52]	-0,09 [-0,38; 0,10]	-1,40 [-1,77; -0,92]	1,18 [0,73; 1,91]	0,38 [0,03; 1,39]
SD	975 [753; 1159]	1143 [903; 1313]	1629 [1470; 1769]	1070 [916; 1434]	1019 [805; 1237]	1813 [1646; 2092]
KURT	5,12 [2,89; 7,12]	0,01 [-0,95; 3,31]	-0,51 [-1,04; -0,22]	2,88 [1,73; 4,52]	1,92 [-0,51; 4,27]	-0,98 [-1,17; 1,01]
	k	p	t	g	b	d
	N=68	N=149	N=61	N=28	N=16	N=17
CoG	3154 [2711; 3631]	3497 [2954; 3978]	4599 [3872; 5077]	2800 [2348; 3121]	2828 [2327; 3304]	3626 [3052; 3903]
SKEW	1,33 [0,56; 1,91]	0,61 [0,19; 1,13]	-0,003 [-0,54; 0,48]	1,82 [1,19; 2,83]	1,19 [0,79; 1,72]	0,48 [-0,12; 0,93]
SD	1276 [1043; 1512]	1606 [1386; 1802]	1753 [1417; 1982]	1004 [833; 1382]	1600 [1345; 1707]	1710 [1672; 1851]
KURT	0,97 [-0,57; 3,82]	-0,19 [-0,88; 0,88]	-0,68 [-1,12; 0,19]	3,95 [0,73; 10,38]	0,81 [-0,32; 3,22]	-0,52 [-0,96; -0,05]

Par rapport aux données de nos études précédentes (voir **Figure 70**), nous remarquons sur les fricatives des valeurs globalement plus élevées pour le CoG, avec un /s/ et un /z/ qui se démarquent comme précédemment des autres fricatives (voir **Figure 71**). Pour le SKEW, le /s/

et le /z/ montrent une valeur négative, alors que dans nos études précédentes, ces deux fricatives se démarquaient par un SKEW plus élevé. Le SD est globalement moins élevé, avec le /f/ et le /v/ se démarquant toujours par un SD plus élevé que les autres fricatives. Enfin, comme précédemment, les fricatives /s/ et /z/ se démarquent par un KURT élevé.

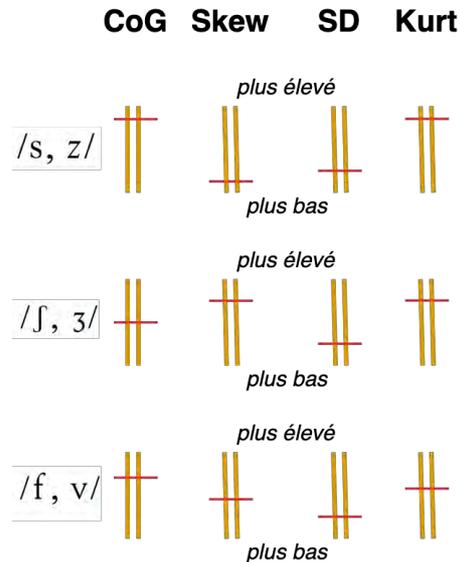


Figure 70 – Représentation schématique des valeurs des moments spectraux pour les fricatives dans le texte par rapport à celles extraites des phrases porteuses

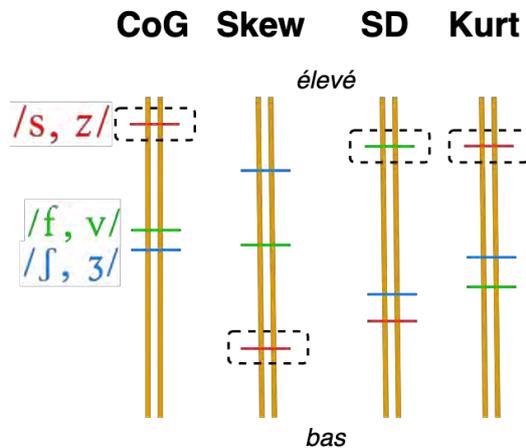


Figure 71 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les fricatives dans le texte ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres

Sur les plosives, la démarcation du /k/ et du /g/ par son CoG bas précédemment observée se voit diminuée. Les valeurs de SKEW sont comparables avec celles précédemment mesurées (voir [Figure 72](#)). Le SD est cette fois le plus bas pour ces mêmes phonèmes, alors qu'un SD bas caractérisait les plosives /t/ et /d/ dans nos études précédentes (voir [Figure 73](#)). Le KURT montre également des valeurs et tendances dissimilaires, le /p/ et le /b/ ne se démarquant plus par leur KURT négatif.

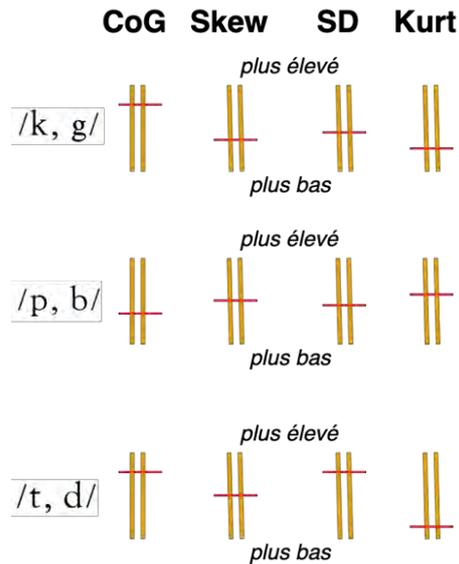


Figure 72 – Représentation schématique des valeurs des moments spectraux pour les plosives dans le texte par rapport à celles extraites des phrases porteuses

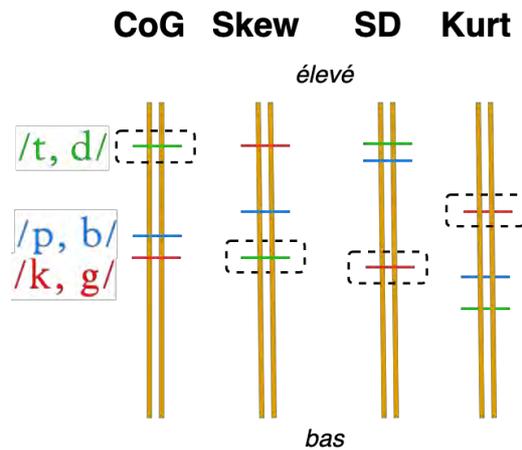


Figure 73 – Représentation schématique des moments spectraux en fonction du lieu d'articulation pour les plosives dans le texte ; encadrés en pointillé, les phonèmes qui se différencient significativement des autres

Nous émettons deux hypothèses quant aux différences de valeurs observées sur notre texte par rapport à nos études précédentes.

La première concerne l'échantillon de parole employé. Dans nos études précédentes, les consonnes étaient extraites de phrases porteuses avec un contexte contrôlé de structure « Le sac /øCø/ convient ». Les consonnes de notre nouveau texte, elles, sont entourées de la voyelle /a/, et le contexte phonémique plus éloigné est variable. Pour rappel, nous avons observé une influence du contexte phonémique sur les moments spectraux dans notre étude pilote sur un extrait de « Monsieur Seguin » (cf. [Annexe O](#)). Nos nouvelles observations, particulièrement les valeurs pour les fricatives, tendent à conforter cette conclusion. En effet — hormis pour le SKEW, dont nous avons déjà souligné que l'interprétation dans les fricatives doit être faite avec précaution au vu de leur profil spectral — les valeurs diffèrent de manière quasiment homogène dans le contexte en /a/, les tendances restant similaires par rapport au contexte en /ø/. Nous avons donc mené une analyse de reproductibilité des moments spectraux à travers les multiples occurrences consonantiques de notre nouveau texte. Pour cela, nous avons employé le test de Friedman, avec des comparaisons multiples à l'aide du test des rangs signés de

Wilcoxon pour comparer chaque moment spectral à travers les multiples occurrences de chaque phonème dans la lecture du texte entier, le cas échéant. Seules quatre différences significatives ont été détectées, toutes sur les plosives : pour le CoG entre la 3^e et la 6^e occurrence du /p/ ($p=0,0468$) et entre la 6^e et la 9^e occurrence du /p/ ($p=0,0494$), entre la 1^e et la 3^e occurrence du /k/ ($p=0,0495$) et entre la 1^e et la 4^e occurrence du /k/ ($p=0,0395$). Ces résultats nous amènent à conclure que l’effet de coarticulation est principalement limité aux phonèmes voisins directs, étant donné que les mesures de moments spectraux restent globalement stables à travers les consonnes en contexte /aCa/. Le contexte plus éloigné, bien que variable, semble ainsi peu influencer sur les moments spectraux. Des études plus approfondies sont nécessaires pour mieux comprendre l’influence de la coarticulation sur les moments spectraux et confirmer qu’un contexte vocalique stable intégré dans un texte permet des mesures spectrales reproductibles malgré le contexte autrement variable.

Notre deuxième hypothèse pour expliquer les différences de valeurs observées concerne la méthode d’analyse. Dans la présente preuve de concept, nous avons préconisé un outil facilement disponible en ligne pour en permettre l’utilisation par les cliniciens et chercheurs de tous horizons. En observant de plus près les résultats de l’alignement forcé WebMAUS Basic et de la méthode FBDS, nous nous sommes rendu compte que — par rapport à l’outil du LIA employé dans nos explorations acoustiques précédentes — cet alignement présente des difficultés pour les plosives, particulièrement pour /p/ et /g/. La frontière de droite du segment consonantique est à de multiples reprises décalée, tantôt incluant un début de la voyelle après le burst (dans le cas du /g/), tantôt excluant totalement le burst du segment consonantique (dans le cas du /p/). Ces décalages perturbent en conséquence la détection du burst par l’algorithme FBDS, qui performe correctement dès lors que l’alignement consonantique est juste. Cette observation peut expliquer les nombreuses différences observées particulièrement pour les plosives. Aussi — la méthode FBDS étant dépendante de la qualité de l’alignement forcé préalable — il est primordial de s’assurer de la bonne performance de l’outil d’alignement utilisé.

Mesures semi-consonantiques

Notons que seule la lecture entière du texte contient la semi-consonne /ɥ/.

Les données descriptives pour les trois mesures sur les six lectures par les 17 locuteurs sont reprises dans le [Tableau 25](#).

Tableau 25 – Médiane [intervalle interquartile] des pentes formantiques des semi-consonnes dans le nouveau texte

	F1 slope		F2 slope		F3 slope	
	médiane	[IIQ]	médiane	[IIQ]	médiane	[IIQ]
/ɥ/ (N=17)	-0,29 Hz/ms	[-0,70 ; 0,43]	0,84 Hz/ms	[-1,13 ; 6,81]	-1,71 Hz/ms	[-8,78 ; 3,79]
/w/ (N=255)	2,02 Hz/ms	[-1,55 ; 4,44]	-1,15 Hz/ms	[-8,80 ; 3,64]	-4,68 Hz/ms	[-11,83 ; 0,31]
/j/ (N=210)	1,34 Hz/ms	[0,26 ; 2,72]	-2,72 Hz/ms	[-6,42 ; -0,34]	-3,40 Hz/ms	[-6,80 ; -0,02]

Note : N = nombre total d’occurrences phonémiques analysées ; IIQ = intervalle interquartile

5.2.2. Objectif 1 : Reproductibilité

Le test des rangs signés de Wilcoxon (avec correction de Bonferroni pour comparaisons multiples) a été utilisé pour comparer les mesures de qualité vocale, les mesures consonantiques et semi-vocaliques sur les différentes répétitions du texte (lectures 1 à 5). Une ANOVA à un

facteur sur mesures répétées a été employée pour examiner l’effet des répétitions sur la mesure de l’aire du triangle vocalique.

Pour la **qualité vocale**, nous observons que les trois mesures (CPPs, *slope* et *tilt*) sont assez stables au fil des répétitions chez nos sujets sains (voir p. ex. [Figure 74](#)). Aucune différence statistiquement significative n’est observée.

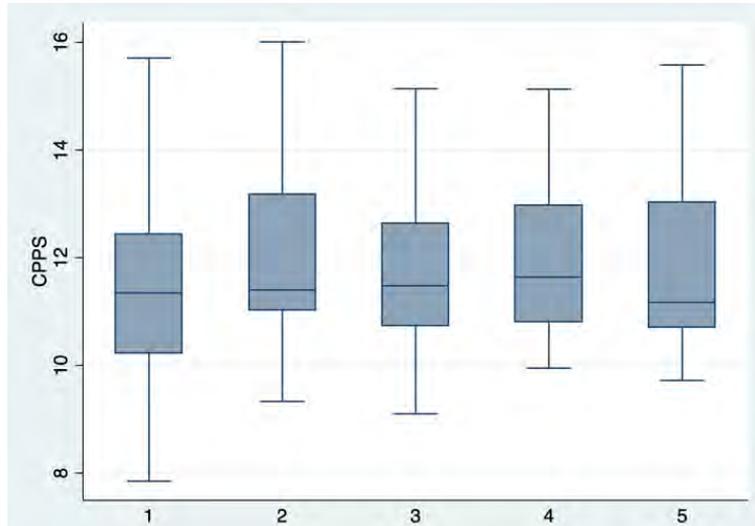


Figure 74 – Boxplots des valeurs du CPPs (en dB) au fil des cinq lectures

Pour la mesure de l’**aire du triangle vocalique**, les résultats montrent que les valeurs sur les répétitions sont statistiquement différentes ($F=25,99$; $p<0,001$). Les comparaisons post-hoc à l’aide du test de Tukey ont montré une seule différence statistiquement significative, entre la deuxième et la cinquième lecture ($p=0,028$) (voir [Figure 75](#)).

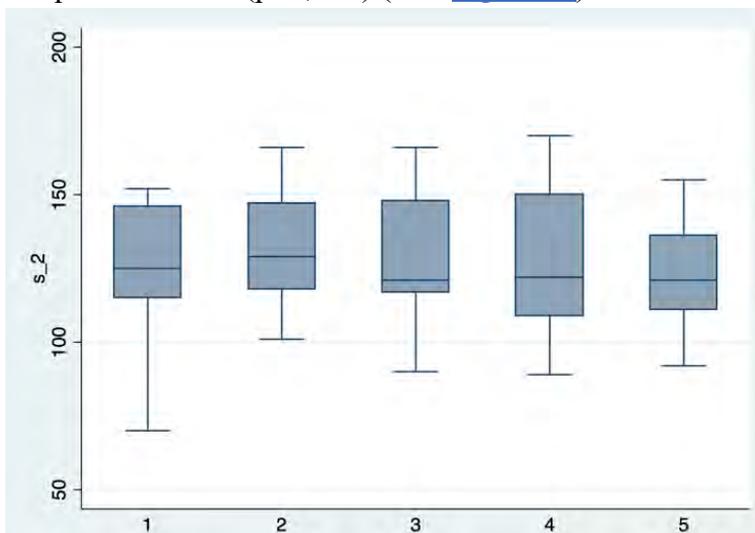


Figure 75 – Boxplots des valeurs de l’aire relative du triangle vocalique (en %) au fil des cinq lectures

Pour les **moments spectraux consonantiques**, étant donné que la première partie de notre texte ne comprend que deux consonnes non-sonantes (/p/ et /b/) dans le contexte /aCa/, seules ces deux occurrences ont pu être comparées à travers les lectures. Aucune différence statistiquement significative n’a été mesurée pour les moments spectraux sur ces deux phonèmes à travers les cinq premières lectures (voir [Figure 76](#)). Ces résultats de reproductibilité restent à compléter par une étude portant sur toutes les consonnes sonantes à travers des productions répétées.

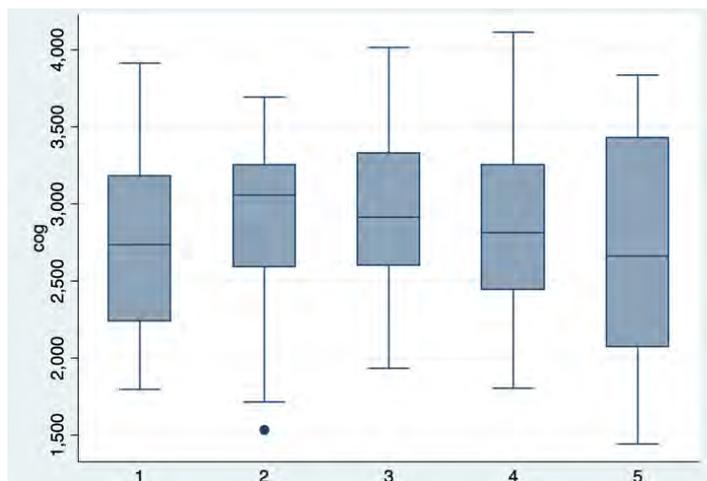


Figure 76 – Boxplots des valeurs du CoG (en Hz) pour la consonne /b/ au fil des cinq lectures

Pour les mesures **semi-consonantiques**, l'analyse de reproductibilité n'a pu être effectuée que pour les semi-consonnes /w/ et /j/, le /ɥ/ n'apparaissant qu'une fois lors de la lecture du texte entier. Afin de pouvoir comparer les cinq premières lectures, seules les occurrences de la première partie ont été retenues pour la lecture du texte entier.

Seule une différence statistiquement significative a été mesurée pour la pente de F3 dans la première occurrence du /w/ (dans « rejoignent »), entre la 3^e et la 5^e lecture ($p=0,002$). Aucune différence significative n'a été mesurée pour la seconde occurrence du /w/ (dans « voyage »), ni pour le /j/ (dans « voyage »). La mesure de la pente de F2, qui est considérée comme une mesure de la vitesse de modification du tractus phonatoire et utilisée en parole pathologique [KENT 89a; MARTEL-SAUVAGEAU 17; YUNUSOVA 12], est donc elle aussi reproductible dans notre nouveau texte (voir [Figure 77](#)).

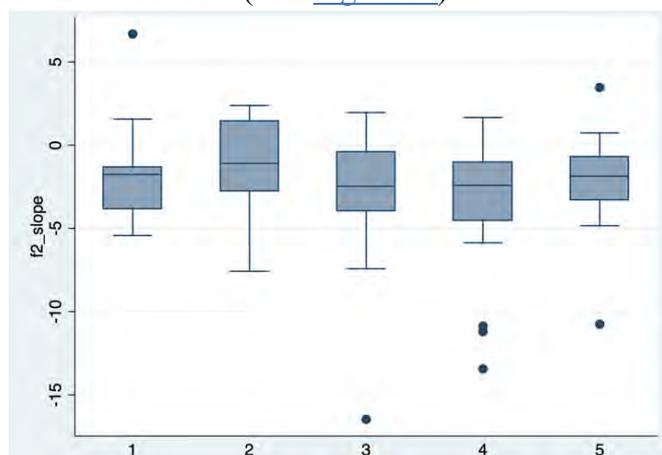


Figure 77 – Boxplots des valeurs de la pente de F2 pour la semi-consonne /j/ au fil des cinq lectures

5.2.3. Objectif 2 : Vitesse de lecture

Afin d'investiguer l'effet de la lecture rapide sur les différentes mesures, nous avons comparé cette dernière lecture aux cinq autres lectures.

Pour la **qualité vocale**, aucune différence statistiquement significative n'a été révélée pour les trois mesures.

Pour la mesure de l'**aire du triangle vocalique**, des différences statistiquement significatives ont été révélées par les comparaisons post-hoc à l'aide du test de Tukey entre la lecture rapide et la deuxième ($p=0,001$), la troisième ($p=0,008$) et la quatrième ($p=0,007$)

lecture. La lecture rapide induit donc une légère «réduction vocalique», l'aire restant cependant supérieure au triangle de référence.

Pour les **moments spectraux consonantiques**, les différences entre la lecture rapide et les cinq autres lectures ne sont pas statistiquement significatives avec la correction de Bonferroni (CoG lecture 2-6 : $p=0,019$; SD lecture 1-6 : $p=0,035$; SD lecture 2-6 : $p=0,031$; SKEW lecture 2-6 : $p=0,017$; KURT lecture 5-6 : $p=0,044$). Pour les bilabiales /p, b/, malgré un CoG et un SD quelque peu plus élevés et un SKEW et un KURT légèrement plus bas, la mesure des moments spectraux semble donc peu influencée par la lecture rapide.

Pour les mesures des **pentés de F1, F2 et F3 sur les semi-consonnes**, bien que de légères différences soient observées pour la pente de F3, celles-ci ne sont pas statistiquement significatives avec la correction de Bonferroni pour les comparaisons multiples.

5.2.4. Objectifs secondaires

Faisabilité et lisibilité

Nous avons calculé un score de facilité de lecture de Flesch de 55,5 (0 = très difficile à lire, 100 = très facile à lire pour les 10 ans et +) (cf. « 3.2.3 [Conception du texte](#) »). Nos enregistrements confirment qu'après une lecture silencieuse du texte entier, sa lecture à voix haute ne semble pas poser de difficultés particulières. En effet, très peu d'hésitations sont observées. Seuls les locuteurs du Québec indiquent que certains mots sont peu ou pas employés au quotidien (p. ex. « bigorneau », « la java » et « macadamia » qui se dit « macadame » en français québécois), sans toutefois que cela n'entrave la fluidité et le naturel de leur lecture à voix haute. Notre nouveau texte permet donc de mettre à l'épreuve certains processus articulatoires tout en respectant la fluidité du débit de lecture ainsi que de multiples variations prosodiques.

Fatigabilité ou échauffement vocal

Lors de la lecture du texte entier, la phrase voisée est lue à deux reprises. Nous pouvons ainsi effectuer les trois mesures de qualité vocale sur les deux occurrences (voir [Tableau 26](#)). Cela permet notamment de mesurer un éventuel changement de qualité vocale dû à la fatigabilité, ou au contraire un échauffement vocal. Le test des rangs signés de Wilcoxon révèle une différence statistiquement significative ($p=0,01$) entre le CPPs de la première et de la seconde occurrence. Aussi, l'observation des données montre que le CPPs de la première occurrence est généralement moins élevé que celui de la seconde. Ceci peut s'expliquer par un effet d'échauffement vocal ou par une habitude au texte, qui peut aussi s'observer dans le [Tableau 22](#).

Tableau 26 – Données descriptives des mesures CPPs, slope et tilt pour les deux occurrences de la phrase entièrement voisée lors de la lecture du texte entier par les 17 locuteurs

	CPPs		Slope		Tilt	
	médiane	[IIQ]	médiane	[IIQ]	médiane	[IIQ]
Occurrence 1	11,44 dB	[10,57 ; 12,64]	-26,41 dB	[-29,36 ; -22,97]	-11,63 dB	[-12,38 ; -11,09]
Occurrence 2	11,88 dB	[10,52 ; 13,12]	-26,58 dB	[-29,66 ; -22,14]	-11,67 dB	[-12 ; -11,4]

Note : IIQ = intervalle interquartile

Effet du genre

Nous avons analysé les différences de valeurs entre les locuteurs masculins et féminins pour les quatre catégories de mesures. Les données descriptives pour les différences statistiquement significatives sont reprises dans le [Tableau 27](#).

Tableau 27 – Données descriptives des mesures pour lesquelles une différence significative de genre a été mesurée

Catégorie	Mesure	Hommes (N=7)	Femmes (N=10)
Qualité vocale	CPPs	<i>Médiane [IIQ]</i> 10,81 dB [10,12 ; 11,71]	<i>Médiane [IIQ]</i> 12,14 dB [10,88 ; 13,68]
Voyelles	Aire relative du triangle	<i>Moyenne [E.T.]</i> 119,95 % [23,49]	<i>Moyenne [E.T.]</i> 132,63 % [19,78]
Semi-consonnes	Pente de F1	<i>Médiane [IIQ]</i> -1,10 Hz/ms [-5,97 ; 1,58]	<i>Médiane [IIQ]</i> 3,43 Hz/ms [1,32 ; 5,0]
	Pente de F3	-7,14 Hz/ms [-13,58 ; -2,62]	-2,89 Hz/ms [-10,50 ; 1,93]

Note : IIQ = intervalle interquartile ; E.T. = écart-type

Pour les **mesures de qualité vocale**, un test de Mann-Whitney a montré une différence statistiquement significative pour le CPPs ($z=4,67$; $p<0,0001$). Ainsi, les locutrices semblent avoir une qualité vocale globale légèrement plus élevée que les hommes. En effet, le pic cepstral représente les pics harmoniques du spectre vocal. Une prééminence de ce pic plus importante indique une composition harmonique spectrale périodique et régulière, qui se traduit par la perception d’une bonne qualité vocale. Dans le cas d’un signal vocal pathologique, les harmoniques se démarquent peu dans le spectre et le pic cepstral est diminué [MURTON 20].

Un test T de Student ($t= 2,93$; $p=0,0043$) a révélé que les locutrices montrent également une **aire relative du triangle vocalique** significativement supérieure (moyenne = 132,63 ; E.T. = 19,78) à celle des locuteurs (moyenne = 119,95 ; E.T. = 23,49). Notons que le script adapte ses calculs en fonction du genre du locuteur.

Pour les **moments spectraux consonantiques**, le test de Kruskal-Walis avec le test de Dunn (avec correction de Bonferroni) pour les comparaisons par paires (phonème par phonème) n’ont révélé aucune différence significative entre les genres pour un même phonème.

Enfin, pour les **mesures semi-consonantiques**, nous avons également employé le test de Kruskal-Walis avec le test de Dunn pour les comparaisons par paires (phonème par phonème, pour chaque mesure, avec correction de Bonferroni). Des différences significatives ont été mesurées pour la pente de F1 et de F3 sur le /w/ ($p<0,001$ et $p=0,0014$, respectivement).

Effet de la variante régionale du français

Nous avons également souhaité comparer les valeurs pour les trois pays. Les données descriptives pour les différences statistiquement significatives sont reprises dans le [Tableau 28](#).

Tableau 28 – Données descriptives des mesures CPPs, slope et tilt pour les locuteurs belges, canadiens et français

Catégorie	Mesure	Belgique (N=6)	Canada (N=3)	France (N=8)
Qualité vocale	CPPs	<i>Médiane [IIQ]</i> 10,83 dB [10,19 ; 12,16]	<i>Médiane [IIQ]</i> 13,68 dB [12,64 ; 15,13]	<i>Médiane [IIQ]</i> 11,4 dB [10,58 ; 12,08]
	Slope	-26,55 dB [-28,71 ; -23,51]	-15,31 dB [-19,58 ; -12,95]	-28,31 dB [-30,39 ; -26,07]
	Tilt	-11,96 dB [-12,36 ; -11,1]	-12,63 dB [-13,24 ; -11,51]	-11,41 dB [-11,8 ; -11,06]
Voyelles	Aire relative du triangle	<i>Moyenne [E.T.]</i> 143,33 % [13,37]	<i>Moyenne [E.T.]</i> 115,17 % [11,16]	<i>Moyenne [E.T.]</i> 119,94 % [23,84]

Note : IIQ = intervalle interquartile ; E.T. = écart-type

Pour les **mesures de qualité vocale**, un test de Kruskal-Wallis révèle des différences significatives pour les trois mesures ($p=0,0001$ pour CPPs et *slope*, $p=0,0009$ pour *tilt*). Les comparaisons multiples à l'aide du test de Dunn (avec correction de Bonferroni) montrent que les locuteurs canadiens présentent un CPPs et un *slope* significativement plus élevés que les locuteurs français et belges ($p<0,0001$ respectivement). La mesure de *slope* est de plus significativement plus basse chez les locuteurs français par rapport aux locuteurs belges ($p<0,03$). Enfin, le *tilt* ne se différencie que chez les locuteurs canadiens par rapport aux locuteurs français ($p=0,0004$). Ainsi, les locuteurs canadiens semblent avoir un spectre vocal plus périodique, mais avec une pente spectrale plus inclinée qui indique une moindre richesse en harmoniques aiguës.

Pour l'**aire relative du triangle** vocalique, une ANOVA à un facteur a montré que les valeurs en fonction du pays sont statistiquement différentes ($F=20,46$; $p<0,001$). Des comparaisons post-hoc à l'aide du test de Tukey ont révélé que les valeurs des locuteurs belges se différencient significativement de celles des français ($p<0,001$) et de celles des canadiens ($p<0,001$). L'aire relative du triangle vocalique des locuteurs belges est donc significativement plus grande.

En ce qui concerne les **moments spectraux consonantiques** et les **mesures semi-consonantiques**, le test de Kruskal-Wallis avec le test de Dunn pour les comparaisons par paires (phonème par phonème, pour chaque moment avec correction de Bonferroni) n'a révélé aucune différence significative entre les trois variantes régionales pour un même phonème.

5.3. Conclusion du chapitre

Le protocole que nous avons mis en place permet une extraction rapide et aisée de quatre types de mesures sur notre nouveau texte de référence : qualité vocale, voyelles, consonnes et semi-consonnes. L'étude de faisabilité que nous avons menée indique qu'il est tout à fait possible d'employer des mesures acoustiques sur notre nouveau texte en employant uniquement des outils libres d'utilisation, et que celles-ci sont globalement reproductibles.

Toutes les mesures acoustiques investiguées se montrent reproductibles à travers les multiples lectures de notre texte. La lecture rapide de ce dernier, quant à elle, induit une légère « réduction vocalique », l'aire relative du triangle restant cependant supérieure au triangle de référence. Les autres mesures semblent peu influencées par la vitesse de lecture.

Un effet du genre a été mesuré pour la qualité vocale (CPPs) et pour l'aire relative du triangle vocalique. Les mesures de moments spectraux consonantiques ne sont quant à elles pas influencées par le genre. De même, alors que la pente de F1 et de F3 pour le /w/ diffère selon le genre, la pente de F2 n'est pas influencée par ce dernier.

Enfin, un effet de la variante régionale du français a été mesuré pour les trois mesures de qualité vocale et pour l'aire relative du triangle vocalique. Les moments spectraux consonantiques et les mesures semi-consonantiques ne diffèrent pas entre les variantes régionales pour un même phonème.

Une étude de validation approfondie des mesures acoustiques sur notre nouveau texte sur des échantillons plus importants dans les trois variantes régionales est nécessaire afin de permettre une éventuelle utilisation en pratique clinique.

Conclusion générale

Conclusion

L'objectif principal de ce travail, à travers un regard pluridisciplinaire alliant les domaines du traitement du signal et de l'orthophonie, était d'étudier et contribuer à l'amélioration des outils de mesure acoustique de l'intelligibilité de la parole. Il répond non seulement à la nécessité d'une évaluation exhaustive et fiable de la parole en pratique clinique, mais également à la nécessité en recherche scientifique d'une bonne connaissance des mesures acoustiques et de leur comportement en parole saine avant leur application à la parole pathologique.

Dans un premier temps, nous avons rédigé des définitions étayées de l'intelligibilité et de la compréhensibilité de la parole. Notre **enquête de consensus internationale Delphi** [POMMEE 21b] nous a permis d'inclure dans ces définitions tous les éléments consensuels recueillis au fil du processus et de prendre en compte les points de vue de différents domaines d'expertise. En plus d'une meilleure compréhension de ces deux termes, cette étude nous a permis de spécifier leur évaluation en décrivant les tâches et mesures qui correspondent le mieux à leur définition.

Nous nous sommes ensuite penchés plus en détail sur la question de la mesure de l'intelligibilité. Nous avons pour cela effectué une analyse des besoins structurée selon les trois piliers de l'Evidence-Based Practice.

Le premier pilier, celui de la **recherche scientifique**, a été abordé au travers d'une **revue systématique de la littérature** [POMMEE 21c] selon la méthodologie PRISMA. L'objectif de cette étude était d'identifier les besoins en termes de mesure de la parole saine dans la littérature scientifique. Nos résultats ont mis en évidence la nature variable et « imprécise » de la parole chez les locuteurs adultes sains. Une meilleure compréhension de ces imprécisions inhérentes est nécessaire pour fournir une ligne de base plus réaliste à l'étude de la parole pathologique. Cette étude souligne donc le besoin de recherches approfondies pour mieux comprendre les comportements des mesures acoustiques chez le locuteur sain et identifier les combinaisons acoustiques pertinentes qui pourraient expliquer les variations perçues de la parole. Notre revue met notamment en avant la nécessité d'investiguer davantage les mesures acoustiques spectrales sur les consonnes, tout en considérant les aspects suprasegmentaux. En plus de la parole elle-même, la terminologie, la méthodologie et les tâches d'évaluation montrent également une grande variabilité à travers les études retenues. Ce constat nous a permis de souligner la nécessité de fournir dans les études des définitions non ambiguës des principaux concepts abordés. De plus, décrire de manière suffisamment détaillée les tâches d'évaluation et les mesures acoustiques employées est primordial pour permettre la réplique, les comparaisons croisées ou méta-analyses, et la mise en commun des données. Les facteurs genre et âge doivent en outre être davantage explicités, en précisant a minima les informations sur leur répartition dans les groupes d'étude.

Le second pilier de l'EBP, celui de l'**expérience clinique**, nous a amenés à aborder l'évaluation de la parole en clinique orthophonique francophone. En effet, ce n'est qu'en prenant en compte les opinions des cliniciens que le transfert en pratique clinique d'un outil

créé par des experts scientifiques peut se faire de manière idéale. Aussi, nous avons mené une **enquête auprès de cliniciens** [POMMEE 21a] concernant les pratiques actuelles et les manques et besoins en termes d'évaluation de la parole. Les résultats de cette étude soulignent la nécessité de proposer de nouveaux outils d'évaluation des troubles de la parole qui sont fiables/reproductibles, faciles à utiliser et financièrement accessibles. L'évaluation s'en trouverait plus simple, précise, rapide, écologique et pertinente dans le contexte des multiples pathologies pouvant affecter la parole, avec des critères d'interprétation et des repères normatifs clairement définis. La création d'un nouvel outil standardisé d'évaluation de la parole répondant à ces critères, incluant notamment des mesures acoustiques, permettrait une bonne adéquation avec les besoins exprimés par les thérapeutes. Cette étude révèle entre autres l'intérêt de mesures acoustiques plus spécifiques au domaine de l'articulation de la parole/des phonèmes, permettant au clinicien d'obtenir une analyse rendant plus directement compte de la performance de production de la parole de ses patients.

Enfin, pour qu'un tel nouvel outil soit fiable et valide, et compte tenu de la familiarité limitée de nombreux cliniciens avec les paramètres d'enregistrement détaillés, une formation plus exhaustive des cliniciens à l'évaluation acoustique de la parole, ainsi qu'une standardisation et une simplification de la procédure d'enregistrement et des conditions de stockage des données semblent nécessaires. En outre, l'évaluation bénéficierait également d'une standardisation des dimensions à évaluer, des tâches, ainsi que des instructions fournies. Sur la base de ces constats, un ensemble de tâches minimales ainsi que des suggestions pour les conditions d'enregistrement ont été fournis à titre de recommandation pour les cliniciens. De plus, les explorations acoustiques menées au cours de ce travail ont été pensées avec une perspective d'applicabilité en contexte clinique pour une évaluation plus objective de la parole.

Notre troisième étude, à l'intersection entre les piliers de la recherche scientifique et de l'expertise, a consisté en des **interviews semi-dirigées auprès de chercheurs** dans différents domaines applicatifs liés à la parole. Tout comme dans la revue systématique et l'enquête aux cliniciens, les données de cette étude ont montré une grande variabilité de la terminologie liée à la parole, de la parole saine, et des méthodologies de mesure de celle-ci. Face à la subjectivité des mesures perceptives et au caractère chronophage de certaines d'entre elles, le besoin de mesures plus « objectives » et fiables a été confirmé (notamment sur les consonnes), mais il reste à identifier les mesures les plus pertinentes pour rendre compte de l'intelligibilité de la parole et des autres concepts. Enfin, les interviews ont révélé que les aspects pratiques (coût, manque de temps, manque de maîtrise des outils informatiques/de connaissances en acoustique) freinent encore le transfert des avancées scientifiques en outils cliniques.

La quatrième étude de notre analyse des besoins adresse le troisième pilier de l'EBP, c.-à-d. le **point de vue du patient**. Via un **questionnaire adressé aux patients** ayant été évalués pour un trouble de la parole, nous avons recueilli leur opinion quant au bilan de parole. Cette étude nous a permis de souligner l'importance de l'enregistrement d'échantillons de parole lors des bilans cliniques initiaux et de suivi, que ce soit pour l'évaluation perceptive ou acoustique. De plus, les patients interrogés ont exprimé le désir d'une meilleure prise en compte des conséquences fonctionnelles et psychosociales de leur trouble de la parole. Conformément à la démarche EBP, ceci souligne l'importance de l'utilisation de questionnaires d'impact fonctionnel et de qualité de vie, en plein essor dans la recherche scientifique actuelle.

Tous les besoins identifiés en termes de mesure de la parole ainsi que nos contributions pour y répondre sont synthétisés dans le [Tableau 12](#).

Après avoir fait le point sur les besoins en termes de mesures de la parole en recherche scientifique et en pratique clinique, nous nous sommes penchés sur les **supports et échantillons de parole** sur lesquels la parole peut être évaluée. En effet, elle peut être appréciée perceptivement ou instrumentalement sur des échantillons de différents niveaux de granularité, plus ou moins fonctionnels. Après un bref état des lieux des outils et supports d'évaluation utilisés en langue française, nous avons décrit plus en détail la contribution des phonèmes à l'intelligibilité, ainsi que l'importance des textes de référence pour l'évaluation de la parole. Au vu des limites des textes existants, nous avons mis en place un groupe de travail international d'experts francophones belges, français et québécois afin de construire un **nouveau texte de référence pour l'évaluation de la parole et de la voix**. Notre étude internationale de consensus Delphi nous a permis de sélectionner les critères les plus importants pour la création de ce texte, qui a ensuite été validé par l'ensemble des membres de notre groupe. Inspiré du texte du protocole MonPaGe [TROUVILLE 21], il a été pensé spécifiquement pour l'évaluation de la parole et de la voix, et rend possible des mesures acoustiques afin de compléter l'évaluation perceptive. Il a notamment été construit afin de pouvoir permettre l'extraction d'un maximum de données sur la lecture de son premier paragraphe. La faisabilité et la reproductibilité de mesures acoustiques sur ce texte ont été abordées dans notre dernier chapitre.

Tout au long de ce travail, nous avons identifié un besoin de mesures moins subjectives et plus reproductibles de la parole. Ce besoin a été identifié non seulement dans la population des cliniciens, mais également dans le domaine de la recherche scientifique sur la parole. Pour y répondre, nous avons abordé dans notre dernier chapitre l'un des objectifs principaux de la recherche en parole : la caractérisation acoustique des phonèmes — dans notre cas plus spécifiquement des consonnes. Nous avons ainsi mené des **analyses spectrales consonantiques**, dans l'objectif de proposer de nouvelles pistes de mesures plus objectives de l'intelligibilité, utilisables en clinique et en recherche.

Après une étude pilote tentant de reproduire une « **banane de la parole** » [POMMEE 19a] — une aire consonantique construite à l'aide d'une détection de pics spectraux — nous avons investigué en détail le comportement des **moments spectraux** en parole saine, dans les consonnes non-sonantes du français [POMMEE 21d]. Ces mesures spectrales considèrent le spectre consonantique dans son entièreté et répondent ainsi à l'intérêt d'une analyse considérant l'aspect global du spectre des consonnes.

Dans un premier temps, nous avons investigué ces mesures sur base d'une **segmentation consonantique manuelle**. Une combinaison des deux paires CoG-SKEW et SD-KURT semble être la meilleure recommandation pour distinguer acoustiquement les différentes consonnes. Les moments spectraux, tout en reflétant la variabilité de la parole saine, permettent par leur combinaison la distinction des phonèmes prononcés. Le CoG est significativement plus élevé dans les fricatives alvéolaires et plus bas dans les plosives palatales. Le SD est significativement plus élevé dans les fricatives labiodentales et les plosives bilabiales, alors que les plosives alvéolaires montrent une dispersion moins importante de la concentration spectrale.

Cependant, la segmentation manuelle pose non seulement le problème de la subjectivité, mais aussi de l'aspect chronophage en vue d'une éventuelle utilisation en pratique clinique.

Ainsi, nous avons investigué l'utilisation des moments spectraux sur base de l'**alignement forcé** de nos stimuli de parole. L'alignement forcé employé seul ne permettant pas une détection suffisamment précise des bursts dans les plosives, nous avons développé un nouvel algorithme combinant cet alignement forcé avec la méthode de **segmentation « forward-backward divergence » (FBDS)** [ANDRE-OBRECHT 88]. Nos résultats montrent que cette méthode permet effectivement une détection fidèle à la segmentation manuelle du burst des plosives et de la fenêtre d'analyse dans un segment stable pour les fricatives. Les distinctions entre les lieux d'articulation à l'aide des moments spectraux issus de notre algorithme sont similaires à ceux issus de la segmentation manuelle. La mesure des moments spectraux est donc réalisable de manière automatisée à l'aide de l'algorithme que nous avons développé. Afin d'en aborder la question de la pertinence clinique, nous avons ensuite exploré leur utilisation **sur des enregistrements de patients** traités pour un cancer de la sphère ORL. Les résultats suggèrent que l'analyse des moments spectraux fournit des informations utiles pour décrire les singularités acoustiques et perceptives de ces locuteurs. En effet, l'étude conjointe des résultats perceptifs et acoustiques de quatre patients a montré que les moments spectraux peuvent être interprétés en lien avec l'intelligibilité perçue. Pour chaque consonne, l'interprétation requiert la prise en compte relative des quatre moments spectraux en regard de ceux des autres consonnes.

Enfin, alors que nos explorations acoustiques précédentes portaient sur des stimuli en contexte phonémique contrôlé, nous avons démontré la faisabilité de **mesures acoustiques sur notre nouveau texte de référence**. Nous avons pour cela rédigé un protocole facile d'utilisation par des professionnels non experts en informatique, permettant l'extraction de mesures de qualité vocale, de mesures vocaliques, de mesures consonantiques et d'une mesure semi-consonantique. La lecture à voix haute de notre texte permet à la fois de mettre à l'épreuve certains processus articulatoires, mais aussi d'obtenir un échantillon de parole avec un débit de lecture fluide et de multiples variations prosodiques. Toutes les mesures acoustiques investiguées se montrent reproductibles à travers les multiples lectures de notre texte. La lecture rapide de ce dernier, quant à elle, induit une légère « réduction vocalique », l'aire relative du triangle restant cependant supérieure au triangle de référence. Les autres mesures semblent peu influencées par la vitesse de lecture. Un effet du genre a été mesuré pour la qualité vocale (CPPs) et pour l'aire relative du triangle vocalique. Les mesures de moments spectraux consonantiques ne sont quant à elles pas influencées par le genre. De même, alors que la pente de F1 et de F3 pour le /w/ diffère selon le genre, la pente de F2 n'est pas influencée par ce dernier. Enfin, un effet de la variante régionale du français a été mesuré pour les trois mesures de qualité vocale et pour l'aire relative du triangle vocalique. Les moments spectraux consonantiques et les mesures semi-consonantiques ne diffèrent pas entre les variantes régionales pour un même phonème.

Cette étude de faisabilité indique qu'il est tout à fait possible d'employer des mesures acoustiques sur notre nouveau texte en employant uniquement des outils libres d'utilisation, et que celles-ci sont globalement reproductibles. Cependant, une étude de validation approfondie des mesures acoustiques sur notre nouveau texte sur des échantillons plus importants dans les trois variantes régionales est nécessaire afin de permettre une éventuelle utilisation en pratique clinique ou de recherche.

Perspectives

De nombreux points enraient encore le développement et le transfert en pratique courante de ces nouvelles mesures plus « objectives » et reproductibles, tant en pratique clinique que dans la recherche scientifique. Aussi, de multiples perspectives se sont dégagées pour de futures études dans le domaine de la parole saine et pathologique, qui permettront de progresser davantage vers la création de nouveaux outils de mesure de l'intelligibilité.

Terminologie

Tout d'abord, de nombreuses ambiguïtés persistent dans la terminologie liée à l'évaluation de la parole. Par exemple, des adjectifs tels que « objective, instrumentale, automatique, quantitative, qualitative » sont fréquemment utilisés pour désigner l'évaluation de la parole, mais leur définition reste floue. D'autres termes tels « la mesure, l'outil, le moyen » d'évaluation peuvent également prêter à confusion. Par conséquent, des études supplémentaires sont nécessaires pour **clarifier les questions liées à la terminologie**, générer des définitions moins équivoques et faciliter les avancées de la recherche sur la parole et la voix grâce à une meilleure communication entre les experts. À cette fin, la méthodologie Delphi semble être un moyen approprié. Dans l'attente d'une terminologie plus consensuelle dans le domaine de l'évaluation de la parole, il est primordial que les auteurs définissent clairement les concepts visés lorsqu'ils introduisent leur recherche, afin d'éviter toute ambiguïté et de permettre à des personnes de différents horizons de comprendre sans équivoque le sens attribué.

Acoustique

Domaine temporel

Concernant les mesures acoustiques, dans ce travail, nous nous sommes concentrés sur les mesures du domaine fréquentiel. À l'avenir, il serait toutefois intéressant d'**investiguer également les mesures du domaine temporel** (comme le délai de voisement ou « voice onset time ») en lien avec l'intelligibilité perçue en parole saine francophone. En effet, les mesures du domaine temporel et du domaine fréquentiel fournissent des données complémentaires [FLOEGEL 20; X. LI 08].

Moments spectraux en parole saine

De nombreuses perspectives se dégagent plus spécifiquement concernant l'exploration des moments spectraux en parole saine. L'étude des moments spectraux montre des résultats variables dans la littérature, ce qui peut s'expliquer à la fois par des différences interlinguistiques, ainsi que par des disparités méthodologiques et par les effets de coarticulation. Au vu du peu d'études disponibles pour les consonnes françaises, il convient donc de souligner le **besoin d'études supplémentaires sur de grands ensembles de données, avec des descriptions méthodologiques détaillées** (p. ex. contexte vocalique, paramètres d'enregistrement et d'analyse acoustique tels que la fréquence d'échantillonnage, les filtres appliqués et la position de la fenêtre d'analyse) pour permettre les comparaisons et la mise en commun des résultats. Entre autres, une **étude de reproductibilité** portant sur toutes les consonnes sonantes à travers des productions répétées est nécessaire afin de confirmer la fiabilité de la mesure des moments spectraux.

Rappelons aussi que des différences de moments spectraux entre parole lue et parole spontanée ont été démontrées [VAN SON 96, 99]. Au vu de l'influence du type de parole et de la coarticulation, les résultats rapportés dans cette étude ne doivent pas être généralisés à la parole spontanée. Bien que cette question ait été brièvement abordée dans notre travail, d'autres études sont nécessaires pour **explorer plus en détail l'influence des phonèmes voisins** (directs et plus éloignés) sur les moments spectraux d'une consonne cible. Ainsi, à des fins cliniques, pour l'instant, un ensemble de phrases avec un contexte vocalique contrôlé semble donc le plus pertinent pour maximiser la fiabilité des mesures. De plus, à la lumière de l'importante variabilité inter- et intralocuteur des mesures acoustiques et de l'influence des paramètres d'enregistrement et d'analyse, pour l'utilisation dans un contexte clinique, il est crucial d'utiliser les mêmes conditions d'enregistrement et les mêmes paramètres pour les mesures pré et post-thérapeutiques.

De plus, étant donné qu'un **effet de l'âge** a été observé sur les moments spectraux [TAYLOR 20], des études ultérieures avec des échantillons plus importants pour différents groupes d'âge devraient permettre d'en analyser l'évolution au fil du vieillissement naturel. Ceci pourra également permettre une meilleure compréhension des comportements articulatoires chez le locuteur sain et de leur lien avec les mesures acoustiques.

Enfin, il serait intéressant d'investiguer plus en détail le **facteur genre**. En effet, nous n'avons pas observé dans notre étude de différences significatives entre les genres pour chaque phonème. Les distinctions entre phonèmes pourraient néanmoins être caractéristiques du genre du locuteur.

Moments spectraux en parole pathologique

D'autres perspectives concernent plus spécifiquement l'application des moments spectraux à la parole pathologique.

Premièrement, dans notre étude de cas, nous avons arbitrairement choisi l'intervalle interquartile des sujets sains comme référence ; il reste encore à déterminer sur des échantillons plus larges et plus homogènes de patients (d'un point de vue perceptif) **quel degré de déviation des moments spectraux a un impact sur la perception** des phonèmes.

À ce jour, bien que les performances soient globalement satisfaisantes, **l'alignement forcé n'est pas pertinent en cas de parole trop fortement altérée**. Une étape de pré-catégorisation du degré d'altération global de la parole permettrait de décider si une analyse au niveau phonémique est envisageable. Le cas échéant, la détection des bursts des plosives par la méthode FBDS étant dépendante de la qualité de l'alignement forcé préalable, le choix de l'outil est primordial.

Pour permettre une interprétation articulatoire des déviations des moments spectraux, de futures études devront **prendre en compte un examen moteur de la sphère orofaciale** des patients, ainsi qu'idéalement des **mesures articulatoires physiologiques**.

De plus, à des fins d'utilisation clinique, il serait également intéressant de **développer des moyens de visualisation des moments spectraux** afin d'en faciliter l'interprétation. Pour cela, nous avons proposé deux pistes. La première consiste en la visualisation, à l'aide des moments spectraux, de toutes les consonnes prononcées par les sujets sains sur un graphique à trois dimensions (voir [Figure 67](#)). Cette méthode permettrait d'examiner si un phonème a correctement été prononcé ou non. Cependant, d'un point de vue clinique, l'amalgamation des

quatre moments spectraux en un point dans un espace 2D ou 3D ne permet aucune interprétation articulatoire ni piste pour le plan thérapeutique. Aussi, une deuxième piste consiste à tracer des distributions « standard » basées sur les données des sujets sains pour chaque consonne. La courbe de la production du phonème par un patient pourrait ensuite y être superposée en combinaison avec des mesures de chevauchement et de probabilités d'identité phonémique (voir [Figure 68](#)).

En outre, les **troubles vocaux et de nasalité**, fréquemment associés aux troubles de la parole dans de nombreuses pathologies, n'ont pas pu être analysés dans ce travail. Pour les confusions consonantiques basées sur le voisement, d'autres **mesures acoustiques complémentaires** aux moments spectraux doivent être utilisées. Pour l'hypernasalité, il reste encore à **analyser tout d'abord en parole saine le comportement des moments spectraux dans les consonnes nasales**, dont la structure spectrale se démarque des autres consonnes par la présence de formants et d'anti-formants. Il serait ainsi intéressant de déterminer en quelle mesure les moments spectraux s'appliquent à la distinction des phonèmes nasaux entre eux (/m, n, ŋ, ɲ/) et entre les phonèmes des paires nasales/non-nasales (p. ex. /b/ et /m/).

Enfin, les moments spectraux ont été appliqués à la parole de patients atteints d'un cancer de la sphère ORL dans ce travail. Des études supplémentaires sont nécessaires pour investiguer ces mesures dans d'autres populations de patients. Il est pour cela primordial de prendre en compte la grande hétérogénéité de la parole dans un même groupe étiologique (degré de sévérité, traitements, processus compensatoires...) et d'adopter un raisonnement physiopathologique afin de maximiser l'homogénéité des symptômes de parole des groupes d'étude.

Mesures acoustiques sur notre nouveau texte

Concernant les mesures acoustiques sur notre nouveau texte de référence, notre étude pilote a montré des résultats prometteurs pour une éventuelle utilisation en pratique clinique ou de recherche. Cependant, une **étude de validation approfondie** des mesures acoustiques sur ce texte sur des échantillons plus importants dans les trois variantes régionales est nécessaire. Cette étude pourra entre autres permettre de valider l'utilisation des moments spectraux avec l'algorithme FBDS associé à un alignement forcé.

Parole saine

Notre dernière perspective concerne une question plus théorique, concernant la définition de la parole saine. Dans nos enregistrements de phrases porteuses, nous avons par exemple choisi d'exclure les sifflements aigus sur les sibilantes (sigmatismes), amenant des valeurs aberrantes de moments spectraux. Pourtant, ces sifflements et d'autres accidents articulatoires sont inhérents à la parole saine — dont la variabilité est encore plus prononcée en lecture de texte avec des contextes phonémiques variés, et d'autant plus en langage spontané. L'exclusion de certaines productions dans l'étude de la parole saine peut donc finalement aller à l'encontre du constat de sa variabilité. Ainsi, il serait intéressant de mener une étude de consensus afin de répondre aux questions épineuses suivantes : **Qu'est-ce que la parole « saine » ?** Lors de son étude, dans quelles conditions peut-on exclure un stimulus présentant des valeurs acoustiques ou une perception « aberrantes », sans pour autant faire fi des imprécisions inhérentes à la parole naturelle ?

Bibliographie

- [ABDELLI-BERUH 04] Abdelli-Beruh, N. B. (2004). The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica*, 61(4), 201–219. <https://doi.org/10.1159/000084158>
- [ABRAHAM 20] Abraham, A. K., Pushpavathi, M., Sreedevi, N., Navya, A., Vikram, C. M. M., Prasanna, S. R. M., & Mahadeva Prasannâ, S. R. (2020). Spectral moment and duration of burst of plosives in speech of children with hearing impairment and typically developing children: A comparative study. *Interspeech*, 4981–4985. <https://doi.org/10.21437/Interspeech.2020-1805>
- [ABRAMSON 17] Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63(1), 75–86. <https://doi.org/10.1016/j.wocn.2017.05.002>
- [ABRY 03] Abry, C. (2003). [B]-[D]-[G] as a universal triangle as acoustically optimal as [I] - [a] - [U]. *15th ICPHS Barcelona*, 727–730.
- [ACHER 14] Acher, A., Perrier, P., Savariaux, C., & Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. *Clinical Linguistics and Phonetics*, 28(4), 241–256. <https://doi.org/10.3109/02699206.2013.802015>
- [ADAMS 74] Adams, M. R., & Reis, R. (1974). Influence of the onset of phonation on the frequency of stuttering: A replication and reevaluation. *Journal of Speech and Hearing Research*, 17(4), 752–754. <https://doi.org/10.1044/jshr.1704.752>
- [AICHERT 04] Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, 88(1), 148–159. [https://doi.org/10.1016/S0093-934X\(03\)00296-7](https://doi.org/10.1016/S0093-934X(03)00296-7)
- [ALEXANDER 09] Alexander, J. M., & Kluender, K. R. (2009). Spectral tilt change in stop consonant perception by listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 52(3), 653–670. [https://doi.org/10.1044/1092-4388\(2008\)08-0038](https://doi.org/10.1044/1092-4388(2008)08-0038)
- [ALI 01] Ali, A. M. A., Van Der Spiegel, J., & Mueller, P. (2001). Acoustic-phonetic features for the automatic classification of stop consonants. *IEEE Transactions on Speech and Audio Processing*, 9(8), 833–841. <https://doi.org/10.1109/89.966086>
- [ALLISON 14] Allison, K. M., & Hustad, K. C. (2014). Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *International Journal of Speech-Language Pathology*, 16(4), 396–407. <https://doi.org/10.3109/17549507.2013.876667>
- [ALLISON 19] Allison, K. M., Yunusova, Y., & Green, J. R. (2019). Shorter sentence length maximizes intelligibility and speech motor performance in persons with dysarthria due to Amyotrophic Lateral Sclerosis. *American Journal of Speech-Language Pathology*, 28(1), 96–107. https://doi.org/10.1044/2018_AJSLP-18-0049
- [ALTAHER 19] Altaher, A. M., Chu, S. Y., Kam, R. binti M., & Razak, R. A. (2019). A report of assessment tools for individuals with dysarthria. *The Open Public Health Journal*, 12(1), 384–386. <https://doi.org/10.2174/1874944501912010384>
- [ANDERSEN 74] Andersen, N. (1974). On the calculation of filter coefficients for maximum entropy spectral analysis. *Geophysics*, 39(1), 69–72. <https://doi.org/10.1190/1.1440413>
- [ANDRE-OBRECHT 88] Andre-Obrecht, R. (1988). A new statistical approach for the automatic segmentation of continuous speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(1), 29–40. <https://doi.org/10.1109/29.1486>
- [ANTONETTI 20] Antonetti, A. E. da S., Siqueira, L. T. D., Gobbo, M. P. de A., Brasolotto, A.

- G., & Silverio, K. C. A. (2020). Relationship of cepstral peak prominence-smoothed and long-term average spectrum with auditory–perceptual analysis. *Applied Sciences*, 10(23), 8598. <https://doi.org/10.3390/app10238598>
- [ASTESANO 18] 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., Pont, O., Pouchoulin, G., Puech, M., 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. *Language Resources and Evaluation Conference (LREC)*.
- [ATAL 71] Atal, B. S., & Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, 50(2B), 637–655. <https://doi.org/10.1121/1.1912679>
- [ATKINSON-CLEMENT 19] Atkinson-Clement, C., Letanneux, A., Baille, G., Cuartero, M. C., Véron-Delor, L., Robieux, C., Berthelot, M., Robert, D., Azulay, J. P., Defebvre, L., Ferreira, J., Eusebio, A., Moreau, C., & Pinto, S. (2019). Psychosocial Impact of Dysarthria: The patient-reported outcome as part of the clinical management. *Neurodegenerative Diseases, Icm*. <https://doi.org/10.1159/000499627>
- [AUDACITY TEAM 99] Audacity Team. (1999). *Audacity*. <https://www.audacityteam.org/>
- [AUSTRALIAN INSTITUTE OF HEALTH AND WELFARE 03] Australian Institute of Health and Welfare. (2003). Disability prevalence and trends. In *Disability series* (Issue 34).
- [AUZOU 98] Auzou, P. (1998). *Évaluation clinique de la dysarthrie*. Orthoédition.
- [AUZOU 06] Auzou, P., & Rolland-Monnoury, V. (2006). *BECD 2006 - Batterie d'Évaluation Clinique de la Dysarthrie*. Orthoédition.
- [AUZOU 19] Auzou, P., & Rolland-Monnoury, V. (2019). *Batterie d'Évaluation Clinique de la Dysarthrie - BECD 2019*. Orthoédition.
- [AVERY 96] Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *The Journal of the Acoustical Society of America*, 99(6), 3738–3748. <https://doi.org/10.1121/1.414970>
- [AWAN 09] Awan, S. N., & Lawson, L. L. (2009). The effect of anchor modality on the reliability of vocal severity ratings. *Journal of Voice*, 23(3), 341–352. <https://doi.org/10.1016/j.jvoice.2007.10.006>
- [BAKEN 00] Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice* (2nd ed.). Cengage Learning.
- [BALAGUER 20A] Balaguer, M., Champenois, M., Farinas, J., Piquier, J., & Woisard, V. (2020a). The (head and neck) carcinologic handicap index: Validation of a modular type questionnaire and its ability to prioritise patients' needs. *European Archives of Oto-Rhino-Laryngology*, 0123456789. <https://doi.org/10.1007/s00405-020-06201-6>
- [BALAGUER 19] Balaguer, M., Farinas, J., Fichaux-Bourin, P., Puech, M., Piquier, J., & Woisard, V. (2019). Validation of the French versions of the Speech Handicap Index and the Phonation Handicap Index in patients treated for cancer of the oral cavity or oropharynx. *Folia Phoniatria et Logopaedica*. <https://doi.org/10.1159/000503448>
- [BALAGUER 17] Balaguer, M., Percodani, J., & Woisard, V. (2017). The Carcinologic Handicap Index (CHI): A disability self-assessment questionnaire for head and neck cancer patients. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 134(6), 399–403. <https://doi.org/10.1016/j.anorl.2017.06.010>
- [BALAGUER, POMMEE 20B] Balaguer, M., Pommée, T., Farinas, J., Piquier, J., & Woisard, V. (2020b). Functional impact of speech disorders in patients treated for oral or oropharyngeal cancer, assessed by perceptual and automatic measurements. *Motor Speech Conference*. <https://doi.org/https://hal.archives-ouvertes.fr/hal-02505380>
- [BALAGUER, POMMEE 20C] Balaguer, M., Pommée, T., Farinas, J., Piquier, J., Woisard, V., &

- Speyer, R. (2020c). Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: Systematic review. *Head & Neck*, 42(1), 111–130. <https://doi.org/10.1002/hed.25949>
- [BALAGUER, POMMEE 21] Balaguer, M., Pommée, T., Farinas, J., Piquier, J., & Woisard, V. (2021). Paramètres perceptifs expliquant la sévérité du trouble de parole mesurée automatiquement en cancérologie ORL. *Rééducation Orthophonique*, 75–92.
- [BALAGUER, POMMEE n.d.] Balaguer, M., Pommée, T., & Woisard, V. (accepted). Instrumental examination of motor speech disorders. In A. am Zehnhoff-Dinnesen, A. Schindler, M.-C. Monfrais-Pfauwadel, K. Neumann, J. Sopko, & P. Zorowka (Eds.), *European manual of medicine – Phoniatics Volume 2: Disorders of speech, literacy development, acquired language and swallowing*. Springer.
- [BALLARD 15] Ballard, K. J., & McCabe, P. (2015). Developmental motor speech disorders. In *The cambridge handbook of communication disorders* (pp. 383–399). <https://doi.org/10.1017/CBO9781139108683.026>
- [BAREFOOT 93] Barefoot, S. M., Bochner, J. H., Johnson, B. A., & vom Eigen, B. A. (1993). Rating deaf speakers' comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology*, 2(3), 31–35. <https://doi.org/10.1044/1058-0360.0203.31>
- [BARRIBALL 94] Barriball, L. K., & While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing*, 19(2), 328–335.
- [BARSTIES 15] Barsties, B., & Maryn, Y. (2015). The improvement of internal consistency of the Acoustic Voice Quality Index. *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, 36(5), 647–656. <https://doi.org/10.1016/j.amjoto.2015.04.012>
- [BARSTIES 16] Barsties, B., & Maryn, Y. (2016). External validation of the Acoustic Voice Quality Index Version 03.01 with extended representativity. *Annals of Otolaryngology, Rhinology & Laryngology*, 125(7), 571–583. <https://doi.org/10.1177/0003489416636131>
- [BATTHYANY 19] Batthyany, C., Maryn, Y., Trauwaen, I., Caelenberghe, E., van Dinther, J., Zarowski, A., & Wuyts, F. (2019). A case of specificity: How does the Acoustic Voice Quality Index perform in normophonic subjects? *Applied Sciences*, 9(12), 2527. <https://doi.org/10.3390/app9122527>
- [BAYLOR 13] Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The communicative participation item bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research*, 56(4), 1190–1208. [https://doi.org/10.1044/1092-4388\(2012\)12-0140](https://doi.org/10.1044/1092-4388(2012)12-0140)
- [BECHET 12] Béchet, M., Hirsch, F., & Sock, R. (2012). F2/F3 d'occlusives sonores chez des locuteurs porteurs de fente palatine. *Actes Des Journées d'Etudes Sur La Parole*, 1(1), 561–568.
- [BEERS 91] Beers, M. H., Ouslander, J. G., Rollinger, I., Reuben, D. B., Brooks, J., & Beck, J. C. (1991). Explicit criteria for determining inappropriate medication use in nursing home residents. *Archives of Internal Medicine*, 151(9), 1825–1832. <https://doi.org/10.1001/archinte.1991.00400090107019>
- [BEHRENS 88] Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: a descriptive analysis. *Journal of Phonetics*, 16(3), 295–298. [https://doi.org/10.1016/s0095-4470\(19\)30504-2](https://doi.org/10.1016/s0095-4470(19)30504-2)
- [BEHRMAN 18] Behrman, A. (2018). *Speech and voice science* (3rd ed.). Plural Publishing.
- [BENJAMIN 97] Benjamin, B. (1997). Speech production of normally aging adults. *Seminars in Speech and Language*, 18(02), 135–141. <https://doi.org/10.1055/s-2008-1064068>
- [BENZEGUIBA 07] Benzeguiba, M., De Mori, R., Deroo, O., Dupon, S., Erbes, T., Jouvét, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V., & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, 49(10–11), 763–786. <https://doi.org/10.1016/j.specom.2007.02.006>

- [BILODEAU-MERCURE 16] Bilodeau-Mercure, M., & Tremblay, P. (2016). Age differences in sequential speech production: Articulatory and physiological factors. *Journal of the American Geriatrics Society*, 64(11), e177–e182. <https://doi.org/10.1111/jgs.14491>
- [BIRKO 15] Birko, S., Dove, E. S., Özdemir, V., & Dalal, K. (2015). Evaluation of nine consensus indices in delphi foresight research and their dependency on delphi survey characteristics: A simulation study and debate on delphi design and interpretation. *PLoS ONE*, 10(8), 1–14. <https://doi.org/10.1371/journal.pone.0135162>
- [BLACKLOCK 04] Blacklock, O. (2004). *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods*. University of Southampton.
- [BLANC 14] Blanc, E., & Giusti, L. (2014). *Évaluation de l'intelligibilité de la parole dans les dysarthries: adaptation en français de la version révisée du Frenchay Dysarthria Assessment [FDA-2]*. Marseille.
- [BLUMSTEIN 79] Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66(4), 1001–1017. <https://doi.org/10.1121/1.383319>
- [BOERING 61] Boering, G., Hoeksema, P. E., Huffstadt, A. J. C., Moolenaar-Bijl, A. J., Ritsema Van Eck, C. R., & Visser, H. K. A. (1961). *Over de behandeling van patienten met cheilognathopalatoschisis*. Academisch Ziekenhuis.
- [BOERSMA 01] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9–10), 341–345.
- [BONATTI 05] Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16(6), 451–459. <https://doi.org/10.1111/j.0956-7976.2005.01556.x>
- [BOND 94] Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, 14(4), 325–337. [https://doi.org/10.1016/0167-6393\(94\)90026-4](https://doi.org/10.1016/0167-6393(94)90026-4)
- [BONNEAU 08] Bonneau, A., & Laprie, Y. (2008). Selective acoustic cues for French voiceless stop consonants. *The Journal of the Acoustical Society of America*, 123(6), 4482–4497. <https://doi.org/10.1121/1.2916693>
- [BRADLOW 96] 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)
- [BRANCAZIO 98] Brancazio, L., & Fowler, C. A. (1998). On the relevance of locus equations for production and perception of stop consonants. *Perception and Psychophysics*, 60(1), 24–50. <https://doi.org/10.3758/BF03211916>
- [BRINCA 15] Brinca, L., Batista, A. P., Tavares, A. I., Pinto, P. N., & Araújo, L. (2015). The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *Journal of Voice*, 29(6), 776.e7–776.e14. <https://doi.org/10.1016/j.jvoice.2015.01.007>
- [BROWN 45] Brown, S. F. (1945). The loci of stutterings in the speech sequence. *Journal of Speech Disorders*, 10(3), 181–192. <https://doi.org/10.1044/jshd.1003.181>
- [BUDER 96] Buder, E. H., Kent, R. D., Kent, J. F., Milenkovic, P., & Workinger, M. (1996). FORMOFFA: An automated formant, moment, fundamental frequency, amplitude analysis of normal and disordered speech. *Clinical Linguistics and Phonetics*, 10(1), 31–54. <https://doi.org/10.3109/02699209608985160>
- [BUNTON 01] Bunton, K., & Weismer, G. (2001). The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *Journal of*

- Speech, Language, and Hearing Research*, 44(6), 1215–1228.
[https://doi.org/10.1044/1092-4388\(2001/095\)](https://doi.org/10.1044/1092-4388(2001/095))
- [BÜRKI 08] Bürki, A., Gendrot, C., Gravier, G., Linares, G., & Fougeron, C. (2008). Aligement automatique et analyse phonétique : comparaison de différents systèmes pour l’analyse du schwa. *Revue TAL, ATALA (Association Pour Le Traitement Automatique Des Langues)*, 49(3), 165–197.
- [BURNS 11] Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery*, 128(1), 305–310. <https://doi.org/10.1097/PRS.0b013e318219c171>
- [BUTCHER 89] Butcher, A. (1989). Measuring coarticulation and variability in tongue contact patterns. *Clinical Linguistics and Phonetics*, 3(1), 39–47. <https://doi.org/10.3109/02699208908985269>
- [BUTLER 19] Butler, C. (2019). Working the ‘wise’ in speech and language therapy: Evidence-based practice, biopolitics and ‘pastoral labour.’ *Social Science and Medicine*, 230(June 2018), 1–8. <https://doi.org/10.1016/j.socscimed.2019.03.038>
- [CAMARGO 19] Camargo, Z., Madureira, S., Reis, N., Rilliard, A., Camargo, Z., Madureira, S., Reis, N., Rilliard, A., Camargo, Z., Madureira, S., Reis, N., & Rilliard, A. (2019). The phonetic approach of voice qualities : challenges in corresponding perceptual to acoustic descriptions. In *Subsidia. Tools and resources for speech sciences*.
- [CARAMAZZA 00] Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, 403(6768), 428–430. <https://doi.org/10.1038/35000206>
- [CARMICHAEL 07] Carmichael, J. (2007). *Introducing objective acoustic metrics for the Frenchay Dysarthria Assessment procedure*. University of Sheffield. <https://doi.org/10.13140/RG.2.1.4037.8960>
- [CARREIRAS 08a] Carreiras, M., Gillon-Dowens, M., Vergara, M., & Perea, M. (2008a). Are vowels and consonants processed differently? event-related potential evidence with a delayed letter paradigm. *Journal of Cognitive Neuroscience*, 21(2), 275–288. <https://doi.org/10.1162/jocn.2008.21023>
- [CARREIRAS 08b] Carreiras, M., & Price, C. J. (2008b). Brain activation for consonants and vowels. *Cerebral Cortex*, 18(7), 1727–1735. <https://doi.org/10.1093/cercor/bhm202>
- [CATTELAÏN 19] Cattelain, T. (2019). *Production des consonnes plosives du français : du contrôle des bruits de plosion*. Université Grenoble Alpes.
- [CHALMERS 19] Chalmers, J., & Armour, M. (2019). The delphi technique. *Handbook of Research Methods in Health Social Sciences*, 715–735. https://doi.org/10.1007/978-981-10-5251-4_99
- [CHANG 17] Chang, S.-A., Park, M.-H., Oh, S.-H., & Lee, K.-M. (2017). Influence of upper frequency boundary on consonant perception in normal hearing and cochlear implant adults. *Audiology and Speech Research*, 13(3), 231–239. <https://doi.org/10.21848/asr.2017.13.3.231>
- [CHEN 11] Chen, F., & Loizou, P. C. (2011). Predicting the intelligibility of vocoded and wideband Mandarin Chinese. *The Journal of the Acoustical Society of America*, 129(5), 3281–3290. <https://doi.org/10.1121/1.3570957>
- [CHEN 13] Chen, F., Wong, L. L. N., & Wong, E. Y. W. (2013). Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility. *The Journal of the Acoustical Society of America*, 134(2), EL178–EL184. <https://doi.org/10.1121/1.4812820>
- [CHIAL 03] Chial, M. R. (2003). Suggestions for computer-based audio recording of speech samples for perceptual and acoustic analyses. *Phonology Project Technical Report No. 13*.
- [CHIN 90] Chin, J., Sato, P. A., & Mann, J. M. (1990). Projections of HIV infections and AIDS

- cases to the year 2000. *Bulletin of the World Health Organization*, 68(1), 1–11.
- [CHO 19] Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72, 52–65. <https://doi.org/10.1016/j.wocn.2018.11.002>
- [COLE 96] Cole, R., Yan, Y., Mak, B., & Fanty, M. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *IEE Proceedings - Vision, Image, and Signal Processing*, 2, 853–856. <https://doi.org/10.1121/1.417028>
- [COLEMAN 71] Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, 14(3), 565–577. <http://www.ncbi.nlm.nih.gov/pubmed/5163891>
- [COLLIS 12] Collis, J., & Bloch, S. (2012). Survey of UK speech and language therapists' assessment and treatment practices for people with progressive dysarthria. *International Journal of Language and Communication Disorders*, 47(6), 725–737. <https://doi.org/10.1111/j.1460-6984.2012.00183.x>
- [COMBESURE 81] Combescure, P. (1981). 20 listes de dix phrases phonétiquement équilibrées. *Revue d'acoustique*, 14(56), 34–38.
- [CONNAGHAN 17] Connaghan, K. P., & Patel, R. (2017). The impact of contrastive stress on vowel acoustics and intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research*, 60(1), 38–50. https://doi.org/10.1044/2016_JSLHR-S-15-0291
- [CONWAY 15] Conway, A., & Walshe, M. (2015). Management of non-progressive dysarthria: Practice patterns of speech and language therapists in the Republic of Ireland. *International Journal of Language and Communication Disorders*, 50(3), 374–388. <https://doi.org/10.1111/1460-6984.12143>
- [N. J. COX 09] Cox, N. J. (2009). Speaking Stata: Creating and varying box plots. *The Stata Journal*, 9(3), 478–496. <https://doi.org/https://doi.org/10.1177/1536867X0900900309>
- [R. M. COX 87] Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Intelligibility of average talkers in typical listening environments. *Journal of the Acoustical Society of America*, 81(5), 1598–1608. <https://doi.org/10.1121/1.394512>
- [CROCHEMORE 01] Crochemore, E., & Vannier, F. (2001). Analyse phonétique de la parole dysarthrique. In P. Auzou, C. Özsancak, & V. Brun (Eds.), *Les dysarthries* (pp. 71–82). Masson.
- [CUNNINGHAM 19] Cunningham, B. J., Kwok, E., Turkstra, L., & Oram Cardy, J. (2019). Establishing consensus among community clinicians on how to categorize and define preschoolers' speech and language impairments at assessment. *Journal of Communication Disorders*, 82(November-December), 105925. <https://doi.org/10.1016/j.jcomdis.2019.105925>
- [CUTLER 02] Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language*, 46(2), 296–322. <https://doi.org/10.1006/jmla.2001.2814>
- [CUTLER 00] Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection : Cross-linguistic comparisons. *Memory & Cognition*, 28(5), 746–755.
- [D'ALESSANDRO 18] D'Alessandro, D., & Fougeron, C. (2018). Réduction de la coarticulation et vieillissement. *Journées d'Étude Sur La Parole*, 410–418. <https://hal.archives-ouvertes.fr/hal-02436286>
- [D'ALESSANDRO 19] D'Alessandro, D., Pernon, M., Fougeron, C., Laganaro, M., Alessandro, D., Pernon, M., Fougeron, C., Laganaro, M., & Vtov, A. (2019). Anticipatory VtoV coarticulation in French in several motor speech disorders. *Phonetics and Phonology in Europe*. <https://hal.archives-ouvertes.fr/hal-02427864>
- [DALKEY 63] Dalkey, N., & Helmer, O. (1963). An experimental application of Deplhi method

- to use of experts. In *Management science* (Vol. 9, pp. 458–467).
- [DARLEY 69a] Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12(2), 246–269. <https://doi.org/10.1044/jshr.1202.246>
- [DARLEY 69b] Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, 12(3), 462–496. <https://doi.org/10.1044/jshr.1203.462>
- [DARLEY 75] Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders* (3rd editio). W.B. Saunders Company.
- [DAUDET 69] Daudet, A. (1869). La chèvre de monsieur Seguin. In J. Hetzel (Ed.), *Lettres de mon moulin* (pp. 37–50).
- [DE BODT 06] De Bodt, M. S., Guns, C., & Van Nuffelen, G. (2006). *NSVO: Nederlandstalig Spraakverstaanbaarheidsonderzoek*. Vlaamse Vereniging voor Logopedisten.
- [DE BRUIJN 09] De Bruijn, M. J., Ten Bosch, L., Kuik, D. J., Quené, H., Langendijk, J. A., Leemans, C. R., & Verdonck-De Leeuw, I. M. (2009). Objective acoustic-phonetic speech analysis in patients treated for oral or oropharyngeal cancer. *Folia Phoniatica et Logopaedica*, 61(3), 180–187. <https://doi.org/10.1159/000219953>
- [DE MAUPASSANT 82] de Maupassant, G. (1882). Pierrot. *Le Gaulois*.
- [DE MIRANDA 19] de Miranda, V. S. G., Marcolino, M. A. Z., Rech, R. S., Barbosa, L. de R., & Fischer, G. B. (2019). Evidence-based speech therapy: The role of systematic revisions. *Codas*, 31(2), 1–2. <https://doi.org/10.1590/2317-1782/20182018167>
- [DEGROOTE 11] Degroote, G., Simon, J., Borel, S., & Crevier-Buchman, L. (2011). The French version of Speech Handicap Index: Validation and comparison with the Voice Handicap Index. *Folia Phoniatica et Logopaedica*, 64(1), 20–25. <https://doi.org/10.1159/000328982>
- [DELATTRE 55] Delattre, P., Cooper, F. S., Liberman, A. M., & Gerstman, L. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4), 769–773. <https://doi.org/10.1121/1.1917767>
- [DELGADO-HERNANDEZ 18] Delgado-Hernández, J., León-Gómez, N. M., Izquierdo-Arteaga, L. M., & Llanos-Fumero, Y. (2018). Análisis cepstral de la voz normal y patológica en adultos españoles. Medida de la prominencia del pico cepstral suavizado en vocales sostenidas versus habla conectada. *Acta Otorrinolaringológica Española*, 69(3), 134–140. <https://doi.org/10.1016/j.otorri.2017.05.006>
- [DELIYSKI 05] Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, 19(1), 15–28. <https://doi.org/10.1016/j.jvoice.2004.07.003>
- [DENMAN 19] Denman, D., Kim, J. H., Munro, N., Speyer, R., & Cordier, R. (2019). Describing language assessments for school-aged children: A Delphi study. *International Journal of Speech-Language Pathology*, 21(6), 602–612. <https://doi.org/10.1080/17549507.2018.1552716>
- [DERDEMEZIS 16] Derdemezis, E., Vorperian, H. K., Kent, R. D., Fourakis, M., Reinicke, E. L., & Bolt, D. M. (2016). Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. *American Journal of Speech-Language Pathology*, 25(3), 335–354. https://doi.org/10.1044/2015_AJSLP-15-0020
- [DETERDING 06] Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, 36(2), 187–196. <https://doi.org/10.1017/S0025100306002544>
- [DI CRISTO 12] Di Cristo, A. (2012). Le pouvoir de la prosodie ou la revanche de Cendrillon. In L. Baqué & M. Estrada (Eds.), *L'homme communicant* (pp. 95–114). CIPA.
- [DIAMOND 14] Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C.,

- Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401–409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>
- [DING 03] Ding, Z. Q., McLoughlin, I. V., & Tan, E. C. (2003). Extension of proposal of standards for intelligibility tests of Chinese speech: CDRT-tone. *IEE Proceedings - Vision, Image, and Signal Processing*, 150(1), 1. <https://doi.org/10.1049/ip-vis:20030161>
- [DIRECTION DE LA RECHERCHE, DES ÉTUDES, & DE L'ÉVALUATION ET DES STATISTIQUES 19] Direction de la Recherche, des Études, & de l'Évaluation et des Statistiques. (2019). *Les orthophonistes : données statistiques*. http://www.data.drees.sante.gouv.fr/ReportFolders/reportFolders.aspx?IF_ActivePath=P_490,497,970,1026
- [DIRECTION GENERALE SOINS DE SANTE 18] Direction générale Soins de Santé. (2018). *Statistiques annuelles des professionnels de soins de santé en Belgique*. https://organesdeconcertation.sante.belgique.be/sites/default/files/documents/statan_2018_fr_2.pdf
- [DISSEN 19] Dissen, Y., Goldberger, J., & Keshet, J. (2019). Formant estimation and tracking: A deep learning approach. *The Journal of the Acoustical Society of America*, 145(2), 642–653. <https://doi.org/10.1121/1.5088048>
- [DITTNER 10] Dittner, J., Lepage, B., Woisard, V., Kergadallan, M., Boisteux, K., Robart, E., & Welby-Gieusse, M. (2010). Élaboration et validation d'un test quantitatif d'intelligibilité pour les troubles pathologiques de la production de la parole. *Revue de Laryngologie Otologie Rhinologie*, 131(1), 9–14.
- [DODD 07] Dodd, B. (2007). Evidence-based practice and speech-language pathology: Strengths, weaknesses, opportunities and threats. *Folia Phoniatrica et Logopaedica*, 59(3), 118–129. <https://doi.org/10.1159/000101770>
- [DOLLAGHAN 07] Dollaghan, C. (2007). *The Handbook of evidence-based practice in communication disorders*. Paul H. Brooks Publishing Co.
- [DORSEY 07] Dorsey, E. R., Constantinescu, R., Thompson, J. P., Biglan, K. M., Holloway, R. G., Kiebertz, K., Marshall, F. J., Ravina, B. M., Schifitto, G., Siderowf, A., & Tanner, C. M. (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5), 384–386. <https://doi.org/10.1212/01.wnl.0000247740.47667.03>
- [DROMEY 03] Dromey, C. (2003). Spectral measures and perceptual ratings of hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology*, 11(2), 85–94. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0038362549&partnerID=tZOtx3y1>
- [DUFFY 16] Duffy, J. R. (2016). Functional speech disorders: clinical manifestations, diagnosis, and management. In *Handbook of clinical neurology* (1st ed., Vol. 139). Elsevier B.V. <https://doi.org/10.1016/B978-0-12-801772-2.00033-3>
- [DURIEUX 18] Durieux, N., Maillart, C., Donneau, A. F., & Pasleau, F. (2018). Controlled before-after study to evaluate change in evidence-based practice of speech and language therapy students. *Health Information and Libraries Journal*, 35(3), 213–226. <https://doi.org/10.1111/hir.12224>
- [DWIVEDI 16] Dwivedi, R. C., St.Rose, S., Chisholm, E. J., Clarke, P. M., Kerawala, C. J., Nutting, C. M., Rhys-Evans, P. H., Kazi, R., & Harrington, K. J. (2016). Acoustic parameters of speech: Lack of correlation with perceptual and questionnaire-based speech evaluation in patients with oral and oropharyngeal cancer treated with primary surgery. *Head & Neck*, 38(5), 670–676. <https://doi.org/10.1002/hed.23956>
- [DYKSTRA 07] Dykstra, A. D., Hakel, M. E., & Adams, S. G. (2007). Application of the ICF in reduced speech intelligibility in dysarthria. *Seminars in Speech and Language*, 28(4), 301–

311. <https://doi.org/10.1055/s-2007-986527>
- [ENDERBY 08] Enderby, P., & Palmer, R. (2008). *FDA-2: Frenchay Dysarthria Assessment - Second Edition*. Pro-Ed.
- [ERINGIS 14] Eringis, D., & Tamulevičius, G. (2014). Improving speech recognition rate through analysis parameters. *Electrical, Control and Communication Engineering*, 5(1), 61–66. <https://doi.org/10.2478/ecce-2014-0009>
- [EVERS 98] Evers, V., Reetz, H., & Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics*, 26(4), 345–370. <https://doi.org/10.1006/jpho.1998.0079>
- [EVIDENCE-BASED MEDICINE WORKING GROUP 92] Evidence-Based Medicine Working Group. (1992). Evidence-Based Medicine : A new approach to teaching the practice of medicine. *JAMA*, 268(17), 2420–2425. <https://doi.org/10.1001/jama.1992.03490170092032>
- [EVITTS 16] Evitts, P. M., Starmer, H., Teets, K., Montgomery, C., Calhoun, L., Schulze, A., MacKenzie, J., & Adams, L. (2016). The impact of dysphonic voices on healthy listeners: Listener reaction times, speech intelligibility, and listener comprehension. *American Journal of Speech-Language Pathology*, 25(4), 561–575. https://doi.org/10.1044/2016_AJSLP-14-0183
- [FAIRBANKS 60] Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd editio). Harper & Row.
- [FANT 60] Fant, G. (1960). *Acoustic theory of speech production*. Mouton.
- [FANT 69] Fant, G. (1969). Quarterly progress and status report: Stops in CV-syllables. *STL-QPSR*, 10(4), 1–25.
- [FANT 73] Fant, G. (1973). *Speech sounds and features*. MIT Press.
- [FASSER 92] Fasser, C. E., Smith, Q. W., & Luchi, R. J. (1992). Geriatrics fellows' perceptions of the quality of their research training. *Academic Medicine*, 67(10), 696–698. <https://doi.org/10.1097/00001888-199210000-00017>
- [FELT TIP INC 08] Felt Tip Inc. (2008). *Sound Studio*. <https://feltpip.com/ss/>
- [FERGUSON 07] Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, 50(5), 1241–1255. [https://doi.org/10.1044/1092-4388\(2007/087\)](https://doi.org/10.1044/1092-4388(2007/087))
- [FERGUSON 14] Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 135(6), 3570–3584. <https://doi.org/10.1121/1.4874596>
- [FICHAUX-BOURIN 09] Fichaux-Bourin, P., Woisard, V., Grand, S., Puech, M., & Bodin, S. (2009). Validation of a self assessment for speech disorders (Phonation Handicap Index). *Revue de Laryngologie Otologie Rhinologie*, 130(1), 45–51.
- [FINEOUT-OVERHOLT 05] Fineout-Overholt, E., & Johnston, L. (2005). Teaching EBP: Asking searchable, answerable clinical questions. *Worldviews on Evidence-Based Nursing*, 2(3), 157–160. <https://doi.org/10.1111/j.1741-6787.2005.00032.x>
- [FITCH 00] Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Science*, 4(7), 258–267. [https://doi.org/https://doi.org/10.1016/S1364-6613\(00\)01494-7](https://doi.org/https://doi.org/10.1016/S1364-6613(00)01494-7)
- [FLANAGAN 72] Flanagan, J. L. (1972). *Speech analysis, synthesis and perception* (2nd ed.). Springer.
- [FLEGE 92] Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, 92(1), 128–143. <https://doi.org/10.1121/1.404278>
- [FLEMMING 09] Flemming, E. (2009). The phonetics of schwa vowels. In D. Minkova (Ed.), *Phonological weakness in English* (pp. 78–95). Palgrave Macmillan UK.

- https://doi.org/10.1007/978-0-230-29686-2_5
- [FLESCH 48] Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- [A.R. FLETCHER 17] Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2017). Assessing vowel centralization in dysarthria: A comparison of methods. *Journal of Speech, Language, and Hearing Research*, 60(2), 341–354. https://doi.org/10.1044/2016_JSLHR-S-15-0355
- [S.G. FLETCHER 72] Fletcher, S. G. (1972). Contingencies for bioelectronic modification of nasality. *The Journal of Speech and Hearing Disorders*, 37(3), 329–346. <https://doi.org/10.1044/jshd.3703.329>
- [FLIPSEN 99] Flipsen, P. J., Shriberg, L., Weismer, G., Karlsson, H., & McSweeney, J. (1999). Acoustic characteristics of /s/ in adolescents. *Journal of Speech, Language, and Hearing Research*, 42, 663–677.
- [FLOEGEL 20] Floegel, M., Fuchs, S., & Kell, C. A. (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nature Communications*, 11(1), 1–12. <https://doi.org/10.1038/s41467-020-16743-2>
- [FOGERTY 10] Fogerty, D., & Humes, L. E. (2010). Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors. *The Journal of the Acoustical Society of America*, 128(5), 3114–3125. <https://doi.org/10.1121/1.3493439>
- [FOGERTY 12] Fogerty, D., Kewley-Port, D., & Humes, L. E. (2012). The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss. *The Journal of the Acoustical Society of America*, 132(3), 1667–1678. <https://doi.org/10.1121/1.4739463>
- [FORREST 88] Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84(1), 115–123. <https://doi.org/10.1121/1.396977>
- [FOSTER 13] Foster, A. M., Worrall, L. E., Rose, M. L., & O'Halloran, R. (2013). Turning the tide: Putting acute aphasia management back on the agenda through evidence-based practice. *Aphasiology*, 27(4), 420–443. <https://doi.org/10.1080/02687038.2013.770818>
- [FOUGERON 19] Fougeron, C., Delvaux, V., Ménard, L., & Laganaro, M. (2019). The MonPaGe_HA database for the documentation of spoken French throughout adulthood. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 4301–4306.
- [FOURNIER 51] Fournier, J.-E. (1951). *Audiométrie vocale : les épreuves d'intelligibilité et leurs applications au diagnostic, à l'expertise et à la correction prothétique des surdités*. Maloine.
- [FRANCIS 00] Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception and Psychophysics*, 62(8), 1668–1680. <https://doi.org/10.3758/BF03212164>
- [FREDOUILLE 19] Fredouille, C., Ghio, A., Laaridh, I., Lalain, M., Woisard, V., Fredouille, C., Ghio, A., Laaridh, I., Lalain, M., & Woisard, V. (2019). Acoustic-phonetic decoding for speech intelligibility evaluation in the context of head and neck cancers. *International Congress of Phonetic Sciences*, 3051–3055.
- [GALLUZZI 15] Galluzzi, C., Bureca, I., Guariglia, C., & Romani, C. (2015). Phonological simplifications, apraxia of speech and the interaction between phonological and phonetic processing. *Neuropsychologia*, 71, 64–83. <https://doi.org/10.1016/j.neuropsychologia.2015.03.007>
- [GENDROT 05] Gendrot, C., & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *9th European Conference on Speech Communication and Technology*, 1,

- 2453–2456.
- [GENTIL 92] Gentil, M. (1992). Phonetic intelligibility testing in dysarthria for the use of french language clinicians. *Clinical Linguistics and Phonetics*, 6(3), 179–189. <https://doi.org/10.3109/02699209208985529>
- [GHIO 20] 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 16] Ghio, A., Giusti, L., Blanc, E., Pinto, S., Lalain, M., Robert, D., Fredouille, C., & Woisard, V. (2016). Quels tests d’intelligibilité pour évaluer les troubles de production de la parole? *Journées d’Étude Sur La Parole*, 4(1), 589–596. <https://hal.archives-ouvertes.fr/hal-01372037>
- [GHIO 18] 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. *Journées d’Étude Sur La Parole*, 285–293. <https://doi.org/10.21437/jep.2018-33>
- [GHIO 17] Ghio, A., Lalain, M., Giusti, L., Robert, D., Rebourg, M., André, C., Fredouille, C., Ghio, A., Lalain, M., Giusti, L., Robert, D., & Pouchoulin, G. (2017). Du décodage acoustico-phonétique pour mesurer l’intelligibilité de locuteurs atteints de troubles de production de la parole. *Journées de Phonétique Clinique*. <https://hal.archives-ouvertes.fr/hal-01615016>
- [GIBSON 19] Gibson, M., & Gil, J. (2019). *Romance phonetics and phonology*. Oxford University Press.
- [GOLDMAN 11] Goldman, J. (2011). EasyAlign : an automatic phonetic alignment tool under Praat. *Interspeech*. <http://archive-ouverte.unige.ch/unige:18188>
- [GORDON 02] Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2), 141–174. <https://doi.org/10.1017/S0025100302001020>
- [GOTSMANN 11] Gottsmann, F., & Harwardt, C. (2011). Investigating robustness of spectral moments on normal-and high-effort speech. *Interspeech, August*, 2937–2940.
- [GUENTHER 95] Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594–621. <https://doi.org/10.1037/0033-295X.102.3.594>
- [GUREVICH 17] Gurevich, N., & Scamihorn, S. L. (2017). Speech-language pathologists’ use of intelligibility measures in adults with dysarthria. *American Journal of Speech-Language Pathology*, 26(3), 873–892. https://doi.org/10.1044/2017_AJSLP-16-0112
- [HALEY 18] Haley, K. L., Cunningham, K. T., Eaton, C. T., & Jacks, A. (2018). Error consistency in acquired apraxia of speech with aphasia: Effects of the analysis unit. *Journal of Speech, Language, and Hearing Research*, 61(2), 210–226. https://doi.org/10.1044/2017_JSLHR-S-16-0381
- [HALLE 57] Halle, M., Hughes, G. W., & Radley, J. -P. A. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, 29(1), 107–116. <https://doi.org/10.1121/1.1908634>
- [HARDCASTLE 09] Hardcastle, B., & Tjaden, K. (2009). Coarticulation and speech impairment. *The Handbook of Clinical Linguistics*, 506–524. <https://doi.org/10.1002/9781444301007.ch32>
- [HARMEGNIES 88] Harmegnies, B. (1988). *Contribution à la caractérisation de la qualité vocale. Analyses plurielles de spectres moyens à long terme de parole*. Université de Mons-Hainaut.
- [HARTMAN 95] Hartman, F. T., & Baldwin, A. (1995). Using technology to improve Delphi

- method. *Journal of Computing in Civil Engineering*, 9(4), 244–249. [https://doi.org/10.1061/\(asce\)0887-3801\(1995\)9:4\(244\)](https://doi.org/10.1061/(asce)0887-3801(1995)9:4(244))
- [HAZAN 17] Hazan, V. (2017). Speech communication across the life span. *Acoustics Today*, 13(1), 36–43. <https://acousticstoday.org/wp-content/uploads/2017/01/Speech-Communication-Across-the-Life-Span-Valerie-Hazan.pdf>
- [HAZAN 04] Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*, 116(5), 3108–3118. <https://doi.org/10.1121/1.1806826>
- [HAZAN 18] Hazan, V., Tuomainen, O., Tu, L., Kim, J., Davis, C., Brungart, D., & Sheffield, B. (2018). How do aging and age-related hearing loss affect the ability to communicate effectively in challenging communicative conditions? *Hearing Research*, 369, 33–41. <https://doi.org/10.1016/j.heares.2018.06.009>
- [HEBA 21] Heba, A. (2021). *Reconnaissance automatique de la parole à large vocabulaire : des approches hybrides aux approches End-to-End*. Université Toulouse III Paul Sabatier. <https://hal.archives-ouvertes.fr/tel-03269807>
- [HEMAN-ACKAH 03] Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., Horman, M., & Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology, Rhinology & Laryngology*, 112(4), 324–333. <https://doi.org/10.1177/000348940311200406>
- [HERNANDEZ 19] Hernandez, A., Lee, H., & Chung, M. (2019). Acoustic analysis of fricatives in dysarthric speakers with cerebral palsy. *Phonetics and Speech Sciences*, 11(3), 23–29. <https://doi.org/10.13064/ksss.2019.11.3.023>
- [HILLENBRAND 94] Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778. <https://doi.org/10.1044/jshr.3704.769>
- [HILLENBRAND 95] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099–3111. <https://doi.org/10.1121/1.411872>
- [HILLENBRAND 96] Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2), 311–321. <https://doi.org/10.1044/jshr.3902.311>
- [HILLENBRAND 03] Hillenbrand, J. (2003). American English: Southern Michigan. *Journal of the International Phonetic Association*, 33(1), 121–126. <https://doi.org/10.1017/S0025100303001221>
- [HILLENBRAND 01] Hillenbrand, J., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, 109(2), 748–763. <https://doi.org/10.1121/1.1337959>
- [HINKLE 03] Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.).
- [HODGE 10] Hodge, M., & Whitehill, T. (2010). Intelligibility impairments. *The Handbook of Language and Speech Disorders*, 99–114. <https://doi.org/10.1002/9781444318975.ch4>
- [HOGIKYAN 99] Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V- RQOL). *Journal of Voice*, 13(4), 557–569. [https://doi.org/10.1016/S0892-1997\(99\)80010-1](https://doi.org/10.1016/S0892-1997(99)80010-1)
- [HOHOFF 03] Hohoff, A., Seifert, E., Fillion, D., Stamm, T., Heinecke, A., & Ehmer, U. (2003). Speech performance in lingual orthodontic patients measured by sonagraphy and auditive analysis. *American Journal of Orthodontics and Dentofacial Orthopedics*, 123(2), 146–152. <https://doi.org/10.1067/mod.2003.12>
- [HONDA 08] Honda, K. (2008). Physiological processes of speech production. In J. Benesty, M. M. Sondhi, & Y. A. Huang (Eds.), *Springer handbook of speech processing* (pp. 7–26).

- Springer. https://doi.org/https://doi.org/10.1007/978-3-540-49127-9_2
- [HONDA 19] Honda, K. (2019). The speech mechanism. In M. A. Redford (Ed.), *The handbook of speech production* (p. 60). John Wiley & Sons, Inc.
- [HOOD 80] Hood, J. D., & Poole, J. P. (1980). Influence of the speaker and other factors affecting speech intelligibility. *International Journal of Audiology*, 19(5), 434–455. <https://doi.org/10.3109/00206098009070077>
- [HOOPER 09] Hooper, C. R., & Cralidis, A. (2009). Normal changes in the speech of older adults: You’ve still got what it takes; it just takes a little longer! *Perspectives on Gerontology*, 14(2), 47. <https://doi.org/10.1044/gero14.2.47>
- [HSIEH 05] Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- [HSU 07] Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10). <https://doi.org/10.7275/pdz9-th90>
- [HUINCK 04] Huinck, W. J., Van Lieshout, P. H. H. M., Peters, H. F. M., & Hulstijn, W. (2004). Gestural overlap in consonant clusters: Effects on the fluent speech of stuttering and non-stuttering subjects. *Journal of Fluency Disorders*, 29(1), 3–25. <https://doi.org/10.1016/j.jfludis.2003.09.001>
- [HUSTAD 06] Hustad, K. C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech-Language Pathology*, 15(3), 268–277. [https://doi.org/10.1044/1058-0360\(2006/025\)](https://doi.org/10.1044/1058-0360(2006/025))
- [HUSTAD 08] Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 51(3), 562–573. [https://doi.org/10.1044/1092-4388\(2008/040\)](https://doi.org/10.1044/1092-4388(2008/040))
- [IEEE 69] IEEE. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3), 225–246.
- [IMBERT 10] Imbert, G. (2010). L’entretien semi-directif : à la frontière de la santé publique et de l’anthropologie. *Recherche En Soins Infirmiers*, 3(102), 23–34.
- [INTERNATIONAL PHONETIC ASSOCIATION 99] International Phonetic Association. (1999). *Handbook of the International Phonetic Association*. Cambridge University Press.
- [ISKAROUS 11] Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129(2), 944–954. <https://doi.org/10.1121/1.3514537>
- [JACOBSON 97] 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>
- [JAKIELSKI 98] Jakielski, K. J. (1998). *Motor organization in the acquisition of consonant clusters*. University of Texas at Austin. <https://www.proquest.com/openview/e3c556421c6b8f8dbeebf23f42fe1fdd/1?pq-origsite=gscholar&cbl=18750&diss=y>
- [JAKIELSKI 00] Jakielski, K. J. (2000). Quantifying phonetic complexity in words: An experimental index. *Annual Child Phonology Conference*.
- [JAKOBSON 51] Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis: The distinctive features and their correlates*. The MIT Press.
- [JAN 07] Jan, M. (2007). L’évaluation instrumentale de la dysarthrie en France. In P. Auzou, V. Rolland-Monnoury, S. Pinto, & C. Ozsancak (Eds.), *Les dysarthries* (pp. 259–269). Solal.

- [JANNEDY 17] Jannedy, S., & Weirich, M. (2017). Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. *The Journal of the Acoustical Society of America*, 142(1), 395–405. <https://doi.org/10.1121/1.4991347>
- [JESUS 02] Jesus, L. M. T., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30(3), 437–464. <https://doi.org/10.1006/jpho.2002.0169>
- [JESUS 15] Jesus, L. M. T., Valente, A. R. S., & Hall, A. (2015). Is the Portuguese version of the passage “The North Wind and the Sun” phonetically balanced? *Journal of the International Phonetic Association*, 45(1), 1–11. <https://doi.org/10.1017/S0025100314000255>
- [JOHNSON 93] Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2), 701–714. <https://doi.org/10.1121/1.406887>
- [JONES 95] Jones, J., & Hunter, D. (1995). Qualitative research: Consensus methods for medical and health services research. *Bmj*, 311(7001), 376. <https://doi.org/10.1136/bmj.311.7001.376>
- [JONGMAN 00] Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252. <https://doi.org/10.1121/1.1288413>
- [KAHN 11] Kahn, J. (2011). *Parole de locuteur : performance et confiance en identification biométrique vocale*. Université d’Avignon. <https://tel.archives-ouvertes.fr/tel-00995071/>
- [KATZ 00] Katz, W. F. (2000). Anticipatory coarticulation and aphasia: Implications for phonetic theories. *Journal of Phonetics*, 28(3), 313–334. <https://doi.org/10.1006/jpho.2000.0118>
- [KATZ 91] Katz, W. F., Kripke, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of Speech, Language, and Hearing Research*, 34(6), 1222–1232. <https://doi.org/10.1044/jshr.3406.1222>
- [KAY 12] Kay, T. S. (2012). *Spectral analysis of stop consonants in individuals with dysarthria secondary to stroke*. Louisiana State University.
- [KENNINGTON 15] Kennington, C., Iida, R., Tokunaga, T., & Schlangen, D. (2015). Incrementally tracking reference in human/human dialogue using linguistic and extra-linguistic information. *NAACL HLT, June*, 272–282. <https://doi.org/10.3115/v1/n15-1031>
- [KENT 83] Kent, R. D. (1983). The segmental organization of speech. In *The production of speech* (pp. 57–89). Springer New York. https://doi.org/10.1007/978-1-4613-8202-7_4
- [KENT 92] Kent, R. D. (1992). *Intelligibility in speech disorders*. John Benjamins Publishing Company. <https://doi.org/10.1075/sspcl.1>
- [KENT 00] Kent, R. D. (2000). Research on speech motor control and its disorders. *Journal of Communication Disorders*, 33(5), 391–428. [https://doi.org/10.1016/s0021-9924\(00\)00023-x](https://doi.org/10.1016/s0021-9924(00)00023-x)
- [KENT 89a] Kent, R. D., Kent, J. F., Weismer, G., Martin, R. E., Sufit, R. L., Brooks, B. R., & Rosenbek, J. C. (1989a). Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects. *Clinical Linguistics & Phonetics*, 3(4), 347–358. <https://doi.org/10.3109/02699208908985295>
- [KENT 09] Kent, R. D., & Kim, Y. (2009). Acoustic analysis of speech. *The Handbook of Clinical Linguistics*, 360–380. <https://doi.org/10.1002/9781444301007.ch22>
- [KENT 89b] Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989b). 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 99] 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(3), 141–186. [https://doi.org/10.1016/S0021-9924\(99\)00004-0](https://doi.org/10.1016/S0021-9924(99)00004-0)
- [KEWLEY-PORT 07] 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>
- [KHAN 03] Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *JRSM*, 96(3), 118–121. <https://doi.org/10.1258/jrsm.96.3.118>
- [H. KIM 10] Kim, H., Martin, K., Hasegawa-Johnson, M., & Perlman, A. (2010). Frequency of consonant articulation errors in dysarthric speech. *Clinical Linguistics and Phonetics*, 24(10), 759–770. <https://doi.org/10.3109/02699206.2010.497238>
- [Y. KIM 17] Kim, Y., & Choi, Y. (2017). A cross-language study of acoustic predictors of speech intelligibility in individuals with Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 60(9), 2506–2518. https://doi.org/10.1044/2017_jslhrs-16-0121
- [Y. KIM 11] Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research*, 54(2), 417–429. [https://doi.org/10.1044/1092-4388\(2010/10-0020\)](https://doi.org/10.1044/1092-4388(2010/10-0020))
- [Y. KIM 09] Kim, Y., Weismer, G., Kent, R. D., & Duffy, J. R. (2009). Statistical models of F2 slope in relation to severity of dysarthria. *Folia Phoniatrica et Logopaedica*, 61(6), 329–335. <https://doi.org/10.1159/000252849>
- [KLANGPORNKUN 13] Klangpornkun, N., Onsuwan, C., Tantibundhit, C., & Pitathawatchai, P. (2013). Predictions from "speech banana" and audiograms: Assessment of hearing deficits in Thai hearing loss patients. *The Journal of the Acoustical Society of America*, 134(5), 4132–4132. <https://doi.org/10.1121/1.4831179>
- [KMET 04] Kmet, L. M., Lee, R. C., & Cook, L. S. (2004). Standard quality assessment criteria for evaluating primary research papers from a variety of fields. *HTA Initiative*, 13(February), 1–22.
- [KOENIG 13] Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, 56(4), 1175–1189. [https://doi.org/10.1044/1092-4388\(2012/12-0038\)](https://doi.org/10.1044/1092-4388(2012/12-0038))
- [KONG 12] Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40(6), 725–744. <https://doi.org/10.1016/j.wocn.2012.07.002>
- [KÖRKKÖ 15] Körkkö, P. (2015). Spectral moments analysis of /s/ coarticulation development in Finnish-speaking children. *ICPhS*. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0470.pdf>
- [KUO 16] Kuo, C., & Weismer, G. (2016). Vowel reduction across tasks for male speakers of American English. *The Journal of the Acoustical Society of America*, 140(1), 369–383. <https://doi.org/10.1121/1.4955310>
- [KURUVILLA-DUGDALE 18] Kuruvilla-Dugdale, M., Custer, C., Heidrick, L., Barohn, R., & Govindarajan, R. (2018). A phonetic complexity-based approach for intelligibility and articulatory precision testing: A preliminary study on talkers with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research: JSLHR*, 61(9), 2205–

2214. https://doi.org/10.1044/2018_JSLHR-S-17-0462
- [KURUVILLA-DUGDALE 20] Kuruvilla-Dugdale, M., Dietrich, M., McKinley, J. D., & Deroche, C. (2020). An exploratory model of speech intelligibility for healthy aging based on phonatory and articulatory measures. *Journal of Communication Disorders*, 87(March), 105995. <https://doi.org/10.1016/j.jcomdis.2020.105995>
- [LAFON 64] Lafon, J.-C. (1964). *Le test phonétique et la mesure de l'audition*. Éditions Centrex.
- [LAGANARO 21] Laganaro, M., Fougeron, C., Pernon, M., Levêque, N., Borel, S., Fournet, M., Catalano Chiuvé, S., Lopez, U., Trouville, R., Ménard, L., Burkhard, P. R., Assal, F., & Delvaux, V. (2021). Sensitivity and specificity of an acoustic- and perceptual-based tool for assessing motor speech disorders in French: the MonPaGe-screening protocol. *Clinical Linguistics & Phonetics*, 1–16. <https://doi.org/10.1080/02699206.2020.1865460>
- [LALAIN 20] 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
- [LAM 16] Lam, J., & Tjaden, K. (2016). Clear speech variants: An acoustic study in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 59(4), 631–646. https://doi.org/10.1044/2015_JSLHR-S-15-0216
- [LAM 12] Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, 55(6), 1807–1821. [https://doi.org/10.1044/1092-4388\(2012/11-0154\)](https://doi.org/10.1044/1092-4388(2012/11-0154))
- [LANDIS 77] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <http://www.ncbi.nlm.nih.gov/pubmed/843571>
- [LANSFORD 14] Lansford, K. L., & Liss, J. M. (2014). Vowel acoustics in dysarthria: Mapping to perception. *Journal of Speech, Language, and Hearing Research*, 57(1), 68–80. [https://doi.org/10.1044/1092-4388\(2013/12-0263\)](https://doi.org/10.1044/1092-4388(2013/12-0263))
- [LAVOIE 02] Lavoie, L. M. (2002). Subphonemic and suballophonic consonant variation: The role of the phoneme inventory. *ZAS Papers in Linguistics*, 28, 39–54.
- [H. LEE 14] Lee, H., Gambette, P., Barkat-Defradas, M. (2014). iPhocomp : calcul automatique de l'indice de complexité phonétique de Jakielski. *Journées d'Étude Sur La Parole*, 622–630.
- [J. LEE 14] Lee, J., Hustad, K. C., & Weismer, G. (2014). Predicting speech intelligibility with a multiple speech subsystems approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research*, 57(5), 1666–1678. https://doi.org/10.1044/2014_JSLHR-S-13-0292
- [LETRILLIART 11] Letrilliart, L., & Vanmeerbeek, M. (2011). À la recherche du consensus : Quelle méthode utiliser ? *Exercer*, 99(99), 170–177.
- [LEVELT 93] Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1–15). Walther de Gruyter.
- [LEVEQUE 16] Lévêque, N., Laganaro, M., Fougeron, C., Delvaux, V., Pernon, M., Borel, S., & Catalano, S. (2016). MonPaGe : un protocole informatisé d'évaluation de la parole pathologique en langue française. *Revue Neurologique*, 172(August 2017), A162–A163. <https://doi.org/10.1016/j.neurol.2016.01.386>
- [F. LI 10] Li, F., Menon, A., & Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *The Journal of the Acoustical Society of America*, 127(4), 2599–2610. <https://doi.org/10.1121/1.3295689>
- [X. LI 08] Li, X., Ning, Z., Brashears, R., & Rife, K. (2008). Relative contributions of spectral and temporal cues for speech recognition in patients with sensorineural hearing loss.

- Journal of Otology*, 3(2), 84–91. [https://doi.org/10.1016/S1672-2930\(08\)50019-5](https://doi.org/10.1016/S1672-2930(08)50019-5)
- [LIBERATI 09] Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*, 62(10), e1-34. <https://doi.org/10.1016/j.jclinepi.2009.06.006>
- [LIBERMAN 67] Liberman, 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>
- [LIBERMAN 58] Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1(3), 153–167. <https://doi.org/10.1177/002383095800100301>
- [LIBERMAN 52] Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*, 65(4), 497. <https://doi.org/10.2307/1418032>
- [LIKERT 32] Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1–55.
- [LINDBLUM 90] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16
- [LINSTONE 02a] Linstone, H. A., & Turoff, M. (2002a). Computers and the future of Delphi: Introduction. In H. A. Linstone & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 483–489).
- [LINSTONE 02b] Linstone, H. A., & Turoff, M. (2002b). *The Delphi method: Techniques and applications* (M. Turoff & H. A. Linstone (eds.)). New Jersey Institute of Technology.
- [LIU 05] Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, 117(6), 3879–3889. <https://doi.org/10.1121/1.1898623>
- [LIU 00] Liu, H.-M., Tseng, C.-H., & Tsao, F.-M. (2000). Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clinical Linguistics & Phonetics*, 14(6), 447–464. <https://doi.org/10.1080/026992000415877>
- [LOUSADA 12] Lousada, M. L., Jesus, L. M. T., & Pape, D. (2012). Estimaco das caractersticas espectrais relacionadas com o ponto de articulao de oclusivas utilizando a tcnica multitaper. *Documentao e Estudos Em Lingustica Terica e Aplicada*, 28(1), 1–26. <https://doi.org/10.1590/delta.v28i1.4240>
- [LOWELL 11] Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2011). Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25(5), e223–e232. <https://doi.org/10.1016/j.jvoice.2010.06.007>
- [LUDLOW 19] Ludlow, C. L., Kent, R. D., & Gray, L. C. (2019). Acoustic analysis of speech. In C. L. Ludlow, R. D. Kent, & L. C. Gray (Eds.), *Measuring voice, speech, and swallowing in the clinic and laboratory* (pp. 161–218). Plural Publishing.
- [MA 01] Ma, E. P.-M., & Yiu, E. M.-L. (2001). Voice Activity and Participation Profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language, and Hearing Research*, 44(3), 511–524. <https://doi.org/10.1092-4388/01/4403-0511>
- [MACKENZIE 20] Mackenzie, L., & Turton, D. (2020). Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard*, 6(s1), 1–14. <https://doi.org/10.1515/lingvan-2018-0061>
- [MANIWA 09] Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly

- spoken English fricatives. *The Journal of the Acoustical Society of America*, 125(6), 3962–3973. <https://doi.org/10.1121/1.2990715>
- [MANN 80] Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics*, 28(3), 213–228. <https://doi.org/10.3758/BF03204377>
- [MARKS 02] Marks, E. A., Moates, D., Bond, Z., & Vasquez, L. (2002). Vowel mutability: The case of monolingual Spanish listeners and bilingual Spanish-English listeners. *Southwest Journal of Linguistics*, 21(2), 73.
- [MARTEL-SAUVAGEAU 20] Martel-Sauvageau, V., Breton, M., & Chabot, A. (2020). The impact of clear speech on the spectral properties and intelligibility of fricatives-vowel sequences in speakers with Parkinson's disease. *Motor Speech Conference*.
- [MARTEL-SAUVAGEAU 17] Martel-Sauvageau, V., & Tjaden, K. (2017). Vocalic transitions as markers of speech acoustic changes with STN-DBS in Parkinson's disease. *Journal of Communication Disorders*, 70(12), 1–11. <https://doi.org/10.1016/j.jcomdis.2017.10.001>
- [MARTENS 10a] Martens, H., Van Nuffelen, G., & De Bodt, M. (2010a). De ontwikkeling van de fonetisch uitgebalanceerde standaardtekst. *Logopedie*, 23(5), 31–36.
- [MARTENS 10b] Martens, H., Van Nuffelen, G., & De Bodt, M. (2010b). *Nederlandstalig Spraakverstaanbaarheidsonderzoek - Zinsniveau (NSVO-Z) [Dutch Sentence Intelligibility Assessment (DSIA)]*. Vlaamse Vereniging voor Logopedisten.
- [MARYN 09] Maryn, Y., Corthals, P., De Bodt, M., Van Cauwenberge, P., & Deliyski, D. (2009). Perturbation measures of voice: A comparative study between Multi-Dimensional Voice Program and Praat. *Folia Phoniatrica et Logopaedica*, 61(4), 217–226. <https://doi.org/10.1159/000227999>
- [MAUCLAIR 13] Mauclair, J., Koenig, L., Robert, M., & Gatignol, P. (2013). Burst-based features for the classification of pathological voices. *Interspeech*, 2167–2171.
- [MAUROVA PAILLEREAU 16] Maurová Paillereau, N. (2016). Do isolated vowels represent vowel targets in French? An acoustic study on coarticulation. *SHS Web of Conferences*, 27, 09003. <https://doi.org/10.1051/shsconf/20162709003>
- [MAYER 11] Mayer, J. (2011). *Spectral moments berechnen*. <http://praatpfanne.lingphon.net/praat-ressourcen/praat-skripte>
- [McCLOY 12] McCloy, D., Wright, R., & Souza, P. (2012). Modeling intrinsic intelligibility variation: Vowel-space size and structure. *Proceedings of Meetings on Acoustics*, 18(1). <https://doi.org/10.1121/1.4870070>
- [McCLOY 15] McCloy, D., Wright, R., & Souza, P. (2015). Talker versus dialect effects on speech intelligibility: A symmetrical study. *Language and Speech*, 58(3), 371–386. <https://doi.org/10.1177/0023830914559234>
- [MCCURTIN 12] McCurtin, A., & Roddam, H. (2012). Evidence-based practice: SLTs under siege or opportunity for growth? the use and nature of research evidence in the profession. *International Journal of Language and Communication Disorders*, 47(1), 11–26. <https://doi.org/10.1111/j.1460-6984.2011.00074.x>
- [MCLEOD 12] McLeod, S., Harrisson, L. J., & McCormack, J. (2012). *Intelligibility in Context Scale*. Charles Sturt University.
- [MCMILLAN 14] McMillan, S. S., Kelly, F., Sav, A., Kendall, E., King, M. A., Whitty, J. A., & Wheeler, A. J. (2014). Using the nominal group technique: how to analyse across multiple groups. *Health Services and Outcomes Research Methodology*, 14(3), 92–108. <https://doi.org/10.1007/s10742-014-0121-1>
- [MCMILLAN 16] McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*, 38(3), 655–662. <https://doi.org/10.1007/s11096-016-0257-x>
- [MCPHERSON 18] McPherson, S., Reese, C., & Wendler, M. C. (2018). Methodology update:

- Delphi studies. *Nursing Research*, 67(5), 404–410. <https://doi.org/10.1097/NNR.0000000000000297>
- [MENIN-SICARD 01] Menin-Sicard, A., & Sicard, E. (2001). *Vocalab*. GERIP.
- [MENIN-SICARD 06] Menin-Sicard, A., & Sicard, E. (2006). *Diadolab*. GERIP.
- [METZ 90] Metz, D. E., Schiavetti, N., Samar, V. J., & Sitler, R. W. (1990). Acoustic dimensions of hearing-impaired speakers' intelligibility: Segmental and suprasegmental characteristics. *Journal of Speech and Hearing Research*, 33(3), 476–487. <https://doi.org/10.1044/jshr.3303.476>
- [MEUNIER 07] Meunier, C. (2007). Phonétique acoustique. In P. Auzou, V. Rolland-Monnoury, S. Pinto, & C. Ozsancak (Eds.), *Les dysarthries* (pp. 164–173). Solal.
- [MEYER 11] Meyer, S., & Ptok, M. (2011). Zur objektiven Beurteilung der Zischlautbildung. *Laryngo- Rhino- Otologie*, 90(12), 753–759. <https://doi.org/10.1055/s-0031-1275279>
- [MEYNADIER 13] Meynadier, Y. (2013). Éléments de phonétique acoustique. In N. Nguyen & M. Adda-Decker (Eds.), *Méthodes et outils pour l'analyse phonétique des grands corpus oraux* (pp. 25–83). Hermès.
- [MILLER 13] Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language and Communication Disorders*, 48(6), 601–612. <https://doi.org/10.1111/1460-6984.12061>
- [MILLER 17] Miller, N., & Bloch, S. (2017). A survey of speech–language therapy provision for people with post-stroke dysarthria in the UK. *International Journal of Language and Communication Disorders*, 52(6), 800–815. <https://doi.org/10.1111/1460-6984.12316>
- [MILLER 11] Miller, N., Deane, K. H. O., Jones, D., Noble, E., & Gibb, C. (2011). National survey of speech and language therapy provision for people with Parkinson's disease in the United Kingdom: Therapists' practices. *International Journal of Language and Communication Disorders*, 46(2), 189–201. <https://doi.org/10.3109/13682822.2010.484849>
- [MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE 20] Ministère de l'Enseignement supérieur et de la Recherche. (2020). *Etudes & métiers : Professions de santé*. https://cedies.public.lu/dam-assets/pictures/pdf-publications/etudes_metiers/sante-2020.pdf
- [MOHER 09] Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- [MOOSHAMMER 07] Mooshammer, C., Hoole, P., & Geumann, A. (2007). Jaw and order. *Language and Speech*, 50(2), 145–176. <https://doi.org/10.1177/00238309070500020101>
- [MUNSON 04] Munson, B. (2004). Children and adults : Evidence spectral mean. *Journal of Speech, Language, and Hearing Research*, 47(February), 58–69.
- [MURTON 20] Murton, O., Hillman, R., & Mehta, D. (2020). Cepstral peak prominence values for clinical voice evaluation. *American Journal of Speech-Language Pathology*, 29(3), 1596–1607. https://doi.org/10.1044/2020_AJSLP-20-00001
- [NEEL 08] Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, 51(3), 574–585. [https://doi.org/10.1044/1092-4388\(2008/041\)](https://doi.org/10.1044/1092-4388(2008/041))
- [NEEL 15] Neel, A. T., Palmer, P. M., Sprouls, G., & Morrison, L. (2015). Muscle weakness and speech in oculopharyngeal muscular dystrophy. *Journal of Speech, Language, and Hearing Research*, 58(1), 1–12. https://doi.org/10.1044/2014_JSLHR-S-13-0172
- [NESPOR 03] Nespors, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, 2(2), 203–229. <https://doi.org/10.1418/10879>
- [NEWMAN 03] Newman, R. S. (2003). Using links between speech perception and speech

- production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, 113(5), 2850–2860. <https://doi.org/10.1121/1.1567280>
- [NGUYEN 01] Nguyen, N. (2001). Rôle de la coarticulation dans la reconnaissance des mots. *Année Psychologique*, 101(1), 125. <https://doi.org/10.3406/psy.2001.29719>
- [NHMRC 99] NHMRC. (1999). A guide to the development, implementation and evaluation of clinical practice guidelines. In *National Health and Medical Research Council (NHMRC)*. https://doi.org/10.1007/978-3-642-16373-9_2
- [NIRGIANAKI 14] Nirgianaki, E. (2014). Acoustic characteristics of Greek fricatives. *The Journal of the Acoustical Society of America*, 135(5), 2964–2976. <https://doi.org/10.1121/1.4870487>
- [NISSEN 03] Nissen, S. (2003). An acoustic analysis of voiceless obstruents produced by adults and typically developing children. Ohio State University. https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1041225568&disposition=inline
- [NISSEN 05] Nissen, S., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118(4), 2570–2578. <https://doi.org/10.1121/1.2010407>
- [NORTHERN 84] Northern, J. L., & Downs, M. P. (1984). *Hearing in Children*. Lippincott Williams & Wilkins.
- [OBSERVATIURE SUISSE DE LA SANTE 16] Observatoire suisse de la santé. (2016). *Personnel de santé en Suisse: Etat des lieux et projections à l'horizon 2030*. https://www.obsan.admin.ch/sites/default/files/publications/2017/obsan_71_rapport_corr.pdf
- [ORGANISATION MONDIALE DE LA SANTE 01] Organisation mondiale de la santé. (2001). *Classification internationale du fonctionnement, du handicap et de la santé: CIF*.
- [OWREN 06] Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *The Journal of the Acoustical Society of America*, 119(3), 1727–1739. <https://doi.org/10.1121/1.2161431>
- [PAGE 21] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- [PALMER 05] Palmer, R. (2005). The management of dysarthria by practising speech and language therapists. In *An evaluation of speech and language therapy for chronic dysarthria*. University of Sheffield.
- [PARNELL 87] Parnell, M. M., & Amerman, J. D. (1987). Perception of oral diadochokinetic performances in elderly adults. *Journal of Communication Disorders*, 20(4), 339–351. [https://doi.org/10.1016/0021-9924\(87\)90015-3](https://doi.org/10.1016/0021-9924(87)90015-3)
- [R. PATEL 13] Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. (2013). “The Caterpillar”: A novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Pathology*, 22(1), 1–9. [https://doi.org/10.1044/1058-0360\(2012/11-0134\)](https://doi.org/10.1044/1058-0360(2012/11-0134))
- [R. R. PATEL 18] Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyiski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, 27(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009
- [PECKELS 73] Peckels, J. P., & Rossi, M. (1973). Le test de diagnostic par paires minimales.

- Revue d'acoustique*, 27, 245–262.
- [PEETERS 04] Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO 1st Project Report*, 54, 1–25. <http://www.citeulike.org/group/1854/article/1562527>
- [PEPIOT 15] Pépiot, E. (2015). Voice, speech and gender. *Corela*, HS-16, 0–13. <https://doi.org/10.4000/corela.3783>
- [PERNON 20] Pernon, M., Lévêque, N., Delvaux, V., Assal, F., Borel, S., Fougeron, C., Trouville, R., & Laganaro, M. (2020). MonPaGe, un outil de screening francophone informatisé d'évaluation perceptive et acoustique des troubles moteurs de la parole (dysarthries, apraxie de la parole. *Rééducation Orthophonique*, 281, 169–198.
- [PERRY 80] Perry, S., & Kalberer, J. T. (1980). The NIH consensus-development program and the assessment of health-care technologies. *New England Journal of Medicine*, 303(3), 169–172. <https://doi.org/10.1056/NEJM198007173030334>
- [PETERSON 52] Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- [PHADKE 20] Phadke, K. V., Laukkanen, A.-M., Ilomäki, I., Kankare, E., Geneid, A., & Švec, J. G. (2020). Cepstral and perceptual investigations in female teachers with functionally healthy voice. *Journal of Voice*, 34(3), 485.e33-485.e43. <https://doi.org/10.1016/j.jvoice.2018.09.010>
- [PIACENTINI 14] Piacentini, V., Mauri, I., Cattaneo, D., Gilardone, M., Montesano, A., & Schindler, A. (2014). Relationship between quality of life and dysarthria in patients with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation*, 95(11), 2047–2054. <https://doi.org/10.1016/j.apmr.2014.04.023>
- [PLICHTA 02] Plichta, B. (2002). Best practices in the acquisition, processing, and analysis of acoustic speech signals. *U. Penn Working Papers in Linguistics*, 8(3), 209–222. <http://flint.matrix.msu.edu/extras/Audio-technology.pdf%0Ahttp://repository.upenn.edu/pwpl/vol8/iss3/16%0Ahttp://repository.upenn.edu/pwpl/vol8/>
- [POMMÉE 20] Pommée, T., & Balaguer, M. (2020). Intérêts cliniques des mesures de la parole pathologique adulte. *Séminaire SDORMP FORM Café Ortho*.
- [POMMÉE 21a] Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021a). Assessment of adult speech disorders: current situation and needs in French-speaking clinical practice. *Logopedics Phoniatrics Vocology*, 0(0), 1–15. <https://doi.org/10.1080/14015439.2020.1870245>
- [POMMÉE 21b] Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021b). Intelligibility and comprehensibility: A Delphi consensus study. *International Journal of Language and Communication Disorders*. <https://doi.org/https://doi.org/10.1111/1460-6984.12672>
- [POMMÉE 21c] Pommée, T., Balaguer, M., Pinquier, J., Mauclair, J., Woisard, V., & Speyer, R. (2021c). Relationship between phoneme-level spectral acoustics and speech intelligibility in healthy speech: a systematic review. *Speech, Language and Hearing*, 24(2), 105–132. <https://doi.org/10.1080/2050571X.2021.1913300>
- [POMMÉE n.d.] Pommée, T., Balaguer, M., & Woisard, V. (accepted). Clinical examination of motor speech disorders. In A. am Zehnhoff-Dinnesen, A. Schindler, M.-C. Monfrais-Pfauwadel, K. Neumann, J. Sopko, & P. Zorowka (Eds.), *European manual of medicine – Phoniatrics Volume 2: Disorders of speech, literacy development, acquired language and swallowing*. Springer.
- [POMMÉE 19a] Pommée, T., Mauclair, J., Woisard, V., Farinas, J., & Pinquier, J. (2019a). Génération de la «banane de la parole» en vue d'une évaluation objective de

- l'intelligibilité. *Journées de Phonétique Clinique*. <https://doi.org/https://oatao.univ-toulouse.fr/24885/>
- [POMMEE 21d] Pommée, T., Pinquier, J., Mauclair, J., & Woisard, V. (2021d). Exploration des moments spectraux comme mesure acoustique sur les fricatives et plosives en Français. *Séminaire AFCP*.
- [POMMEE 19b] Pommée, T., Woisard, V., Mauclair, J., & Pinquier, J. (2019b). Pertinence clinique des mesures d'intelligibilité de la parole. *Ecole d'été En Logopédie-Orthophonie*. <https://doi.org/https://hal.archives-ouvertes.fr/hal-03003824>
- [PORZEL 11] Porzel, R. (2011). *Contextual computing*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-17396-7>
- [POUCHOULIN 07] Pouchoulin, G., Fredouille, C., Ghio, A., Azzarello, M., Giovanni, A., Pouchoulin, G., Fredouille, C., Ghio, A., & Azzarello, M. (2007). Modélisation statistique et informations pertinentes pour la caractérisation des voix pathologiques (dysphonies). *Journées d'Étude Sur La Parole*. https://dumas.ccsd.cnrs.fr/SANTE_PUB_INSERM/hal-00136742v1
- [POWELL 03] Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, 41(4), 376–382. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>
- [POWELL 06] Powell, T. (2006). A comparison of English reading passages for elicitation of speech samples from clinical populations. *Clinical Linguistics and Phonetics*, 20(2–3), 91–97. <https://doi.org/10.1080/02699200400026488>
- [REIDY 15] Reidy, P. F. (2015). A comparison of spectral estimation methods for the analysis of sibilant fricatives. *The Journal of the Acoustical Society of America*, 137(4), EL248–EL254. <https://doi.org/10.1121/1.4915064>
- [J. REILLY 12] Reilly, J., & Fisher, J. L. (2012). Sherlock Holmes and the strange case of the missing attribution: A historical note on “The Grandfather Passage.” *Journal of Speech, Language, and Hearing Research*, 55(1), 84–88. [https://doi.org/10.1044/1092-4388\(2011/11-0158\)](https://doi.org/10.1044/1092-4388(2011/11-0158))
- [K. J. REILLY 13] Reilly, K. J., & Spencer, K. A. (2013). Sequence complexity effects on speech production in healthy speakers and speakers with hypokinetic or ataxic dysarthria. *PLoS ONE*, 8(10), 1–14. <https://doi.org/10.1371/journal.pone.0077450>
- [REPP 89] Repp, B. H., & Lin, H. B. (1989). Acoustic properties and perception of stop consonant release transients. *Journal of the Acoustical Society of America*, 85(1), 379–396. <https://doi.org/10.1121/1.397689>
- [REVIS 04] Révis, J. (2004). *L'analyse perceptive des dysphonies : approche phonétique de l'évaluation vocale*. Université Aix-Marseille II. <http://www.theses.fr/2004AIX20675>
- [RINKEL 08] Rinkel, R. N., Leeuw, I. M. V., van Reij, E. J., Aaronson, N. K., & Leemans, C. R. (2008). Speech Handicap Index in patients with oral and pharyngeal cancer: Better understanding of patients' complaints. *Head & Neck*, 30(7), 868–874. <https://doi.org/10.1002/hed.20795>
- [ROSENBEK 84] Rosenbek, J. C., Kent, R. D., & Lapointe, L. L. (1984). Apraxia of speech: An overview and some perspectives. In J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds.), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 1–72). College-Hill Press.
- [ROSSI 98] Rossi, M., & Peter-Defare, É. (1998). *Les lapsus ou Comment notre fourche a langué*. Presses Universitaires de France.
- [ROULSTONE 15] Roulstone, S. (2015). Exploring the relationship between client perspectives, clinical expertise and research evidence. *International Journal of Speech-Language Pathology*, 17(3), 211–221. <https://doi.org/10.3109/17549507.2015.1016112>
- [RUMBACH 19] Rumbach, A. F., Finch, E., & Stevenson, G. (2019). What are the usual assessment practices in adult non-progressive dysarthria rehabilitation? A survey of

- Australian dysarthria practice patterns. *Journal of Communication Disorders*, 79(March), 46–57. <https://doi.org/10.1016/j.jcomdis.2019.03.002>
- [RYCROFT-MALONE 04] Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2004). What counts as evidence in evidence-based practice? *Journal of Advanced Nursing*, 47(1), 81–90. <https://doi.org/10.1111/j.1365-2648.2004.03068.x>
- [SACKETT 96] Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ*, 312(7023), 71–72. <https://doi.org/10.1136/bmj.312.7023.71>
- [SANDOVAL 13] Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, 134(5), EL477–EL483. <https://doi.org/10.1121/1.4826150>
- [SAPIR 09] Sapir, S., Połczyńska, M., & Tobin, Y. (2009). Why does the vowel space area as an acoustic metric fail to differentiate dysarthric from normal vowel articulation and what can be done about it? *Poznan Studies in Contemporary Linguistics*, 45(2), 301–311. <https://doi.org/10.2478/v10010-009-0018-2>
- [SAPIR 10] Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research*, 53(1), 114–125. [https://doi.org/10.1044/1092-4388\(2009/08-0184\)](https://doi.org/10.1044/1092-4388(2009/08-0184))
- [SAPIR 11] Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2011). Acoustic metrics of vowel articulation in Parkinson's disease: Vowel space area (VSA) vs. Vowel articulation index (VAI). *Models and Analysis of Vocal Emissions for Biomedical Applications - 7th International Workshop, MAVEBA 2011*, 9, 173–175.
- [SASISEKARAN 14] Sasisekaran, J. (2014). Exploring the link between stuttering and phonology: A review and implications for treatment. *Seminars in Speech and Language*, 35(2), 95–113. <https://doi.org/10.1055/s-0034-1371754>
- [SAUDER 17] Sauder, C., Bretl, M., & Eadie, T. (2017). Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice*, 31(5), 557–566. <https://doi.org/10.1016/j.jvoice.2017.01.006>
- [SCARBEL 12] Scarbel, L. (2012). *Etude acoustique de la production de consonnes et de voyelles du français par des enfants porteurs d'un implant cochléaire*. Université Grenoble Alpes.
- [SCHARDT 07] Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, 7, 1–6. <https://doi.org/10.1186/1472-6947-7-16>
- [SCHIAVETTI 92] Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement and management* (p. 11). John Benjamins Publishing Company. <https://doi.org/10.1075/sspcl.1.02sch>
- [SCHILLER 06] Schiller, N. O. (2006). Phonological encoding in speech production. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, 53–60.
- [SCHLOSSER 06] Schlosser, R. W., & O'Neil-Pirozzi, T. M. (2006). Problem formulation in evidence-based reviews. *Contemporary Issues in Communication Science and Disorders*, 33, 5–10.
- [SCHOW 18] Schow, R. L., & Nerbonne, M. A. (2018). *Introduction to audiologic rehabilitation* (7th editio). Pearson.
- [SCHRAMM 54] Schramm, W. L. (1954). How communication works. In W. L. Schramm (Ed.), *The process and effects of mass communication* (pp. 3–26). University of Illinois Press.

- [SCHULLER 13] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S. (2013). Paralinguistics in speech and language - State-of-the-art and the challenge. *Computer Speech and Language*, 27(1), 4–39. <https://doi.org/10.1016/j.csl.2012.02.005>
- [SCHULLER 19] Schuller, B., Weninger, F., Zhang, Y., Ringeval, F., Batliner, A., Steidl, S., Eyben, F., Marchi, E., Vinciarelli, A., Scherer, K., Chetouani, M., & Mortillaro, M. (2019). Affective and behavioural computing: Lessons learnt from the First Computational Paralinguistics Challenge. *Computer Speech and Language*, 53, 156–180. <https://doi.org/10.1016/j.csl.2018.02.004>
- [SHANNON 02] Shannon, R. V. (2002). The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *American Journal of Audiology*, 11(2), 124–127. [https://doi.org/10.1044/1059-0889\(2002\)013](https://doi.org/10.1044/1059-0889(2002)013)
- [SHUEY 89] Shuey, E. M. (1989). Intelligibility of older versus younger adults' CVC productions. *Journal of Communication Disorders*, 22(6), 437–444. [https://doi.org/10.1016/0021-9924\(89\)90036-1](https://doi.org/10.1016/0021-9924(89)90036-1)
- [SINHA 11] Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*, 8(1). <https://doi.org/10.1371/journal.pmed.1000393>
- [SKINNER 78] Skinner, M. W. (1978). The hearing of speech during language acquisition. *Otolaryngologic Clinics of North America*, 11(3), 631–650.
- [SKODDA 13] Skodda, S., Grönheit, W., Mancinelli, N., & Schlegel, U. (2013). Progression of voice and speech impairment in the course of Parkinson's disease: A longitudinal study. *Parkinson's Disease*, 2013, 1–8. <https://doi.org/10.1155/2013/389195>
- [SKODDA 12] Skodda, S., Grönheit, W., & Schlegel, U. (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS ONE*, 7(2). <https://doi.org/10.1371/journal.pone.0032132>
- [SMILJANIC 09] Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, 3(1), 236–264. <https://doi.org/10.1111/j.1749-818X.2008.00112.x>
- [SMITH 14] Smith, L. K., & Goberman, A. M. (2014). Long-time average spectrum in individuals with Parkinson disease. *NeuroRehabilitation*, 35(1), 77–88. <https://doi.org/10.3233/NRE-141102>
- [SRAMKOVA 15] Šrámková, H., Granqvist, S., Herbst, C. T., & Švec, J. G. (2015). The softest sound levels of the human voice in normal subjects. *The Journal of the Acoustical Society of America*, 137(1), 407–418. <https://doi.org/10.1121/1.4904538>
- [STAIGER 12] Staiger, A., Finger-Berg, W., Aichert, I., & Ziegler, W. (2012). Error variability in apraxia of speech: A matter of controversy. *Journal of Speech, Language, and Hearing Research*, 55(5). [https://doi.org/10.1044/1092-4388\(2012\)11-0319](https://doi.org/10.1044/1092-4388(2012)11-0319)
- [STEVENS 98] Stevens, K. N. (1998). *Acoustic Phonetics*. MIT Press.
- [STEVENS 78] Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5), 1358–1368. <https://doi.org/10.1121/1.382102>
- [STEVENS 63] Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research*, 6(2), 111–128. <https://doi.org/10.1044/jshr.0602.111>
- [STEWART 01] Stewart, M. (2001). Towards a global definition of patient centred care: The patient should be the judge of patient centred care. *British Medical Journal*, 322(7284), 444–445. <https://doi.org/10.1136/bmj.322.7284.444>
- [STIPANCIC 16] Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility

- measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research*, 59(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271
- [STORY 17] Story, B. H., & Bunton, K. (2017). Vowel space density as an indicator of speech performance. *The Journal of the Acoustical Society of America*, 141(5), EL458–EL464. <https://doi.org/10.1121/1.4983342>
- [SUNDARA 05] Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America*, 118(2), 1026–1037. <https://doi.org/10.1121/1.1953270>
- [SUOMI 85] Suomi, K. (1985). The vowel-dependence of gross spectral cues to place of articulation of stop consonants in CV syllables. *Journal of Phonetics*, 13(3), 267–285. [https://doi.org/10.1016/s0095-4470\(19\)30759-4](https://doi.org/10.1016/s0095-4470(19)30759-4)
- [SUSSMAN 12] Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research*, 55(4), 1208–1219. [https://doi.org/10.1044/1092-4388\(2011/11-0048\)](https://doi.org/10.1044/1092-4388(2011/11-0048))
- [SVEC 10] Švec, J. G., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, 19(4), 356–368. [https://doi.org/10.1044/1058-0360\(2010/09-0091\)](https://doi.org/10.1044/1058-0360(2010/09-0091))
- [SVEC 18] Švec, J. G., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research*, 61(3), 441–461. https://doi.org/10.1044/2017_JSLHR-S-17-0095
- [SWANSON 10] Swanson, J. A., Schmitz, D., & Chung, K. C. (2010). How to practice Evidence-Based Medicine. *Plastic and Reconstructive Surgery*, 126(1), 286–294. <https://doi.org/10.1097/PRS.0b013e3181dc54ee>
- [TAYLOR 20] Taylor, S., Dromey, C., Nissen, S., Tanner, K., Eggett, D., & Corbin-Lewis, K. (2020). Age-related changes in speech and voice: Spectral and cepstral measures. *Journal of Speech, Language, and Hearing Research*, 63(3), 647–660. https://doi.org/10.1044/2019_JSLHR-19-00028
- [TESTON 04] Teston, B. (2004). L'évaluation instrumentale des dysphonies : Etat actuel et perspectives. In A. Giovanni (Ed.), *Le bilan d'une dysphonie* (pp. 105–169). Solal.
- [TJADEN 17] Tjaden, K., & Martel-Sauvageau, V. (2017). Consonant acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear and loud speaking conditions. *American Journal of Speech-Language Pathology*, 26(2Special Issue), 569–582. https://doi.org/10.1044/2017_AJSLP-16-0090
- [TJADEN 10] Tjaden, K., Sussman, J. E., Liu, G., & Wilding, G. (2010). Long-term average spectral (LTAS) measures of dysarthria and their relationship to perceived severity. *Journal of Medical Speech-Language Pathology*, 18(4), 125–132. <https://doi.org/10.1016/j.artmed.2015.09.007.Information>
- [TJADEN 14] Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research*, 57(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372
- [TJADEN 97] Tjaden, K., & Turner, G. S. (1997). Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, 40(6), 1358–1372. <https://doi.org/10.1044/jslhr.4006.1358>
- [TJADEN 04] Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria. *Journal of Speech, Language, and Hearing Research*, 47(4), 766–783. [https://doi.org/10.1044/1092-4388\(2004/058\)](https://doi.org/10.1044/1092-4388(2004/058))
- [TOMIAK 90] Tomiak, G. R. (1990). *An acoustic and perceptual analysis of the spectral*

- moments invariant with voiceless fricative obstruents*. State University of New York at Buffalo.
- [TREMBLAY 17] Tremblay, P., Sato, M., & Deschamps, I. (2017). Age differences in the motor control of speech: An fMRI study of healthy aging. *Human Brain Mapping*, 38(5), 2751–2771. <https://doi.org/10.1002/hbm.23558>
- [TROUVILLE 21] Trouville, R., Delvaux, V., Fougeron, C., & Laganaro, M. (2021). *Logiciel d'évaluation de la parole (version screening) MonPaGe-2.0.s*. <https://lpp.in2p3.fr/monpage/>
- [TUBACH 90] Tubach, J.-P., & Louis-Jean, B. (1990). *Un corpus de transcription phonétique (300.000 phones) : Constitution et exploitation statistique*. Ecole nationale supérieure des télécommunications.
- [TURNER 95] Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, 38(5), 1001–1013. <https://doi.org/10.1044/jshr.3805.1001>
- [TUROFF 96] Turoff, M., & Hiltz, S. R. (1996). Computer-based Delphi processes. In M. Adler & E. Ziglio (Eds.), *Gazing into the oracle: The Delphi method and its application to social policy and public health* (pp. 56–85). Jessica Kingsley Publishers.
- [VALLABHA 02] Vallabha, G. K., & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication*, 38(1–2), 141–160. [https://doi.org/10.1016/S0167-6393\(01\)00049-8](https://doi.org/10.1016/S0167-6393(01)00049-8)
- [VAN DE WEIJER 91] Van de Weijer, J., & Slis, I. (1991). Nasaliteitsmeting met de Nasometer. *Logopedie En Foniatrie*, 63, 97–101.
- [VAN DEN BERG 89] Van den Berg, R. J. H. (1989). Perception of voicing in Dutch two-obstruent sequences: Covariation of voicing cues. *Speech Communication*, 8(1), 17–25. [https://doi.org/10.1016/0167-6393\(89\)90064-2](https://doi.org/10.1016/0167-6393(89)90064-2)
- [VAN DER MAATEN 08] van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2605), 2579–2605.
- [VAN LIERDE 12] Van Lierde, K., Browaeys, H., Corthals, P., Mussche, P., Van Kerkhoven, E., & De Bruyn, H. (2012). Comparison of speech intelligibility, articulation and oromyofunctional behaviour in subjects with single-tooth implants, fixed implant prosthetics or conventional removable prostheses. *Journal of Oral Rehabilitation*, 39(4), 285–293. <https://doi.org/10.1111/j.1365-2842.2011.02282.x>
- [VAN OOIJEN 96] Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory and Cognition*, 24(5), 573–583. <https://doi.org/10.3758/BF03201084>
- [VAN RIPER 63] Van Riper, C. (1963). *Speech correction: Principles and methods* (4th ed.). Prentice-Hall.
- [VAN SON 18] Van Son, R. J. J. H., Middag, C., & Demuynck, K. (2018). Vowel space as a tool to evaluate articulation problems. *Interspeech*, 357–361.
- [VAN SON 96] Van Son, R. J. J. H., & Pols, L. C. W. (1996). Acoustic profile of consonant reduction. *ICSLP*, 3(January), 1529–1532. <https://doi.org/10.1109/icslp.1996.607908>
- [VAN SON 99] Van Son, R. J. J. H., & Pols, L. C. W. (1999). Acoustic description of consonant reduction. *Speech Communication*, 28(2), 125–140. [https://doi.org/10.1016/S0167-6393\(99\)00009-6](https://doi.org/10.1016/S0167-6393(99)00009-6)
- [VAN ZUNDERT 98] van Zundert, M., Rietveld, T., & Mugge, A. M. (1998). Normgegevens voor het gebruik van de nasometer bij jonge kinderen. *Stem-, Spraak- En Taalpathologie*, 7(4), 237–247.
- [VASILESCU 04] Vasilescu, I., Candea, M., & Adda-decker, M. (2004). Hésitations autonomes dans 8 langues : une étude acoustique et perceptive. *Colloque MIDL 2004, Modélisations*

- Pour l'identification Des Langues et Des Variétés Dialectales, 29-30 Novembre, 29–30.*
- [VILAIN 15] Vilain, C., Berthommier, F., & Boe, L.-J. (2015). A brief history of articulatory-acoustic vowel representation. *First International Workshop on the History of Speech Communication Research, 1*, 148–159.
- [VOIERS 83] Voiers, W. D. (1983). Evaluating processed speech using the Diagnostic Rhyme Test. *Speech Technology*, 30–39.
- [VOLIN 11] Volín, J., & Zimmermann, J. (2011). Spectral slope parameters and detection of word stress. *Proceedings of the Technical Computing Prague*.
- [VON DER GRACHT 12] von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technological Forecasting and Social Change*, 79(8), 1525–1536. <https://doi.org/10.1016/j.techfore.2012.04.013>
- [VRANCEANU 09] Vranceanu, A. M., Cooper, C., & Ring, D. (2009). Integrating patient values into evidence-based practice: Effective communication for shared decision-making. *Hand Clinics*, 25(1), 83–96. <https://doi.org/10.1016/j.hcl.2008.09.003>
- [WADE 04] Wade, D. T., & Halligan, P. W. (2004). Do biomedical models of illness make for good healthcare systems? *Bmj*, 329(7479), 1398–1401. <https://doi.org/10.1136/bmj.329.7479.1398>
- [WALSH 05] Walsh, R. (2005). Meaning and purpose: A conceptual model for speech pathology terminology. *International Journal of Speech-Language Pathology*, 7(2), 65–76. <https://doi.org/10.1080/14417040500125285>
- [WALSH 06] Walsh, R. (2006). A history of the terminology of communication sciences and disorders. In *Australian Federal Government Department of Education, Science and Training*. <https://doi.org/10.13140/RG.2.2.15577.52329>
- [WALSHE 11] Walshe, M., & Miller, N. (2011). Living with acquired dysarthria: The speaker's perspective. *Disability and Rehabilitation*, 33(3), 195–203. <https://doi.org/10.3109/09638288.2010.511685>
- [WALSHE 08] Walshe, M., Miller, N., Leahy, M., & Murray, A. (2008). Intelligibility of dysarthric speech: Perceptions of speakers and listeners. *International Journal of Language and Communication Disorders*, 43(6), 633–648. <https://doi.org/10.1080/13682820801887117>
- [WALSHE 09] Walshe, M., Peach, R. K., & Miller, N. (2009). Dysarthria Impact Profile: Development of a scale to measure psychosocial effects. *International Journal of Language & Communication Disorders*, 44(5), 693–715. <https://doi.org/10.1080/13682820802317536>
- [WATTS 17] Watts, C. R., Awan, S. N., & Maryn, Y. (2017). A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice*, 31(3), 387.e1-387.e10. <https://doi.org/10.1016/j.jvoice.2016.09.012>
- [WEISMER 08] Weismer, G. (2008). Speech intelligibility. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), *The Handbook of Clinical Linguistics* (pp. 568–582). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781444301007.ch35>
- [WEISMER 01] 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/10.1159/000052649>
- [WEISMER 00] Weismer, G., Laures, J. S., Jeng, J., Kent, R. D., & Kent, J. F. (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatica et Logopaedica*, 53(706), 201–219.
- [WEISMER 92] Weismer, G., Martin, R., Kent, R. D., & Kent, J. F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *The Journal of the Acoustical Society of America*, 91(2), 1085–1098. <https://doi.org/10.1121/1.402635>

- [WEISMER 95] Weismer, G., Tjaden, K., & Kent, R. D. (1995). Can articulatory behavior in motor speech disorders be accounted for by theories of normal speech production? *Journal of Phonetics*, 23(1–2), 149–164. [https://doi.org/10.1016/S0095-4470\(95\)80039-5](https://doi.org/10.1016/S0095-4470(95)80039-5)
- [WELLS 83] Wells, G. B. (1983). A feature analysis of stuttered phonemes. *Journal of Fluency Disorders*, 8(2), 119–124. [https://doi.org/10.1016/0094-730X\(83\)90024-4](https://doi.org/10.1016/0094-730X(83)90024-4)
- [WHALEN 81] Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[S] boundary. *Journal of the Acoustical Society of America*, 69(1), 275–282. <https://doi.org/10.1121/1.385348>
- [WHITEHILL 02] Whitehill, T. L. (2002). Assessing intelligibility in speakers with cleft palate: A critical review of the literature. *Cleft Palate-Craniofacial Journal*, 39(1), 50–58. https://doi.org/https://dx.doi.org/10.1597/1545-1569_2002_039_0050_aaiswc_2.0.co_2
- [WHITFIELD 14] Whitfield, J., & Goberman, A. M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson’s disease. *Journal of Communication Disorders*, 51, 19–28. <https://doi.org/10.1016/j.jcomdis.2014.06.005>
- [WHITFIELD 17] Whitfield, J., & Goberman, A. (2017). Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International Journal of Speech-Language Pathology*, 19(2), 184–194. <https://doi.org/10.1080/17549507.2016.1193897>
- [WHITING 08] Whiting, L. (2008). Semi-structured interviews for novice researchers. *Nursing Standard*, 22(23), 35–40. <http://cat.inist.fr/?aModele=afficheN&cpsidt=20079971>
- [WINHOLTZ 97] Winholtz, W. S., & Titze, I. R. (1997). Miniature head-mounted microphone for voice perturbation analysis. *Journal of Speech, Language, and Hearing Research*, 40(4), 894–899. <https://doi.org/10.1044/jslhr.4004.894>
- [WINN 20] Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, 147(2), 852–866. <https://doi.org/10.1121/10.0000692>
- [WOISARD 21] 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., Pinquier, J., Pouchoulin, G., Puech, M., Robert, D., & Roger, V. (2021). C2SI corpus: A database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers. *Language Resources and Evaluation*, 55(1), 173–190. <https://doi.org/10.1007/s10579-020-09496-3>
- [WOISARD 10] Woisard, V., & Lepage, B. (2010). Perception of speech disorders: Difference between the degree of intelligibility and the degree of severity. *Audiological Medicine*, 8(4), 171–178. <https://doi.org/10.3109/1651386X.2010.525375>
- [WOLK 00] Wolk, L., Blomgren, M., & Smith, A. B. (2000). The frequency of simultaneous disfluency and phonological errors in children: A preliminary investigation. *Journal of Fluency Disorders*, 25(4), 269–281. [https://doi.org/10.1016/S0094-730X\(00\)00076-0](https://doi.org/10.1016/S0094-730X(00)00076-0)
- [WOOD 71] Wood, K. S. (1971). Terminology and nomenclature. In L. E. Travis (Ed.), *Handbook of speech pathology and audiology* (p. 3). Prentice-Hall.
- [WRAY 17] Wray, F., & Clarke, D. (2017). Longer-term needs of stroke survivors with communication difficulties living in the community: A systematic review and thematic synthesis of qualitative studies. *BMJ Open*, 7(10), 1–18. <https://doi.org/10.1136/bmjopen-2017-017944>
- [YANG 96] Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24(2), 245–261. <https://doi.org/10.1006/jpho.1996.0013>
- [YIU 94] Yiu, E. M., van Hasselt, C. A., Williams, S. R., & Woo, J. K. S. (1994). Speech intelligibility in tone language (Chinese) laryngectomy speakers. *International Journal of Language & Communication Disorders*, 29(4), 339–347.

- <https://doi.org/10.3109/13682829409031287>
- [YOON 15] Yoon, T.-J. (2015). A Corpus-based study on the effects of gender on voiceless fricatives in American English. *Phonetics and Speech Sciences*, 7(1), 117–124. <https://doi.org/10.13064/ksss.2015.7.1.117>
- [YORKSTON 84] Yorkston, K. M., Beukelman, D. R., & Traynor, C. (1984). *Assessment of intelligibility in dysarthric speech*. Pro-Ed.
- [YORKSTON 96] Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 5(1), 55–66. <https://doi.org/10.1044/1058-0360.0501.55>
- [YUNUSOVA 12] Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatica et Logopaedica*, 64(2), 94–102. <https://doi.org/10.1159/000336890>
- [YUNUSOVA 05] Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research*, 48(6), 1294–1310. [https://doi.org/10.1044/1092-4388\(2005/090\)](https://doi.org/10.1044/1092-4388(2005/090))
- [ZIEGLER 15] Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, 69, 24–39. <https://doi.org/10.1016/j.cortex.2015.04.001>
- [ZUE 76] Zue, V. W. (1976). *Acoustic characteristics of stop consonants: a controlled study* [Massachusetts Institute of Technology]. <http://hdl.handle.net/1721.1/29469>

Annexes

Annexe A – Versions originales en anglais des différents énoncés et définitions de l'étude Delphi concernant la terminologie.....	222
Annexe B – Description du panel d'experts ayant participé à l'étude Delphi	225
Annexe C – Niveaux d'intelligibilité dans la boucle de communication et exemples de potentiels facteurs perturbateurs	228
Annexe D – Tableau de résultats : Description des 22 études incluses dans la revue systématique.....	229
Annexe E – Définitions et formules (le cas échéant) des mesures acoustiques utilisées dans les études de la revue systématique.....	241
Annexe F– Tentative de comparaison croisée des résultats acoustiques de la revue systématique.....	245
Annexe G – Ensemble minimal de tâches et recommandations pour les enregistrements dans le cadre de l'évaluation des troubles de la parole chez l'adulte.....	248
Annexe H – Questionnaire destiné à des patients évalués pour un trouble de la parole	252
Annexe I – Membres du groupe de travail pour la création d'un nouveau texte pour l'évaluation de la parole et de la voix	257
Annexe J – Supports de production de parole courants en français, anglais et néerlandais ..	258
Annexe K – Texte original MonPaGe [Fougeron 19].....	262
Annexe L – Transcriptions phonétiques du nouveau texte pour l'évaluation de la parole et de la voix	263
Annexe M – Critères contrôlés dans la création du texte entier pour l'évaluation de la parole et de la voix.....	265
Annexe N – Spectres lissés pour les consonnes non-sonantes du français (segmentation manuelle)	273
Annexe O – Analyse de variation des moments spectraux selon le contexte phonémique....	275
Annexe P – Algorithme de détection de fenêtre d'analyse des moments spectraux pour les consonnes non-sonantes.....	292
Annexe Q – Boîtes à moustaches des valeurs des moments spectraux issus des deux méthodes de segmentation	295
Annexe R – Spectres lissés pour les consonnes non-sonantes du français (méthode FBDS) 298	
Annexe S – Spectres lissés pour les consonnes non-sonantes françaises de patients avec cancer ORL	300

Annexe A – Versions originales en anglais des différents énoncés et définitions de l'étude Delphi concernant la terminologie

1. Définitions de l'intelligibilité et de la compréhensibilité dans la littérature :

- [HODGE 10] :
 “**Intelligibility**, or how understandable one’s speech is to another, is a functional indicator of oral communication competence. It reflects a talker’s ability to convert language to a physical signal (speech) and a listener’s ability to perceive and decode this signal to recover the meaning of the talker’s message.”
- [HUSTAD 08] :
 « **Intelligibility** refers to how well a speaker’s acoustic signal can be accurately recovered by a listener. »
- [YORKSTON 96] :
 “The term **intelligibility** refers to the degree to which the acoustic signal (the utterance produced by the dysarthric speaker) is understood by a listener. [...] The concepts of comprehensibility and intelligibility may be distinguished by the fact that **comprehensibility** incorporates signal-independent information such as syntax, semantics, and physical context.”
- [BAREFOOT 93] :
 “[...] comprehensibility is defined as the extent to which a listener understands utterances produced by a speaker in a communication context. In our view, comprehensibility pertains to the domains of both speech and language, whereas intelligibility pertains principally to the domain of speech. The primary distinction between comprehensibility and intelligibility is that comprehensibility is intended to account for communication features of utterances that extend beyond the auditory-acoustic domain. **Comprehensibility**, in our use of the term, explicitly incorporates contextual features such as syntax, semantics, and pragmatics, and involves face-to-face communication activity in which meaningful utterances are produced by talkers and processed by listeners.”

2. Énoncés présentés au tour 2 de l'étude Delphi et pourcentages d'accord

A. Terminology

Concept	Assertion	Agreement
<i>Synonyms</i>	- “Intelligibility and comprehensibility are two terms relating to speech but are not synonyms.”	97% (33/34)
	- “Intelligibility and comprehensibility refer to the speech production abilities.”	76% (26/34)
	- “ The assessment of intelligibility and comprehensibility should not take into account the perceptual abilities of the listener. ”	41% (14/34)
<i>Message reconstruction</i>	- “Intelligibility and comprehensibility both allow the reconstruction of a message by the listener.”	97% (33/34)
	- “More precisely, the intelligibility of a message is specifically carried by the acoustic signal.”	97% (33/34)
<i>Functional communication</i>	- “Intelligibility and comprehensibility contribute to communication.”	100% (34/34)

	- “Comprehensibility refers more to the functional dimension of communication than intelligibility.”	94% (32/34)
	- “Comprehensibility refers to functional communication and is therefore not limited to the face-to-face context (e.g. audio telephone conversations).”	94% (32/34)
<i>Phonetic-acoustic production</i>	- “For intelligibility, the reconstruction of the message is made possible by the speaker’s phonetic-acoustic production ability.”	97% (33/34)
<i>Acoustic-phonetic decoding</i>	- “For intelligibility, the reconstruction of the message is made possible by the listener’s acoustic-phonetic decoding capabilities.”	97% (33/34)
	- “The most relevant unit of analysis for intelligibility is the phoneme.”	85% (29/34)
<i>Contextual elements</i>	- “Comprehensibility combines, in addition to acoustic-phonetic decoding, contextual elements, such as the linguistic or the non-verbal context.”	94% (32/34)
	- “Intelligibility, which relates to acoustic-phonetic decoding, is therefore a component of comprehensibility.”	94% (32/34)
	- “Intelligibility also includes signal-independent elements.”	76% (26/34)

B. Speech measures

Concept	Assertion	Agreement
<i>Perceptual measures</i>	- “Word-level ratings also relate to intelligibility.”	100% (34/34)
	- “Perceptually, comprehensibility is best assessed using meaning-related ratings.”	94% (32/34)
	- “Unpredictable sentences allow for a perceptual assessment of intelligibility.”	91% (31/34)
	- “Perceptually, intelligibility is best assessed using %-correct scores on low-level items (phonemes, syllables, pseudowords).”	76% (26/34)
<i>“Objective” measures</i>	- “Supra-segmental objective measures (timing, intonation, stress...) relate to both intelligibility and comprehensibility.”	91% (31/34)
	- “Voice quality also relates to intelligibility, as it contributes to the acoustic-phonetic decoding.”	85% (29/34)
	- “No objective measure is suitable to directly assess comprehensibility.”	76% (26/34)
	- “Intelligibility is best assessed using phoneme-level acoustic measures.”	74% (25/34)

3. Résultat final de l’étude Delphi : définition étayée de l’intelligibilité et de la compréhension

Intelligibility and comprehensibility are two terms relating to speech, but they are not synonyms. They both refer to the assessment of the speaker’s production abilities and both contribute to communication. Hence, while speech production is targeted, the listener’s speech perception factors cannot be dismissed (i.e., listener’s hearing loss should be excluded at least).

Intelligibility refers to the reconstruction of an *utterance* at the acoustic-phonetic level, intelligibility-related information is thus carried by the acoustic signal (i.e. intelligibility focuses on signal-dependent information). This reconstruction is made possible both by the speaker's phonetic-acoustic production ability and by the listeners acoustic-phonetic decoding skills. Perceptually, intelligibility is best analyzed on low-predictability stimuli: phonemes, syllables, pseudo-words, but also words (in minimal pairs) and unpredictable sentences for a more functional assessment taking coarticulation and phrase-level symptoms into account (e.g. respiration and prosody), as long as top-down cognitive compensation processes of the listener are avoided (i.e., no help from semantic or linguistic context).

Objectively, intelligibility can be assessed using consonant, vowel and glide acoustics (incl. inter-phoneme formant transitions), be they on isolated phonemes, or embedded in syllables, in (pseudo-)words or in sentences. Furthermore, in some cases, voice quality also contributes to intelligibility, as it plays a role in certain phonemic contrasts. Supra-segmental parameters (e.g., objectively assessed by speech rate or stress) also contribute to intelligibility.

Comprehensibility refers to the reconstruction of a *message* at the semantic-discursive level, subsequent to the acoustic-phonetic reconstruction. Therefore, intelligibility is a component of comprehensibility. In addition to the acoustic-phonetic decoding, it also includes signal-independent, contextual elements such as the linguistic or the non-verbal context. However, one can be comprehensible without all low-level units necessarily being accurately decoded; therefore, while intelligibility affects comprehensibility, the latter is, however, not fully dependent on it.

Comprehensibility refers to the more functional dimension of communication and is perceptually best assessed using meaning-related ratings (i.e., taking into account top-down cognitive processes which might compensate for degraded acoustic-phonetic information).

Nowadays, no objective instrumental measure is yet suitable to assess comprehensibility per se (i.e., the transmission of the overall meaning of the message). However, some suprasegmental parameters contribute to comprehensibility and can be objectively assessed (e.g., timing and intonation measures).

Annexe B – Description du panel d'experts ayant participé à l'étude Delphi

Tour	1	2	3
Participants	N= 40	N= 34	N= 33
Domaines d'activité	N (% ¹)	N (%)	N (%)
Troubles de la parole (TP)	38 (95 %)	32 (94,1 %)	31 (93,9 %)
TP seuls	13 (32,5 %)	11 (32,4 %)	11 (33,3 %)
TP et voix/fluence	25 (62,5 %)	21 (61,8 %)	20 (60,6 %)
Fluence seule	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Profession ²			
Orthophoniste	29 (72,5 %)	24 (70,6 %)	23 (69,7 %)
Linguiste	9 (22,5 %)	8 (23,5 %)	8 (24,2 %)
ORL/phonniatre	8 (20 %)	7 (20,6 %)	7 (21,2 %)
Informaticien	7 (17,5 %)	7 (20,6 %)	7 (21,2 %)
Psychologue/neuropsychologue	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Audiologiste	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Neuroscientifique	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Autres (FLE, performance/comédie/chant)	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Années d'expérience en parole/voix			
>20 années	13 (32,5 %)	12 (35,3 %)	12 (36,4 %)
15-20 années	6 (15 %)	5 (14,7 %)	5 (15,2 %)
10-15 années	4 (10 %)	3 (8,8 %)	3 (9,1 %)
5-10 années	10 (25 %)	8 (23,5 %)	8 (24,2 %)
<5 années	7 (17,5 %)	6 (17,6 %)	5 (15,2 %)
Activité professionnelle			
Premier rang (activité principale)			
Recherche	14 (35 %)	14 (41,2 %)	14 (42,4 %)
Clinique	13 (32,5 %)	10 (29,4 %)	9 (27,3 %)
Académique	11 (27,5 %)	9 (26,5 %)	9 (27,3 %)
Industriel	2 (5 %)	1 (2,9 %)	1 (3 %)
Combinaisons	34 (85 %)	32 (94,1 %)	31 (93,9 %)
Activité clinique seule	4 (10 %)	3 (8,8 %)	2 (6,1 %)
Activité clinique et Recherche	21 (53 %)	18 (52,9 %)	18 (54,5 %)
Recherche et Académique	28 (70 %)	26 (76,5 %)	26 (78,8 %)
Niveau d'éducation			
Troisième cycle (Doctorat)	23 (57,5 %)	21 (61,8 %)	21 (63,6 %)
Deuxième cycle (Master)	16 (40 %)	13 (38,2 %)	12 (36,4 %)
Dont doctorants	3 (7,5 %)	3 (8,8 %)	3 (9,1 %)
Premier cycle (Bachelier)	1 (2,5 %)	0 (0 %)	0 (0 %)
Année de diplomation			
Moyenne (E.T.)	2009,6 (7,88)	2010,3 (7,72)	2010,1 (7,72)
Médiane (EIQ)	2012 (11,5)	2014 (12)	2014 (12)
Titre académique (équivalents du système américain)			

Associate Professor	16 (40 %)	15 (44,1 %)	15 (45,5 %)
Aucune activité académique	8 (20 %)	6 (17,6 %)	5 (15,2 %)
Lecturer	5 (12,5%)	4 (11,8%)	4 (12,1%)
Pas de titre académique (p. ex. doctorant)	3 (7,5 %)	3 (8,8 %)	3 (9,1 %)
Professor	3 (7,5%)	3 (8,8%)	3 (9,1%)
Clinical Instructor	2 (5 %)	0 (0 %)	0 (0 %)
Assistant Professor	1 (2,5%)	1 (2,9%)	1 (3%)
Professor Emeritus	1 (2,5%)	1 (2,9%)	1 (3%)
Senior Lecturer	1 (2,5%)	1 (2,9%)	1 (3%)
Pays			
France	7 (17,5 %)	7 (20,6 %)	7 (21,2 %)
Royaume-Uni	7 (17,5 %)	6 (17,6 %)	6 (18,2 %)
Allemagne	6 (15 %)	5 (14,7 %)	5 (15,2 %)
Canada	3 (7,5 %)	2 (5,9 %)	2 (6,1 %)
Australie	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Belgique	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Finlande	2 (5 %)	1 (2,9 %)	0 (0 %)
Malte	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Colombie	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Hong Kong	1 (2,5 %)	0 (0 %)	0 (0 %)
Inde	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Pays-Bas	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Pakistan	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Espagne	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Suède	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Suisse	1 (2,5 %)	0 (0 %)	0 (0 %)
États-Unis	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Langue principale²			
Anglais	25 (62,5 %)	22 (64,7 %)	22 (66,7 %)
Français	12 (30 %)	11 (32,4 %)	11 (33,3 %)
Allemand	5 (12,5 %)	4 (11,8 %)	4 (12,1 %)
Finnois	2 (5 %)	1 (2,9 %)	0 (0 %)
Espagnol	2 (5 %)	2 (5,9 %)	2 (6,1 %)
Chinois	2 (5 %)	1 (2,9 %)	1 (3 %)
Catalan	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Néerlandais	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Hindi	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Kannada	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Malayalam	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Maltais	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Portugais	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Suédois	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Urdu	1 (2,5 %)	1 (2,9 %)	1 (3 %)
Population principale de patients³			
Groupes d'âge			
% moyen pour l'ensemble des participants			
Enfants (<10 ans)	24,8 %	22,2 %	20,6 %
Adolescents (10-18 ans)	10,1 %	7,8 %	7,9 %
Adultes jeunes (19-35 ans)	13,1 %	11,5 %	11,7 %
Adultes d'âge moyen (36-65 ans)	20,2 %	22,3 %	22,7 %
Personnes âgées (65+)	31,8 %	36,2 %	37,1 %
Pathologies			

Pathologies neurodégénératives	22,1 %	25,7 %	26,4 %
Pathologies neurologiques acquises	15,4 %	17,1 %	17,2 %
Autres (liés à la parole/voix)	13,6 %	8,8 %	7,7 %
Oncologie	8,9 %	10,4 %	10,8 %
Troubles de la voix d'origine fonctionnelle	5,6 %	5,4 %	5,2 %
Autres TP d'origine structurelle	5,2 %	5,7 %	5,9 %
Troubles de la fluence	5,1 %	4,7 %	4,7 %
Autres (non liés à la parole/voix)	4,9 %	2,6 %	2,1 %
Troubles de la voix d'origine structurelle	4,6 %	4,1 %	4,1 %
Déficience auditive	4,5 %	4,6 %	4,7 %
Troubles de la voix d'origine neurologique	4,1 %	4,6 %	4,7 %
Polyhandicap	3,8 %	3,8 %	3,9 %
Tumeurs neurologiques	1,8 %	2,1 %	2,1 %
Iatrogénie	0,4 %	0,4 %	0,5 %

¹ En raison des arrondissements des chiffres à une décimale, des écarts minimes par rapport à 100 % sont possibles

² Les combinaisons étaient autorisées, les totaux par colonne ne s'élèvent donc pas forcément à 100 %.

³ Chaque participant a réparti 100 % sur toutes les catégories données ; les résultats sont des pourcentages moyens sur l'ensemble des participants pour chaque catégorie

Annexe C – Niveaux d’intelligibilité dans la boucle de communication et exemples de potentiels facteurs perturbateurs

<i>Niveaux de la boucle</i>			<i>Exemples de facteurs pouvant affecter l’intelligibilité</i>
Encodage du message	Linguistique et Paralinguistique	Encodage phonologique	<i>Physiologique</i> : processus phonologiques dans le développement de la parole, lapsus (paraphasies phonologiques) <i>Pathologique</i> : troubles phonologiques
		Encodage prosodique	<i>Physiologique</i> : phonèmes dans les syllabes non accentuées vs accentuées ; différences individuelles dans l’encodage prosodique <i>Pathologique</i> : aprosodie, troubles de la fluence
	Acoustique-phonétique	Planification motrice, encodage de séquences de mouvements (systèmes respiratoire, phonatoire et articulatoire)	<i>Physiologique</i> : lapsus (paraphasies phonétiques) <i>Pathologique</i> : dyspraxie verbale développementale, apraxie acquise de la parole, atteintes cérébelleuses, troubles de la fluence
Énoncé	Exécution motrice	Coordination et exécution des mouvements (systèmes nerveux central et périphérique et muscles) pour produire les séquences de phonèmes	<i>Physiologique</i> : compromis entre les limites physiologiques du système de production de la parole et les nombreuses demandes de la parole spontanée (p. ex. hypoarticulation dans la théorie H&H [LINDBLOM 90] ; coarticulation) <i>Pathologie</i> : dysarthrie, déficits structurels après un traumatisme ou une intervention chirurgicale
	Paramètres suprasegmentaux	Prosodie et paralinguistique (p. ex. débit de parole et paramètres vocaux)	<i>Physiologique</i> : stress, tempérament <i>Pathologique</i> : troubles de la fluence, dysphonie
	Analyse auditive	Réception et traitement du signal acoustique par le système auditif périphérique	<i>Physiologique</i> : presbycusie <i>Pathologique</i> : déficit auditif, surdité
Décodage du message	Acoustico-phonétique	Mise en correspondance des caractéristiques acoustiques avec les traits phonétiques pour identifier les séquences de phonèmes	<i>Physiologique</i> : perception de contrastes phonémiques dans une langue étrangère <i>Pathologique</i> : troubles du traitement auditif central de type déficit du décodage auditif
	Linguistique et Paralinguistique	Décodage phonologique	<i>Pathologique</i> : Troubles phonologiques
		Décodage prosodique	<i>Pathologique</i> : Troubles du traitement auditif central de type déficit prosodique ; troubles du spectre autistique

Annexe D – Tableau de résultats : Description des 22 études incluses dans la revue systématique

A. Études décrivant les associations entre les variables acoustiques et la perception auditive chez les locuteurs sains							
Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Conclusions
A.1. Mesure du résultat : Association entre mesure(s) acoustique(s) et échelle(s) d'évaluation perceptive							
A01. MCRÆ 02	III-2	20/24 83 % haute	N=13 (9 H, 4 F) Âge : $\mu=67$ (intervalle 59-80) Langue : Anglais américain	- Voyelles /i, a, u, æ/ délimitées par des consonnes plosives - Fricatives /s, ʃ/ en positions initiales et finales de mots	- Voyelles : Aire de l'espace vocalique (VSA ; quadrilatère, utilisant les fréquences de F1 et F2 au point médian temporel) - Consonnes : Différence /ʃ/-/s/ du coefficient du 1 ^{er} moment ; un coefficient du 1 ^{er} moment plus bas suggère une constriction plus postérieure, plus lâche ou un arrondissement labial accru dans /ʃ/	Sévérité globale de la parole : estimation directe de magnitude (DME) avec un modulus d'une valeur de 100 (= modérément sévère)	<i>Association :</i> - Régression entre l'aire de l'espace vocalique et sévérité globale de la parole : non significatif - Régression entre la différence de 1 ^{er} moment et la sévérité globale de la parole : non significatif <i>Données descriptives :</i> Moyenne (intervalle) - Sévérité globale de la parole (DME) : 28 (2-51) - VSA : NR (graphiques de quadrilatères vocaliques) - Différence de 1 ^{er} moment : NR (graphiques)
A02. HAZAN 04	III-3	21/24 88 % haute	N=33 (15 H, 18 F) Âge F $\mu=33,11$ (ET=10,9) ; H $\mu=30,7$ (ET=10,5) Langue : Anglais britannique	Voyelles /i, u, æ/ dans des mots monosyllabiques CVC	- Mesures de formants de voyelles : • F1 et F2 dans la partie stable de la voyelle • Distance euclidienne entre F1 et F2 pour chaque voyelle - Mesures de l'espace vocalique : • Différence de fréquence de F1 pour /i/ et /æ/ • Différence de fréquence de F2 pour /i/ et /u/	- Pourcentage d'identification correcte (tâche NR) - Évaluations subjectives (échelle à 7-points, 1-7 : 7= score le plus élevé sur l'attribut positif de la paire) : marmonné-précis (mumbly-precise), déplaisant-plaisant (unpleasant-pleasant), assourdi-clair	<i>Association :</i> - Mesures de formants de voyelles : • F1, F2 et distance euclidienne F1-F2 : pas de corrélation significative avec le pourcentage d'identification correcte, ni avec les échelles d'évaluation subjective - Mesures de l'espace vocalique : • Différence F1 /i/-/æ/ : non corrélée au pourcentage d'identification correcte, ni avec aucune des évaluations subjectives • Différence F2 /i/-/u/ : corrélation significative avec le pourcentage d'identification correcte ($r=0,401$;

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Conclusions
A03. NEEL 08	III-3	18/22 82 % haute	N=93 (45 H, 48 F) Âge : NR Langue : Anglais américain (dialecte Michigan/Upper Midwest)	Voyelles /i, ɪ, e, ε, æ, a, ʌ, o, u, u/ en contexte /hVd/	- Mesures globales : • F1 et F2 moyens à travers les 10 voyelles • Quantité moyenne de mouvement des formants, moyennée sur les 10 voyelles : Somme de la distance euclidienne dans l'espace F1xF2 entre le début et l'état stable de la voyelle, et de la distance euclidienne entre l'état stable et la fin de la voyelle - Mesures fines : • Aire de l'espace vocalique (VSA, quadrilatère ; deux triangles : /i, æ, u/ et /æ, u, a/) • Distance moyenne entre les voyelles (dispersion vocalique) : distance euclidienne moyenne entre chaque paire de voyelles	Pourcentage d'identification correcte sur l'ensemble des 10 voyelles pour chaque locuteur	(muffled-clear), voilé-non voilé (husky-not husky), rauque-non rauque (creaky-not creaky), nasal-non nasal, aigu pour une femme/un homme-grave pour une femme/un homme, ténu-riche (thin-rich), faible-puissant (weak-powerful) et brusque-fluide (harsh-smooth) p=0,006) et avec deux échelles d'évaluation (rs= NR ; p<0,01) : marmonné-précis, assourdi-clair <i>Données descriptives :</i> • Pourcentage d'identification correcte : NR (graphiques) • Échelles d'évaluation subjective : NR (graphiques) • Mesures de formants et d'espace vocaliques : NR <i>Association :</i> - Aucune des mesures acoustiques, individuellement ou combinées, n'explique plus de 18 % de la variance de l'identification des voyelles - Prédicteurs significatifs de l'identification de voyelles : • Hommes : VSA (r ² =0,12 ; p<0,02) et intervalle F1 (r ² =0,14 ; p<0,01) • Femmes : VSA (r ² =0,09 ; p<0,02) et quantité moyenne de mouvement des formants (r ² =0,11 ; p<0,02) - Prédicteurs non significatifs : F1, F2, mouvement des formants chez les hommes, distance entre les voyelles, intervalle F1 chez les femmes, intervalle F2, rapport dynamique <i>Données descriptives :</i> Moyenne (ET) : • Scores d'identification de voyelles : H=95,6 % (4,0 %) ; F=96,8 % (2,6) • F1 (Bark) : H=5,04 (0,20) ; F=5,88 (0,30) • F2 (Bark) : H=13,05 (0,37) ; F=14,70 (0,53) • Mouvement des formants : H=1,43 (0,29) ; F=1,99 (0,47) • VSA : H=18,57 (4,13) ; F=25,07 (6,55)

- Intervalles F1 et F2 : soustraction de la valeur F1/F2 la plus basse de la valeur la plus haute
- Rapport dynamique (distinctivité entre les voyelles avec des trajectoires dynamiques et statiques) : distance euclidienne moyenne (du début à l'état stable et à la fin de la voyelle dans l'espace F1 x F2) couverte par les 3 voyelles les plus dynamiques (/æ, ʌ, ʊ/) divisée par la distance couverte par les 3 voyelles les plus statiques (/i, ε, u/)
- Distance entre voyelles : H=4,54 (0,50) ; F=5,46 (0,66)
- Intervalle F1 (Bark) : H=3,83 (0,59) ; F=4,32 (0,80)
- Intervalle F2 (Bark) : H=9,37 (1,04) ; F=11,15 (1,06)
- Rapport dynamique : H=1,97 (0,53) ; F=2,24 (0,59)

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Conclusions
A04. FERGUSON 14	III-3	20/22 91 % haute	N=41 (20 H, 21 F) Âge : intervalle 18-45 Langue : Anglais	Voyelles /i, ɪ, e, ε, æ, a, ʌ, o, ʊ, u/ en contexte /bʌd/, en parole conversationnelle (CON) et en parole surarticulée (clear speech, CL)	- Valeurs stationnaires de F1 et F2 - Valeurs dynamiques : • Longueur de vecteur (VL) = Distance euclidienne du vecteur dans l'espace F1xF2 connectant les valeurs des formants à 20 % et 80 % de la durée de la voyelle • Longueur de trajectoire (TL) = Somme des longueurs de quatre sections de voyelle temporellement équidistantes : 20 %-35 %, 35 %-50 %, 50 %-65 %, 65 %-80 %	Pourcentage d'identification correcte de voyelles : sélection parmi dix ensembles de trois mots-clés [p. ex. « (1) feet, thief, bead, (2) sit, rib, bid » etc.]	<i>Association :</i> - F1 stationnaire : Corrélation significative et forte avec l'identification de voyelles (Z=4,5 ; p<0,0001) - F2 stationnaire : Corrélation significative et avec l'identification de voyelles (Z=4,6 ; p<0,0001) - VL : Pentés de régression significativement positives avec l'identification de voyelles (Z=5,1 ; p<0,001) : Les voyelles avec un changement plus important dans l'espace F1xF2 sont mieux identifiées - TL : Effet non significatif sur la précision de l'identification de voyelles : une trajectoire plus courbée dans l'espace vocalique n'affecte pas l'identification de voyelles <i>Données descriptives :</i> - Identification de voyelles : NR - F1 : NR (graphiques) - F2 : NR (graphiques) - VL moyen (Barks) : CL=1,27 ; CON=1,1 ; ratio=0,09 - TL moyen (Barks) : CL=0,50 ; CON=0,49 ; ratio=1

A05. WHITFIELD 17	III-3	21/24 88 % haute	N=10 5 H (Âge : $\mu=24,40$, intervalle 20-36) 5 F (Âge : $\mu=24,30$; intervalle 18-29) Langue : Anglais américain standard	Voyelles cardinales /i, a, u, æ/ dans des phrases, en parole conversationnell e (CON) et en parole surarticulée (clear speech, CL)	- Aire de l'espace vocalique acoustico-articulatoire (AAVS) : racine carrée de la variance généralisée de tous les formants de voyelles échantillonnés dans le graphe de coordonnées F1xF2 Cette mesure a été effectuée sur une phrase contenant toutes les voyelles cardinales, et sur deux phrases du Rainbow Passage. - Aire de l'espace vocalique traditionnel (VSA) : valeurs formantiques mesurées durant l'état stable des noyaux vocaliques des mots « stack, key, blue, box » (dans la phrase contenant toutes les voyelles cardinales)	Évaluation de la clarté de parole sur une échelle visuelle analogique de 100 mm (0-100 : pas clair – très clair)	<i>Association</i> : Corrélation significative entre les scores de différence perceptifs (conversationnel vs clair) et le changement relatif de l'AAVS ($r=0,67$; $r^2=0,45$; $p<0,01$) <i>Données descriptives</i> : Moyenne (ET) : A. Phrase avec voyelles cardinales - Évaluation de clarté de la parole (mm) : • CON : H=66,33 (2,14) ; F=71,97 (8,82) • CL : H=80,10 (4,85) ; F=85,80 (3,72) - AAVS (kHz) : • CON : H=31,59 (3,77) ; F=77,17 (16,43) • CL : H= 43,43 (6,67) ; F=107,17 (19,87) - VSA (kHz) : H=200,81(23,65) ; F= 577,74(94,11) B. Phrase 1 du Rainbow Passage - Évaluation de clarté de la parole (mm) : • CON : H=61,80 (7,55) ; F=70,73 (4,93) • CL : H= 74,20 (3,72) ; F=80,57 (8,77) - AAVS (kHz) : • CON : H=24,45 (6,54) ; F=61,83 (6,44) • CL : H=31,10 (7,05) ; F=83,23 (18,60) C. Phrase 3 du Rainbow Passage - Évaluation de clarté de la parole (mm) : • CON : H=64,60(11,16) ; F=66,57(11,88) • CL : H=75,63 (11,12) ; F=87,03 (9,05) - AAVS (kHz) : • CON : H=27,98 (5,06) ; F=68,83 (6,86) • CL : H=35,37 (8,13) ; F=93,81 (20,21)
-------------------------	-------	------------------------	--	--	--	---	---

A.2. Mesure du résultat : Association indirecte entre mesure(s) acoustique(s) et évaluation perceptive

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Conclusions
A06. KATZ 91	III-3	20/24 83 % haute	N=10 (5 H, 5 F) Âge : $\mu=32$ (ET=6,7 ; intervalle 26-45)	/s/ dans /su/ et /si/	- Énergie centroïde de la fricative à 30 ms et 100 ms avant sa fin : supposé indiquer des résonances des	En entendant seulement le son /s/, identification de la syllabe d'origine	- Scores d'identification de syllabes : NR - Énergie centroïde moyenne (Hz) 30 ms avant la fin : • /si/ : 5524 • /su/ : 5134

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Conclusions
							<p>Langue : Anglais</p> <p>cavités antérieures (indication du degré de mouvement labial anticipatoire)</p> <p>- Pics spectraux de la fricative à 30 ms avant sa fin, anticipant F2 de la voyelle</p> <ul style="list-style-type: none"> • Ratio /si-/su/ : 1,08 <p>- Énergie centroïde moyenne (Hz) 100 ms avant la fin :</p> <ul style="list-style-type: none"> • /si/ : 6806 • /su/ : 6182 • Ratio /si-/su/ : 1,10 <p>Valeurs pour /i/ significativement plus élevées que pour /u/, (reflétant la coarticulation labiale), F=30,9 ; p<0,001 ; Valeurs pour la fenêtre 100 ms significativement plus élevées que pour la fenêtre 30 ms (F=107,2 ; p<0,001)</p> <p>- Pics spectraux moyens en fin de la fricative :</p> <ul style="list-style-type: none"> • /si/ : 1999 • /su/ : 1866 • Ratio /si-/su/ : 1,07 <p>Effets de contexte vocalique importants, valeurs pour /si/ plus élevées que pour /su/ (reflétant les effets de coarticulation linguale et labiale), F=101,2 ; p<0,001</p>
A07. FLEGE 92	III-3	21/24 88 % haute	N=30 (15 H, 15 F) 10 natifs Anglais (Âge : $\mu=27,9$; ET=6,5) 10 natifs Espagnol « expérimentés » (Âge : $\mu=28,2$; ET=5,7) 10 natifs Mandarin « expérimentés » (Âge : $\mu=28,1$; ET=3,0) Langue : Anglais américain	Contraste /t-/d/ en fins de mots dans des paires minimales de CVC (/bvt/-/bvd/ et /svt/-/svd/) contenant soit /i, ɪ, ε, æ/.	Fréquence finale de F1 dans les 45 dernières ms des voyelles	Pourcentage d'identification correcte pour les plosives /t-/d/ en position finale de mots ; calcul de scores A' basés sur le nombre d'identifications correctes de /t/ et de fausses alertes (identification incorrecte de /d/ comme non voisé), sur des mots /bVC/	<p>- Locuteurs natifs anglais :</p> <ul style="list-style-type: none"> • En général, fréquence finale de F1 significativement plus élevée dans les mots se terminant en /t/ que dans les mots se terminant par /d/ • Pour paires /bvt/-/bvd/ (valeurs de F entre 11,9 et 20,8, p<0,01), fréquence finale de F1 en moyenne plus élevée de 68 Hz pour /t/ • Pour paires /svt/-/svd/ (F= 76,7 ; p<0,01), fréquence finale de F1 en moyenne plus élevée de 73 Hz pour /t/ <p>- Locuteurs expérimentés non natifs de l'anglais :</p> <ul style="list-style-type: none"> • Différences acoustiques non significatives pour /bvt/-/bvd/ • Pour /svt/-/svd/, différences significatives (543 vs 518 Hz) pour les sujets mandarins (F= 11,6 ; p<0,01), non significatives pour les sujets espagnols <p>Chez les locuteurs expérimentés non natifs de l'anglais, la fréquence finale de F1 représentait 1,5 %</p>

de la variance de l'identification /t/-/d/ (F=4,67 ; p=0,032)

Données descriptives :

- Pourcentage correct d'identification : taux global de 68 % à 71 %. Taux plus élevés pour /t/ que /d/ (82 % vs 65 %). Scores corrects significativement plus élevés pour plosives des locuteurs natifs anglais (A'=0,953) que pour les locuteurs expérimentés et non expérimentés espagnols et mandarins (A'=0,751 ; 0,720 ; 0,668 et 0,679, respectivement).
- Fréquence finale de F1 : NR (graphiques)

A08. BUNTON 01	III-2	21/24 88 % haute	N=10 (5 H, 5 F) Âge : intervalle 68-77 Langue : Anglais américain (dialecte Upper Midwest)	Voyelles hautes vs basses (hauteur de langue) : paires de voyelles /ε-æ/ et /i-ε/ dans des mots	Différences entre F0-F1 et entre F1-F2 (mesurées à 50 % de la durée totale de la voyelle)	Pourcentage d'identification correcte de mots, format choix multiple (paires minimales ou quasi minimales)	Dans les productions correctement perçues - Différences F0-F1 statistiquement significatives (/ε/-/æ/ : U = 222,0 ; p<0,001 ; /i/-/ε/ : U = 194,0 ; p<0,001) - Différences F1-F2 statistiquement significatives (/ε/-/æ/ : U = 179,5 ; p<0,001 ; /i/-/ε/ : U = 116,0 ; p<0,001) Ces mesures sont donc considérées comme étant liées au contraste perçu de hauteur linguale. <i>Données descriptives :</i> - Pourcentage d'identification correcte : 96,44 % (94,38 %-98,38 %) - F0-F1, F1-F2 : NR (graphiques)
-------------------	-------	------------------------	--	---	---	--	--

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Conclusions
A09. FERGUSON 07	III-3	20/24 83 % haute	N=12 (6 H, 6 F) Âge : μ=32,75 (intervalle 20-45) Langue : Anglais américain	Voyelles /i, ɪ, e, ε, æ, a, ʌ, o, u, u/ en contexte /bVd/, en parole conversationnelle (CON) et en parole surarticulée (clear speech, CL)	- Valeurs formantiques stationnaires : • Périmètre : dimension totale de l'espace vocalique ; somme de 4 distances euclidiennes entre voyelles adjacentes (/i/-/æ/, /æ/-/a/, /a/-/u/, and /u/-/i/) • Intervalle F1 : différence entre le F1 moyen des voyelles basses /æ, a/ et le F1 des voyelles hautes /i, u/	Pourcentage de voyelles correctes en CL et en CON : Création des groupes « Sans Bénéfice » (« No Benefit », NB) et « Grand Avantage » (« Big Benefit », BB) selon le gain d'intelligibilité des voyelles (BB = effet important de la parole	<i>Association :</i> Différences de bénéfice de parole surarticulée (différences entre les groupes de locuteurs NB et BB) : - Valeurs formantiques stationnaires : • Périmètres : Pas de différence significative, F = 0,66 ; p = 0,44 • Intervalle F1 : Pas de différence significative, F = 0,497 ; p = 0,5 • Intervalle F2 antérieur et postérieur : Pas de différence significative, F = 0,29 ; p = 0,59 ; F = 2,1 ; p = 0,15

- F2 antérieur : valeur moyenne de F2 pour /i, ɪ, e, ε, æ/
- F2 postérieur : valeur moyenne de F2 pour /ɑ, ʌ, o, ɔ, u/

surarticulée pour les auditeurs, en comparaison à la parole conversationnelle)

- Mouvement dynamique des formants :
- Changement spectral (λ) : somme du déplacement absolu (de 20 % à 80 %) des fréquences de F1 et F2
- Angle spectral (Ω) : Somme, en radians, des valeurs absolues des angles de F1 et F2, calculés comme les arc tangents de la différence entre les fréquences des formants à 20 % et à 80 % de la voyelle, divisés par la durée entre ces deux points

- Mouvement dynamique des formants :
- Changement spectral : Significativement plus important pour NB, $F = 10,86$; $p < 0,01$
- Angle spectral : Pas de différence significative, $F = 3,06$, $p = 0,08$

Données descriptives :

- Pourcentage d'identification correcte de voyelles :
 - BB : CL=79,2 ; CON=60,6 ; différence = 18,6
 - NB : CL= 67,0 ; CON=68,1 ; différence = - 1,1
- Valeurs formantiques stationnaires (toutes mesures en Barks) :
- Périmètres :
 - BB : CL=13,77 ; CON=12,65 ; différence=1,12
 - NB : CL=14,26 ; CON=13,85 ; différence=0,41
- Intervalle F1 :
 - BB : CL=3,07 ; CON=2,79 ; différence=0,28
 - NB : CL=3,12 ; CON=3,16 ; différence=-0,04
- Intervalle F2 antérieur :
 - BB : CL=13,13 ; CON=12,82 ; différence=0,31
 - NB : CL=13,2 ; CON=13,06 ; différence=0,14
- Intervalle F2 postérieur :
 - BB : CL=10,13 ; CON=10,26 ; différence=-0,13
 - NB : CL=9,75 ; CON=9,86 ; différence=-0,11
- Mouvement dynamique des formants :
- Changement spectral (Barks) :
 - BB : CL=1,88 ; CON=1,56 ; différence=0,32
 - NB : CL=2,27 ; CON=2,12 ; différence=0,15
- Angle spectral (radians) :
 - BB : CL=0,89 ; CON=1,04 ; différence=-0,15
 - NB : CL=1,03 ; CON=1,12 ; différence=-0,09

B. Études présentant des données descriptives sur les variables acoustiques chez des locuteurs sains

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Données descriptives concernant les locuteurs sains
A10. WEISMER 92	III-2	17/24 71 % bonne	N=15 (H) Âge : $\mu=72$ (intervalle 68-80) Langue : Anglais américain	Voyelles (NR) dans 12 mots monosyllabiques (CV, CVC, CCVC, CVCC, VC) : wax, sigh, sip, ship, sew, coat, row, cash, hail, ate, shoot, and blend.	Trajectoires formantiques F1-F2 : • Étendue de transition (TE) : degré de changement en fréquence le long du segment de transition d'une trajectoire • Pente ou vitesse de transition : TE/TD (TD : durée du segment de transition) • Fréquence de début (SF) : fréquence initiale du segment de transition	Pourcentage d'identification correcte de mots : sélection parmi quatre mots possibles	<i>Perceptif</i> - Pourcentage d'identification correcte : NR <i>Acoustique</i> - Données descriptives fournies pour toutes les mesures acoustiques : pour chacun des 12 mots, pour F1 (Tableau 2, page 1094) et pour F2 (Tableau 3, page 1095)
A11. HOHOFF 03	III-3	20/24 83 % haute	N=23 (6 H, 17 F) Âge : $\mu=35,1$ (ET=10,3 ; intervalle 19,6-57,1) Langue : Français standard	/s/ dans le mot « soleil »	Fréquence limite supérieure (UBF) du bruit de friction : Fréquence maximale de la bande, intervalle de gris maximum dans le spectrogramme à large bande	Échelle de Likert à 5 points (1-5 : production du son /s/ non pathologique – hautement pathologique)	Moyenne (ET ; intervalle) : <i>Perceptif</i> - échelle Likert pour le son /s/ : 1,43 (0,46 ; 1,00-3,20) <i>Acoustique</i> - UBF son /s/ : 12 961,48 Hz (585,77 ; 11 454 - 13 898)
A12. YUNUSOVA 05	III-2	20/24 83 % haute	N=10 (7 H, 3 F) Âge H : $\mu=56,9$; F : $\mu=57,3$ Langue : Anglais américain (dialecte Upper Midwest)	- Tous les segments vocaliques d'une lecture à voix haute - Fricatives /s, \int / en position initiale de mots	- Segments vocaliques : écart interquartile (EIQ) de F2 pour chaque groupe respiratoire - Fricatives : Différences de 1 ^{er} moment entre /s/-/ \int /	- Niveau phrastique : estimation directe de magnitude, modulus de 100 (facilité de compréhension) - Niveau lexical : Pourcentage d'identification correcte de mots	Moyenne (ET) : <i>Perceptif</i> - Intelligibilité phrases (DME) : 222,38 (23,81 ; 175-244) - Intelligibilité mots (%) : 98,66 (1,12 ; 96,6-99,9) <i>Acoustique</i> - EIQ F2 (Hz) : 500 (73) - Différence de 1 ^{er} moment /s/-/ \int / (Hz) : 1,3 (0,7)
A13. DE BRUIJN 09	III-2	21/24 88 % haute	N=18 [« appariés en genre et en âge » au groupe d'étude ; groupe d'étude H=55 %, F=45 %,	Voyelles cardinales /a, i, u/ et consonne vélaire /x/ dans des mots	- F1 et F2, et espace vocalique (vowel space) - Pente spectrale pour /x/	Évaluations de l'intelligibilité (échelle à 10 points, 1-10 : faible - bonne intelligibilité), articulation et	Moyenne (ET) : <i>Perceptif</i> - Jugements d'intelligibilité : NR <i>Acoustique</i> - F1 /i/ : 296 Hz (49)

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Données descriptives concernant les locuteurs sains
			Âge $\mu=53,8$ (ET=8,7)] Langue : Néerlandais			résonance nasale (échelle à 4 points, 1-4 : qualité de parole normale-déviante)	- F2 /i/ : 2325 Hz (248) - Aire du triangle vocalique : 0,213 Hz ² (0,11) - F1 et F2 /a, u/ : NR - Pente spectrale pour /x/ : NR
A14. VAN LIERDE 12	III-2	20/24 83 % haute	N=9 (ratio H/F NR) $\mu=47,6$ (22-61) Langue : Néerlandais	Son /s/	2 ^e moment spectral : dispersion (écart-type) des fréquences autour du centre de gravité	- Transcription phonétique - Évaluation de l'intelligibilité globale sur une échelle ordinale à 4 points (0-3 : normal – sévèrement altérée)	<i>Perceptif</i> - Transcription phonétique (traduit de l'anglais) : « Tous les sujets étaient capables de produire toutes les voyelles et consonnes néerlandaises... les caractéristiques phonétiques étaient normales » - Intelligibilité (traduit de l'anglais) : « Le groupe contrôle avait une intelligibilité normale » <i>Acoustique</i> - 2 ^e moment spectral sur /s/ : 2081 Hz
A15. SKODDA 13	III-2	21/24 88 % haute	N=60 (30 H, 30 F) Âge : $\mu=66,87$ (médiane=67,5 ; ET=7,1 ; intervalle 55-80) Langue : Allemand	Voyelles cardinales /a, i, u/ dans des mots	Indice d'articulation vocalique (VAI)	Évaluations sur une échelle à 4 points : intelligibilité (0-3 : bonne-faible intelligibilité) et sur une échelle à 5 points : articulation (0-4 : articulation normale – intelligibilité nettement réduite)	Moyenne (ET) : <i>Perceptif</i> - Jugement d'intelligibilité : 0,08 (0,28) - Jugement d'articulation : 0,07 (0,25) <i>Acoustique</i> - VAI : H=0,767 (0,058) ; F=0,874 (0,062)
A16. WHITFIELD 14	III-2	23/24 96 % haute	N=10 5 H (Âge : $\mu=65,8$; intervalle 57-73), 5 F (Âge : $\mu=71,8$; intervalle 58-81) Langue : Anglais américain standard	Tous les segments voisés du premier paragraphe du Rainbow Passage	Espace vocalique acoustico-articulaire (AAVS) : racine carrée de la variance généralisée des données F1-F2, résultant en un espace elliptique représentant la variabilité moyenne bivariée dans l'espace F1xF2	Évaluation de la clarté de la parole sur une échelle visuelle analogique de 100 mm (0-100 : pas clair-très clair)	Moyenne (ET) : <i>Perceptif</i> - Jugement de clarté de parole (mm) : H=63mm (9,92) ; F=64,60mm (17,96) <i>Acoustique</i> - AAVS (kHz ²) : H=38,45 (5,20) ; F= 64,59 (9,77)
A17. NEEL 15	III-2	22/22 100 % haute	N=12 (4 H, 8 F) Âge : intervalle 52-69 Langue : Anglais	- /s/ initial dans les mots « sip, seep, see » - Voyelles /a, i, u/ dans les mots « heed, hod, who'd »	- Consonne : 1 ^{er} moment spectral au centre de la fricative /s/ = moyenne pondérée des fréquences des pics spectraux (mesure de précision du placement lingual)	À l'aide d'une échelle visuelle analogique de 100 mm (0-100 : pas d'altération – altération sévère) : intelligibilité globale, précision articulaire	Moyenne (ET ; intervalle) : <i>Perceptif</i> - Jugement d'intelligibilité : 4,4 - Jugement de précision articulaire : NR (graphiques) <i>Acoustique</i> - Consonne : 1 ^{er} moment spectral sur /s/ (Hz) :

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Données descriptives concernant les locuteurs sains
					- Voyelles : • intervalle F1 (valeur F1 la plus basse des 3 voyelles soustraite de la valeur la plus élevée) • intervalle F2 • Aire de l'espace vocalique (VSA)		6962,6 (1282,6 ; 4700–8756) - Voyelles : • Intervalle F1 (Hz) : 448,9 (83,9 ; 286–532) • Intervalle F2 (Hz) : 1552,8 (197,8 ; 1309–1899) • VSA (Hz ²) : 334 262 (98,557 ; 192980–526903)
A18. DWIVEDI 16	III-2	24/24 100 % haute	N=51 (32 H, 19 F) Âge : $\mu=54,4$ (ET=9,3) Langue : Anglais	Voyelle tenue /i/ (portion moyenne stable)	F1 et F2	London Speech Evaluation, échelle à 4 points (0-3 : normal-altération sévère) : intelligibilité (« impression auditive-perceptive de compréhensibilité »), articulation, note globale	Moyenne (ET) : <i>Perceptif</i> - London Speech Evaluation : NR <i>Acoustique</i> - F1 : H=315,9 Hz (170,7) ; F=353,8 Hz (78,3) - F2 : H=1782,6 (846,2) ; F=2111,5 (986,7)
A19. CONNAGHAN 17	III-2	21/24 88 % haute	N=15 (9 H, 6 F) Âge : $\mu=36$ (intervalle 22-59) Langue : Anglais américain	Voyelles hautes /i/, /ɪ/ et voyelle basse /æ/ dans des mots accentués et non accentués	- F1 et F2 - Distance euclidienne (ED) entre les centroïdes F1xF2 des voyelles de chaque paire (/æ-i/, /æ-ɪ/, /i-ɪ/), en guise de mesure de dispersion vocalique - Fonction de densité de probabilité (PDF) : probabilité relative que chaque échantillon vocalique provienne de la zone F1 × F2 de la voyelle cible	Pourcentage d'identification correcte pour chaque voyelle	Moyenne (ET) : <i>Perceptif</i> - Pourcentage d'identification correcte de voyelles : NR <i>Acoustique (moyenné sur toutes les voyelles)</i> <i>Mots accentués</i> : - F1(Hz) : 545,0 (208,4) ; F2(Hz) : 2178,8 (336,3) - ED(Hz) : 536,1 (129,6) - PDF : 0,99 (0,06) <i>Mots non accentués</i> : - F1 : 507,2 (169,9) ; F2 : 2068,0 (320,1) - ED : 476,9 (113,6) - PDF : 0,87 (0,30) Les données pour chaque voyelle sont également disponibles (Tableau 3, p.44).
A20. FLETCHER 17	III-2	22/24 92 % haute	N=17 (11 H, 6 F) Âge : $\mu=66$ (ET=NR) Langue :	Voyelles cardinales /a:/, /i:/, /o:/ dans des mots	- Aire de l'espace vocalique (VSA) - Ratio de centralisation des formants (FCR) :	Évaluations sur des échelles visuelles analogiques :	Moyenne (ET) : <i>Perceptif</i> - Jugement d'intelligibilité : 0,877 (0,110) - Jugement de précision de parole : 0,987 (0,198)

Référence	Design d'étude ¹	QualSyst ²	Population saine [N, Genre, Âge (années), Langue]	Échantillon de parole pour l'acoustique (phonème cible)	Paramètres acoustiques (Définitions)	Mesures perceptives	Données descriptives concernant les locuteurs sains
			Anglais Néozélandais		Les deux mesures ont été effectuées en Barks et en Hertz, au point médian de la voyelle et au point où il y avait le moins de mouvement dans les trajectoires des formants (= point flexible).	Groupe d'auditeurs 1 : Intelligibilité = Facilité à comprendre le locuteur (0-100 : facile-difficile) Groupe d'auditeurs 2 : Précision de la parole (0-100 : précis-imprécis)	<i>Acoustique</i> - VSA(Bark ²) flexible : H=10,91(2,64) ; F=13,87(2,64) - VSA(Bark ²) médian : H=7,76(2,16) ; F=10,79(1,68) - VSA(Hz ²) flexible : H=243,21(69,88) ; F=385,82(103,11) - VSA(Hz ²) médian : H=174,73(55,40) ; F=295,13(61,61) - FCR(Bark) flexible : H=1,19(0,07) ; F=1,19(0,03) - FCR(Bark) médian : H=1,28(0,07) ; F=1,25(0,03) - FCR(Hz) flexible : H=0,97(0,07) ; F=0,92(0,03) - FCR(Hz) médian : H=1,06(0,08) ; F=0,98(0,03)
A21. KIM 17	III-2	22/24 92 % haute	N=24 (14 H, 10 F) 12 Anglais américain Âge : médiane=59 (intervalle 49-85) 12 Coréen Âge : médiane=NR (intervalle 52-72)	Voyelles /a, i, u/ dans des mots	Espace vocalique acoustique (AVS) dérivé des fréquences de F1 et F2 au point temporel central de la voyelle :	Évaluation de l'intelligibilité sur une échelle de 10 points à intervalles égaux (1-10 : totalement inintelligible – complètement intelligible)	<i>Perceptif</i> - Jugement d'intelligibilité : NR <i>Acoustique</i> - AVS : Moyenne en log (ET) : • Locuteurs Anglais : 5,21 (0,09) • Locuteurs Coréen : 5,48 (0,19)
A22. MARTEL- SAUVAGEAU 17	III-2	23/24 96 % haute	N=8 (3 H, 5 F) Âge : « Apparié en âge +/- 2 ans » (Âges NR) Langue : Français Québécois	- Contextes semi-vocaliques (glides) : /w a, j a, ε j/ dans des mots - CVCV dans une phrase porteuse, avec les voyelles cibles /i, u, a/ et les consonnes /b, d, g/	- Semi-voyelles (glides) : pentes de F2 = changement global de fréquence divisé par la durée de la transition - CVCV : • Équations du locus (LE) : fonction de régression linéaire utilisant F2 en début et F2 au centre de la voyelle ; pour /b, d, g/ (k et c : constantes) • Distinctivité des LE : Distinctivité entre les LE des trois lieux d'articulation, mesurée en utilisant les paramètres constants (k, c) des équations comme dimensions d'un espace de	Intelligibilité globale évaluée sur une échelle visuelle analogique de 229 mm (0-229 : n'a rien compris – a tout compris)	Moyenne (ET) : <i>Perceptif</i> - Intelligibilité globale : NR (graphiques) <i>Acoustique</i> - Pentes de F2 (Hz/ms) : • /w a/ : 12,95(1,97) • /j a/ : -5,78(1,53) • /ε j/ : 5,03(1,06) - Équations du locus : • /b/ : pente=0,67, ordonnée=506,6 • /d/ : pente=0,29, ordonnée=1331,2 • /g/ : pente=1,04, ordonnée=134,2 - Distinctivité des LE : 0,05

locus triangulaire. L'aire de cet espace est ensuite calculée en utilisant les distances euclidiennes entre les coordonnées de /b/-/d- /g/

Note : H = hommes ; F = femmes ; NR = non renseigné ; ET = écart-type

¹Les designs d'études sont présentés selon la hiérarchie du NHMRC [NHMRC 99] : Niveau I Revues systématiques ; Niveau II Essais contrôlés randomisés ; Niveau III-1 Essais contrôlés pseudorandomisés ; Niveau III-2 Études comparatives avec groupe contrôle parallèle et allocation non randomisée (études de cohorte, études cas-témoins, ou séries chronologiques interrompues avec un groupe contrôle) ; Niveau III-3 Études comparatives sans groupe contrôle parallèle, avec groupe contrôle historique, deux ou plusieurs études à un seul bras, ou séries chronologiques interrompues sans groupe contrôle parallèle ; Niveau IV Séries de cas.

²Les lignes directrices d'interprétation du score de qualité méthodologique QualSyst [KMET 04] sont : haute >80 % ; bonne 60–79 % ; appropriée 50–59 % ; mauvaise <50 %

Annexe E – Définitions et formules (le cas échéant) des mesures acoustiques utilisées dans les études de la revue systématique

Mesures sur les voyelles

Mesures formantiques stationnaires

- (1) Aire de l'espace vocalique (« Vowel Space Area ») [A. R. FLETCHER 17] : les valeurs des premier et deuxième formants des voyelles cardinales de la langue étudiée sont utilisées comme coordonnées dans un espace F1/F2 pour construire un triangle ou un quadrilatère de voyelles. L'aire du triangle ou du quadrilatère résultant est ensuite calculée à l'aide de formules classiques telles que :

$$Hz^2 = 0,5 \times |F1[v_1] \times (F2[v_2] - F2[v_3]) + F1[v_3] \times (F2[v_1] - F2[v_2]) + F1[v_2] \times (F2[v_3] - F2[v_1])|$$

(où v1, v2 et v3 sont les voyelles cardinales)

- (2) Aire de l'espace vocalique acoustico-articulatoire (« Articulatory–Acoustic Vowel Space ») [WHITFIELD 14, 17] : « Cet espace est calculé comme la racine carrée de la variance généralisée de tous les formants de voyelles échantillonnés dans le graphe de coordonnées F1-F2. La variance généralisée de l'AAVS est calculée comme le produit de la variance des données de F1, de la variance des données de F2 et de la portion de la variance non partagée entre elles. La racine carrée de la variance généralisée fournit une mesure de la variabilité des formants qui est l'équivalent d'un écart-type bivarié dans l'espace F1-F2. Par conséquent, une augmentation de l'étendue ou de la dispersion des valeurs F1 ou F2 dans un énoncé produirait une AAVS plus grande » (traduit de [WHITFIELD 17]).
- (3) Mesures stationnaires de F1 et F2 : les premier et second formants sont extraits, généralement au point médian temporel. Ils peuvent ensuite être comparés, par exemple entre voyelles, ou entre groupes de locuteurs.
- (4) Intervalles de F1 et F2 : soustraction de la valeur F1/F2 la plus basse de la valeur la plus élevée.
- (5) Différence F0-F1 [BUNTON 01] : distance euclidienne entre la fréquence fondamentale et le premier formant.
- (6) Différence F1-F2 [BUNTON 01; HAZAN 04] : distance euclidienne entre le premier et le second formants.
- (7) Distance euclidienne entre paires de voyelles dans l'espace F1xF2 [CONNAGHAN 17; NEEL 08] : distance euclidienne entre les centroïdes F1xF2 des voyelles de chaque paire de voyelles, en guise de mesure de la dispersion vocalique.

- (8) Indice d'articulation vocalique (« Vowel Articulation Index », VAI) [SKODDA 12] : « paramètre de substitution des fréquences des premier et second formants ($F1$ et $F2$) des trois voyelles cardinales /a/, /i/ et /u/ » (traduit de l'anglais) :

$$VAI = \frac{F2[i] + F1[a]}{F1[i] + F1[u] + F2[u] + F2[a]}$$

- (9) Ratio de centralisation des formants (« Formant Centralization Ratio », FCR) [A. R. FLETCHER 17] : valeur réciproque du VAI [SKODDA 12], une mesure qui « pondère les formants qui sont susceptibles d'augmenter en raison de la centralisation des voyelles par rapport aux formants qui sont susceptibles de diminuer » (traduit de l'anglais) :

$$FCR = \frac{F2[o:] + F2[e:] + F1[i:] + F1[e:]}{F2[i:] + F1[o:]}$$

(utilisation des voyelles cardinales de l'anglais néo-zélandais)

- (10) Fonction de densité de probabilité (« Probability Density Function ») [CONNAGHAN 17] : probabilité relative qu'un échantillon de voyelle provienne de la zone de la voyelle cible dans l'espace $F1 \times F2$
- (11) Fréquence de début [WEISMER 92] : la fréquence de démarrage (en Hertz) du segment de transition (cf. (18))

Mesures formantiques dynamiques

- (12) Changement spectral (« spectral change ») [FERGUSON 07] : « la somme, en Barks, du déplacement absolu de la fréquence des formants pour $F1$ et $F2$. Ainsi, λ est calculé comme suit :

$$\lambda (\text{Barks}) = |F1_{80} - F1_{20}| + |F2_{80} - F2_{20}|$$

où $F1_{20}$, $F1_{80}$, $F2_{20}$, et $F2_{80}$ sont les valeurs de $F1$ et $F2$ en Barks à 20 % et 80 % de la durée de la voyelle. » (traduit de l'anglais)

- (13) Angle spectral (« spectral angle ») [FERGUSON 07] : l'angle spectral (ou inclinaison, « tilt ») est calculé pour chaque voyelle en comparant $F1$ et $F2$ à 80 % de la durée de la voyelle à la fréquence mesurée à 20 % de la durée de la voyelle. L'angle θ en radians pour chaque formant n est d'abord calculé comme l'arctangente de la différence entre la fréquence du formant à 80 % et à 20 % de la durée de la voyelle, divisée par la durée séparant ces deux points, exprimée en décisecondes. L'angle spectral est la somme en radians des valeurs absolues des deux angles des formants :

$$\theta (\text{radians}) = \left| \arctan\left(\frac{F1_{80} - F1_{20}}{\frac{time_{80} - time_{20}}{100}}\right) \right| + \left| \arctan\left(\frac{F2_{80} - F2_{20}}{\frac{time_{80} - time_{20}}{100}}\right) \right|$$

où $F1_{20}$, $F1_{80}$, $F2_{20}$, et $F2_{80}$ sont les valeurs de $F1$ et $F2$ en Barks at 20 % et 80 % de la durée de la voyelle.

- (14) Mouvement formantique moyen sur l'ensemble des voyelles [NEEL 08] : pour chaque voyelle, la somme de « la distance euclidienne dans l'espace $F1 \times F2$ en Barks entre le début de la voyelle (20 % de la durée de la voyelle) et l'état stable [...] et la distance euclidienne entre l'état stable de la voyelle et la finale (80 % de la durée de la voyelle) »

est calculée. La moyenne de ces distances est ensuite calculée sur les différentes voyelles de chaque locuteur. » (traduit de l'anglais)

$$\sqrt{(F1_{50} - F1_{20})^2 + (F2_{50} - F2_{20})^2} + \sqrt{(F1_{80} - F1_{50})^2 + (F2_{80} - F2_{50})^2}$$

où $F1_{20}$, $F1_{50}$, $F1_{80}$, $F2_{20}$, $F2_{50}$ et $F2_{80}$ sont les valeurs en Barks de F1 et F2 à 20 %, 50 % et 80 % de la durée de la voyelle.

- (15) Rapport dynamique (« dynamic ratio ») [NEEL 08] : un indicateur composite basé sur des mesures dynamiques ; distinctivité en Barks entre les voyelles avec des trajectoires dynamiques et statiques ; distance euclidienne moyenne (des débuts de voyelles aux états stables aux finales dans l'espace $F1 \times F2$ en Barks, voir 13) couverte par les trois voyelles les plus dynamiques (/æ, ʌ, ʊ/) divisée par la distance couverte par les trois voyelles les plus statiques (/i, ε, u/)
- (16) Longueur de vecteur (« vector length », VL) [FERGUSON 14] : la distance euclidienne dans l'espace $F1 \times F2$ entre le début de la voyelle (20 % de la durée de la voyelle) et la finale (80 % de la durée de la voyelle) :

$$VL (Barks) = \sqrt{(F1_{80} - F1_{20})^2 + (F2_{80} - F2_{20})^2}$$

où $F1_{20}$, $F1_{80}$, $F2_{20}$ et $F2_{80}$ sont les valeurs en Barks de F1 et F2 à 20 % et 80 % de la durée de la voyelle.

- (17) Longueur de trajectoire (« trajectory length », TL) [FERGUSON 14] : la somme en Barks des quatre distances euclidiennes entre les sections de voyelles 20 %-35 %, 35 %-50 %, 50 %-65 % et 65 %-80 % :

$$vowel\ section\ length\ VSL_n (Barks) = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2}$$

$$trajectory\ length\ TL (Barks) = \sum_{n=1}^4 VSL_n$$

- (18) Étendue de la transition (« transition extent ») [WEISMER 92] : le degré de changement de fréquence (en Hertz) le long du segment de transition d'une trajectoire. Le début et la fin de ce segment sont « la première et la dernière paire temps-fréquence, respectivement, pour lesquelles l'incrément ultérieur de 20 ms [est] associé à un changement d'au moins 20 Hz. » (traduit de l'anglais)
- (19) Pente ou vitesse de transition (« transition rate or slope ») [WEISMER 92] : la division de l'étendue de la transition (en Hertz) par la durée (en ms) du segment de transition

Mesures sur les semi-voyelles

- (1) Pente de F2 (« F2 slope ») [A. R. FLETCHER 17] : le déplacement global de fréquence en Hertz (étendue de la transition) dans une semi-voyelle, divisé par la durée de la transition (en ms), en guise de mesure de la vitesse de modification du tractus phonatoire.

Mesures sur les consonnes

- (1) Moments spectraux : dans l'analyse des moments spectraux, le spectre consonantique (fricatives ou plosives) est considéré comme une distribution statistique, qui peut être décrite par quatre mesures : le premier moment (centre de gravité) est la « fréquence qui divise le spectre en deux moitiés, de telle sorte que la quantité d'énergie dans les régions de haute fréquence est égale à celle de la région de basse fréquence » (traduit de [YOON 15]). Le deuxième moment (écart-type) mesure la dispersion de l'énergie spectrale autour de ce centre de gravité. Le troisième moment spectral (« skewness ») fait référence à l'asymétrie de la distribution de l'énergie par rapport à la moyenne. Par exemple, une asymétrie nulle est mesurée dans les distributions symétriques, tandis que les valeurs positives indiquent la queue droite (valeurs au-dessus du mode) s'étend plus loin que la queue gauche [JONGMAN 00]. Le quatrième moment (« kurtosis ») mesure à quel point la distribution est pointue.
- (2) Pic spectral fricatif du son /s/ [KATZ 91] : déterminé à partir des spectres LPC à court terme à 30 ms avant la finale de la fricative, anticipant le F2 des voyelles /i/ and /u/.
- (3) Fréquence limite supérieure (« Upper Boundary Frequency ») [HOHOFF 03] : la fréquence la plus élevée du bruit de friction, la plage la plus grise dans le spectrogramme à large bande.
- (4) Pente spectrale (« spectral slope ») de la fricative /x/ [DE BRUIJN 09] : mesure du déclin de l'énergie spectrale des basses aux hautes fréquences du spectre, calculée par régression linéaire [PEETERS 04].
- (5) Fréquence de F1 en finale [FLEGE 92] : fréquence du premier formant mesurée dans les 45 dernières ms des voyelles /i, ɪ, ε, æ/ avant les plosives /t/ et /d/.
- (6) Équations du locus (« locus equations ») [MARTEL-SAUVAGEAU 17] : une « métrique acoustique alternative pour caractériser les caractéristiques de transitions segmentales » (traduit de l'anglais), à savoir la fonction de régression linéaire utilisant F2 au début de la voyelle et F2 au milieu :
- $$F2_{onset} = k \times F2_{mid} + c$$
- (7) Distinctivité des équations de locus (« locus equation distinctiveness ») [MARTEL-SAUVAGEAU 17] : la distinctivité entre les équations de locus correspondant aux trois lieux d'articulation /b, d, g/, mesurée en utilisant les paramètres constants k et c des équations comme dimensions d'un espace de locus triangulaire. L'aire de cet espace est ensuite calculée en utilisant les distances euclidiennes entre les coordonnées /b-/d-/g/.

Annexe F – Tentative de comparaison croisée des résultats acoustiques de la revue systématique

Les mesures acoustiques qui peuvent être comparées entre les études sont principalement vocaliques, à l'exception de la fréquence centroïde (premier moment spectral) sur la fricative /s/ utilisée dans les études [KATZ 91] et [NEEL 15]. Dans les autres études portant sur les consonnes, les données incomplètes concernant les groupes témoins, les diverses mesures et les différentes méthodologies ne permettent pas une analyse comparative.

Une tentative de comparaison des résultats de mesures acoustiques similaires utilisées dans les différentes études est présentée dans le tableau C.1. Nous pouvons observer que même si plusieurs études utilisent la même mesure, la population étudiée, l'échantillon phonémique, la méthode d'extraction et de calcul et la présentation des résultats sont très différents et parfois non communiqués (selon l'objectif de chaque étude), ce qui rend difficile la mise en relation des valeurs obtenues. Si nous examinons les mesures stationnaires de premier et second formant, par exemple, l'étude [NEEL 08] utilise l'échelle de Bark, tandis que les études [DE BRUIJN 09], [DWIVEDI 16] et [CONNAGHAN 17] utilisent le Hertz. La méthode d'extraction des formants n'est rapportée que dans l'étude [DWIVEDI 16]. Cependant, les valeurs des formants peuvent différer en fonction de la méthode d'extraction (p. ex. codage prédictif linéaire, transformée de Fourier rapide, analyse cepstrale) et des paramètres d'extraction/analyse (tels que le type de fenêtre d'analyse, la taille de la trame, le pas d'avancement et les paramètres spécifiques à chaque méthode) [DERDEMEZIS 16; ERINGIS 14]. Dès lors, ce manque d'informations ne permet pas de répliquer la méthodologie de l'étude ni de réaliser des analyses comparatives. De plus, l'étude [DE BRUIJN 09] a été réalisée sur des échantillons néerlandais, tandis que les études [DWIVEDI 16] et [CONNAGHAN 17] ont utilisé des échantillons anglais, ce qui peut avoir un impact sur la prononciation des voyelles. Enfin, la population étudiée dans l'étude [CONNAGHAN 17] est presque 20 ans plus jeune que celle des études [DE BRUIJN 09] et [DWIVEDI 16].

En outre, pour les mesures de l'aire de l'espace vocalique (VSA), nous constatons une incohérence dans les unités utilisées par les différents auteurs : quatre études utilisent des unités de Hertz au carré, mais les valeurs sont très dissemblables. L'étude [CONNAGHAN 17] donne des valeurs inférieures à 1Hz^2 , alors que l'étude [NEEL 15] montre des valeurs supérieures à 30000Hz^2 , et les études [A. R. FLETCHER 17] et [Y. KIM 17] rapportent des valeurs entre 150 et 300Hz^2 . Pour les valeurs de l'aire de l'espace vocalique acoustico-articulatoire (AAVS), malgré le fait que l'auteur principal soit le même dans les deux papiers, l'étude [WHITFIELD 17] utilise le kHz, alors que l'étude [WHITFIELD 14] utilise le kHz^2 . Pourtant, dans les deux études, des valeurs entre 25 et 65 sont rapportées. Ces observations soulignent la nécessité d'être précis dans la description des mesures acoustiques et dans le rapport des résultats.

Tableau F.1 – Mesures acoustiques utilisées dans différentes études et leurs résultats à des fins de comparaison

Mesure	Étude	Résultat	Unité Échantillon	Extraction	Langue	Âge	N
F1 et F2	[NEEL 08]	F1 : H=5,04(0,20) ; F=5,88(0,30)	Bark /i, ɪ, e, ε, æ,	?	Anglais américain (Michigan/Upper Midwest dialect)	NR	N=93 (45 H, 48 F)

		F2 : H=13,05(0,37); F=14,70(0,53)		α, ʌ, o, u, u/ (poolés)				
[DE BRUIJN 09]	F1 : 296(49)	Hz	/i/	?	Néerlandais	Apparié au groupe d'étude : μ=53,8 (ET=8,7)	N=18 (Apparié au groupe d'étude : H=55 %, F=45 %)	
	F2 : 2325(248)							
[DWIVEDI 16]	F1 : H=315,9(170,7); F=353,8(78,3)	Hz	/i/	LPC	Anglais	μ=54,4 (ET=9,3)	N=51 (32 H, 19 F)	
	F2 : H=1782,6(846,2); F=2111,5(986,7)							
[CONNAGHAN 17]	F1[i] : accentué= 345,7(47,2) non accentué= 350,8(50,3)	Hz	/i, ɪ, æ/	? (fenêtre = 30 ms)	Anglais américain	μ=36 (intervalle 22-59)	N=15 (9 H, 6 F)	
	F2[i] : accentué= 2508,4(243,0) non accentué= 2377,3(239,5)							
Intervalles F1 et F2	[NEEL 08] Intervalle F1 : H=3,83(0,59); F=4,32(0,80)	Bark	/i, ɪ, e, ε, æ, α, ʌ, o, u, u/ (poolés)	?	Anglais américain (Michigan/Upper Midwest dialect)	NR	N=93 (45 H, 48 F)	
	Intervalle F2 : H=9,37(1,04); F=11,15(1,06)							
	[NEEL 15] Intervalle F1 : 448,9 (83,9 ; 286-532)	Hz	/i, a, u/	?	Anglais	intervalle 52-69	N=12 (4 H, 8 F)	
	Intervalle F2 : 1552,8 (197,8 ; 1309-1899)							
VSA	[NEEL 08] H=18,57 (4,13); F=25,07 (6,55)	?	quadrilatère /i, a, u, æ/	?	Anglais américain (Michigan/Upper Midwest dialect)	NR	N=93 (45 H, 48 F)	
	[WHITFIELD 17] H=200,81(23,65); F=577,74(94,11)	kHz	quadrilatère /i, a, u, æ/	LPC (méthode Burg, taille de fenêtre = 50 ms ; pas = 1 ms)	Anglais américain standard	H : μ=24,40 ; intervalle 20-36 F : μ=24,30 ; intervalle 18-29	N=10 (5 H, 5 F)	
	[DE BRUIJN 09] 0,213(0,11)	Hz ²	triangle /i, a, u/	?	Néerlandais	Apparié au groupe d'étude : μ=53,8 (ET=8,7)	N=18 (Apparié au groupe d'étude : H=55 %, F=45 %)	
	[NEEL 15] 334,262 (98,557; 192,980-526,903)	Hz ²	triangle /i, a, u/	?	Anglais	intervalle 52-69	N=12 (4 H, 8 F)	
	[CONNAGHAN 17] Point flexible : H=10,91(2,64); F=13,87(2,64)	Bark ²	triangle /a:, i:, o:/	LPC (méthode Burg, taille de fenêtre = 25 ms, pas = 6,25 ms)	Anglais néo-Zélandais	μ=66	N=17 (11 H, 6 F)	
	Point médian temporel : H=7,76(2,16); F=10,79(1,68)							
	Point flexible : H=243,21(69,88); F=385,82(103,11)	Hz ²						

		Point médian temporel :					
		H=174,73(55,40) ; F=295,13(61,61)					
[Y. KIM 17]	Locuteurs anglais : 5,21 (0,09) Locuteurs coréens : 5,48 (0,19)	Log (Hz ²)	triangle /i, a, u/	?	Anglais américain Coréen	Anglais : médiane=59 (intervalle 49-85) Coréen : médiane =NR (intervalle 52-72)	N=24 (14 H, 10 F) 12 Anglais 12 Coréens
AAVS	[WHITFIELD 17] Conversationnel : H=27,98(5,06) ; F=68,83(6,86) Clair : H=35,37(8,13) ; F=93,81(20,21)	kHz	quadrilatère /i, a, u, æ/	LPC (méthode Burg, taille de fenêtre = 50 ms, pas = 1 ms)	Anglais américain standard	H : μ =24,40 ; intervalle 20-36 F : μ =24,30 ; intervalle 18-29	N=10 (5 H, 5 F)
	[WHITFIELD 14] H=38,45 (5,20) ; F= 64,59 (9,77)	kHz ²	Tous les segments voisins du 1er paragraphe du Rainbow Passage	LPC (méthode Burg, taille de fenêtre = 50 ms, pas = 1 ms)	Anglais américain	H : μ =65,8 ; intervalle 57-73 F : μ =71,8 ; intervalle 58-81	N=10 (5 H, 5 F)
Fréquence centroïde	[KATZ 91] À 30 ms avant la finale de la fricative : /si/=5524 ; /su/=5134 À 100 ms avant la finale de la fricative : /si/=6806 ; /su/=6182	Hz	/s/ dans /s u/ et /si/	DFT (taille de fenêtre = 20 ms ; 30 ms et 100 ms avant la finale de la fricative)	Anglais	μ =32 (ET=6,7 ; intervalle 26-45)	N=10 (5 H, 5 F)
	[NEEL 15] 6962,6 (1282,6 ; 4700-8756)	Hz	/s/ initial dans les mots « sip, seep, see »	? (taille de fenêtre = 20 ms ; au centre de la fricative)	Anglais	intervalle 52-69	N=12 (4 H, 8 F)

Note : H = hommes ; F = femmes ; F1/F2 = premier et second formants ; VSA = aire de l'espace vocalique (vowel space area) ; AAVS = aire de l'espace vocalique acoustico-articulatoire (articulatory-acoustic vowel space) ; LPC = codage linéaire prédictif (linear predictive coding) ; DFT = transformée de Fourier discrète (discrete Fourier transform) ; NR = non renseigné ; μ = moyenne ; ET = écart-type

Annexe G – Ensemble minimal de tâches et recommandations pour les enregistrements dans le cadre de l'évaluation des troubles de la parole chez l'adulte

Niveaux CIF	Dimensions d'analyse	Tâches	Mesures
Structures et fonctions du corps	Information contextuelle	Conversation guidée avec le patient et les autres parties prenantes (pairs, professionnels de la santé...)	Anamnèse du patient (diagnostic et antécédents médicaux, déficits associés, âge, contexte de vie/travail, partenaires de communication, principales plaintes, besoins et priorités, facilitateurs et obstacles à la communication...)
	Sensibilité et motricité	Tâches de motricité et de sensibilité orofaciales verbales et non verbales. Respiration au repos et pendant la parole Posture	Force, vitesse, amplitude, précision et coordination des mouvements de la langue, des lèvres, de la mâchoire, des joues, du velum et du visage Mesures aérodynamiques Coordination pneumophonique
	Phonation	Voyelles tenues et échantillons de parole continue	Mesure objective de la qualité vocale, niveau de pression sonore et hauteur du son, temps maximum phonatoire Évaluation subjective de la qualité globale de la voix et des paramètres vocaux (p. ex. GRBAS-I, CAPE-V)
	Articulation	Répétition de phonèmes/syllabes Diadococinésies (pataka) Répétition de mots multisyllabiques simples ou complexes Parole automatique	Inventaire phonémique Planification et programmation motrice vs exécution
Activités	Intelligibilité	Répétition ou lecture de : - pseudomots - paires minimales de mots pour tester des contrastes phonétiques spécifiques (p. ex. TPI en français) - phrases sémantiquement non prédictibles	Pourcentage de stimuli corrects à partir de la transcription (phonèmes/syllabes/pseudomots, mots) N.B. Idéalement, transcription par un collègue pour éviter la familiarité avec les stimuli et avec la parole du patient ; larges collections/bases de pseudomots pour éviter la mémorisation, avec des similarités/paires minimales
	Compréhensibilité	Tâches de compréhensibilité de phrases (p. ex. tâche	Évaluation de la compréhensibilité, de la sévérité, du caractère naturel et de la nasalité de

		de vérification de phrases, sélection d'images... ; idéalement, stimuli inconnus de l'auditeur) Lecture de textes Parole (semi)spontanée	la parole (voir Darley, 69a pour d'autres critères d'évaluation) Détérioration dans le temps/fatigabilité
	Suprasegmental	Tâches spécifiques liées à la prosodie Chunking/syntaxe (p. ex. Profiling Elements of Prosody in Speech-Communication, PEPS-C)	Intonation and intentions : utilisation de phrases déclaratives/interrogatives/exclamatives/impératives Expression des émotions Expression de l'accentuation contrastive/Focus Découpage syntaxique Rythme de la parole
Participation	Impact psychosocial	Speech Handicap Index, Phonation Handicap Index, Dysarthria Impact Profile ou autres questionnaires spécifiques (p. ex. Functional Assessment of Cancer Therapy Head & Neck)	Impact global et dimensions spécifiques

Conditions d'enregistrement

[CHIAL 03; DELIYSKI 05; R. R. PATEL 18; PLICHTA 02; ŠRÁMKOVÁ 15; ŠVEC 10, 18; WINHOLTZ 97]

À des fins de recherche, tous les paramètres suivants doivent être tenus constants ; à des fins cliniques, les paramètres doivent être tenus constants pour chaque patient (p. ex. le niveau d'entrée peut parfois nécessiter une adaptation en fonction du profil du patient, comme pour les chanteurs ou pour les patients ayant un niveau d'intensité vocale très bas, mais il doit ensuite être réglé de manière identique pour chaque enregistrement suivant de ces mêmes patients).

Microphone :

- microphone de type serre-tête pour assurer une distance constante entre le microphone et la bouche et maximiser le rapport signal/bruit
- microphone cardioïde (directionnel) pour limiter l'effet du bruit ambiant et maximiser le rapport signal/bruit (à des fins de recherche, microphone omnidirectionnel pour une meilleure précision en termes de niveau de pression sonore et de fréquence)

- type de transducteur (conversion de la pression acoustique en signal électrique) : condensateur
- réponse en fréquence : gamme large et plate (c.-à-d. variation de <2 dB), idéalement 10 Hz-20 000 Hz ; au moins 50 Hz-8000 Hz pour capturer tout le spectre de la parole/voix
- plage dynamique : 15-130 dB ; le niveau de bruit interne du microphone (voir les spécifications du fabricant) doit être inférieur d'au moins 15 dB à la phonation la plus douce, c.-à-d. un niveau de bruit de max. 30 dBA lorsque le microphone est utilisé à une distance de 5 cm de la bouche
- utilisez des embouts en mousse pour éviter les turbulences sur les plosives, les fricatives ou le bruit du flux d'air
- 4–10 cm de la bouche du patient à un angle de 45-90° (essayez de toujours utiliser la même distance, car cela a un impact important sur les mesures d'intensité) ; la distance optimale pour une réponse en fréquence plate (nécessaire pour les mesures spectrales) devrait être indiquée dans les spécifications du microphone

Préamplificateur :

Le préamplificateur amplifie l'amplitude/le voltage faible du microphone :

- utilisez un préamplificateur qui a la même réponse en fréquence et la même plage dynamique que le microphone
- vérifiez l'impédance de terminaison minimale (« minimum terminating impedance ») dans les spécifications du microphone
- le microphone à condensateur doit être alimenté par l'intermédiaire de l'« alimentation fantôme » (« phantom power supply »), généralement 48 V
- niveau d'entrée : réglez-le de manière à ce que les niveaux vocaux maximums soient proches mais inférieurs aux niveaux maximums enregistrables du dispositif (évitez la saturation)

Enregistrement numérique :

- fréquence d'échantillonnage minimale : 44,1 kHz
- résolution minimale en amplitude : 16 bits
- format de fichier : fichier .wav (pas de compression ; ne PAS utiliser .mp3)
- mode d'enregistrement : mono

Suggestions générales :

- Essayez de vous en tenir à un seul logiciel d'enregistrement/d'analyse : Praat (gratuit), Audacity (gratuit), Computerized Speech Lab (commercial)...
- Calibrez le microphone pour les mesures d'intensité : enregistrez un son pur stable provenant d'un haut-parleur ou une voyelle tenue avec le microphone de type serre-tête et utilisez simultanément un sonomètre à la même distance ; comparez l'intensité mesurée par le microphone (par exemple en utilisant Praat) avec la valeur du sonomètre, qui vous donne la valeur de référence. Si, par exemple, Praat indique 65 dB SPL alors que le sonomètre indique 60 dB SPL, pour toutes les mesures

ultérieures effectuées avec exactement la même distance et les mêmes paramètres, soustrayez 5 dB de la mesure de l'ordinateur pour obtenir le niveau de pression acoustique « réel ».

- Effectuez les enregistrements dans une pièce calme. Si aucune pièce insonorisée n'est disponible, mesurez le bruit ambiant à l'aide d'un sonomètre ; en considérant que le rapport signal/bruit doit être supérieur à 42 dB pour des mesures acoustiques fiables (bien que ≥ 30 dB soit acceptable), et que l'intensité moyenne de la voix conversationnelle est d'environ 60 dB, le bruit ambiant doit être inférieur à 18 dB pour permettre un rapport signal/bruit optimal.
- Évitez les sources de bruit telles que les ventilateurs/climatisations, les souffleries d'ordinateurs, les bourdonnements de néons... mais aussi les smartphones et autres accessoires connectés comme les montres connectées, qui peuvent provoquer des interférences sonores.

Annexe H – Questionnaire destiné à des patients évalués pour un trouble de la parole



Institut de Recherche
en Informatique de Toulouse
UMR 5515 - IRIT - UFR 5722



UNIVERSITÉ
TOULOUSE III
PAUL SABATIER

Hôpitaux de Toulouse



État des lieux de l'évaluation orthophonique des troubles de la production de la parole chez l'adulte : Point de vue du patient

Questionnaire à destination du patient

Durée de complétion estimée : 15 minutes

Informations personnelles

1. Vous êtes : un homme
 une femme
2. Quel est votre âge ?
3. Dans quel département habitez-vous ?
4. Vous vivez : en ville
 en banlieue
 à la campagne

Votre maladie

5. Vos difficultés pour parler sont dues à :
 - une attaque cérébrale (AVC)
 - une tumeur au cerveau
 - une maladie de Parkinson
 - une sclérose latérale amyotrophique
 - un cancer de la bouche et/ou de la gorge
 - une maladie génétique
 - une surdité
 - un autre problème :
6. Le début de cette maladie date :
 - d'il y a moins de 6 mois
 - d'entre 6 mois et 1 an
 - d'entre 1 et 2 ans
 - d'entre 2 et 5 ans
 - d'entre 5 et 10 ans
 - de plus de 10 ans

Votre trouble de la parole

Un « **trouble de la parole** » est une difficulté à articuler les sons et à produire une parole qui soit compréhensible par la personne à qui l'on parle.
Il s'agit d'un problème d'articulation ou de prononciation, mais il ne s'agit pas d'un problème pour trouver ses mots, ou d'un bégaiement.

7. Les premiers signes de votre **trouble de la parole** sont apparus il y a :
- moins de 6 mois
 - entre 6 mois et 1 an
 - entre 1 et 2 ans
 - entre 2 et 5 ans
 - entre 5 et 10 ans
 - plus de 10 ans

Nous appellerons « **bilan** » l'évaluation de votre parole par l'orthophoniste.
Le **premier bilan** est l'évaluation de votre parole par l'orthophoniste la première fois que vous l'avez vu(e).
Le **dernier bilan** est la dernière évaluation de votre parole par un(e) orthophoniste.

Votre premier suivi en orthophonie par rapport au trouble de la parole

8. La **première fois** que vous avez vu un orthophoniste par rapport à votre problème de parole, c'était :
- votre orthophoniste actuel(le)
 - un(e) autre orthophoniste qui travaillait :
 - en cabinet
 - à l'hôpital
 - dans un autre établissement
- Préciser si possible :
9. Le **premier bilan** de votre parole date d'il y a :
- moins de 6 mois
 - entre 6 mois et 1 an
 - entre 1 et 2 ans
 - entre 2 et 5 ans
 - entre 5 et 10 ans
 - plus de 10 ans
10. Est-ce que le dernier bilan de parole était aussi le premier ?
- Oui, il s'agissait de mon premier bilan de parole
 - Non, j'en ai eu un autre avant
 - Non, j'en ai eu plusieurs autres avant. Combien ?

11. Avant le premier bilan de parole, aviez-vous déjà été suivi en orthophonie ?

- Oui
 Non

Si oui, pour quelle raison ?

Votre suivi actuel en orthophonie par rapport au trouble de la parole

12. Actuellement, vous êtes suivi par :

- un(e) orthophoniste en cabinet
 un(e) orthophoniste à l'hôpital
 un(e) orthophoniste dans un autre établissement

Préciser si possible :

13. Qu'est-ce qui vous a amené à consulter cet orthophoniste en particulier ?

- Je connaissais déjà personnellement cet orthophoniste
 La proximité vis-à-vis de mon domicile
 Le bouche-à-oreille (recommandation par des amis, la famille...)
 La recommandation par un autre professionnel
 Le hasard
 Autre :

14. L'orthophoniste qui a effectué votre **dernier** bilan de parole exerçait :

- en ville
 en banlieue
 à la campagne

15. Le **dernier** bilan de votre parole date d'il y a :

- moins de 6 mois
 entre 6 mois et 1 an
 entre 1 et 2 ans
 entre 2 et 5 ans
 entre 5 et 10 ans
 plus de 10 ans

16. Combien de temps a duré votre **dernier** bilan de parole ?

- Moins d'un quart d'heure
 Entre 15 et 30 minutes
 Entre 30 et 60 minutes
 Plus d'une heure

17. Avez-vous eu l'impression que votre **dernier** bilan de parole avec l'orthophoniste vous avait fatigué ?

- Oui, beaucoup
 Oui, un peu
 Non, pas trop
 Non, pas du tout

18. L'orthophoniste vous a-t-il expliqué pourquoi il vous proposait les exercices que vous avez réalisés lors du bilan ?
- Non, pas du tout
 - Non, pas beaucoup, j'ai eu très peu d'informations / seulement parce que j'ai posé des questions
 - Oui, mais j'aurais aimé davantage d'informations
 - Oui, les explications étaient très claires
19. Avez-vous eu l'impression que les exercices proposés lors du **dernier** bilan étaient adaptés à votre problème et à vos besoins ?
- Oui, ils étaient tout-à-fait adaptés à mon cas
 - Non, il m'a semblé que le bilan était général et pas ou peu adapté à mon cas
20. Lors du **dernier** bilan, l'orthophoniste a-t-il enregistré votre parole ?
- Oui
 - Non
- Si oui, vous a-t-il dit pourquoi ?
- Oui
 - Non
21. L'orthophoniste vous a-t-il demandé d'évaluer vous-même votre parole ?
- Oui
 - Non
- Si oui, comment ?
- En plaçant un trait ou une croix sur une ligne
 - En donnant un nombre, par exemple entre 0 et 10
 - Autre :
22. Les conséquences de mon problème de parole (pour demander des choses aux gens, pour discuter avec eux, pour faire des activités de groupe,...) ont été prises en compte lors du bilan :
- Tout-à-fait d'accord : *Les conséquences de mon problème de parole ont été **complètement** prises en compte lors du bilan.*
 - Assez d'accord : *Les conséquences de mon problème de parole ont été **un peu** prises en compte lors du bilan.*
 - Pas trop d'accord : *Les conséquences de mon problème de parole n'ont **pas trop** été prises en compte lors du bilan.*
 - Pas du tout d'accord : *Les conséquences de mon problème de parole n'ont **pas du tout** été prises en compte lors du bilan.*

23. Dans le **dernier** bilan, l'orthophoniste vous a-t-il/elle demandé si vos problèmes pour parler :

Vous agacent :

- Oui, ça m'a été demandé
 Non

Vous sapent le moral :

- Oui, ça m'a été demandé
 Non

Vous empêchent de faire des choses que vous aimeriez faire :

- Oui, ça m'a été demandé
 Non

Font que vous évitez de discuter ou de passer du temps avec des gens :

- Oui, ça m'a été demandé
 Non

Vous gênent pour travailler et gagner de l'argent :

- Oui, ça m'a été demandé
 Non

24. Vous a-t-on déjà demandé de remplir un questionnaire concernant vos problèmes de parole et leurs conséquences ?

- Oui, à chaque bilan
 Oui, uniquement lors du premier bilan
 Oui, uniquement lors du dernier bilan
 Non, jamais

25. J'aimerais que les conséquences de mon problème de parole (pour demander des choses aux gens, pour discuter avec eux, pour faire des activités de groupe,...) soient davantage prises en compte lors du bilan :

- Tout-à-fait d'accord
 Assez d'accord
 Pas trop d'accord
 Pas du tout d'accord

26. Avez-vous eu l'impression que le dernier bilan a permis de prendre en compte tout ce qui vous gênait dans votre trouble de la parole ?

- Oui
 Non

Si non, quels sont les éléments qui n'ont pas été pris en compte dans l'évaluation de votre parole ?

.....

27. Les résultats du bilan vous ont été communiqués de manière facilement compréhensible :

- Oui
 Non

Annexe I – Membres du groupe de travail pour la création d'un nouveau texte pour l'évaluation de la parole et de la voix

- Corine Astésano : Enseignant-Chercheur, Maître de Conférences - HDR en sciences du langage à l'Université Toulouse – Jean Jaurès ; domaines de recherche : prosodie, phonétique expérimentale, neuropsychologie du langage et de la musique, parole pathologique ;
- Liziane Bouvier : orthophoniste, doctorante à l'Université Laval, Québec ; domaines de recherche : évaluation et intervention en apraxie de la parole en contexte neurodégénératif ;
- Véronique Delvaux : phonéticienne, Chargée de cours à l'Université de Mons et à l'Université Libre de Bruxelles, Chercheur qualifié FNRS à l'UMONS ; domaines de recherche : nasalité vocalique, contrôle et connaissance phonétique, imitation et convergence phonétique, production et perception de la parole en langues étrangères, évaluation de la parole en phonétique clinique ;
- Cécile Fougeron : phonéticienne, Professeur des Universités et HDR, directrice adjointe du Laboratoire de Phonétique et Phonologie, Paris; domaines de recherche : planification phonétique, programmation motrice et exécution de la parole et leurs troubles, corrélats phonétiques des troubles moteurs de la parole, sources de variations dans la parole et manifestations phonétiques de l'organisation prosodique ;
- Muriel Lalain : orthophoniste, chargée de recherche au Laboratoire Parole et Langage de l'Université Aix-Marseille ; domaines de recherche : notamment intelligibilité de la parole en production et en perception ainsi que ses troubles ;
- Vincent Martel-Sauvageau : orthophoniste, chercheur à la faculté de médecine de l'Université Laval, Québec, professeur adjoint ; domaines de recherche : propriétés acoustiques fines de la parole saine et pathologique (notamment dans la maladie de Parkinson), marqueurs de l'intelligibilité de la parole, plus particulièrement dans le contexte du français québécois ;
- Dominique Morsomme : logopède au CHU de Liège, spécialisée en voix parlée et chantée, professeure à l'Université de Liège et responsable de l'Unité Logopédie de la Voix ; domaines de recherche : évaluation et prise en charge de la voix parlée et de la voix chantée ;
- Timothy Pommée : logopède/orthophoniste, doctorant à l'Institut de Recherche en Informatique de Toulouse ; domaine de recherche : mesures d'intelligibilité de la parole ;
- Danièle Robert : médecin ORL – phoniatre, Praticien Hospitalier, Marseille, ses consultations sont principalement dédiées aux troubles de la parole et de la déglutition ;
- Virginie Woisard : médecin ORL – phoniatre, Praticien Hospitalier Professeur Associé et Directrice du Centre de Formation Universitaire en Orthophonie de Toulouse ; domaines de recherche : troubles de la parole, de la déglutition et de la voix.

Annexe J – Supports de production de parole courants en français, anglais et néerlandais

Français

Dans les batteries de tests standardisées, nous retrouvons comme supports de production de parole :

- le texte du protocole MonPaGe [LAGANARO 21] : non équilibré phonétiquement, car différents objectifs étaient visés, nécessitant notamment la présence répétée des voyelles cardinales /a, i, u/ (pour la création du triangle vocalique, mais aussi pour l'analyse de la variabilité vocalique), la présence de structures permettant d'analyser la coarticulation voyelle-à-voyelle, la présence d'éléments de séries automatiques (pour comparer la production en séries vs de mots isolés dans le texte) et une longueur limitée à 200 mots, rendant l'équilibrage impossible.
- dans la BECD [AUZOU 19] :
 - Score d'Intelligibilité : lecture de 10 mots (choisis aléatoirement parmi 50), 10 phrases (choisies aléatoirement parmi 50) et discours spontané (5 minutes de conversation) ;
 - Test Phonétique d'Intelligibilité : 5 listes de 52 items (ordre différent), mots bisyllabiques, choix multiple parmi 4, 4x13 contrastes phonétiques ;
 - Analyse phonétique : répétition des phonèmes français (voyelles isolées, glides avec /a/ ou /i/, consonnes avec /a/) ; 37 répétitions de mots simples comprenant 30 phonèmes en position initiale, médiane et finale chacun (sauf 2 phonèmes qui n'apparaissent pas en position finale en français) ; 25 répétitions de mots complexes avec glides et clusters ;
- dans la FDA-2 [GHIO 20] :
 - 10 mots parmi 101 ; critères de sélection : fréquence lexicale (>10 occurrences/million de mots), position des consonnes dans le mot (toutes les consonnes françaises dans chaque position et les groupes de consonnes les plus fréquents en position initiale), listes de mots de 1, 2, 3 ou 4 syllabes, structure phonotactique constante dans chaque liste de mots courts, variété des sons vocaliques dans chaque liste, contrôle du phonème final pour les mots bisyllabiques, point d'unicité repoussé au plus loin pour les mots longs ;
 - 51 phrases courtes : non répétitives (25 débuts différents, moins prévisibles), structures morphosyntaxiques variées, temps et modes variables (10 interrogatives, 6 exclamatives, 10 impératives et 25 déclaratives), mot cible dans chaque phrase avec une structure phonétique variable mais contrôlée (chaque consonne apparaît en position initiale et finale), occurrence du mot cible >10/million, tous les mots cibles sont absents du corpus de mots.
- dans l'épreuve de décodage acoustico-phonétique [GHIO 18] : pseudo-mots, c.-à-d. des non-mots qui respectent les structures phonotactiques fréquentes du français ; 52 éléments sont choisis au hasard dans une base de données de 89346 formes possibles ; chaque liste créée de 52 éléments est phonétiquement équilibrée ; structure : C(C)1V1C(C)2V2, où C(C)i est une consonne isolée ou un cluster ;
- dans le test de diagnostic par paires minimales [PECKELS 73] : 216 paires de mots ; utilisé uniquement dans la recherche car chronophage ; cible seulement les consonnes initiales.

En plus de ces supports, il existe aussi des listes de phrases et des textes qui ne sont pas intégrés dans des épreuves standardisées, tels :

- la chèvre de Monsieur Seguin [DAUDET 69] : non équilibré phonétiquement (œuvre littéraire) ;
- les trois textes de Harmegnies [HARMEGNIES 88] : phonétiquement équilibrés, 277 phonèmes, mais l'emploi de mots peu fréquents et de structures syntaxiques peu communes induit des perturbations dans la fluidité de la lecture ; difficile surtout pour les enfants de moins de 10 et pour les personnes peu lettrées ; induit également des problèmes dans l'utilisation de la prosodie ;
- la Bise et le Soleil (traduction de « The North Wind and the Sun ») [INTERNATIONAL PHONETIC ASSOCIATION 99] : non équilibré phonétiquement dans ses traductions multiples ;
- Listes de phrases de Combescure [COMBESCURE 81] : 20 listes de 10 phrases, équilibrage phonétique par liste, mais difficultés pour les analyses prosodiques ;
- un extrait de « Pierrot » de Guy de Maupassant (BECD) [DE MAUPASSANT 82] : non équilibré phonétiquement (œuvre littéraire), a été utilisé car il correspond à une minute de temps de parole ; émotionnellement chargé ;
- les phrases Lonchamp et Vaissière : deux phrases qui contiennent tous les phonèmes du français : « *Au loin un gosse trouve, dans la belle nuit complice, une merveilleuse et fraîche jeune campagne.* » et « *Il faut déjà que vous sachiez que les bords de telles rues ne sont qu'un peu glissants le matin à Zermatt.* »

Anglais

Voici une liste non exhaustive des supports les plus communément utilisés en anglais ; ceux-ci ne sont pas intégrés dans des batteries standardisées d'évaluation de la parole :

- Rainbow Passage [FAIRBANKS 60] : le texte entier a été construit pour inclure presque tous les phonèmes anglais, *a priori* équilibré phonétiquement ; certains mots sont difficiles à lire ; pas de charge émotionnelle ; long, donc généralement seul un court extrait est utilisé ;
- My Grandfather [VAN RIPER 63] : comprend presque tous les phonèmes anglais (à l'exception des plosives glottales et de /j/), plus court que le Rainbow Passage (132 mots, moins de 3 minutes) ; « inventaire varié de phonèmes anglais apparaissant à la fois isolés et imbriqués dans un ensemble de groupes phonotactiquement improbables (par exemple, frock, zest). Outre des exigences articulatoires éprouvantes, [il] assaille son lecteur de complexité syntaxique et sémantique » [J. REILLY 12] ; seulement des déclaratives et une exclamation [R. PATEL 13] ;
- The Grandfather Passage [DARLEY 75] : version légèrement modifiée de « My Grandfather » ; chevauchement de 88 % [J. REILLY 12] ; répertoire phonétique complet ;
- Zoo Passage [S. G. FLETCHER 72] : texte entièrement oral, exclusion de consonnes nasales
- The Caterpillar [R. PATEL 13] : contient tous les phonèmes anglais, une couverture phonotactique complète ; des déclaratives, des exclamations et des interrogations ; des mots de longueur croissante (fast/faster, amuse/amusement...) ; de nombreuses répétitions de mots ; inclusion de formes de mots et de phrases pour examiner le contrôle respiratoire, phonatoire, articulatoire, résonatoire et prosodique ; vocabulaire contemporain et syntaxe simple pour se concentrer sur les capacités de production de la parole tout en minimisant la charge cognitive ;
- The Harvard Sentences [IEEE 69] : 72 listes phonétiquement équilibrées de 10 phrases ; largement utilisées pour tester les systèmes de télécommunication ; obsolètes, le contenu de

certaines d'entre elles n'est pas approprié au regard des normes actuelles (p. ex. sexisme : « Les écoles pour dames enseignent le charme et la grâce ») ; charge émotionnelle (« Tous deux ont perdu la vie dans la tempête qui faisait rage »).

- Textes de Kuo & Weismer [KUO 16] : 6 textes pour l'évaluation des voyelles dans différents contextes hVd et CVC.

Néerlandais

En néerlandais, comme en français, nous retrouvons des supports de production de parole dans les batteries de test standardisées :

- dans le Dutch Intelligibility Assessment (NSVO) [DE BODT 06] : construit pour mesurer l'intelligibilité du locuteur au niveau des phonèmes et pour donner au clinicien un aperçu du type d'erreurs articulatoires segmentales commises par le locuteur ; 50 mots monosyllabiques CVC, principalement non sensés, mais facilement prononçables ; trois sous-ensembles : A) consonnes néerlandaises en position initiale (19 mots), B) consonnes en position finale (15 mots), C) voyelles et diphtongues en position médiane (15 mots) ; 35 variantes de chaque sous-test, choisies au hasard pour chaque évaluation ;
- dans le Dutch Sentence Intelligibility Assessment (NSVO-Z) [MARTENS 10b] : 1 200 phrases sémantiquement imprévisibles ; pour chaque évaluation, 18 phrases sont choisies aléatoirement et doivent être lues à voix haute ;

En plus de ces supports, il existe aussi des listes de phrases et des textes qui ne sont pas intégrés dans des épreuves standardisées, tels :

- De Noordenwind en de Zon (traduction de « The Northwind and the sun » [INTERNATIONAL PHONETIC ASSOCIATION 99]) : non équilibré phonétiquement dans ses traductions multiples ;
- Texte de [VAN ZUNDERT 98] : créé à l'origine pour analyser la nasalité sur les consonnes ; trois textes, avec un contenu nasal croissant (faible = « texte oral », normal, élevé) ; dans le texte « normal », le pourcentage de consonnes orales par rapport aux consonnes nasales est proche de la fréquence des phonèmes en néerlandais ; facile à répéter par les enfants car les phrases ne sont pas trop complexes ;
- [VAN DE WEIJER 91] :
 - Texte oral : similaire au « Zoo passage » [S. G. FLETCHER 72], qui exclut également les consonnes nasales ;
 - Texte oro-nasal : « Papa en Marloes » ; correspond exactement au « Rainbow passage » anglais [FAIRBANKS 60], contient 11,5 % de consonnes nasales ; 8 phrases ; seul texte dit « phonétiquement équilibré » en néerlandais ; utilisé en clinique d'orthophonie flamande et néerlandophone ; Critique [MARTENS, 10a] : équilibre phonétique basé sur un corpus de langage écrit ; facile à lire par les enfants, mais peu adapté aux adultes ; pas facile à répéter car il n'y a pas d'intrigue (passage descriptif) ; marqué régionalement (p. ex. « Marloes » fréquent aux Pays-Bas mais pas en Flandre, choix des prépositions et du vocabulaire) ;
 - Texte nasal : rapport consonnes nasales/consonnes totales=57% (86/152)
- Moolenaar-Bijl passage [BOERING 61] : 3 phrases orales pour l'évaluation de l'hypernasalité, facilement répétées par de jeunes enfants ;
- « De Koning » [MARTENS 10a] : créé méthodologiquement ; court (37 phrases, moyenne 3 de minutes) ; se concentre sur les plosives et les voyelles, parce que les plosives non voisées

et la réduction vocalique (F1, F2) sont faciles à mesurer ; pour les plosives, dans la dysarthrie, l'air s'échappe déjà pendant l'intervalle de silence avec le résultat que l'explosion est faible ou absente ; l'air qui s'échappe pendant l'intervalle de silence cause du bruit et est facilement mesurable (quantité d'intensité dans l'intervalle silencieux par rapport à l'intensité d'un locuteur sain) ; voyelles /e, o, a/ (moins extrêmes, permettent plus de variation) ; plosives en position finale des mots et des phrases, parce que la quantité de bruit dans l'intervalle silencieux est mieux mesurée dans cette condition ; les plosives et les voyelles apparaissent dans les mots de contenu (plus fréquents) ; prise en compte de l'accentuation phrastique et lexicale, de la position des mots, du lieu et du mode d'articulation, de la longueur des mots (6-44 mots) et des phrases, des types de mots et de la nasalité (deux phrases entièrement nasales et deux phrases entièrement orales) ; il s'agit d'un texte clair et simple, avec un contenu adapté aux adultes ; contient des déclaratives, des exclamations, des phrases interrogatives et des ordres ; les processus phonologiques en langage spontané ont été pris en compte (c.-à-d. des mots qui changent de sens en cas de processus phonologique).

Annexe K – Texte original MonPaGe [FOUGERON 19]

Lundi, le chat, le loup et Papa vont à Bali. Les copains sont tout contents.

Mardi, Papy y va aussi. Il dit : « Je n'ai pas un sou ! Qui va prendre soin de moi ? » « Moi ! » dit le chat, « moi ! » dit le loup. « Vous ? », Papy réfléchit.

Mercredi, Papy dit : « Toi, le chat, tu es doux, tu es chou, tu n'as pas de poux ! Mais pas ce loup : il a une cape rouge et je n'aime pas ce gars-là ! »

Jeudi, le chat et Papy se baladent à Bali. Papa glisse ! Aïe ! Ouille ! Son cou craque, son coude claque, c'est la débâcle !

Vendredi, Papa a mal. Il pleure, il crie ! « Toi, Papy, aide-moi, trouve le nain ! » « Un nain ? On n'en a jamais vu par ici ? ! »

Samedi matin, le chat va voir son ami le loup et lui dit : « Aide-moi à soigner Papa ! »

Samedi soir, le loup lui donne sa recette magique : « Coupe un oignon, cache-le sous la souche, et lorsque le lilas fleurira, Papa sera guéri ! » Abracadabra, ça y est, on a réussi !

Dimanche, le chat tout doux, le loup magicien, Papa et Papy quittent Bali. Les copains sont tout contents.

Annexe L – Transcriptions phonétiques du nouveau texte pour l'évaluation de la parole et de la voix

Transcription phonétique en français de France et de Belgique

lœdi/lœdi matē alis e sō papa vō(t) a malibu
 laba il rœzwan(ə)/rœzwan(ə) papi aprɛ/aprɛ œ/ɛ vwajaʒ(ə) sū susi
 il fɛ/fe ʃo mɛ/me la briz(ə) leʒɛr(ə) e lɛr iɔde/jode də/dø la mɛr lɛ/le raviv(ə)
 vɛr midi alis (s)ɛ(k)sklam(ə)/(s)ɛ(k)sklam(ə) ʒɛ/ʒe vrɛmā/vrɛmā trɛ/tre trɛ/tre fɛ
 papi lɛ/le gid(ə) albɔ vit(ə) vɛr œ/ɛ kafe lyksyø o/ɔ bɔr dœ/dø lo lœ/lø bigɔrno
 sale
 mardi il vō(t) a la plɑʒ(ə)
 il nja pa(z)œ/ɛ nɔɑʒ(ə) dā lə/lø sjɛl papa sɛtɛrɔʒ(ə)/sɛtɛrɔʒ(ə) avō nu pri la krɛm(ə)
 sɔlɛr(ə)/sɔlɛr(ə)
 bjɛ syr rɛpō alis(ə)
 mɛrkrœdi/mɛrkrœdi papa e papi sə/sø balad(ə) ā bavardā
 pādā sə/sø tā alis(ə) (s)ə/(s)ø detā ā lizā œ/ɛ rɔmā/romā e māʒ(ə) œ/ɛ bōbō alanan(s)
 ʒœdi ɛl va fɛr(ə) œ/ɛ dʒɔgiŋ /dʒɔgiŋ
 papa lɔi/lwi kri(j)(ə) nu partō fɛr(ə) kɛlkœ/kɛlkø zafɑ
 o magazɛ papi afɛt(ə) dɛ/de nwa də/dø makadamja
 vādrœdi/vādrœdi il vizit(ə) œ/ɛ myze dar apstrɛ/apstre
 papa sɛkstazi(j)(ə)/sɛkstazi(j)(ə) dəvā/døvā œ/ɛ splādidi(ə) tablɔ e dəmād(ə)/dəmād(ə)
 ki a dō(k)kre.e sɛt œvr(ə)
 sam(ə)/ødi matē alis(ə) (s)ātrɛn(ə) pur la swarɛ karaokɛ/karaokɛ ā rɛpɛtā/rɛpɛtā
 rapid(ə)/ømā patakapatakapataka
 sam(ə)/ødi swar il fɛt(ə) lœr dɛpar ā dāsū la ʒava su lə/lø lila
 kɔm a larivɛ il fɛ/fe ʃo mɛ/me la briz(ə) leʒɛr(ə) e lɛr iɔde/jode də/dø la mɛr lɛ/le
 raviv(ə)
 dimāʃ(ə) alis(ə) papa e papi kit(ə) malibu
 il rātr(ə) afamɛ
 a tabl(ə)ilja də/dø la pidza garni(j)(ə) e dɛ/de laʒan(ə) o ʃāpinō
 rasazje il sɛ(k)sklam(ə)/sɛ(k)sklam(ə) kɛl sɛʒur ɛkstr(a)ɔrdinɛr(ə)/ɛkstr(a)ɔrdinɛr(ə)

*Transcription phonétique en français québécois*³⁸

Note : pour les notations avec « / », la première variante est plus familière, alors que la seconde est plus formelle. Les symboles entre (parenthèses) sont généralement élidés.

lœdʒi matē alis e sō papa vō (t)ɑ malibu
labɑ il ɛʒwan papi apɛ(z) œ vwajaʒ sɑ susi
il fɛ ʃo mɛ la bɔiz leʒɛ/ɛɛ e laɛ/ɛɛ i.ɔde d(ə) la maɛ/ɛɛ le/ɛ ɛaviv
vɛɛ/ɛɛ midʒi alis sɛksklam ʒe vɛmā tɛɛ tɛɛ fɛ
papi le/ɛ gid albɔ vit vɛɛ/ɛɛ œ kafe lyksy.ø/lyksyø o bɔɛ dɑ lo lə bigɔɛno sale
 maɛdʒi il vō(t) a la plɑʒ
 il ni ɑ/nja pa œ nu.ɑʒ dā l(ə) sjɛl papa sɛtɛ/ɛɛʒ avō nu pri la kɛɛm sɔlaɛ/ɛɛ
 bjɛ svɛ rɛpō/rɛpō alis
 mɛɛkrœdʒi papa e papi s(ə) balad ā bavardā
 pādā sə tā alis (s)ə detā ā lizā œ ɛmā e māʒ œ bōbō a lanana

³⁸ Merci à Vincent Martel Sauvageau (Ph.D., Université Laval) et à Liziane Bouvier (Ph.D., Université de Toronto) pour cette transcription.

3øðzi el va faε/εv œ dʒogɪŋ
 papa lɪi kvi nu pavtõ faε/εv kelkə zaʃa
 o magazē papi afet de/ε nwa d(ə) makadamja
 vādʒəðzi il vizit œ myze dav abʒ/bstvε
 papa sekstazi dəvã œ splāðzid tablo e dəmãd/dəmãn ki a/kja dõ(k) kvε.e set œv(v)
 samðzi matē alis sātʒaε/en puv la swavε kara.oke ã vεpε/etã vapid(ə)mã pataka/a pataka/a pataka/a
 samðzi swav il faε/εt laœv depav/depav ã dāsã la java su l(ə) lila
 kɔm a lavive il fe fo mε la bʒiz leʒaε/εv e laε/εv iode d(ə) la maε/εv le/ε kvaviv
 dzimãʃ alis papa e papi kɪ malibu
 il vāt(v) afame
 a tab(l) il i/j a d(ə) la pɪða gavni e de/ε laʒav o ʃãpɪŋõ
 vasazje il seksklam kel sezov ekstv(a)vðzinaε/ε(v) / εgstv(a)vðzinaε/ε(v)

Annexe M – Critères contrôlés dans la création du texte entier pour l'évaluation de la parole et de la voix

Les critères faisant partie des 10 critères « phares » à contrôler selon l'étude Delphi sont identifiés en gras.

Critère à contrôler	Mise en application	Source
Niveau phonémique		
Inventaire phonémique complet	Inclusion de tous les phonèmes de la langue française	Delphi
Équilibre phonétique	Corrélation Spearman de 0,78 entre les fréquences dans le texte et dans la langue française selon [TUBACH 90], $p < 0,0001$; test U de Mann-Whitney : $U = 515$, $p = 0,25$ → distribution des fréquences identique sur le texte et en langue française (voir Tableau K.1.)	
Prise en compte des positions dans le mot et du contexte phonémique	<p>Voyelles : inclusion dans les contextes, CVCV/CV.CVC ; C=occlusive sourde, R ou L (voir Tableau K.2.)</p> <p>Consonnes :</p> <ul style="list-style-type: none"> - inclusion de toutes les consonnes non-sonantes dans un même contexte phonémique : aCa - inclusion de contrastes de paires minimales avec contraste de voisement (pas/bas, faire/vers, guide/quittent) <p>Inclusion d'occlusives sourdes et de clusters consonantiques en initiale de mots et d'enchaînements consonne sourde – consonne sonore :</p> <ul style="list-style-type: none"> - Occlusives sourdes initiales : papa (x6), papy (x5), café, plage, pas, pris, crème, pendant, temps, crie, partons, quelques, tableau, qui, créé, pour, karaoké, pataka (x3), comme, quittent, table, pizza, quel - Clusters initiaux : brise (x2), vraiment, très (x2), plage, pris, crème, jogging, crie, trop, splendide, créé - CsourdeCsonore : après, très (x2), plage, pris, crème, mercredi, crie, abstrait, splendide, créé, entraîne, rentrent, exclame(nt) (x2), extraordinaire (semi-consonnes non comptées) 	
Inclusion de clusters consonantiques	<p>Bi-consonantiques : 26</p> <p>Tri-consonantiques : 3 (/mɛrkrœdi/, /ɛkstasi/, /splãdid/)</p> <p>Quadri-consonantiques : 4 (/ɛksklam/ x2, /apstrɛ/, /ɛkstraɔrdinɛr/)</p>	
Inclusion d'éléments permettant de mesurer la coarticulation voyelle-à-voyelle	« Papa » vs « Papy »	

Répétitions multiples des voyelles cardinales	/a/ : 115 /i/ : 59 /u/ : 8	MonPaGe
Inclusion de mots avec des glides pour la dynamique articulatoire	/w/ : rejoignent, voyage, noix, soirée, soir /j/ : voyage, n'y a, ciel, bien, macadamia, il y a, rassasiés /ɥ/ : luxueux, nuages	
Inclusion d'éléments permettant d'analyser le contraste nasal/oral	<p>Consonnes :</p> <p>/p a, b a/ vs /m a/ :</p> <p>- /p a/ : « Papa », « Papy », « pas », « partons », « pataka », « départ »</p> <p>- /b a/ : « là-bas », « baladent », « bavardant »</p> <p>- /m a/ : « matin », « Malibou », « mardi », « magasin », « makadamia »</p> <p>/p i, b i/ vs /m i/ :</p> <p>- /p i/ : « Papy », « rapide », « pizza », « champignons »</p> <p>- /b i/ : « Bigorneau »</p> <p>- /m i/ : « midi »</p> <p>/t a, d a/ vs /n a/ :</p> <p>- /t a/ : « extasie », « tableau », pataka », « table »</p> <p>- /d a/ : « makadamia », « d'art »</p> <p>- /n a/ : « ananas »</p> <p>/d i/ vs /n i/ :</p> <p>- /d i/ : « lundi », « midi », « mardi », « mercredi », « jeudi », « vendredi », « splendide », « samedi », « dimanche », « extraordinaire »</p> <p>- /n i/ : « garnie »</p> <p>Voyelles :</p> <p>/s a/ vs /s ā/ :</p> <p>- /s a/ : « salé », « samedi », « rassasiés »</p> <p>- /s ā/ : « sans », « dansant »</p> <p>/p a/ vs /p ā/ :</p> <p>- /p a/ : « Papa », « Papy », « pas », « partons », « pataka », « départ »</p> <p>- /p ā/ « pendant »</p> <p>/m a/ vs /m ā/ :</p> <p>- /m a/ : « matin », « Malibou », « mardi », « magasin », « macadamia »</p> <p>- /m ā/ : « vraiment », « roman », mangeant », « demande », « rapidement », « dimanche »</p> <p>/d a/ vs /d ā/ :</p> <p>- /d a/ : « makadamia », « d'art »</p> <p>- /d ā/ : « dans », « bavardant », « pendant », « dansant »</p> <p>/t a/ vs /t ā/ :</p> <p>- /t a/ : « extasie », « tableau », « pataka », « table »</p> <p>- /t ā/ : « temps », « détend », « répétant »</p> <p>/v a/ vs /v ā/ :</p>	

	<p>- /v a/ : « bavardant », « va », « java » - /v ã/ : « vendredi », « devant »</p> <p>/ʃ a/ vs /ʃ ã/ : « achats » / « champignons »</p> <p>/z a/ vs /z ã/ : « pizza », « lasagne » / « lisant »</p> <p>/l a/ vs /l ã/ : - /l a/ : « la », « là », « exclame(nt) », « plage », « balade », « lila », « lasagne » - /l ã/ : « splendide »</p> <p>/r a/ vs /r ã/ : - /r a/ : « ravivent », « karaoké », « rapidement », « rassasiés » - /r ã/ : « rentrent »</p> <p>/f ε/ vs /f ě/ : « fait », « faire », « fêtent » / « faim »</p> <p>/s ε/ vs /s ě/ : « cette » / « s'interrogent »</p> <p>/s o/ɔ/ vs /s ɔ̃/ : « solaire » / « son »</p> <p>/b o/ɔ/ vs /b ɔ̃/ : « bord » / « bonbon »</p>	
<p>Inclusion de /i – e – ε – a/ pour la mesure de la différence F1-F2 (contraste hauteur de langue)</p>	<p>Prise en compte des phénomènes d'affrication en français québécois, des variations régionales de prononciation vocalique en France et en Belgique ainsi que de la facilité de segmentation (selon le contexte phonémique, p. ex. voyelle précédée d'une plosive ou fricative sourde) :</p> <p>- /i/ de « <u>m</u>idi », « <u>M</u>alibou », « <u>B</u>igorneau », « <u>v</u>isitent », « <u>l</u>ila », « <u>a</u>rrivée » (pas d'affrication, syllabe ouverte en français québécois) ou « <u>P</u>apy », « <u>q</u>ui », « <u>p</u>izza », « <u>ch</u>ampignons » (occlusives sourdes)</p> <p>- /e/ de « <u>A</u>lice et son » (/s e s/), « <u>i</u>odé de » (/d e d/), « <u>d</u>étend » « <u>d</u>épart » ou « <u>k</u>araoké »</p> <p>- /ε/ de « <u>v</u>ers », « <u>f</u>aire », « <u>q</u>uelques », « <u>a</u>chète », « <u>c</u>ette », « <u>f</u>ête » ou « <u>q</u>uel » (prononciation /ε/ même en France)</p> <p>- /a/ de « <u>P</u>apa », « <u>P</u>api », « <u>c</u>afé », « <u>p</u>as », « <u>p</u>artons », « <u>a</u>chats », « <u>m</u>acadamia », « <u>e</u>xtasie », « <u>t</u>ableau », « <u>s</u>amedi », « <u>k</u>araoké », « <u>p</u>ataka », « <u>d</u>épart », « <u>a</u>ffamés », « <u>à</u> table », « <u>r</u>assasiés »</p>	Revue systématique
<p>Inclusion de la consonne /s/ suivie des voyelles cardinales /sa, si, su/ pour la mesure du 1^{er} moment spectral sur la fricative</p>	<p>/sa/ : « <u>s</u>alé », « <u>s</u>amedi » (x2), « <u>r</u>assasiés » /si/ : « <u>s</u>oucis » /su/ : « <u>s</u>oucis », « <u>s</u>ous »</p>	
<p>Inclusion de /t,d/ après /i/ (mesure de F1 en offset de /i/ devant /t,d/, indice de voisement)</p>	<p>/id/ : « <u>g</u>uide », « <u>m</u>idi », « <u>s</u>plendide », « <u>r</u>apide », « <u>p</u>izza » /it/ : « <u>v</u>ite », « <u>v</u>isitent », « <u>q</u>uittent »</p>	

Enchaînement de voyelles et de fricatives sourdes (flux d'air continu avec voisement intermittent, poser la voix sur le souffle)	« sans soucis », « il fait chaud »	Groupe de travail
Inclusions de mots avec voyelle initiale, en position initiale de phrase ou après un signe de ponctuation induisant une pause : incitent d'éventuels coups de glotte (forçage vocal)	Début de phrase : « il fait chaud », « il n'y a pas », « avons-nous pris », « au magasin », « ils rentrent », « à table » Après un signe de ponctuation (liaison possible ou coup de glotte probable chez hypertoniques) : « Alice » (x5), « ils rejoignent », « elle va faire », « ils visitent », « ils fêtent », « il fait chaud », « il y a », « ils s'exclament »	
Niveau lexical		
Contrôle de la complexité articulatoire/phonétique des mots	Inclusion de mots simples et de mots complexes ; complexité calculée à l'aide de l'Indice de Complexité Phonétique [H. LEE 14], intégrant huit paramètres dans son calcul : - Min : 0 ; Max : 16 - Moyenne : 3 ; E.T. : 2,5 - Médiane : 3 ; EIQ : 3,75 16 % des mots ne présentent aucun paramètre de complexité ; 45 % présentent entre 1 et 3 paramètres de complexité, 39 % présentent plus de 4 paramètres de complexité.	Delphi
Inclusion de mots répétés à plusieurs reprises et de paires de mots de longueur croissante (apraxie vs dysarthrie)	Mots répétés : Alice (x6), Papa (x6), Papy (x5), pataka (x3), matin (x2), vont (x2), malibou (x2), fait (x2), chaud (x2), mais (x2), brise (x2), légère (x2), l'air (x2), iodé (x2), mer (x2), ravivent (2), vers (x2), très (x2), faire (x2), samedi (x2), s'exclame(nt) (x2) Paires croissantes : + table/tableau + fait/faire/fêtent + achats/achète + soir/soirée + dans/dansant	
Couplage avec des items de tâches d'évaluation isolées	Inclusion des jours de la semaine (langage automatique) et de diadococinésies (« pataka »)	MonPaGe

Contrôle de la fréquence lexicale	<p>Inclusion de mots fréquents et de mots rares. La fréquence lexicale a été évaluée à l'aide de la base de données Lexique 3.83 (http://www.lexique.org/; fréquence des lemmes dans les sous-titres de films) :</p> <ul style="list-style-type: none"> - 25 mots rares (<10 occurrences / million d'occurrences) (11 %) dont 20 très rares (<5 occurrences / million d'occurrences) : abstrait, ananas, bigorneau, exclame(nt), extasie, garnie, iodé, java, jogging, karaoké, lasagnes, lilas, luxueux, macadamia, rassasiés, ravivent, salé ; achats, affamés, brise, papy, solaire - 128 mots très fréquents (>200 occurrences / million d'occurrences) (56 %) : a, à, achète, ai, air, alors, après, au, aux, bien, ce, cette, comme, dans, de, demande, des, devant, donc, eau, elle, en, et, faire, fait, il, ils, la, le, les, leur, lisant, lui, là-bas, mais, mange, matin, nous, papa, partons, pas, pendant, pour, pris, quel, quelques, qui, quittent, rentrent, répond, sans, se, soir, son, sous, sûr, temps, très, un, va, vers, vite, vont, vraiment, y 	Groupe de travail
Niveau phrastique		
Prosodie : focus, contours intonatifs, discours rapporté, modalités...	<p>Modalités :</p> <ul style="list-style-type: none"> - Interrogative : 2 - Déclaratives : 17 - Exclamative : 4 - Discours rapporté : 7 <p>Limitation du nombre de points d'exclamation pour éviter d'induire une interprétation trop théâtrale (par rapport au texte MonPaGe).</p> <p>Utilisation de tournures de phrases plutôt que juste de la ponctuation pour induire les modulations prosodiques (p. ex. « s'exclame : », « s'interroge : »)</p> <p>Focus (p. ex. « vraiment très très faim »)</p>	Delphi
Inclure des phrases à durée variable (soutien respiratoire, groupes de respiration, fatigabilité)	<p>Nombre de mots :</p> <ul style="list-style-type: none"> - Min : 3 ; Max : 18 - Moyenne : 9,9 ; E.T. : 4,2 - Médiane : 9 ; EIQ : 7 <p>Nombre de syllabes (minimum) :</p> <ul style="list-style-type: none"> - Min : 5 ; Max : 32 - Moyenne : 15,2 ; E.T. : 6,6 - Médiane : 15 ; EIQ : 6 <p>Nombre de phonèmes (minimum) :</p> <ul style="list-style-type: none"> - Min : 11 ; Max : 67 - Moyenne : 33,3 ; E.T. : 13,4 <p>Médiane : 32 ; EIQ : 13,5</p>	Delphi
Inclure une séquence identique en début et en fin de texte pour les effets de fatigue	« mais la brise légère et l'air iodé de la mer les ravivent. », séquence placée en début et en fin du texte, mais ni en première ni en dernière phrase pour éviter effets prosodiques (p. ex. relâchement final)	MonPaGe
Inclusion d'une phrase à prédominance nasale vs orale (insuffisance vélaire, cleft palate...)	<p>Phrase à prédominance nasale : « Pendant ce temps, Alice se détend en lisant un roman et mange un bonbon à l'ananas. »</p> <p>Phrase à prédominance orale : « la brise légère et l'air iodé de la (m)er les ravivent » ; « Papy les guide alors vite vers (un) café luxueux au bord de l'eau : Le Bigor(n)eau Salé. » ; « pataka pataka pataka »</p>	MonPaGe

Inclure des segments 100 % voisins	« mais la brise légère et l'air iodé de la mer les ravivent. » (x2) « en lisant un roman et mange un bonbon à l'ananas. »	Groupe de travail
Niveau global		
Longueur globale du texte : environ 200 mots	23 phrases, 227 mots, 350 syllabes (minimum ³⁹) avec possibilité de passation rapide sur 58 mots	Delphi
Thématique contemporaine, aussi neutre que possible	Évitement d'éléments induisant une charge émotionnelle (notamment douleur et guérison dans le texte MonPaGe) ; registre courant (ni soutenu, ni familier) ; contenu et thématique non infantilisant	Groupe de travail
Lisibilité	Score de facilité de lecture de Flesch (0 = très difficile à lire, 100 = très facile à lire pour les 10 ans et +) : 55,5 Formule de calcul [FLESCHE 48] : $206,835 - 1,015 \times \frac{\text{mots totaux}}{\text{phrases totales}} - 84,6 \times \frac{\text{syllabes totales}}{\text{mots totaux}}$	Groupe de travail

Tableau M.1. – Fréquences d'occurrence des phonèmes dans le nouveau texte et dans la langue française (N=766)

Consonnes											
	NTxt	FTxt (%)	FFr (%)		NTxt	FTxt (%)	FFr (%)		NTxt	FTxt (%)	FFr (%)
p	40	5,22	3,74	t	25	3,26	5,22	k	22	2,87	4,22
b	14	1,83	1,05	d	39	5,09	4,13	g	5	0,65	0,55
m	28	3,66	3,30	f	8	1,04	1,50	s	35	4,57	6,10
n	11	1,44	2,79	v	20	2,61	2,27	z	11	1,44	1,85
ʃ	6	0,78	0,47	l	63	8,22	5,74	ʒ	3	0,39	0,08
ʒ	13	1,70	1,59	r	58	7,57	7,22	ʒ	1	0,13	/
Semi-voyelles											
ɥ	2	0,26	0,46	w	5	0,65	1,08	j	7	0,91	1,77
				ɥ/w	1	0,13		i/j	2	0,26	/
Voyelles											
	NTxt	FTxt (%)	FFr (%)		NTxt	FTxt (%)	FFr (%)		NTxt	FTxt (%)	FFr (%)
i	59	7,70	5,50	y	3	0,39	1,90	u	8	1,04	2,02
e	26	3,39	7,19	ø	2	0,26	0,45	o	7	0,91	1,07
ε	22	2,87	3,81	œ/ə	2	0,26	5,59	ɔ	6	0,78	2,06
e/ε	21	2,74	/	ø/œ/ə	17	2,22	/	o / ɔ	7	0,91	/
a	115	15,01	7,98	œ/ε	9	1,17	0,54	ē	6	0,78	1,26
ā	27	3,52	3,55					ō	10	1,31	1,95

Notes : NTxt = nombre d'occurrences dans le texte ; FTxt = fréquence d'occurrence dans le texte ; FFr = fréquence d'occurrence dans la langue française (Tubach & Boé, 1990)

³⁹ Le nombre exact de syllabes et de phonèmes peut varier en fonction des différences de prononciation (liaisons et variations régionales)

Tableau M.2. – Contextes consonantiques des voyelles

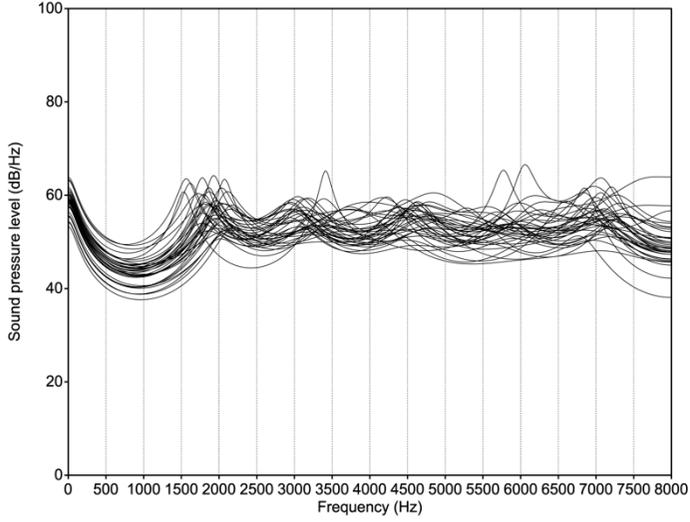
N total	Voy	Mots monosyllabiques, syllabes fermées			Mots bisyllabiques, syllabe finale fermée	
		CVC	CCVC	VC	(C)CV.CVC(C)	
12	a	1 dar	4 plaʒ, nʒaʒ, swaʁ, tabl		7 rəʒwaŋ, vwajaʒ, balad, depar, lazaŋ, eksklam (x2)	
15	i	3 gid, vit, kit	1 briz	1 il	10 alis (x6), raviv, dʒəŋiŋ, vizit, splâdid	
2	u	1 pur			1 seʒur	
19	ɛ	1 lɛr (x2), mer (x2), vɛr (x2), fɛr (x2), sɛt, fɛt, kɛl	2 sjɛl, krɛm	1 ɛl	5 leʒɛr (x2), sɔlɛr, aʒɛt, âtrɛn	
3	ɔ	2 bɔr, kɔm			1 alɔr	
2	œ	1 lœʁ		1 œvr		
1	y	1 syr				
4	ã	2 mãʒ, rãtr			2 dimãʃ, dãmãd	
1	ɔ̃	1 dɔ̃k				
N total	Voy	Mots monosyllabiques, syllabes ouvertes			Mots bisyllabiques, syllabe finale ouverte	
		V	(C)CV		(CC)V(C).CV	
33	a	8 a	occl non voisées	1 pa	7 papa (x6), tablo	
			occl voisées	1 ba		
			fric non voisées			1 aʃa
			fric voisées	1 va		2 zava, pidza
			semi-voyelles, liquides et nasales	1 la (x10), 1 nwa		1 lila
18	i	1 i	occl non voisées	1 ki	5 papi (x5)	
			occl voisées		6 lœdi, maʁdi, ʒœdi, sam(ə)di (x2), midi	
			fric non voisées		1 susi	
			fric voisées			
			semi-voyelles, liquides et nasales	3 pri, lʁi, kri	1 garni	
2	u		occl non voisées			
			occl voisées			
			fric non voisées	1 su		
			fric voisées			
			semi-voyelles, liquides et nasales	1 nu		
6	o	2 o		3 fo (x2), lo	1 tablo	
17	e	8 e		4 kre	5 kafe, sale, jode, myze, sware	
1	∅				1 lyksq∅	

4	ẽ		2	fẽ, bjẽ	2	matẽ (x2)	
8	õ		3	sõ, võ (x2)	5	avõ, partõ, repõ, bõbõ, partõ	
14	ã	4	ã	3	sã, tã, dã	7	vremã, pãdã, detã, lizã, romã, dævã, dãsã
1	y				1	myze	
Mots trisyllabiques, syllabe finale ouverte							
18		malibu (x2), bigorno, vãdædi, mækkædi, bavardã, anana, magazẽ, ekstazi, repetã, rapidãmã, pataka (x3), arive, afame, fãpiñõ, rasazje					
Mots quadrisyllabique (constitués seulement de syllabes ouvertes)							
2		makadamja, karaoke					

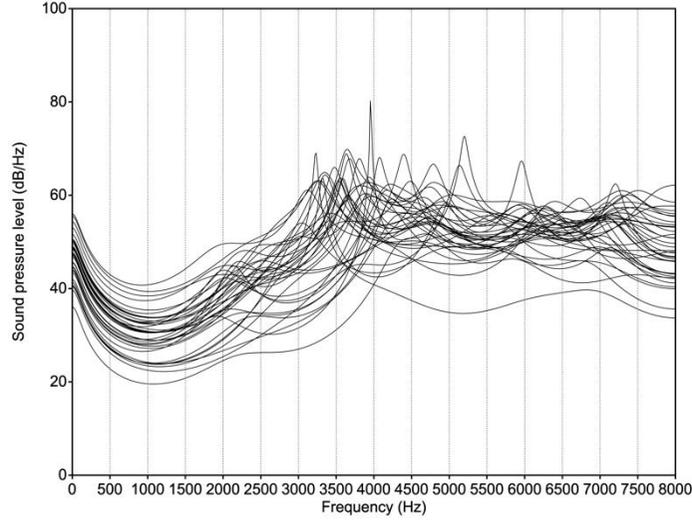
Note : en vert, les voyelles cardinales ; N total = nombre total d'occurrences ; Voy = voyelle

Annexe N – Spectres lissés pour les consonnes non-sonantes du français (segmentation manuelle)

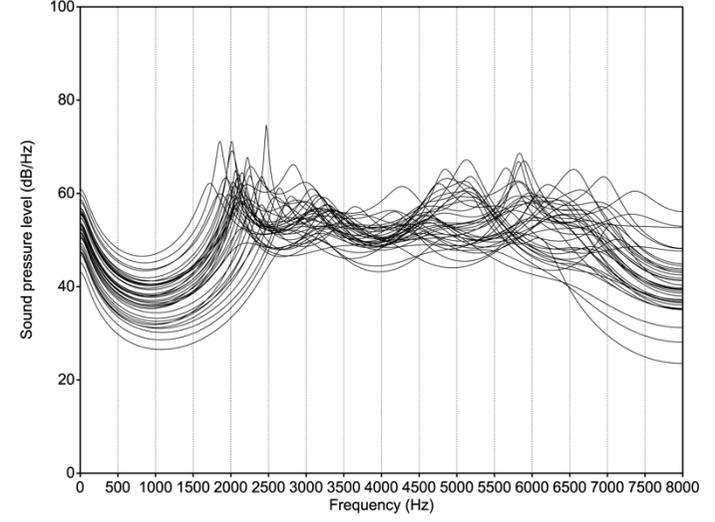
/f/



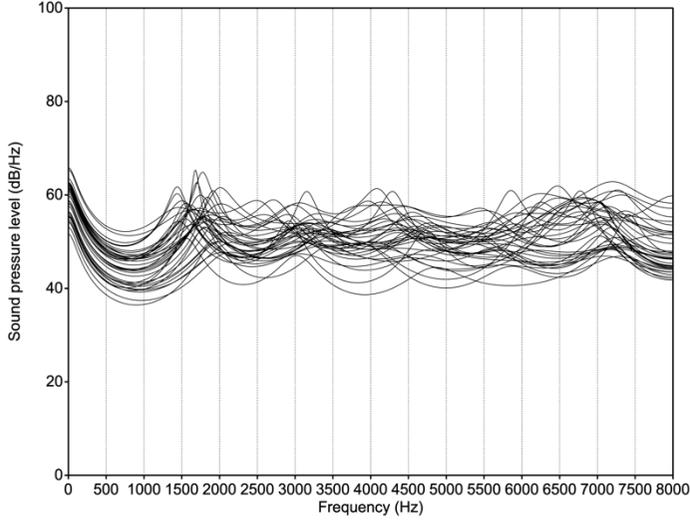
/s/



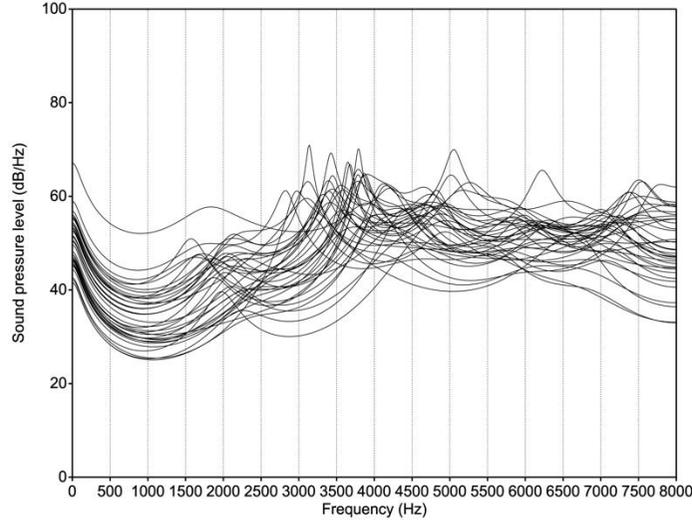
/ʃ/



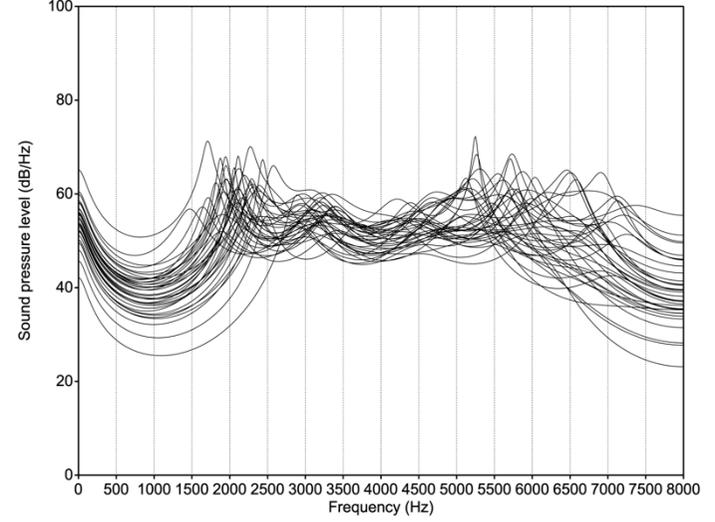
/v/



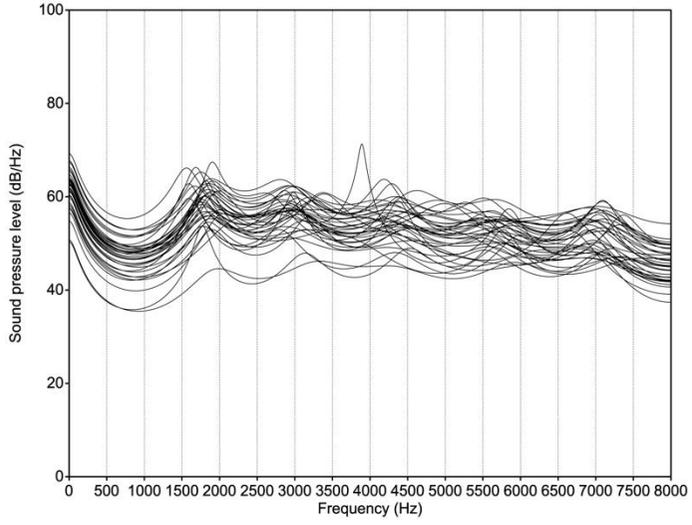
/z/



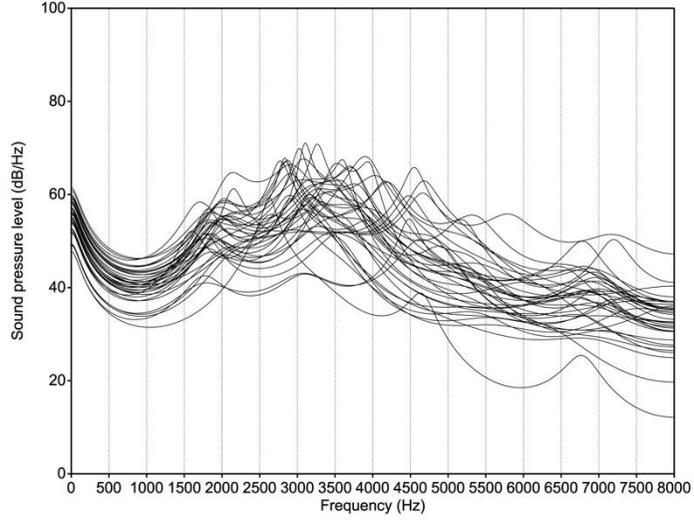
/ʒ/



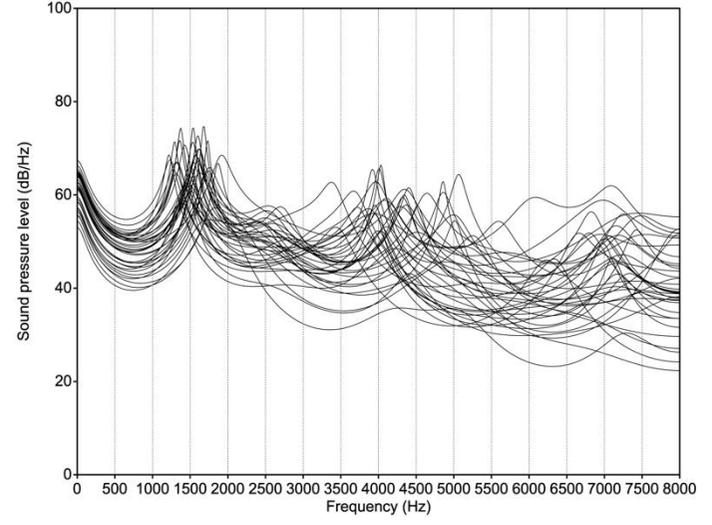
/p/



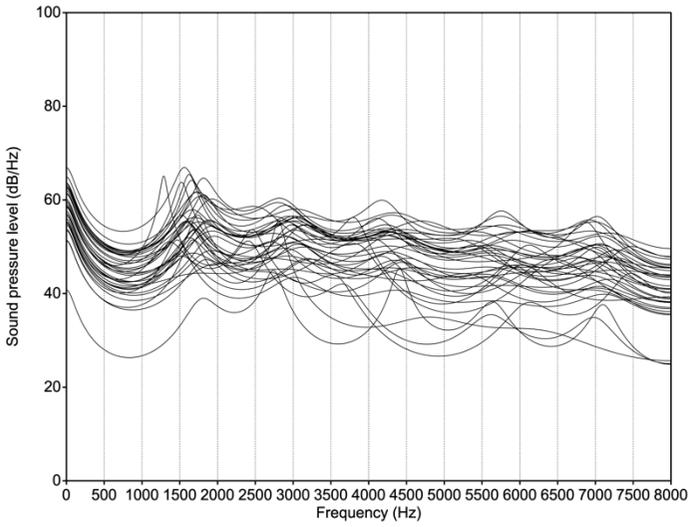
/t/



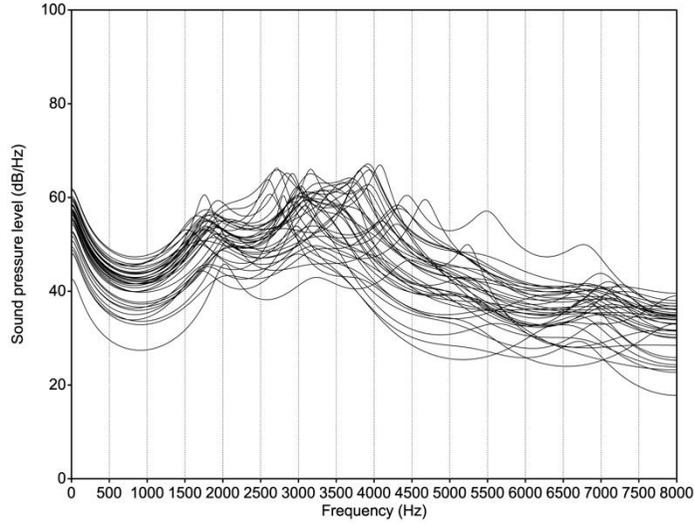
/k/



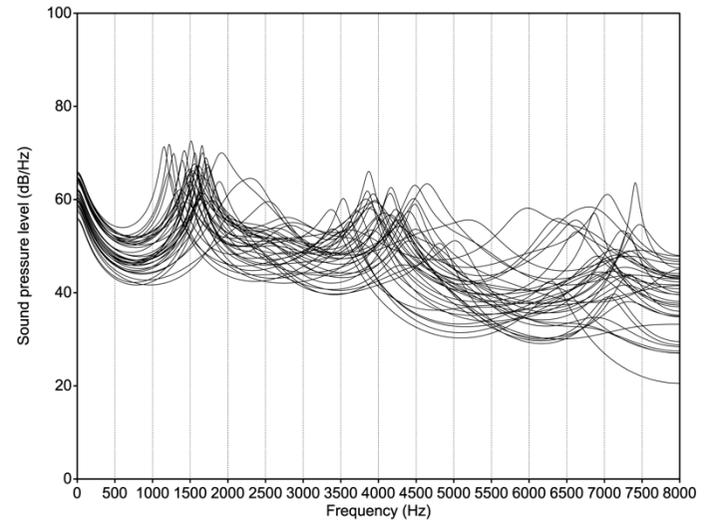
/b/



/d/



/g/



Annexe O – Analyse de variation des moments spectraux selon le contexte phonémique

Échantillons de parole :

- Dix enregistrements de sujets sains (tirés de notre base de données décrite dans le manuscrit) : 5 hommes, 5 femmes
 - o Phrases porteuses « Le sac /øCø/ convient »
 - o Extrait de « La chèvre de Monsieur Seguin » (en vert : plosives ; en bleu : fricatives) :
*Monsieur **S**egu**i**n n'**a**vait **j**amais eu **d**e bonheur avec ses **ch**èvres. Il les **per**dait **t**outes **d**e la même **f**açon : Un **b**eau **m**atin, elles **c**assaient leur **c**orde, **s**'en allaient **d**ans la montagne, et là-haut le loup les **m**angeait. Ni les **c**aresses **d**e leur maître, ni la **p**eur **d**u loup, rien ne les **r**etenait. **C'**étaient, **p**araît-il, **d**es **ch**èvres **i**ndépendantes, voulant à **t**out **p**rix le **g**rand air et la **l**iberté.*
- Segmentation manuelle

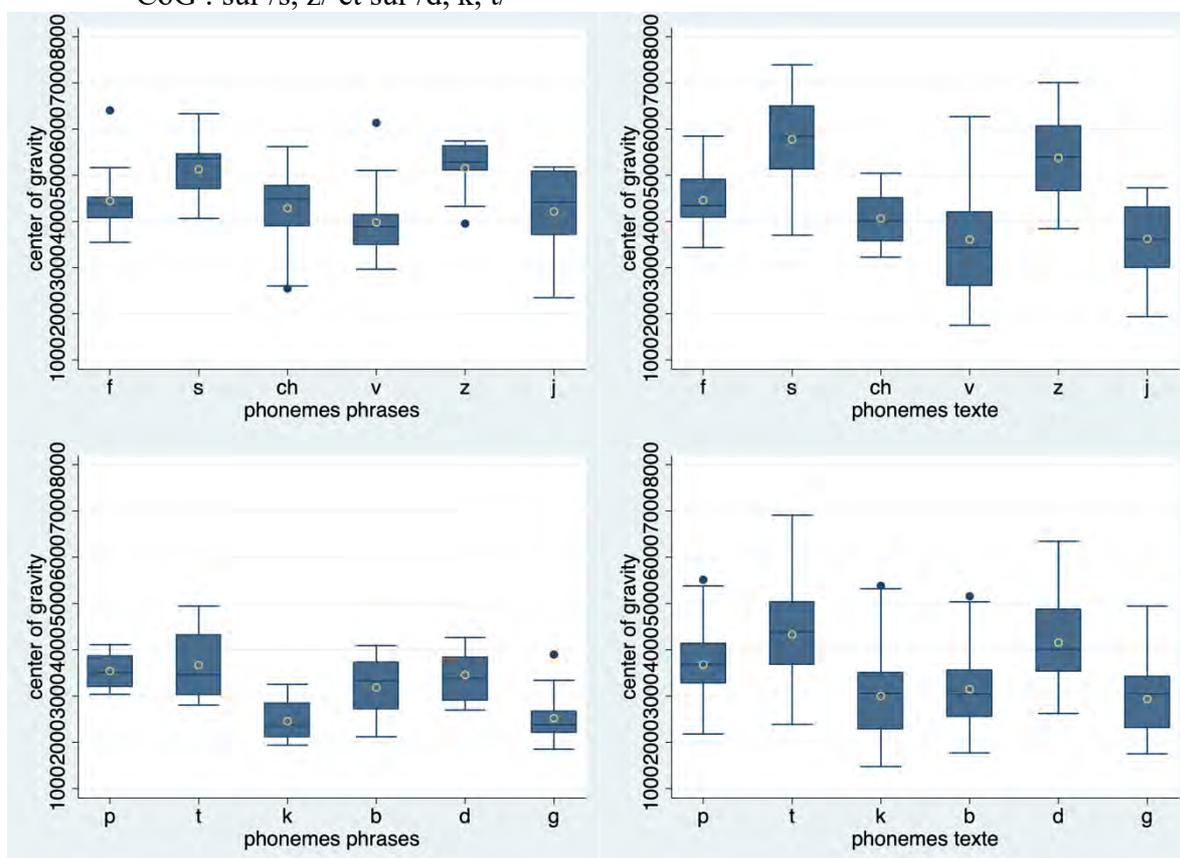
Question de recherche 1 :

« Y a-t-il une différence significative quant aux moments spectraux dans un contexte vocalique contrôlé (phrases porteuses) par rapport à un contexte phonémique plus varié (texte) ? »

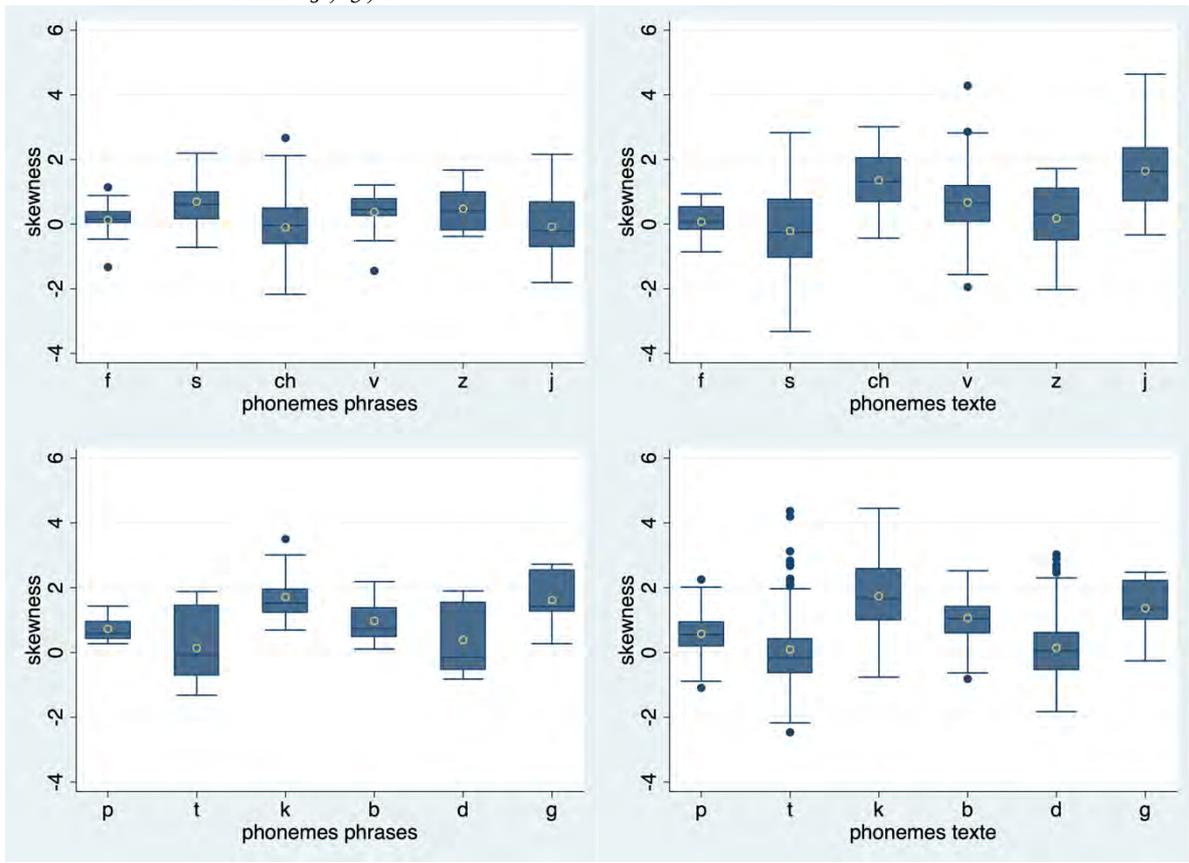
Analyse : Test des rangs signés de Wilcoxon (non paramétrique, échantillons appariés) pour comparer les valeurs des moments spectraux pour chaque phonème dans les phrases porteuses et dans le texte (valeurs moyennées sur les multiples répétitions dans le texte)

Résultats : Des différences significatives ont été trouvées entre les valeurs sur les phrases et sur le texte pour :

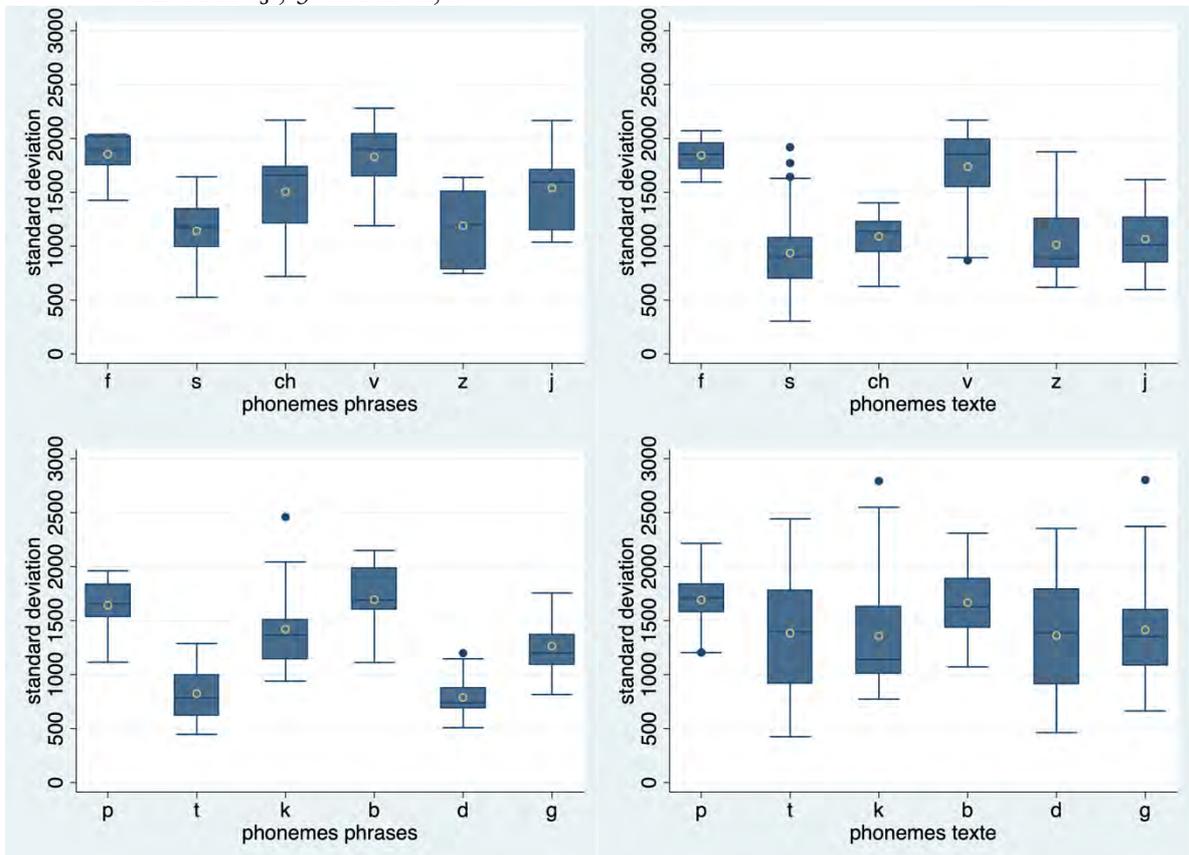
- CoG : sur /s, z/ et sur /d, k, t/



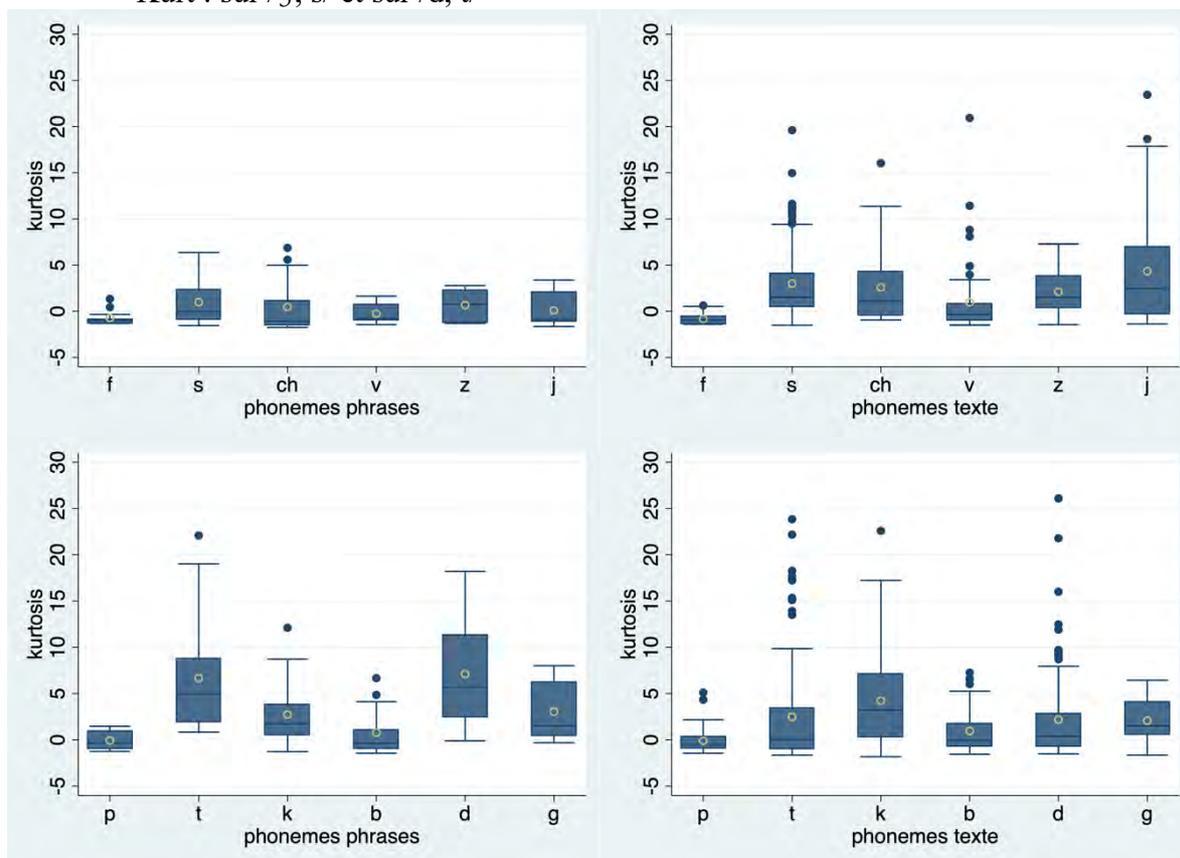
- SKEW : sur /ʃ, ʒ, s/



- SD : sur /ʃ, ʒ/ et sur /d, t/



- Kurt : sur /ʒ, s/ et sur /d, t/



Conclusions :

- Nous observons une plus grande variabilité des moments spectraux dans le texte, par rapport aux phrases.
- L'articulation des consonnes apico-alvéolaires (/t, d, s, z/) présente une énergie spectrale centroïde (CoG) légèrement plus élevée dans le texte ; les fricatives palato-alvéolaires (/ʃ, ʒ/) présentent un SKEW positif dans le texte, l'apico-alvéolaire /s/ un SKEW négatif (tendance inversée par rapport aux phrases).
- Les fricatives palato-alvéolaires (/ʃ, ʒ/) présentent un spectre légèrement plus pointu dans le texte (SD moins élevé, KURT plus élevé). Pour les plosives apico-alvéolaires /t, d/, nous observons des résultats contradictoires avec un SD plus élevé dans le texte, ainsi qu'un KURT également plus élevé. L'écart-type très large des valeurs de SD et de KURT sur ces deux phonèmes indique une grande variabilité des valeurs qui peut expliquer cette observation au vu de la taille réduite de l'échantillon.

Une analyse plus fine prenant en compte les différents contextes phonémiques dans le texte semble nécessaire pour mieux comprendre ces observations.

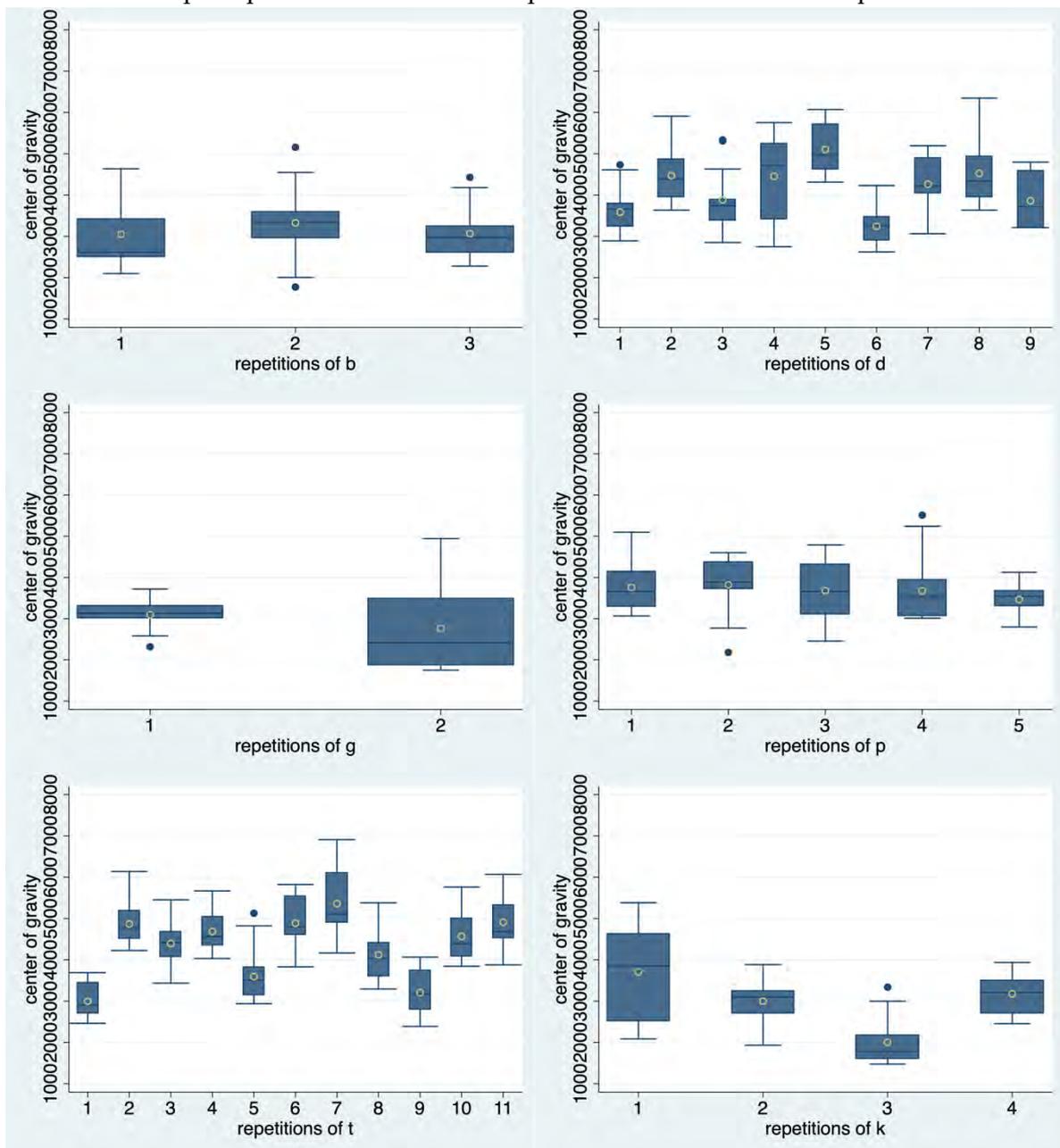
Question de recherche 2 :

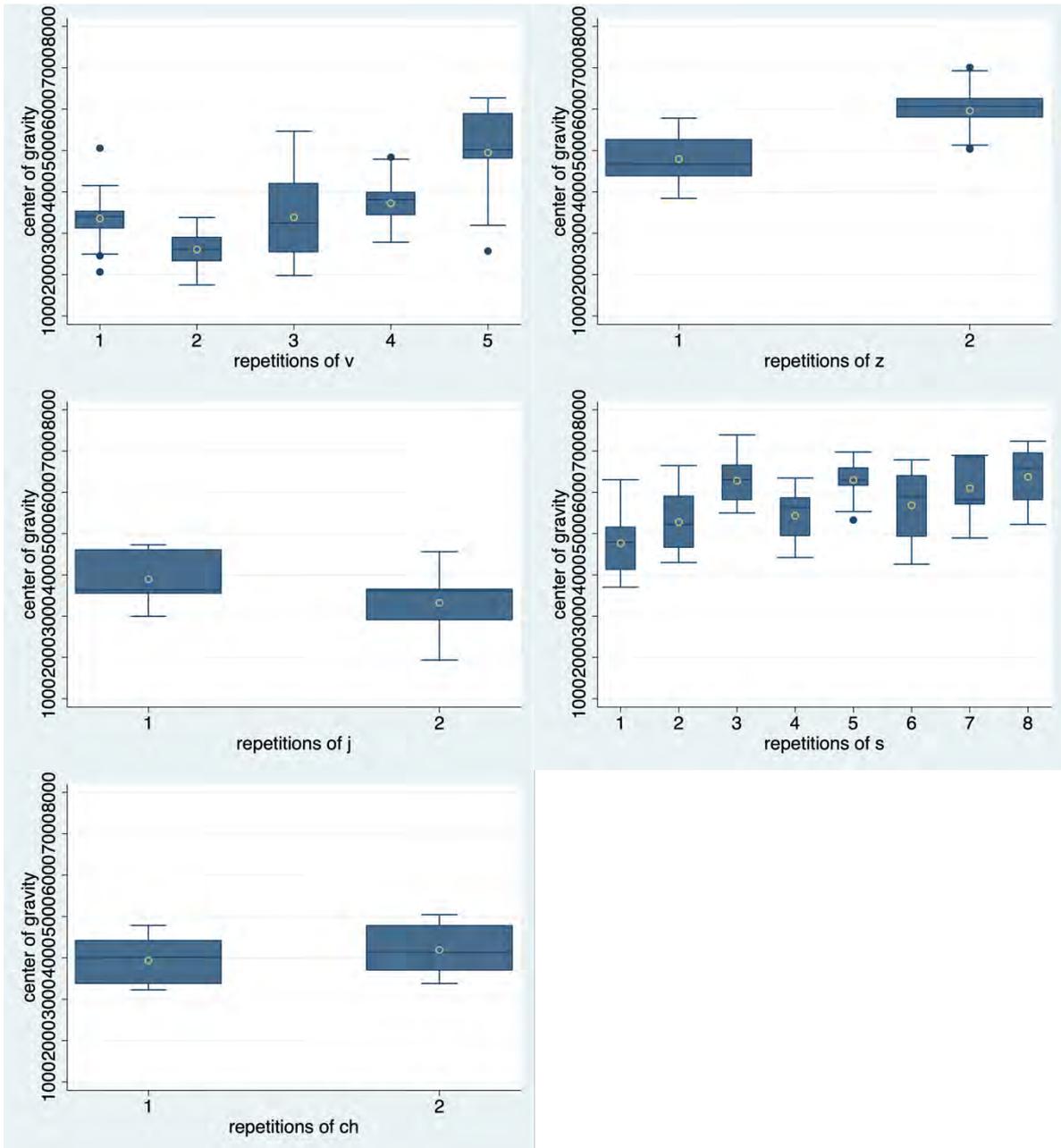
« Quelle est plus spécifiquement l'influence du contexte phonémique sur les moments spectraux dans les différentes répétitions d'un même phonème à l'intérieur du texte ? »

Analyse : Test de Friedman (non-paramétrique, échantillons appariés avec plus de deux niveaux) pour comparer les valeurs des moments spectraux pour chaque phonème dans les phrases porteuses et dans le texte (valeurs moyennées sur les multiples répétitions dans le texte)

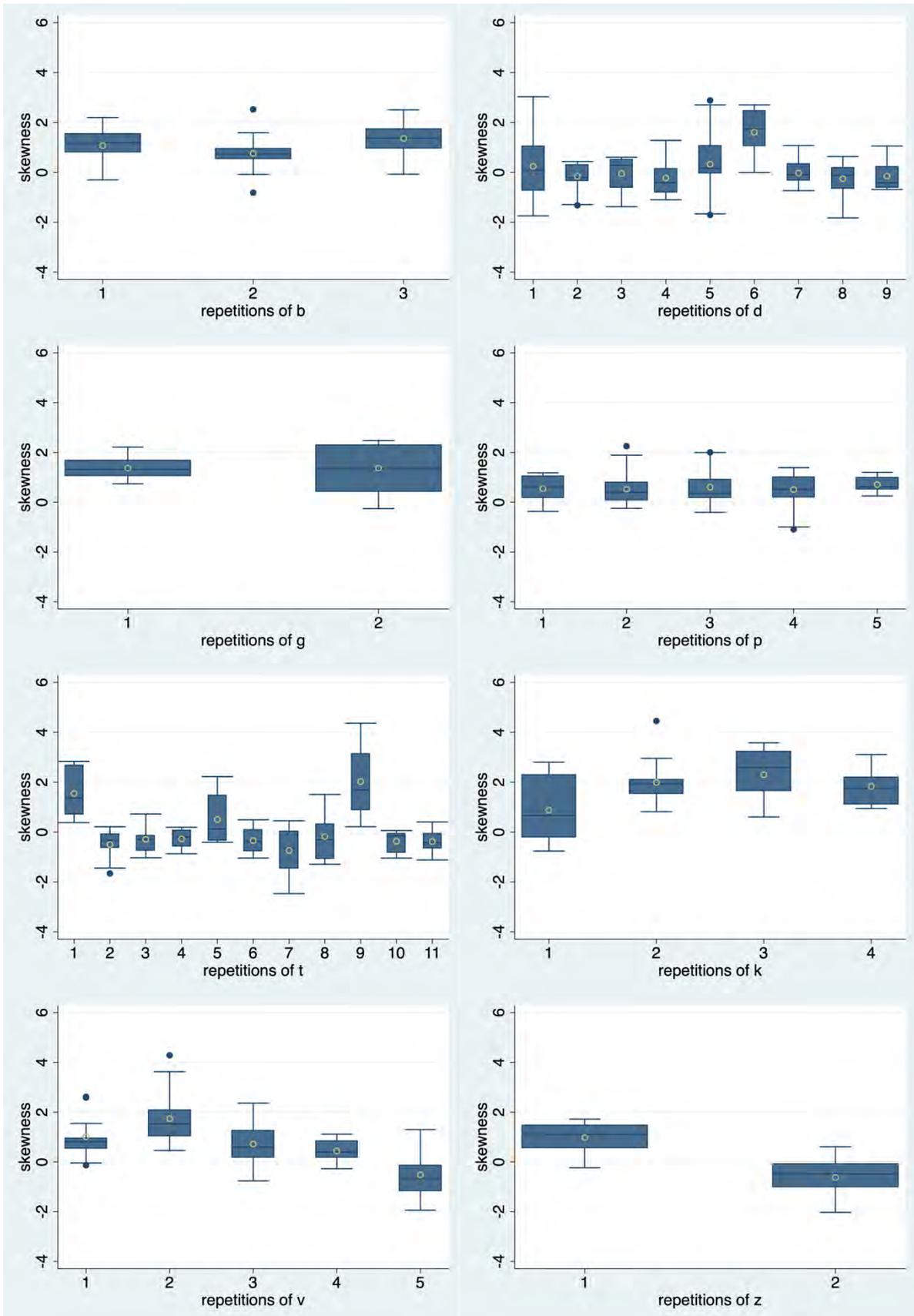
Résultats : Nous présentons tout d'abord les boxplots des quatre moments spectraux à travers les multiples occurrences consonantiques dans le texte. Notons que le /f/ n'apparaît qu'une fois dans ce texte et n'est donc pas représenté. Nous présentons ensuite l'analyse de l'influence des différents contextes phonémiques.

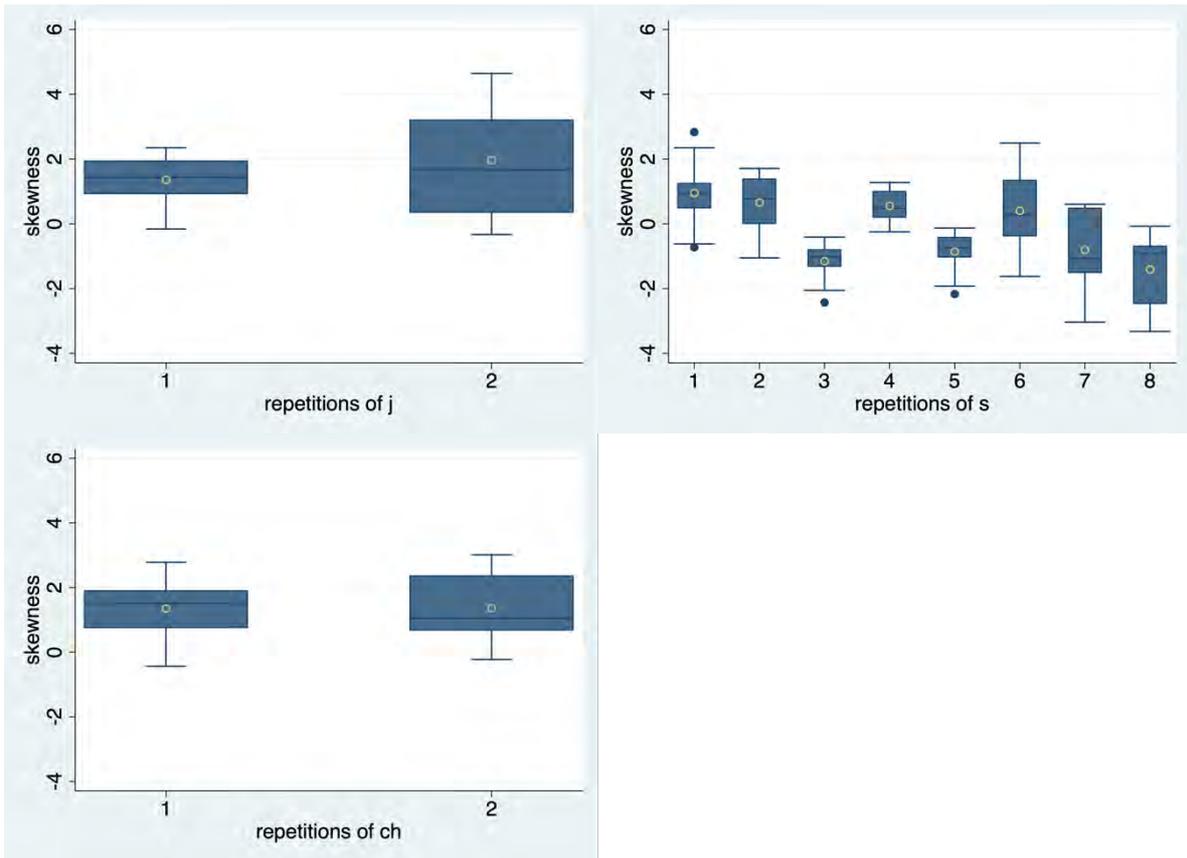
- Boxplots pour CoG dans les multiples occurrences consonantiques :



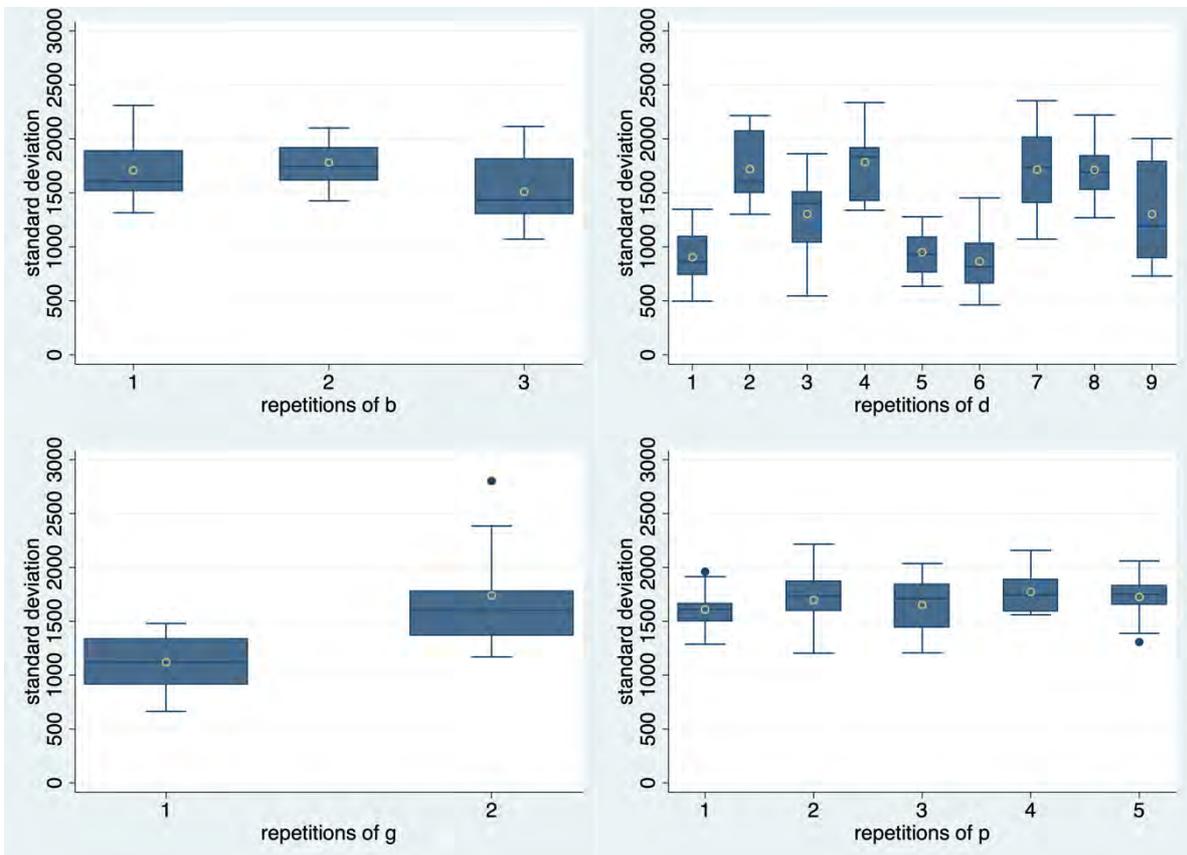


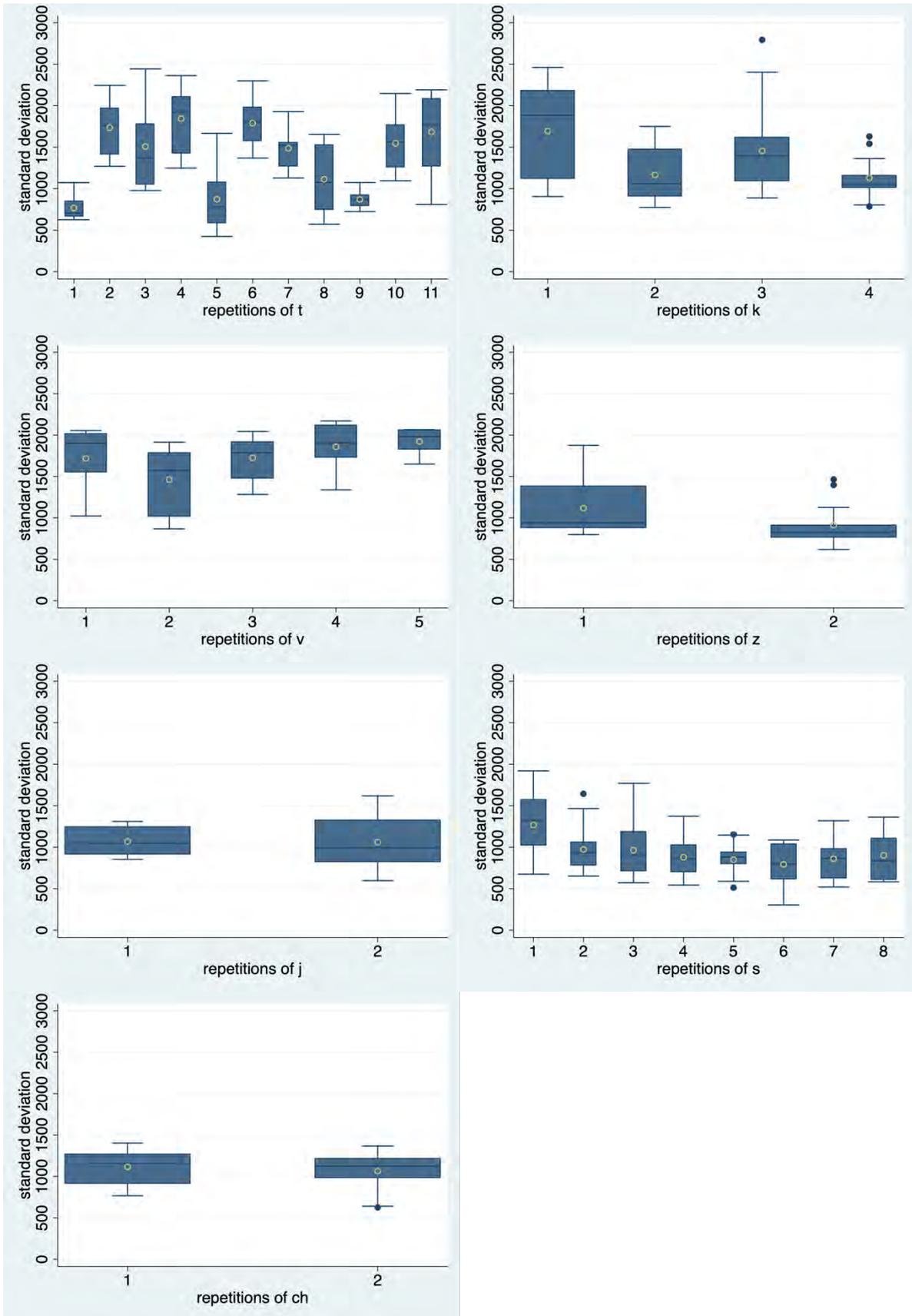
- Boxplots pour SKEW dans les multiples occurrences consonantiques :



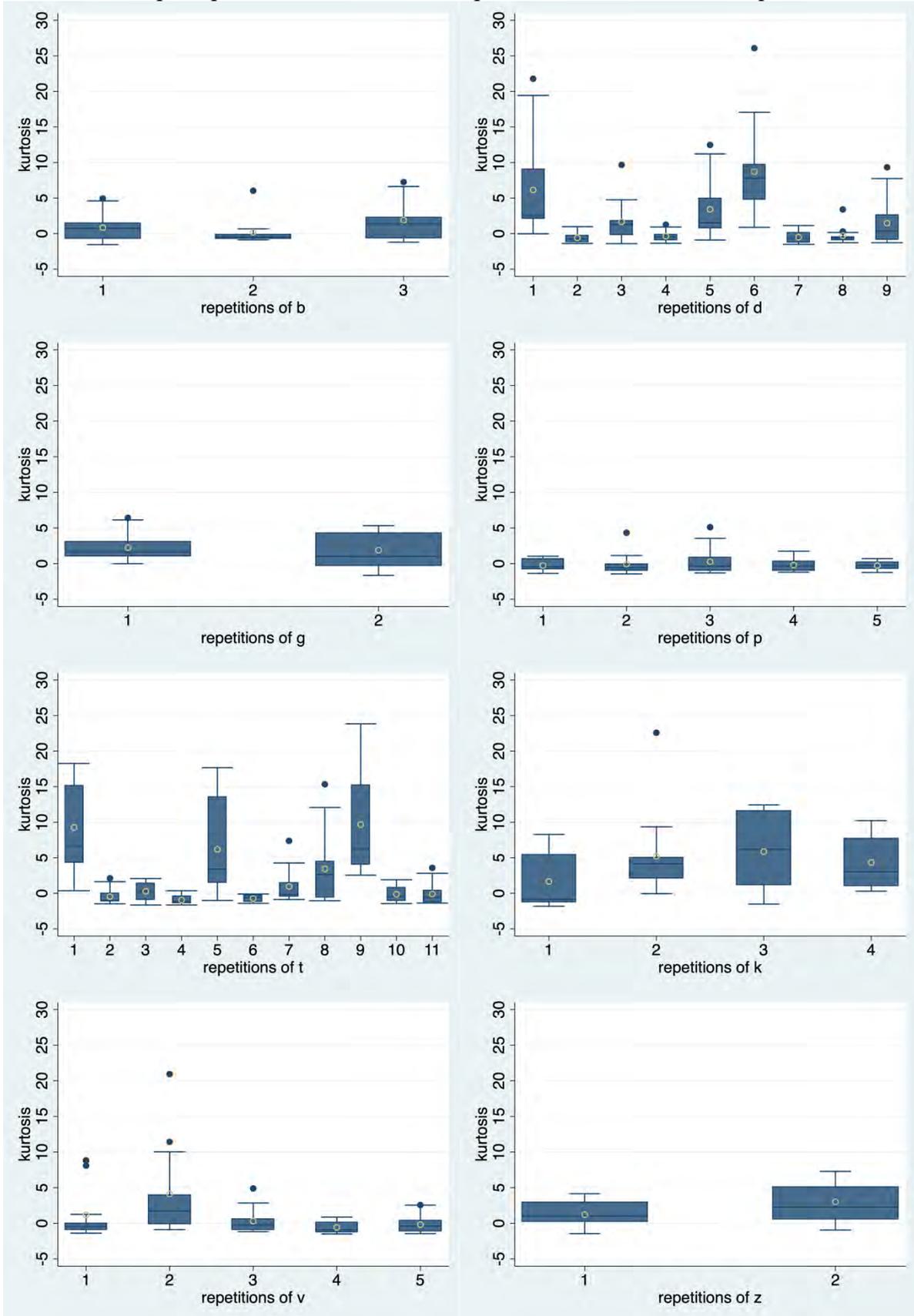


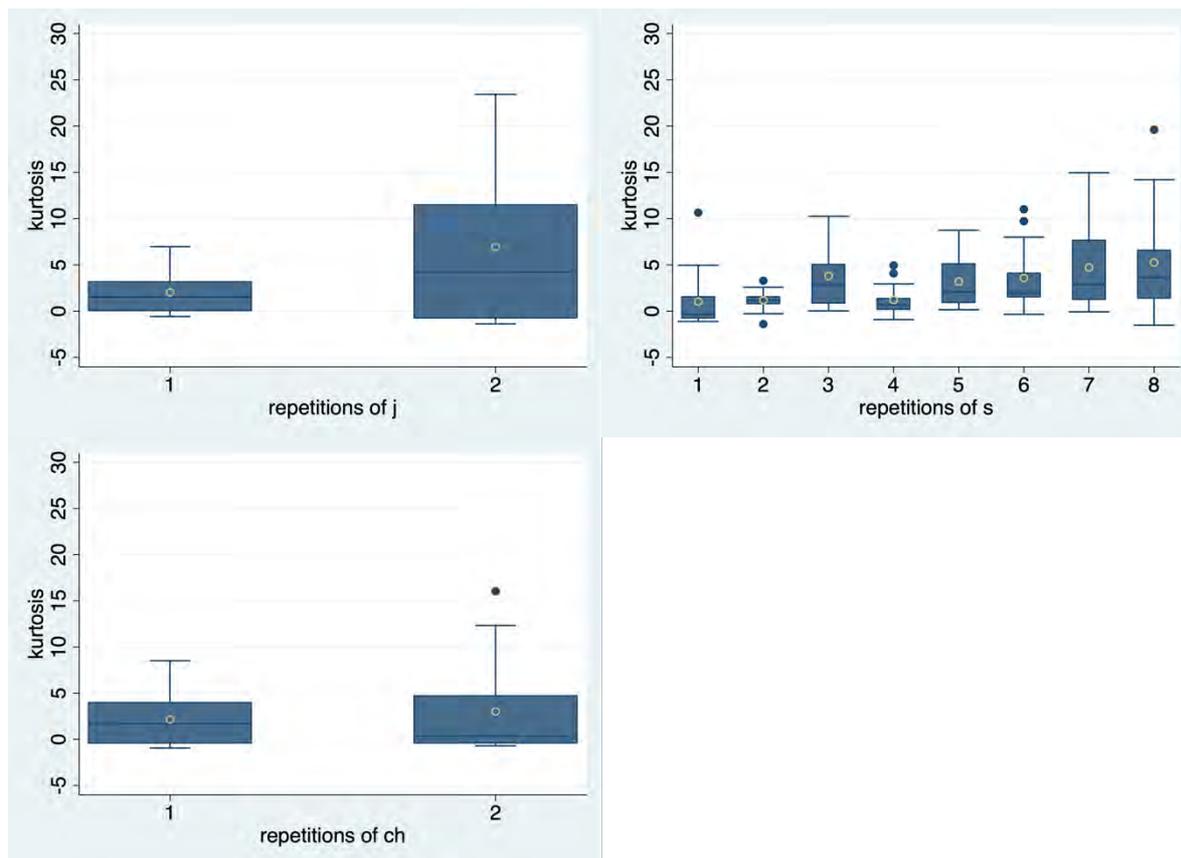
- Boxplots pour SD dans les multiples occurrences consonantiques :





- Boxplots pour KURT dans les multiples occurrences consonantiques :





Analyse de l'influence des contextes phonémique

Pour analyser l'influence des contextes phonémiques des différentes occurrences de chaque consonne dans le texte, nous avons mis en commun les résultats du test de Friedman avec les données descriptives (médianes de chaque moment spectral dans chaque occurrence consonantique, avec aide visuelle des boxplots) ainsi que l'observation des phonèmes qui entourent chaque occurrence. Nous allons maintenant présenter les résultats des analyses statistiques et des observations qualitatives (aidées de codes couleurs), pour chaque consonne.

Légende :

- Labialisation (« Labial. ») : - = non, + = oui
- Aperture : ++ = ouvert, + = semi-ouvert, 0 = neutre, - = semi-fermé, -- = fermé
- Nasalisation (« Nasal ») : - = non, + = oui
- Antériorité (« Antérieur. ») : ++ = antérieur max., + = antérieur, 0 = central, - = postérieur, -- = postérieur max.
- Occ. = occurrence
- \lrcorner = silence précédant ou suivant la consonne cible (p. ex. consonne en début de phrase).

- /b/ :

		occ.			occ.		
	œ	1	o	i	2	ε	
Labial.	+		+		-		-
Aperture	+		-	---			+
Nasal.	+		-		-		-
Antérieur.	+		--	++			++

Test de Friedman – différences à $\alpha=0,05$: aucune

Observation qualitative : Les bilabiales semblent relativement peu influencées par le contexte phonémique (/p).

- /d/⁴⁰ :

		occ.			occ.			occ.			occ.		
	y	1	œ/ø	ε r	2	ε/e	ɔr	3]	ε/e	4	ã	
Labial.	+		+	-		-	+			-		+	
Aperture	--		+/-			+/-				+/-		++	
Nasal.	-		-	-		-	-			-		+	
Antérieur.	+		+	--		++	--			++		--	

		occ.			occ.			occ.			occ.			occ.	
	ε s	5	œ/ø	œ r	6	y	i]	7	ε/e	ẽ	8	e	ã	9	ã
Labial.	-		+	+		+	-		-	-		-	+		+
Aperture	-		+/-			--			+/-	+		-	++		++
Nasal.	-		-	-		-	-		-	+		-	+		+
Antérieur.	+		+	+		+	++		++	+		++	--		--

Test de Friedman – différences à $\alpha=0,05$:

- CoG : 1≠2 ; 2≠3 ; 4≠5 ; 5≠6 ; 6≠7 ; 8≠9 ; 1≠4 ; 1≠5 ; 1≠7 ; 1≠8 ; 2≠5 ; 2≠6 ; 2≠9 ; 3≠5 ; 3≠6 ; 4≠6 ; 4≠9 ; 5≠7 ; 5≠8 ; 5≠9 ; 6≠8 ; 6≠9 ; 7≠9

→ groupements : 1-3-6 (bas) ; 2-7-8 (élevés) ; 4 très variable ; 5 se différencie de tous (élevé) ; 6 se différencie de tous (bas) sauf de 1

Observation qualitative : CoG plus bas si contexte labialisé ; CoG plus élevé si contexte non labialisé ; influence du /s/ (qui a lui-même un CoG élevé) ?

- SKEW : 4≠5 ; 5≠6 ; 6≠7 ; 1≠6 ; 2≠6 ; 3≠6 ; 4≠6 ; 6≠8 ; 6≠9 : 6 se différencie de tous (élevé)

		occ.			occ.			occ.			occ.		
	y	1	œ/ø	ε r	2	ε/e	ɔr	3]	ε/e	4	ã	
Labial.	+		+	-		-	+			-		+	
Aperture	--		+/-			+/-				+/-		++	
Nasal.	-		-	-		-	-			-		+	
Antérieur.	+		+	--		++	--			++		--	

⁴⁰ L'occurrence de /t/ et de /d/ dans « il les perdait toutes de la même façon » a été exclue des analyses, étant donné la variabilité des prononciations (parfois /t u t d ə/ avec rupture, parfois « /t u d ə/ » avec fusion du /t/ dans le /d/).

	occ.		occ.		occ.		occ.		occ.
ε	5	œ/ø	6	y	7	ε/e	8	e	9
s				il/ɹ					
Labial.	-	+	+	+	-	-	-	-	+
Aperture	-	+/-		--		+/-	+	-	++
Nasal.	-	-	-	-	-	-	+	-	+
Antérieur.	+	+	+	+	++	++	+	++	--

Test de Friedman – différences à α=0,05 :

- SD : 1≠2 ; 2≠3 ; 3≠4 ; 4≠5 ; 6≠7 ; 8≠9 ; 1≠3 ; 1≠4 ; 1≠7 ; 1≠8 ; 1≠9 ; 2≠5 ; 2≠6 ; 2≠9 ; 3≠5 ; 3≠6 ; 3≠7 ; 3≠8 ; 4≠9 ; 5≠7 ; 5≠8 ; 5≠9 ; 6≠8 ; 6≠9 ; 7≠9

→ groupements : 1-5-6 (bas) ; 2-4-7-8 (élevés) ; 3-9 (moyens)

Observation qualitative : SD plus bas si contexte labialisé ; SD plus élevé si contexte non labialisé ; plus élevé avec ε/e (antérieurs et non labialisés)

- KURT : 1≠2 ; 2≠3 ; 4≠5 ; 5≠6 ; 6≠7 ; 1≠3 ; 1≠4 ; 1≠7 ; 1≠8 ; 1≠9 ; 2≠5 ; 2≠6 ; 2≠9 ; 3≠6 ; 3≠7 ; 4≠6 ; 5≠7 ; 5≠8 ; 6≠8 ; 6≠9 ; 7≠9

→ groupements : 1-5-6 (élevés) ; 2-3-4-7-8-9 (bas) ; 6 se différencie de tous (élevé) sauf 1 ; 1 très variable

- /g/ :

	occ.		occ.
œ/ø	1	ẽ	2
Labial.	+	-	+
Aperture	+/-	+	+/-
Nasal.	-	+	-
Antérieur.	+	+	-

Test de Friedman – différences à α=0,05 :

- CoG : aucune (2 plus variable)
- SKEW : aucune (2 plus variable)
- SD : 1 plus bas que 2
- KURT : aucune

Observation qualitative : Le phonème qui suit a peu d'influence sur le /g/ (étant donné que le précédent est identique) ? Ou les palato-vélaires sont relativement peu influencées par le contexte phonémique ?

- /p/ :

	occ.		occ.		occ.		occ.		occ.
ε/e	1	ε	2	œ	3	a	4	ã	5
Labial.	-	-	-	+	-	-	-	+	+
Aperture	+/-	+	++	+	+/-	++	-	++	--
Nasal.	-	-	-	-	-	-	-	+	-
Antérieur.	++	++	+	+	++	+	++	--	--

Test de Friedman – différences à $\alpha=0,05$:

- **CoG** : aucune
- **SKEW** : aucune
- **SD** : 1 \neq 4
- **KURT** : aucune

Observation qualitative : Les bilabiales semblent relativement peu influencées par le contexte phonémique (/ / b).

- /t/ :

		occ.		occ.		occ.		occ.		occ.		occ.		occ.				
	ϵ/e	1	u	a	2	$\tilde{\epsilon}$	$\tilde{\text{õ}}$	3	a	ϵ	4	r	$\text{œ}/\text{ø}$	5	$\text{œ}/\text{ø}$	ϵ/e	6	ϵ/e
Labial.	-		+	-		-	+		-		-		+		+	-		-
Aperture	+/-		--	++		+	+		++	+			+/-		+/-	+		+/-
Nasal.	-		-	-		+	+		-	-			-		-	-		-
Antérieur.	++		--	+		+	--		+	++			+		+	++		++

		occ.		occ.		occ.		occ.		occ.		occ.					
	ϵ/e	7	i	$\tilde{\text{ã}}$	8	$\text{ɹ}/\text{v}$	u	a	9	u	$\tilde{\text{ã}}$	10	ϵ	ϵ	r	11	e
Labial.	-		-	+		-	+		-	+		-	-		-		-
Aperture	+/-		--	++		++			--	++		++	+				-
Nasal.	-		-	+		-	-		-	+		-	-		-		-
Antérieur.	++		++	--		-	--		+	--		--	++		-		++

Test de Friedman – différences à $\alpha=0,05$:

- **CoG** : 1 \neq 2 ; 4 \neq 5 ; 5 \neq 6 ; 6 \neq 7 ; 7 \neq 8 ; 8 \neq 9 ; 9 \neq 10 ; 1 \neq 3 ; 1 \neq 4 ; 1 \neq 5 ; 1 \neq 6 ; 1 \neq 7 ; 1 \neq 8 ; 1 \neq 10 ; 1 \neq 11 ; 2 \neq 5 ; 2 \neq 8 ; 2 \neq 9 ; 3 \neq 5 ; 3 \neq 6 ; 3 \neq 7 ; 3 \neq 9 ; 3 \neq 11 ; 4 \neq 7 ; 4 \neq 8 ; 4 \neq 9 ; 5 \neq 7 ; 5 \neq 8 ; 5 \neq 9 ; 5 \neq 10 ; 5 \neq 11 ; 6 \neq 8 ; 6 \neq 9 ; 7 \neq 9 ; 7 \neq 10 ; 8 \neq 10 ; 8 \neq 11 ; 9 \neq 11

→ **groupements** : 1-9 (bas), se différencie de tous les autres ; 5 se différencie de tous (bas) ; 8 se différencie de tous (moyen) sauf de 3 ; 2-3-4-6-10-11 (élevés) ; 7 se différencie de tous (élevé+) sauf de 2 et 11

Observation qualitative : **CoG plus bas** quand suivi d'une voyelle labialisée (1, 5, 9) ; **CoG plus élevé** quand suivi d'une voyelle non labialisée

- **SKEW** : 1 \neq 2 ; 5 \neq 6 ; 7 \neq 8 ; 8 \neq 9 ; 9 \neq 10 ; 1 \neq 3 ; 1 \neq 4 ; 1 \neq 5 ; 1 \neq 6 ; 1 \neq 7 ; 1 \neq 8 ; 1 \neq 10 ; 1 \neq 11 ; 2 \neq 4 ; 2 \neq 5 ; 2 \neq 9 ; 3 \neq 9 ; 4 \neq 9 ; 5 \neq 7 ; 5 \neq 8 ; 5 \neq 9 ; 5 \neq 10 ; 6 \neq 9 ; 7 \neq 9 ; 9 \neq 11

→ **groupements** : 1-9 (élevés et variables) se différencie de tous les autres ; 2-3-4-6-7-8-10-11 (bas) ; 5 et 7 plus variables

Observation qualitative : **SKEW plus élevé** quand suivi d'une voyelle labialisée (1, 5, 9) ; **SKEW plus bas** quand suivi d'une voyelle non labialisée

- **SD** : 1 \neq 2 ; 3 \neq 4 ; 4 \neq 5 ; 5 \neq 6 ; 7 \neq 8 ; 9 \neq 10 ; 1 \neq 3 ; 1 \neq 4 ; 1 \neq 6 ; 1 \neq 7 ; 1 \neq 8 ; 1 \neq 9 ; 1 \neq 10 ; 1 \neq 11 ; 2 \neq 5 ; 2 \neq 8 ; 2 \neq 9 ; 2 \neq 10 ; 3 \neq 5 ; 3 \neq 9 ; 4 \neq 7 ; 4 \neq 8 ; 4 \neq 9 ; 4 \neq 10 ; 5 \neq 7 ; 5 \neq 10 ; 5 \neq 11 ; 6 \neq 8 ; 6 \neq 9 ; 6 \neq 10 ; 7 \neq 9 ; 8 \neq 10 ; 8 \neq 11 ; 9 \neq 11

→ **groupements** : 1-5-8-9 (bas, 8 variable) ; 2-3-4-6-7-10-11 (plus élevés, variabilité)

Observation qualitative : **SD plus bas** quand suivi d'une voyelle labialisée (1, 5, 9) ; **SD plus élevé** quand suivi d'une voyelle non labialisée

- **KURT** : 1 \neq 2 ; 3 \neq 4 ; 4 \neq 5 ; 5 \neq 6 ; 6 \neq 7 ; 8 \neq 9 ; 9 \neq 10 ; 1 \neq 3 ; 1 \neq 4 ; 1 \neq 6 ; 1 \neq 7 ; 1 \neq 8 ; 1 \neq 10 ; 1 \neq 11 ; 2 \neq 5 ; 2 \neq 8 ; 2 \neq 9 ; 3 \neq 5 ; 3 \neq 6 ; 3 \neq 8 ; 3 \neq 9 ; 4 \neq 7 ; 4 \neq 8 ; 4 \neq 9 ; 4 \neq 10 ; 4 \neq 11 ; 5 \neq 7 ; 5 \neq 10 ; 5 \neq 11 ; 6 \neq 8 ; 6 \neq 9 ; 7 \neq 9 ; 8 \neq 10 ; 8 \neq 11 ; 9 \neq 11

→ **groupements** : 1-5-9 (élevés, variables) ; 8 moyen ; 2-3-4-6-7-10-11 (bas)

Observation qualitative : KURT plus élevé quand suivi d'une voyelle labialisée (1,5,9) ; KURT plus bas quand suivi d'une voyelle non labialisée

- /k/ :

	ε	occ. 1	sε/e	εl	occ. 2	a	œr	occ. 3	ɔ	ε/e	occ. 4	a
Labial.	-		-	-		-	+		+	-		-
Aperture	++		-			++			+	+/-		++
Nasal.	-		-	-		-	-		-	-		-
Antérieur.	++		+			+	-		--	++		+

Test de Friedman – différences à α=0,05 :

- CoG : 2≠3 ; 3≠4 ; 1≠3

→ groupements : 3 plus bas ; 1-2-4 (1 très variable)

- SKEW : aucune (1 très variable)

Observation qualitative : CoG plus bas si contexte labialisé et postérieur

- SD : 1≠2 ; 1≠4

→ groupements : 1 plus élevé (mais très variable) ; influence du s ?

- KURT : aucune

- /v/ :

	a	occ. 1	ε/e	a	occ. 2	ε	ε	occ. 3	r]	ε	occ. 4	r œ] u	occ. 5
Labial.	-		-	-		-	-		-	-		+	+	+
Aperture	++		+/-	++		+	+		-	+		-	-	--
Nasal.	-		-	-		-	-		-	-		-	-	-
Antérieur.	+		++	+		++	++			++		-	--	--

Test de Friedman – différences à α=0,05 :

- CoG : 1≠2 ; 4≠5 ; 1≠5 ; 2≠4 ; 2≠5 ; 3≠5

→ groupements : 1-3-4 (bas) ; 5 diffère de tous (élevé) ; 2 diffère (bas) de tous sauf 3 (qui est + variable)

Observation qualitative : CoG plus élevé quand contexte labialisé et postérieur ; CoG plus bas quand contexte non-labialisé et antérieur

- SKEW : 2≠3 ; 4≠5 ; 1≠5 ; 2≠4 ; 2≠5 ; 3≠5

→ groupements : 1-3-4 ; 5 diffère de tous (bas) ; 2 diffère (élevé) de tous sauf 1

Observation qualitative : SKEW plus bas quand contexte labialisé et postérieur

- SD : 2≠4 ; 2≠5

→ groupements : 1-3-4-5 ; 2 plus bas mais variable

- KURT : 3≠4 ; 2≠4

→ groupements : 2 plus élevé mais variable

- /z/ :

		occ.			occ.		
	ε/e	1	y		œ/ø	2	ẽ
Labial.	-				+		
Aperture	+/-				--		
Nasal.	-				-		
Antérieur.	++				+		

Test de Friedman – différences à α=0,05 :

- **CoG** : 2 plus élevé

Observation qualitative : **CoG plus élevé si suivi d'une voyelle non-labialisée, ouverte et nasale**

- **SKEW** : 2 plus bas

Observation qualitative : **SKEW plus bas si suivi d'une voyelle non-labialisée, ouverte et nasale**

- **SD** : 2 plus bas

- **KURT** : aucune

- /ʒ/ :

		occ.			occ.		
	ε/e	1	a		ã	2	ε/e
Labial.	-				-		
Aperture	+/-				++		
Nasal.	-				-		
Antérieur.	++				+		

Test de Friedman – différences à α=0,05 :

- **CoG** : 2 plus bas

Observation qualitative : **CoG plus bas si précédé d'une voyelle postérieure, labialisée et nasale**

- **SKEW** : aucune (2 plus variable)

- **SD** : aucune

- **KURT** : aucune (2 plus variable)

- /s/ :

		occ.		occ.		occ.		occ.	
	œ/ø	1	jø ø	2	œ/ø ε k	3	ε/e a	4	õ
Labial.	+								
Aperture	+/-								
Nasal.	-								
Antérieur.	+								

		occ.		occ.		occ.		occ.	
	a	5	ε/e	6	ã	7	d œ/ø	8	e
Labial.	-								
Aperture	++								
Nasal.	-								
Antérieur.	+								

Test de Friedman – différences à $\alpha=0,05$:

- **CoG** : 1≠2 ; 2≠3 ; 3≠4 ; 4≠5 ; 5≠6 ; 6≠7 ; 1≠3 ; 1≠4 ; 1≠5 ; 1≠6 ; 1≠7 ; 1≠8 ; 2≠5 ; 2≠7 ; 2≠8 ; 3≠6 ; 4≠7 ; 4≠8 ; 6≠8

→ **groupements** : 1 diffère de tous (bas) ; 2-4-6 (moyen) ; 3-5-7-8 (élevés)

Observation qualitative : CoG plus élevé si contexte non labialisé ; CoG plus bas en contexte labialisé ; influence de la semi-voyelle j ?

- **SKEW** : 2≠3 ; 3≠4 ; 4≠5 ; 5≠6 ; 6≠7 ; 1≠3 ; 1≠5 ; 1≠7 ; 1≠8 ; 2≠5 ; 2≠7 ; 2≠8 ; 3≠5 ; 3≠6 ; 4≠7 ; 4≠8 ; 6≠8

→ **groupements** : 1-2-4-6 (haut) ; 3-5-7-8 (bas)

Observation qualitative : SKEW plus bas si contexte non labialisé ; SKEW plus élevé si contexte labialisé ; influence de la semi-voyelle j ?

- **SD** : 1≠2 ; 1≠4 ; 1≠5 ; 1≠6 ; 1≠7 ; 1≠8

→ 1 plus élevé, les autres tous similaires

Observation qualitative : influence de la semi-voyelle j ?

- **KURT** : 2≠3 ; 3≠4 ; 1≠8 ; 2≠5 ; 2≠7 ; 2≠8 ; 4≠6 ; 4≠7 ; 4≠8

→ 1 plus bas, 7-8 plus élevés et variables ; 2-4 plus bas et peu variables

Observation qualitative : influence de la semi-voyelle j ?

- /ʃ/ :

		occ.			occ.	
	ε/e	1	ε	ε/e	2	ε
Labial.	-		-	-		-
Aperture	+/-		+	+/-		+
Nasal.	-		-	-		-
Antérieur.	++		++	++		++

Test de Friedman – différences à $\alpha=0,05$:

- **CoG** : 2 plus élevé
- **SKEW** : aucune
- **SD** : aucune
- **KURT** : aucune

Observation qualitative : Les valeurs sont similaires lorsque le contexte phonémique est identique.

Conclusions :

- Un contexte phonémique labialisé semble entraîner une baisse de la composante spectrale principale (CoG plus bas, Skew plus élevé) pour toutes les consonnes analysées sauf le /v/. De plus, les spectres semblent plus pointus (SD plus bas, KURT plus élevé) lorsque la consonne se présente en contexte labialisé, pour les apico-alvéolaires /d/ et /t/.
- Un contexte phonémique nasalisé ainsi que l'inclusion de la consonne dans un cluster consonantique entraînent une plus grande variabilité des moments spectraux.
- L'influence de l'antériorité des phonèmes voisins — théoriquement liée notamment au CoG — reste à analyser plus en détail sur un plus grand échantillon de données et sur des contextes contrôlés à cet effet.
- Aucune différence significative sur les moments spectraux à travers les différentes occurrences n'est observée pour les plosives bilabiales, ni pour le /g/. Nous n'avons donc pas assez de preuves pour conclure une influence du contexte sur les valeurs des moments spectraux dans ces phonèmes.

Une connaissance plus approfondie de l'influence du contexte phonémique permettrait de nuancer les interprétations des analyses spectrales. Cette étude est une étude pilote sur le sujet. Des études avec un échantillon plus large de locuteurs ainsi que de plus nombreuses occurrences phonémiques dans des contextes davantage contrôlés permettraient de consolider les analyses d'influence du contexte phonémique. De plus, l'utilisation d'un algorithme de segmentation automatique aidé d'un alignement forcé, par sa systématique, permettrait de limiter la variabilité due à la segmentation manuelle.

Toutefois, nos résultats préliminaires indiquent une influence quasi certaine du contexte vocalique, même si la nature de cette dernière reste à préciser. Ces observations nous amènent à préconiser l'utilisation de phrases porteuses avec un contexte vocalique fixe et contrôlé pour l'évaluation clinique.

Annexe P – Algorithme de détection de fenêtre d'analyse des moments spectraux pour les consonnes non-sonantes

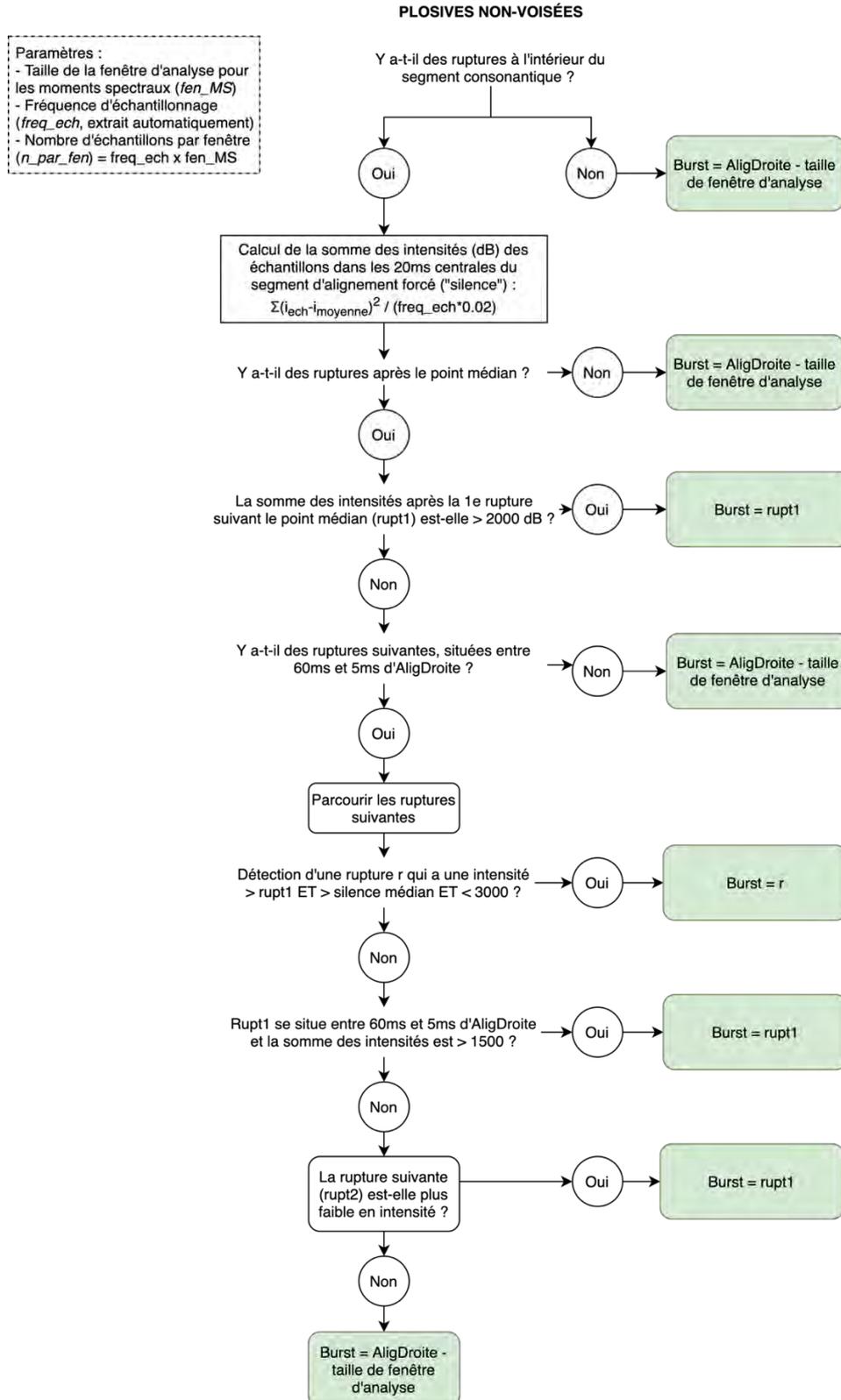


Figure 78 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux (c.-à-d. du burst) pour les plosives non voisées ; *AligDroite* = frontière droite du segment consonantique de l'alignement forcé

Paramètres :
- Taille de la fenêtre d'analyse pour les moments spectraux (*fen_MS*)

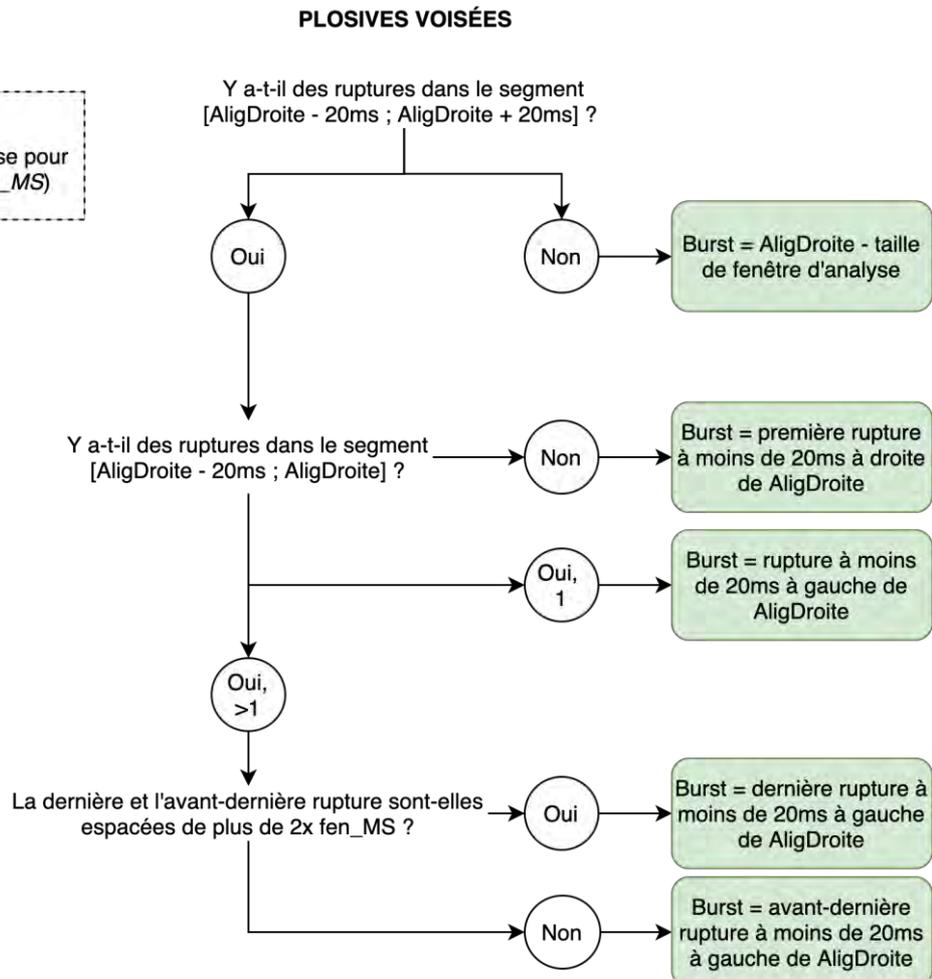


Figure 79 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux (c.-à-d. du burst) pour les plosives voisées ; AligDroite = frontière droite du segment consonantique de l'alignement forcé

FRICATIVES NON-VOISÉES

Paramètres :
- Taille de la fenêtre d'analyse pour les moments spectraux (*fen_MS*)

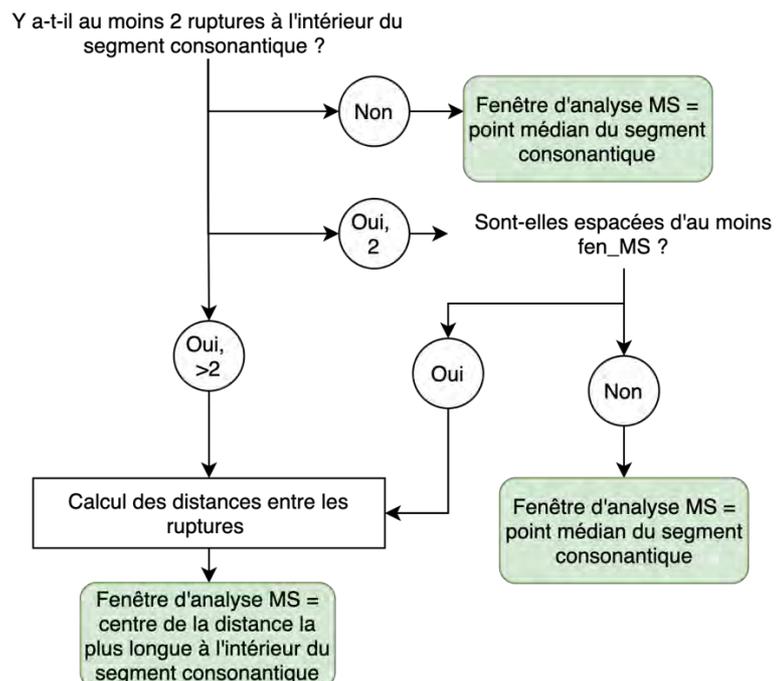


Figure 80 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux pour les fricatives non voisées

FRICATIVES VOISÉES

Paramètres :
- Taille de la fenêtre d'analyse pour les moments spectraux (*fen_MS*)

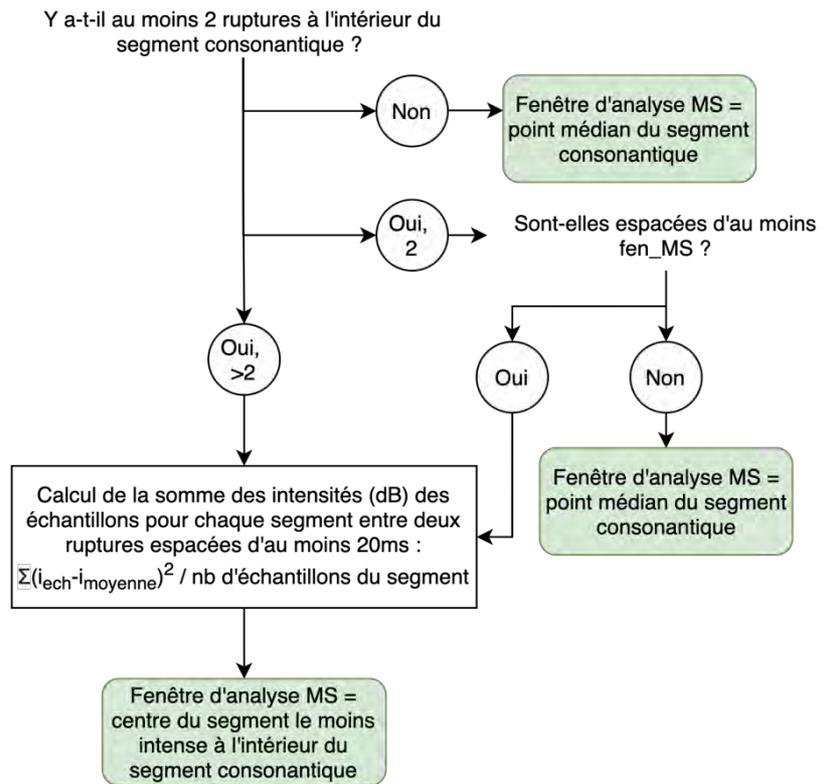
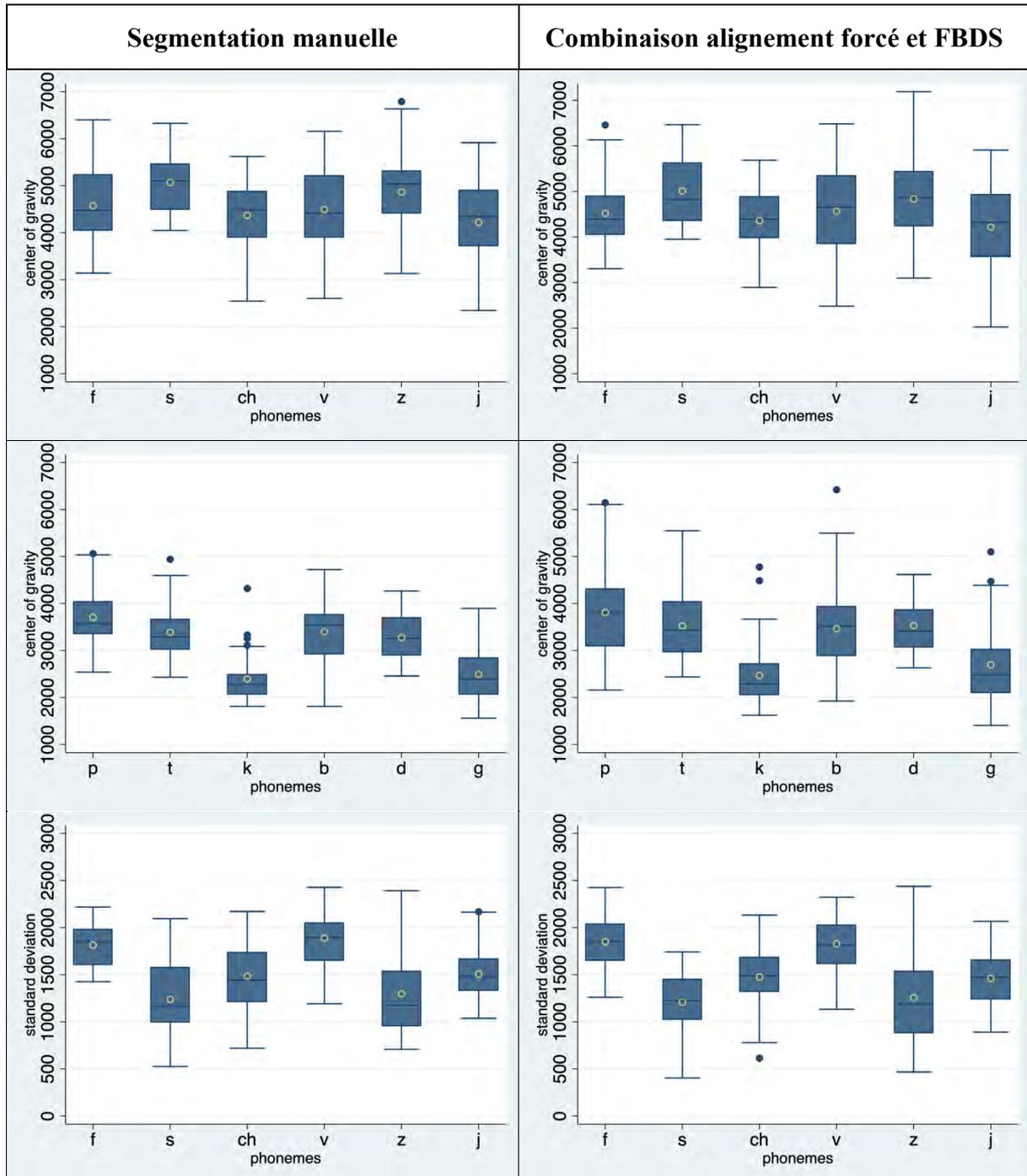
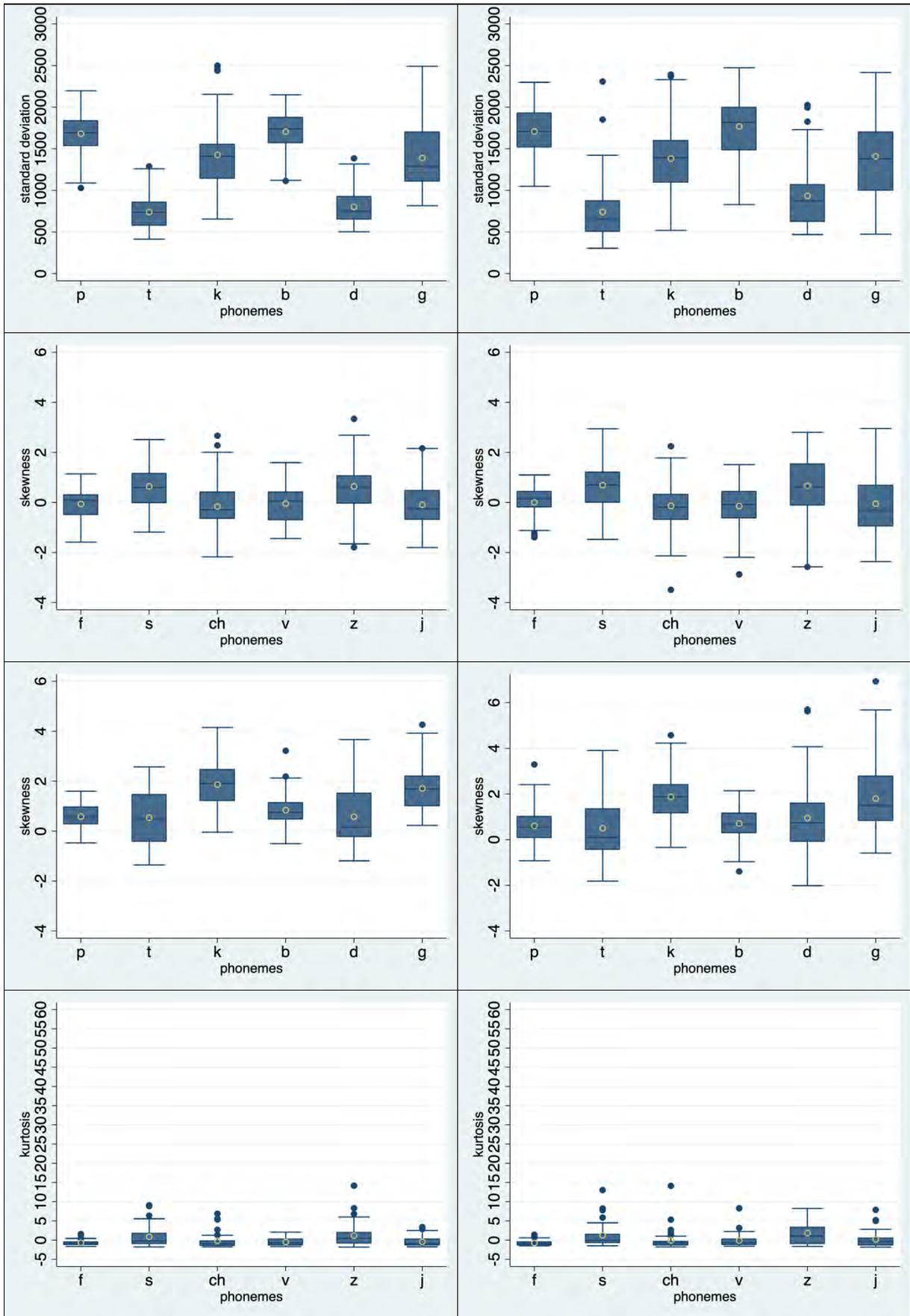


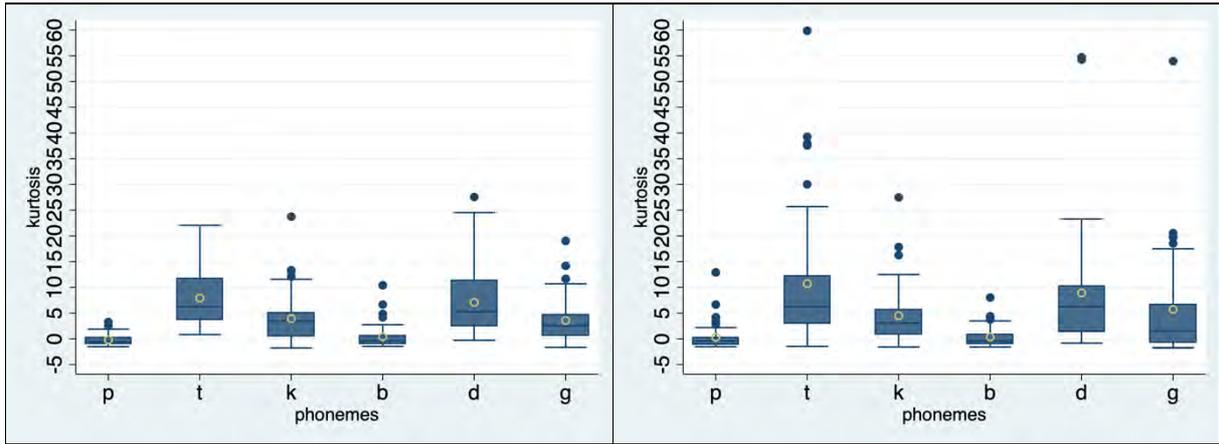
Figure 81 – Algorithme de détection de la fenêtre optimale d'analyse des moments spectraux pour les fricatives voisées

Annexe Q – Boîtes à moustaches des valeurs des moments spectraux issus des deux méthodes de segmentation

Les boîtes bleues représentent l'écart interquartile, la ligne bleue à l'intérieur des boîtes la médiane, le cercle jaune la moyenne, les moustaches incluent tous les points situés dans un intervalle de 1,5 EIQ du quartile supérieur/inférieur, les points bleus représentent les valeurs qui se trouvent en dehors de ces limites.

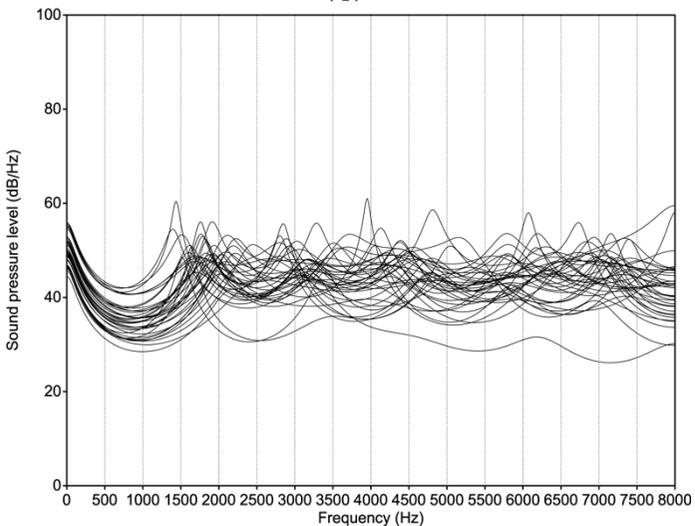




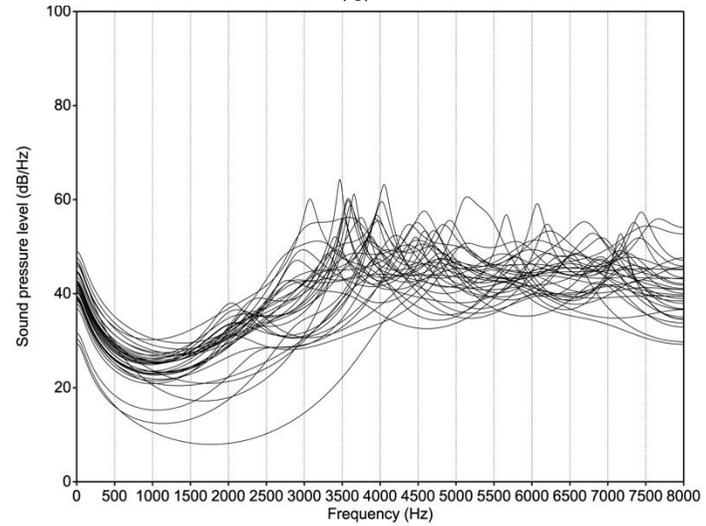


Annexe R – Spectres lissés pour les consonnes non-sonantes du français (méthode FBDS)

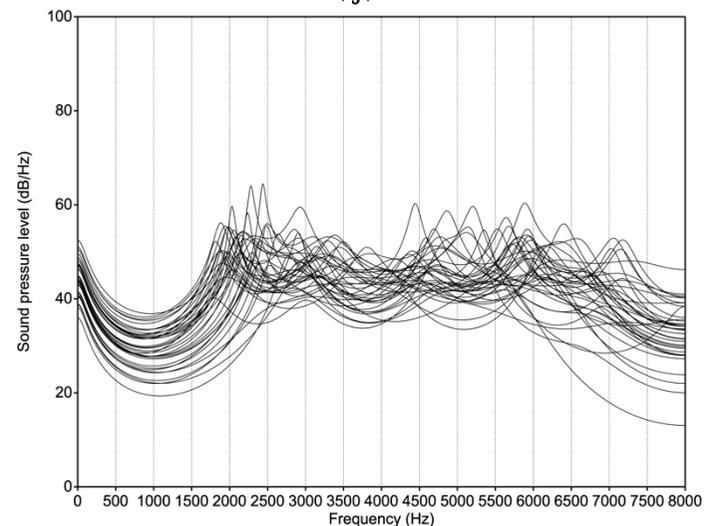
/f/



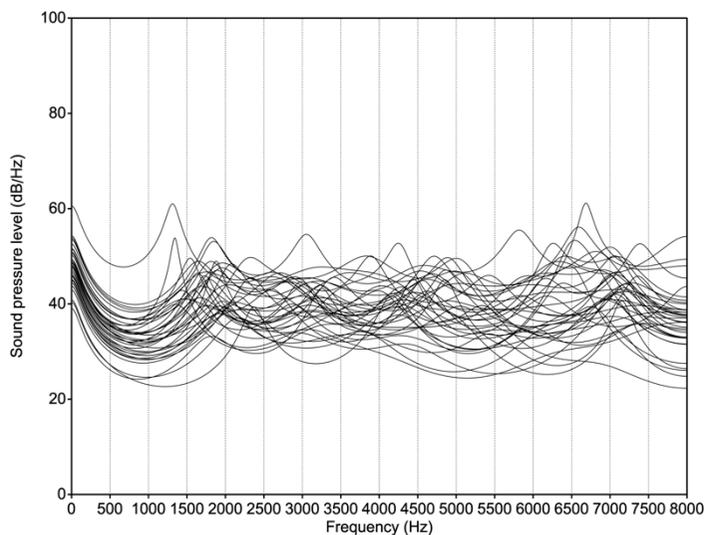
/s/



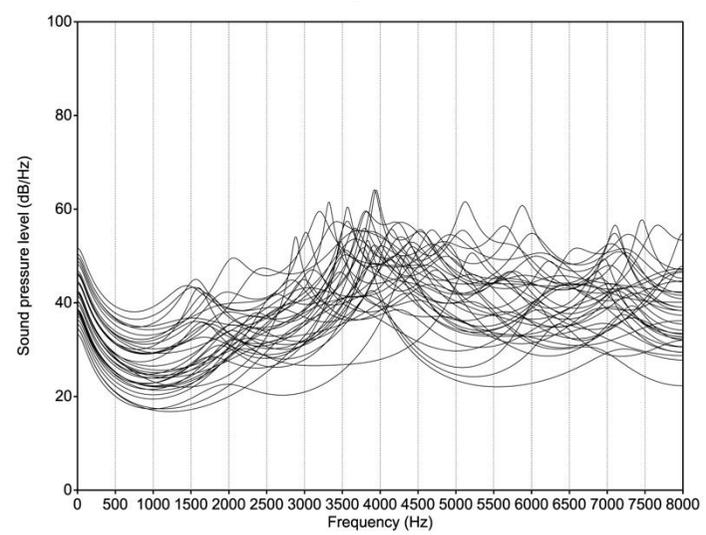
/ʃ/



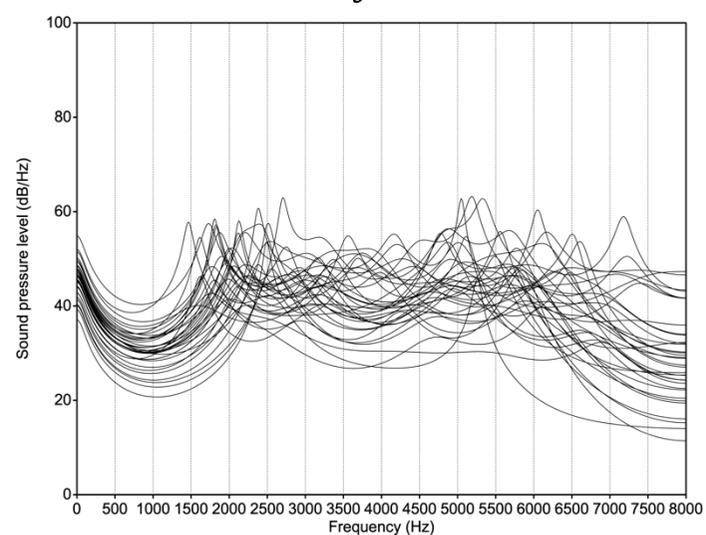
/v/



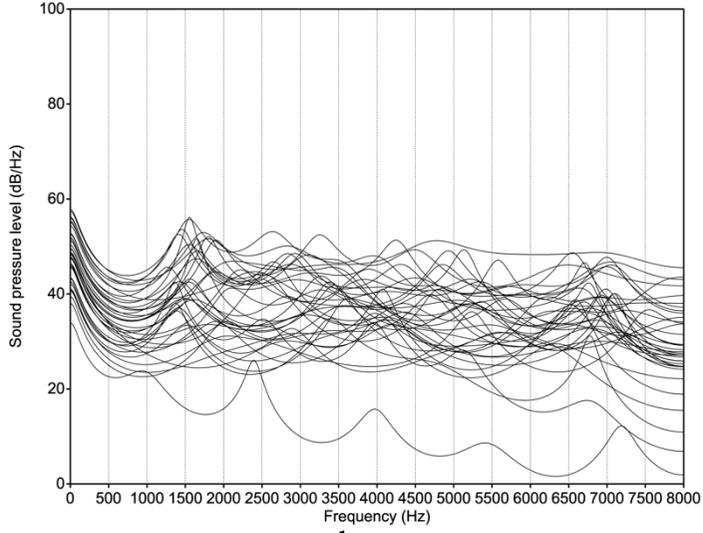
/z/



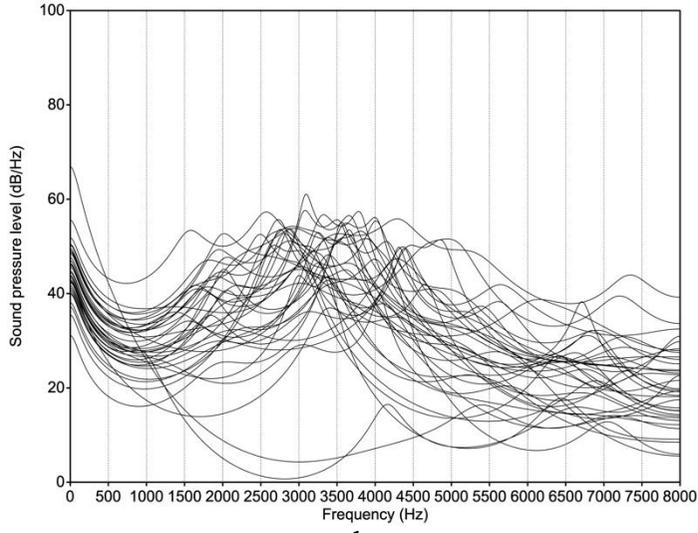
/ʒ/



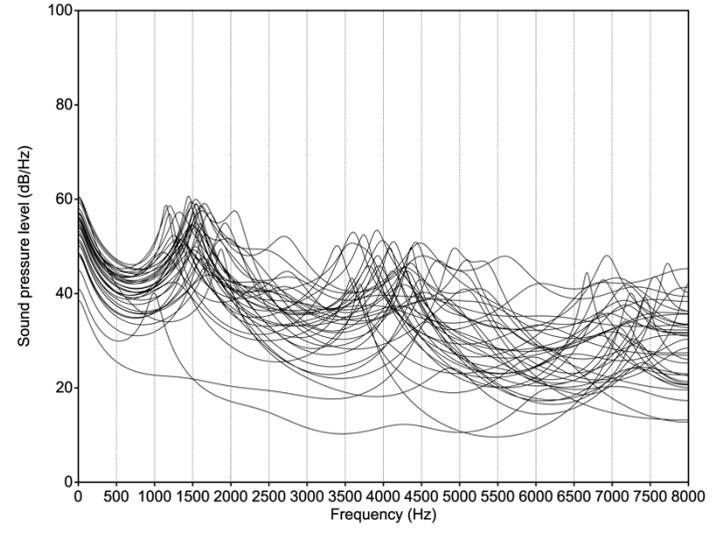
/p/



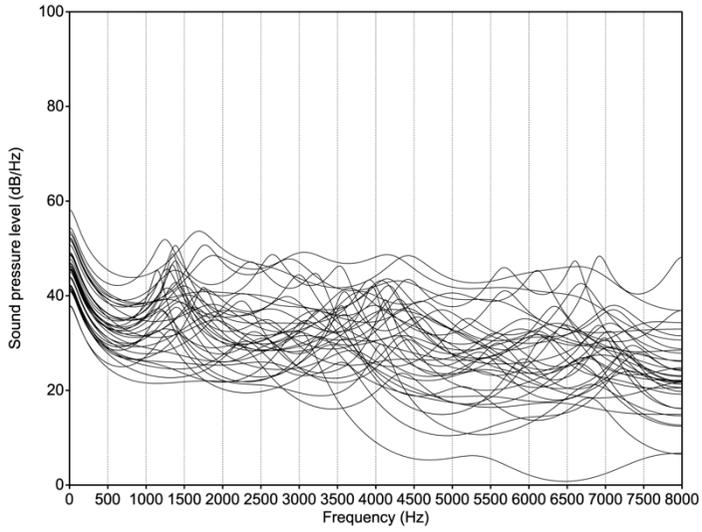
/t/



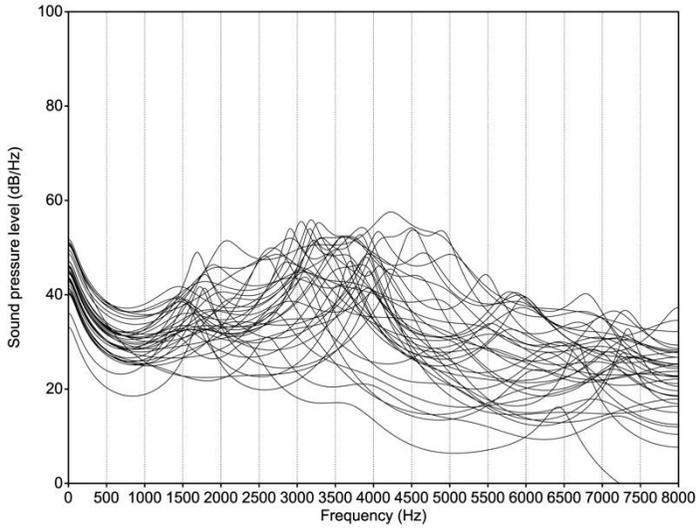
/k/



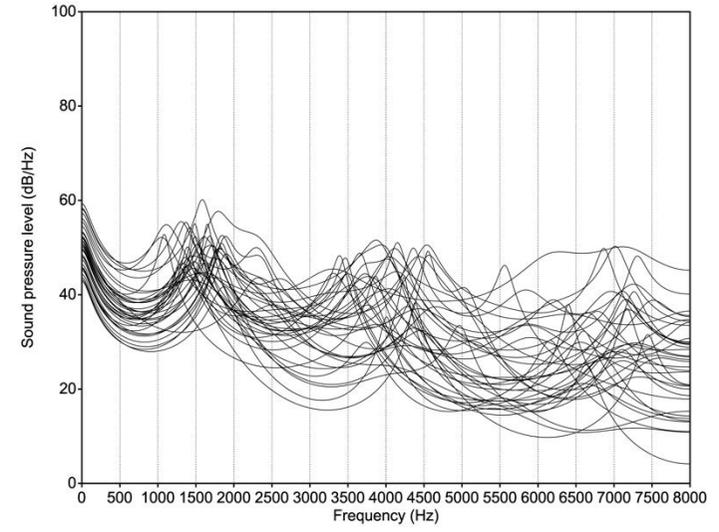
/b/



/d/

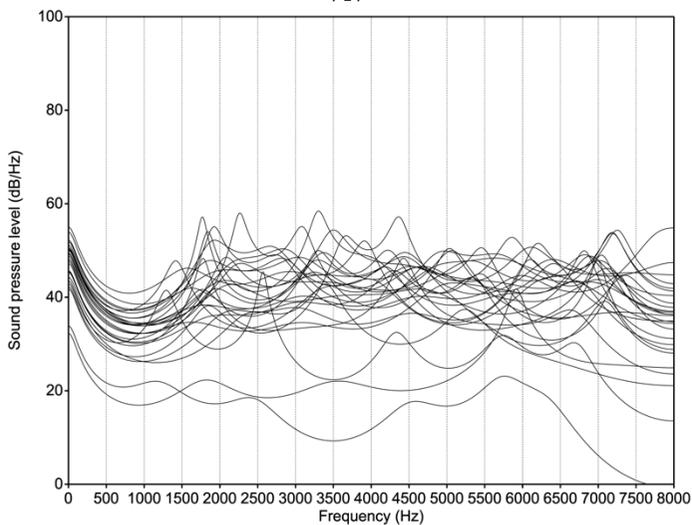


/g/

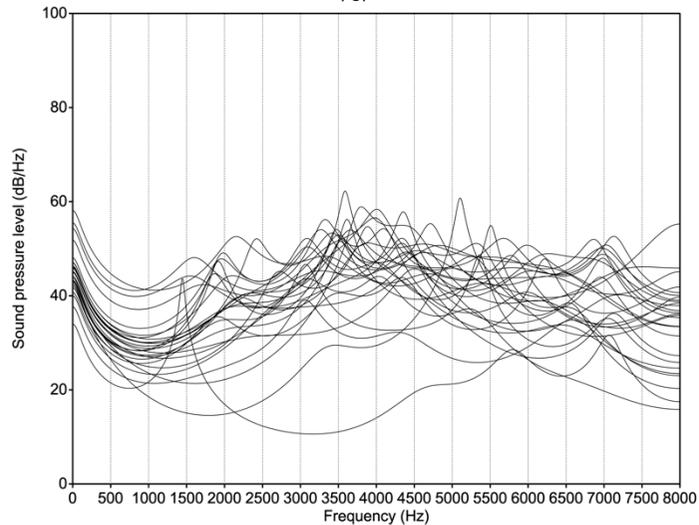


Annexe S – Spectres lissés pour les consonnes non-sonantes françaises de patients avec cancer ORL

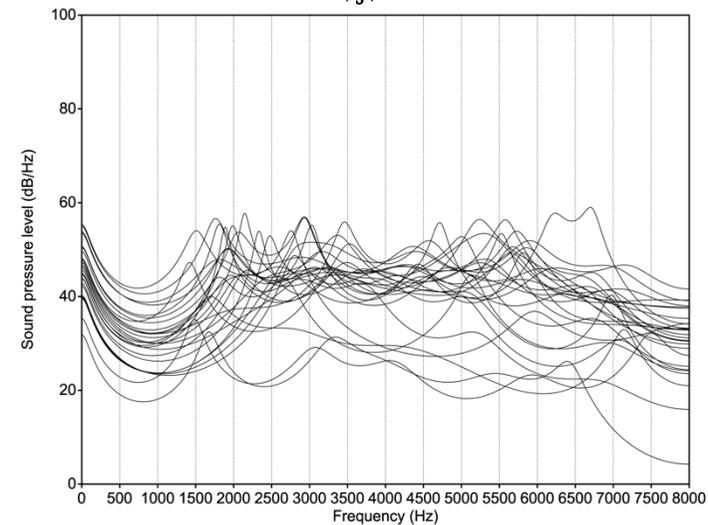
/f/



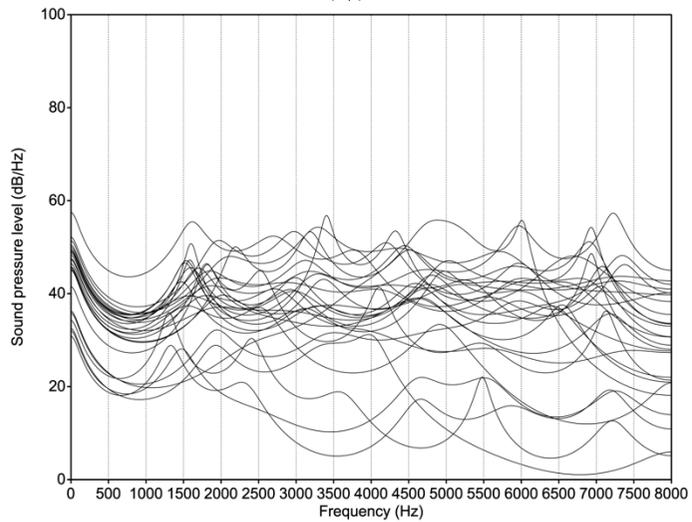
/s/



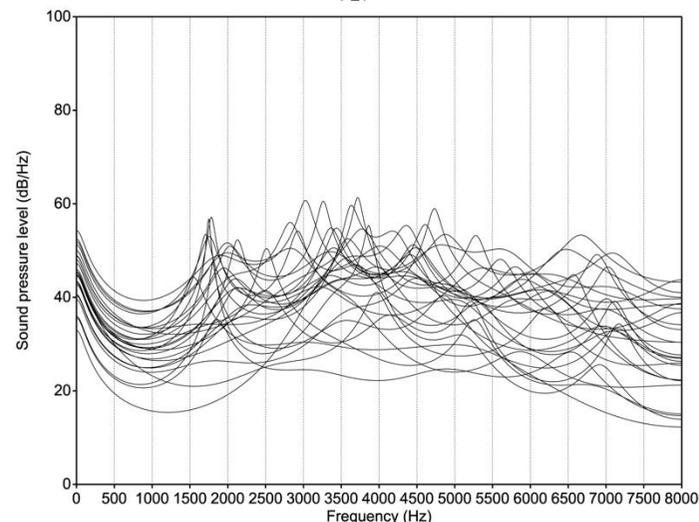
/ʃ/



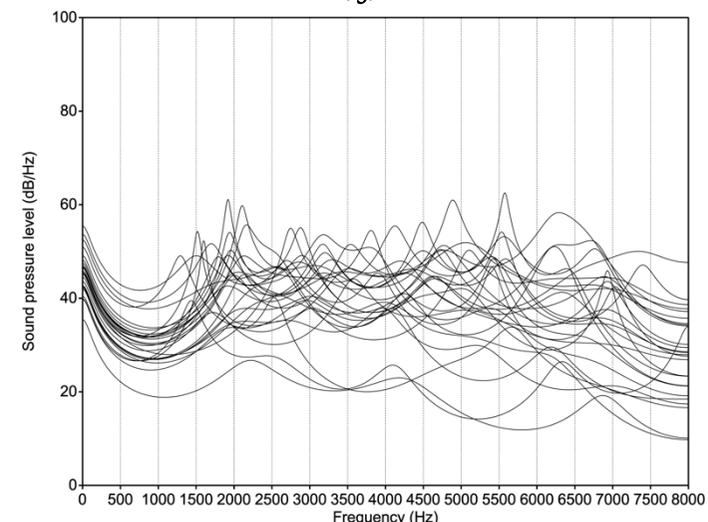
/v/



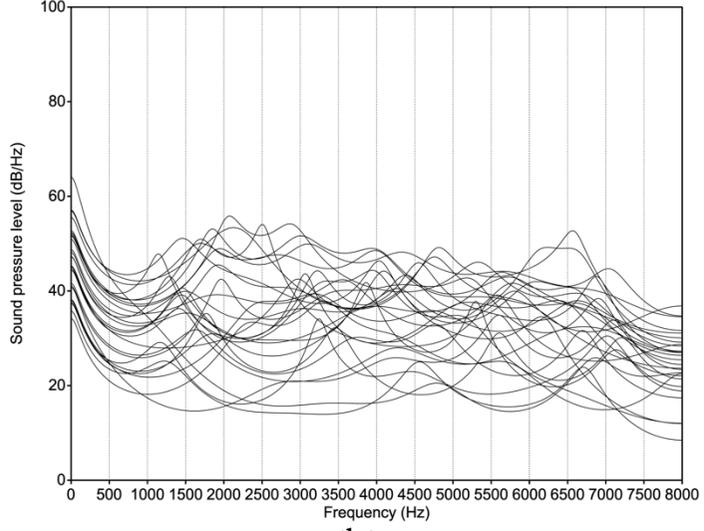
/z/



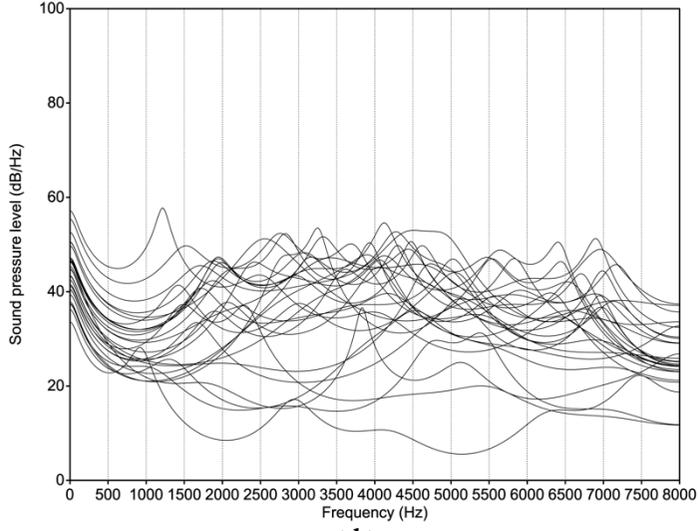
/ʒ/



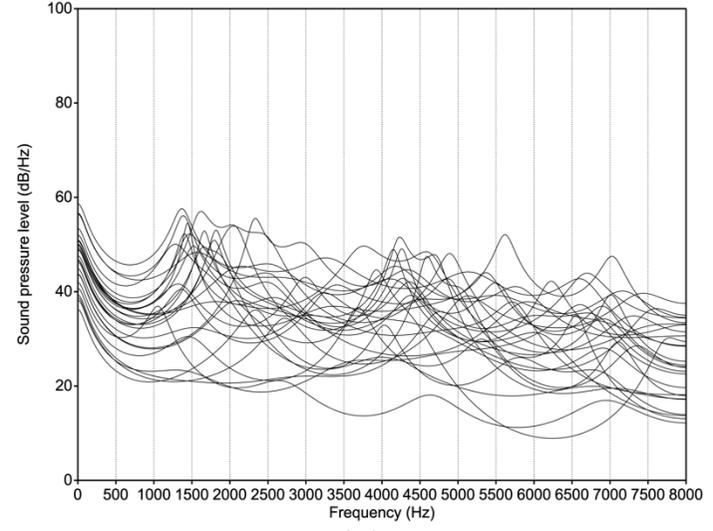
/p/



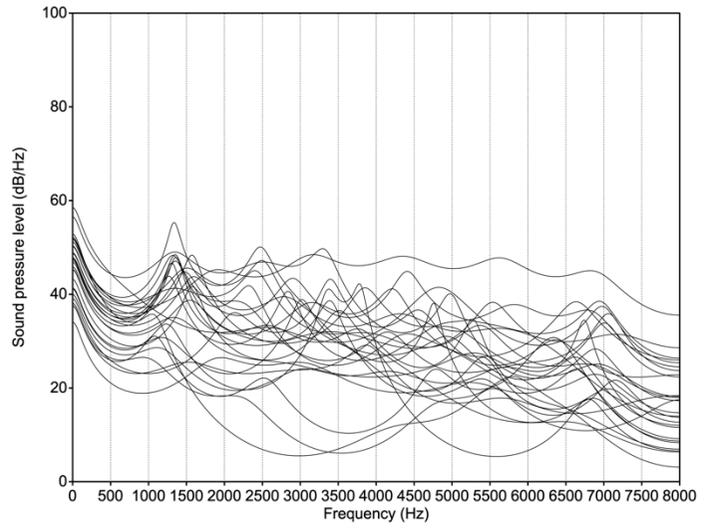
/t/



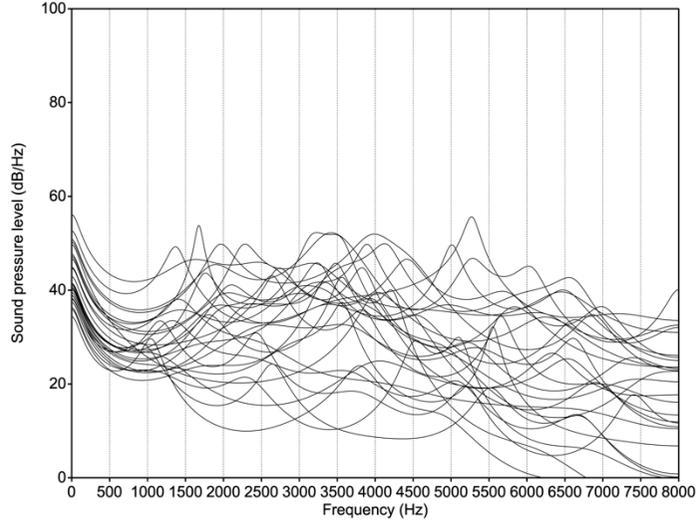
/k/



/b/



/d/



/g/

