



Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé

Mathieu Balaguer

► To cite this version:

Mathieu Balaguer. Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé. Sciences de l'information et de la communication. Université Paul Sabatier - Toulouse III, 2021. Français. NNT : 2021TOU30109 . tel-03557511

HAL Id: tel-03557511

<https://theses.hal.science/tel-03557511>

Submitted on 4 Feb 2022

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
Mathieu BALAGUER

Le 12 octobre 2021

**MESURE DE L'ALTERATION DE LA COMMUNICATION PAR ANALYSES
AUTOMATIQUES DE LA PAROLE SPONTANÉE APRES TRAITEMENT
D'UN CANCER ORAL OU OROPHARYNGE.**

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, Virginie WOISARD et Jérôme FARINAS

Jury

Mme Nathalie HENRICH BERNARDONI, Rapporteuse
M. Emmanuel BABIN, Rapporteur
M. Jérôme FARINAS, Examineur
M. Julien PINQUIER, Directeur de thèse
Mme Virginie WOISARD, Co-directrice de thèse
M. Rudolph SOCK, Président

Université Toulouse III Paul Sabatier

École doctorale 475 EDMITT

Mathématiques, Informatique, Télécommunications de Toulouse

IRIT Institut de Recherche en Informatique de Toulouse – UMR 5505

Équipe SAMoVA

Structuration, Analyse et Modélisation de documents Vidéo et Audio

Discipline Informatique et Télécommunications

**Mesure de l'altération
de la communication
par analyses automatiques
de la parole spontanée
après traitement
d'un cancer oral ou oropharyngé**

Mathieu Balaguer

Composition du jury

M. Rudolph Sock	LiLPa Université de Strasbourg	Président de jury
Mme Nathalie Henrich Bernardoni	Laboratoire GIPSA Lab (Unité mixte du CNRS, de Grenoble-INP et de l'Université de Grenoble-Alpes)	Rapporteuse
M. Emmanuel Babin	CHU Caen, Unité INSERM U1086 ANTICIPE « Cancers et prévention »	Rapporteur
M. Jérôme Farinas	Université Toulouse III Paul Sabatier	Co-encadrant
M. Julien Pinquier	Université Toulouse III Paul Sabatier	Directeur de thèse
Mme Virginie Woisard	CHU Toulouse, Laboratoire LNPL Université Jean Jaurès Toulouse II	Directrice de thèse

Adresse du laboratoire

IRIT – Institut de Recherche en Informatique de Toulouse UMR 5505

Université Toulouse III Paul Sabatier

Équipe SAMoVA – Structuration, Analyse et Modélisation de documents Vidéo et Audio

118 Route de Narbonne, F-31062 Toulouse Cedex 9

Remerciements

« *Nous avons fait comme Christophe Colomb... euh... non, comme Samothrace... enfin, un type de cette époque : “Veni, vidi, vici !”* »^a.

En premier lieu, je tiens à chaleureusement remercier Rudolph Sock, Président de jury, Nathalie Henrich Bernardoni et Emmanuel Babin qui m’ont fait l’honneur d’évaluer mon travail de thèse. Au plaisir de vous recroiser pour continuer ces discussions passionnantes.

Faire confiance à un orthophoniste, titulaire d’un Master en épidémiologie clinique, pour réaliser un Doctorat en informatique et télécommunications... Peu auraient eu le courage de se lancer !

Merci à Julien Pinquier : tu as su, avec bienveillance, pédagogie (et patience aussi), m’accompagner à la découverte de ce monde... toute une Affaire ! (« *C’est inouï !... C’est prodigieux !... C’est incroyable !...* »^b) Je n’aurais pas pu espérer meilleur Directeur de thèse, tant ta curiosité, ton ouverture d’esprit et ton esprit d’équipe m’ont permis de me sentir capable de relever les défis même dans les moments de doute.

Merci également à Jérôme Farinas, tu m’as fait confiance dès le départ de ce projet, et tu m’as patiemment guidé à travers les méandres de Bash sans jamais te lasser, ni t’agacer (« *Mais enfin, capitaine, nous direz-vous ce que cela signifie ? ...* »^c)... Grâce à toi, j’ai pu prendre le vol 714 pour l’analyse du signal de parole.

Au Professeur Tournesol et au Capitaine Haddock, donc !

Merci également à Virginie Woisard. Ton sens clinique, ta capacité phénoménale à avoir toujours dix analyses d’avance (« *Judicieuse question ! ...* »^d) et à gérer ~~dix~~ cent projets à la fois – tous de véritables Bijoux – m’épatent prodigieusement. Si j’en suis là, c’est grâce à toi, et à une proposition que tu m’as faite il y a maintenant près de 10 ans. C’était le point de départ de toute une aventure (en vrac : Master, Oncoréhabilitation, C2SI, CRIL, CFUO, UVD, DU Éducation thérapeutique... j’en oublie !) durant laquelle j’ai toujours pu compter sur toi, notamment dans les moments les plus difficiles.

Au-delà de mes directeur, directrice et encadrant de thèse, je remercie tout particulièrement pour la richesse de nos échanges et travaux (« *Il y a du vrai dans ce que vous dites...* »^e) :

Renée Speyer, pour votre disponibilité et votre rigueur méthodologique. Je vous dois la première partie de ma thèse, et même si le travail pour réaliser une revue systématique est gigantesque, cela semble bien plus facile à vos côtés. Tusen takk ! ;

Benoît Lepage : depuis le Master, votre collaboration m’a été précieuse, tant pour les questions statistiques et méthodologiques que pour l’écriture et le suivi des projets de recherche associés (notamment DAPADAF-E) ;

Florence Sordes pour les échanges autour des cercles sociaux et de la dynamique psychosociale des patients ;

Guillaume Cabanac pour les éclairages hyper précis et pertinents tout au long du suivi et les boosts de motivation quand c’était nécessaire.

Je remercie particulièrement M. le Professeur Serrano, pour votre soutien dès le tout début, tant au niveau hospitalier qu’universitaire.

Peu de doctorants peuvent se targuer de faire partie d'une équipe aussi dynamique, stimulante et bienveillante que l'équipe SAMoVA. Merci :

À tous les permanents et doctorants que j'ai pu croiser sur mon chemin ;

Je dirais même plus, à Timothy (« *Oui, botus et mouche cousue : c'est votre denise.* »^f). Cette thèse n'aurait clairement pas été la même sans un logopède (devenu orthophoniste !) avec qui faire le tour du monde. Ravi d'avoir pu éviter ensemble les peaux de bananes des différents scripts qui nous ont souvent laissés perplexes...

À Lucile, du bureau 222 ³/₄, pour sa patience, sa disponibilité, ses explications toujours claires sur le TDNN-HMM basé sur Kaldi (!), et ses thés écoresponsables ;

À Lila, toujours dans la famille du TDNN-HMM, informaticienne et qui a toutes les qualités de l'orthophoniste ;

À Régine André-Obrecht pour ses nombreux et excellents conseils et ses encouragements ;

À Christine Sénac pour son accueil dans son bureau, les moments de rire et de sourire et les petits mots avant soutenance qui font plaisir ;

À l'équipe « parole pathologique » pour les échanges constructifs et moments passés ensemble à droite, à gauche et sur la route : Julie, Robin, Vincent et Sebastião ;

Au reste de l'équipe, ceux qui ont fait de cette thèse un vrai moment de plaisir : Estelle, Leo, Verdiana, Jim, Antoine, Etienne, Gautier, Amélie, Benjamin, les déjà docteurs Sébastien, Nicolas, Abdel, Isabelle, Hervé, Philippe, Thomas, Imed et Lionel.

Merci aux membres des projets RUGBI, pour les échanges et éclairages lors de nos réunions (« *“Santhéodoriens”, je vous ai compris !* »^g) : Muriel, Alain, Christine, Anna et Marie du LPL, Corinne (avec deux « n ») et Sondes du LIA, Corine (avec un seul « n ») et Halima du nouveau LNPL.

Merci aussi aux participants et partenaires du projet DAPADAF-E, une belle aventure est encore devant nous : les équipes du LPL, du LIA, mais aussi de l'AP-HM (Camille, Danièle, Alexia...), et du CHU de Toulouse (Florine, Nadège et David).

Du côté clinique, je tiens à remercier très chaleureusement toutes les personnes du CHU de Toulouse qui m'ont permis de réaliser ce parcours.

Je pense notamment à Catherine Le Saux, toujours enthousiaste et d'un grand support : merci d'avoir cru en moi et de m'avoir donné les moyens de parvenir à réaliser ce travail. Merci aussi à André Weider pour votre soutien à la recherche en soins, et Ghislaine Mendes pour tes encouragements dès le début. J'ai également une pensée pour les membres de la plateforme CARES, avec qui je partage l'objectif ambitieux mais passionnant de contribuer à la recherche en soins.

Merci aux collègues orthophonistes qui ont assuré le suivi des patients : Stéphanie, Leslie et Gabrielle. Mais j'ai une pensée particulière pour Anaïs et Clémence, mes relais (et quels relais !) dans l'équipe, toujours partantes pour faire des enregistrements, collecter des questionnaires, élaborer des projets et gérer les péripéties des inclusions (« *C'est bien notre chance ! Pour une fois que nous tenions des coupables, il faut qu'ils s'arrangent pour être innocents ! ...* »^h) !

J'ai également une pensée pour mes collègues soignants et soignantes, notamment Cathy et Aurélie dans leur aide à l'inclusion des sujets : travailler avec vous est un vrai bonheur ! Merci aussi à Sabine et Pascale pour m'avoir permis d'aiguiser mes réflexions cliniques à leurs côtés.

Bien évidemment, ce travail ne serait rien sans la participation active de tous les patients qui se sont prêtés au jeu des enregistrements de parole et des 25 (!) pages de questionnaires à compléter. Je les remercie vivement.

Merci aussi aux licornes du CFUO de Toulouse pour le travail en commun, les moments de relâchements indispensables aussi : Florence, Lola et Aline (ou comment soulever des montagnes et les transformer en arcs-en-ciel).

Une pensée très (très) particulière aussi pour la Team Bombasses. Babeth et Laura, je n'oublierai jamais que si j'en suis là aujourd'hui, c'est grâce à vous, sur qui j'ai pu compter dans les moments les plus critiques. C'est également vous qui avez été les pionnières de l'utilisation des GIF dans nos messageries ! Je n'ai jamais eu l'impression d'avoir 80 ou 5 535 kms entre nous. Je suis tellement fier de vous et de vos parcours de vie, professionnels et personnels.

Merci aussi à ce collègue spécial, celui avec qui j'ai partagé le plus longtemps mon bureau en ces temps de confinements, ce collègue à moustaches qui dit « miaou », j'ai nommé : Oslo (« *Ha ! Ha ! Ha ! Il commence à grésiller sérieusement du trolley, ce pauvre Dupond ! ...* »ⁱ).

Comment conclure sans remercier celles et ceux qui me soutiennent, m'encouragent, affrontent mes moments de doute et d'angoisse : ma famille. Christine, Caroline, Sébastien, Quentin : Merci à vous d'être là, simplement.

Merci enfin à toi, Romu, pour ton amour et ton soutien au quotidien depuis maintenant plus de 13 ans. Que c'est passé vite !

« *Eih bennek, eih blavek.* »^j

^a Dupont, à Tintin : *L'Île noire* (1938), Hergé, éd. Casterman, coll. « Tintin », 1965, t. 7, p. 60

^b Tournesol : *On a marché sur la Lune*, Hergé, éd. Casterman, coll. « Tintin », 1954, t. 17, p. 20

^c Dupont, à Haddock : *Le Trésor de Rackham le Rouge*, Hergé, éd. Casterman, coll. « Tintin », 1944, t. 12, p. 22-23

^d Dupont, à Dupond : *Les Bijoux de la Castafiore*, Hergé, éd. Casterman, coll. « Tintin », 1963, t. 21, p. 45

^e Dupont, à Tintin : *Le Secret de la Licorne*, Hergé, éd. Casterman, coll. « Tintin », 1943, t. 11, p. 10

^f Dupont, à Dupond : *L'Affaire Tournesol*, Hergé, éd. Casterman, coll. « Tintin », 1956, t. 18, p. 13

« Dupond : Motus et bouche cousue : c'est notre devise.

Dupont : Oui, botus et mouche cousue : c'est votre devise. »

^g Dupont, à Dupond : *Tintin et les Picaros*, Hergé, éd. Casterman, coll. « Tintin », 1976, t. 23, p. 60

« Dupond : Tu n'aurais pas une parole historique, par hasard ?

Dupont : Euh... "Santhéodoriens, je vous ai compris"... Ça irait, tu crois ? ... »

^h Dupont, à Dupond : *Les Bijoux de la Castafiore*, Hergé, éd. Casterman, coll. « Tintin », 1963, t. 21, p. 60

ⁱ Dupont : *Objectif Lune*, Hergé, éd. Casterman, coll. « Tintin », 1953, t. 16, p. 24 (Dupont)

^j *Le Sceptre d'Ottokar* (1939), Hergé, éd. Casterman, coll. « Tintin », 1947, t. 8, p. 21

Devise de la Syldavie, pays imaginaire créé par Hergé, qui signifie à peu près en langue syldave « J'y suis, j'y reste ! »

... Et merci à Hergé, bien évidemment ! (source : https://fr.wikiquote.org/wiki/Les_Aventures_de_Tintin, consulté le 17/10/2021).

Mesure de l'altération
de la communication
par analyses automatiques
de la parole spontanée
après traitement
d'un cancer oral ou oropharyngé

Mathieu Balaguer

Table des matières

Table des figures	4
Liste des tableaux	6
Glossaire et liste des abréviations	8
Introduction	11
1 État de l’art : revue systématique de littérature	21
1.1 Matériel et méthode	21
1.1.1 Protocole et enregistrement du protocole	21
1.1.2 Critères d’éligibilité	22
1.1.3 Sources de données et stratégies de recherche	22
1.1.4 Qualité méthodologique et niveau de preuve	24
1.1.5 Extraction des données	25
1.2 Résultats	25
1.2.1 Sélection des études	25
1.2.2 Évaluation de la qualité	25
1.2.3 Données bibliométriques	27
1.2.4 Caractéristiques des participants aux études sélectionnées	28
1.2.5 Critères de comparaison	31
1.2.6 Échantillons de parole	32
1.2.7 Mesures acoustiques	33
1.3 Discussion	36
1.3.1 Choix des paramètres acoustiques / automatiques	36
1.3.2 Échantillons de parole	40
1.3.3 Limites de l’étude	40
1.4 Conclusion	41
2 Conception et constitution du corpus	42
2.1 Participants	42
2.1.1 Critères d’inclusion et de non-inclusion	43
2.1.2 Recrutement	44
2.2 Évaluation de l’impact sur la communication	44
2.3 Évaluation des facteurs associés au trouble de parole	48
et à la communication	48
2.3.1 Déficits consécutifs au traitement de la pathologie cancérologique ORL	48
2.3.2 Anxiété / Dépression	49
2.3.3 Statut cognitif	49
2.3.4 Besoins de communication liés aux cercles sociaux	50
2.3.5 Auto-perception du trouble de parole	53
2.3.6 Qualité de vie en cancérologie ORL	54

2.4	Enregistrements de parole et choix des tâches	56
2.5	Déroulement des passations	57
2.5.1	Participation des sujets	58
2.5.2	Caractéristiques des sujets inclus	59
2.6	Analyses statistiques : principes généraux	62
2.7	Conclusion	63
3	Analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé	65
3.1	Recherche et extraction des paramètres	66
3.1.1	Préparation des données	67
3.1.2	Signal de parole	68
3.1.3	Niveau articulatoire – Phonétique articulatoire	71
3.1.4	Niveau articulatoire – Composante phonémique	78
3.1.5	Composante morpho-syntaxique – Composante lexicale	84
3.1.6	Composante morpho-syntaxique – Composante grammaticale	93
3.1.7	Composante pragma-sémantique et discursive	96
3.1.8	Synthèse de l'extraction des paramètres	100
3.2	Sélection des paramètres	101
3.3	Conclusion	105
4	Prédiction de l'altération de la communication	106
4.1	Construction du score holistique de communication	106
4.1.1	Présentation des items	107
4.1.2	Sélection des items	109
4.1.3	Validité d'apparence	111
4.1.4	Validité de construit	112
4.1.5	Construction du score holistique	113
4.1.6	Validation croisée à cinq blocs	117
4.1.7	Validation complémentaire par analyse en profils latents	117
4.1.8	Discussion	119
4.2	Modélisations de l'impact sur la communication au moyen des paramètres automatiques	121
4.2.1	Analyse bivariée entre SHC et paramètres automatiques	121
4.2.2	Modélisation prédictive au moyen des seuls paramètres automatiques	122
4.2.3	Modélisation prédictive avec facteurs associés	127
4.3	Conclusion	131
	Conclusion	133
	Bibliographie	144
	Annexes	167

Table des figures

Figure 1 – Taux d’incidence et de mortalité des cancers Lèvre-Bouche- Pharynx en France selon l’année (taux standardisés monde TSM) – Échelle logarithmique – Source : Santé publique France, 2019.....	11
Figure 2 – Classification internationale du fonctionnement, du handicap et de la santé, et interactions entre composantes (OMS, 2001).....	13
Figure 3 - Modèle conceptuel de qualité de vie globale, et relations entre les différents niveaux de mesure qui le constituent (Wilson & Paul, 1995)	14
Figure 4 - Diagramme du processus de revue, selon la méthodologie PRISMA, adapté de Moher (Moher et al., 2009).....	24
Figure 5 - Localisation des affiliations des auteurs, et collaborations entre équipes de recherche	27
Figure 6 - Nombre d'articles publiés par année	28
Figure 7 - Localisations tumorales	29
Figure 8 - Proportion détaillée des tumeurs selon leur taille (critère T de la classification TNM) dans les articles retenus (attention : les articles de de Bruijn (de Bruijn et al., 2011, 2009) ne différencient pas les tumeurs T3 et T4 entre les sujets).....	30
Figure 9 - Nombre d'études selon les catégories d'analyse (la somme est supérieure au nombre total car certaines études analysent différentes catégories).....	33
Figure 10 - Modèle psycholinguistique de Caron (1989), cité par François & Nespoulous (2011) et Devevey & Kunz (2018)	45
Figure 11 - Déroulement de la participation des sujets.....	58
Figure 12 - Scores de sévérité et d'intelligibilité obtenus par les sujets, classés par score croissant de sévérité.....	62
Figure 13 - Structure du corpus, et mise en lien avec les modélisations conceptuelles.....	64
Figure 14 - Schéma général de l'analyse des paramètres automatiques extraits de la parole spontanée selon le modèle psycholinguistique de Caron.....	67
Figure 15 - Durée totale des fichiers (<i>duration</i> , en gras) et durée de parole (<i>phontimew</i> , en italique), en secondes	72
Figure 16 - Métriques d'évaluation : en abscisses, le paramétrage du TDNNf « lexical » (les deux premiers chiffres correspondent au WIP – 00 : WIP=0 ; 05 : WIP=0,5 ; 10 : WIP=1 –, les deux derniers au LMWT)	86

Figure 17 - Performances du système sous le paramétrage WIP=0 ; LMWT = 11 selon la sévérité perceptive (lot 1 uniquement). L'axe des ordonnées de droite est relatif à la sévérité perceptive, l'axe de gauche aux autres variables (dont l'accuracy).....	87
Figure 18 - Graphique de corrélation entre accuracy et sévérité du trouble évaluée de façon perceptive (lot 1 uniquement)	87
Figure 19 - Comparaison des index de Guiraud, calculés sur les fichiers complets et sur les quarts de fichiers du lot 1	92
Figure 20 - Comparaison des scores de densité lexicale pour le lot 1, calculés sur le fichier complet ou sur une minute (à partir d'une transcription perceptive ou d'une sortie automatique)	93
Figure 21 - Représentation des mots ayant au moins dix occurrences chez au moins un des sujets (les mots ayant au moins 50 occurrences ont été étiquetés)	99
Figure 22 - Schéma synthétique de la méthodologie d'extraction des paramètres après analyses automatiques du signal de parole spontanée.....	100
Figure 23 - Processus de sélection des paramètres issus des analyses automatiques de la parole spontanée	102
Figure 24 - Processus de construction du score de communication.....	110
Figure 25 - Scores I-CVI et Kappas d'agrément avec cut-off (ligne rouge)	112
Figure 26 - Différences entre analyse en composantes principales et analyse factorielle selon K. Grace-Martin (https://www.theanalysisfactor.com/the-fundamental-difference-between-principal-component-analysis-and-factor-analysis/).....	114
Figure 27 - Scree plot (tracé d'éboulis) des valeurs propres des facteurs suite à l'analyse factorielle en PCF.....	115
Figure 28 - Scores holistiques de communication obtenus par sujet	116
Figure 29 - Nuage de points représentant le SHC et le SHC prédit par la validation croisée, avec droite de régression	117
Figure 30 - Analyses bivariées entre le score holistique de communication et les paramètres automatiques dont les coefficients de corrélation de Spearman sont significatifs au seuil de 5 %.	122
Figure 31 - Nuage de points entre SHC et valeurs prédites sur les 25 sujets de l'échantillon	125
Figure 32 - Nuage de points entre SHC et valeurs prédites par la validation croisée à cinq blocs	125
Figure 33 - Comparaison des performances entre modélisations du SHC	129
Figure 34 - Synthèse de notre étude et mise en lien avec le modèle de Wilson.....	133

Liste des tableaux

Tableau 1 - Bases de données et syntaxe de recherche (descripteurs et texte libre)	23
Tableau 2 - Niveau de preuve (NHMRC) et qualité méthodologique (QualSyst) des 22 articles inclus.....	26
Tableau 3 - Nombre total de participants dans les études retenues	28
Tableau 4 - Critères de comparaison choisis par les auteurs des études incluses	32
Tableau 5 - Constitution des échantillons de parole	33
Tableau 6 - Résultats de l'évaluation perceptive de la sévérité du trouble et de l'intelligibilité.....	61
Tableau 7 - ICC des paramètres perceptifs.....	61
Tableau 8 - Description statistique des paramètres de macrovariabilité (intensité en décibels, fréquence en hertz sauf <i>f0cent</i> en cents).....	69
Tableau 9 - Description statistique des paramètres spectraux et cepstraux (en décibels).....	69
Tableau 10 - Description statistique des mesures temporelles calculées sur les résultats de WebRTC-VAD (les durées sont données en secondes)	73
Tableau 11 - Description statistique des mesures temporelles extraites par le script « Syllable nuclei » de Praat	75
Tableau 12 - Description statistique des mesures extraites de SpeechTools	77
Tableau 13 - Description statistique des mesures phonémiques extraites avec le système TDNNf-HMM.....	82
Tableau 14 - Valeurs de performances du système avec le paramétrage retenu {WIP = 0 ; LMWT = 11}	86
Tableau 15 - Description statistique des mesures lexicales calculés selon les sorties du système TDNNf-HMM « lexical »	89
Tableau 16 - Description statistique des scores de confiances mesurés au niveau lexical selon les sorties du système TDNNf-HMM « lexical ».....	91
Tableau 17 - Description statistique des scores « Index de Guiraud » et « Densité lexicale »	92
Tableau 18 - Description statistique des scores issus des analyses par classes grammaticales	94
Tableau 19 - Description statistique des scores issus des analyses par formes actives et supplémentaires (Iramuteq)	96
Tableau 20 Description statistique des classes thématiques retrouvées par le logiciel Iramuteq	97
Tableau 21 - Valeurs des scores de sentiment par segment de 5 120 caractères.....	98

Tableau 22 Description statistique des scores de sentiment (Microsoft Azure).....	100
Tableau 23 - Items inter-corrélés et choix des items retenus selon leur dispersion (en gras).....	113
Tableau 24 - Valeurs AIC et BIC des modèles GSEM (Analyse en profils latents)	118
Tableau 25 - Comparaison des classes latentes prédites par l'analyse LPA et du score SHC.....	119
Tableau 26 - Résultats de la modélisation du SHC par les paramètres automatiques et les facteurs associés	129

Glossaire et liste des abréviations

- ACP** : Analyse en Composantes Principales (Principal Component Analysis).
Méthode d'analyse de données permettant la projection dans un espace exprimant au mieux la variabilité.
- AIC** : Akaike Information Criterion (Critère d'information d'Akaike). Mesure de la qualité d'un modèle statistique.
- AFSOS** : Association Francophone des Soins Oncologiques de Support
- ANN** : Artificial Neural Network (Réseau de neurones artificiel). Méthode d'apprentissage automatique basée sur la modélisation de neurones formels.
- API** : Application Programming Interface (Interface de programmation applicative). Ensemble de méthodes permettant à un logiciel de proposer des services.
- BECD** : Batterie d'Évaluation Clinique de la Dysarthrie
- BIC** : Bayesian Information Criterion (Critère d'information bayésien). Critère de sélection de modèle parmi un ensemble fini de modèles.
- CB** : Cavité buccale
- CHI** : Carcinologic Handicap Index
- CIF** : Classification internationale du fonctionnement, du handicap et de la santé
- CPIB** : Communicative Participation Item Bank
- CPPS** : Smoothed Cepstral Peak Proeminence (Proéminence du pic cepstral lissé)
- DIP** : Dysarthria Impact Profile
- DNN** : Deep Neural Network (Réseau de neurones profond). Méthode d'apprentissage automatique à base de neurones pouvant modéliser des systèmes complexes.
- EBP** : Evidence-Based Practice (Pratique fondée sur les preuves)
- ECES** : Évaluation de la Constitution des Cercles Sociaux
- ECVB** : Échelle de Communication Verbale de Bordeaux
- EDP** : Échelle de Détresse Psychologique
- EIQ** : Écart Interquartile
- EORTC** : European Organization for Research and Treatment of Cancer (QLQ-C30 : Core Questionnaire ; QLQ-H&N35 : Head and Neck Module)
- ET** : Écart-type
- F0** : Fréquence Fondamentale
- HAD** : Hospital Anxiety and Depression scale
- HMM** : Hidden Markov Model (Modèle de Markov caché). Automates très utilisés en reconnaissance de la parole pour modéliser les évolutions temporelles des sons.

HTK : Hidden Markov Model Toolkit. Boîte à outil développée par l'université de Cambridge pour modéliser des HMM.

ICC : Intraclass Correlation Coefficient (Coefficient de Corrélation Intraclasses)

LASSO : Least Absolute Shrinkage and Selection Operator. Méthode de contraction des coefficients de régression statistique.

LMWT : Language Model Weight. Paramètre permettant de contrôler l'influence du modèle de langage d'un système de reconnaissance automatique de la parole.

LPA : Latent Profile Analysis (Analyse en profils latents)

LPC : Linear Prediction Coding (Coefficient de prédiction linéaire). Méthode utilisée principalement dans le traitement de la parole pour représenter l'enveloppe spectrale d'un signal numérique de parole sous forme comprimée, en utilisant les informations d'un modèle auto-régressif.

LTAS : Long Terme Average Spectrum (Spectre moyen à long terme)

MAR : Missing At Random. Données manquantes dépendant des valeurs observées mais pas des valeurs attendues (valeurs réelles) : l'absence de donnée dépend des caractéristiques du sujet, mais n'est pas liée au traitement.

MCAR : Missing Completely At Random. Données manquantes ne dépendant ni des valeurs observées, ni des valeurs attendues : rien ne distingue les sujets avec données manquantes des autres.

MFCC : Mel-Frequency Cepstrum Coefficient (Coefficient cepstral). Vecteur synthétique permettant de représenter les informations sur un court extrait de signal de parole.

MoCA : Montréal Cognitive Assessment

MNAR : Missing Not At Random. Données manquantes dépendant de la vraie valeur attendue : l'absence de donnée est liée au traitement.

MRTSC : Multi Résolution Sinusoidal Transform Coding. Représentation en ondelettes d'un signal.

MSCC : Mel-S transform Cepstrum Coefficient (Coefficient cepstral). La transformation S est une méthode d'analyse temps-fréquence qui combine les avantages de la transformée en ondelettes avec la transformée de Fourier à court terme (meilleur débruitage, résolution temporelle et localisation temps-fréquence) (Fang et al., 2017)

OP : Oropharynx

ORL : Oto-Rhino-Laryngologie

PCF : Principal-Component Factor

PCM : Pulse-Code Modulation. Format d'un signal audio échantillonné non compressé et non chiffré.

PER : Phone Error Rate (Taux d'erreur de phones)

PLS : Partial Least Squares (Moindres carrés partiels)

PHI : Phonation Handicap Index

PRO : Patients Reported Outcome

QDV : Qualité de vie

QFS : Questionnaire de Fonctionnement Social

RCP : Régression en Composantes Principales

RMSE : Root-mean-square error (Erreur quadratique moyenne)

SHC : Score Holistique de Communication

SVM : Support Vector Machine (Séparateurs à vaste marge). Méthode de classification automatique cherchant à discriminer le plus possible les classes à séparer.

TDNN : Time Delay Neural Network. Méthode de classification prenant en compte différentes échelles temporelles.

TDNNf : Time Delay Neural Network Factorization. Méthode dérivée de TDNN.

TNM : Classification TNM des cancers selon leur extension anatomique (T = Tumeur : volume tumoral et fixation aux organes voisins ; N = Territoire ganglionnaire – Node en anglais ; M = métastases)

VAD : Voice Activity Detector (Détecteur d'activité vocale). Algorithme utilisé pour repérer les zones présentant une activité vocale.

VADS : Voies Aéro-Digestives Supérieures

VOT : Voice Onset Time. Temps qui s'écoule entre le relâchement et le voisement d'une consonne occlusive.

VSA : Vowel Space Area (Aire de l'espace vocalique).

WIP : Word Insertion Penalty. Pénalité d'insertion de mots. Paramètre utilisé pour contrôler le décodage d'un système de reconnaissance de la parole.

Introduction

Cancers de la cavité buccale et de l'oropharynx et troubles de parole

Les cancers de la cavité buccale et de l'oropharynx ont une incidence élevée en France¹ : plus de 13 500 nouveaux cas ont été déclarés en 2018 (dont environ trois quarts chez les hommes). Il s'agit du troisième cancer le plus incident, notamment chez les hommes avec 18,3 cas pour 100 000 personnes-années (taux standardisé mondial, contre 5,8 chez les femmes). Dans le même temps, la mortalité consécutive aux cancers atteignant les lèvres, la bouche ou le pharynx diminue, avec un taux de 4,9 pour 100 000 en 2018 (contre 13,4 pour 100 000 en 1990) (voir Figure 1).

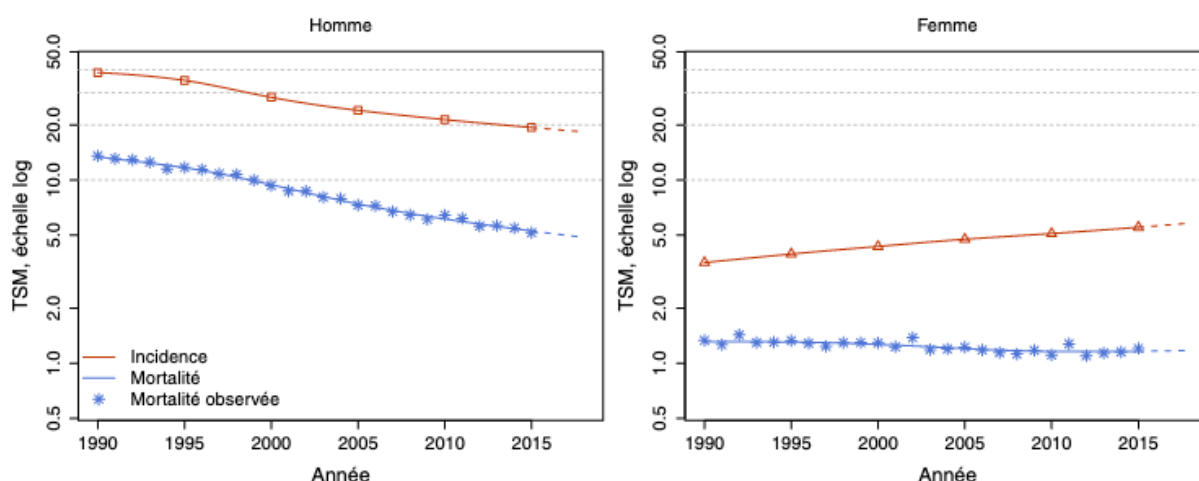


Figure 1 – Taux d'incidence et de mortalité des cancers Lèvre-Bouche-Pharynx en France selon l'année (taux standardisés monde TSM) – Échelle logarithmique – Source : Santé publique France, 2019

¹ Rapport Santé publique France, juillet 2019 :

<https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-sein/documents/rapport-synthese/estimations-nationales-de-l-incidence-et-de-la-mortalite-par-cancer-en-france-metropolitaine-entre-1990-et-2018-volume-1-tumeurs-solides-etud>, consulté le 07/05/2021

Ainsi, l'allongement de la durée de vie consécutive au cancer fait que les patients vivent désormais plus longtemps avec les séquelles du cancer et des traitements. Associé au fait que ces cancers restent fréquents, les séquelles de la pathologie cancéreuse de la cavité buccale et de l'oropharynx sont un véritable problème de santé publique. En raison de leur localisation, ils ont un impact sur les capacités de parole des sujets atteints (Balaguer et al., 2019b; Mlynarek et al., 2008). Certains facteurs cliniques de gravité du trouble de parole ont d'ailleurs été identifiés : les tumeurs de plus gros volume sont plus délétères (Borggreven et al., 2007; Colangelo et al., 2000), les tumeurs localisées dans la cavité buccale dégradent davantage la parole que les tumeurs oropharyngées (DeNittis et al., 2001; Stelzle et al., 2013) conformément au lieu d'articulation de la majorité des phonèmes, ou encore le traitement entrepris où la chirurgie est la plus délétère (Balaguer et al., 2019a; Barrett et al., 2004) bien que la radiothérapie dégrade aussi les capacités de parole (Vainshtein et al., 2014). Le délai depuis le diagnostic ou le traitement influence la sévérité perçue du trouble de parole par les patients (Bolt et al., 2016; Eadie et al., 2018; Vainshtein et al., 2014). Enfin, l'âge peut influencer les capacités de production de parole après traitement (Matsui et al., 2007).

Impact du trouble de parole sur les capacités de communication

Dès 2014, l'AFSOS (Association Francophone des Soins Oncologiques de Support) émettait des recommandations quant à la prise en charge des troubles de parole dans le référentiel « *Déficiences liées aux cancers des VADS* »². Les troubles de la voix et de la parole doivent être pris en charge précocement en post-opératoire, « *dès que la situation du patient le permet* », avec pour objectifs l'évaluation du handicap et l'évaluation des besoins de communication. Dans ce contexte, il convient de s'intéresser, au-delà des déficits analytiques et dynamiques, aux répercussions fonctionnelles et psychosociales post-traitement du trouble de parole.

Pour cela, divers cadres conceptuels, s'appuyant sur des modèles bio-psycho-sociaux ont été développés, permettent de mieux comprendre l'impact des actes thérapeutiques en cancérologie ORL sur la qualité de vie des patients (Borggreven et al., 2007).

² <https://www.afsos.org/wp-content/uploads/2015/12/Ind-Reeduc-4-deficiences-liees-VADS-2015-03-16-1-2.pdf>, consulté le 15/10/2021

Un des modèles les plus connus est la Classification internationale du fonctionnement, du handicap et de la santé (CIF, voir Figure 2) proposée par l'OMS en 2001 (Winance, 2008; World Health Organization, 2001). Dans le champ de la parole, les déficits organiques et anatomiques concernent les capacités motrices et sensitives des organes de la parole et leur coordination dans le cadre de la respiration (au repos et au cours de la parole), la phonation (c'est-à-dire la vibration des plis vocaux au niveau laryngé permettant la production de son, ainsi que la résonance) et de l'articulation des sons (Dykstra et al., 2007). Les limitations d'activités concernent les conséquences des déficits analytiques sur la parole d'une part, mais aussi des limitations induites par le trouble de parole sur les capacités de communication d'autre part. Les restrictions de participation sont relatives à la manière avec laquelle le sujet interagit avec son environnement mais aussi les restrictions induites par le trouble de parole sur sa dynamique psychosociale. Ainsi, outre le trouble de parole à proprement parler, les altérations de la communication et la modification de la dynamique psychosociale sont à envisager de façon spécifique en cancérologie ORL.

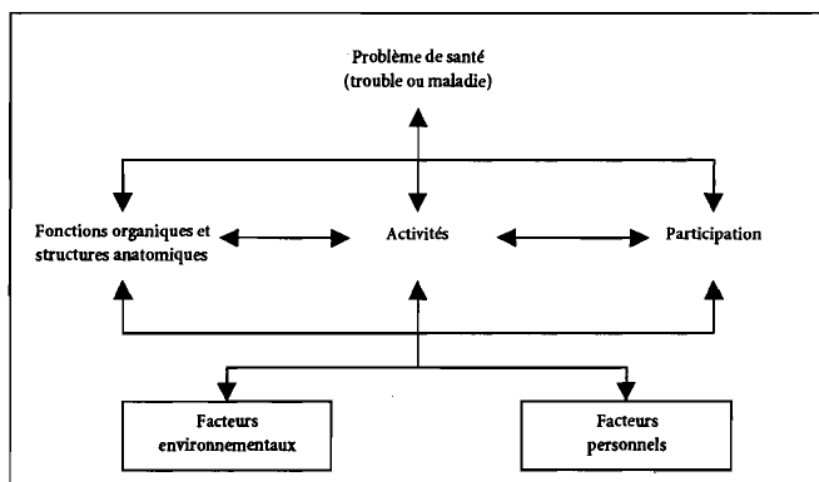


Figure 2 – Classification internationale du fonctionnement, du handicap et de la santé, et interactions entre composantes (OMS, 2001)

Des modélisations conceptuelles complémentaires ont été proposées pour préciser les différents niveaux en jeu dans la dynamique fonctionnelle et psychosociale, et établir les relations causales pouvant exister entre eux (Murphy et al., 2007; Wilson & Cleary, 1995) : les facteurs biologiques et physiologiques, le statut symptomatique, le statut fonctionnel, les perceptions générales de la santé et la qualité de vie globale (voir Figure 3).

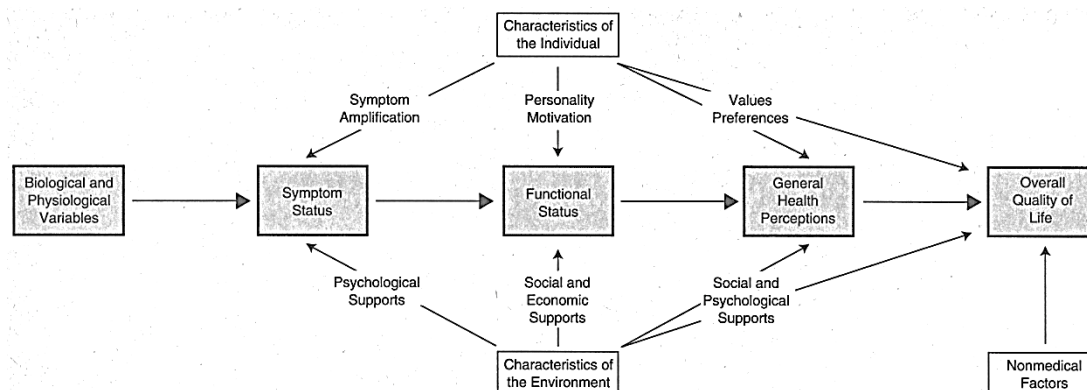


Figure 3 - Modèle conceptuel de qualité de vie globale, et relations entre les différents niveaux de mesure qui le constituent (Wilson & Paul, 1995)

Plusieurs études se sont ainsi intéressées aux conséquences fonctionnelles du trouble de parole sur la communication. De façon générale, les troubles de parole en contexte cancérologique contribuent à l'altération de la dynamique sociale des patients (Reich, 2009), dont la communication est une des parties essentielles. En cancérologie ORL, une corrélation à 0,29 est retrouvée entre l'acceptabilité de la parole ressentie par les patients laryngectomisés totaux et leur participation à la communication (Eadie et al., 2016). Les études menées spécifiquement auprès des patients traités pour un cancer de la cavité buccale ou de l'oropharynx ciblent principalement l'impact psychosocial et non fonctionnel, en retrouvant des liens entre qualité de vie et déficits de parole (Borggreven et al., 2007; Mlynarek et al., 2008; Rinkel et al., 2008).

Évaluation de la parole et de la communication en cancérologie ORL

En pratique clinique, l'évaluation de la production de parole (Fontan, 2012) est une étape essentielle dans la détermination des stratégies thérapeutiques pour le patient. Les objectifs ciblent en particulier les trois niveaux de fonctions, d'activités et de participation. Ils visent à limiter les déficiences, en restaurant ou optimisant une mobilité la plus efficace et la plus précise possible des structures restantes et reconstruites, par l'optimisation de la sensibilité intra-orale et de la production motrice de la parole (points d'articulation, enchaînements...). Les activités doivent redevenir les plus fonctionnelles possibles, en favorisant la récupération en termes d'intelligibilité, de compréhensibilité et de niveau de communication, ou en mettant si nécessaire en place des compensations pour optimiser les capacités de communication (communication améliorée et alternative). Enfin les restrictions de participation

doivent être réduites, en adaptant les stratégies de communication aux besoins de communication, selon la constitution des cercles sociaux et de l'entourage des patients, la vie sociale (familiale et professionnelle), les loisirs...

Pour autant, un état des lieux de la prise en compte des conséquences fonctionnelles et psychosociales auprès de cliniciens et de patients a montré que ces dimensions sont encore peu évaluées de façon systématique en pratique courante.

Nous avons sondé 119 praticiens prenant en charge des patients atteints d'un trouble de la parole au moyen d'un questionnaire. Parmi eux, 43 % présentaient un trouble de parole dû à une atteinte des fonctions motrices oro-faciales dont un cancer. Cette enquête a montré que l'impact fonctionnel au quotidien était finalement peu pris en compte par les cliniciens (Pommée et al., 2021a) : un quart n'évaluent pas dans leurs bilans les répercussions psychosociales du trouble de parole ressenties par les patients. De plus, quand ces répercussions sont évaluées, les outils utilisés sont souvent inadéquats. Les cliniciens rapportent dans 70 % des cas l'utilisation du Voice Handicap Index (Jacobson et al., 1997; Woisard et al., 2007) qui est un autoquestionnaire concernant les conséquences d'un trouble vocal et non de parole. Pour autant, des questionnaires plus spécifiques relatifs à la parole ont été validés et sont disponibles en français : le Speech Handicap Index ou SHI (Degroote et al., 2012; Rinkel et al., 2008), ou plus récemment le Phonation Handicap Index ou PHI (Balaguer et al., 2020b), validé spécifiquement en cancérologie ORL. Cependant, les répercussions psychosociales sont de plus en plus prises en compte dans les batteries d'évaluation clinique à disposition des orthophonistes : le SHI est par exemple inclus dans la version 2019 de la Batterie Clinique d'Évaluation de la Dysarthrie (BECD).

Du point de vue des patients, nous avons conduit une étude préliminaire auprès de 11 sujets (dont quatre ayant des troubles de la parole consécutivement à un cancer ORL) pour établir un état des lieux du bilan orthophonique des troubles de la parole du point de vue du patient. Les résultats montrent que la majorité (6/11) auraient souhaité que les conséquences du trouble de parole aient été davantage prises en compte lors de leur évaluation, deux participants considérant d'ailleurs qu'elles n'ont pas été suffisamment prises en compte.

Une seconde problématique de cette évaluation est relative aux moyens d'évaluations utilisés. D'ailleurs, bien qu'il s'agisse du symptôme le plus courant dans les cancers de la cavité buccale et de l'oropharynx (Plisson et al., 2017), peu d'outils sont disponibles pour l'évaluation des troubles de parole (Ghio et al., 2016).

À l'heure actuelle, l'évaluation clinique, principalement réalisée par des orthophonistes, reste essentiellement perceptive (Middag et al., 2014; Pommée et al., 2021a). Les thérapeutes s'appuient sur différentes méthodes de mesure pour quantifier la perception de la parole : échelles visuelles analogiques, échelles de Likert, estimations directes de la magnitude (Carmichael, 2007; Kent et al., 1989), transcription du stimulus (Hustad, 2006), ou encore évaluation de la compréhension de la parole (Hustad, 2008). Parmi toutes ces mesures, le score de sévérité du trouble de la parole perceptif sur une tâche de description d'image (c'est-à-dire de parole semi-spontanée) est le score le plus pertinent, dans le sens où il présente une fiabilité inter-juges satisfaisante, une limitation de l'effet plafond dans l'évaluation des troubles légers, et une dispersion suffisante pour permettre la mise en évidence de variabilité inter-individus (Balaguer et al., 2019b). Une standardisation de ces évaluations permettrait une meilleure reproductibilité de la mesure, mais les outils standardisés sont encore peu usités (Gurevich & Scamihorn, 2017) et peu sont validés dans le champ de la cancérologie. La Batterie d'Évaluation Clinique de la Dysarthrie ou BECD (Auzou & Rolland-Monnoury, 2019), le Frenchay Dysarthria Assessment (Enderby, 1980; Ghio et al., 2020) ou l'évaluation issue du projet MonPaGe (Pernon et al., 2020) sont validés auprès de patients neurologiques. En cancérologie, des batteries d'évaluation analytique et dynamique existent. Mais elles permettent une mesure en amont de la fonction de parole, au niveau des déficiences sous-jacentes : l'échelle de Motricité Bucco-Linguo-Faciale ou MBLF (Gatignol & Lannadère, 2010), ou la Grille d'Évaluation Motrice de la Cavité Orale ou GEMCO, non encore publiée (Pluvy, 2018).

Cette évaluation perceptive présente néanmoins de nombreuses limites. D'une part, le manque de standardisation de cette évaluation rend les mesures peu reproductibles et peu stables dans le temps ou entre cliniciens (Pommée et al., 2021a). En effet, la tâche-support de production de la parole (sur une voyelle tenue, la lecture d'un texte ou la description d'une image), les conditions d'enregistrement de la parole (quant à l'équipement d'enregistrement) ou le manque de consensus terminologique sur les concepts évalués – intelligibilité (Hustad, 2008; Lindblom, 1990), compréhensibilité (Ghio et al., 2018; Woisard et al., 2013), sévérité du trouble (Kent et al., 1989; Yorkston et al., 1996) – conduisent à des mesures très diverses et peu comparables. De plus, l'évaluation perceptive réalisée par un auditeur, même expert, n'est pas toujours représentative de la perception par les patients eux-mêmes de leur trouble de parole. En effet, le pourcentage d'accord exact entre l'évaluation perceptive externe et l'auto-évaluation du trouble de parole n'est que de 44,9 % (Gray et al., 2012). Enfin, l'évaluation perceptive quantitative d'un trouble de parole est variable selon le juge qui l'évalue, en raison des effets

d'expertise ou des différences de référents internes (Kim et al., 2011; Kuo & Tjaden, 2016; Middag, 2013). Au-delà de cette variabilité interindividuelle, des effets de variabilité intra-individu peuvent également être retrouvés : un même juge peut attribuer des scores différents selon le contexte d'évaluation, sa disponibilité mentale ou son habitude à la parole pathologique (Balaguer et al., 2019b; Fex, 1992).

Afin de pallier les biais induits par l'évaluation perceptive, de nouveaux outils d'évaluation automatique instrumentale de la parole se développent. Ils visent essentiellement à extraire du signal de parole des paramètres de caractérisation des dégradations acoustiques (Van Nuffelen et al., 2009).

En pratique clinique, différents outils sont utilisés, tels que le gratuiciel Praat³ (Boersma & van Heuven, 2001), le logiciel DIADOLAB (Menin-Sicard & Sicard, 2019) ou encore le système « Évaluation Vocale Assistée » EVA-2 (Ghio et al., 2008).

L'évaluation automatique de la parole pathologique est, de plus, un champ de recherche en plein essor. Ciblante l'analyse automatique de la parole après cancer, plusieurs projets nationaux ont été conduits ou sont encore en cours en France. Le projet INCA C2SI Carcinologic Speech Severity Index (Woisard et al., 2021) a permis d'aboutir à la construction d'un score automatique à partir de paramètres extraits de différentes tâches de production de parole chez des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. Le projet RUGBI⁴, encore en cours, vise à améliorer la mesure du déficit d'intelligibilité par une identification plus spécifique des unités linguistiques pertinentes d'un point de vue acoustique et prosodique, et l'identification de tâches linguistiques sensibles. À un niveau plus segmental, le projet DAPADAF-E⁵ (2020-2023) vise à évaluer la fiabilité d'une tâche de décodage acoustico-phonétique et sa validité de critère par rapport au diagnostic des déficits anatomiques et dynamiques oro-pharyngés, dans les troubles de la parole des sujets traités pour un cancer de la cavité buccale et de l'oropharynx. Enfin, au niveau européen, le consortium TAPAS⁶ regroupe des chercheurs et

³ Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.42, retrieved 15 April 2021 from <http://www.praat.org/>

⁴ <https://www.irit.fr/rugbi>, financé par l'Agence Nationale de la Recherche (ANR-18-CE45-0008), consulté le 07/05/2021

⁵ Financé par la DGOS (Direction Générale de l'Offre de Soins : PHRI-09-0004)

⁶ <https://www.tapas-etn-eu.org>, financé par le programme Horizon 2020 de l'Union européenne, et la bourse Marie Skłodowska-Curie (n°766287), consulté le 07/05/2021

cliniciens de différents pays européens autour de trois axes d'étude de la parole pathologique (dont les troubles de parole d'étiologie cancérologique) : détection des troubles de parole, prise en charge et outils basés sur la parole pour l'aide à la vie quotidienne.

Objectif du travail

Les troubles de la parole sont ainsi une problématique à laquelle les patients traités pour un cancer de la cavité buccale ou de l'oropharynx sont fréquemment confrontés. Une évaluation globale du trouble de parole tenant compte de l'impact qu'ils peuvent avoir au quotidien dans la communication avec leurs pairs est essentielle. Peu d'outils existent à cet effet à l'heure actuelle, et une évaluation basée sur des analyses automatiques de la parole permettrait de pallier les biais rencontrés dans le cadre des mesures perceptives.

L'objectif de notre étude est ainsi de **mesurer l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé**. En d'autres termes, il s'agit de trouver dans le signal de parole spontanée quels paramètres issus d'analyses automatiques permettent de prédire l'impact du trouble de parole sur les capacités fonctionnelles de communication des patients.

Un premier verrou scientifique concerne la mesure de l'impact du trouble de parole sur la communication. Actuellement, aucun indicateur n'est validé en cancérologie ORL pour cette mesure. Un score de référence doit donc être construit au moyen d'indicateurs disponibles, par exemple en proposant des autoquestionnaires aux patients.

Un deuxième verrou concerne la mesure du trouble de parole. Cette mesure doit être automatique afin de pallier les biais liés à l'évaluation perceptive. Or, d'une part, les mesures acoustiques et automatiques étudiées à l'heure actuelles sont principalement de bas niveau (au niveau phonémique notamment). Pourtant, des mesures de plus haut niveau (lexical, grammatical, discursif notamment) portent probablement des indices complémentaires sur l'impact sur la communication voire la qualité de vie, conformément au modèle de Wilson (Wilson & Cleary, 1995). D'autre part, les échantillons de parole étudiés concernent des tâches de parole contrainte, qui est peu représentative des productions usuelles des patients. Les outils automatiques existants et utilisables sur de la parole spontanée doivent ainsi être recherchés. Leur adaptation doit également être envisagée pour en améliorer leur pertinence et leurs performances.

Enfin, un dernier verrou concerne la prise en compte de facteurs associés au trouble de parole, à la communication et à la qualité de vie dans la prédiction de l'impact fonctionnel. En effet, les modélisations conceptuelles montrent que le lien entre trouble de parole et communication n'est pas direct et exclusif, et que de multiples facteurs individuels (perception des symptômes, déficits associés) et environnementaux (contexte social et psychologique) doivent être considérés dans cette relation. L'extension de la prédiction à la qualité de vie dans ce cadre conceptuel aboutit aux mêmes constats. Une recherche des outils existants permettant la mesure de ces facteurs associés doit ainsi être conduite. Une validation d'outils est à prévoir en cas d'absence de questionnaire validé ou d'outil insuffisamment précis dans le contexte cancérologique. La prédiction de l'altération de la communication après traitement d'un cancer oral ou oropharyngé pourra alors être menée, au moyen d'indicateurs mixtes : venant à la fois du signal de parole spontanée, mais aussi des facteurs associés.

Ce travail a pu être réalisé dans le cadre de l'étude RUGBI « *Recherche d'unités linguistiques pertinentes pour améliorer la mesure de l'intelligibilité de la parole altérée par des troubles de production pathologique* », concernant la prise en compte des aspects fonctionnels et de qualité de vie des patients présentant un trouble de parole. Ce projet a été approuvé par le Comité de Protection des Personnes (CPP : Ouest IV, 19/02/2020, référence 11/20_3). Après une première année de thèse réalisée grâce à un financement institutionnel de mon centre hospitalier de rattachement (Hôpitaux de Toulouse), le projet RUGBI a financé les deux dernières années de mon travail de recherche. Ma participation à ce projet m'a également permis de collaborer avec les différents partenaires et d'enrichir mes analyses et réflexions à leur contact : Laboratoire Octogone Lordat (Université Toulouse II Jean Jaurès), Laboratoire Parole et Langage (Aix-en-Provence), Laboratoire Informatique d'Avignon et Centre Hospitalier Universitaire de Toulouse.

Quatre parties composeront ce manuscrit.

La première partie concernera un état de l'art sur les effets des cancers oraux et oropharyngés sur la parole au moyen d'analyses acoustiques et automatiques. Pour cela, nous avons conduit une revue systématique de la littérature. Elle mettra en évidence un manque d'études sur deux points-clés : peu de paramètres linguistiques hors niveau phonético-phonologique sont considérés, et la parole spontanée est très peu étudiée alors qu'elle est pourtant la production la plus proche de la parole « quotidienne » des sujets.

Nous détaillerons dans une deuxième partie le corpus que nous avons constitué pour répondre à notre objectif. Ce corpus comprend des données issues d'autoquestionnaires complétés par les sujets, relatives à la perception de

leur trouble de parole, de leurs capacités de communication mais aussi des facteurs pouvant influencer la parole ou la communication. Il comprend également des enregistrements de parole des sujets, incluant l'enregistrement de parole spontanée, nous permettant de conduire les analyses à partir d'analyses automatiques du signal.

Une troisième partie sera relative à la recherche et à la sélection des paramètres issus d'analyses automatiques de la parole spontanée, selon les niveaux linguistiques et de communications définis par les modèles psycholinguistiques : signal de parole, composante articulatoire et phonémique, niveau morpho-syntaxique et dimensions pragma-sémantique et discursive.

Enfin, une quatrième et dernière partie concernera la prédiction de l'impact fonctionnel sur la communication au moyen des paramètres automatiques précédemment sélectionnés. Après avoir construit un score holistique de communication à partir des données issues des questionnaires complétés par les sujets, différentes modélisations ont été menées pour sélectionner les paramètres permettant la meilleure prédiction de ce score.

1 État de l'art : revue systématique de littérature

En raison des limites liées à l'évaluation perceptive de la parole, de nouveaux outils d'évaluation de la parole, basés sur des données objectives se développent (Berisha et al., 2013).

L'objectif de ce chapitre est de réaliser une revue systématique de la littérature concernant l'utilisation des outils acoustiques et automatiques sur le signal de parole pour mesurer les effets des cancers de la cavité buccale et de l'oropharynx. Cette revue permettra d'obtenir un état des lieux actuel et exhaustif de ces outils pouvant être utilisés dans le cadre de notre travail.

1.1 Matériel et méthode

1.1.1 Protocole et enregistrement du protocole

La méthodologie de recherche et l'analyse des résultats de cette revue systématique ont été guidées par les lignes directrices PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). La méthodologie PRISMA est conçue pour guider les chercheurs dans la réalisation d'une présentation rigoureuse et transparente des revues systématiques (Liberati et al., 2009; Moher et al., 2009).

Cette étude est enregistrée dans PROSPERO⁷ (plateforme d'enregistrement en ligne de revues de littérature, financée par le National Institute for Health Research et mise en place par le département académique CRD de l'Université de York) sous le numéro d'enregistrement CRD42019129598.

⁷ <https://www.crd.york.ac.uk/prospero/>, consulté le 10/05/2021

1.1.2 Critères d'éligibilité

Pour être inclus dans cette revue systématique, les articles devaient décrire les effets des cancers de la cavité buccale et de l'oropharynx sur la parole au moyen d'analyses acoustiques et/ou automatiques.

Ainsi, seuls les articles respectant les critères suivants ont été inclus :

- L'évaluation de la parole au sens large, combinant les concepts d'intelligibilité, c'est-à-dire de production de parole de bas niveau (Hustad, 2008; Keintz et al., 2007) et de compréhensibilité (Kent et al., 1989). Le trouble de parole est défini comme le déficit fonctionnel de la parole dégradant les capacités à interagir avec un tiers (Lindblom, 1990) ;
- L'utilisation de termes relatifs à l'acoustique ou au traitement automatique de la parole, comme « acoustic analysis », « phonetics », « signal processing », « sound spectrography » ...
- L'inclusion de patients traités d'un cancer de la cavité buccale ou de l'oropharynx.

Les critères d'exclusion étaient :

- L'absence de larynx original chez les sujets de l'étude : exclusion des patients ayant subi une laryngectomie totale ou partielle, ou porteurs de prothèse laryngée (l'absence de larynx conduisant à une altération de la voix et non directement de la parole) ;
- Les études incluant des populations d'enfants ;
- Tout article n'étant pas un article original, comme les résumés, les actes de conférences, les revues...
- Les études de cas ;
- Les articles non publiés en anglais.

1.1.3 Sources de données et stratégies de recherche

Une recherche de la littérature a été effectuée dans deux bases de données électroniques différentes, afin de récupérer l'ensemble des ressources pertinentes : PubMed et Embase. Ces deux bases de données ont été sélectionnées en fonction du sujet de cette recherche. À noter qu'une troisième base de données, Web of Science, n'a pas permis de récupérer de nouvelle référence.

Toutes les publications datées jusqu'au 4 décembre 2018, date de la recherche, ont été incluses, sans aucune limitation temporelle concernant les dates de publication.

Les termes de recherche sont listés dans le Tableau 1.

Tableau 1 - Bases de données et syntaxe de recherche (descripteurs et texte libre)

Base de données	Syntaxe de recherche (descripteurs et texte libre)	Nombre de références trouvées
PubMed	((("Speech"[Mesh] OR "Speech Sound Disorder"[Mesh] OR "Speech Disorders"[Mesh] OR "Articulation Disorders"[Mesh] OR "Voice"[Mesh] OR "Voice Quality"[Mesh] OR "Voice Disorders"[Mesh] OR "Hoarseness"[Mesh] OR "Aphonia"[Mesh] OR "Dysphonia"[Mesh] OR "Phonation"[Mesh]) OR "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]) AND ("Pharyngeal Neoplasms"[Mesh] OR "Mouth Neoplasms"[Mesh] OR "Oropharyngeal Neoplasms"[Mesh] OR "Facial Neoplasms"[Mesh] OR "Head and Neck Neoplasms"[Mesh] OR "Laryngeal Neoplasms"[Mesh] OR "Hypopharyngeal Neoplasms"[Mesh]))	296
Embase	((speech/ OR speech sound disorder/ OR voice/ OR dysphonia/ OR aphonia/ OR voice disorder/ OR hoarseness/ OR phonation/ OR speech intelligibility/) OR (Intelligibil*.ab. OR Intelligibil*.ti. OR comprehensibil*.ab. OR comprehensibil*.ti. OR understandabil*.ab. OR understandabil*.ti.)) AND (voice analysis/ OR voice onset time/ OR voice parameter/ OR 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/) AND ("head and neck cancer"/ OR "head and neck tumor"/ OR oropharynx tumor/ OR pharynx tumor/ OR oropharynx cancer/ OR oropharynx carcinoma/ OR pharynx carcinoma/ OR oropharynx squamous cell carcinoma/ OR pharynx cancer/ OR pharynx tumor/ OR mouth cancer/ OR mouth tumor/ OR salivary gland tumor/ OR tongue tumor/ OR tonsil tumor/ OR mouth carcinoma/ OR "head and neck carcinoma"/ OR mouth squamous cell carcinoma/ OR salivary gland carcinoma/ OR tongue carcinoma/ OR tonsil carcinoma/ OR face tumor/ OR face cancer/ OR larynx cancer/ OR larynx tumor/ OR larynx carcinoma/ OR hypopharynx cancer/ OR hypopharynx tumor/ OR hypopharynx carcinoma/ OR hypopharynx squamous cell carcinoma/)	262

Afin de limiter tout biais de sélection, l'ensemble des résumés récupérés dans les deux bases de données a été évalué par deux évaluateurs indépendants

ayant chacun contribué à la revue systématique de littérature de l'autre dont il est second auteur (Pommée et al., 2021b). Les différences d'opinion quant à l'éligibilité des articles ont fait l'objet de réunions de concertation afin d'aboutir à un consensus.

Un diagramme général représentant le processus de sélection conformément à la méthodologie PRISMA est à retrouver Figure 4.

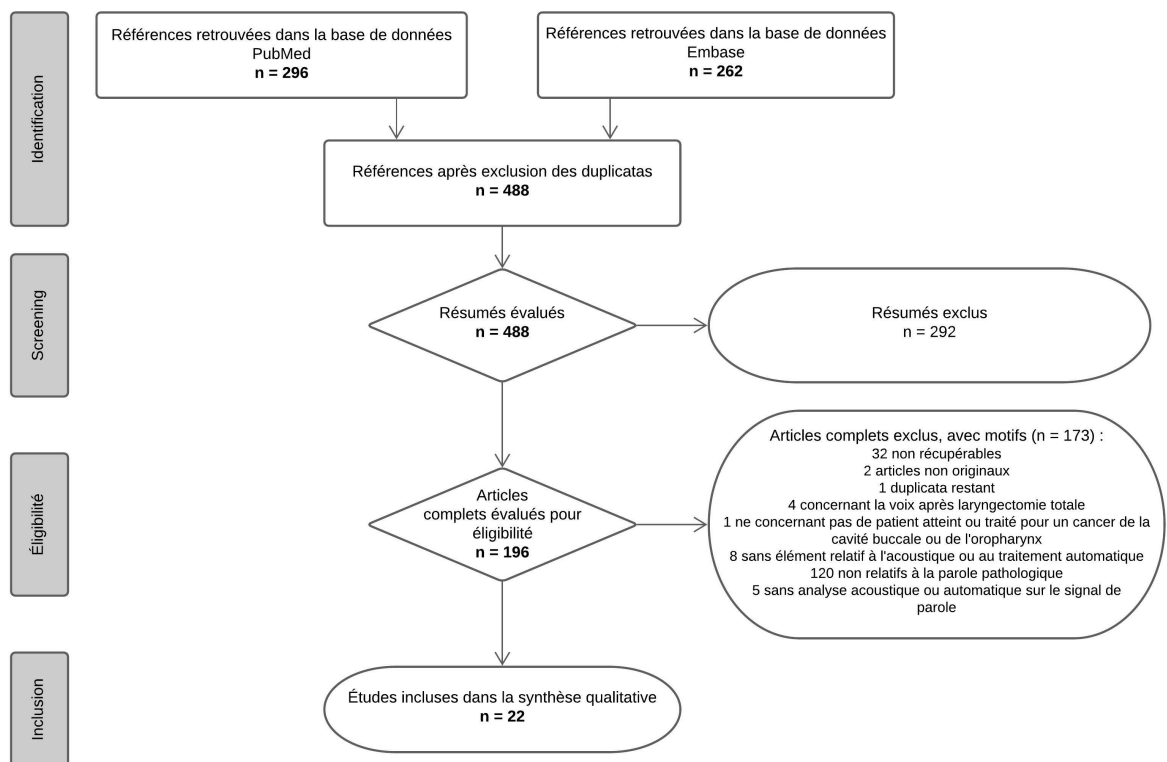


Figure 4 - Diagramme du processus de revue, selon la méthodologie PRISMA, adapté de Moher (Moher et al., 2009)

1.1.4 Qualité méthodologique et niveau de preuve

Le National Health and Medical Research Council (NHMRC) Evidence Hierarchy a été utilisé pour évaluer le niveau de preuve, de I (« Revues systématiques ») à IV (« Études de cas ») (NHMRC, 1999).

L'outil d'évaluation critique QualSyst de Kmet et al. (Kmet et al., 2004) fournit des moyens systématiques, reproductibles et quantitatifs afin d'évaluer la qualité méthodologique de la recherche sur un large éventail de modèles d'étude. Un score QualSyst supérieur à 80 % est interprété comme témoignant

d'une forte qualité, 60-79 % comme une bonne qualité, 50-59 % comme une qualité correcte, et moins de 50 % comme une qualité méthodologique médiocre. Les études de qualité méthodologique médiocre ont été exclues de l'analyse.

1.1.5 Extraction des données

Après évaluation de la qualité méthodologique, nous avons extrait des articles restants les données suivantes : nombre de participants à l'étude et leurs caractéristiques (âge, diagnostic et langue parlée), paramètres acoustiques / issus du traitement automatique du signal de parole (et leurs définitions), critère(s) de comparaison, échantillon de parole et principales conclusions des auteurs.

En outre, des données bibliométriques géographiques ont été extraites à l'aide de l'outil Netscity⁸ (projet Netscience du Labex SMS, Toulouse, France).

1.2 Résultats

1.2.1 Sélection des études

Un total de 488 références a été récupéré dans les deux bases de données électroniques.

Au cours d'une première étape, les deux examinateurs indépendants ont passé en revue l'ensemble des 488 références : nous avons vérifié l'éligibilité en évaluant chaque critère d'inclusion et d'exclusion pour chaque résumé. À l'issue de cette étape, 196 articles en texte intégral ont été récupérés, puis une nouvelle vérification de l'éligibilité selon la même méthodologie a été menée pour chaque article intégral. Finalement, 22 articles répondaient aux critères d'inclusion et ont été inclus dans cette étude (Figure 4).

1.2.2 Évaluation de la qualité

La qualité globale des études, telle qu'évaluée au moyen l'outil QualSyst, allait de « bonne » à « forte » : quatre études étant classées comme « bonnes » et 18 comme « fortes ».

⁸ <https://www.geotests.net/netscitypg/index.php>, consulté le 10/05/2021

Sur la base de la hiérarchie des preuves du NHMRC, 20 études ont été classées comme des preuves de niveau III (14 comme III-3 : « *Études comparatives avec 2 ou plusieurs études à un seul bras* » ; six comme III-2 : « *Études comparatives avec contrôles simultanés et allocation non randomisée (études de cohorte), ou études cas-témoins* »), et deux comme des preuves de niveau IV (« *Série de cas* »). Aucun article présentant un faible niveau de preuve n'a dû être exclu. Les évaluations des 22 articles inclus sont présentées dans le Tableau 2.

Tableau 2 - Niveau de preuve (NHMRC) et qualité méthodologique (QualSyst) des 22 articles inclus

Référence	Score QualSyst (%) *	Qualité méthodologique	Niveau de preuve NHMRC †
(Acher et al., 2014)	15/20 (75 %)	Bonne	IV
(Chung et al., 2011)	20/22 (91 %)	Forte	III-3
(de Bruijn et al., 2009)	21/24 (88 %)	Forte	III-2
(de Bruijn et al., 2011)	20/22 (91 %)	Forte	III-2
(de Carvalho-Teles et al., 2008)	18/22 (82 %)	Forte	III-3
(Dwivedi et al., 2016)	24/24 (100 %)	Forte	III-2
(Fang et al., 2017)	14/20 (70 %)	Bonne	III-2
(Ha et al., 2016)	20/22 (91 %)	Forte	III-2
(Jacobi et al., 2013)	19/20 (95 %)	Forte	III-3
(Kazi et al., 2007)	20/20 (100 %)	Forte	III-2
(Kim et al., 2014)	19/22 (86 %)	Forte	III-3
(Knuuttila & Juhani Pukander, 1999)	16/20 (80 %)	Forte	III-3
(Kraaijenga et al., 2015)	19/22 (86 %)	Forte	III-3
(Kumar et al., 2013)	16/22 (73 %)	Bonne	III-3
(Laaksonen et al., 2011)	21/22 (95 %)	Forte	III-3
(Laaksonen et al., 2010)	17/20 (85 %)	Forte	III-3
(Markkanen-Leppänen et al., 2005)	22/22 (100 %)	Forte	III-3
(Moerman et al., 2003)	14/22 (64 %)	Bonne	IV
(Seikaly et al., 2003)	17/20 (85 %)	Forte	III-3
(Takatsu et al., 2017)	19/20 (95 %)	Forte	III-3
(Wakumoto et al., 1996)	18/20 (90 %)	Forte	III-3
(Yoshida et al., 2000)	19/20 (95 %)	Forte	III-3

* Qualité méthodologique : Forte > 80 % ; Bonne 60–79 % ; Correcte 50–59 % ; Faible < 50 %.

† Niveaux de hiérarchie des preuves (NHMRC – légende complète en **Annexe I**) : Niveau III-2 Études comparatives avec groupe contrôlé et allocation des sujets non randomisée (études de cohorte), études cas témoins, ou séries chronologiques

interrompues avec groupe contrôle ; Niveau III-3 Études comparatives avec cohorte contrôle historique, deux études or plus à bras unique, ou séries chronologiques interrompues sans groupe contrôle ; Niveau IV Études de cas.

Le tableau complet des résultats des 22 études retenues se trouve en **Annexe I**.

1.2.3 Données bibliométriques

Le champ d'étude concernant l'analyse des paramètres acoustiques et issus du traitement automatique du signal de parole des patients traités pour un cancer de la cavité buccale ou de l'oropharynx concerne principalement des équipes situées dans trois zones géographiques : l'Europe occidentale (principalement les Pays-Bas), l'Amérique du Nord et l'Extrême-Orient (Japon et Corée du Sud). Quelques collaborations entre équipes sont notées : entre la Finlande et le Canada, et entre la Corée du Sud et les États-Unis (Figure 5).

Cette répartition géographique aura une influence sur les langues des échantillons de parole collectés.

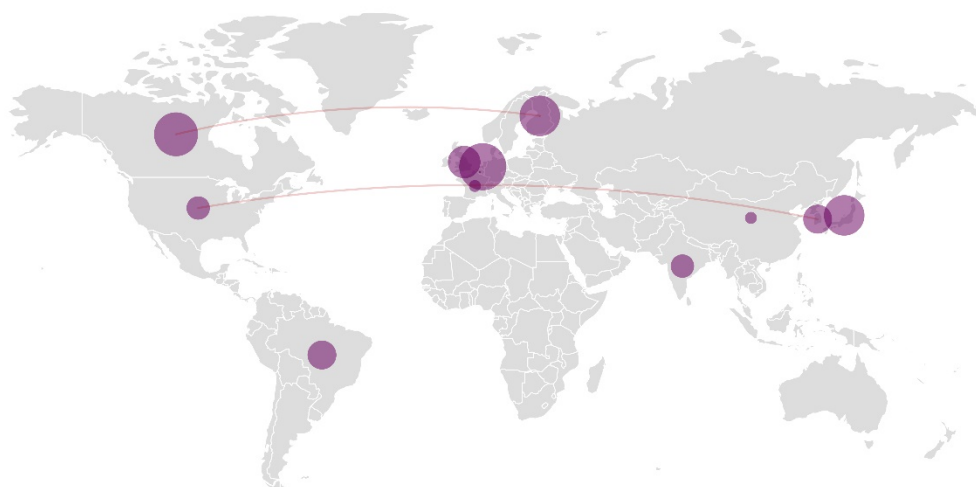


Figure 5 - Localisation des affiliations des auteurs, et collaborations entre équipes de recherche

La plupart des études retenues dans cette revue ont été publiées depuis 2010 (13/22, 59 %). Le traitement automatique de la parole et l'utilisation d'outils d'apprentissage dans le cadre de l'évaluation dans un contexte oncologique est plus récent, et a commencé vers 2010 (Figure 6). Le domaine de l'analyse automatique de la parole pathologique est donc en pleine expansion, en raison du développement récent de nouveaux outils et de nouvelles mesures.

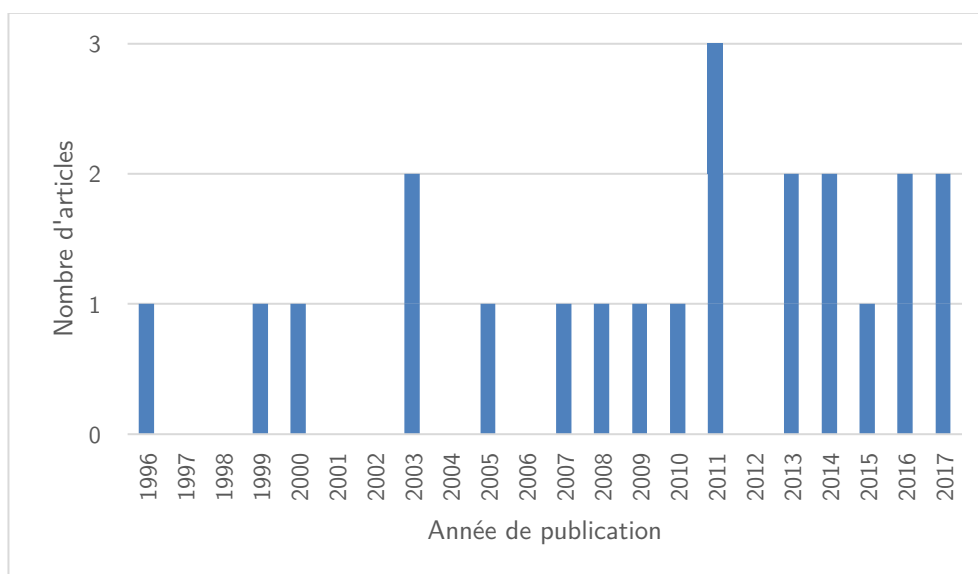


Figure 6 - Nombre d'articles publiés par année

1.2.4 Caractéristiques des participants aux études sélectionnées

Parmi les 22 études, dix incluent plus de 20 patients (Chung et al., 2011; de Bruijn et al., 2011, 2009; de Carvalho-Teles et al., 2008; Dwivedi et al., 2016; Jacobi et al., 2013; Kazi et al., 2007; Kim et al., 2014; Markkanen-Leppänen et al., 2005; Takatsu et al., 2017), et également dix incluent entre deux et 18 sujets (Acher et al., 2014; Ha et al., 2016; Knuuttila & Juhani Pukander, 1999; Kumar et al., 2013; Laaksonen et al., 2011, 2010; Moerman et al., 2003; Seikaly et al., 2003; Wakumoto et al., 1996; Yoshida et al., 2000). Deux articles ne rapportent pas le nombre de sujets impliqués dans l'analyse (Fang et al., 2017; Kraaijenga et al., 2015). Les détails sont donnés dans le Tableau 3.

Tableau 3 - Nombre total de participants dans les études retenues

Nombre de participants	Nombre d'études (%)
2 à 5 participants	3 (14 %)
6 à 10 participants	2 (9 %)
11 à 20 participants	5 (23 %)
21 à 50 participants	4 (18 %)
51 à 62 participants	6 (27 %)
Non précisé	2 (9 %)

Notons qu'à l'exception de deux études (Acher et al., 2014; Yoshida et al., 2000), les sujets inclus sont majoritairement des hommes.

Deux études (Fang et al., 2017; Kim et al., 2014) utilisent des données rétrospectives.

Tous les participants des 22 études avaient été traités d'un cancer de la cavité buccale ou de l'oropharynx au moment de l'étude (voir Figure 7).

Au total, 11 études (50 %) portent sur des patients dont la localisation tumorale se situait dans la cavité buccale uniquement. Les sites anatomiques concernent principalement la langue (9/11) : après traitement par glossectomie totale (de Carvalho-Teles et al., 2008; Knuuttila & Juhani Pukander, 1999; Laaksonen et al., 2011, 2010; Takatsu et al., 2017; Wakumoto et al., 1996), ou partielle (Acher et al., 2014; Ha et al., 2016; Kazi et al., 2007). Les deux autres études portent sur des tumeurs maxillaires (Kumar et al., 2013; Yoshida et al., 2000).

Six études (27 %) incluent à la fois des patients traités pour un cancer de la cavité buccale et d'autres traités pour un cancer de l'oropharynx (de Bruijn et al., 2011, 2009; Dwivedi et al., 2016; Fang et al., 2017; Kim et al., 2014; Markkanen-Leppänen et al., 2005).

Seulement cinq (23 %) n'incluent que des patients avec une localisation tumorale oropharyngée. Deux s'intéressent à des patients présentant une extension tumorale au palais mou (Moerman et al., 2003; Seikaly et al., 2003). Les trois autres concernent l'amygdale, seule (Chung et al., 2011) ou en comparaison avec la zone de la base de la langue (Jacobi et al., 2013; Kraaijenga et al., 2015).

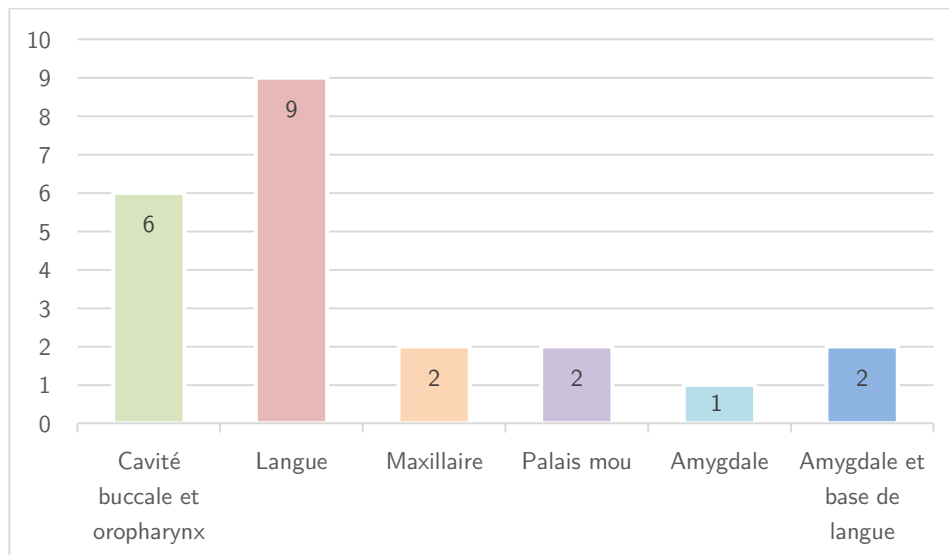


Figure 7 - Localisations tumorales

En ce qui concerne la taille de la tumeur, 12 études (54 %) incluent des tumeurs de petite taille (T1 + T2⁹) (Chung et al., 2011; de Bruijn et al., 2011, 2009; Dwivedi et al., 2016; Ha et al., 2016; Jacobi et al., 2013; Knuuttila & Juhani Pukander, 1999; Laaksonen et al., 2011, 2010; Markkanen-Leppänen et al., 2005; Takatsu et al., 2017; Wakumoto et al., 1996). Trois études (14 %) se concentrent sur des tumeurs plus importantes (T3 + T4) (Fang et al., 2017; Kim et al., 2014; Seikaly et al., 2003). Une étude (5 %) inclut autant de sujets présentant des T1 + T2 que des tumeurs T3 + T4 (T3 + T4) (Moerman et al., 2003). Enfin, six études (27 %) ne rapportent pas la taille de la tumeur des participants (Acher et al., 2014; de Carvalho-Teles et al., 2008; Kazi et al., 2007; Kraaijenga et al., 2015; Kumar et al., 2013; Yoshida et al., 2000).

La Figure 8 montre la proportion détaillée des tailles des tumeurs dans les études qui ont rapporté cette information.

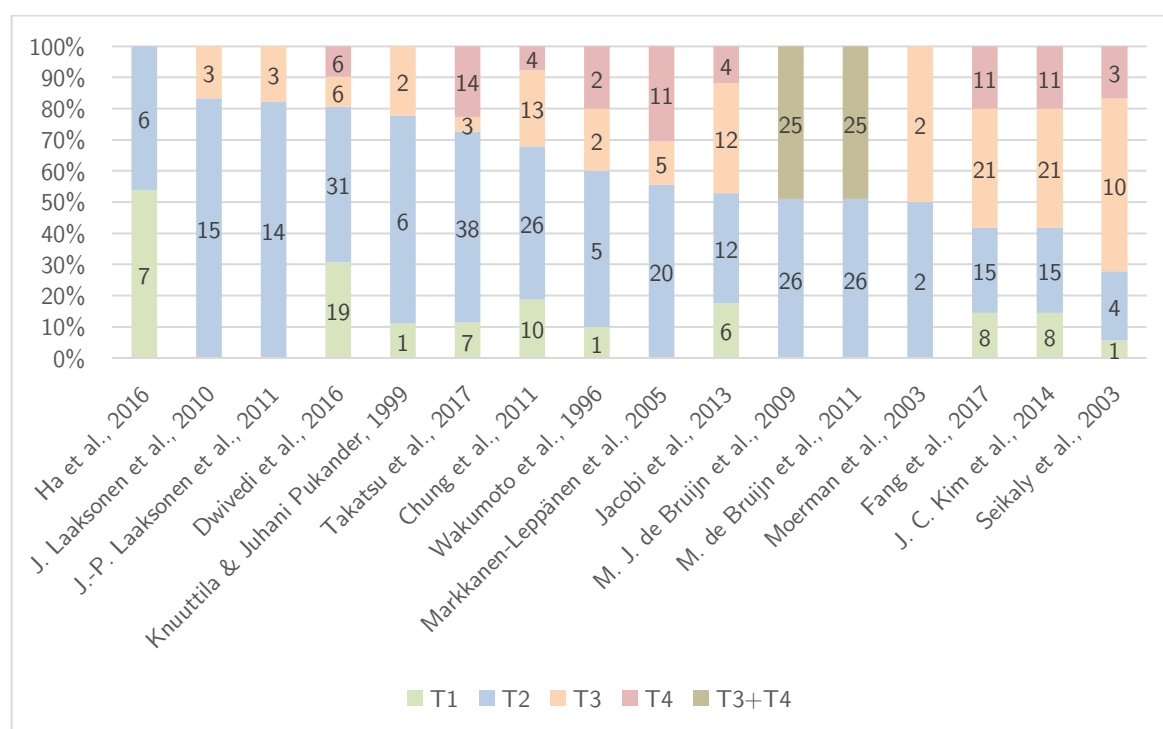


Figure 8 - Proportion détaillée des tumeurs selon leur taille (critère T de la classification TNM) dans les articles retenus (attention : les articles de de Bruijn (de Bruijn et al., 2011, 2009) ne différencient pas les tumeurs T3 et T4 entre les sujets)

⁹ Classification TNM: <https://www.onco-hdf.fr/app/uploads/2019/11/R%C3%A9f%C3%A9rentiel-r%C3%A9gional-VADS-VF.pdf>, consulté le 10/05/2021

Parmi les 22 études incluses, 20 (91 %) portent sur des patients traités chirurgicalement. Parmi elles, la chirurgie a été réalisée exclusivement (sans information rapportée sur le traitement complémentaire) dans 14 études (Chung et al., 2011; de Bruijn et al., 2011, 2009; de Carvalho-Teles et al., 2008; Dwivedi et al., 2016; Ha et al., 2016; Kazi et al., 2007; Kumar et al., 2013; Laaksonen et al., 2011, 2010; Markkanen-Leppänen et al., 2005; Takatsu et al., 2017; Wakumoto et al., 1996; Yoshida et al., 2000), dont sept avec reconstructions chirurgicales. La chirurgie a été associée à d'autres méthodes de traitement comme la radiothérapie ou la chimiothérapie dans six études (Acher et al., 2014; Fang et al., 2017; Kim et al., 2014; Knuuttila & Juhani Pukander, 1999; Moerman et al., 2003; Seikaly et al., 2003). Enfin, une radiochimiothérapie seule a été réalisée pour les participants de deux études (9 %) (Jacobi et al., 2013; Kraaijenga et al., 2015).

Les principales langues parlées par les sujets et donc retrouvées dans les échantillons de parole sont l'anglais (Dwivedi et al., 2016; Ha et al., 2016; Laaksonen et al., 2011, 2010) – y compris l'anglais américain et l'anglais canadien – et le néerlandais dans quatre études (de Bruijn et al., 2011, 2009; Jacobi et al., 2013; Kim et al., 2014). Les autres études sont toutes réalisées dans des langues différentes : français (Acher et al., 2014), portugais (de Carvalho-Teles et al., 2008), allemand (Fang et al., 2017), finnois (Knuuttila & Juhani Pukander, 1999), hindi (Kumar et al., 2013), japonais (Takatsu et al., 2017) et coréen (Chung et al., 2011). Sept études ne mentionnent pas la langue (Kazi et al., 2007; Kraaijenga et al., 2015; Markkanen-Leppänen et al., 2005; Moerman et al., 2003; Seikaly et al., 2003; Wakumoto et al., 1996; Yoshida et al., 2000).

1.2.5 Critères de comparaison

Les différents critères de comparaison utilisés dans les études sont présentés dans le Tableau 4.

Six études (27 %) comparent des mesures acoustiques ou automatiques avec des scores perceptifs. Ce sont des scores d'intelligibilité attribués par des juges à l'aide d'une échelle ordinale de type Likert, soit de façon globale (Acher et al., 2014), soit sur des paramètres spécifiques tels que l'articulation, la nasalité ou la « faiblesse » (de Bruijn et al., 2011, 2009; Dwivedi et al., 2016; Yoshida et al., 2000). Une étude utilise le pourcentage de consonnes correctement identifiées (Chung et al., 2011).

Tableau 4 - Critères de comparaison choisis par les auteurs des études incluses

Critère de comparaison	Nombre d'études (%)
Évaluation perceptive	6 (27 %)
Mesure globale de l'intelligibilité	1
Paramètres spécifiques : articulation, nasalité, « faiblesse »	4
Pourcentage d'identification correcte des consonnes	1
Différences entre paramètres acoustiques / automatiques	5 (23 %)
Formants	1
Paramètres spectraux	2
Paramètres issus d'un autre analyse acoustique ou normes existantes	2
Paramètres acoustiques / automatiques identiques avant / après traitement	3 (14 %)
Paramètres acoustiques / automatiques identiques entre sujets et contrôles	8 (36 %)

Cinq études (23 %) étudient la performance des mesures acoustiques soit en analysant les différences entre les paramètres étudiés, soit en comparant les résultats avec des données existantes : comparaison des formants (de Carvalho-Teles et al., 2008), comparaison de la performance de deux paramètres spectraux (Fang et al., 2017; Kim et al., 2014) ou comparaison avec les mêmes paramètres d'autres logiciels ou avec des normes existantes (Kumar et al., 2013; Moerman et al., 2003).

Trois études (14 %) comparent les paramètres acoustiques avant et après traitement (Jacobi et al., 2013; Knuuttila & Juhani Pukander, 1999; Kraaijenga et al., 2015). Enfin, huit études (36 %) comparent les mêmes paramètres entre un groupe de sujets et un groupe contrôle (Ha et al., 2016; Kazi et al., 2007; Laaksonen et al., 2011, 2010; Markkanen-Leppänen et al., 2005; Seikaly et al., 2003; Takatsu et al., 2017; Wakumoto et al., 1996).

1.2.6 Échantillons de parole

Quatorze études mesurent des paramètres acoustiques issus de l'analyse de phonèmes isolés. Plus précisément, huit analysent des voyelles isolées (Acher et al., 2014; de Carvalho-Teles et al., 2008; Dwivedi et al., 2016; Kazi et al., 2007; Knuuttila & Juhani Pukander, 1999; Kraaijenga et al., 2015; Kumar et al., 2013; Yoshida et al., 2000), trois analysent des phonèmes extraits d'un texte lu (de Bruijn et al., 2011, 2009; Laaksonen et al., 2011) et deux des phonèmes issus de mots isolés (Ha et al., 2016; Laaksonen et al., 2010). Une étude analyse à la fois des phonèmes isolés et d'autres extraits de mots (analyse des formants et de leurs transitions (Takatsu et al., 2017)).

L'échantillon de parole d'une étude consiste en la production de syllabes (Wakumoto et al., 1996), tandis qu'une autre contient des productions de diadococinésies (Jacobi et al., 2013).

Une étude réalise des analyses au niveau de la phrase (Markkanen-Leppänen et al., 2005), et quatre réalisent des mesures sur un texte lu (Chung et al., 2011; Fang et al., 2017; Kim et al., 2014; Moerman et al., 2003).

Enfin, une étude ne rapporte pas la composition de son échantillon de parole (Seikaly et al., 2003).

Ces résultats sont présentés dans le Tableau 5.

Tableau 5 - Constitution des échantillons de parole

Échantillon de parole	Nombre d'études (%)
Phonèmes isolés	14 (64 %)
Voyelles tenues	8
Phonèmes extraits d'un texte lu	3
Phonèmes extraits de mots isolés	2
Combinaison de voyelles tenues et de phonèmes extraits de mots	1
Syllabes et diadococinésies	2 (9 %)
Phrases	1 (5 %)
Texte lu	4 (17 %)
Non précisé	1 (5 %)

1.2.7 Mesures acoustiques

Les paramètres acoustiques analysés dans les études incluses, rapportés ci-dessous, sont présentés en **Annexe II**.

La Figure 9 représente la distribution des unités d'analyse dans les articles.

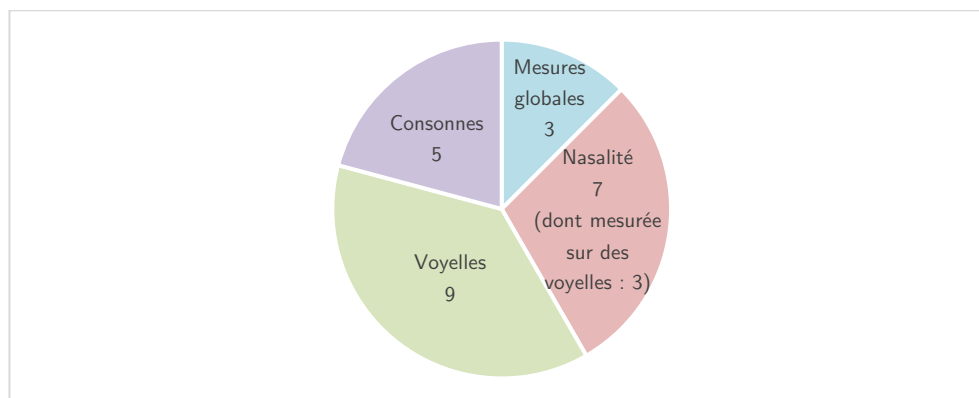


Figure 9 - Nombre d'études selon les catégories d'analyse (la somme est supérieure au nombre total car certaines études analysent différentes catégories)

1.2.7.1 Nasalité (7 articles)

Sept articles se concentrent sur l'analyse de la nasalité. Trois études ont analysé la nasalité sur des voyelles (Jacobi et al., 2013; Kraaijenga et al., 2015; Kumar et al., 2013), les autres l'ont étudiée de façon plus globale sur des phrases (Markkanen-Leppänen et al., 2005) et deux sur un texte lu (Chung et al., 2011; Moerman et al., 2003). Une étude n'indique pas l'unité d'analyse utilisée (Seikaly et al., 2003).

La plupart des études calculent un score de nasalité en utilisant un logiciel dédié : Praat (Kraaijenga et al., 2015), Dr. Speech (Kumar et al., 2013). D'autres utilisent des nasomètres (Markkanen-Leppänen et al., 2005; Moerman et al., 2003; Seikaly et al., 2003). Le rapport de l'énergie acoustique provenant de la cavité nasale et de la cavité orale est calculé dans deux études (Chung et al., 2011; Jacobi et al., 2013).

Le score de nasalité présente une association significative avec l'évaluation perceptive, en cas de résection étendue ou de reconstruction du palais mou (Chung et al., 2011). Quatre autres études montrent une augmentation de la nasalité après traitement (Jacobi et al., 2013; Kraaijenga et al., 2015; Kumar et al., 2013; Seikaly et al., 2003). Une étude montre que les tumeurs de la cavité orale n'ont pas d'impact significatif sur la nasalité contrairement aux tumeurs oropharyngées (Markkanen-Leppänen et al., 2005).

1.2.7.2 Voyelles (9 articles)

Neuf articles étudient les premiers et seconds formants (F1 et F2) des voyelles (de Bruijn et al., 2009; de Carvalho-Teles et al., 2008; Dwivedi et al., 2016; Ha et al., 2016; Kazi et al., 2007; Knuuttila & Juhani Pukander, 1999; Laaksonen et al., 2010; Takatsu et al., 2017; Yoshida et al., 2000). Parmi celles-ci, trois étudient également le F3 (de Carvalho-Teles et al., 2008; Ha et al., 2016; Kazi et al., 2007) et une analyse les formants jusqu'au F12 (Yoshida et al., 2000). L'aire du triangle vocalique (Vowel Space Area, VSA) est utilisée dans deux études (de Bruijn et al., 2009; Takatsu et al., 2017) et la pente spectrale (ou slope) n'est retrouvée que dans une seule (Takatsu et al., 2017).

Quatre études examinent les différences acoustiques avant et après traitement. Après une chirurgie de la langue, des différences significatives sont trouvées dans F1 et F2 (Knuuttila & Juhani Pukander, 1999; Laaksonen et al., 2010), F1 étant généralement augmenté et F2 abaissé. Les mesures acoustiques sont impactées par la reconstruction locale (Takatsu et al., 2017), ainsi que par la mise en place de prothèses d'abaissement du palais, dont il est démontré qu'elles modifient F1, F2 et F3 chez les patients traités pour une glossectomie subtotale (de Carvalho-Teles et al., 2008). Deux études montrent une corrélation entre les mesures acoustiques et l'intelligibilité perçue : F2 du /i/ et la taille de l'air vocalique sont liés aux évaluations de l'intelligibilité ($r = 0,39$,

$p < 0,05$) et de l'articulation ($r = 0,42$, $p < 0,05$) (de Bruijn et al., 2009), et F7 et F12 du /i/ sont également fortement corrélés aux évaluations perceptives ($r = 0,84$) (Yoshida et al., 2000). Une seule étude ne trouve pas de corrélation significative entre mesures acoustiques et évaluation perceptive sur F0, F1 et F2 (Dwivedi et al., 2016).

Les études comparant les sujets et les témoins sains trouvent que F2 et F3 sont plus faibles dans le groupe des patients (Ha et al., 2016). Chez les femmes, des corrélations significatives sont trouvées entre les sujets et les contrôles pour F2 et F3, mais seulement pour F1 chez les hommes (Kazi et al., 2007).

1.2.7.3 Consonnes (5 articles)

Certaines études analysent les moments spectraux sur les plosives et les fricatives : le centre de gravité/moyenne spectrale (Acher et al., 2014; Jacobi et al., 2013; Laaksonen et al., 2011) et l'asymétrie spectrale (skewness) (Acher et al., 2014; de Bruijn et al., 2009; Laaksonen et al., 2011). Le Klatt Voice Onset Time (V.O.T.) est également analysé sur les deux groupes de consonnes dans une étude (Acher et al., 2014).

Sur les plosives, la durée de relâchement de la pression de l'air est mesurée dans deux études (de Bruijn et al., 2009; Jacobi et al., 2013). La fréquence d'énergie de pointe (consonant peak energy frequency) de la consonne /t/ et la transition des formants dans la syllabe /ta/ sont analysées dans une étude (Wakumoto et al., 1996).

Sur les fricatives, la durée de friction et l'énergie de bande sont calculées dans deux études (Jacobi et al., 2013; Laaksonen et al., 2011). Dans une étude (Jacobi et al., 2013), F1, F2 et F3 sont mesurées sur les consonnes /l/ et /r/.

Les résultats montrent que la durée du relâchement de la pression de l'air sur /k/ est liée aux mesures d'intelligibilité et d'articulation (de Bruijn et al., 2009). De même, le centre de gravité et l'asymétrie spectrale sont corrélés avec l'évaluation perceptive dans des contextes spécifiques (contexte iCi et aCa) (Acher et al., 2014). La comparaison pré et post-traitement permet de considérer la moyenne et l'asymétrie spectrales comme de bonnes mesures des effets à court terme, mais la durée de friction sur /s, z/ ne semble pas être pertinente pour les effets à long terme. Un an après la radiochimiothérapie, le pic fréquentiel de l'explosion (ou burst) du /k/ est réduite, F3 est significativement plus élevé avec une intensité plus faible sur /l/, et une fréquence de l'explosion est significativement plus élevée sur /t/ (Jacobi et al., 2013). Dans différents contextes, le V.O.T. semble congruent avec l'évaluation perceptive (Acher et al., 2014). Enfin, la variance des formants F2-F3 à la transition entre plosive et voyelle revient à la normale après chirurgie, et la fréquence d'énergie de pointe des consonnes est plus basse qu'avant la chirurgie pour certains sujets (Wakumoto et al., 1996).

1.2.7.4 Mesures globales (3 articles)

Deux articles étudient la performance de différentes caractéristiques acoustiques, calculées à partir de corpus existants, afin de classer la parole en deux catégories (intelligible / inintelligible). Les caractéristiques étudiées sont les MFCC (Mel-Frequency Cepstrum Coefficients) et MSCC (Mel S-transform Cepstrum Coefficients) dans (Fang et al., 2017), et le MRSTC (Multi Resolution Sinusoidal Transform Coding) dans (Kim et al., 2014). Ces paramètres sont transmis à différents algorithmes de classification qui produisent une décision binaire sur l'intelligibilité : l'article de (Kim et al., 2014) utilise un classificateur basé sur des techniques de régression, l'article de (Fang et al., 2017) utilise un séparateur à vaste marge (SVM : Support Vector Machine). Un troisième article utilise un réseau de neurones artificiel (ANN : Artificial Neural Network) pour prédire la qualité de l'articulation et la nasalité (de Bruijn et al., 2011).

Les MSCC donnent de meilleurs résultats que les MFCC dans la classification de la parole intelligible et inintelligible sur des corpus rétrospectifs, et les MRSTC montrent une meilleure classification lorsqu'ils sont soumis à un SVM (Kim et al., 2014). Les ANN permettent de prédire la qualité d'articulation perçue sur /a/, ainsi que l'hypernasalité perceptive sur /i/ et /u/ (de Bruijn et al., 2011).

1.3 Discussion

L'objectif principal de cette étude était de réaliser une revue systématique de la littérature scientifique étudiant les effets du cancer de la cavité buccale ou de l'oropharynx sur la parole, en utilisant des paramètres acoustiques ou issus d'un traitement automatique du signal.

Deux grandes lignes de réflexion émergent de l'analyse des 22 articles sélectionnés, concernant le choix des paramètres, et l'unité d'analyse choisie pour évaluer la parole.

1.3.1 Choix des paramètres acoustiques / automatiques

En regardant les analyses les plus utilisées dans les études retenues pour cette revue, deux domaines principaux peuvent être repérés : les mesures de nasalité et l'analyse acoustique des voyelles.

Les analyses selon le type de traitement et la langue utilisée n'ont pas révélé de tendance, notamment en raison du faible nombre d'études et de patients dans chaque catégorie. Seules deux langues se retrouvent dans plus

d'un article : l'anglais et le néerlandais (quatre études). Parmi elles, deux analysent les mêmes voyelles cardinales et les deux premiers formants (de Bruijn et al., 2009; Dwivedi et al., 2016) : elles montrent des résultats différents concernant la corrélation entre ces scores et l'évaluation perceptive de l'intelligibilité. D'autres études sont donc nécessaires pour étudier spécifiquement l'effet de la constitution phonémique d'une langue sur l'intelligibilité des patients après traitement.

La localisation de la tumeur joue un rôle dans le choix du paramètre étudié. La plupart des études incluant des patients traités pour un cancer de l'oropharynx utilise la mesure de la nasalité comme l'un des critères de mesure de parole. Parmi ces études, cinq n'incluent que le cancer de l'oropharynx (Chung et al., 2011; de Bruijn et al., 2011; Kraaijenga et al., 2015; Moerman et al., 2003; Seikaly et al., 2003) et deux incluent également des patients ayant été traités par une chirurgie de la cavité buccale (Jacobi et al., 2013; Markkanen-Leppänen et al., 2005). La pathologie oropharyngée, du fait de sa localisation, a un impact sur la dynamique des structures anatomiques impliquées dans la nasalité de la parole, notamment par son effet sur le palais mou ou l'amygdale.

Au-delà de la localisation tumorale, l'étude de Mayo (Mayo & Mayo, 2011) a montré que la langue parlée influence également la valeur moyenne de nasalité. Dans notre étude toutefois, il est difficile de conclure sur l'utilisation d'outils automatiques mesurant la nasalité selon la langue. En effet, parmi les sept études s'y intéressant, seules trois rapportent la langue utilisée dans l'échantillon de parole : une en néerlandais (Jacobi et al., 2013), une en hindi (Kumar et al., 2013) et une en coréen (Chung et al., 2011). Pour autant, la question du choix et de l'adaptation des outils acoustiques et automatiques à une langue ou à un groupe de langues selon sa construction linguistique reste une piste à explorer.

La majorité des études incluant des patients traités pour un cancer de la cavité buccale analyse les voyelles et les consonnes. Si la nasalité est principalement évaluée au niveau de la phrase ou du texte, la plupart des autres analyses se concentrent en revanche sur les caractéristiques acoustiques de voyelles, produites isolément ou plus rarement extraites de syllabes ou de la parole continue. Les analyses sont principalement effectuées sur les premiers formants, dont nous savons qu'ils sont directement impactés par la pathologie orale : l'ouverture de la mâchoire modifie le premier formant et la position de la langue modifie le deuxième formant. Les études faisant le lien entre ces mesures formantiques et l'intelligibilité perçue (des critères de comparaison perceptive sont utilisés dans trois articles sur neuf traitant des mesures de formants) mettent en avant l'intérêt de trois paramètres principaux : l'aire du triangle vocalique (de Bruijn et al., 2009), le deuxième formant sur la voyelle

/i/ et les scores de nasalité issus de réseaux de neurones calculés sur la voyelle /i/ (de Bruijn et al., 2011).

Les caractéristiques acoustiques des consonnes induisent l'utilisation de différents paramètres acoustiques spécifiques. Sur les plosives [p t k], l'analyse spectrale de l'explosion et le relâchement de la pression de l'air semblent pertinents (de Bruijn et al., 2009; Jacobi et al., 2013). Le centre de gravité, la pente spectrale et l'énergie de bande sont plus couramment utilisés pour les fricatives (de Bruijn et al., 2009; Jacobi et al., 2013; Laaksonen et al., 2010; Wakumoto et al., 1996).

Ainsi, les paramètres acoustiques analysés dépendent de la localisation de la tumeur : les analyses sur les voyelles et les consonnes concernent principalement les patients traités pour un cancer oral, alors que l'étude de la nasalité concerne surtout les patients traités pour un cancer de l'oropharynx. Ceci est conforme aux conséquences sur la parole, liées aux changements morphologiques et dynamiques consécutifs au traitement. Il semble donc approprié d'adapter le choix des paramètres acoustiques et automatiques à la pathologie présentée par le patient lors de l'évaluation clinique.

Plusieurs paramètres acoustiques permettent de différencier les sujets atteints d'un cancer de la cavité buccale ou de l'oropharynx, des témoins sains. C'est le cas de l'analyse des formants principalement dans les cancers de la cavité orale (Ha et al., 2016; Kazi et al., 2007; Laaksonen et al., 2011; Takatsu et al., 2017; Wakumoto et al., 1996), mais aussi des scores de nasalité dans deux études (Markkanen-Leppänen et al., 2005; Seikaly et al., 2003). La validité clinique de ces mesures a ainsi été vérifiée. D'autres paramètres permettent de mesurer un changement avant / après traitement, comme les fréquences de l'explosion sur /t/ et /k/ (Jacobi et al., 2013) ou les scores de nasalité (Kraaijenga et al., 2015) pour les patients atteints de cancer de l'oropharynx, et des F1 et F2 pour les patients atteints de cancer de la cavité buccale (Knuuttila & Juhani Pukander, 1999). Ces paramètres montrent donc une bonne sensibilité au changement. Cependant, la question de la mesure de référence à laquelle comparer ces paramètres (permettant de vérifier la validité de critère) est essentielle. Six études choisissent des mesures perceptives comme référence, mais les limites liées à la variabilité des mesures humaines laissent ouverte cette question.

En ce qui concerne la taille de la tumeur, la parole dans le contexte des tumeurs de petite taille est principalement analysée sur les voyelles (principalement l'analyse des formants), et sur les consonnes (moments spectraux). La nasalité n'est étudiée que dans une seule étude sur les voyelles, au moyen d'un réseau de neurones (de Bruijn et al., 2011).

Les trois études incluant des tumeurs plus importantes (Fang et al., 2017; Kim et al., 2014; Seikaly et al., 2003) utilisent majoritairement des extractions de coefficients issus du signal audio (MFCC, MSCC, MRSTC) (Fang et al., 2017; Kim et al., 2014).

L'utilisation des techniques de traitement automatique du signal de parole (notamment au moyen de réseaux de neurones) est assez récente dans le domaine de l'évaluation de la parole et montre des performances prometteuses en termes de classification binaire intelligible / inintelligible, avec le jugement perceptif comme critère de validation externe.

La taille de la tumeur, en fonction de son impact sur les structures anatomiques impliquées dans la production de la parole, semble déterminer les critères acoustiques sélectionnés pour les analyses. Les paramètres acoustiques spécifiques aux phonèmes sont donc principalement utilisés dans les tumeurs de petits volumes, ayant un impact moindre sur la dynamique de la parole. En ce qui concerne les tumeurs de plus grand volume, les études recherchent des paramètres de qualité de la parole plus généraux pour catégoriser la parole comme intelligible ou inintelligible. Une raison possible motivant ce choix serait liée au fait que les tumeurs de plus gros volume dégradent trop fortement les caractéristiques acoustico-phonétiques de la parole, limitant l'analyse au niveau segmental.

Ainsi, un lien étroit semble exister entre les paramètres acoustiques / automatiques et la localisation de la tumeur, ainsi qu'entre ces paramètres et la taille de la tumeur. De plus, il existe une grande variabilité des paramètres acoustiques utilisés dans les différentes études, principalement au niveau segmental. L'utilisation de paramètres cepstraux et d'outils d'apprentissage automatique permet une analyse continue de la parole, mais ces techniques sont encore très récentes et la recherche doit être développée. Actuellement, les paramètres acoustiques et automatiques semblent être pertinents pour compléter l'évaluation perceptive de la parole, réalisée dans la pratique courante. Il semble donc opportun d'étudier des modèles d'analyse plus complets permettant non seulement d'étudier la parole des patients en fonction de leurs performances d'intelligibilité (c'est-à-dire de performances de production phonémique au niveau segmental), mais aussi de façon plus large en intégrant des paramètres de plus haut niveau pouvant être liés aux autres dimensions linguistiques (niveau lexical, morphologique...) altérant d'un point de vue fonctionnel les capacités de communication.

1.3.2 Échantillons de parole

L'analyse des échantillons de parole sur lesquels sont mesurés les paramètres acoustiques et automatiques montre une prédominance de l'étude de phonèmes isolés (voyelles ou consonnes). Les phrases ou les textes sont plutôt utilisés pour la mesure des coefficients cepstraux (tels que MFCC ou MSCC) ou de la nasalité.

Cependant, d'un point de vue fonctionnel, l'analyse de la parole spontanée serait le moyen le plus proche de mesurer les performances de parole dans la vie quotidienne du patient. Dans notre revue, nous constatons qu'il n'y a pas d'études sur de telles tâches, ni sur de la parole semi-spontanée comme la description d'images, ni sur de la véritable parole spontanée comme la parole conversationnelle.

1.3.3 Limites de l'étude

Deux bases de données ont été interrogées dans le cadre de cette revue systématique (PubMed et Embase). Le Web of Science (WoS) a également été utilisé, mais aucune nouvelle référence n'a été trouvée (c'est-à-dire que tous les articles trouvés dans le WoS étaient des doublons des entrées de PubMed et Embase). Cependant, il n'est pas exclu que d'autres études existent en dehors du champ de cette recherche.

Comme pour toute revue de littérature, un possible biais de publication ne peut pas être contrôlé. Il est en effet possible que d'autres études aient porté sur l'analyse acoustique / automatique de la parole en contexte cancérologique ORL mais qu'elles n'aient pas été publiées par manque de résultats probants, et de fait ne se retrouvent pas dans notre étude.

Parmi les 22 articles sélectionnés, deux études ont été réalisées sur des corpus identiques ou très similaires : (de Bruijn et al., 2011, 2009), et (Laaksonen et al., 2011, 2010). Cependant, les deux ont été retenus car les objectifs principaux étaient différents et complémentaires : (de Bruijn et al., 2009) s'est concentré sur l'analyse des formants alors que (de Bruijn et al., 2011) a utilisé des modèles plus complexes à base de réseaux de neurones ; (Laaksonen et al., 2011) a étudié l'analyse des moments spectraux sur les consonnes, alors que (Laaksonen et al., 2010) a étudié les formants des voyelles.

La grande variabilité des études incluses souligne la nécessité de développer des outils standardisés d'évaluation acoustique chez les patients traités pour un cancer oral ou oropharyngé. La standardisation peut permettre de réaliser des évaluations plus précises et plus fiables dans le diagnostic du trouble de la parole et de sa sévérité, mais aussi dans les comparaisons intra-individuelles dans le suivi des patients.

1.4 Conclusion

La revue systématique de littérature réalisée dans le cadre de ce travail de doctorat (Balaguer et al., 2020e) nous a permis d'aboutir à un état des lieux exhaustif des outils acoustiques ou automatiques permettant un traitement automatique du signal de parole pour décrire les effets des cancers de la cavité buccale ou de l'oropharynx.

Après recherche dans trois bases de données (PubMed et Embase, le WoS n'ayant pas permis de récupérer de nouvelle référence), 488 références relatives aux effets mesurés acoustiquement ou automatiquement des cancers oraux ou oropharyngés sur la parole ont été récupérées. Après vérification de la conformité des résumés à nos critères d'inclusion, 292 références ont été exclues. Nous avons évalué ainsi un total de 196 articles complets. Cent soixante-treize articles ont été exclus car ils ne respectaient pas nos critères d'inclusion, n'étaient pas récupérables, non originaux ou étaient des duplicatas d'articles déjà présents dans la sélection. Nous avons ainsi abouti à retenir 22 articles pour la suite de nos analyses, tous présentant une bonne ou forte qualité méthodologique (évaluée au moyen du QualSyst).

La plupart des études se sont concentrées sur la qualité du décodage acoustico-phonétique pour mesurer la parole : neuf articles prenant les mesures sur des voyelles et cinq sur les consonnes. L'analyse acoustique et automatique de la parole est ainsi principalement réalisée au niveau segmental, dans un contexte contrôlé de parole, comme peut l'être la lecture de texte. Dans notre revue de littérature, aucune étude n'a étudié la parole spontanée, basée sur des productions non contraintes. L'analyse automatique de cette parole spontanée est plus complexe à réaliser car elle ne permet aucune référence à laquelle comparer la parole du patient en l'absence de possibilité d'alignement automatique, et qu'elle comporte de nombreuses dimensions linguistiques associées (phonémique, lexicale, syntaxique, pragmatique ou discursive). Pourtant, ce contexte de production est le plus proche des situations de communication vécues par les patients au quotidien. C'est alors dans la parole spontanée que se manifesterait l'impact fonctionnel du trouble de la parole par la diminution de la capacité du sujet à transmettre un message. Les mesures automatiques sur la parole spontanée (comme la parole conversationnelle) doivent ainsi être développées.

Cette revue systématique de littérature nous a ainsi amenés à concevoir un nouveau corpus pour collecter des données de parole spontanée d'une part, mais aussi qui inclut des outils de mesure des capacités de communication. La conception et la constitution de ce corpus sont présentées dans la suite de ce manuscrit (voir chapitre 2).

2 Conception et constitution du corpus

Afin de répondre à l'objectif concernant la mesure de l'impact du trouble de parole sur la communication des patients traités pour un cancer de la cavité buccale et de l'oropharynx, j'ai constitué un nouveau corpus. La constitution de ce corpus a été d'une part déterminée par la nécessité d'étudier la parole spontanée des sujets en constituant des échantillons de parole en situation conversationnelle (ce qui n'est pas encore couramment fait comme l'a montrée la revue systématique de littérature). D'autre part, dans le cadre de l'évaluation et conformément à la Classification Internationale du Fonctionnement, du handicap et de la santé, l'ensemble des dimensions fonctionnelles (niveau des activités) et psychosociales (niveau des participations) doivent être prises en compte, notamment les aspects relatifs à la communication pouvant être altérée par les troubles de parole.

Ainsi, ce corpus comprend deux volets principaux. Un premier volet est constitué d'autoquestionnaires et d'un test de dépistage cognitif (sur les fonctions langagières, mnésiques, exécutives et d'orientation spatio-temporelle) fournissant des données sur le trouble de parole, les capacités de communication, la qualité de vie et les facteurs associés. Le second volet comprend des enregistrements de parole sur lesquels de multiples paramètres seront mesurés.

Le caractère novateur de ce corpus réside dans les enregistrements de véritable parole spontanée, qui serviront de support à nos analyses ultérieures.

2.1 Participants

Les patients venant en consultation ou hospitalisation (de jour ou en soins de suite et de réadaptation) en secteur ORL ont été inclus après avoir vérifié les critères d'inclusion, et les avoir informés sur cette recherche.

Un objectif de 21 sujets minimum était initialement fixé, dans la mesure où 50 % des études retenues dans la revue systématique incluaient 21 participants ou plus. Afin d'avoir un échantillon plus large et permettant une meilleure puissance statistique relative, nous envisagions d'inclure 30 sujets.

2.1.1 Critères d'inclusion et de non-inclusion

Ont été inclus :

- Les patients âgés d'au moins 18 ans affiliés ou bénéficiaires du régime de la sécurité sociale ;
- Les patients francophones natifs afin d'éviter les effets liés à un développement linguistique dans une langue différente du français ;
- Les patients traités pour un cancer de la cavité buccale ou de l'oropharynx, que le traitement soit chirurgical et/ou radiothérapique et/ou chimiothérapique ;
- Les patients venant en consultation de suivi, en hospitalisation de jour ou en hospitalisation en soins de suite et réadaptation ;
- Les patients en rémission clinique depuis au moins 6 mois, afin d'évaluer le caractère chronique et stable des troubles et de leur impact fonctionnel au quotidien ;
- Les patients présentant des tumeurs de toute taille (T1 à T4) ;
- Enfin, les patients présentant un trouble perceptible de la parole en situation conversationnelle ou non, afin que même les déficits de parole les plus fins soient analysés.

N'ont pas été inclus :

- Les patients dont la fatigabilité ne permet pas la passation des épreuves d'évaluation, autrement dit ne pouvant pas maintenir l'effort attentionnel la durée des épreuves, soit environ 90 minutes ;
- Les patients présentant une pathologie associée potentiellement responsable de troubles de parole ou de la fluence, telles que le bégaiement, un trouble de parole d'origine trouble neurologique... ;
- Les patients présentant des troubles cognitifs dont la sévérité est plus que modérée (troubles caractérisés par un score inférieur ou égal à 16/30 à un test de dépistage cognitif : la MoCA¹⁰), pouvant interférer avec la réalisation des épreuves de l'enregistrement de parole, la complétion des autoquestionnaires et l'entretien semi-dirigé ;
- Les patients auxquels il sera impossible de donner des informations éclairées et de s'assurer de leur compliance en raison d'une santé physique ou psychologique altérée ;
- Enfin, les patients participant à une autre recherche comprenant une période d'exclusion toujours en cours.

¹⁰ <https://www.mocatest.org/faq/>, consulté le 14/05/2021

2.1.2 Recrutement

Le recrutement des sujets s'est déroulé sur une période totale de 15 mois, en deux périodes d'inclusion : une première période de cinq mois et demi et une seconde période de six mois et demi.

Un intervalle de trois mois a séparé ces deux périodes, durant lequel les recrutements ont été suspendus en raison de la période de confinement sanitaire lors de la crise pandémique du Covid-19. Cette période de suspension correspondait à deux mois de confinement strict, et un mois post-confinement durant lequel aucun patient n'a été inclus en raison de la modification de la dynamique sociale induite par la situation sanitaire, et pouvant induire un biais de réponse (notamment quant à l'état anxiodépressif ou aux interactions avec les cercles sociaux).

Ce recrutement en deux temps a ainsi mené à la constitution de deux sous-groupes : un lot 1 de patients recrutés avant la crise sanitaire (soit avant mi-mars 2020), et un lot 2 de patients inclus après la période de trois mois de suspension (soit entre mi-juin et fin décembre 2020). La comparabilité des deux lots a également été étudiée, afin de nous assurer de l'absence de différence significative entre les deux groupes quant à nos variables d'intérêt.

2.2 Évaluation de l'impact sur la communication

De nombreux modèles théoriques appliqués à la communication ont été développés depuis une cinquantaine d'années. En raison des différentes conceptualisations appliquées à la communication, définie comme « *la transmission d'un contenu de pensée entre un locuteur et un allocutaire* » (François & Nespoulous, 2011), trois grands types de modélisation ont été proposés (Picard, 1992).

Les premiers modèles développés étaient des modèles dits « techniques ». Ils ciblent principalement la transmission du signal. Un des modèles les plus connus est celui de Shannon et Weaver (Shannon & Weaver, 1964), dans lequel la communication est envisagée sous la forme d'un transfert d'un message d'une source à un destinataire sous la forme d'un signal. Ces modèles assez réducteurs prennent peu en compte le contexte de communication, à l'exception du bruit ambiant pouvant perturber la transmission du message.

Les modèles linguistiques, dont le modèle de Jakobson (Jakobson, 1963) est l'un des plus connus, associe au message des fonctions linguistiques permettant

une communication efficace : les fonctions émotive, référentielle, poétique, conative, phatique et métalinguistique. Dans ce cadre, la communication s'inscrit pleinement dans un contexte d'élocution entre destinataire et destinataire, avec une intention de transmission du message par le destinataire à un destinataire en prenant en compte le contenu du message et les connaissances préalables du contexte.

Enfin, les modèles psychosociologiques élargissent la notion de message transmis aux éléments non-verbaux de communication peu pris en compte dans les modèles précédents. Ils ciblent davantage les contextes de communication, en intégrant à la communication le cadre (contexte spatial et temporel de l'échange), l'institution (contexte culturel de l'échange) et le nombre de participants.

D'autres modèles, combinant des concepts relatifs aux trois grandes familles présentées ci-dessus, ont été développés.

Les modèles psycholinguistiques, comme le modèle de Caron (voir Figure 10) (Devevey & Kunz, 2018; François & Nespoulous, 2011), envisagent le message sous la forme d'un signal acoustique encodé par un émetteur et décodé par un récepteur, selon trois niveaux symétriques de production et de réception. Ces modèles, dits « naïfs » en raison de leur compréhension relativement intuitive, se composent d'un niveau conceptuel lié au contexte de communication et à la motivation communicationnelle, d'un niveau catégoriel lié à la représentation morpho-syntaxique des signes linguistiques, et d'un dernier niveau articulatoire comprenant à la fois la mise en œuvre des structures phonologiques à produire dans la parole mais aussi de la production articulatoire du son cible du côté du locuteur. Ces modèles sont séquentiels, allant dans un sens top-down du point de vue du locuteur (production).

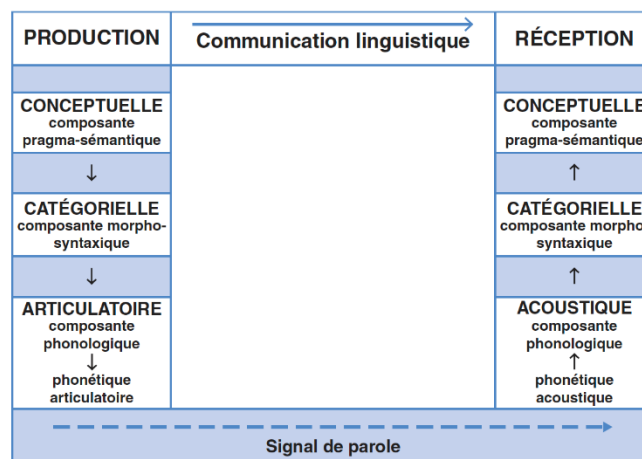


Figure 10 - Modèle psycholinguistique de Caron (1989), cité par François & Nespoulous (2011) et Devevey & Kunz (2018)

Ainsi, la parole s'inscrit complètement dans un contexte de communication, dont elle est un des vecteurs de transmission du message. Les troubles de parole viennent perturber cet échange en dégradant les aspects acoustico-phonétiques reçus par l'interlocuteur. Mais les processus de réception étant « bottom-up », la dynamique communicationnelle sera dégradée de façon plus large, les niveaux supérieurs catégoriel et conceptuel étant alors également perturbés par effet séquentiel.

Pour autant, peu d'outils existent à l'heure actuelle pour évaluer l'impact sur la communication des troubles de parole consécutifs à un cancer ORL.

Le CPIB (Communicative Participation Item Bank) a été développé en 2008 (Baylor et al., 2009) sous la forme d'un questionnaire à destination des patients. Il vise à évaluer les capacités de communication dans des situations de parole à partir d'items répondus par le patient sous la forme d'une échelle de Likert à cinq niveaux (pas du tout, un peu, assez, beaucoup, énormément) (Baylor et al., 2013). Si le CPIB a été initialement créé pour être utilisé auprès de patients adultes présentant des troubles d'origine neurologique, des études plus récentes ont permis d'étudier les performances de cet outil auprès de patients atteints d'un cancer des VADS, principalement auprès de patients laryngectomisés totaux (Bolt et al., 2016; Eadie et al., 2018, 2014).

En français, deux échelles et questionnaires sont validés, mais seulement auprès d'une population adulte atteinte de pathologies neurologiques. Pour autant, nous avons choisi de les retenir dans le cadre de la constitution de notre corpus car ils permettent une mesure écologique (c'est-à-dire proche des situations concrètes de la vie réelle quotidienne) de la communication, tout en tenant compte de la dimension psychoaffective.

L'Échelle de Communication Verbale de Bordeaux ou ECVB (Mazaux et al., 2006) est un questionnaire constitué de 34 items répartis en sept dimensions correspondant à des situations de communication quotidienne : expression des intentions (trois items), conversation (sept items), téléphone (sept items), achats (quatre items), relations sociales (cinq items), lecture (quatre items), écriture (quatre items). Initialement validé auprès de patients ayant subi un AVC (Accident Vasculaire Cérébral), les dimensions évaluées concernent tous les aspects de communication pouvant être altérés, que ce soit sur la communication orale ou écrite. Chaque item est coté sur le principe d'une échelle de Likert à quatre niveaux (jamais, rarement, souvent, très souvent), permettant le calcul d'un score par item, d'un sous-score par dimension et d'un score global. Plus les scores sont faibles, plus la gêne est importante. Si dans sa version validée en neurologie adulte l'ECVB est complété par l'orthophoniste

(ou tout autre examinateur) au cours d'un entretien semi-dirigé avec le patient, nous avons fait le choix de proposer également une complétion par le patient lui-même, comme cela se fait habituellement dans les questionnaires PRO (Patients Reported Outcomes) (Doward & McKenna, 2004). Dans un contexte différent de la neurologie adulte, où une grande partie des sujets évalués présente des troubles cognitifs et linguistiques pouvant interférer avec l'autocomplétion de questionnaires, cette double cotation de l'ECVB permettra de comparer les perceptions du patient quant à sa communication quotidienne avec la perception externe d'un examinateur selon ce que lui rapportent les patients durant l'entretien.

Le Dysarthria Impact Profile ou DIP, initialement validé en anglais (Walshe et al., 2009), a été traduit et validé en français (Letanneux et al., 2013). À destination des sujets atteints par une maladie de Parkinson, il se compose de 48 items en quatre dimensions : « l'effet de la dysarthrie sur moi en tant qu'individu » (12 items), « l'acceptation de ma dysarthrie » (10 items), « comment je perçois la réaction des autres face à ma parole » (14 items), « comment la dysarthrie affecte ma communication avec les autres » (12 items). Ces 48 items sont également présentés sous forme d'une échelle de Likert à cinq niveaux (fortement en désaccord, en accord, ne sais pas, en désaccord, fortement en désaccord), ce qui permet le calcul d'un score par item et d'un sous-score par dimension. Enfin, une dernière dimension concerne la dysarthrie parmi l'ensemble des troubles du patient, où il est demandé au répondant de classer la dysarthrie par importance de préoccupation aux côtés d'autres préoccupations ressenties par le patient.

À l'heure actuelle, il n'existe aucun outil validé en français auprès de population atteinte d'un cancer de la cavité buccale ou de l'oropharynx, permettant de mesurer l'impact du trouble de parole sur la communication.

Nous n'avons pas retenu le CPIB dans notre étude : d'une part, il est moins exhaustif que l'ECVB et le DIP en raison de son caractère très court (dix questions) ; d'autre part, les concepts sous-tendus par les dix items sont repris dans les deux autres questionnaires. Ainsi, s'il semble pertinent de prévoir une validation de ces deux outils dans cette population, l'ECVB et le DIP seront ici utilisés afin de construire un score holistique de communication en cancérologie ORL.

2.3 Évaluation des facteurs associés au trouble de parole et à la communication

Si le lien entre parole et communication est étroit, les troubles de parole s'inscrivent dans un contexte complexe. Comme présenté dans la modélisation de Wilson (Wilson & Cleary, 1995), de multiples facteurs peuvent être associés à la fois au trouble de parole (statut symptomatique) et aux capacités de communication (statut fonctionnel).

Les caractéristiques de l'individu renvoient à l'ensemble des déficits associés et des autres altérations du carrefour aérodigestif, mais aussi au statut cognitif du sujet et à son état anxiodépressif, ce dernier étant aussi étroitement lié à la motivation personnelle. Les caractéristiques de l'environnement, notamment à propos du soutien social et psychologique, sont à mettre en relation avec la constitution des cercles sociaux qui impactent les besoins de communication des sujets. Les perceptions de santé relatives au trouble de la parole seront en lien avec l'auto-perception de ce trouble par le sujet. Enfin, les données individuelles comme la catégorie socio-professionnelle sont liées au soutien économique et social, les données cliniques et de traitement correspondent à la symptomatologie (caractéristiques individuelles).

Une évaluation complète du trouble de parole, prenant à la fois en compte les trois niveaux décrits par la CIF, doit ainsi être multidimensionnelle. De multiples questionnaires ayant pour objectif d'évaluer les facteurs associés au trouble de parole, à la communication ou à la qualité de vie ont été développés, et présentent des formes et des performances très hétérogènes (Ojo et al., 2012). Cette section présente les outils choisis ainsi que les raisons théoriques ayant conduit à ces choix.

2.3.1 Déficits consécutifs au traitement de la pathologie cancérologique ORL

Différentes études ont montré l'influence des déficits en cancérologie ORL. Par exemple, la douleur, la fatigue, la réduction de mobilité cervico-scapulaire ou encore les troubles de la déglutition sont des déficits fréquemment sources de plaintes par les patients (Balaguer et al., 2017, 2021a). L'étude de Chen (Chen et al., 2015) montre un lien entre difficultés de communication après cancer ORL et image corporelle. Cette même étude trouve un lien significatif entre difficultés de communication et intensité des symptômes associés.

En cancérologie ORL, des questionnaires de type Patients Concerns Inventory (PCI) ont été développés (Rogers et al., 2009) mais ne sont pas

validés à l'heure actuelle en français. Ainsi, pour quantifier les déficits associés fonctionnels (douleur, déglutition, nutrition, respiration, phonation, audition, vision, olfaction-gustation, changements de l'aspect physique, mouvements du cou et des épaules, psychosocial), le Carcinologic Handicap Index ou CHI (Balaguer et al., 2017, 2021a) a été retenu. Ce questionnaire de type Patients Concerns Inventory est rapide, transversal et global. Il a, de plus, été validé récemment en français.

2.3.2 Anxiété / Dépression

La dépression est une pathologie retrouvée chez près d'un cinquième des patients (Rieke et al., 2017). L'état anxiodépressif modifie d'ailleurs à la fois les capacités de communication des sujets atteints (Eadie et al., 2018), mais a aussi un impact sur la qualité de la production de parole (Böhm et al., 2016).

Deux questionnaires évaluant l'anxiété et la dépression ont été retenus, conformément aux recommandations de l'AFSOS dans le référentiel « *Psycho-oncologie : anxiété et troubles anxieux en cancérologie* »¹¹. Il s'agit de l'échelle HAD « Hospital Anxiety and Depression Scale » (Zigmond & Snalth, 1983) permettant d'obtenir un sous-score d'anxiété et un sous-score de dépression, et de l'Échelle de Détresse Psychologique ou EDP (Nordmann-Dolbeault, 2009) pour laquelle le sujet doit positionner sur une ligne allant de « pas de détresse » à « détresse très importante » son état psychologique estimé de la semaine précédente.

2.3.3 Statut cognitif

Le statut cognitif peut être dégradé après traitement d'un cancer ORL (Gan et al., 2011). Le statut cognitif auto-rapporté (au moyen d'un questionnaire) est d'ailleurs la seconde variable associée aux compétences de communication, après la sévérité du trouble de parole, parmi l'âge, le niveau d'éducation, la détresse émotionnelle, l'utilisation de la parole, les troubles de déglutition et la perte auditive (Bolt et al., 2016). Toutefois, cette étude a porté sur une population majoritairement traitée par laryngectomie totale (62 %) et à distance éloignée du diagnostic (huit ans en moyenne). L'étude d'Eadie (Eadie et al., 2018) retrouve également une corrélation à 0,25 entre

¹¹ https://www.afsos.org/wp-content/uploads/2016/09/anxiete_AFSOS_141021.pdf, consulté le 16/05/2021

l'état cognitif auto-rapporté et les capacités de communication mesurées au moyen du CPIB. Là encore, la majorité des participants ont subi une laryngectomie totale (53,4 %).

Plus généralement, près de la moitié des patients traités pour un cancer ORL présentent des scores à la batterie de dépistage cognitif MoCA inférieurs au score pathologique fixé à 26/30 six mois après le traitement (Regier et al., 2019). L'état cognitif s'améliore ensuite légèrement mais la proportion de patients ayant un score inférieur au seuil reste élevé à 18 mois (40 %). Un lien significatif est également retrouvé entre statut cognitif et durée de traitement d'autre part (Williams et al., 2017b).

L'AFSOS, dans son référentiel « *Troubles cognitifs et cancers* »¹² présente les préconisations du groupe de réflexion sur les évaluations cognitives en oncologie (Grec-Onco). Deux batteries globales pouvant être utilisées régulièrement au cours du suivi afin d'établir le statut cognitif des patients sont référencées : la Mattis (Mattis, 1976) et la MoCA (Rambeau et al., 2019). La MoCA, validée dans de nombreuses langues dont le français, a été retenue en raison de son caractère rapide (seulement 30 items) et exhaustif. Elle permet de mesurer six dimensions : les capacités visuospatiales et exécutives, le langage (dénomination, répétition, évocation lexicale), la mémoire à court terme et le rappel, l'attention, les capacités d'abstraction et l'orientation spatio-temporelle (cette fonction n'est d'ailleurs pas évaluée dans la Mattis). Dans notre étude, la MoCA a eu une utilité double : vérifier l'éligibilité des sujets (en excluant les sujets trop dégradés d'un point de vue cognitif), et fournir des informations sur le statut cognitif général des sujets inclus.

2.3.4 Besoins de communication liés aux cercles sociaux

La communication induisant un échange entre locuteur et interlocuteur(s), la constitution des cercles sociaux a une influence sur les besoins et capacités de communication des locuteurs.

Le soutien social perçu est lié aux capacités de communication des patients avec leurs pairs (Eadie et al., 2018). D'un autre côté, l'isolement social est lié au retrait dans les situations de communication, en raison d'une intelligibilité dégradée notamment pour les patients laryngectomisés (Danker et al., 2010)

Mais la constitution des cercles sociaux joue aussi un rôle important sur les autres facteurs associés. Un lien existe entre constitution des cercles sociaux et

¹² https://www.afsos.org/wp-content/uploads/2020/12/Troubles-cognitifs-et-cancer_AFSOS.pdf, consulté le 16/05/2021

facteurs de risques en cancérologie ORL. Une faible diversité des cercles sociaux est significativement associée à une dépendance à la consommation alcoolique (Mowbray et al., 2014), tabagique, et à un risque majoré d'infection des voies aéro-digestives supérieures (Cohen et al., 1997). La taille du réseau de soutien formel et la satisfaction quant à ce réseau modifie la perception de santé en cancérologie. La santé générale est perçue différemment selon la taille des réseaux de support formels et de la satisfaction des sujets quant à ces réseaux (Payne et al., 2019). Au niveau psychocognitif, le risque relatif de développer une démence est significativement plus important chez des individus isolés et chez ceux vivant seuls et n'étant pas mariés (Fratiglioni et al., 2000). De plus, le statut psychologique est lui aussi lié à la constitution des cercles sociaux (Kawachi, 2001). La faiblesse des relations sociales augmente le niveau de détresse psychologique, et abaisse la motivation pour prendre soin de soi (Kroenke, 2018). L'isolement social augmente significativement le risque de dépression, alors que le soutien social le diminue (Pearlin & Johnson, 1977).

Différents outils sont disponibles pour mesurer les cercles et réseaux sociaux des patients. Certains évaluent le support disponible en situation d'anxiété chez des étudiants universitaires (St-Jean-Trudel et al., 2005), la satisfaction et la disponibilité de ce support social chez des adultes sains (Bruchon-Schweitzer et al., 2003), ou encore l'ajustement social (Weissman et al., 1981). D'autres ciblent la fréquence à laquelle différents types de relations ont été utilisées pour remplir les fonctions de trois systèmes comportementaux (attachement, soins et affiliation) chez 200 adolescents (Furman & Buhrmester, 2009). Un dernier questionnaire mesure la fréquence et la satisfaction des comportements sociaux d'une population psychiatrique adulte dans différents types d'activités quotidiennes (Zanello et al., 2006).

Mais ces questionnaires ont deux limites majeures. D'un côté, ils ne sont pas validés auprès de populations pathologiques, à l'exception du Questionnaire de Fonctionnement Social ou QFS (Zanello et al., 2006). D'un autre côté, aucun n'évalue spécifiquement la constitution des cercles sociaux, en termes de taille de chaque cercle (nombre de participants à chaque cercle), de fréquence des relations avec chaque cercle, et de satisfaction quant à la fréquence et à la qualité des échanges avec chaque cercle. Or, la constitution des cercles sociaux n'influence pas seulement les besoins de communication des sujets, mais contribue également à leur dynamique psychosociale en termes de relations interpersonnelles.

Ainsi, dans le cadre de notre étude, nous avons fait le choix de retenir le seul questionnaire validé auprès d'une population pathologique même s'il ne s'agissait pas de patients adultes en cancérologie ORL : le QFS. Ce questionnaire permet de mesurer la fréquence et la satisfaction des

comportements sociaux d'une population adulte psychiatrique, au moyen de l'étude de huit dimensions (activités, tâches de la vie quotidienne, loisirs, relations familiales et de couple, relations extrafamiliales, gestion financière et administrative, santé générale et vie collective et informations), sous deux aspects à chaque fois : fréquence de l'interaction, et satisfaction quant à l'interaction.

Pour compléter cette analyse, nous avons créé la grille ECCS (Évaluation de la Constitution des Cercles Sociaux) (**Annexe III**). L'objectif de ce questionnaire est de pouvoir être utilisé en pratique clinique courante dans une population traitée pour un cancer des VADS.

Un comité d'experts composé de deux orthophonistes, d'un phoniatre, de trois chercheurs en informatique et d'un chercheur en sciences sociales/psychologie a participé à la construction de l'ECCS. Un cercle social est défini comme un ensemble de personnes, de connexions, de « points communs » et de normes qui sont mutuellement reconnus comme partagés (Bidart et al., 2011). Six cercles sociaux ont été retenus. Trois font partie du réseau proche du patient : le cercle privé, défini comme les personnes avec lesquelles on vit au quotidien, la famille (famille que l'on ne voit pas forcément au quotidien mais avec laquelle on a des contacts étroits et réguliers), et les amis (personnes pour lesquelles on éprouve de la sympathie, que l'on connaît bien et avec lesquelles on a des contacts étroits et réguliers). Les trois autres concernent des individus avec lesquelles les relations sont moins étroites : les connaissances (personnes que l'on a rencontrées, avec lesquelles on a des contacts réguliers mais pas de relation particulière, ou que l'on connaît peu), les soignants et la communauté (personnes que l'on connaît, auxquelles on dit « bonjour » et « au revoir », avec lesquelles on partage des intérêts communs mais pas de relation particulière), et les étrangers. Quatre critères d'évaluation, partiellement basés sur le modèle du questionnaire QFS, ont également été définis par le comité d'experts. Le premier critère concerne le nombre de personnes considérées comme faisant partie de ce cercle. Le deuxième critère concerne la régularité des contacts. Le troisième concerne la satisfaction globale de la fréquence des contacts, et le dernier critère concerne la satisfaction globale des échanges avec les personnes des différents cercles.

Deux de ces critères ont été mesurés sur une échelle de fréquence. Le nombre de personnes dans chaque cercle a été évalué pour les cercles privé, famille et amis. Aucune quantification n'a été demandée pour les autres cercles sociaux, en raison de la difficulté de les quantifier de manière fiable. Pour le cercle privé, quatre réponses ont été proposées : 0, 1, 2 à 4, ou 5 et plus. Pour la famille et les amis, cinq réponses ont été proposées : 0, 1, 2 à 4, 5 à 9, ou 10 et plus. Le score attribué à ce critère correspond à la première valeur de la

réponse proposée (c'est-à-dire 0, 1, 2, 5 ou 10). La régularité des contacts a été mesurée pour les six cercles sur une échelle à cinq points : tous les jours (5), plusieurs fois par semaine (4), plusieurs fois par mois (3), plusieurs fois par an (2), une fois par an ou moins (1). Les deux autres critères de « satisfaction » (fréquence et échanges) ont été mesurés pour les six cercles sociaux sur une échelle de Likert en 11 points, allant de -5 (fréquence estimée des échanges trop faible/très insatisfaisante), à +5 (fréquence estimée des échanges trop élevée/très satisfaisante), 0 étant considéré comme une valeur neutre (suffisante/neutre). Le score attribué pour l'évaluation de ces critères de satisfaction correspond à la valeur de l'item coché par le répondant.

Une fois élaboré, nous avons conduit une étude préliminaire sur 22 patients dont 13 avaient été traités pour un cancer de la cavité buccale ou de l'oropharynx (les neuf autres étaient traités pour un cancer du larynx ou de l'hypopharynx). Selon les recommandations du COSMIN (Mokkink et al., 2010), la validité de construit, la validité clinique (comparaison avec un groupe de 21 sujets sains) et la validité de critère (correspondant aux comparaisons des performances de l'ECCS avec celles d'outils déjà validés : QFS, HAD, PHI et EORTC QLQ-H&N35) ont été étudiées.

Finalement, cette étude préliminaire montre que l'ECCS est valide pour être utilisé en pratique clinique courante. Il permet d'évaluer la dynamique sociale autour des patients concernant la constitution de leurs cercles sociaux. Cette évaluation permettrait non seulement d'adapter la stratégie de soins aux ressources humaines autour du patient, mais aussi de prévenir l'isolement social et les conséquences fonctionnelles qui peuvent en découler. Enfin, une étude plus large devrait être menée dans un premier temps sur un échantillon plus grand pour vérifier la généralisabilité des résultats, et dans un second temps sur l'influence de la constitution des cercles sociaux sur d'autres pathologies chroniques, telles que les AVC et les pathologies neurodégénératives. De plus, une étude secondaire ciblant la fiabilité temporelle reste à conduire.

Dans ces conditions, nous avons choisi d'associer l'ECCS au QFS pour évaluer la constitution et les échanges avec les membres des différents cercles sociaux autour des sujets.

2.3.5 Auto-perception du trouble de parole

La perception qu'ont les patients de leur trouble de parole a une influence majeure sur leurs capacités de communication. C'est en effet le critère principal en termes de significativité sur la participation aux activités de communication des patients après cancer des VADS (Bolt et al., 2016). L'auto-évaluation de la

sévérité du trouble de parole est aussi le meilleur prédicteur de l'impact communicationnel dans cette population à plus de deux ans du diagnostic (Eadie et al., 2018).

Au-delà des conséquences sur les capacités de communication, la perception de trouble de parole a également une influence sur l'état anxiodépressif des patients ainsi que sur leurs performances cognitives (Eadie et al., 2018).

Pour mesurer l'auto-perception du trouble de parole, nous avons choisi d'utiliser le Phonation Handicap Index (ou PHI) (Balaguer et al., 2020b). Il a été préféré au SHI (Speech Handicap Index) car il est plus court et présente des performances équivalentes dans notre population d'intérêt (Balaguer et al., 2020b). Le PHI est composé de 15 items regroupés en trois domaines : Signes physiques, spécifiques de la parole concernant les modifications de la voix, de la prosodie et de l'articulation ; Retentissement fonctionnel envisageant les répercussions lors des formulations les plus élémentaires jusqu'à l'expression des émotions et de la communication en groupe ; Retentissement psycho-social reflétant le handicap social ressenti par le patient. Un sous-score est calculé par domaine. Le PHI est également constitué de trois questions complémentaires, que le patient doit évaluer de 0 à 10. Elles concernent la sévérité perçue du trouble de parole, la gêne à produire une parole compréhensible et enfin le handicap relatif à la parole ressenti dans la vie quotidienne.

2.3.6 Qualité de vie en cancérologie ORL

Si nous nous intéressons aux restrictions de participation, la corrélation entre trouble de parole et qualité de vie en cancérologie ORL n'est que faible à modérée.

Thomas (Thomas et al., 2009) retrouve une corrélation modérée entre intelligibilité et qualité de vie ($r = 0,52$), tandis que cette corrélation est plus faible dans d'autres études, entre 0,24 et 0,28 (Borggreven et al., 2007). La qualité de vie globale est définie comme le bien-être général global et le développement personnel évalués objectivement et subjectivement dans les dimensions physiques, matérielles, sociales et émotionnelles, pondérés par un ensemble de valeurs personnelles (Felce & Perry, 1995; Karimi & Brazier, 2016).

La qualité de vie relative à la santé correspond à la façon dont une personne fonctionne dans sa vie et son bien-être perçu dans les domaines physique, mental et social de la santé spécifiquement (Hays & Reeve, 2008; Torrance, 1987), ou encore les aspects du bien-être perçu par l'individu liés ou affectés par la présence d'une maladie ou d'un traitement (Ebrahim, 1995). Les

corrélations entre qualité de vie relative à la parole et intelligibilité sont statistiquement significatives, mais le lien est plus fort dans le cas de l'intelligibilité évaluée au niveau de la phrase (Intelligibilité sur phrase et UWQOL Speech : $r = 0,662$; $p = 0,0001$) qu'au niveau des mots (Intelligibilité de mots et UWQOL Speech : $r = 0,416$; $p = 0,0001$) (Meyer et al., 2004). Concernant la qualité de vie globale, la corrélation reste faible à modérée avec les autoquestionnaires mesurant les atteintes fonctionnelles liées à la parole (Balaguer et al., 2020b; Dwivedi et al., 2010; Karnell et al., 2000; Rinkel et al., 2008).

Ainsi, l'absence de lien fort entre trouble de la parole et qualité de vie conduit à la nécessité de prendre en compte les facteurs associés à ces deux concepts afin d'en expliquer plus précisément la relation.

Parmi les déficits associés, certaines caractéristiques individuelles comme le niveau de revenu et le statut marital sont également des facteurs en jeu dans la qualité de vie des patients en cancérologie ORL (Liao et al., 2019), de même que le niveau d'éducation (Tahani et al., 2017).

D'un point de vue plus clinique, la qualité de vie est également impactée par le traitement et les difficultés diagnostiques éventuelles (Min Ang et al., 2019), ou le délai depuis le traitement avec une amélioration progressive de la qualité de vie à partir de six mois après traitement (Mair et al., 2017). La présence de troubles de déglutition associés entraînant une alimentation entérale par sonde de gastrostomie dégrade significativement la qualité de vie (Terrell et al., 2005).

Même si le lien est modéré entre qualité de vie et anxiété (Hassanein et al., 2001), il est plus fort entre qualité de vie et dépression (Dzebo et al., 2017).

Le statut cognitif est également significativement lié à la qualité de vie (Williams et al., 2017a), et le bien-être social est lié au statut cognitif général et aux performances langagières, tandis que le bien-être fonctionnel est liés aux capacités mnésiques (Williams et al., 2017b).

Enfin, l'isolement social induit une moins bonne qualité de vie après cancer (Kroenke et al., 2013). Spécifiquement par rapport à la parole, le support social modifie la qualité de vie relative à la parole (Karnell et al., 2007).

Concernant l'évaluation de la qualité de vie, différents outils existent en cancérologie ORL (Heutte et al., 2014). Nous avons choisi de retenir les grilles de référence européennes de l'European Organisation for Research and Treatment of Cancer (EORTC).

Le questionnaire EORTC QLQ-C30 est un questionnaire généraliste (Aaronson et al., 1993). Il est constitué de 30 items sous la forme d'échelle de

Likert permettant le calcul d'un score de statut de santé globale, de cinq scores fonctionnels (physique, rôle, émotionnel, cognitif et social) et de neuf scores d'échelles de symptômes (fatigue, nausée et vomissements, douleur, dyspnée, insomnie, perte d'appétit, constipation, diarrhée, difficultés financières).

Le questionnaire EORTC QLQ-H&N35 (Bjordal et al., 1999) est un questionnaire complémentaire à l'EORTC QLQ-C30. Il est composé de 35 items, permettant d'évaluer des dimensions variées souvent dégradées par la pathologie cancérologique ORL : douleur, déglutition, sens (goût et odorat), parole, alimentation dans ses aspects sociaux, contact social, sexualité, état dentaire, ouverture buccale, sécheresse buccale, salive collante, toux, sensation d'être malade, utilisation d'antalgiques, de compléments nutritionnels, alimentation entérale, perte ou gain de poids.

2.4 Enregistrements de parole et choix des tâches

Le second volet du corpus correspond aux enregistrements de parole sur lesquels divers paramètres ont été mesurés, et qui seront présentés dans la suite de ce manuscrit.

Les tâches de production de parole ont été sélectionnées afin d'obtenir une représentation la plus exhaustive possible des différents aspects de la parole des sujets.

Six tâches de production de parole ont été proposées à chaque sujet.

Parmi elles, cinq correspondent à des tâches de parole contrainte :

- La réalisation d'un [a] confortable tenu durant quelques secondes ;
- La lecture d'une série de 17 phrases porteuses du type « Le sac øCø convient » où C représente chacune des 17 consonnes du français en contexte vocalique stable. Cette tâche permettra de réaliser des analyses acoustiques sur les consonnes dans le cadre d'une autre étude qui ne sera pas détaillée ici ;
- La lecture du texte de référence de « La chèvre de Monsieur Seguin » ;
- Deux phrases de Combescure « *La voiture s'est arrêtée au feu rouge* » et « *Annie s'ennuie loin de ses parents* », qui contiennent les deux semi consonnes /w/ et /ɥ/ absentes du texte de référence ;
- La répétition d'une liste de 52 pseudo-mots dans le cadre d'une épreuve de décodage acoustico-phonétique, mise au point par le Laboratoire Parole et Langage d'Aix-en-Provence (dans le cadre du projet INCa C2SI Grant n°2014-135) (Ghio et al., 2018) et récemment validée (Rebourg et al., 2020).

Ces cinq tâches ne feront pas l'objet d'analyses dans notre étude. Pour autant, j'ai malgré tout souhaité les inclure afin que des analyses sur de la parole contrainte puissent être menées dans des études ultérieures nécessitant également de la parole non contrainte.

La sixième tâche correspond à l'enregistrement de l'entretien semi-dirigé mené par l'examineur avec le patient au cours de la passation de l'ECVB. Ce contexte correspond à une véritable situation de parole spontanée, non contrainte (Prins & Bastiaanse, 2004). D'une part, la revue systématique de littérature a montré l'absence d'étude sur de la parole non contrainte. D'autre part, nous faisons l'hypothèse que les indices sur les capacités de communication seraient retrouvés dans la parole spontanée qui est la plus représentative de l'expression orale dans la vie quotidienne (Knuijt et al., 2017).

Les enregistrements de parole ont été réalisés dans une salle d'entretien non anéchoïque, de façon à rester au plus près possible des évaluations cliniques habituelles. Aucun bruit extérieur ni intérieur (climatisation, ventilation...) ne devait être perceptible afin de ne pas perturber la qualité de l'enregistrement.

Les échantillons de parole ont été enregistrés sur un enregistreur numérique ZOOM H4N Pro (fréquence d'échantillonnage 48 kHz, résolution de 16 bits, mode mono). Les fichiers sonores ont été enregistrés au format PCM linéaire (exporté en WAVE). Le micro serre-tête (Thomann T.Bone HC 444 TWS) était placé à 6 cm de la bouche du sujet, positionné de face sous le niveau de la lèvre inférieure et au niveau de la commissure labiale droite, afin d'avoir une fiabilité des mesures d'intensité par un maintien de la distance microphone bouche. De plus, dans un objectif ultérieur d'applicabilité clinique des résultats de notre étude, nous avons souhaité utiliser du matériel facile d'accès et d'utilisation, et abordable financièrement pour des cliniciens.

En raison de la situation sanitaire, la mousse du micro serre-tête a été changée entre chaque sujet (le patient ne devant pas porter de masque afin de préserver la qualité de l'audio), et le micro a été désinfecté.

2.5 Déroulement des passations

Comme indiqué précédemment, les passations se sont déroulées sur deux périodes (pré- et post-confinement de mars 2020).

J'ai procédé à l'inclusion et à l'ensemble des passations (questionnaires et enregistrements de parole) pour les sujets du lot 1. Concernant le lot 2, en raison des restrictions d'accès hospitalier liées au contexte sanitaire, mes

collègues orthophonistes ont pris le relais de l'inclusion et des enregistrements. Ceci a ainsi nécessité un temps préalable de formation de mes collègues, afin de s'assurer de la standardisation de la procédure.

2.5.1 Participation des sujets

Le déroulement global de la participation des sujets est représenté en Figure 11.

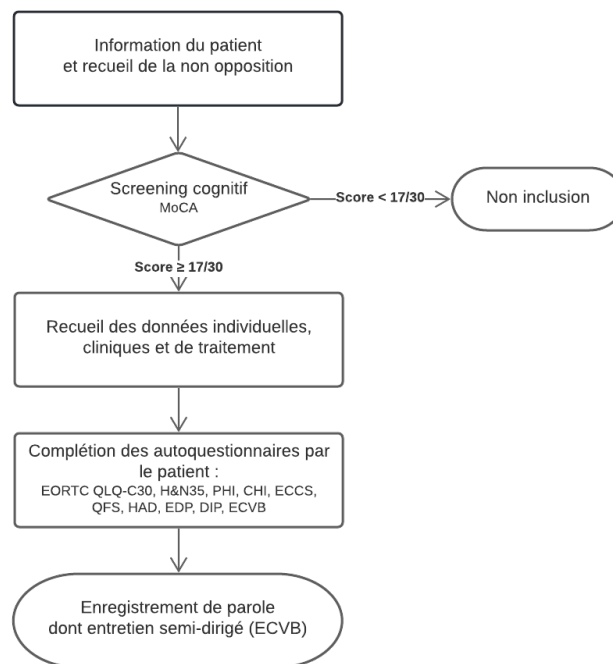


Figure 11 - Déroulement de la participation des sujets

À l'arrivée des sujets et après les avoir informés sur l'étude, un test de dépistage cognitif leur a été proposé au moyen de la MoCA : il a permis de confirmer l'inclusion du sujet dans l'étude (un score inférieur ou égal à 16 entraînant une non-inclusion, en raison d'une altération de l'état cognitif pouvant perturber la complétion des questionnaires). En cas d'inclusion, un code de pseudonymisation a été attribué à chaque participant.

Les sujets inclus ont ensuite complété les autoquestionnaires permettant l'évaluation de la qualité de vie (EORTC QLQ-C30 et H&N35), de l'impact des déficits associés (CHI), d'anxiété et de dépression (HAD, EDP), de la satisfaction quant aux interactions avec les cercles sociaux du sujet (QFS, ECVB), du handicap de parole (PHI), ainsi que les questionnaires relatifs à la dynamique communicationnelle (DIP, ECVB).

Puis, ils ont bénéficié d'enregistrements de leur parole, dont un sur un entretien semi-dirigé. Cet entretien a permis à l'examineur de compléter sa propre version de l'ECVB (appelée version « examinateur ») sans avoir pris connaissance des éléments rapportés préalablement par le sujet lors de l'autocomplétion de l'ECVB (appelée version « patient »). La comparaison ultérieure des résultats entre les deux versions « examinateur » et « patient » a permis d'analyser les différences éventuelles d'impact sur la communication selon les points de vue interne (patient) et externe (examineur).

La participation totale des sujets se déroulait au cours d'une même journée.

Une durée de 90 minutes a été préalablement estimée par sujet, pour son information et le recueil de sa non-opposition, la complétion des questionnaires, et les enregistrements de parole.

2.5.2 Caractéristiques des sujets inclus

Vingt-sept sujets ont été recrutés dans cette étude.

Deux sujets recrutés n'ont pas été inclus. Un sujet présentait un score inférieur au seuil fixé à 16/30 au dépistage cognitif. Un autre a été exclu des analyses en raison d'un problème technique ayant conduit à une troncature de l'enregistrement audio de la tâche de parole spontanée, biaisant les scores issus du traitement de cette tâche (notamment des mesures temporelles ou des inventaires phonémiques ou lexicaux).

Finalement, 25 patients ont été inclus dans cette étude : 12 font partie du lot 1 « avant confinement », et 13 font partie du lot 2 « après confinement ». Des tests statistiques non paramétriques (en raison de l'effectif réduit de notre corpus) ont été menés pour évaluer la comparabilité des deux lots :

- Test de Mann-Whitney pour les variables quantitatives : âge, délai depuis la fin de traitement ;
- Et test du Chi-2 ou test de Fisher (si les effectifs théoriques sont inférieurs à 5) pour les variables qualitatives : genre, départements de naissance et de résidence, localisation de la tumeur, réalisation d'une chirurgie, d'une radiothérapie, d'une chimiothérapie, caractéristiques de la tumeur selon la classification TNM, la catégorie socio-professionnelle, le statut marital et le statut auditif.

Les caractéristiques de ces 25 sujets, ainsi que les résultats des tests statistiques de comparaison des sujets des deux lots se retrouvent en **Annexe IV**. Aucune différence significative au seuil de 5 % n'a été retrouvé entre les

deux lots : les deux lots de cette étude sont comparables entre termes de caractéristiques individuelles, cliniques et de traitement.

2.5.2.1 Description générale des sujets

L'échantillon est composé d'une majorité d'hommes (60 %), d'âge médian 67 ans (écart interquartile = 12) dont la plupart réside dans un des départements de la région Occitanie (88 %). Plus de la moitié des sujets est marié (60 %) ou pacsé (8 %). Près de la moitié (48 %) est retraité. Une minorité des sujets rapporte une plainte auditive sans appareillage (24 %).

Au niveau clinique, 56 % des sujets ont été traités pour une tumeur de la cavité buccale (oropharynx : 40 %, cavité buccale + oropharynx : 4 %). La classification TNM a pu être retrouvée dans les dossiers pour 19 sujets, parmi lesquels une majorité (57,9 %) a été traitée d'une tumeur de gros volume (T3 ou T4).

Un traitement chirurgical a été réalisé dans 88 % des cas, une radiothérapie dans 96 % des cas et une chimiothérapie dans 60 % des cas. 84 % des sujets ont été traités par une association chirurgie-radiothérapie. Le délai médian depuis la fin du traitement est de 40 mois (3 ans et 4 mois), avec des valeurs extrêmes allant de 6 mois à 564 mois (47 ans).

2.5.2.2 Évaluations perceptives de la parole

Une évaluation perceptive a été menée pour obtenir une mesure de référence de la sévérité du trouble de la parole des sujets (« vérité terrain »).

Pour réaliser cette tâche, un jury d'auditeurs experts a été recruté. Il s'agissait d'un médecin phoniatre et de cinq orthophonistes exerçant en service d'ORL / cancérologie, et aguerris à l'évaluation des troubles de parole.

Conformément aux évaluations perceptives conduites précédemment dans le cadre des projets C2SI et RUGBI, chaque auditeur devait évaluer deux paramètres globaux et quatre paramètres spécifiques sur les enregistrements de parole spontanée (c'est-à-dire sur les enregistrements de l'entretien semi-dirigé).

Les deux paramètres globaux, la sévérité du trouble et l'intelligibilité de la parole, ont été évalués de 0 (trouble très sévère / inintelligible) à 10 (absence de trouble / totalement intelligible) (voir Tableau 6). Les quatre paramètres spécifiques (qualité de la voix, résonance, prosodie, altérations phonémiques) ont été évalués de 0 (trouble majeur) à 3 (absence de trouble).

Tableau 6 - Résultats de l'évaluation perceptive de la sévérité du trouble et de l'intelligibilité

	Total	Lot 1	Lot 2	Comparaison inter-lots Valeurs p (test de Mann-Whitney)
Sévérité	25	12	13	0,98
Minimum	0,5	0,67	0,5	
Maximum	10	10	9,5	
Moyenne	5,89	5,75	6,01	
Écart-type	2,77	2,95	2,72	
Quartile 1	3,5	2,92	5,17	
Médiane	6,33	6,33	6,17	
Quartile 3	8	7,83	8,67	
Écart interquartile	4,5	4,91	3,5	
Intelligibilité	25	12	13	0,96
Minimum	1,17	1,17	1,5	
Maximum	10	10	10	
Moyenne	6,87	6,76	6,96	
Écart-type	2,71	2,85	2,68	
Quartile 1	4,83	4,42	5,83	
Médiane	7,17	7,5	7,17	
Quartile 3	9,17	8,83	9,33	
Écart interquartile	4,33	4,42	3,5	

Pour chacun des six paramètres, un coefficient de corrélation intraclasse (ICC) a été calculé (voir Tableau 7). Si les scores globaux et les altérations phonémiques présentent des ICC élevés (supérieurs à 0,80) témoignant d'une homogénéité forte des cotations entre auditeurs, les ICC sont moindres pour les autres paramètres : il est modéré pour la résonance, et faible (inférieur à 0,40) pour la qualité de la voix et la prosodie.

Tableau 7 - ICC des paramètres perceptifs

Paramètre	ICC	Intervalle de confiance à 95 % de l'ICC
Sévérité du trouble de parole	0,85	[0,77 ; 0,93]
Intelligibilité	0,82	[0,72 ; 0,91]
Qualité de la voix	0,34	[0,15 ; 0,52]
Résonance	0,56	[0,39 ; 0,74]
Prosodie	0,26	[0,09 ; 0,44]
Altérations phonémiques	0,82	[0,73 ; 0,91]

Les sujets présentent ainsi des troubles de la parole de tous niveaux de sévérité. Certains sujets présentent une parole estimée comme très dégradée perceptivement (sujets CMH25 et CMB11 par exemple avec des scores de sévérité inférieurs à 1/10), d'autres ne semblent pas présenter perceptivement de trouble (sujet PFG13 avec un score de sévérité à 10/10) (voir Figure 12).

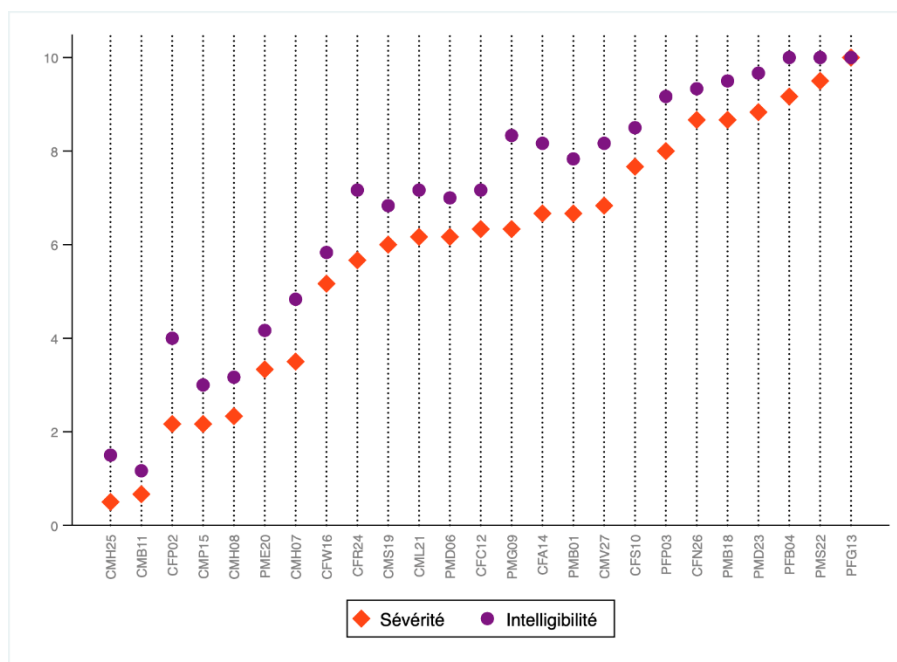


Figure 12 - Scores de sévérité et d'intelligibilité obtenus par les sujets, classés par score croissant de sévérité

De plus, il semble que si les mesures globales du trouble de parole sont homogènes chez les auditeurs experts, les mesures spécifiques sont plus variables. Les outils d'analyse automatique de la parole permettraient alors d'extraire du signal des paramètres plus fins, moins sujets à la variabilité interindividuelle concernant l'acoustique de la voix, le rythme ou le débit, voire même permettraient des mesures plus précises et détaillées du niveau phonémique.

2.6 Analyses statistiques : principes généraux

Toutes les analyses statistiques ont été menées au moyen du logiciel Stata 16.1 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.).

En raison de l'effectif réduit de l'échantillon ($n < 30$), les tests statistiques réalisés dans le cadre de ce manuscrit seront non-paramétriques, sauf mention contraire.

Tout au long de cette étude, nous avons choisi un seuil de significativité de 5 %.

Lors de la constitution de la base de données, certaines données manquantes de complétion de questionnaires ont été retrouvées (Mohan & Pearl, 2014; Rubin, 1976).

De façon générale, les données manquantes de type MNAR (Missing Not At Random) n'ont pas été imputées et ont été laissées manquantes pour ne pas introduire de biais de réponse. Les données MAR (Missing At Random) et MCAR (Missing Completely At Random) ont, quant à elles, été imputées par la moyenne du domaine du questionnaire auquel elles appartiennent, à condition qu'au moins 50 % des items de ce domaine ne soient pas manquants. Dans le cas contraire, ces données ont été laissées manquantes.

2.7 Conclusion

Nous avons constitué un nouveau corpus de parole de patients traités pour un cancer de la cavité buccale ou de l'oropharynx. Le caractère novateur de ce corpus réside dans l'inclusion d'un enregistrement de parole spontanée des patients, au cours d'un entretien semi-dirigé. Nous avons également réalisé des enregistrements sur d'autres tâches de production de parole, permettant des analyses complémentaires si nécessaire (production d'un /a/ tenu, lectures de phrases et de texte, répétition de pseudo-mots).

Notre corpus comprend de nombreuses données individuelles, cliniques et de traitement, ainsi que des résultats d'un test de dépistage cognitif (MoCA) et d'autoquestionnaires évaluant les déficits consécutifs au traitement de la pathologie (CHI), le statut anxiodépressif (HAD et EDP), les besoins de communication liés aux cercles sociaux (ECCS – que nous avons élaboré – et QFS), l'auto-perception du trouble de parole (PHI), la communication (ECVB, DIP) et la qualité de vie globale et relative à la santé (EORTC QLQ-C30, H&N35).

Malgré l'impact de la crise sanitaire sur le recrutement (ayant contraint à une période de suspension des inclusions), 25 sujets ont été inclus. Pour chacun, nous disposons donc de six enregistrements de parole (correspondant aux six tâches de production de parole), et des 11 autoquestionnaires complétés.

La structure globale du corpus, avec les liens entre les niveaux du modèle de Wilson et les outils utilisés est présentée en Figure 13.

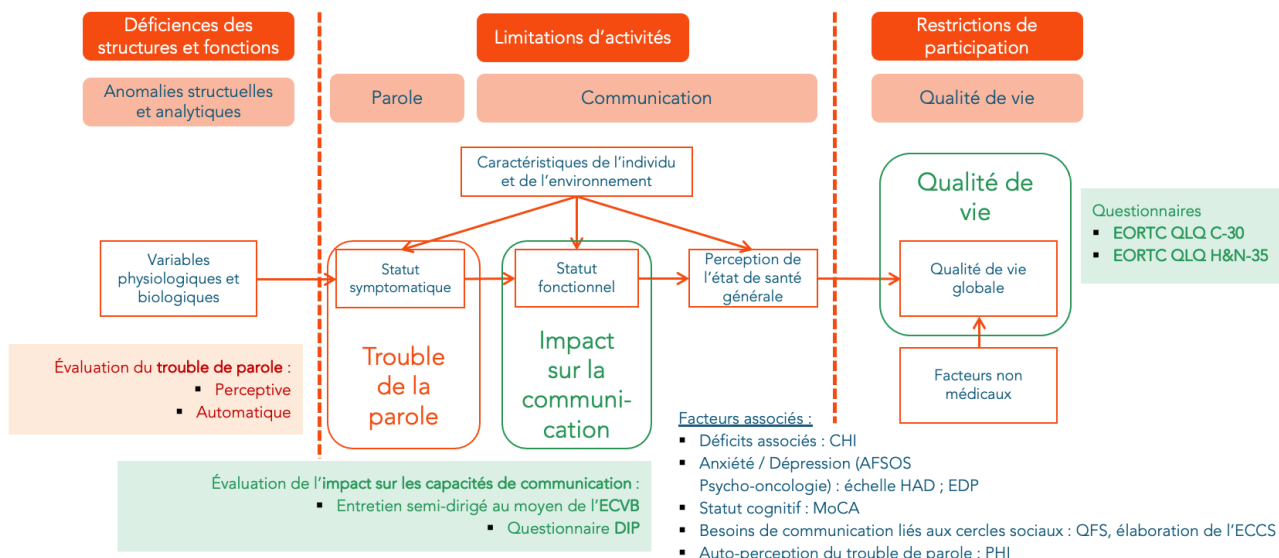


Figure 13 - Structure du corpus, et mise en lien avec les modélisations conceptuelles

Nous avons également constitué une base de données, contenant les résultats issus des autoquestionnaires et les paramètres issus d'analyses automatiques du signal de parole. Afin de pouvoir être exploitée ultérieurement, cette base de données a été construite selon les principes du « data tidying » (Wickham, 2014), qui consiste à optimiser la structure des bases de données pour en faciliter les analyses.

Le corpus étant collecté et disponible, nous allons présenter le travail que nous avons mené concernant l'étude de la parole spontanée au moyen d'analyses automatiques du signal de parole (voir chapitre 3).

3 Analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé

Nous avons souligné la nécessité de prendre en considération les capacités de communication impactées par le trouble de parole dans le cadre de l'évaluation clinique. Bien que la parole spontanée soit le mode de production le plus proche de la parole « quotidienne » des sujets, peu d'études ont porté sur l'analyse automatique de ce type de parole, hormis une étude qui a cherché à identifier automatiquement la parole après cancer de la cavité buccale comparativement à la parole de sujets sains par analyse de la parole spontanée (Halpern et al., 2020). Des analyses automatiques menées sur de la parole spontanée permettraient de s'affranchir des limites de l'évaluation perceptive. Elles fourniraient également des éléments complémentaires à l'évaluation à l'oreille nue, conduisant à une mesure plus fine des conséquences communicationnelles du trouble de parole.

L'objectif de cette partie est d'étudier quels paramètres peuvent être extraits de la parole spontanée par des analyses automatiques après traitement d'un cancer de la cavité buccale ou de l'oropharynx.

Dans un premier temps, nous explorerons différents paramètres issus d'analyses automatiques du signal de parole spontanée utilisables dans l'étude des différents niveaux linguistiques. Pour cela, nous avons cherché quels outils étaient disponibles à l'heure actuelle, puis nous les avons mis en œuvre sur les échantillons de parole de notre corpus. Certains de ces outils ont nécessité quelques adaptations, que nous détaillerons. Une réflexion complémentaire est également proposée pour certains, en vue d'en choisir le paramétrage le plus adapté. Dans un second temps, nous aborderons le processus de sélection ayant abouti à retenir les paramètres les plus pertinents.

3.1 Recherche et extraction des paramètres

Dans cette partie, j'ai extrait des paramètres issus de multiples outils d'analyse automatique nécessitant parfois d'être adaptés au contexte d'étude de la parole spontanée pathologique. Pour rappel, cette parole spontanée correspond à l'enregistrement de l'entretien semi-dirigé entre le sujet et l'examineur lors de la complétion de la version « examineur » de l'ECVB (voir section 2.5.1).

Pour construire notre méthodologie de recherche des paramètres, nous nous sommes basés sur le modèle psycholinguistique de Caron (Devevey & Kunz, 2018; François & Nespoulous, 2011). Nous intéressés à la production de parole, nous avons cherché à extraire des paramètres dans les trois grandes dimensions de production de ce modèle : conceptuelle, catégorielle et articulatoire. Nous avons également cherché à étudier le signal de parole, support de l'échange oral dans le cadre de la communication linguistique.

Ainsi, le signal de parole sera étudié au moyen de paramètres de macrovariabilité de l'intensité et de la fréquence, mais également au moyen de paramètres spectraux et cepstraux.

Le niveau articulatoire concernera, dans sa composante de phonétique articulatoire, les mesures temporelles telles que la durée de parole ou le débit de parole et d'articulation, mais aussi l'étude de segments vocaliques et non vocaliques. Dans sa composante phonologique, les mesures phonémiques seront prises en compte, avec l'étude de la reconnaissance des phonèmes dans la parole des sujets (inventaire phonémique, classes phonétiques).

Le niveau catégoriel sera envisagé selon deux volets. Le premier étudiera le niveau lexical, via l'inventaire des mots reconnus par un système de reconnaissance automatique, et les caractéristiques associées (longueur et fréquence des mots, diversité et densité lexicale). Le second envisagera le niveau grammatical par l'étude des classes grammaticales des mots reconnus.

Enfin, le niveau conceptuel sera étudié. Les thématiques abordées seront analysées par l'étude des sorties du système de reconnaissance au moyen d'une classification hiérarchique descendante. Le sentiment général issu du contenu de l'entretien sera également étudié.

Le schéma général d'analyse des paramètres automatiques est présenté en Figure 14.

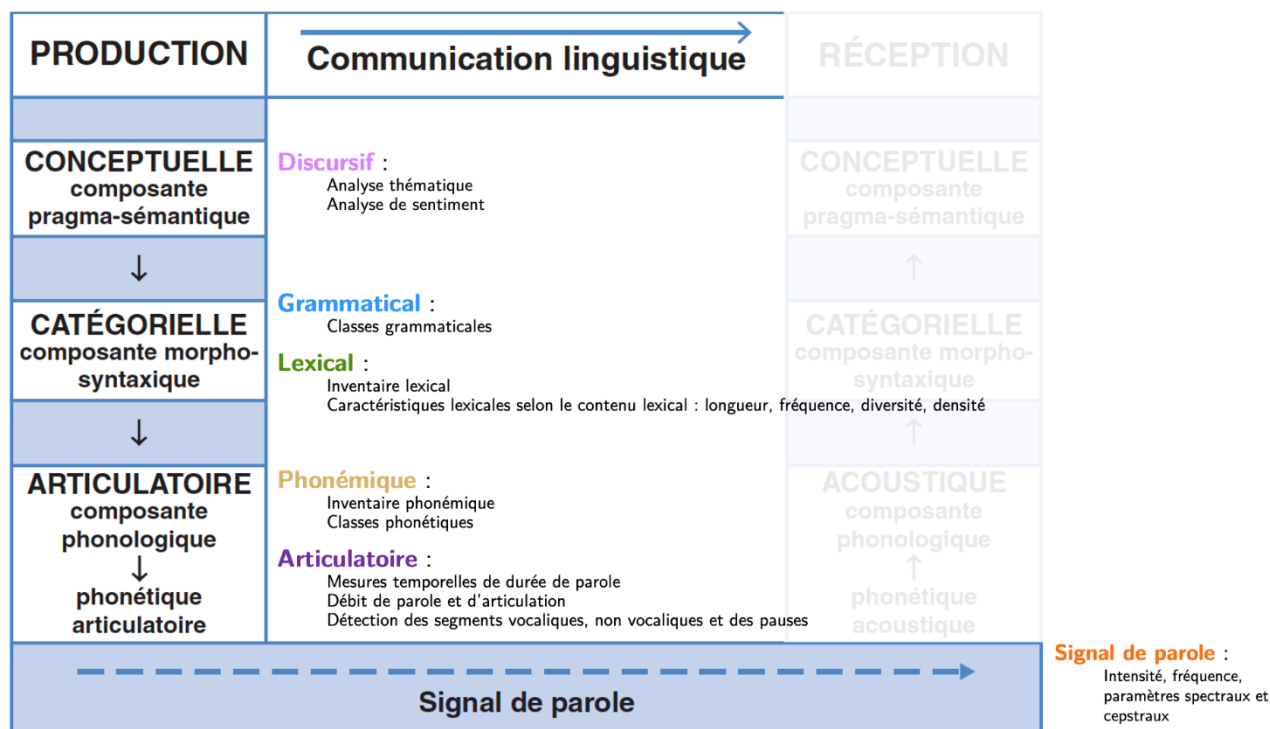


Figure 14 - Schéma général de l'analyse des paramètres automatiques extraits de la parole spontanée selon le modèle psycholinguistique de Caron

3.1.1 Préparation des données

Pour leur traitement, les fichiers audios ont été rééchantillonnés en 16 kHz avec une résolution en 16 bits, toujours en mono, en vue des analyses acoustiques de parole et de l'utilisation de systèmes de reconnaissance de la parole selon le format habituellement utilisé pour la transcription automatique de la parole (Milenkovic et al., 2020).

Nos échantillons de parole spontanée correspondant à l'enregistrement d'un entretien semi-dirigé entre examinateur et sujet, une segmentation préalable des fichiers a été nécessaire. Les segments de parole du sujet ont été déterminés, tout comme les segments « non utiles » correspondant aux pauses ou aux segments de parole de l'examineur (se caractérisant par une faible énergie grâce à l'utilisation du micro serre-tête).

Si cette détermination des segments de parole du sujet est parfois automatiquement gérée par certains outils d'analyses automatiques qui excluent automatiquement les zones de faible énergie, d'autres outils ne procèdent pas de la même manière, et nécessitent en entrée uniquement la parole du patient.

Pour cela, j'ai utilisé l'outil Google WebRTC-VAD¹³. Il s'agit d'un détecteur d'activité vocale (VAD, Voice Activity Detector). Ce VAD a été utilisé en raison de son caractère performant, rapide et libre.

Le paramétrage d'agressivité du VAD a été fixé à 3, soit sa valeur maximale, afin de s'assurer de l'exclusion de l'ensemble des segments de parole de l'examineur.

3.1.2 Signal de parole

La prise en considération du signal acoustique de parole est importante. En effet, en contexte cancérologique, la tumeur ou le traitement entrepris vont modifier la capacité de déformation et la souplesse des cavités de résonance et des articulateurs. Ces altérations conduiront à une modification du signal acoustique véhiculant la parole.

Les analyses de cette section ont été menées au moyen du logiciel Praat¹⁴. Le script utilisé a été adapté de l'étude concernant la validation de l'Acoustic Voice Quality Index (Maryn et al., 2010; Pommée et al., 2020).

Ce script procède à l'extraction des paramètres de façon séquentielle : il recherche, extrait et concatène les segments voisés de la parole continue dans un nouveau fichier appelé « OnlyVoiced ». Puis il calcule la proéminence du pic cepstral lissé (CPPS, Smoothed Cepstral Peak Frequency), la pente du spectre moyen à long terme et l'inclinaison de la courbe de tendance à travers le spectre sur le fichier « OnlyVoiced ». Nous avons également ajouté à ce script le calcul de mesures fréquentielles et d'intensité.

3.1.2.1 F0 et intensité

Les premiers paramètres concernent des indicateurs de macrovariabilité de la fréquence fondamentale et de l'intensité. Ce choix a été motivé pour trois raisons principales. D'une part, ils sont utilisés couramment en pratique clinique (Feenaughty et al., 2014). D'autre part, les mesures de qualité vocale peuvent être informatives sur l'altération de la parole car la source phonatoire laryngée contribue aux contrastes phonétiques voisés / non voisés (par exemple entre /s/ et /z/). Enfin, la modulation de la fréquence fondamentale et de l'intensité au cours d'un énoncé est également un des phénomènes en jeu dans la prosodie de la parole (Pépiot, 2014).

¹³ <https://github.com/wiseman/py-webrtcvad>, consulté le 18/05/2021

¹⁴ <https://www.fon.hum.uva.nl/praat/>, consulté le 18/05/2021

Ainsi, l'écart-type et l'étendue ont été calculés pour ces deux paramètres. Afin de limiter l'effet de sexe, un calcul du centième du demi-ton tempéré a également été mesuré pour la fréquence fondamentale : la valeur c en cents de l'intervalle entre la fréquence maximale et la fréquence minimale retrouvées sur un échantillon de parole se calcule ainsi :

$$c = 1200 \times \log_2 \left(\frac{F0_{max}}{F0_{min}} \right)$$

Les résultats sont présentés dans le Tableau 8.

Tableau 8 - Description statistique des paramètres de macrovariabilité (intensité en décibels, fréquence en hertz sauf $f0cent$ en cents)

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ (Écart Inter-quartile)	Valeur min	Valeur max
<i>intrange</i>	Intensité – étendue	27,27	3,77	26,91	6,56	20,80	34,68
<i>intsd</i>	Intensité – écart-type	3,17	0,24	3,18	0,29	2,71	3,73
<i>f0range</i>	F0 – étendue	282,01	70,40	304,96	97,50	145,26	357,42
<i>f0sd</i>	F0 – écart-type	29,72	9,53	28,76	15,29	16,72	47,72
<i>f0min</i>	F0 – Valeur minimum	69,39	4,03	67,73	7,94	63,38	75,65
<i>f0max</i>	F0 – Valeur maximum	351,40	68,84	379,88	94,33	216,27	423,69
<i>f0cent</i>	F0 cent	2 772,92	426,07	2 955,56	476,00	1 918,15	3 275,13

3.1.2.2 Paramètres spectraux et cepstraux

Les autres paramètres (spectraux et cepstraux) ont été calculés en tant qu'indicateurs plus globaux de la qualité vocale, élément constitutif du signal de parole (Maryn et al., 2009) (voir Tableau 9).

Tableau 9 - Description statistique des paramètres spectraux et cepstraux (en décibels)

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>cpps</i>	Proéminence du pic cepstral lissé	10,25	1,90	10,24	1,83	6,00	13,52
<i>slope</i>	Pente spectrale	-30,44	3,63	-31,12	2,85	-34,90	-18,18
<i>tilt</i>	Inclinaison de la courbe à travers le spectre	-11,02	1,00	-11,06	1,18	-13,51	-8,80

Deux paramètres spectraux ont été calculés sur le spectre moyen à long terme (LTAS : Long-Term Average Spectrum). Il s'agit d'un spectre global issu d'une moyenne de spectres (carré de la magnitude de la transformée de

Fourier d'un signal, (Liang et al., 2013)) calculés à intervalles réguliers de 200 ms.

La pente spectrale (ou slope), calculée en dB, correspond au rapport entre l'énergie des basses fréquences (inférieurs à 100 Hz) et celle des fréquences hautes (supérieures à 100 Hz) à travers le spectre moyen à long terme (Volín & Zimmermann, 2011).

L'inclinaison de la courbe à travers le spectre (ou tilt) correspond à la diminution de l'intensité des harmoniques supérieurs, par rapport à l'intensité de la fréquence fondamentale. En d'autres termes, elle se calcule par la différence d'énergie de la droite de régression entre les basses et les hautes fréquences à travers le spectre moyen à long terme (Mendoza et al., 1996).

Ces mesures du spectre moyen à long terme sont particulièrement pertinentes dans notre étude car elles peuvent être utilisées sur des échantillons de parole continue (Lowell et al., 2011).

Un paramètre a été calculé par l'étude du cepstre (transformation inverse du logarithme du spectre) : la proéminence du pic cepstral lissé ou CPPS (Smoothed Cepstral Peak Proeminence), calculée par l'amplitude logarithmique de la première rhamonique du cepstre de la parole (Hillenbrand et al., 1994). Cette mesure acoustique correspond à un indicateur de la périodicité du spectre vocal pour évaluer la qualité « globale » de la voix (Heman-Ackah et al., 2003). Initialement développé sans lissage du cepstre (Hillenbrand et al., 1994) pour l'évaluation des voix soufflées, le lissage permet d'obtenir de meilleures corrélations avec la qualité vocale.

À l'instar des mesures du domaine spectral calculées sur le spectre moyen à long terme, le CPPS peut être utilisé sur de la parole continue. Des corrélations jusqu'à $|r| = 0,73$ sont retrouvées entre le CPPS issu d'un traitement automatique d'une phrase complète issue d'un texte lu et les mesures perceptives, c'est-à-dire des corrélations plus élevées que lorsque ce paramètre est mesuré sur des voyelles isolées (Haderlein et al., 2012).

Compte tenu de la nature de la tâche-support (parole spontanée au cours de l'entretien), nous avons cherché à vérifier si les paramètres spectraux et cepstraux pouvaient être calculés sur l'ensemble du fichier, ou s'il était pertinent de plutôt les calculer sur des segments de durée plus limitée.

Cette analyse a porté sur les fichiers du lot 1 (voir section 2.1.2). Ainsi, pour chacun des trois paramètres, ont été calculés :

- Un score global sur l'ensemble du fichier « OnlyVoiced » ;
- Des scores sur des segments de 5, 10 et 20 secondes du fichier « OnlyVoiced » : dans un premier temps, sans chevauchement des

fichiers, puis dans un second temps avec un chevauchement d'une seconde ;

- Enfin, des scores uniquement sur les dix premières secondes de chaque fichier « OnlyVoiced », sur les dix dernières, et sur les dix du milieu.

La pente spectrale est significativement différente selon la position et la durée du segment de calcul, qu'il y ait chevauchement ou non (test de Wilcoxon : $p < 0,05$). L'inclinaison de la courbe à travers le spectre ne montre quant à elle aucune différence significative (test de Wilcoxon : $p > 0,05$).

Concernant le CPPS, le test de Wilcoxon ne montre pas de différence significative ($p > 0,05$) sur des segments isolés de 5, 10 et 20 secondes par rapport à la valeur calculée sur le fichier global. De plus, aucune différence significative n'est non plus retrouvée sur les segments isolés des 10 premières, 10 dernières secondes et 10 secondes du milieu du fichier.

Ainsi, à l'exception de la pente spectrale, les autres mesures spectrales et cepstrales sur de la parole spontanée ne semblent pas être influencées par la durée ou la position du segment de calcul. Conformément aux données de la littérature (Lowell et al., 2011), nous choisissons de conserver les mesures calculées sur l'ensemble des fichiers pour la suite des analyses.

3.1.3 Niveau articulatoire – Phonétique articulatoire

3.1.3.1 Mesures temporelles de durée de parole

La durée globale des fichiers en secondes a été récupérée (paramètre noté *duration*, voir Figure 15).

Le début et la fin de l'enregistrement ont été déterminés par l'examineur. L'enregistrement était lancé au début de l'entretien dès que les premiers échanges relatifs à l'impact du trouble de parole sur la communication étaient entrepris. L'examineur coupait l'enregistrement quand ni le sujet ni l'examineur n'avait de nouvel élément à apporter pour poursuivre l'échange.

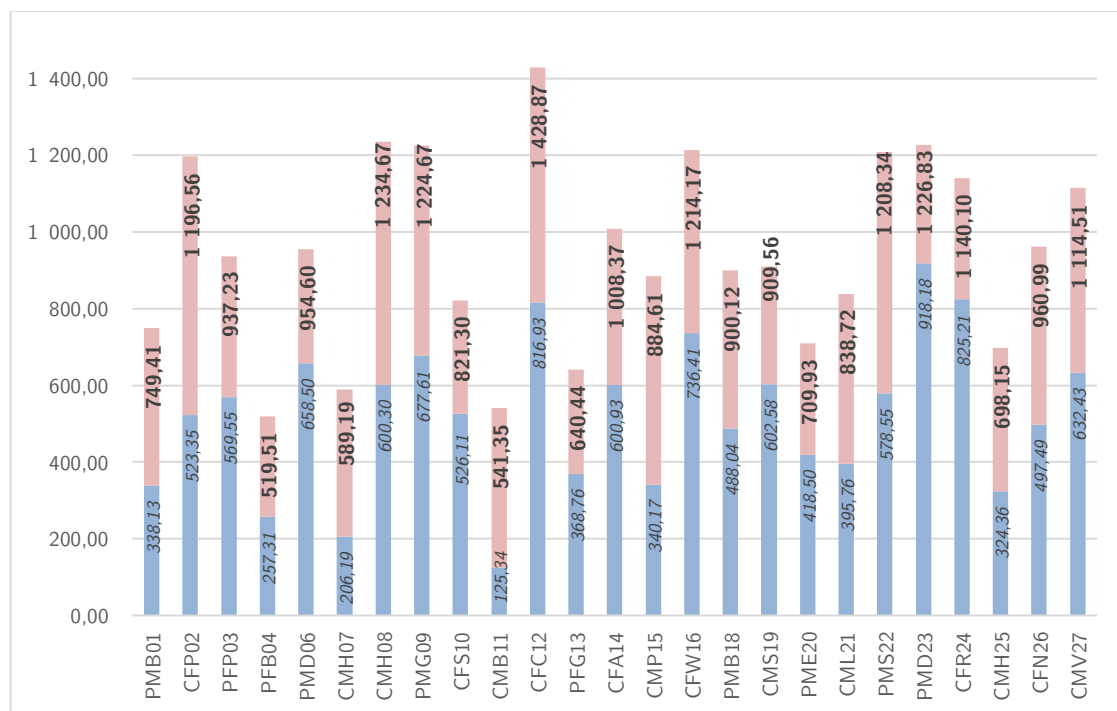


Figure 15 - Durée totale des fichiers (*duration*, en gras) et durée de parole (*phontimew*, en italique), en secondes

Des mesures temporelles concernant la structure de l'échange ont aussi été récupérées au moyen des sorties de l'outil WebRTC-VAD présenté précédemment (voir section 3.1.1, voir Tableau 10). Il s'agit de :

- La durée totale de parole en secondes (*phontimew* : correspondant à la somme de la durée de chaque segment de parole, par sujet) (valeurs en italique dans la Figure 15) ;
- La proportion de durée de parole dans l'enregistrement ($spkprop = \frac{phontimew}{duration}$) ;
- La durée globale de production entre le début de la première prise de parole et la fin de la dernière prise de parole en secondes, incluant les pauses (*ndurw*) ;
- La proportion de durée parlée sur la durée globale de production ($spkdurw = \frac{phontimew}{ndurw}$) ;
- Le nombre de segments de parole produits (*nsegw*) au total, et ce même nombre normalisé en secondes par la durée globale de production ($nsegsecw = \frac{nsegw}{ndurw}$) ;
- La durée moyenne des segments de parole en secondes (*ndursegw*).

Tableau 10 - Description statistique des mesures temporelles calculées sur les résultats de WebRTC-VAD (les durées sont données en secondes)

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>duration</i>	Durée totale du fichier	946,09	245,21	937,23	447,14	519,51	1 428,87
<i>phontimew</i>	Durée totale de parole de chaque sujet	521,07	194,54	526,11	263,67	125,34	918,18
<i>spkprop</i>	Proportion de parole dans le fichier	0,54	0,12	0,55	0,13	0,23	0,75
<i>ndurw</i>	Durée globale de production du sujet entre le début de la première prise de parole et la fin de la dernière prise de parole, incluant les pauses	937,36	244,25	930,84	442,26	517,11	1 421,97
<i>spkdurw</i>	Proportion de durée parlée sur la durée globale de production	0,54	0,12	0,56	0,13	0,23	0,75
<i>nsegw</i>	Nombre de segments de parole	241,76	65,11	253,00	77,00	118,00	385,00
<i>nsegsecw</i>	..., par seconde	0,26	0,04	0,26	0,05	0,19	0,37
<i>ndursegw</i>	Durée moyenne des segments de parole	2,13	0,53	2,10	0,78	1,06	3,00

3.1.3.2 Débit de parole et d'articulation

Le script Praat « Syllable nuclei » (de Jong & Wempe, 2009) a été utilisé pour récupérer de nouveaux indicateurs temporels. Il permet de détecter les noyaux syllabiques et ainsi de mesurer automatiquement le débit d'articulation et de parole. Ces deux paramètres sont couramment évalués en pratique courante auprès de patients traités pour un cancer de la cavité buccale ou de l'oropharynx (Dwivedi et al., 2009, 2012). Il permet aussi de récupérer d'autres paramètres temporels, dont certains sont proches de ceux issus de WebRTC-VAD mais avec des méthodes de calcul différentes.

Ce script fonctionne en quatre étapes. Dans un premier temps, les fichiers sonores sont filtrés avec une limite inférieure à 100 Hz et une limite supérieure à 5 000 Hz (fenêtrage de Hanning). Puis, les segments vocaux dans l'enregistrement de la parole continue sont détectés, extraits et concaténés, pour aboutir à la génération de deux fichiers : un fichier « OnlySilence » contenant tous les segments de silences mais aussi de parole de l'examineur, et un fichier « OnlySound » contenant le reste des segments de production du sujet. Les fichiers sonores sont ensuite normalisés en intensité. Enfin, les noyaux syllabiques sont détectés selon trois critères : l'intensité, le contour

d'intensité et le voisement. Les noyaux syllabiques sont ainsi repérés par un pic voisé caractérisé par une intensité supérieure à 2 dB à l'intensité médiane, et supérieure à 4 dB à l'intensité précédent immédiatement le pic.

Neuf paramètres sont ainsi extraits (voir Tableau 11) :

- Le nombre de syllabes détectées (soit le nombre de noyaux syllabiques détectés ; $nsylp$) ;
- Le nombre de pauses détectées ($npausep$) ;
- La durée globale de production voisée du sujet entre le début de la première prise de parole et la fin de la dernière prise de parole en secondes, incluant les pauses ($ndurp$) ;
- La durée totale de parole voisée de chaque sujet en secondes ($phontimep$) ;
- La durée totale des silences en secondes ($silp = ndurp - phontimep$) ;
- La proportion de durée de parole voisée du sujet dans l'enregistrement ($spkdurp$) ;
- Le débit de parole en syllabes par seconde ($spratep = \frac{nsylp}{ndurp}$), le débit de parole se calcule en prenant en compte les pauses du discours ;
- Le débit d'articulation en syllabes par secondes ($artratep = \frac{nsylp}{phontimep}$), il se calcule en excluant les pauses du discours ;
- Enfin, la durée moyenne d'une syllabe en secondes ($asdp = \frac{phontimep}{nsylp} = \frac{1}{artratep}$).

L'utilisation des deux outils WebRTC-VAD et « Syllable nuclei » de Praat conduit à obtenir des paramètres proches : la durée globale de production du sujet entre le début de la première prise de parole et la fin de la dernière prise de parole du sujet incluant les pauses ($ndur-p/w$), la durée totale de parole ($phontime-p/w$) et la proportion de parole dans l'enregistrement ($spkdur-p/w$).

Nous avons réalisé un test de Wilcoxon en série appariée pour comparer les scores issus de ces deux outils. Les durées globales de production ($ndur$) ne sont pas significativement différentes ($p = 0,10$), tandis qu'une différence significative est retrouvée sur la durée globale de parole ($phontime$, $p < 0,001$) et sur la proportion de parole dans l'enregistrement ($spkdur$, $p < 0,001$).

L'existence de différences significatives entre les paramètres issus de WebRTC-VAD et du script « Syllable nuclei » de Praat montre que les modalités d'extraction des segments voisés et de parole réalisée par les deux outils sont différentes. Dans ce contexte, l'ensemble des paramètres de WebRTC-VAD et de « Syllable nuclei » seront conservés pour la suite des analyses.

Tableau 11 - Description statistique des mesures temporelles extraites par le script « Syllable nuclei » de Praat

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>nsylp</i>	Nombre de syllabes / nombre de noyaux syllabiques	1 231,40	506,92	1 280,00	645,00	135,00	2 332,00
<i>npausep</i>	Nombre de pauses	283,20	87,82	284,00	114,00	124,00	460,00
<i>ndurp</i>	Durée globale de production voisée du sujet entre le début de la première prise de parole et la fin de la dernière prise de parole en secondes, incluant les pauses	934,06	242,09	930,28	425,96	516,67	1 416,30
<i>phontimep</i>	Durée totale de parole voisée, en secondes	377,09	160,42	406,92	179,06	45,11	763,51
<i>silp</i>	Durée totale des silences, en secondes	556,97	145,83	515,06	229,39	322,20	863,27
<i>spkdurp</i>	Proportion de parole voisée dans l'enregistrement	0,39	0,11	0,43	0,12	0,09	0,63
<i>spratep</i>	Débit de parole (syllabes/seconde)	1,29	0,40	1,26	0,55	0,25	2,01
<i>artratep</i>	Débit d'articulation (syllabe/seconde)	3,30	0,46	3,33	0,45	2,52	4,48
<i>asdp</i>	Durée moyenne d'une syllabe en secondes	0,31	0,04	0,30	0,04	0,22	0,40

3.1.3.3 Détection des segments vocaliques et pauses

Nous avons utilisé un dernier outil pour évaluer la structure du signal de parole en segments vocaliques et non vocaliques : SpeechTools. L'algorithme utilisé est une version optimisée d'un extracteur de pseudo-syllabes présenté dans le cadre de travaux universitaires (Farinas, 2002; Pellegrino, 1998). Cet algorithme, facile d'utilisation car demandant peu de réglages, permet de décomposer le signal de parole des enregistrements, sans nécessiter d'avoir recours à un système de transcription de la parole.

L'algorithme commence à extraire la fréquence fondamentale grâce à l'outil externe Reaper¹⁵. Puis, il réalise une segmentation automatique issue de

¹⁵ <https://www.reaper.fm/>, consulté le 18/05/2021

l'algorithme « Forward-Backward Divergence » (Andre-Obrecht, 1988) au moyen de l'outil externe *diverg*. Il s'agit d'une étude statistique du signal dans le cadre temporel, où une segmentation infra-phonémique est réalisée, les segments correspondant à des parties stables ou transitoires des phonèmes (autrement appelés segments quasi-stationnaires).

Un partitionnement en segments vocaux et non vocaux est réalisé au moyen de WebRTC-VAD, intégré à cet algorithme. Le segment le moins intense en termes d'énergie est détecté et les autres segments sont classés comme silence ou activité selon un seuil adaptatif. Puis, les segments vocaliques sont repérés à partir d'une analyse spectrale, en détectant les segments vocaux de basse fréquence (grâce à la courbe de cumul d'énergie réduite).

Les pseudo-phonèmes sont alors construits. Un segment vocalique contenant la fréquence fondamentale et n'ayant pas de composante en hautes fréquences correspond à une pseudo-voyelle, un segment non-vocalique à une pseudo-consonne. L'algorithme labellise également les pseudo-plosives correspondant à des silences inférieurs à 150 ms (comme dans les sons /p/, /t/ et /k/). Les pseudo-silences sont des silences dont la durée est supérieure à 150 ms. Nous appellerons dans la suite de ce travail « pseudo-phonèmes », les pseudo-consonnes, pseudo-voyelles, pseudo-plosives et pseudo-silences.

Cet algorithme permet également de constituer des pseudo-syllabes, autrement dit des groupes d'un nombre indéfinis de pseudo-phonèmes se finissant par une voyelle. Bien que cette structure soit la structure syllabique la plus fréquente dans le monde (Dauer, 1983), nous utiliserons les données de plus petite granularité, au niveau des pseudo-phonèmes, car ils sont plus précis pour étudier la structure temporelle de la parole.

Vingt-six paramètres ont été extraits (voir Tableau 12) :

- Le nombre total de pseudo-phonèmes (*tot_cvps*) et le nombre total de pseudo-consonnes (*c_cvps*), pseudo-plosives (*p_cvps*), pseudo-voyelles (*v_cvps*) et pseudo-silences (*s_cvps*) : cinq paramètres ;
- Les mêmes valeurs normalisées avec nombre par seconde ($Xs_cvps = \frac{X_cvps}{phontimew}$) : cinq paramètres ;
- La proportion de pseudo-consonnes, pseudo-plosives, pseudo-consonnes et pseudo-silences (*propX_cvps*) : quatre paramètres ;
- La somme (*sumdurX*), la moyenne (*meandurX*) et l'écart-type (*sddurX*) des pseudo-consonnes, pseudo-plosives, pseudo-voyelles et pseudo-silences : 12 paramètres.

Tableau 12 - Description statistique des mesures extraites de SpeechTools

Code	Paramètre	Moyenne	Écart-type	Médiane	EQ	Valeur min	Valeur max
<i>c_cvps</i>	Nombre de pseudo-consonnes total	4 774,48	2 249,96	4 831,00	3 432,00	927,00	9 916,00
<i>cs_cvps</i>	..., par seconde	9,00	2,08	9,34	2,07	4,50	11,57
<i>p_cvps</i>	Nombre de pseudo-plosives total	771,84	304,80	794,00	368,00	248,00	1 404,00
<i>ps_cvps</i>	..., par seconde	1,54	0,44	1,46	0,59	0,67	2,53
<i>s_cvps</i>	Nombre de pseudo-silences total	75,12	31,19	67,00	39,00	28,00	156,00
<i>ss_cvps</i>	..., par seconde	0,16	0,06	0,14	0,08	0,07	0,32
<i>v_cvps</i>	Nombre de pseudo-voyelles total	934,96	435,04	986,00	486,00	174,00	2 060,00
<i>vs_cvps</i>	..., par seconde	1,75	0,35	1,70	0,59	1,12	2,41
<i>tot_cvps</i>	Nombre de pseudo-phonèmes total	6 556,40	2 840,81	6 817,00	3 091,00	1 639,00	12 982,00
<i>tots_cvps</i>	..., par seconde	12,45	2,05	13,07	2,63	8,18	15,20
<i>propc_cvps</i>	Proportion de pseudo-consonnes parmi les pseudo-phonèmes	0,71	0,07	0,73	0,07	0,54	0,80
<i>propp_cvps</i>	Proportion de pseudo-plosives parmi les pseudo-phonèmes	0,13	0,05	0,11	0,07	0,05	0,26
<i>props_cvps</i>	Proportion de pseudo-silences parmi les pseudo-phonèmes	0,01	0,01	0,01	0,01	0,01	0,04
<i>propv_cvps</i>	Proportion de pseudo-voyelles parmi les pseudo-phonèmes	0,14	0,02	0,15	0,02	0,10	0,19
<i>sumdurc</i>	Somme de la durée des pseudo-consonnes	187,90	84,90	193,77	107,54	44,63	370,49
<i>meandurc</i>	Moyenne de la durée des pseudo-consonnes	0,04	0,00	0,04	0,00	0,03	0,05
<i>sddurc</i>	Écart-type de la durée des pseudo-consonnes	0,03	0,01	0,03	0,01	0,02	0,04
<i>sumdurp</i>	Somme de la durée des pseudo-plosives	45,29	17,19	46,10	24,82	14,13	78,78
<i>meandurp</i>	Moyenne de la durée des pseudo-plosives	0,06	0,00	0,06	0,00	0,05	0,07
<i>sddurp</i>	Écart-type de la durée des pseudo-plosives	0,03	0,00	0,03	0,00	0,03	0,04
<i>sumdurs</i>	Somme de la durée des pseudo-silences	15,17	6,66	13,58	6,86	5,27	34,18
<i>meandurs</i>	Moyenne de la durée des pseudo-silences	0,20	0,01	0,20	0,01	0,19	0,22

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>sddurs</i>	Écart-type de la durée des pseudo-silences	0,05	0,02	0,05	0,02	0,03	0,10
<i>sumdurv</i>	Somme de la durée des pseudo-voyelles	235,42	97,83	228,84	143,20	35,51	434,50
<i>meandurv</i>	Moyenne de la durée des pseudo-voyelles	0,26	0,07	0,24	0,07	0,18	0,48
<i>sddurv</i>	Écart-type de la durée des pseudo-voyelles	0,24	0,06	0,23	0,05	0,15	0,40

3.1.4 Niveau articulatoire – Composante phonémique

L'analyse phonémique de la parole spontanée pose un problème majeur : si des analyses acoustiques précises peuvent être menées sur de la parole contrainte grâce à un alignement (par exemple sur des productions de phonèmes isolés ou de la lecture de texte), il n'est pas envisageable de transcrire phonétiquement le discours spontané issu d'entretiens pour cet alignement. En effet, la durée des enregistrements d'une part et la qualité de la parole post-traitement cancérologique d'autre part rendraient ce processus trop coûteux et peu fiable.

Nous avons ainsi pris le parti d'utiliser un système de reconnaissance de la parole permettant d'obtenir des séries de phonèmes sur lesquelles baser nos analyses ultérieures.

3.1.4.1 Présentation du système de reconnaissance automatique de la parole

Quelques éléments de base concernant le fonctionnement des systèmes de reconnaissance automatique de la parole sont donnés en **Annexe V**.

Les modèles utilisés pour la reconnaissance de la parole s'appuient sur des modèles de Markov cachés (HMM, Hidden Markov Model), initialement liés à des distributions gaussiennes simples ou à des mélanges de lois gaussiennes (GMM, Gaussian Mixture Models). Mais l'évolution de la recherche dans ce domaine a permis de remplacer progressivement ces modèles GMM par des architectures de modèles acoustiques hybrides s'appuyant sur des réseaux de neurones profonds (DNN-HMM, Deep Neural Network-Hidden Markov Model) (Saikia, 2017). Toutefois, les modèles GMM-HMM restent encore utilisés spécifiquement pour les alignements permettant l'entraînement des DNN-HMM.

Parmi les modèles de type DNN-HMM, certains ont été développés dans un objectif de reconnaissance de phones : les TDNN-HMM (Time-Delay Neural

Network) (Sugiyama et al., 1991; Waibel et al., 1989). Parmi les TDNN, ce sont les TDNNf « factorisés » (Povey et al., 2018, 2016) qui font référence à l'heure actuelle en matière de reconnaissance de la parole, et qui sont utilisés pour la reconnaissance de parole non typique comme la parole d'enfants par exemple (Gelin et al., 2021; Wu et al., 2019).

Le modèle TDNNf-HMM utilisé dans le cadre de cette étude a été développé par Gelin (Gelin et al., 2021) au moyen de la boîte à outils Kaldi (Pecchinetti et al., 2015; Povey et al., 2011). Ce modèle à 12 couches cachées présente un Phone Error Rate (ou taux d'erreurs phonèmes) à 23,5 % sur un corpus d'adultes typiques (Gelin et al., 2021).

La base de données en ligne Common Voice¹⁶ a été utilisée pour entraîner le TDNNf-HMM, à partir d'alignements générés au niveau phonémique issus d'un modèle GMM-HMM se basant sur des MFCC de dimension 40 à trois états. Il s'agit d'un modèle monophone, permettant de n'avoir aucun a priori sur le contexte de coarticulation appris sur des personnes non pathologiques et donc ne pas biaiser la reconnaissance. Ce dernier donne de meilleurs résultats qu'un modèle triphone en raison des distorsions phonétiques de la parole non typique (Gelin et al., 2021).

En français, le corpus d'entraînement est constitué de 148,9 heures d'enregistrement de texte lu, par 1 276 locuteurs, pour une moyenne de 420 secondes d'enregistrement par locuteur (corpus Common Voice, septembre 2019).

Nous avons fait le choix de conserver un entraînement sur de la parole adulte typique. D'une part, notre objectif n'est pas ici de proposer le meilleur modèle de reconnaissance automatique de la parole cancérologique, mais une analyse de phonèmes reconnus pourrait donner des indices sur un écart à un modèle acoustique « normal » ou « typique », plus intéressant dans le cadre de l'impact du trouble de parole sur la communication. D'autre part, en raison de son caractère ouvert (chacun peut s'enregistrer et contribuer à l'enrichissement de cette base de données Common Voice) impliquant des environnements et des matériels d'enregistrements différents selon les contributeurs, ce corpus semble adapté à notre étude en raison de son caractère variable, même si la cause en est différente (parole pathologique dans notre cas).

¹⁶ <https://voice.mozilla.org/fr>, consulté le 18/05/2021

Pour le décodage de notre corpus, nous avons fourni au système les segments de parole déterminés par WebRTC-VAD, afin de limiter les effets de résidus de parole de l'examineur pouvant altérer la séquence des phonèmes reconnus.

Dans chaque trame de 25 ms (avec un pas de 10 ms), le phone (ou manière de prononcer un son) s'approchant le plus des caractéristiques acoustiques portées par le signal sera retenu, en fonction du modèle acoustique établi lors de l'étape d'entraînement (partie HMM du TDNNf). Le système produira en sortie le phonème associé au phone, parmi 33 possibilités : 18 consonnes, 12 voyelles (un seul phonème pour /a/ et /ɑ/, un seul pour /ø/, /œ/ et /ə/, et un seul pour /ē/ et /œ̃/ qui ne sont donc pas distingués), et trois semi-consonnes.

De plus, à chaque phonème reconnu est associé un score de confiance obtenu au moyen d'une méthode MBR (Minimum Bayes Risk) se basant sur la séquence du chemin retenu ayant abouti à la reconnaissance du phonème parmi l'ensemble des chemins possibles (correspondant aux autres hypothèses les plus probables) (Xu et al., 2011).

3.1.4.2 Paramétrage du système

Lors du lancement du système TDNNf-HMM, deux paramètres doivent être préalablement fixés : le WIP (Word Insertion Penalty), et le LMWT (Language Model Weights).

Le WIP permet de modifier les poids du modèle acoustique. Il joue sur la transition entre les unités reconnues, et permet d'ajuster la taille des unités de sortie (Nejedlová, 2002). Le WIP peut prendre trois valeurs pénalisant la reconnaissance d'unités courtes : 0 (pas de pénalité), 0,5 et 1 (le système est ici poussé à ne pas préférer les unités courtes). Ici, seules des sorties courtes (les unités phonémiques) seront attendues. Ainsi, un WIP à 0 a été retenu afin de ne pas pénaliser l'insertion de phonèmes.

Le LMWT permet de spécifier à quel point le modèle de langage va aider le modèle acoustique pendant le décodage. Il varie entre 7 (poids minimal du modèle de langage) et 17 (poids maximal), en prenant des valeurs entières. Nous avons opté pour un LMWT à 7 car nous souhaitons un traitement le plus brut possible, sans optimisation trop importante liée au modèle de langage qui risquerait de fausser les résultats au niveau phonémique pur.

Une vérification empirique des sorties pour chacun des 33 paramétrages (trois pour le WIP et 11 pour le LMWT) a permis de confirmer ce postulat théorique et le choix du paramétrage {WIP = 0 ; LMWT = 7}.

3.1.4.3 Inventaire phonémique

Selon les sorties du système, plusieurs scores ont été calculés. Ils permettent de dresser un inventaire des phonèmes reconnus par le système dans la parole des sujets.

Certains ont été normalisés, en divisant les valeurs brutes par la durée totale de parole de chaque sujet (*phontimeuw*). Cette normalisation aboutit à des valeurs par seconde, permettant de limiter les effets de durée d'enregistrement (un enregistrement long contenant davantage de productions du sujet, donc davantage de phonèmes reconnus) (voir Tableau 13).

Un premier groupe de scores concerne des scores globaux :

- Le nombre total de phonèmes reconnus par sujet (*sumphon*), ainsi que sa valeur normalisée par seconde (*sumphons*) ;
- Le nombre de phonèmes différents reconnus, indicateur de la variété phonémique (*difphon*) : valeur pouvant aller de 0 (aucun phonème reconnu) à 33 (au moins une occurrence de chaque phonème a été reconnue).

Un second groupe de phonèmes est issu de regroupements phonétiques :

- Le nombre total de consonnes reconnues (*csn*) et sa valeur normalisée (*csns*) ;
- Parmi les consonnes, le nombre d'occlusives (valeur brute *occ*, et valeur normalisée *occs*) et de fricatives (valeur brute *fri*, et valeur normalisée *fris*), ainsi que les taux d'occlusives (*tx_occcsn*) et de fricatives (*tx_fricasn*) parmi le nombre total de consonnes reconnues ;
- Le nombre total de voyelles reconnues (*voy*), et sa valeur normalisée (*voys*), ainsi que le taux de voyelles reconnues parmi l'ensemble des phonèmes reconnus ($txvoy = \frac{voy}{sumphon}$), et le taux de voyelles nasales parmi les voyelles reconnues (*txvnasal*) ;
- Le nombre total de semi-consonnes reconnues (*semics*) et sa valeur normalisée (*semicsns*) ;
- Le taux de phonèmes nasaux, consonnes et voyelles, parmi l'ensemble des phonèmes reconnus (*txnasal*) ;
- Le nombre total de sonantes (*sonant* : l m n ɲ ŋ ʀ j ɥ w) et de non-sonantes (*nonsonant* : p b t d k g f v s z ʃ ʒ), ainsi que leurs valeurs normalisées (respectivement *sonants* et *nonsonants*), et les taux de chacun parmi les consonnes reconnues (*tx_sonant* et *tx_nonsonant*).

3.1.4.4 Scores de confiance à la sortie du système

Les scores de confiance sont récupérés à la sortie du système, soit après la reconnaissance opérée par le TDNNf-HMM et les optimisations réalisées par Kaldi. Le programme calcule les scores en récupérant le phonème associé à la lattice (chemin représentant des séquences de phonèmes suffisamment probables pour un énoncé spécifique) correspondant à la meilleure estimation finale¹⁷.

Un dernier groupe de paramètres récupérés avec le système concerne ainsi les scores de confiance relatifs aux phonèmes reconnus (voir Tableau 13) :

- Le score moyen global de confiance sur l'ensemble des phonèmes reconnus (*conf*) ;
- Un score moyen global de confiance par groupe : consonnes (*confc*), voyelles (*confv*), semi-consonnes (*confs*), occlusives (*confo*) et fricatives (*confff*).

Tableau 13 - Description statistique des mesures phonémiques extraites avec le système TDNNf-HMM

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>sumphon</i>	Nombre total de phonèmes reconnus	2 503,04	1 328,30	2 740,00	1 794,00	141,00	6 075,00
<i>sumphons</i>	..., par seconde	4,55	1,56	4,78	2,40	1,12	7,49
<i>difphon</i>	Nombre de phonèmes différents reconnus	29,20	5,57	32,00	3,00	5,00	32,00
<i>csn</i>	Nombre total de consonnes reconnues	1 230,52	675,64	1 324,00	972,00	21,00	2 924,00
<i>csns</i>	..., par seconde	2,23	0,89	2,34	1,27	0,17	4,05
<i>voy</i>	Nombre total de voyelles reconnues	1 212,16	633,13	1 367,00	872,00	120,00	3 045,00
<i>voys</i>	..., par seconde	2,22	0,65	2,33	1,06	0,96	3,32
<i>semics</i>	Nombre total de semi-consonnes reconnues	60,36	50,15	56,00	64,00	0,00	217,00
<i>semicsns</i>	..., par seconde	0,11	0,08	0,11	0,11	0,00	0,36
<i>txvoy</i>	Taux de voyelles reconnues	0,51	0,09	0,49	0,04	0,43	0,85
<i>txnasal</i>	Taux de phonèmes nasaux reconnus	0,19	0,06	0,19	0,06	0,06	0,37
<i>txvnasal</i>	Taux de voyelles nasales reconnues par les voyelles	0,18	0,10	0,17	0,09	0,05	0,44

¹⁷ <https://white.ucc.asn.au/Kaldi-Notes/tidigits/eval>, consulté le 18/05/2021

ANALYSES AUTOMATIQUES DE LA PAROLE SPONTANÉE APRES TRAITEMENT D'UN CANCER
ORAL OU OROPHARYNGE

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>occ</i>	Nombre total d'occlusives reconnues	329,76	262,74	325,00	343,00	0,00	1 048,00
<i>occs</i>	..., par seconde	0,58	0,41	0,56	0,62	0,00	1,51
<i>fri</i>	Nombre total de fricatives reconnues	440,76	227,99	457,00	303,00	0,00	904,00
<i>fris</i>	..., par seconde	0,80	0,29	0,85	0,31	0,00	1,15
<i>tx_occcsn</i>	Taux d'occlusives reconnues parmi les consonnes	0,23	0,12	0,27	0,20	0,00	0,37
<i>tx_fricsn</i>	Taux de fricatives reconnues parmi les consonnes	0,36	0,14	0,34	0,16	0,00	0,78
<i>sonant</i>	Nombre total de sonantes reconnues	520,36	276,49	578,00	412,00	21,00	1 078,00
<i>sonants</i>	..., par seconde	0,96	0,38	1,00	0,48	0,17	1,62
<i>nonsonant</i>	Nombre total de non-sonantes reconnues	770,52	449,60	760,00	665,00	0,00	1 952,00
<i>nonsonants</i>	..., par seconde	1,38	0,60	1,33	0,76	0,00	2,61
<i>tx_sonant</i>	Taux de sonantes reconnues parmi les consonnes	0,46	0,14	0,42	0,11	0,23	1,00
<i>tx_nsonant</i>	Taux de non-sonantes reconnues parmi les consonnes	0,59	0,14	0,62	0,13	0,00	0,78
<i>conf</i>	Score moyen global de confiance	0,84	0,02	0,84	0,03	0,78	0,89
<i>confc</i>	Score moyen global de confiance des consonnes	0,87	0,04	0,88	0,03	0,76	0,93
<i>confv</i>	Score moyen global de confiance des voyelles	0,80	0,03	0,80	0,02	0,77	0,91
<i>confs</i>	Score moyen global de confiance des semi-consonnes	0,76	0,04	0,76	0,04	0,65	0,84
<i>confo</i>	Score moyen global de confiance des occlusives	0,87	0,07	0,90	0,09	0,72	0,95
<i>conff</i>	Score moyen global de confiance des fricatives	0,88	0,04	0,88	0,05	0,79	0,93

3.1.5 Composante morpho-syntaxique – Composante lexicale

3.1.5.1 Paramétrage du système

Afin de procéder aux analyses des enregistrements au niveau lexical, nous avons utilisé le même système de reconnaissance de la parole : le TDNNf-HMM, appris sur le corpus Common Voice.

Le fonctionnement du système utilisé ici (que nous appellerons TDNNf « lexical ») est quasi-identique à celui décrit précédemment. Néanmoins, quelques adaptations ont dû être réalisées pour la reconnaissance de mots au lieu de phonèmes.

Concernant l'architecture du système, le modèle GMM-HMM utilisé pour générer les alignements est un modèle triphone. C'est une architecture classique pour la reconnaissance de la parole grand vocabulaire en français, car elle permet de prendre en compte les enchaînements phonémiques du français. De plus, dans cette adaptation, le modèle de langage et le dictionnaire de prononciation (contenant 165 363 mots) permettent la reconnaissance de mots en fonction des phonèmes reconnus.

Nous avons également opté pour fournir en entrée du système les fichiers complets, contrairement au TDNNf « phonémique » pour lequel les segments de parole issus de la segmentation par WebRTC-VAD étaient utilisés. En effet, un pré-test a montré que l'utilisation des segments définis par le VAD limitait le travail du modèle de langage, avec des résultats sensiblement différents quand l'analyse était lancée sur l'ensemble de l'audio ou uniquement sur les segments prédéterminés. Ceci est peut-être dû à la segmentation « agressive » du VAD aboutissant à des segments courts et multiples perturbant la reconnaissance plus globale du message produit par le modèle de langage.

Nous avons également mené une analyse complémentaire sur le paramétrage du WIP et du LMWT. Nos hypothèses émises pour l'analyse phonémique nécessitant une analyse la plus brute possible, sans optimisation, n'est en effet plus valable au niveau lexical. Nous nous sommes alors demandé quelles valeurs attribuer à ces deux paramètres du système.

Pour cela, nous avons choisi d'étudier les performances du système sur une minute des enregistrements du lot 1, en comparant les sorties avec une référence issue d'une transcription perceptive du discours (Young et al., 2015).

J'ai ainsi transcrit manuellement une minute des enregistrements des 12 patients du lot 1, entre la 301^{ème} et la 360^{ème} seconde du fichier : une sélection de la fenêtre temporelle en tout début d'enregistrement aurait pu ne pas être représentative de la parole usuelle du sujet (en raison des phénomènes contextuels de découverte de l'examineur ou de début d'entretien), et des

effets de fatigabilité de parole auraient pu survenir en fin d'enregistrement. Cette transcription a été comparée aux 33 sorties du TDNNf « lexical » (une sortie par paramétrage WIP / LMWT ; pour rappel, le WIP peut prendre trois valeurs différentes et le LMWT 11 valeurs différentes). Le paramétrage ayant les sorties les plus proches de la « vérité terrain » transcrite sera alors retenu pour la suite des analyses.

En comparant les contenus des fichiers contenant les transcriptions perceptives et ceux des fichiers relatifs aux sorties du TDNNf, la boîte à outils HTK (Young et al., 2015) permet d'obtenir le nombre de mots correctement reconnus (H), les substitutions (S), les délétions (D) et les insertions (I) de mots, (N) correspondant au nombre total de mots cibles (présents dans la transcription perceptive).

En fonction de ces scores, HTK permet de calculer les indicateurs suivants :

- L'accuracy (ou exactitude) :

$$Accuracy = \left(\frac{H - I}{N} \right) \times 100 = \left(\frac{H - I}{H + S + D} \right) \times 100$$

- Le pourcentage de mots corrects (ou rappel), soit la proportion de mots correctement reconnus parmi les mots cibles :

$$Rappel = \left(\frac{H}{N} \right) \times 100 = \left(\frac{H}{H + S + D} \right) \times 100$$

- La précision, soit la proportion de mots corrects parmi les mots reconnus :

$$Précision = \left(\frac{H}{H + S + I} \right) \times 100$$

- Le taux d'erreurs mots, c'est-à-dire la proportion de mots mal reconnus parmi le nombre total de mots cibles :

$$Taux\ d'erreur\ mots = \left(\frac{S + D + I}{H + S + D} \right) \times 100$$

Le paramétrage présentant la meilleure accuracy est le paramétrage {WIP = 0 ; LMWT = 11} avec une accuracy à 15,49 % (voir Figure 16).

Nous avons ainsi choisi de retenir ce paramétrage dont les valeurs détaillées se situent dans le Tableau 14.

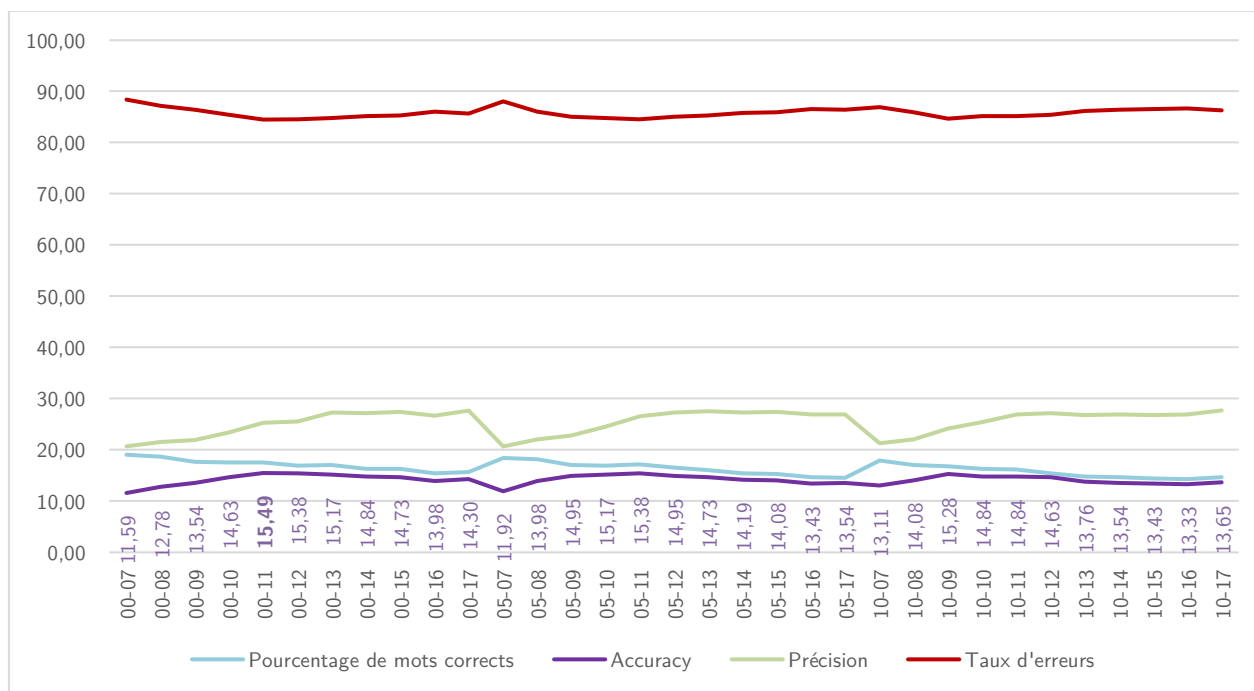


Figure 16 - Métriques d'évaluation : en abscisses, le paramétrage du TDNNf « lexical » (les deux premiers chiffres correspondent au WIP – 00 : WIP=0 ; 05 : WIP=0,5 ; 10 : WIP=1 –, les deux derniers au LMWT)

Même si ce paramétrage présente les meilleures performances en termes d'accuracy parmi les 33 testés, les performances sont malgré tout globalement modérées et le taux d'erreurs mots est très élevé (84,51 %).

Tableau 14 - Valeurs de performances du système avec le paramétrage retenu {WIP = 0 ; LMWT = 11}

Paramètre	Valeur
H : Nombre de mots correctement reconnus	162
D : Nombre de délétions	303
S : Nombre de substitutions	458
I : Nombre d'insertions	19
N : Nombre total de mots cibles	923
Accuracy	15,49
Rappel	17,55
Précision	25,35
Taux d'erreurs mots	84,51

Nous avons ainsi cherché à savoir si les performances du système sous ce paramétrage étaient liées à la sévérité du trouble de parole présenté par les

patients, et si dans ce cas une analyse en sous-groupe selon la sévérité serait pertinente. Aucun écart net selon la sévérité perçue par le jury d'experts ne se dégage (voir Figure 17).

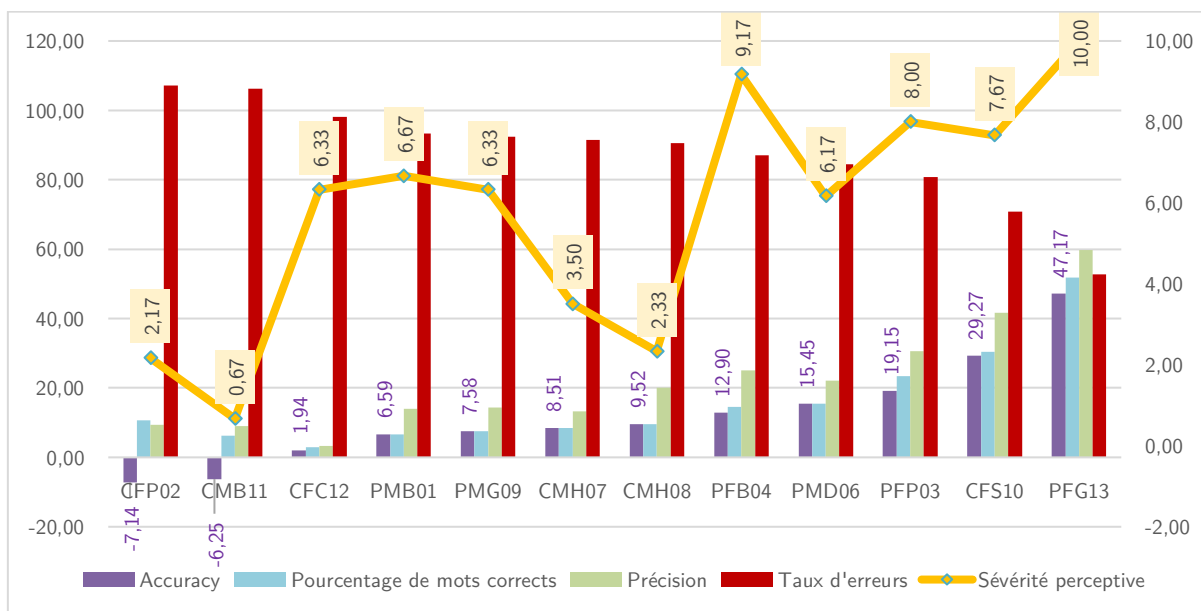


Figure 17 - Performances du système sous le paramétrage WIP=0 ; LMWT = 11 selon la sévérité perceptive (lot 1 uniquement). L'axe des ordonnées de droite est relatif à la sévérité perceptive, l'axe de gauche aux autres variables (dont l'accuracy)

La corrélation entre sévérité perceptive et accuracy est de $r_s = 0,71$ ($p = 0,009$) : il ne semble donc pas exister d'écart significatif de performance d'accuracy en fonction de la sévérité du trouble de parole (voir Figure 18).

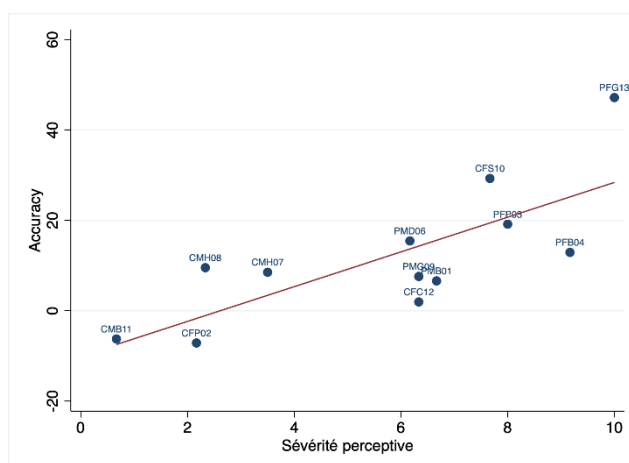


Figure 18 - Graphique de corrélation entre accuracy et sévérité du trouble évaluée de façon perceptive (lot 1 uniquement)

Au vu de ces résultats, il ne semble pas pertinent de mener d'analyse en sous-groupe selon la sévérité du trouble de parole. Les résultats obtenus par notre système sous le paramétrage $\{WIP = 0 ; LMWT = 11\}$ peuvent donc être utilisés pour tous les sujets, indépendamment de leur niveau d'atteinte de parole.

Ainsi, malgré le choix du paramétrage présentant la meilleure accuracy, les performances du système montrent des limites importantes liées à la reconnaissance de la parole pathologique, avec un taux d'erreurs mots très élevé et une accuracy réduite. Toutefois, la corrélation élevée entre sévérité perceptive et accuracy d'une part ($r_s = 0,71$), et sévérité perceptive et précision d'autre part ($r_s = 0,74$) nous conduisent à conserver les sorties du moteur de reconnaissance pour l'ensemble des analyses suivantes.

3.1.5.2 Inventaire lexical et caractéristiques lexicales

Le système a ainsi permis d'établir un inventaire lexical des mots reconnus par sujet (voir Tableau 15). Plusieurs scores ont été mesurés, sous l'hypothèse que les troubles de parole induisent une altération de la longueur et des caractéristiques phonologiques des mots utilisés comme c'est le cas notamment dans l'apraxie de la parole en neurologie (Galluzzi et al., 2015).

Ainsi, un inventaire lexical a été dressé, avec par sujet :

- Le nombre total de mots reconnus (*nword*) ainsi que sa valeur normalisée par seconde ($nwords = \frac{nword}{phontimew}$) ;
- Le nombre total de mots différents reconnus par sujet (*nworddiff*) et sa valeur normalisée par seconde (*nworddiffs*).

Les sorties du système ont aussi permis de calculer le nombre de caractères moyen par mot par sujet (*meanwordlen*) et l'écart-type du nombre de caractères par mot par sujet (*sdwordlen*).

Nous avons également utilisé la base de données en ligne Lexique¹⁸ version 3.83 (New et al., 2001) afin de réaliser des statistiques complémentaires sur les mots reconnus :

- Le nombre moyen de phonèmes par mot (*meannbph*) ;
- Le nombre moyen de syllabes par mot (*meannbsyl*) ;

¹⁸ www.lexique.org, consulté le 18/05/2021

- La distance moyenne de Levenshtein phonologique par mot (*meanpld*) : nombre de transformations phonologiques (insertions, délétions, substitutions) pour arriver d'une chaîne de phonèmes à une autre (Yarkoni et al., 2008), calculée sur les 20 mots phonologiquement les plus proches de la cible ;
- Le nombre moyen d'homophones par mot (*meanhomoph*) ;
- La fréquence moyenne du lemme (unité lexicale de base ayant un signifiant et un signifié unique, autrement dit la forme canonique du mot : par exemple les deux mots *chat* et *chats* correspondent à un seul lemme *chat*, *meanfreqlem*) et du mot à proprement parler (*meanfreq*), par million selon le corpus de films de lexique.org ;
- Nous avons pris le parti de répertorier également les mots reconnus par le système n'étant pas présents dans la base Lexique. Il s'agit principalement de noms communs de lieux ou de personnes, qu'il nous semble intéressant de prendre en compte dans la constitution du discours du sujet (*mis_values*).

Nous avons enfin calculé des scores de structure phonologique, relatifs à l'inclusion d'au moins une structure diconsonantique dans une syllabe par mot :

- Le nombre total de mots reconnus contenant au moins une structure diconsonantique (*wordwcc*), ainsi que la valeur standardisée par seconde ($wordwccs = \frac{wordwcc}{phontimew}$) ;
- La proportion de mots reconnus contenant au moins une structure diconsonantique parmi l'ensemble des mots reconnus ($propwordwcc = \frac{wordwcc}{nword}$).

Tableau 15 - Description statistique des mesures lexicales calculés selon les sorties du système TDNNf-HMM « lexical »

Code	Paramètre	Moyen- ne	Écart -type	Médiane	EIQ	Valeur min	Valeur max
<i>nword</i>	Nombre total de mots reconnus	990,44	512,16	943,00	679,00	128,00	2 552,00
<i>nwords</i>	..., par seconde	1,84	0,51	1,79	0,48	1,02	3,33
<i>nworddiff</i>	Nombre de mots différents reconnus	407,84	180,09	423,00	227,00	25,00	860,00
<i>nworddiffs</i>	..., par seconde	0,77	0,23	0,78	0,26	0,20	1,27
<i>meanwordlen</i>	Nombre de caractère moyen par mot	4,18	0,46	4,26	0,19	2,39	4,82
<i>sdwordlen</i>	Écart-type du nombre de caractère moyen par mot	2,24	0,32	2,31	0,43	1,23	2,64
<i>mis_values</i>	Nombre de mots reconnus et non référencés dans lexique.org	30,96	18,14	28,00	27,00	0,00	78,00

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>meannbph</i>	Nombre moyen de phonèmes par mot	2,77	0,33	2,83	0,25	1,42	3,22
<i>meanbnsyl</i>	Nombre moyen de syllabes par mot	1,33	0,08	1,34	0,08	1,11	1,45
<i>meanpld</i>	Distance de Levenshtein phonologique moyenne par mot	1,08	0,07	1,09	0,07	0,84	1,19
<i>meanhomoph</i>	Nombre d'homophones moyen par mot	5,30	0,57	5,13	0,30	4,82	7,90
<i>meanfreqlem</i>	Fréquence moyenne du lemme, par million	3 577,46	581,90	3 484,37	727,89	2 470,14	4 850,39
<i>meanfreq</i>	Fréquence moyenne du mot, par million	2 539,02	543,63	2 468,11	521,74	1 561,02	4 504,54
<i>wordwcc</i>	Nombre de mots contenant au moins une structure diconsonantique	65,80	53,49	61,00	61,00	1,00	196,00
<i>wordwccs</i>	..., par seconde	0,12	0,08	0,10	0,08	0,01	0,40
<i>propwordwcc</i>	Proportion de mots contenant au moins une structure diconsonantique	0,06	0,03	0,06	0,03	0,01	0,13

3.1.5.3 Scores de confiance

Comme au niveau lexical, des scores de confiance associés aux mots reconnus sont produits par le système et peuvent être étudiés (voir Tableau 16) :

- Le score de confiance moyen par sujet (*meanconf*), et son écart-type (*sdconf*) ;
- Le nombre total de mots reconnus par sujet pondéré par le score de confiance associé à chaque mot ($nwordpond = nword \times meanconf$), ainsi que le nombre pondéré moyen de mots reconnus par seconde par sujet ($nwordponds = \frac{nwordpond}{phontime}$) ;
- Devant l'absence de recommandations et de consensus, un seuil de 0,60 a été arbitrairement choisi comme seuil de « certitude relative » du système quant à la reconnaissance du mot. Ont été ainsi calculés le nombre total de mots reconnus ayant un score de confiance supérieur à 0,60 (*nwordconf*) et la valeur normalisée par seconde (*nwordconfs*), ainsi que la proportion de mots ayant un score de confiance supérieur à 0,60 parmi l'ensemble des mots reconnus (*propnwordconf*).

Tableau 16 - Description statistique des scores de confiances mesurés au niveau lexical selon les sorties du système TDNNf-HMM « lexical »

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>meanconf</i>	Score de confiance moyen	0,67	0,04	0,66	0,04	0,61	0,78
<i>sdconf</i>	Écart-type du score de confiance moyen	0,25	0,01	0,25	0,01	0,22	0,26
<i>nwordpond</i>	Nombre total de mots reconnus pondérés par le score de confiance	665,02	358,43	714,72	457,10	100,45	1 781,60
<i>nwordponds</i>	..., par seconde	1,24	0,40	1,12	0,49	0,71	2,45
<i>nwordconf</i>	Nombre total de mots reconnus avec score de confiance > 0,60	583,00	322,78	582,00	404,00	103,00	1 587,00
<i>nwordconfs</i>	..., par seconde	1,09	0,39	0,97	0,40	0,62	2,29
<i>propnwordconf</i>	Proportion du nombre de mots reconnus avec score de confiance > 0,60 parmi l'ensemble des mots reconnus	0,59	0,07	0,56	0,06	0,49	0,80

3.1.5.4 Diversité et densité lexicales

À partir des sorties du TDNNf « lexical », deux scores complémentaires ont été calculés.

L'étape initiale consiste à lemmatiser les mots avant toute analyse, afin de minimiser les potentiels effets de longueur (de la sortie du TDNNf « lexical »). La seconde étape concerne le calcul de scores de diversité et de densité lexicales.

L'index de Guiraud est une mesure de la diversité lexicale (Guiraud, 1954). Il est calculé au moyen de la formule suivante :

$$\text{Index de Guiraud} = \frac{V}{\sqrt{N}}$$

où V est le nombre de lemmes différents, et N le nombre total de lemmes.

La diversité lexicale (Johansson, 2009) permet d'estimer la quantité d'informations « signifiantes » dans un texte. Pour cela, les mots sont classés soit en « mots pleins » (ou mots ayant un sens plein : noms, verbes, adjectifs et adverbes), soit en « mots outils » (autres catégories de mots : articles, pronoms, conjonctions...). Il se calcule ainsi :

$$\text{Densité lexicale} = \frac{P}{N}$$

où P représente le nom de lemmes ayant un sens plein, et N le nombre total de lemmes.

Les résultats sont présentés dans le Tableau 17.

Tableau 17 - Description statistique des scores « Index de Guiraud » et « Densité lexicale »

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>guiraud</i>	Index de Guiraud	12,43	3,35	13,26	2,68	2,12	17,02
<i>lex_dens</i>	Densité lexicale	0,60	0,06	0,59	0,07	0,45	0,71

Des analyses complémentaires ont été menées sur les données du lot 1, pour s'assurer de la pertinence de calculer ces scores sur le fichier complet plutôt que sur des segments spécifiques.

Concernant la diversité lexicale, nous avons calculé l'index de Guiraud non seulement sur le fichier complet, mais également sur quatre segments de durée égale préalablement partitionnés par sujet, chacun correspondant à une durée d'un quart de la durée globale de l'enregistrement (voir Figure 19).

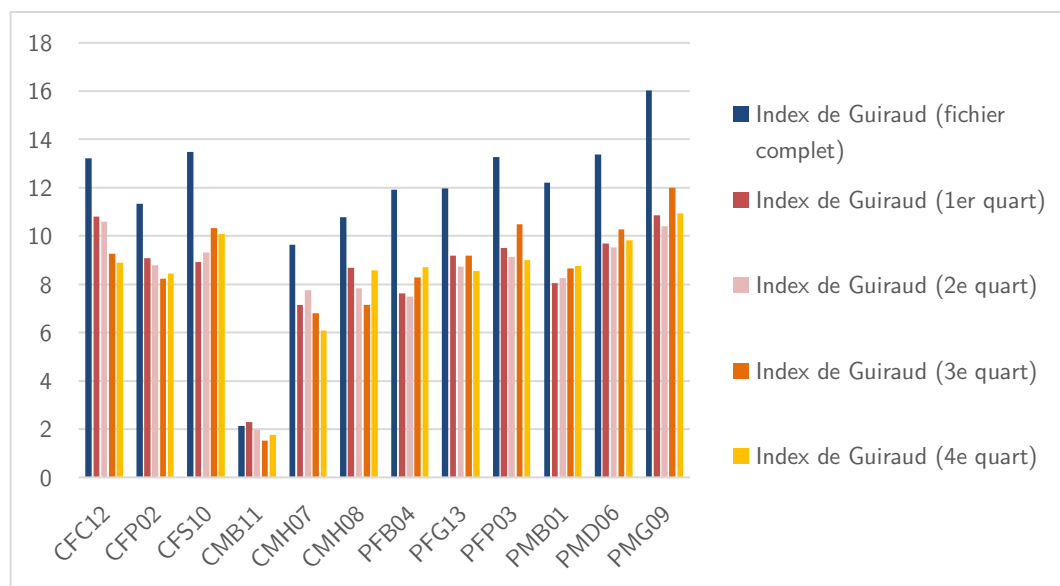


Figure 19 - Comparaison des index de Guiraud, calculés sur les fichiers complets et sur les quarts de fichiers du lot 1

Nous remarquons que la diversité lexicale est plus importante sur le fichier complet que sur chacun des quarts, à l'exception du sujet CMB11. Ce sujet présente en effet un profil particulier : il s'agit du sujet présentant le trouble de parole perçivement jugé comme le plus sévère (0,67/10), et en raison de cette sévérité du trouble, les performances du moteur de reconnaissance sont très faibles. Seuls 141 phonèmes sont reconnus au total tout le long de

l'entretien dont seulement cinq phonèmes différents (110/141 sont reconnus comme des /a/), et au niveau lexical un seul mot est reconnu correct sur les 16 mots cibles attendus.

Les index de Guiraud sur les quarts sont homogènes pour chaque sujet. Pour cette raison, nous conserverons l'index de Guiraud calculé sur l'ensemble du fichier : l'entretien complet est probablement plus riche en thématiques abordées que chacun de ses segments, conduisant à une plus grande diversité lexicale.

Les scores de densité lexicale issus de l'analyse sur la minute transcrite manuellement (perceptivement) et sur la même minute extraite du système automatique sont, là encore, très proches et ne présentent pas de différence significative ($p = 0,27$, test de Wilcoxon). Le score de densité lexicale issu des sorties du TDNNf « lexical » est donc fiable. Une analyse similaire entre score issu de la minute « automatique » et de l'ensemble du fichier ne montre pas non plus de différence significative ($p = 0,12$, test de Wilcoxon). Ainsi, à l'instar de l'index de Guiraud, nous retenons le score de densité lexicale calculé sur l'ensemble du fichier (voir Figure 20).

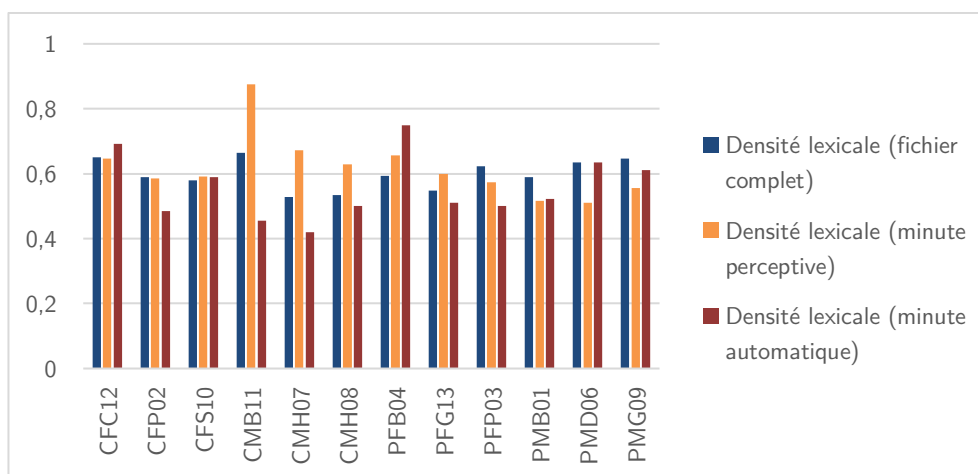


Figure 20 - Comparaison des scores de densité lexicale pour le lot 1, calculés sur le fichier complet ou sur une minute (à partir d'une transcription perceptive ou d'une sortie automatique)

3.1.6 Composante morpho-syntaxique – Composante grammaticale

Nous avons ensuite procédé à une analyse grammaticale des sorties du TDNNf « lexical », toujours au moyen des ressources répertoriées dans la base de données Lexique (voir Tableau 18). Un discours contenant des classes

grammaticales variées pourrait en effet témoigner d'un discours plus riche, varié, et donc d'une plus grande aisance à la production orale.

Nous avons d'abord référencé les classes grammaticales de chaque mot. Toutefois, certains mots peuvent être rattachés à plusieurs classes : c'est par exemple le cas du mot « *en* » pouvant être un adverbe, une préposition ou un pronom selon le contexte syntaxique.

Ainsi, un premier score concerne la somme de l'ensemble des classes grammaticales par sujet (*totalcgram*), valeur qui pourra être supérieure au nombre total de mots reconnus par le système en raison de la possibilité de rattachement de certains mots à plusieurs classes grammaticales.

Nous avons aussi calculé le rapport du nombre de classes grammaticales sur le nombre de mots reconnus ($cgramwords = \frac{totalcgram}{nword}$) : plus le sujet utilise des mots pouvant appartenir à plusieurs classes grammaticales, plus le rapport sera élevé.

Nous avons ensuite calculé pour chaque classe grammaticale (noms, adverbes, auxiliaires, conjonctions, liaisons euphoniques, onomatopées, prépositions, verbes, adjectifs, articles et pronoms) le nombre total de mots qui y sont rattachés (code : *nom_classe_grammaticale*), sa valeur standardisée (nombre de mots reconnus appartenant à cette classe grammaticale par seconde : $[nom_classe_grammaticale]s = \frac{nom_classe_grammaticale}{phonetimew}$), ainsi que la proportion de mots reconnus appartenant à cette classe grammaticale sur le nombre total de classes grammaticales reconnus ($prop[nom_classe_grammaticale] = \frac{nom_classe_grammaticale}{totalcgram}$).

Tableau 18 - Description statistique des scores issus des analyses par classes grammaticales

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>totalcgram</i>	Nombre total de classes grammaticales reconnues	1 831,84	941,07	1 797,00	1 270,00	180,00	4 709,00
<i>cgramwords</i>	Rapport classes grammaticales reconnues par nombre total de mots	1,85	0,15	1,86	0,16	1,41	2,24
<i>nom</i>	Nombre total de noms	504,32	272,98	497,00	337,00	22,00	1 288,00
<i>noms</i>	..., par seconde	0,92	0,31	0,94	0,34	0,18	1,76
<i>propnom</i>	Proportion de noms	0,27	0,04	0,28	0,03	0,12	0,31
<i>adv</i>	Nombre total d'adverbes	159,72	91,20	178,00	131,00	10,00	406,00
<i>advs</i>	..., par seconde	0,29	0,12	0,30	0,19	0,08	0,63
<i>propadv</i>	Proportion d'adverbes	0,08	0,02	0,09	0,04	0,04	0,13
<i>aux</i>	Nombre total d'auxiliaires	72,44	42,36	67,00	62,00	4,00	187,00

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>auxs</i>	..., par seconde	0,13	0,06	0,13	0,08	0,03	0,25
<i>propaux</i>	Proportion d'auxiliaires	0,04	0,01	0,04	0,01	0,02	0,09
<i>conj</i>	Nombre total de conjonctions	44,44	29,49	40,00	24,00	1,00	144,00
<i>conjs</i>	..., par seconde	0,08	0,04	0,08	0,04	0,01	0,16
<i>propcon</i>	Proportion de conjonctions	0,02	0,01	0,02	0,01	0,01	0,04
<i>lia</i>	Nombre total de liaisons euphoniques	0,00	0,00	0,00	0,00	0,00	0,00
<i>proplia</i>	Proportion de liaisons euphoniques	0,00	0,00	0,00	0,00	0,00	0,00
<i>ono</i>	Nombre total d'onomatopées	6,84	11,82	4,00	3,00	0,00	63,00
<i>onos</i>	..., par seconde	0,03	0,10	0,01	0,01	0,00	0,50
<i>propono</i>	Proportion d'onomatopées	0,02	0,07	0,00	0,00	0,00	0,35
<i>prep</i>	Nombre total de prépositions	85,52	57,60	76,00	54,00	11,00	259,00
<i>preps</i>	..., par seconde	0,17	0,09	0,14	0,08	0,03	0,47
<i>proppre</i>	Proportion de prépositions	0,05	0,03	0,05	0,01	0,01	0,16
<i>verb</i>	Nombre total de verbes	263,44	144,62	260,00	208,00	9,00	619,00
<i>verbs</i>	..., par seconde	0,47	0,15	0,46	0,19	0,07	0,83
<i>propver</i>	Proportion de verbes	0,14	0,03	0,14	0,03	0,05	0,19
<i>adj</i>	Nombre total d'adjectifs	280,20	145,80	268,00	205,00	14,00	734,00
<i>adjs</i>	..., par seconde	0,52	0,17	0,52	0,24	0,11	0,96
<i>propadj</i>	Proportion d'adjectifs	0,15	0,02	0,15	0,03	0,08	0,19
<i>art</i>	Nombre total d'articles	142,52	74,22	137,00	85,00	12,00	370,00
<i>arts</i>	..., par seconde	0,27	0,10	0,28	0,11	0,10	0,56
<i>propart</i>	Proportion d'articles	0,08	0,02	0,08	0,02	0,04	0,12
<i>pron</i>	Nombre total de pronoms	272,40	134,09	284,00	158,00	17,00	686,00
<i>prons</i>	..., par seconde	0,51	0,16	0,49	0,15	0,14	0,93
<i>proppro</i>	Proportion de pronoms	0,15	0,02	0,15	0,02	0,09	0,20

Parmi les outils existants, nous avons ensuite choisi d'utiliser le logiciel libre Iramuteq¹⁹ (interface du logiciel R pour les analyses multidimensionnelles de textes et de questionnaires, s'appuyant sur le langage Python). Iramuteq est couramment utilisé en France pour les analyses de contenu de corpus textuels et les analyses lexico-grammaticales (au moyen de procédés de lemmatisation). Les sorties du TDNNf « lexical » sont fournies en entrée à Iramuteq qui procède à une lemmatisation des suites de mots par sujet. Puis les lemmes sont classés en « formes actives » correspondant aux formes pleines des mots (adjectifs, adverbes, noms communs, verbes, formes non reconnues), ou en

¹⁹ <http://www.iramuteq.org/>, consulté le 18/05/2021

« formes supplémentaires » correspondant aux mots outils (autres : articles, auxiliaires, conjonctions, pronoms...).

Nous avons ainsi pu calculer le nombre total de formes actives reconnues (*ira_formact*, ainsi que la valeur standardisée par seconde : $ira_acts = \frac{ira_formact}{phontimew}$), le nombre total de forme supplémentaires reconnues et sa valeur standardisée (*ira_formsup* ; *ira_sups*), et enfin la proportion de formes actives parmi l'ensemble des lemmes ($propactsup = \frac{ira_formact}{(ira_formact + ira_formsup)}$). L'ensemble des résultats se trouve Tableau 19.

Tableau 19 - Description statistique des scores issus des analyses par formes actives et supplémentaires (Iramuteq)

Code	Paramètre	Moyen- ne	Écart- type	Mé- diane	EIQ	Valeur min	Valeur max
<i>ira_formact</i>	Nombre total de formes actives	260,12	126,64	261,00	175,00	11,00	581,00
<i>ira_acts</i>	..., par seconde	0,48	0,15	0,48	0,16	0,09	0,86
<i>ira_formsup</i>	Nombre total de formes supplémentaires	97,40	32,68	95,00	36,00	13,00	163,00
<i>ira_sups</i>	..., par seconde	0,20	0,07	0,18	0,09	0,09	0,35
<i>propactsup</i>	Proportion de formes actives parmi l'ensemble des lemmes reconnus	0,70	0,08	0,71	0,09	0,46	0,79

3.1.7 Composante pragma-sémantique et discursive

Un dernier axe d'analyse a concerné l'étude du niveau conceptuel du modèle psycholinguistique, sous deux aspects : la variété thématique du discours et le sentiment se dégageant du discours.

3.1.7.1 Analyse thématique

Nous avons mené une analyse en Classification hiérarchique descendante (Reinert, 1990) afin d'étudier la variété thématique. Le principe de cette analyse, qui se base aussi sur les sorties du TDNNf « lexical », est de partitionner itérativement la suite de mots reconnus pour chaque sujet en deux parties en maximisant à chaque partition un certain critère. Les mots apparaissant régulièrement dans le même contexte lexical proche seront ainsi regroupés.

Le logiciel Iramuteq a également été choisi pour cette analyse thématique.

Le nombre de classes issues de cette classification hiérarchique descendante est reportée dans le Tableau 20.

Tableau 20 Description statistique des classes thématiques retrouvées par le logiciel Iramuteq

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>ira_nbclass</i>	Nombre de classes thématiques par sujet	5,56	1,20	6,00	1,00	3,00	8,00

3.1.7.2 Analyse de sentiment

Nous avons alors mené une dernière analyse concernant le sentiment global qui se dégage des sorties du TDNNf « lexical ».

La base Lexique fournit également des données concernant la valence émotionnelle et l'éveil (ou « arousal »)²⁰ des mots contenus d'un discours. Malheureusement, ces données n'ont pas pu être utilisées dans le cadre de notre étude : les listes disponibles sont de taille modeste (1 286 mots inclus dans les bases de Lexique) et seuls 152 mots ou 263 lemmes reconnus par le TDNNf « lexical » sont référencés (soit respectivement 1,28 % et 2,20 %). Ces valeurs sont trop faibles pour pouvoir être analysées.

Peu d'outils sont disponibles gratuitement en français pour ce type d'analyse. Nous nous sommes orientés vers l'API (Application Programming Interface) cognitive Microsoft Azure²¹ : elle est facile d'accès (aucune expertise n'est requise en machine learning pour faire fonctionner l'algorithme) et elle s'appuie sur des algorithmes développés à grande échelle sur 15 langues. Ces services directement accessibles en ligne comprennent un module Analyse de texte dont l'analyse de sentiment fait partie. Cette API permet d'aboutir à une valeur comprise entre 0 (sentiment négatif) et 1 (sentiment positif), une valeur à 0,5 correspondant à un sentiment neutre (énoncé factuel).

Initialement développée pour étudier le sentiment se dégageant d'avis en ligne, cette API a été développée pour analyser des corpus de 5 120 caractères maximums. Or, 12 de nos 25 sujets ont un nombre de sorties du TDNNf « lexical » supérieur à ce seuil. Nous avons ainsi calculé une valeur globale de sentiment en faisant une moyenne pondérée des scores de sentiments de chaque segment de 5 120 caractères maximum par le nombre de caractères de chaque

²⁰ http://www.lexique.org/?page_id=492, consulté le 18/05/2021

²¹ <https://azure.microsoft.com/fr-fr/services/cognitive-services/text-analytics/>, consulté le 18/05/2021

extrait. La segmentation s'est faite manuellement, de manière à ne pas couper un mot en deux (voir Tableau 21).

Tableau 21 - Valeurs des scores de sentiment par segment de 5 120 caractères

Sujet	Nombre de caractères total	Segment 1		Segment 2		Segment 3		Sentiment global pondéré
		Nombre de caractères	Senti-ment	Nombre de caractères	Senti-ment	Nombre de caractères	Senti-ment	
PMD06	5 703	5 120	0,490	583	0,410			0,482
PMG09	7 071	5 118	0,511	1 953	0,529			0,516
CFS10	5 894	5 119	0,490	775	0,432			0,483
CFC12	7 557	5 120	0,490	2 437	0,480			0,487
CFW16	8 389	5 117	0,490	3 272	0,484			0,488
PMB18	8 552	5 120	0,489	3 432	0,483			0,487
CMS19	5 738	5 120	0,490	618	0,557			0,497
PMS22	5 803	5 115	0,489	688	0,429			0,482
PMD23	12 817	5 119	0,490	5 118	0,490	2580	0,480	0,488
CFR24	7 153	5 116	0,489	2 037	0,473			0,484
CFN26	5 864	5 117	0,510	747	0,569			0,517
CMV27	6 262	5 117	0,490	1 145	0,455			0,483

Nous remarquons ainsi des différences de valeurs de sentiment chez certains sujets selon la longueur du segment. L'écart est notamment supérieur à 10 % entre le sentiment calculé sur le segment 2 et celui calculé sur le segment 1 (de 5120 caractères ou près) pour quatre (PMD06, CFS10, CMS19, CFN26) des cinq sujets (+ PMS22) dont la longueur du segment 2 est inférieure à 20 % de la longueur du segment 1.

Nous nous sommes alors demandé si la longueur du segment d'étude n'aurait pas tendance à neutraliser la polarité du sentiment.

Pour cela, nous avons réalisé un pré-test sur certains sujets du lot 1.

Nous avons cherché à mesurer le sentiment sur des segments qui seraient partitionnés au moyen de mots « séparateurs » intervenant régulièrement au cours du discours, et qui auraient un sens en termes de segmentation d'énoncés. Les mots reconnus par le TDNNf « lexical » et ayant au moins dix occurrences pour au moins un des sujets ont été représentés Figure 21.

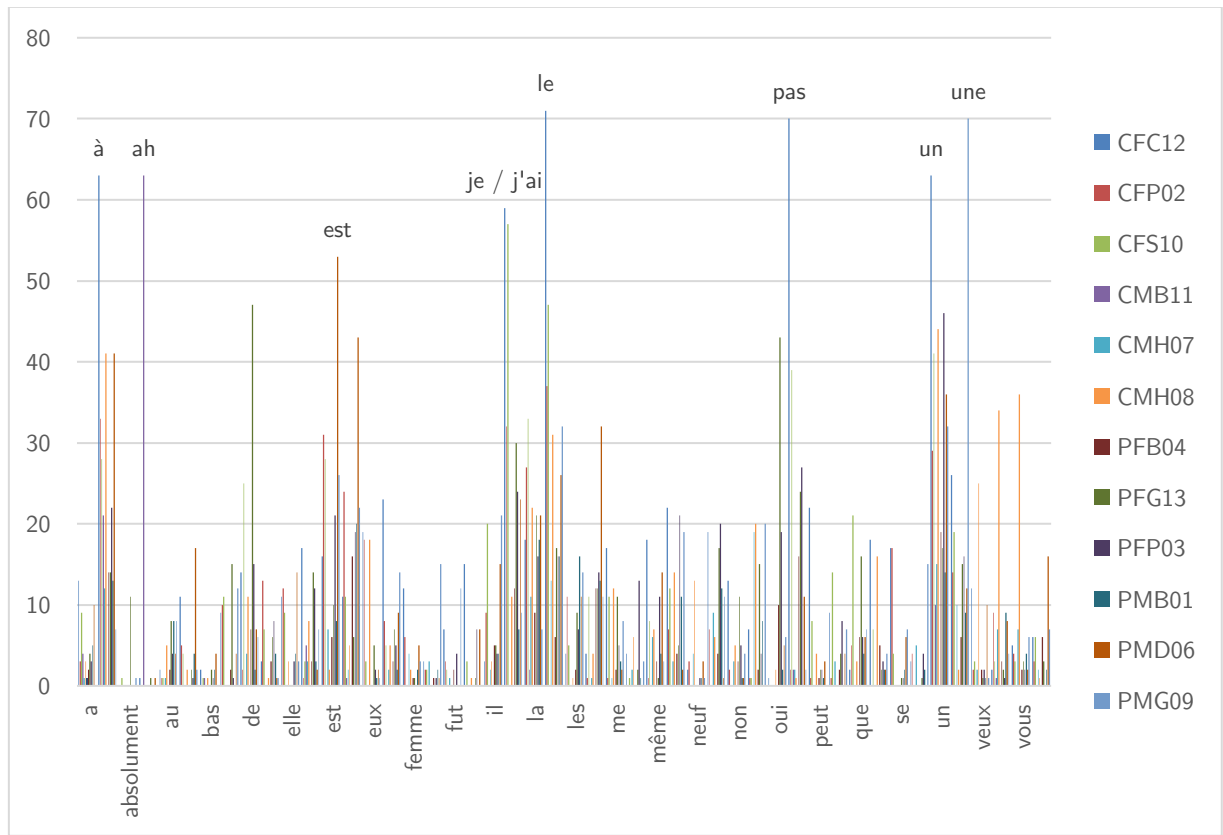


Figure 21 - Représentation des mots ayant au moins dix occurrences chez au moins un des sujets (les mots ayant au moins 50 occurrences ont été étiquetés)

Nous avons alors choisi de segmenter les corpus en utilisant comme séparateur « *je* » (ou « *j'* »), en raison de sa fréquence d'occurrence, et du fait qu'il soit fréquemment utilisé en début d'énoncé.

Pour sept des huit sujets du lot 1 ayant moins de 35 occurrences du séparateur « *je / j'* », ont été calculés :

- La moyenne brute (non pondérée) des scores de sentiments sur les segments définis ;
- La moyenne des scores de sentiment sur les segments nouvellement déterminés, pondérée par le nombre de caractères de chaque segment.

Un test de Wilcoxon a ensuite été réalisé : il ne montre pas de différence significative entre la moyenne non pondérée ($p = 0,29$) ou la moyenne pondérée ($p = 0,45$) issue de la segmentation, et le score global de sentiment que nous avons défini plus tôt.

La longueur du segment sur lequel le score de sentiment est calculé ne semble donc pas modifier de façon significative le score de sentiment. Malgré la

faible puissance de l'analyse statistique précédente, nous choisissons de conserver le score global de sentiment pour la suite des analyses (voir Tableau 22).

Tableau 22 Description statistique des scores de sentiment (Microsoft Azure)

Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
<i>sentiment</i>	Score de sentiment	0,50	0,02	0,49	0,03	0,48	0,58

3.1.8 Synthèse de l'extraction des paramètres

Une synthèse reprenant la méthodologie de recherche et d'extraction des paramètres à partir du signal de parole spontanée est présentée dans la Figure 22.

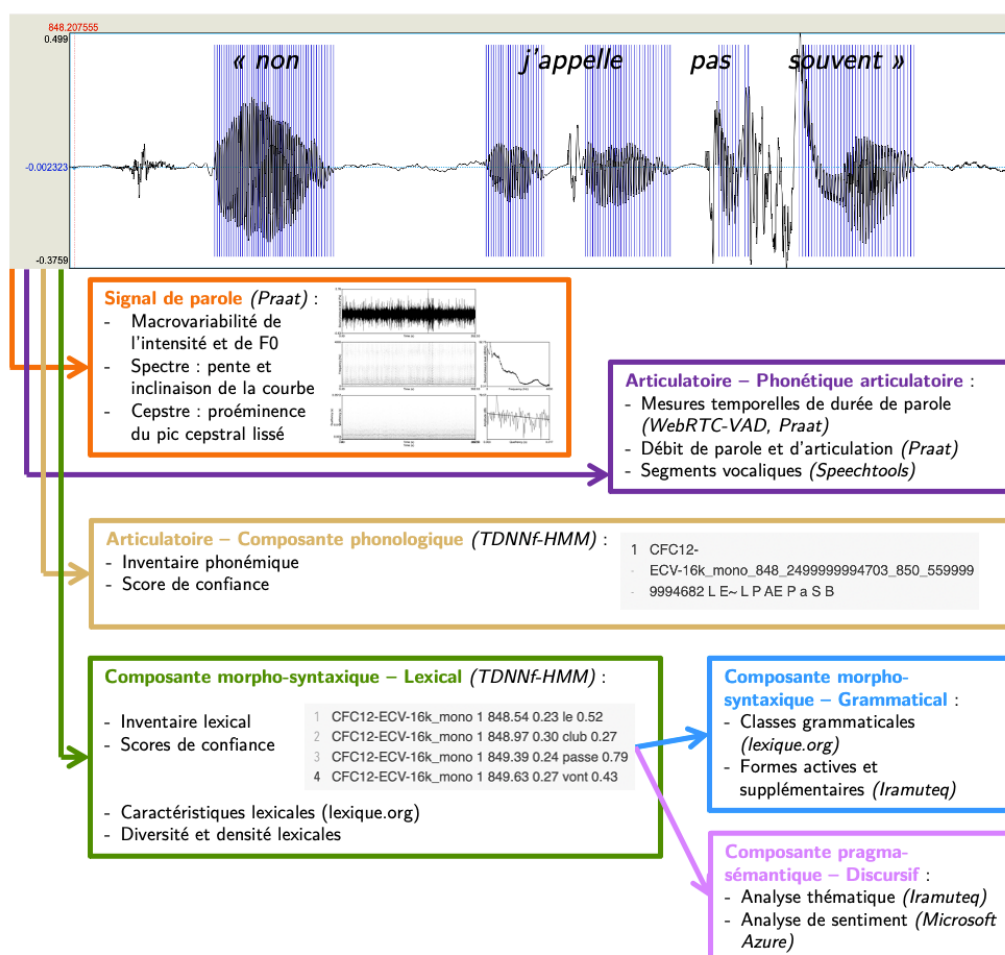


Figure 22 - Schéma synthétique de la méthodologie d'extraction des paramètres après analyses automatiques du signal de parole spontanée

À partir du signal de parole, les paramètres de macrovariabilité de l'intensité et de la fréquence fondamentale, les paramètres spectraux et cepstraux ont été extraits. Nous avons également pu obtenir des mesures de phonétique articulatoire, au niveau temporel (durée de parole, débit de parole et d'articulation), mais aussi des segments vocaliques. L'utilisation d'un système de reconnaissance de la parole a permis : d'une part, l'obtention de sorties phonémiques associées à des scores de confiance ; d'autre part, l'obtention de sorties lexicales là encore associées à des scores de confiance. L'analyse des niveaux morpho-syntaxique (grammatical) et pragma-sémantique (discursif) s'est faite sur la base des sorties lexicales, pour obtenir des mesures concernant les classes grammaticales, les formes actives et supplémentaires, mais aussi les thématiques ou le sentiment se dégageant des sorties lexicales.

3.2 Sélection des paramètres

À l'issue de ce processus d'extraction mené sur la parole spontanée, nous avons recueilli un total de 149 paramètres.

Parmi ces paramètres, nous avons conduit une première sélection sur deux grands axes : éliminer les paramètres les moins pertinents, et ceux redondants avec d'autres. Nous visons ici à réduire le nombre final de paramètres pouvant être utilisés pour répondre à notre objectif principal : prédire l'impact du trouble de parole sur les capacités de communication au moyen de ces paramètres issus d'analyses automatiques de la parole spontanée.

Le processus global de sélection présenté Figure 23 est détaillé dans la suite de cette section, ainsi qu'en **Annexe VI**.

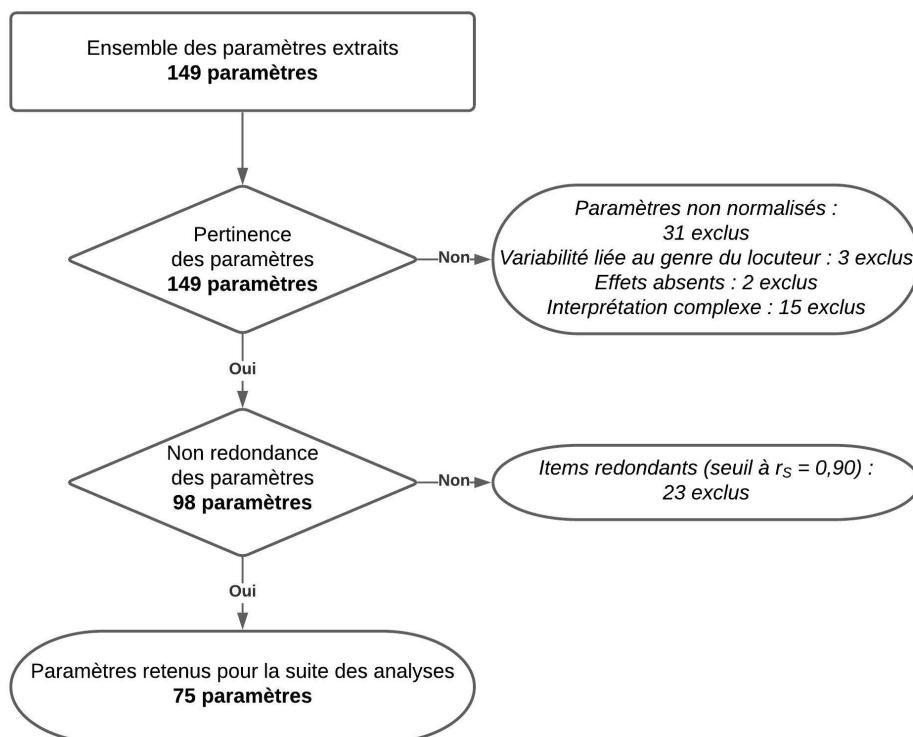


Figure 23 - Processus de sélection des paramètres issus des analyses automatiques de la parole spontanée

3.2.1.1 Sélection des paramètres pertinents

Parmi tous les paramètres précédemment extraits, un choix méthodologique est à mener pour retenir ceux étant pertinents. Un paramètre est considéré comme pertinent s'il respecte trois principes de validité :

- S'il permet une mesure d'un effet présent dans notre échantillon,
- Si la mesure présente un risque de biais limité,
- Et enfin si sa mesure est d'interprétation accessible (si cette mesure a du sens, critère évalué par un consensus de deux experts en traitement du signal m'ayant accompagné dans cette tâche).

Deux paramètres ici mesurent des effets absents de notre échantillon : le nombre total de liaisons euphoniques (*lia*) et la proportion de liaisons euphoniques (*proplia*) issus du système de reconnaissance lexical. En effet, aucune liaison euphonique (par exemple : « l » dans « l'on dit que... », ou « t » dans « va-t-on... ») n'a été reconnue par le système. Ces deux paramètres ont donc été exclus de la suite des analyses.

Trente-et-un paramètres non normalisés ont également été exclus, afin d'éviter un effet de longueur du discours pouvant biaiser la mesure cible. Toutefois, les 31 valeurs normalisées correspondant à ces 31 paramètres non normalisés ont quant à elles bien été conservées.

Au niveau du signal de parole, trois mesures liées à la fréquence fondamentale peuvent être dépendantes du genre du locuteur : le maximum ($f0max$), le minimum ($f0min$) et l'étendue ($f0range$). Dans la mesure où nous avons mesuré la fréquence fondamentale par rapport au rôle que joue sa variabilité dans la prosodie, les deux paramètres maximum et minimum ont ainsi été exclus n'ayant pas d'intérêt spécifique d'interprétation. Cette mesure de variabilité est donnée par la mesure de son étendue ($f0range$) mais aussi par le centième du demi-ton tempéré ($f0cent$), ces deux paramètres étant fortement corrélés entre eux ($r_s = 0,94$). Nous avons ainsi cherché à comparer ces deux mesures en procédant à une analyse stratifiée de chacun de ces paramètres chez les hommes et chez les femmes, via un test de Mann-Whitney. Une différence significative de l'étendue de la fréquence fondamentale ($p = 0,02$) existe selon le genre du locuteur, ce qui n'est pas le cas pour le centième du demi-ton tempéré ($p = 0,07$). Ces résultats nous conduisent donc à préférer ce dernier paramètre, et à exclure l'étendue de la fréquence fondamentale de la suite des analyses.

Au niveau des mesures issues de la détection des segments vocaliques et des pauses, nous avons choisi d'exclure de la suite des analyses les paramètres relatifs aux pseudo-consonnes. Elles sont construites « par exclusion », c'est-à-dire des segments qui ne sont pas vocaliques mais qui ne sont pas des pauses, ce qui rend leur interprétation particulièrement difficile. Les pseudo-plosives ont également été exclues, car leur définition (pauses d'une durée inférieure à 150 ms) ne permet pas de s'assurer de la mesure d'une « vraie » plosive. Cela conduit à ne pas conserver 13 paramètres supplémentaires pour la suite des analyses.

La durée totale des silences ($silp$) n'a pas été conservée, car sa mesure est étroitement liée à deux autres paramètres dont elle est la différence ($silp = ndurp - phontimep$).

Un dernier paramètre n'a pas été retenu : le nombre total de catégories grammaticales rattachées aux mots reconnus ($totalcgram$). Ce paramètre a permis de calculer la proportion des différences classes grammaticales, mais n'a pas d'interprétation directe et évidente en lui-même.

Finalement, 51 paramètres ont été exclus à cette étape. Il reste ainsi 98 paramètres (66,4 %) respectant le critère de pertinence.

3.2.1.2 Sélection des items non redondants

Une étude statistique a permis d'éliminer les items trop redondants. Pour cela, nous avons réalisé une matrice de corrélation inter-items de Spearman. Nous avons établi un seuil à 0,90 : aucun des paramètres ne doit être corrélé à 0,90 ou plus avec un autre paramètre finalement retenu.

Là encore, le comité composé de deux universitaires experts en informatique et traitement du signal m'a assisté pour choisir les items à conserver. Nous avons procédé à une démarche de sélection des paramètres inter-corrélés par bloc. Tous les paramètres inter-corrélés à 0,90 ou plus ont été inclus dans un bloc et un seul des paramètres de ce bloc a été conservé.

Le choix s'est opéré sur un principe de parcimonie : ont été gardés prioritairement les paramètres se retrouvant dans le plus grand nombre de blocs. Par exemple, la durée globale de production du sujet entre le début de la première prise de parole et la fin de la dernière prise de parole, incluant les pauses (*ndurw*) a été conservée car elle fait partie d'un bloc inter-corrélé avec *duration* et *ndurp* (qui, de fait, ont été exclus), et d'un autre bloc avec *duration* et *sumdurv* (exclus également – ces deux blocs sont dissociés car la corrélation entre *ndurp* et *sumdurv* est de $r_s = 0,89$).

Pour les autres blocs, le choix des experts s'est fait sur le paramètre le plus aisément interprétable.

À la suite de ce processus, 23 paramètres supplémentaires n'ont pas été conservés car ils sont considérés comme trop redondants.

Le détail des motifs d'exclusion des 74 paramètres est donné en **Annexe VI**.

3.2.1.3 Paramètres retenus

Soixante-quinze paramètres issus d'analyses automatiques de la parole sont retenus finalement pour la suite des analyses (**Annexe VII**).

Parmi ces 75 paramètres, seuls cinq (6,7 %) présentent des scores significativement différents entre le lot 1 et le lot 2 (test de Mann-Whitney) : le débit d'articulation (*artratep*, $p = 0,01$), la proportion de pseudo-silences parmi les pseudo-phonèmes (*props_cvps*, $p = 0,03$), le taux d'occlusives parmi les consonnes reconnues (*tx_occcsn*, $p = 0,03$), le score moyen de confiance calculé sur les semi-consonnes (*confs*, $p = 0,03$) et enfin le nombre moyen de phonèmes par mot (*meannbph*, $p = 0,02$).

Les paramètres retenus sont donc très majoritairement comparables entre les lots.

3.3 Conclusion

Dans cette partie, nous avons recherché les outils disponibles permettant une analyse automatique dans le cadre de l'étude de la parole spontanée après traitement d'un cancer de la cavité buccale ou de l'oropharynx. Pour cela, nous nous sommes appuyés sur le modèle psycholinguistique de Caron, afin de déterminer les différents niveaux linguistiques à cibler. Nous avons extrait des paramètres de macrovariabilité de la fréquence fondamentale et de l'intensité, ainsi que des paramètres spectraux et cepstraux. Des mesures temporelles concernant la durée de parole, le débit de parole et d'articulation, les segments vocaliques ont été recueillies au moyen de différents outils (scripts Praat, VAD, SpeechTools). L'utilisation d'un système de reconnaissance automatique de la parole de type TDNN factorisé – HMM nous a permis de dresser un inventaire des productions des sujets en parole spontanée avec les scores de confiance associés, tant au niveau phonémique qu'au niveau lexical. Une étape préalable d'optimisation des paramétrages du système a été nécessaire. Nous avons utilisé des scripts permettant une mesure de la densité et de la diversité lexicales. Enfin, l'utilisation des sorties du TDNNf-HMM lexical nous a permis de réaliser des statistiques sur la structure grammaticale des mots reconnus, le nombre de thématiques du discours et le sentiment se dégageant de ces sorties.

Finalement, 149 paramètres ont été collectés. Une méthodologie de sélection de ces paramètres a ensuite été appliquée afin de ne conserver que ceux pertinents selon un consensus entre experts (exclusion de paramètres non normalisés, variables quant au genre du locuteur, mesurant des effets absents ou d'interprétation complexe). Les 98 paramètres restants devant ne pas être redondants, 23 supplémentaires ont été exclus car ils étaient corrélés à plus de 0,90 avec un ou plusieurs autres paramètres déjà retenus. Finalement, 75 paramètres issus d'analyses automatiques ont été retenus.

Le caractère novateur de cette étude est lié à l'exploration de la parole spontanée en contexte cancérologique au moyen d'outils acoustiques et automatiques. La revue systématique de littérature que nous avons menée (voir chapitre 1) n'avait d'ailleurs pas retrouvé d'étude sur ce type de tâche de production.

Les 75 paramètres que nous avons retenus vont ainsi permettre de répondre à notre objectif de prédiction de l'altération de la communication au moyen de paramètres issus d'analyses automatiques de la parole. Les techniques et méthodes utilisées seront ainsi présentées dans la suite de ce manuscrit (voir chapitre 4).

4 Prédiction de l'altération de la communication

Le corpus que nous avons constitué a permis d'extraire 75 paramètres de l'analyse automatique des enregistrements de parole spontanée. Il comprend également de nombreuses informations quant à la perception par les sujets de leurs capacités de communication, de leur qualité de vie et des facteurs associés à ces deux domaines et au trouble de parole.

L'objectif de cette partie est de prédire l'altération de la communication ressentie par les sujets en fonction des paramètres automatiques du signal de parole.

Dans une première partie, nous construirons un score de référence mesurant les capacités de communication : le score holistique de communication. Dans une seconde partie, nous prédirons ce score au moyen d'une modélisation par les paramètres automatiques seuls, puis en ajustant notre modélisation avec les facteurs associés aux troubles de parole et à la communication.

4.1 Construction du score holistique de communication

Les troubles de parole sont une plainte fréquente des patients traités pour un cancer de la cavité buccale et de l'oropharynx. Or, aucun outil ne permet à ce jour de quantifier de façon globale l'impact du trouble de parole sur la communication quotidienne des sujets traités pour un cancer de la cavité buccale ou de l'oropharynx.

Certains évaluent les limitations d'activités et les restrictions de participation, comme le PHI (Balaguer et al., 2020b) ou le domaine « Phonation » ou « Psychosocial » du CHI (Balaguer et al., 2021a). D'autres items relatifs à la communication, au contact social ou à la parole sont présents dans les deux modules de l'EORTC QLQ (Aronson et al., 1993; Bjordal et al., 1999).

D'autres questionnaires évaluent la dynamique communicationnelle mais ne sont pas validés en cancérologie ORL : c'est le cas de l'ECVB (Mazaux et al.,

2006) et du DIP (Letanneux et al., 2013) ; ou ne sont pas validés en français comme le CPIB (Baylor et al., 2009).

Ces outils aboutissent à des scores par items, regroupés en scores globaux. En raison de leur construction par addition ou moyenne, ces scores et sous-scores dimensionnels sont finalement peu représentatifs des limitations de communication réellement perçues par les patients. De plus, tous ces questionnaires ciblent différents aspects de la communication, mais aucun outil ne permet d'obtenir un score global, holistique, représentatif de l'impact sur la communication des troubles de parole en cancérologie ORL.

Notre objectif est ici de construire un score holistique de communication représentant au mieux l'impact fonctionnel sur la communication des troubles de parole chez les patients traités pour un cancer de la cavité buccale ou de l'oropharynx. Pour cela, nous avons adopté une méthodologie proche de celle utilisée pour la sélection des paramètres issus des analyses du signal de parole.

4.1.1 Présentation des items

Pour construire ce score, une première étape a été de collecter les items relatifs à la parole, à la communication et à la qualité de vie globale et relative à la santé parmi l'ensemble des questionnaires recueillis.

Un total de 174 items a été initialement collecté. Le détail des motifs d'inclusion et d'exclusion des 174 paramètres (que nous allons développer dans la suite de ce manuscrit) est donné en **Annexe VIII**.

Deux questionnaires relatifs à la dynamique communicationnelle ont été intégralement conservés en raison de leur richesse et de leur proximité conceptuelle avec notre objectif de mesure de l'altération des capacités communicationnelles : l'ECVB (34 items) et le DIP (49 items).

Pour rappel (voir section 2.5.1), deux réponses par item à l'ECVB ont été recueillies. Une version « patient » a été complétée par le patient seul avant l'entretien semi-dirigé, de façon à obtenir le point de vue de patient sans intermédiaire et pour enrichir le contenu de la discussion en préparant le sujet aux grandes thématiques abordées. Une version « examinateur » a été complétée par l'orthophoniste au cours de l'entretien, sur la base des informations apportées par le patient. L'orthophoniste ne prenait pas connaissance au préalable des réponses de la version « patient ».

Une étude des résultats entre ces deux versions a été conduite, afin de mener au choix de la version de l'ECVB conservée pour la suite des analyses. Un test de Wilcoxon en série appariée (ECVB « examinateur » *vs.* ECVB

« patient ») a été réalisé pour chacun des 34 items, pour les sept dimensions et pour le score global final.

Finalement, un seul item présente une différence significative (mais très proche du seuil de significativité fixé) entre les deux versions ($p = 0,046$). Il s'agit de l'item *ecvb5* « Avec un membre de votre famille ou un de vos amis, avez-vous des difficultés pour avoir une conversation sur un sujet plus abstrait ou plus compliqué, par exemple : l'actualité, l'éducation, le chômage, la littérature... ? ». Sur cet item, les scores de l'ECVB « patient » sont en moyenne légèrement inférieurs à ceux de l'ECVB « examinateur » (ECVB « sujet » : $m = 1,8$ (écart-type = 0,913), médiane = 2 ; ECVB « examinateur » : $m = 2$ (écart-type = 0,957), médiane = 2).

Un ICC a également été calculé entre les scores « patient » et « examinateur ». Pour sept scores (16,7 %), les résultats sont parfaitement identiques entre les deux versions, et l'ICC n'a de fait pas pu être calculé. L'ICC est supérieur ou égal à 0,95 dans 21 scores (50 %). Les ICC sont compris entre 0,85 et 0,95 dans tous les autres cas (14/42, soit 33,3 %).

Les résultats montrent ainsi une équivalence des deux versions « patient » et « examinateur ». Ainsi, afin de permettre au patient de prendre le temps de répondre aux questions et de préparer l'entretien en l'absence de trouble cognitif avéré, il semble intéressant de proposer en cancérologie ORL au patient de compléter lui-même l'ECVB au préalable, et que l'ECVB soit utilisé comme un outil PRO (Patients Reported Outcomes). Dans cette étude, seuls les scores issus de la version « patient » seront ainsi utilisés.

La version originale du DIP, validée en neurologie après de patients atteints de la maladie de Parkinson, contient 48 items. Pour cette étude, nous avons ajouté un item à la section C « Comment je perçois la réaction des autres face à ma parole », afin d'éclaircir une formulation absconse. L'item *dipc6* « Les personnes qui ne me connaissent pas ne me considèrent pas négativement malgré mon problème de parole », a été conservé malgré sa formulation peu explicite, mais a été complété par un nouvel item évaluant le même concept mais avec une formulation différente : l'item *dipc15* « Les inconnus n'ont pas d'a priori négatif sur moi malgré mon problème de parole ».

Nous avons enfin également conservé quelques items d'autres questionnaires liés à notre objectif : l'ensemble des 18 items du PHI évaluant le handicap de parole perçu, huit items du CHI (les quatre du domaine « Phonation » et les quatre du domaine « Psychosocial ») ainsi que les 30 items de l'EORTC QLQ-C30 et les 35 de l'EORTC QLQ-H&N35 mesurant la qualité de vie globale et relative à la parole.

Au niveau des analyses, un item a nécessité un traitement spécifique de gestion de données manquantes. Il s'agit du premier item du domaine « Phonation » du CHI (*chiph1*) : « *Avez-vous des difficultés pour parler ?* ». Pour un sujet (PMB01), seul le score global du domaine « Phonation » a été retrouvé ($chiph = 15$) mais pas le détail de la ventilation de ce sous-score dans les quatre items qui composent ce domaine. Le sujet PMD06 n'a, quant à lui, répondu à aucune question de ce domaine (données MNAR). L'absence d'imputation pour ces deux sujets de l'item *chiph1* aurait conduit à exclure ces deux sujets des analyses ultérieures en données complètes. Cette solution n'a pas été retenue afin de ne pas réduire l'effectif de l'étude, car cela aurait mené à un nombre de sujets d'analyses à 23. Les techniques d'imputations multiples de ces données manquantes n'ont pas été possibles en raison de la taille réduite de notre effectif. Si les techniques d'imputation simple « hot deck » ou des plus proches voisins n'ont pas non plus été retenues car aucun individu ne présentait de caractéristiques permettant de le considérer suffisamment proche pour imputer ses valeurs, deux autres techniques différentes d'imputation simple ont été utilisées. Pour le sujet PMD06, une imputation par modèle de régression sur l'ensemble du PHI a permis d'attribuer une valeur à chaque item de la dimension « Phonation ». Pour PMB01, comme le score global obtenu au domaine par le sujet était très proche de la valeur maximale (15/16), ce score a été divisé par quatre pour attribuer la valeur 3,75 à chaque item du domaine.

4.1.2 Sélection des items

En raison du format des données où le nombre de variables ($j = 174$ items retenus) est largement supérieur au nombre d'observations ($n = 25$ sujets), une méthodologie de réduction de dimensionnalité a été appliquée (Carreira-Perpinán, 1997; Chadeau-Hyam et al., 2013; Cunningham, 2014).

La détermination des caractéristiques de la variable-cible de cette réduction de dimensionnalité est primordiale car elle va conditionner la méthodologie statistique utilisée.

Ainsi, le score holistique de communication (ou SHC) que nous souhaitons construire est considéré comme une variable latente. En effet, il ne peut pas être directement observé ni mesuré, et il nécessite plusieurs variables manifestes comme indicateurs qui sont, elles, observables et mesurables. Cette variable latente explique les réponses sur les variables mesurées : en d'autres termes, le SHC est à l'origine des valeurs prises par les variables manifestes.

Le SHC sera construit comme une variable latente quantitative. Une seconde étape d'analyse sera de construire un SHC qualitatif pour confirmer les performances de ce score et aider à son analyse.

Les variables manifestes correspondent aux items des questionnaires. Bien que les réponses à ces variables soient construites sous la forme d'échelles de Likert (donc catégorielles ordinales), elles seront traitées dans cette étude selon une approche naïve, c'est-à-dire comme des variables quantitatives. Un test de Shapiro-Wilk, permettant de tester l'hypothèse nulle de normalité de distribution, montre que moins d'un quart de ces variables manifestes n'a pas de distribution normale, ce qui conforte le choix de cette approche.

Une première sélection d'items au moyen d'un jury d'experts, puis une sélection statistique a permis d'aboutir à un nombre réduit de paramètres pertinents, non redondants et suffisamment variables. Une analyse factorielle a été conduite, puis nous avons réalisé une validation croisée de ce score. Enfin, une analyse en profils latents menant à la construction d'un SHC qualitatif a permis d'aider à analyser les performances du SHC (voir Figure 24). Nous allons maintenant développer chacune de ces étapes.

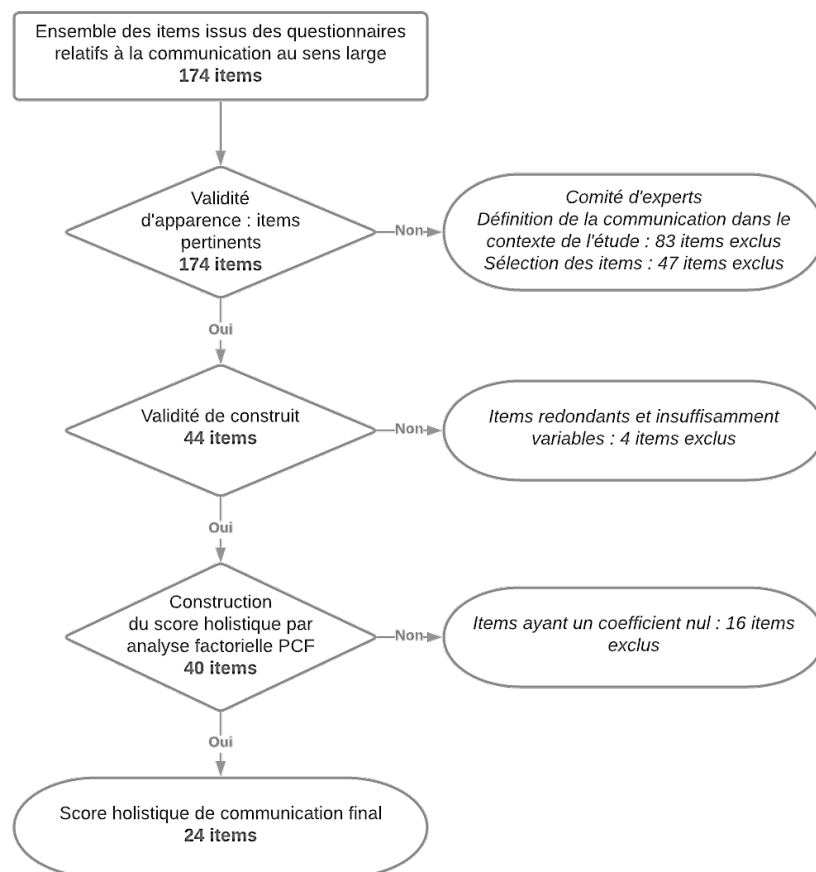


Figure 24 - Processus de construction du score de communication

4.1.3 Validité d'apparence

Nous avons sondé un comité de neuf experts de différentes disciplines concernées par la communication (sciences du langage, médecine, orthophonie, informatique et traitement du signal) pour définir les critères d'inclusion des items dans notre score holistique.

Le critère de **pertinence** des items consistait à conserver ceux qui étaient conformes au cadre conceptuel de l'étude.

Dans un premier temps, le comité a ainsi défini les éléments suivants comme critères de conservation des items :

- Items relatifs à la communication orale : exclusion des items relatifs à la communication écrite sauf s'ils permettent une compensation ou une augmentation de la communication orale ;
- Items relatifs à l'expression : exclusion des troubles de la compréhension seuls ;
- Items relatifs à une interaction avec un tiers, même implicite : interaction entre destinataire et destinataire selon le modèle de Jakobson (Jakobson, 1963).

Cette première étape a permis d'exclure 83 items : ceux relatifs à la lecture, à l'écriture, à la compréhension de conversation, aux items relatifs à la fatigue, à la douleur, à des symptômes spécifiques de la parole tels que la vitesse de parole...

Quatre-vingt-onze items (52,3 %) ont ainsi été retenus pour la suite de l'analyse.

Dans un second temps, un questionnaire en ligne au moyen de l'outil LimeSurvey a été proposé aux neuf membres du comité d'experts²². Les experts devaient indiquer quels items parmi les 91 restants seraient à conserver.

Les items ont été retenus s'ils respectaient un des deux critères suivants :

- L'I-CVI (Item-level Content Validity Index) (Lynn, 1986; Polit et al., 2007) était supérieur à 0,777, ce qui correspondait à un accord de sept experts sur neuf pour conserver l'item ;
- Le Kappa d'agrément était supérieur ou égal à 0,81 (Landis & Koch, 1977) : agrément « presque parfait ».

²² <https://enquetes.univ-tlse3.fr/index.php/998153?lang=fr>, consulté le 02/06/2021

Quarante-quatre items (48,4 %) respectaient l'une de ces deux conditions (voir Figure 25), et ont été retenus pour la suite des analyses.

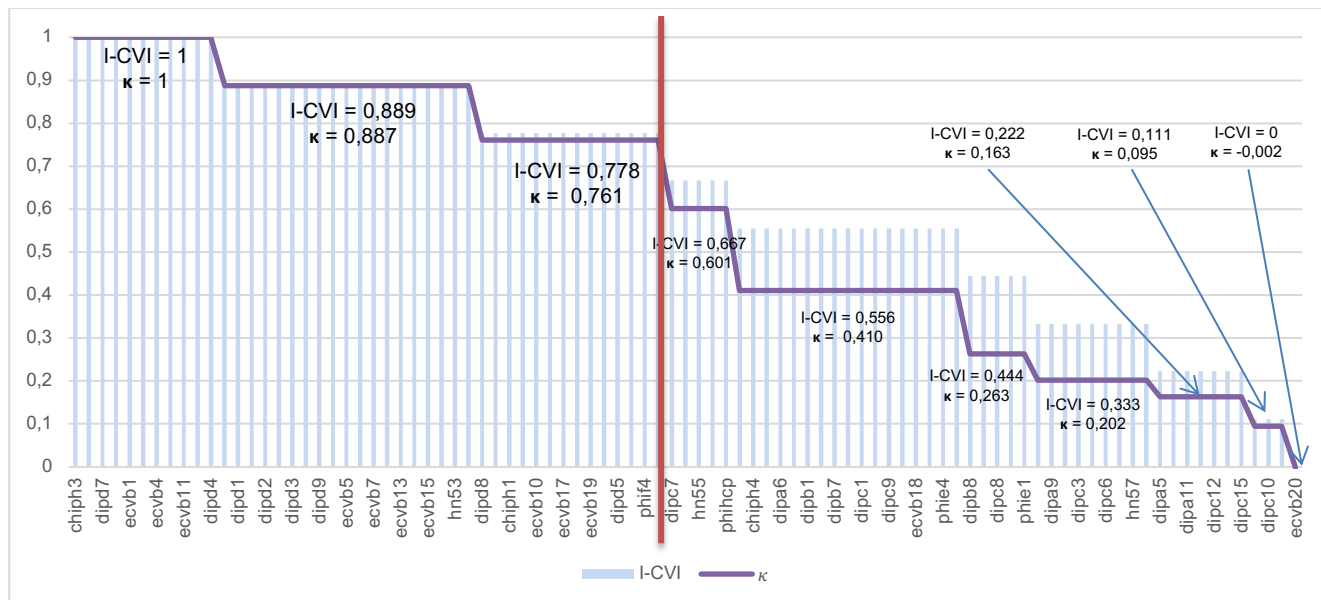


Figure 25 - Scores I-CVI et Kappas d'agrément avec cut-off (ligne rouge)

4.1.4 Validité de construit

Il s'agit ici de savoir si, concernant les concepts qui sont mesurés, les mesures retenues (les 44 items) sont associées entre elles, témoignant d'une redondance inutile de l'information.

Le critère de **non-redondance** des items a été étudié au moyen d'une analyse de la matrice de corrélation inter-items. Les coefficients de Spearman (non paramétriques) ont été utilisés en raison de l'effectif réduit ($n < 30$) de cette étude. Un seuil à 0,90 a été choisi : un seul des items corrélés entre eux à 0,90 ou plus ne peut être retenu pour la suite des analyses.

Dans ce cas, afin d'obtenir des items **suffisamment variables**, et donc permettant une meilleure mesure de la variabilité interindividuelle, seul l'item avec le coefficient de variation (si la distribution de cet item soit une loi normale, vérifiée au moyen du test de Shapiro-Wilk) ou l'index de dispersion (si la distribution ne suit pas de distribution normale) le plus élevé a été retenu.

Quatre items ont été exclus en raison d'une corrélation trop importante avec d'autres items. Quarante items sont finalement retenus pour la construction du score holistique de communication (voir Tableau 23).

Tableau 23 - Items inter-corrélés et choix des items retenus selon leur dispersion (en gras)

Corrélation entre les deux items de la paire	Code item	Item	Coefficient de variation (‡) ou Index de Dispersion (*)
0,95	ecvb11	Avez-vous des difficultés pour téléphoner à votre famille ?	0,67 ‡
	ecvb12	Avez-vous des difficultés pour téléphoner à vos amis ?	0,71 ‡
0,92	chiph3	Parlez-vous moins souvent avec votre famille, vos amis, vos voisins ?	0,77 ‡
	phie2	Mes difficultés de parole limitent ma vie personnelle et sociale	0,84 ‡
0,91	ecvb8	Et avec quelqu'un que vous connaissez peu (le facteur ou un chauffeur de taxi par exemple), êtes-vous gêné(e) pour avoir une conversation sur des sujets simples ? (le temps ; ce que vous avez fait la veille ; les fleurs de votre jardin...) ?	0,87 *
	ecvb10	Avez-vous des difficultés à prendre la parole lorsque vous êtes avec des gens que vous connaissez peu (lors d'un dîner, d'une sortie, d'une soirée...) ?	0,53 ‡
0,90	chiph2	Votre entourage a-t-il des difficultés pour vous comprendre ?	0,61 ‡
	phic4	On me demande de répéter à cause de mon problème pour parler	0,69 ‡

La liste des 40 items retenus est disponible en **Annexe IX**. Une vérification *a posteriori* a permis de vérifier qu'aucun de ces 40 items ne présente de différence significative entre les deux lots de sujets.

4.1.5 Construction du score holistique

Le choix de considérer le score holistique de communication (variable latente) et les variables manifestes comme des variables quantitatives nous a conduits à réaliser une analyse factorielle afin de répondre aux objectifs de cette étude.

L'analyse factorielle est une technique de réduction de données qui vise à capturer le maximum de variance des variables dans un jeu de données plus restreint. Contrairement à l'analyse en composantes principales dont elle est très proche, l'analyse factorielle n'induit pas d'hypothèse de non-corrélation des

facteurs entre eux. Ce type d'analyse se prête ainsi bien à l'élaboration d'une variable latente qui par définition influence les valeurs de toutes les variables manifestes.

L'analyse factorielle vise à sélectionner un nombre optimal de facteurs où le premier facteur capture le maximum de variance commune aux variables, le deuxième facteur le deuxième maximum de variance et ainsi de suite. Dans l'analyse factorielle, les relations entre la variable latente F et les variables manifestes Y sont pondérées par des charges b , calculés de façon optimale par ce type d'analyse. De plus, il existe également un set d'erreurs u correspondant à la variance non expliquée par le facteur (voir Figure 26).

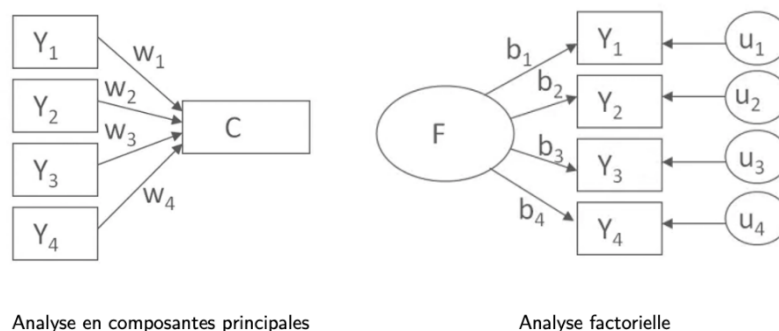


Figure 26 - Différences entre analyse en composantes principales et analyse factorielle selon K. Grace-Martin (<https://www.theanalysisfactor.com/the-fundamental-difference-between-principal-component-analysis-and-factor-analysis/>)

Différents types d'analyses factorielles existent, et c'est une analyse factorielle en PCF (principal-component factor) qui a été retenue pour cette étude (Roscoe et al., 1982). L'analyse factorielle en PCF est privilégiée en réduction de données (Acock, 2018) car elle tente d'expliquer toute la variance : dans ce cas, les variances communes sont fixées à 1, et il n'y a pas de variance unique. Pour cette raison, elle est utilisée quand on souhaite expliquer autant que possible la variance d'un ensemble d'éléments par une seule dimension, autrement dit quand on dispose d'un ensemble d'éléments qui mesurent tous un même concept. Dans ce cas, seul le premier facteur principal sera facteur d'intérêt. Elle se prête donc bien au contexte de cette étude, où un seul score holistique latent quantitatif (le SHC) est recherché parmi l'ensemble des 40 variables manifestes (correspondant aux items sélectionnés).

Afin de pouvoir être utilisée, l'analyse factorielle en PCF présente deux conditions d'application : d'une part, le premier facteur principal retenu doit expliquer une partie « substantielle » de la variance totale pour l'ensemble des items, et d'autre part, la plupart des items doit avoir une charge de 0,4 ou plus sur ce facteur.

Dans cette étude, les conditions d'application sont respectées. La valeur propre du facteur 1 (c'est-à-dire la proportion de variance totale imputable au facteur 1) est de 17,51, contre 4,57 pour le facteur 2 (voir Figure 27). De plus, la proportion de variance expliquée uniquement par le facteur 1 est de 0,44 (0,11 pour le facteur 2). Enfin, 34 items sur 40 ont une charge de 0,4 ou plus sur ce facteur, soit 85 %.

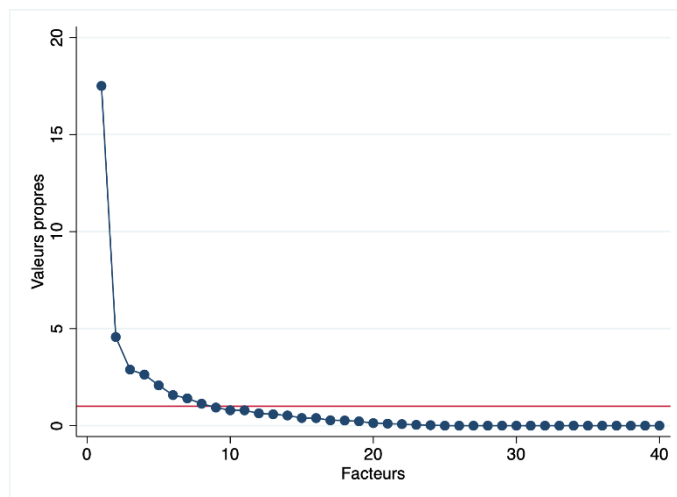


Figure 27 - Scree plot (tracé d'éboulis) des valeurs propres des facteurs suite à l'analyse factorielle en PCF

L'analyse factorielle en PCF étant applicable, une prédiction des valeurs pour le facteur 1, correspondant à la variable latente SHC, a été réalisée. Cette prédiction est issue d'une analyse en régression sur l'ensemble de nouvelles variables créées par estimation du premier facteur. Ainsi, par sujet, le score est prédit par la somme des valeurs standardisées de chaque item pondérées par le coefficient de régression correspondant au facteur 1.

Seize items ont un coefficient égal à zéro, neutralisant la valeur prise par l'item, et donc indiquant qu'ils ne seront pas pris en compte dans le calcul global du SHC. Par conséquent, 24 items sur 40 (60 %) ont des coefficients différents de zéro et seront conservés pour le calcul du SHC.

Pour en faciliter la lisibilité et l'interprétation, le score prédit, initialement centré sur zéro et ayant un écart-type de 1, sera transformé : il sera centré sur 100 avec un écart-type de 10. C'est ce score transformé qui constituera le SHC.

Le SHC se calcule donc ainsi :

$$SHC = \left(\frac{\left[\sum \left(\frac{X_{item} - m_{item}}{s_{item}} \times \beta_{item} \right) \right] - m_{score_initial}}{s_{score_initial}} \times 10 \right) + 100$$

$$= \left(\frac{\left[\sum \left(\frac{X_{item} - m_{item}}{s_{item}} \times \beta_{item} \right) \right] - 0,0353408}{0,999787} \times 10 \right) + 100$$

où X_{item} représente la valeur brute de l'item (score obtenu à l'item par le sujet), m_{item} la moyenne de l'item obtenue dans notre échantillon, s_{item} l'écart-type de l'item obtenu dans notre échantillon, β_{item} représente le coefficient de régression de l'item, $m_{score_initial}$ représente la moyenne du score initialement prédit, $s_{score_initial}$ l'écart-type du score initialement prédit.

Les SHC obtenus sont représentés dans la Figure 28.

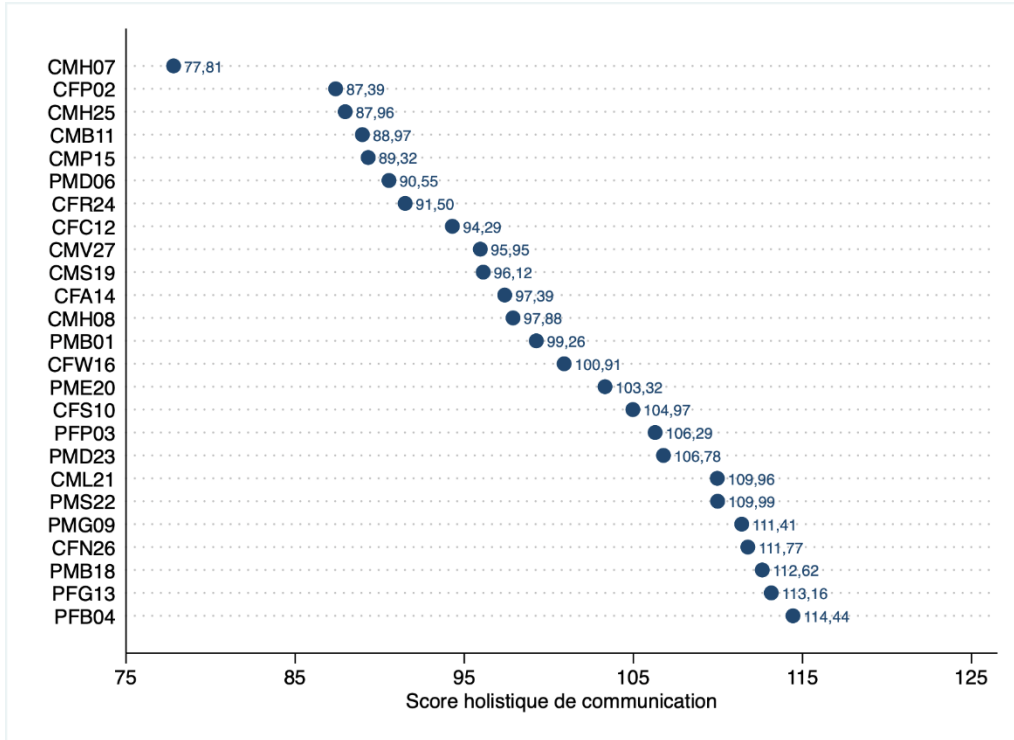


Figure 28 - Scores holistiques de communication obtenus par sujet

La formule complète est ainsi :

$$SHC = (((((ecvb1-2,36)/0,8103497*0,03264) + ((ecvb4-2,08)/0,9539392*0,01566) + ((ecvb5-1,8)/0,9128709*0,18626) + ((ecvb6-$$

$$\begin{aligned}
& 1,4932)/0,8338601*0,10411) + ((ecvb7-2,34)/0,8524999*0,01695) + ((ecvb9- \\
& 1,8)/0,9574271*-0,00532) + ((ecvb12-1,68)/1,215182*-0,01413) + ((ecvb13- \\
& 1,56)/1,356466*0,10219) + ((ecvb16-1,72)/1,137248*0,12197) + ((ecvb17- \\
& 2,3268)/0,8902*0,06856) + ((ecvb19-1,4932)/1,054509*-0,01007) + ((ecvb25- \\
& 1,82)/1,081376*0,15096) + ((ecvb26-2,28)/0,9363048*-0,08137) + ((dipd2- \\
& 2,88)/1,563117*-0,04681) + ((dipd3-2,2)/0,8660254*0,1402) + ((dipd5- \\
& 3,36)/1,113553*-0,03138) + ((dipd6-3,48)/1,262273*0,14321) + ((dipd7- \\
& 2,64)/1,254326*0,05331) + ((dipd9-2,84)/1,344123*-0,12441) + ((hn53-2,2)/1* \\
& -0,071) + ((phif4-2,32)/1,519868*-0,00858) + ((phic1-0,99)/1,251*-0,24677) + \\
& ((phic3-1,76)/1,422439*-0,08623) + ((phic4-1,68)/1,180395*-0,07575))-0,0353408) \\
& / 0,999787 * 10) + 100
\end{aligned}$$

4.1.6 Validation croisée à cinq blocs

Nous avons ensuite conduit une validation croisée à cinq blocs pour vérifier la fiabilité de ce score (soit un entraînement sur 20 observations et une prédiction sur les cinq autres observations, répété cinq fois dans notre cas).

Une corrélation forte à 0,91 est retrouvée entre la vérité terrain (score SHC réellement mesuré) et les valeurs prédites par la validation croisée (voir Figure 29).

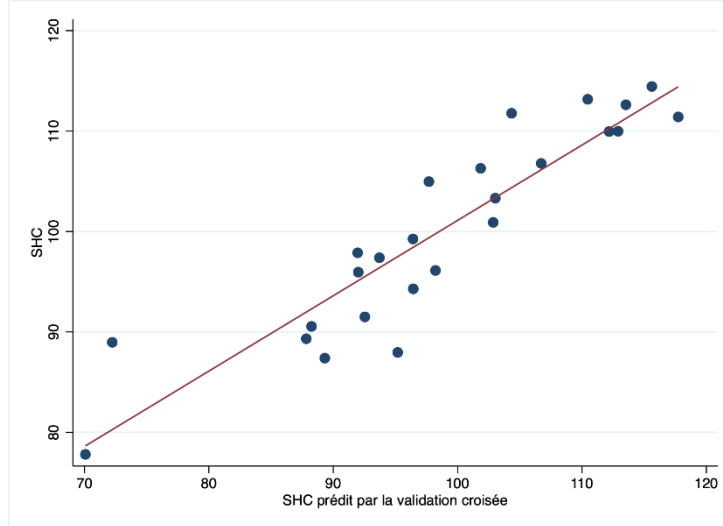


Figure 29 - Nuage de points représentant le SHC et le SHC prédit par la validation croisée, avec droite de régression

4.1.7 Validation complémentaire par analyse en profils latents

Pour enrichir l'interprétation du score SHC, une analyse complémentaire permettant la construction d'un score latent qualitatif a été menée. Les

résultats entre le SHC, score qualitatif issu de l'analyse factorielle, et le score catégoriel issu de cette nouvelle analyse ont été comparés.

Les mêmes 40 items ont été retenus comme variables manifestes du score latent qualitatif que nous abrégerons SHC-qual.

Ce SHC-qual a été élaboré par une analyse en profils latents (Cai, 2012).

L'analyse en profils latents (ou LPA) est basée sur des modèles GSEM (Generalized Structural Equation Modeling, ou modèles d'équations structurelles généralisées), au même titre que les analyses en classes latentes (Coma et al., 2013). Ils permettent de déterminer quels individus appartiennent le plus probablement à un groupe (correspondant à une catégorie de la variable latente) selon les informations portées par d'autres variables.

Concrètement, la LPA adapte un modèle linéaire pour chaque variable observée, et recherche ensuite le meilleur ajustement de modèle à une variable latente à C classes en deux étapes :

- Dans chaque classe, pour chaque variable, une régression linéaire est réalisée. La constante varie selon la classe, les résidus sont estimés ;
- La probabilité d'appartenir à chaque classe du modèle logistique multinomial est estimée par l'exponentielle de la constante de la classe divisée par la somme des exponentielles des constantes de toutes les classes :

$$Pr(C = x) = \frac{e^{\gamma_x}}{\sum_{i=1}^j e^{\gamma_i}}$$

avec C les classes latentes, x la classe prédite, γ les constantes des classes dans le modèle, j le nombre de classes latentes.

Les paramètres de sélection de modèle AIC (Akaike Information Criterion) et BIC (Bayesian Information Criterion) ont été calculés pour un nombre de classes C fixé à deux, trois et quatre classes (voir Tableau 24). Le modèle aboutissant à une variable latente à trois classes a ainsi été retenu, car il présente les critères AIC et BIC les plus faibles (Cameron & Trivedi, 2005).

Tableau 24 - Valeurs AIC et BIC des modèles GSEM (Analyse en profils latents)

Modèle	AIC	BIC
Variable latente à 2 classes	2 811,97	2 959,45
Variable latente à 3 classes	2 729,59	2 927,04
Variable latente à 4 classes	2 747,96	2 995,39

La classe d'appartenance de chaque sujet a ensuite été prédite. Pour chaque sujet, la variable latente catégorielle SHC-qual prendra ainsi une des trois valeurs correspondant à une des trois classes (1, 2 ou 3).

Une comparaison des valeurs des deux variables latentes quantitative (score SHC de référence) et catégorielle (SHC-qual) montre ainsi que la classe 1 du score SHC-qual correspond aux sujets ayant les scores SHC les plus faibles, la classe 2 correspond aux sujets ayant les scores SHC les plus élevés et enfin la classe 3 aux sujets ayant les scores SHC intermédiaires (voir Tableau 25).

Ainsi, l'analyse en profils latents permet d'aboutir à une variable correspondant à un niveau d'impact sur la communication très liée au score SHC. Cette analyse confirme également le fait que le score SHC correspond bien à un niveau d'impact sur la communication, validant donc la construction de notre indicateur holistique.

Tableau 25 - Comparaison des classes latentes prédites par l'analyse LPA et du score SHC

Sujet	Classe prédite	SHC	Sujet	Classe prédite	SHC	Sujet	Classe prédite	SHC
CMH07	1	77,81	CMV27	3	95,95	CFS10	2	104,97
CFP02	1	87,39	CMS19	3	96,12	PFP03	2	106,29
CMH25	1	87,96	CFA14	3	97,39	PMD23	2	106,78
CMB11	1	88,97	CMH08	3	97,88	CML21	2	109,96
CMP15	1	89,32	PMB01	3	99,26	PMS22	2	109,99
PMD06	1	90,55	CFW16	3	100,91	PMG09	2	111,41
CFR24	1	91,50	PME20	3	103,32	CFN26	2	111,77
CFC12	1	94,29				PMB18	2	112,62
						PFG13	2	113,16
						PFB04	2	114,44

4.1.8 Discussion

La construction du SHC a permis d'élaborer un indicateur novateur, prenant en compte les dimensions symptomatologique (par exemple : item *phif4* « Je dois faire un effort pour parler »), interactionnelle (par exemple : item *phic3* « J'ai du mal à communiquer avec des personnes peu familières »), pragmatique (par exemple : *ecvb25* « Au restaurant / café, avez-vous des difficultés pour passer votre commande vous-même ? »), mais aussi psychoaffective (par exemple : *dipd9* « Je me sens à l'aise pour m'exprimer

*dans la plupart des situations, tant à la maison qu'à l'extérieur »). Les stratégies de compensation des difficultés de communication sont également prises en compte (par exemple : *dipd3* « Lorsque les gens ne me comprennent pas, j'essaie de faire passer mon message par d'autres moyens »).*

Dans cette étude, une corrélation de Spearman à 0,79 est retrouvée entre le SHC et le score perceptif de sévérité attribué par le jury de six auditeurs experts. Cette valeur est bien plus élevée que celles retrouvées dans d'autres études. Chez des patients laryngectomisés totaux, la corrélation entre sévérité perçue par les patients de leur trouble de parole et le CPIB est de 0,055 (Eadie et al., 2016). Elle se situe aux alentours de 0,40 chez les patients traités pour un cancer de la cavité buccale ou de l'oropharynx entre un score composite « London Speech Evaluation Scale » (constitué d'auto-évaluations sous forme d'échelles de Likert de l'intelligibilité, de l'articulation, de la nasalité, du débit et de la faiblesse de parole) et deux items du SHI (Dwivedi et al., 2016). Ces différences de performances entre la corrélation forte dans notre étude et les corrélations retrouvés dans la littérature peuvent s'expliquer notamment par les moyens de mesure utilisés, tant du côté du trouble de parole (auto-perception globale ou sur des concepts plus spécifiques mais qui restent malgré tout sujets à une certaine variabilité conceptuelle, contre une mesure externe par un auditeur), que du côté de la communication (échelles de handicap de parole SHI, contre score holistique global).

Dans notre étude, la force du lien retrouvée peut provenir d'un côté de la constitution du jury experts, avec des auditeurs habitués à l'évaluation et au suivi des patients en cancérologie ORL : les experts incluent dans leur mesure du trouble de parole la dimension fonctionnelle communicationnelle, qui est un des besoins rapportés couramment en pratique courante par les patients. D'un autre côté, le SHC est un indicateur de communication volontairement holistique, incluant les notions d'altération de parole à la source des limitations de communication mais aussi les aspects fonctionnels et psychoaffectifs.

Cet indicateur, holistique du fait de sa construction, vient ainsi combler un manque d'outils de mesure des conséquences fonctionnelles des traitements des cancers de la cavité buccale et de l'oropharynx sur la parole et la communication en pratique courante. Il n'est constitué que d'items issus de questionnaires complétés par les patients. La mesure de l'altération de communication est donc auto-rapportée, ce qui est pertinent dans l'objectif clinique d'adapter les stratégies thérapeutiques aux besoins exprimés par les patients.

Malgré sa construction à partir d'un échantillon relativement réduit ($n = 25$), le SHC montre une bonne fiabilité en validation croisée ($r_s = 0,91$).

C'est également un score valide, qui mesure bien le niveau d'impact sur la communication, ce qui a été confirmé par l'analyse en profils latents.

Pour mieux s'assurer d'une généralisation des résultats, cette étude reste à compléter. D'une part, l'augmentation de la taille de l'échantillon permettrait une meilleure puissance, et d'autre part, ce score serait à évaluer sur un nouvel échantillon de patients dans le cadre d'une validation externe.

4.2 Modélisations de l'impact sur la communication au moyen des paramètres automatiques

Maintenant que nous disposons d'un score holistique de communication de référence, nous allons pouvoir étudier les performances des 75 paramètres issus d'une analyse automatique de la parole spontanée, en termes de prédiction de l'altération de la communication rapportée par les sujets.

En effet, même si la corrélation est forte entre sévérité perceptive évaluée par des experts et score holistique de communication dans notre corpus ($r_s = 0,79$), le caractère automatique de la mesure de l'impact du trouble de parole sur la communication permettrait de s'affranchir de la variabilité inter-juges. Une mesure automatique assurerait donc une meilleure reproductibilité en clinique courante où l'évaluation est la plupart du temps menée par un seul auditeur.

4.2.1 Analyse bivariée entre SHC et paramètres automatiques

Nous avons mené une première analyse, en comparant chacun des 75 paramètres automatiques retenus avec le score holistique de communication.

Ces analyses bivariées portent sur l'ensemble des 25 sujets, à l'exception des paramètres « Score moyen global de confiance des occlusives » (*confo*) et « Score moyen global de confiance des fricatives » (*conff*) pour lesquels le sujet CMB11 a été exclu (les valeurs de ces deux paramètres sont manquantes car aucune occlusive ni fricative n'a été reconnue par le TDNNf « phonémique » pour ce sujet).

Nous avons calculé les coefficients de corrélation de Spearman entre le SHC et chacun des paramètres automatiques. Parmi ces 75 paramètres, quatre (5 %) présentent une corrélation supérieure ou égale à 0,70, 13 (17 %) entre 0,50 et 0,70, et 16 (21 %) entre 0,40 et 0,50. Notons que tous les 33 paramètres (44 %)

ayant des coefficients de corrélation supérieurs ou égaux à 0,40 ont une valeur p significative au seuil de 5 % (voir Figure 30).

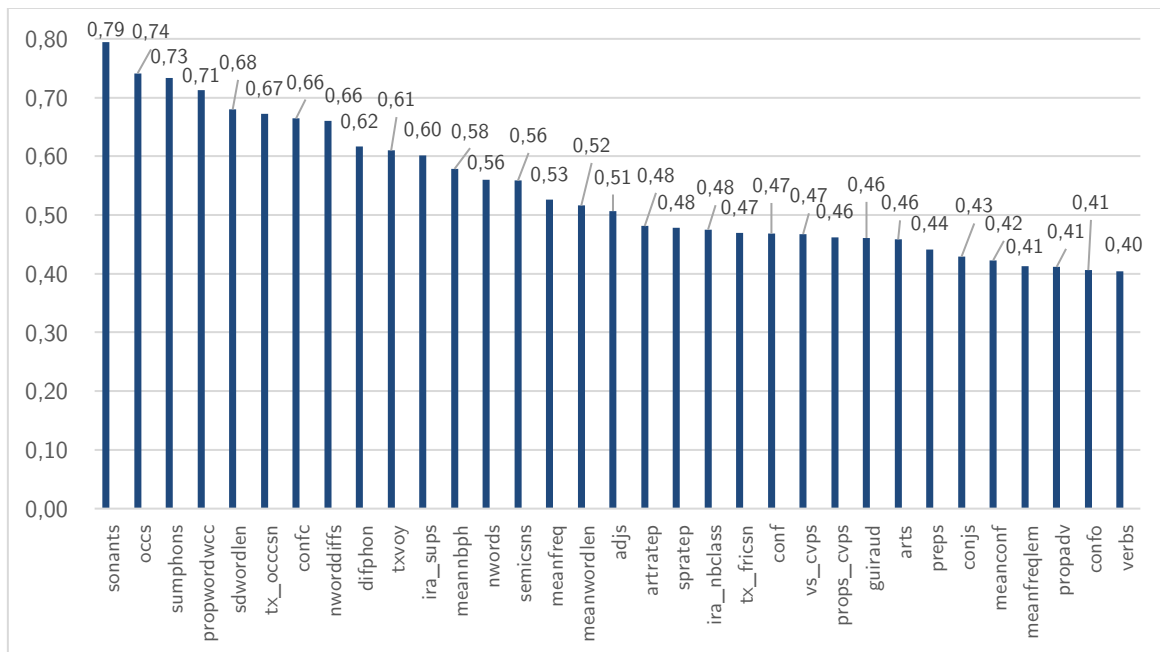


Figure 30 - Analyses bivariées entre le score holistique de communication et les paramètres automatiques dont les coefficients de corrélation de Spearman sont significatifs au seuil de 5 %.

4.2.2 Modélisation prédictive au moyen des seuls paramètres automatiques

Afin de prédire le score holistique de communication au moyen des paramètres issus des analyses automatiques de la parole spontanée, plusieurs approches statistiques ont été réalisées.

Dans un premier temps, nous avons mené une analyse en régression en composantes principales (RCP). Cette approche, classiquement utilisée et guidée par les données, a permis de déterminer 12 composantes latentes aux 75 paramètres automatiques. Une régression en composantes principales sur les 12 composantes déterminées est ensuite réalisée, avec pour variable dépendante le score holistique de communication.

Même si la corrélation est modérée entre valeurs prédites sur l'ensemble de l'échantillon par RCP et SHC ($r_s = 0,65$), cette technique montre de faibles performances en termes de validation croisée à cinq blocs (coefficient de corrélation de Spearman entre score prédit et SHC : $r_s = 0,20$). Cette faiblesse de performance en validation croisée peut s'expliquer par le fait que les 12

composantes prédites selon les 12 plus grandes valeurs propres ne sont pas nécessairement les meilleures variables indépendantes dans la régression en composantes principales pour prédire le SHC, car elles ne sont pas nécessairement corrélées avec le SHC de façon suffisante.

Devant les limites présentées par cette modélisation, nous avons mené une analyse selon l'approche PLS (Partial Least Squares, ou régression des moindres carrés partiels). Cette technique, qui s'appuie sur des modélisations par équations structurelles, permet de modéliser des relations complexes entre variables, dont certaines sont manifestes et d'autres latentes (Tenenhaus et al., 2005; Venturini & Mehmetoglu, 2019).

Le but de cette approche est de prédire des concepts cibles (dans notre cas le SHC) au moyen de variables manifestes et latentes regroupées en construits intermédiaires. Elle fonctionne en deux temps. La première étape constitue le modèle de mesure (à savoir les relations entre variables manifestes et construits latents) fonctionnant sur le principe de l'analyse factorielle et de l'analyse en composantes principales : les variables formatives correspondant aux construits latents vont être calculées en fonction des variables manifestes qui y sont rattachées. La seconde étape constitue le modèle structurel : il met en lien les variables formatives générées à l'étape précédente (correspondant aux construits latents) sur le principe de modèles de régression classiques.

À défaut d'hypothèses théoriques préalables nous permettant de construire nos modèles sur des relations validées entre les concepts, nous avons élaboré notre modèle de mesure à partir de 76 variables manifestes (les 75 issues des analyses automatiques, et le SHC) connectées à 13 variables formatives (12 issues d'une analyse en composantes principales préalable sur les 75 paramètres automatiques, et une dernière de « Communication » constituée uniquement du SHC). Notre modèle structurel reliait bilatéralement nos 12 variables formatives issues de l'analyse en composantes principales (chacune influençait l'autre), tandis que ces 12 variables influençaient unidirectionnellement la variable formative constituée du seul SHC. Une limite de cette approche est liée à la méthodologie d'élaboration des 12 variables formatives issues de l'ACP : comme nous l'avons indiqué, elles ne sont pas forcément les mieux corrélées au SHC et limiteront donc les performances du modèle structurel.

L'analyse PLS montre une forte corrélation entre le SHC et le score prédit sur l'ensemble de l'échantillon ($r_s = 0,94$). Mais bien que cette approche puisse être utilisée sur des échantillons de taille réduite et qu'elle gère la colinéarité éventuelle entre variables manifestes, la validité croisée à cinq blocs présente là encore des performances limitées (corrélation entre SHC et valeur prédite : $r_s = 0,57$), ce qui témoigne d'un probable phénomène de surajustement du modèle aux données. Nous n'avons donc pas retenu cette méthode pour prédire

notre SHC en raison des performances moyennes en validité croisée. Un élargissement de la taille de l'échantillon permettrait peut-être d'améliorer les performances. Enfin, un réajustement du modèle selon de nouveaux concepts théoriques sous-jacents permettrait de mieux expliquer les relations entre (a) variables manifestes et construits latents d'une part, et (b) construits latents entre eux d'autre part. Dans ce cas, la construction des modèles de mesure (a) et structurels (b) serait optimisée, et permettrait d'améliorer ainsi la fiabilité du score prédit.

Nous avons ensuite mené une approche en régression LASSO (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996). Cette méthode de régression pénalisée par un terme d'erreur est utilisée notamment dans des objectifs de prédiction, en permettant de sélectionner des variables. Elle permet de limiter les phénomènes de surajustement en minimisant l'erreur de prédiction par validation croisée, et aboutit à sélectionner des variables qui appartiennent au modèle comprenant le terme d'erreur le plus faible entre valeur réelle et valeur prédite.

Lors de l'utilisation de cette approche, nous avons été confrontés à la présence de deux valeurs manquantes pour le sujet CMB11 (à savoir les scores de confiance sur les occlusives *confo* et les semi-consonnes *confs*) conduisant à exclure ce sujet des analyses, l'analyse LASSO se faisant uniquement sur des données complètes. Afin de ne pas réduire à 24 sujets notre échantillon, nous avons choisi pour cette étape d'imputer ces deux valeurs manquantes. La stratégie d'imputation est liée au caractère MNAR (Missing Not At Random) des données manquantes car l'absence de score vient du fait qu'aucune occlusive ni semi-consonne n'a été reconnu par le TDNNf « phonémique » en raison de la sévérité du trouble de parole du sujet (score de sévérité perceptive à 0,66/10). Plusieurs stratégies d'imputation ont été mises en place : par des valeurs prédéfinies (imputation par les valeurs 0 et 0,5 en raison de la dégradation de la qualité de la parole) et par technique « hot deck » (imputation par les valeurs d'un sujet présentant les scores perceptifs de sévérité les plus proches). Finalement, quelle que soit la stratégie d'imputation choisie, les résultats de la régression LASSO se sont avérés identiques.

La régression LASSO sur nos données aboutit à la sélection d'un modèle à deux variables explicatives (erreur moyenne de prédiction en validation croisée la plus faible à 108,56 pour un coefficient de pénalité λ à 6,91) : le nombre d'occlusives par seconde (*occs*) et le nombre de sonantes par seconde (*sonants*). Ce modèle linéaire à deux variables explicatives prédisant le SHC respecte toutes les conditions d'application (notamment l'homoscédasticité des résidus : $p = 0,29$, test de Breusch-Pagan ; la distribution gaussienne des résidus : $p = 0,84$, test de Shapiro-Wilk ; et l'absence de problème de

colinéarité : facteur d'inflation de la variance à 2,20 inférieur au seuil de 5). Il présente un RMSE (Root-mean-square error ou racine de l'erreur quadratique moyenne) à 6,38. Ce modèle prédit 62,7 % de la variance totale du SHC.

Les valeurs prédites par ce modèle de régression montrent des corrélations élevées avec le SHC. La corrélation de Spearman entre SHC et valeurs prédites sur les 25 sujets de notre échantillon est de $r_s = 0,83$ (voir Figure 31).

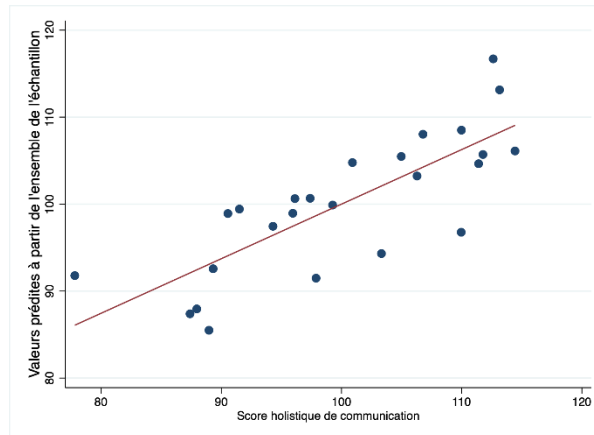


Figure 31 - Nuage de points entre SHC et valeurs prédites sur les 25 sujets de l'échantillon

La validation croisée témoigne là encore de bonnes performances de cette prédiction. Les corrélations sont élevées entre SHC et, d'une part, les valeurs prédites par la validation croisée « leave one out » (entraînement à partir des deux paramètres automatiques sélectionnés sur 24 sujets et prédiction sur le dernier sujet, $r_s = 0,78$), et d'autre part les valeurs prédites par la validation croisée à cinq blocs ($r_s = 0,82$, voir Figure 32).

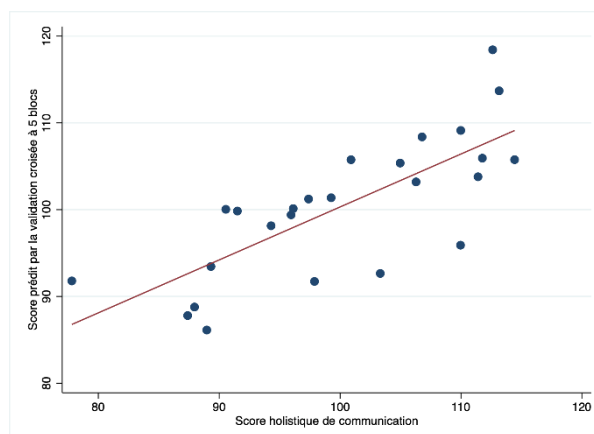


Figure 32 - Nuage de points entre SHC et valeurs prédites par la validation croisée à cinq blocs

Au vu des performances présentées par les différentes techniques utilisées, nous retenons les résultats de la méthode LASSO : le score holistique de communication est le mieux prédit au moyen des deux paramètres automatiques « nombre d'occlusives reconnues par seconde » (*occs*) et « nombre de sonantes reconnues par seconde » (*sonants*), selon la formule suivante :

$$Y_{SHC} = 83,672 + (10,139 \times occs) + (10,891 \times sonants)$$

Les deux meilleurs paramètres prédisant l'altération de la communication sont donc des paramètres issus d'une analyse phonémique du signal de parole. Ces résultats demanderaient à être confirmés sur un échantillon plus conséquent.

Plusieurs raisons peuvent expliquer ce phénomène, liées à la taille de la tumeur, la qualité de la reconnaissance de la parole, et à la construction du SHC.

Parmi les 19 sujets de notre échantillon pour lesquels nous avons pu récupérer les données tumorales, la majorité (11/19, 58 %) sont des patients ayant été traités pour des tumeurs de grande taille (T3 ou T4). Or, les tumeurs de grande taille sont les plus délétères pour la qualité de parole. L'atteinte de bas niveau (au niveau des phonèmes) sera ainsi au premier plan chez ces patients, limitant la qualité de la reconnaissance au niveau lexical et les résultats obtenus aux niveaux supérieurs sur les sorties lexicales.

Sur la minute transcrite, le taux d'erreurs mots du système utilisé est très élevé (à 84,51 %), probablement lié à la mauvaise qualité de la production de la parole pathologique. Il est donc possible que la variabilité des sorties lexicales ne permette pas aux paramètres lexicaux et discursifs d'être liés de façon significative au score holistique de communication, par un effet ricochet bottom-up où les dégradations phonémiques de bas niveau impactent les résultats sur les composantes linguistiques de plus haut niveau.

Une dernière explication est à relier à la construction de notre score holistique de communication. Ce score est auto-rapporté, et il est possible que l'altération perçue par les sujets sur leurs capacités de communication quotidienne soient impactées davantage par les déficits segmentaux porteurs de l'intelligibilité de la parole. En d'autres termes, les sujets attribuent leurs modifications des capacités de communication par leur capacité réduite à articuler correctement un message. Cette interprétation rejoint les résultats retrouvés par Bolt (Bolt et al., 2016), où la sévérité du trouble de parole perçu par les sujets eux-mêmes est le critère principal altérant la participation aux activités de communication des patients après cancer des VADS. Les déficits

suprasegmentaux ne seraient pas retrouvés dans notre étude car ils seraient dégradés par les atteintes phonémiques au niveau articulatoire, qui conditionneraient les choix linguistiques des sujets dans la production de leur parole. Les déficits liés à d'autres niveaux linguistiques seraient peut-être retrouvés en utilisant comme référence un indicateur mixte d'altération de communication, comprenant une part d'évaluation du côté du patient et une part du côté de l'interlocuteur. Les phénomènes linguistiques à la base de la compréhensibilité du locuteur seraient alors peut-être davantage mis en évidence.

4.2.3 Modélisation prédictive avec facteurs associés

Nous avons ensuite cherché à ajuster au mieux notre modèle en intégrant à la modélisation les facteurs associés que nous avons récupérés au moyen des questionnaires d'anxiété / dépression, d'évaluation des déficits associés, des besoins de communication liés aux cercles sociaux, et de statut cognitif.

Pour cela, nous avons conduit une première régression LASSO. Nous avons intégré comme variables explicatives les 75 paramètres issus des analyses automatiques de la parole spontanée mais aussi l'ensemble des items issus des questionnaires n'ayant pas été utilisés pour la constitution du SHC et n'étant pas relatifs à la qualité de vie (129 items : 16 items et scores issus du HAD, un score issu de l'EDP, 50 items issus du CHI avec exclusion du domaine « Phonation » utilisé pour la construction du SHC, trois scores issus du PHI non utilisé pour le SHC, 27 items et scores de l'ECCS, 18 du QFS, et 14 de la MoCA), mais aussi les données de traitement pour lesquelles nous n'avons pas de valeur manquante (quatre items supplémentaires : délai depuis la fin du traitement : *delaittt*, la réalisation ou non d'une chirurgie : *chir*, d'une radiothérapie : *rte*, et la localisation tumorale : *loc*).

Cette méthode globale aboutit à des résultats insatisfaisants, dans la mesure où ils violent les conditions d'application de la régression linéaire (hétéroscédasticité, problèmes de multicollinéarité, distribution des résidus non gaussienne).

Nous avons alors choisi de mener cette modélisation du SHC en deux temps : une première régression LASSO permettant de sélectionner les facteurs associés les plus pertinents dans la prédiction du SHC, puis une modélisation complète incluant les facteurs associés retenus dans le premier temps et les deux paramètres automatiques (nombre d'occlusives *occs* et nombre de sonantes *sonants* reconnus par seconde) retenus dans la section précédente.

La première régression n'incluant que les 133 facteurs associés conduit à en retenir neuf. Quatre sont relatifs à la constitution des cercles sociaux : satisfaction moyenne de la fréquence de contact avec les membres du cycle privé (*eccs1c*), satisfaction moyenne de la qualité des échanges avec les membres du cercle privé (*eccs1d*), satisfaction relative aux relations avec les membres de l'entourage extrafamilial (*qfs52*), satisfaction relative à la manière de gérer les aspects financiers et administratifs (*qfs62*). Deux concernent l'état anxiodépressif : « *Je prends plaisir aux mêmes choses qu'autrefois* » (*had2*), score d'anxiété global (*hada*). Un item concerne les déficits associés au niveau de l'olfaction : « *Avez-vous peur d'un accident à cause de vos problèmes d'odorat ?* » (*chiog2*). Enfin, les deux derniers sont relatifs à l'auto-perception de la gêne et du handicap ressentis quant au trouble de parole : gêne quant à la production d'une parole compréhensible (*phigene*), handicap quotidien ressenti quant au problème de parole (*phihcp*).

L'étape finale consiste ainsi à modéliser par une régression linéaire le score holistique de communication avec les deux paramètres automatiques (*occs* et *sonants*) et les neuf items et scores relatifs aux facteurs associés. Nous avons réalisé une imputation des valeurs manquantes des deux items de l'ECCS par zéro, en raison du caractère neutre de ce score dans le questionnaire (concerne quatre sujets). Une vérification des conditions de régression a conduit à exclure la variable relative à la gêne quant à la production de parole compréhensible (*phigene*) car elle présente un facteur d'inflation de la variance très élevé (à 38,83) ce qui pose un problème de multicollinéarité.

Après exclusion de cette variable (et modélisation du SHC par dix variables explicatives : deux paramètres automatiques et huit facteurs associés), les conditions d'application de la régression sont respectées. La modélisation du SHC se trouve dans le Tableau 26.

Cette modélisation présente un R^2 à 0,836, un R^2 ajusté à 0,720 et une racine de l'erreur quadratique moyenne à 5,30.

La validation croisée à cinq blocs montre une corrélation $r_s = 0,70$ entre le score prédit et le SHC. La validation croisée « leave one out » présente une corrélation plus élevée à $r_s = 0,78$. Enfin, le score prédit sur les 25 sujets est corrélé à $r_s = 0,89$ avec le SHC « vérité terrain ».

Tableau 26 - Résultats de la modélisation du SHC par les paramètres automatiques et les facteurs associés

Code	Paramètre	Coefficient
<i>Constante</i>		73,74
<i>occs</i>	Nombre d'occlusives reconnues par seconde	9,10
<i>sonants</i>	Nombre de sonantes reconnues par seconde	4,07
<i>eccs1c</i>	Satisfaction moyenne de la fréquence de contact avec les membres du cycle privé	1,32
<i>eccs1d</i>	Satisfaction moyenne de la qualité des échanges avec les membres du cercle privé	0,05
<i>qfs52</i>	Satisfaction relative aux relations avec les membres de l'entourage extrafamilial	3,58
<i>qfs62</i>	Satisfaction relative à la manière de gérer les aspects financiers et administratifs	1,56
<i>had2</i>	« Je prends plaisir aux mêmes choses qu'autrefois »	-5,22
<i>hada</i>	Score d'anxiété global	3,39
<i>chiog2</i>	« Avez-vous peur d'un accident à cause de vos problèmes d'odorat ? »	-3,99
<i>phihcp</i>	Handicap quotidien ressenti quant au problème de parole	-0,12

La Figure 33 représente les principaux résultats des modélisations. La modélisation incluant les facteurs associés explique davantage de variance du SHC (84 % contre 63 % pour les seuls paramètres automatiques). Elle est un meilleur prédicteur de l'altération de communication quand le modèle est entraîné sur l'ensemble de l'échantillon (corrélation à 0,89 contre 0,83 pour les seuls paramètres automatiques). Ce gain est cependant lié à de moins bonnes performances en validation croisée à cinq blocs, ce qui limite la fiabilité des résultats de la modélisation avec facteurs associés.

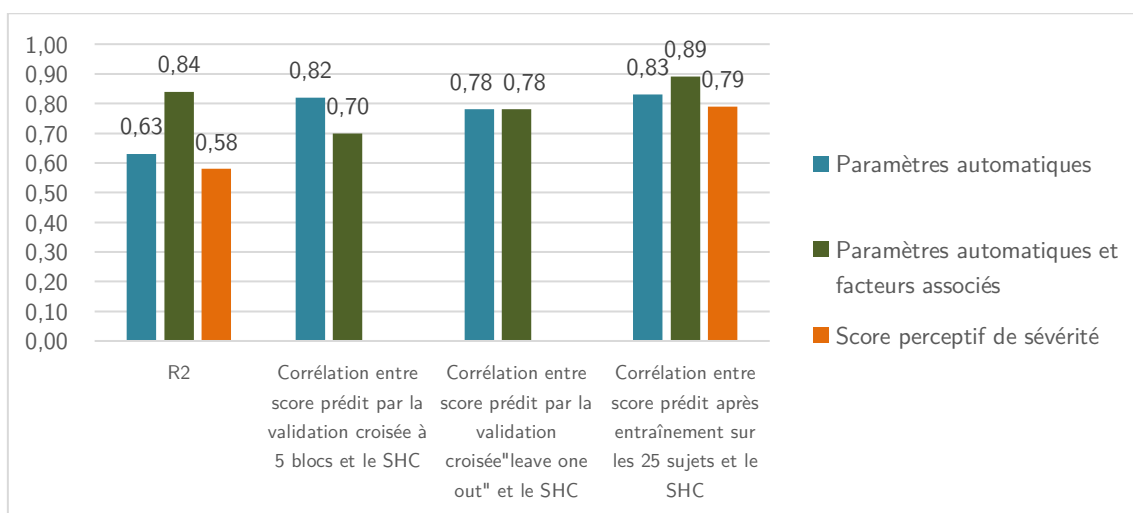


Figure 33 - Comparaison des performances entre modélisations du SHC

Dans tous les cas, l'utilisation des paramètres issus des analyses automatiques de la parole spontanée présente des résultats légèrement meilleurs que le score perceptif de sévérité du trouble de parole. Bien que ce dernier n'ait pas pour objectif de quantifier les altérations communicationnelles des patients à proprement parler, il reste la référence en clinique courante. L'utilisation des analyses automatiques de la parole spontanée présente donc un intérêt majeur pour prédire au mieux l'impact des troubles de parole sur la communication. La fiabilité des modélisations demande à être confirmée là encore par des études de plus grande ampleur.

Nos résultats sont finalement assez proches de ceux retrouvés par Eadie (Eadie et al., 2018), qui avait étudié le rôle des facteurs psychosociaux sur la participation de communication parmi des survivants de cancers ORL (dont 32 sujets soit 36,4 % étaient atteints de cancers de la cavité buccale ou de l'oropharynx / hypopharynx).

Nous retrouvons bien une influence de la sévérité du trouble de parole autoévaluée (*phihcp*), du support social perçu (*eccs1c*, *eccs1d*, *qfs52*, *qfs62*) et de la dépression et de l'anxiété perçues (*had2*, *hada*). Contrairement à Eadie (Eadie et al., 2018), nous ne retrouvons pas d'influence du statut cognitif ni du délai après traitement. Ceci peut s'expliquer par le fait, que dans notre étude, le statut cognitif est évalué au moyen d'un test alors que l'étude de Eadie (Eadie et al., 2018) consistait en une autoévaluation par les sujets de leur statut cognitif : il est donc possible que le ressenti du fonctionnement cognitif modifie davantage les habiletés de communication que le statut réel. Concernant le délai depuis traitement (correspondant au délai depuis diagnostic dans l'étude de Eadie (Eadie et al., 2018), diagnostic et traitement étant de toute façon proches dans le temps dans les pathologies cancérologiques), le délai moyen est moins élevé dans notre étude (7,27 ans contre 12,2 ans). Il est donc possible que la dynamique fonctionnelle de communication ait une importance différente selon la distance au traitement, avec un fonctionnement devenant de moins en moins satisfaisant.

Deux derniers points sont à analyser. Nous n'avons pas mesuré les capacités de résilience des sujets, pourtant facteur significatif chez Eadie (Eadie et al., 2018), ce qui permettrait peut-être d'ajuster nos résultats notamment quant au délai après traitement. Enfin, nous retrouvons un lien significatif entre communication et fonction d'odorat (plus précisément, la fonction d'alerte relative à la crainte d'un accident). Ce point n'avait pas été interrogé par Eadie (Eadie et al., 2018) dans son étude, dans laquelle la majorité des sujets avaient subi une laryngectomie totale (53,7 %), chirurgie qui abolit totalement les fonctions d'olfaction des patients traités. Il est malgré tout possible que les

crainces relatives à la perte de la fonction d'odorat soient malgré tout présentes dans la population des patients laryngectomisés.

4.3 Conclusion

Afin de pouvoir mesurer l'altération de communication chez les patients traités pour un cancer de la cavité buccale ou de l'oropharynx, nous avons développé un nouveau score de référence : le score holistique de communication (SHC). Ce score est volontairement holistique, dans le sens où il est élaboré à partir d'items de questionnaires complétés par les patients sur les différents aspects de la communication : symptomatologique, interactionnelle, pragmatique et psychoaffective.

Pour aboutir à ce score, 174 items relatifs à la communication ont été initialement collectés parmi les autoquestionnaires complétés par les patients (ECVB, DIP, PHI, EORTC QLQ-C35, EORTC H&N35, CHI).

Un comité de neuf experts dont les disciplines sont relatives à la communication (informatique et traitement du signal, sciences du langage, orthophonie et médecine) a permis de vérifier la validité d'apparence en deux temps. Dans un premier temps, une définition consensuelle de la communication orale en contexte cancérologique a été établie, et 83 items n'étant pas conformes à cette définition ont été exclus. Dans un second temps, les 91 items restants ont été proposés aux experts sous forme d'un sondage afin de déterminer si chacun devait être conservé ou non dans la construction du score. Les 47 items présentant des mesures d'agrément trop faibles ($ICVI \leq 0,777$, ou $Kappa \text{ d'agrément} < 0,81$) ont été exclus de la suite des analyses.

Puis, les quatre items redondants (corrélés à plus de 0,90) avec d'autres items retenus ont également été exclus.

Enfin, avec les 40 items restants, une analyse factorielle en principal-component factor a été menée pour aboutir au score holistique de communication constitué des 24 items dont le coefficient de régression sur le facteur 1 dans l'analyse factorielle n'était pas nul.

Le score holistique de communication de référence ayant été construit, nous avons cherché à mesurer l'altération de la communication au moyen du SHC par les paramètres automatiques retenus dans la section 3.2.1.3. Pour cela, la modélisation par régression LASSO présentait les meilleurs résultats en termes de prédiction (corrélation à 0,83 entre SHC prédit et SHC réel) et de généralisabilité ($r_s = 0,82$ entre SHC prédit par validation croisée à cinq blocs et SHC réel). Deux paramètres issus du domaine phonémique sont utilisés dans

la prédiction du SHC : le nombre d'occlusives reconnus par seconde et le nombre de sonantes reconnues par seconde. La formule de calcul du SHC est :

$$Y_{SHC} = 83,672 + (10,139 \times occs) + (10,891 \times sonants)$$

Enfin, dans une perspective plus globale de prise en compte des facteurs associés au trouble de parole ou à la communication selon le modèle de Wilson (Wilson & Cleary, 1995), nous avons modélisé le SHC au moyen d'une régression linéaire incluant à la fois les deux paramètres automatiques retenus suite à une première régression LASSO (*occs* et *sonants*) et les facteurs associés issus eux-mêmes d'une seconde régression LASSO (pour prédire le SHC). Ainsi, le SHC peut être prédit par les deux paramètres automatiques *occs* et *sonants* associés à deux items relatifs à la satisfaction de la fréquence et de la qualité des échanges avec les membres du cercle privé, à la satisfaction relative aux relations avec les membres de l'entourage familial et à la manière de gérer les aspects financiers et administratifs, à l'anxiété globale, à un item relatif au statut dépressif, à un item relatif à l'altération de la fonction d'alerte de l'odorat et enfin au handicap ressenti au quotidien quant au trouble de parole.

$$\begin{aligned} Y_{SHC} = & 73,74 + (9,10 \times occs) + (4,07 \times sonants) + (1,32 \times eccs1c) \\ & + (0,05 \times eccs1d) + (3,58 \times qfs52) + (1,56 \times qfs62) \\ & + (-5,22 \times had2) + (3,39 \times hada) + (-3,99 \times chiog2) \\ & + (-0,12 \times phihcp) \end{aligned}$$

Cette modélisation du SHC avec ces paramètres, bien que légèrement moins généralisable ($r_s = 0,70$ entre SHC prédit en validation croisée à cinq blocs et SHC réel), présente une prédiction encore meilleure ($r_s = 0,89$ entre SHC prédit et SHC réel) de l'altération de la communication chez des patients ayant été traités pour un cancer de la cavité buccale ou de l'oropharynx.

Conclusion

Conclusion générale

La mesure de l'altération de la communication est un enjeu crucial du suivi thérapeutique des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. Le peu d'outils disponibles permettant cette mesure limite la prise en compte des répercussions fonctionnelles et psychosociales des troubles de parole dans le suivi thérapeutique, pourtant au cœur des préoccupations des patients.

Nous avons choisi d'inscrire notre étude dans le cadre conceptuel proposé par Wilson (Wilson & Cleary, 1995), dont une synthèse se retrouve en Figure 34.

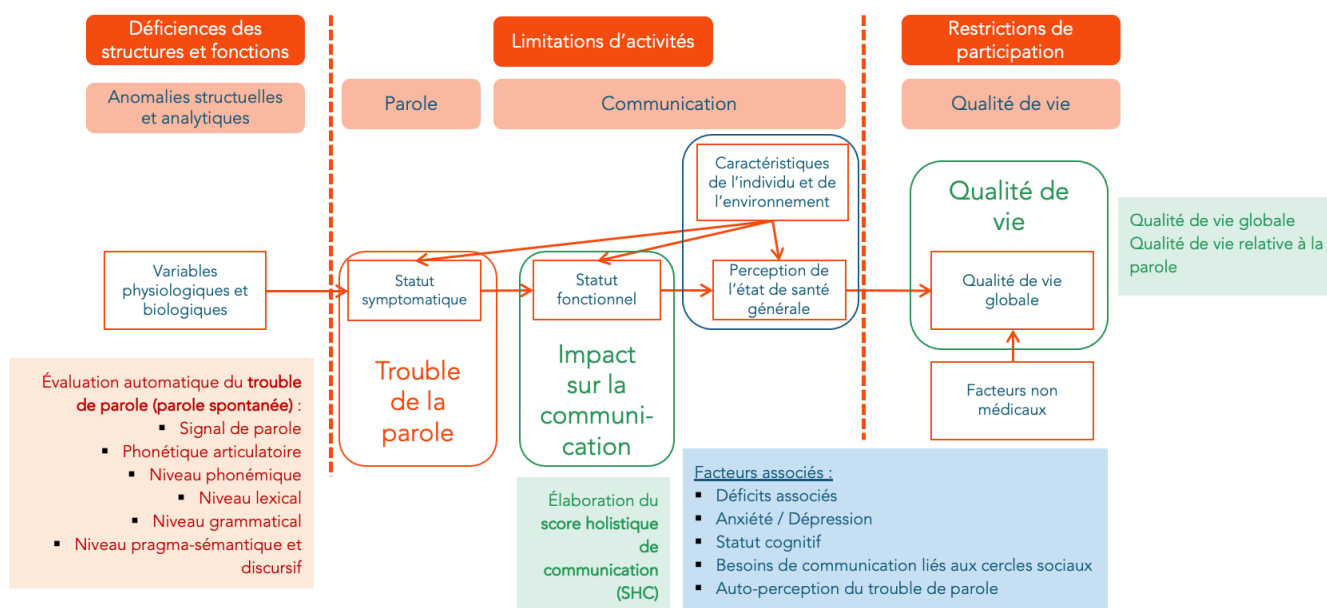


Figure 34 - Synthèse de notre étude et mise en lien avec le modèle de Wilson

Le design global de cette étude a fait l'objet d'une communication affichée concernant l'impact fonctionnel des troubles de la parole chez les patients traités pour un cancer oral ou oropharyngé lors du « 2020 Motor Speech Conference » (Balaguer et al., 2020d). Il a également été présenté en communication affichée lors de l'école d'été en Logopédie-Orthophonie (Balaguer et al., 2020c).

À cet effet, nous avons dû évaluer le trouble de parole, correspondant au statut symptomatique du modèle. Afin de nous affranchir des variabilités liées à l'évaluation perceptive, nous avons opté pour une évaluation automatique de la parole. Nous avons ainsi conduit une revue systématique de littérature qui a abouti à une publication (Balaguer et al., 2020e). Présentée dans le chapitre 1, elle visait à dresser un état des lieux des analyses acoustiques et automatiques utilisées pour mesurer les effets des cancers oraux et oropharyngés sur la production de parole. Deux conclusions principales ont émergé. D'une part, les mesures portaient principalement sur les unités segmentales de la parole (voyelles et consonnes). D'autre part, aucune étude n'avait été menée sur de la parole spontanée, autrement dit sur de la parole non contrainte.

Ces constats nous ont conduits à nous intéresser à l'analyse automatique de la parole spontanée, plus proche des productions quotidiennes des patients que la parole contrainte (telle que des répétitions de mots et des lectures de texte). De plus, cette parole spontanée est le support préférentiel pour récupérer des indices relatifs aux habiletés de communication car elle est produite directement en contexte communicationnel.

Nous avons alors conçu et constitué un nouveau corpus contenant des enregistrements de parole (notamment de parole spontanée) de 25 sujets. Afin de récupérer des informations relatives aux caractéristiques de l'individu et de l'environnement, à la perception de l'état de santé, ou encore à la qualité de vie du modèle de Wilson (Wilson & Cleary, 1995), les données individuelles, cliniques et de traitement, et celles issues de 11 autoquestionnaires et tests ont également été collectées. Ce corpus a été présenté dans le chapitre 2.

Les études de validation des autoquestionnaires CHI et PHI, utilisés dans notre étude, ont été publiées dans deux revues internationales à comité de lecture (Balaguer et al., 2020b, 2021a). Elles ont également été présentées sous forme de communication affichée lors de la 14^{ème} conférence francophone d'épidémiologie clinique (Balaguer et al., 2020a) et sous forme de communication orale lors du congrès mondial de l'« International Association of Logopedics and Phoniatrics (IALP 2019) » (Balaguer et al., 2019c), et du Congrès Inter-Universitaire du Collège des Centres de Formation Universitaire en Orthophonie (Balaguer, 2019).

Dans le chapitre 3, nous avons cherché à extraire des paramètres automatiques de la parole spontanée au moyen de divers outils à tous les niveaux linguistiques du modèle psycholinguistique de Caron (Devevey & Kunz, 2018; François & Nespoulous, 2011). Cent quarante-neuf paramètres ont été récupérés. Certains sont issus de l'analyse acoustique du signal de parole, relatifs aux mesures temporelles de durée, de débit et liés aux segments

vocaliques. L'utilisation d'un système de reconnaissance de la parole de type TDNN factorisé – HMM a permis d'extraire des paramètres phonémiques et lexicaux. Les sorties lexicales ont fait l'objet d'analyses complémentaires au niveau grammatical, thématique et de sentiment. Une sélection a ensuite été conduite pour aboutir à un nombre final de 75 paramètres : un consensus entre experts a permis de retenir les paramètres pertinents et une analyse statistique des corrélations inter-paramètres a permis d'éliminer les items redondants (corrélés à plus de 0,90 avec un autre item).

Devant le manque d'outils disponibles en cancérologie ORL, nous avons ensuite élaboré un score de référence permettant une mesure des capacités de communication présenté en section 4.1 : le score holistique de communication (ou SHC). À partir de 174 items relatifs à la communication issus des autoquestionnaires, un comité d'experts a permis de sélectionner les items pertinents à la suite d'une définition de la communication dans notre contexte d'étude, ainsi que d'une enquête en ligne ayant permis de retenir les items les plus consensuels (au moyen de scores élevés à l'I-CVI ou au kappa d'agrément). Les items trop redondants ont été exclus, sur les mêmes critères que lors de la sélection des paramètres automatiques. Enfin, une analyse factorielle en principal-component factor a permis d'aboutir au SHC (dont l'étendue dans notre corpus va de 77,81 à 114,44 – plus le score est faible, plus la communication est rapportée comme altérée par le sujet). Ce SHC a un véritable intérêt clinique, car il permet d'aboutir à un score global et exhaustif à partir de seulement 24 items, là où précédemment de multiples questionnaires étaient requis – questionnaires n'étant pas tous validés en cancérologie. Il reste encore à en vérifier la validité externe en proposant les items en jeu dans la construction du SHC à un échantillon de nouveaux patients, en deux temps différents permettant également d'en évaluer la fiabilité temporelle.

Ce travail a conduit à la présentation de deux communications affichées. La construction du score holistique de communication a ainsi été présentée à la 15^{ème} conférence francophone d'épidémiologie clinique et aux 28^{èmes} journées des statisticiens des centres de lutte contre le cancer (Balaguer et al., 2021b). Il a également été présenté lors du séminaire 2021 de l'Association Francophone de la Communication Parlée (ACFP) (Balaguer et al., 2021c).

Dans la section 4.2, nous avons cherché à mesurer l'altération de la communication au moyen des paramètres issus des analyses automatiques de la parole spontanée. Une analyse en régression LASSO nous a conduits à une prédiction du SHC au moyen de seulement deux paramètres (le nombre d'occlusives par seconde et le nombre de sonantes par seconde), présentant une corrélation à 0,83 entre le SHC réel (« vérité terrain ») et le SHC prédit par la

modélisation. Les meilleurs paramètres prédicteurs de l'altération de la communication sont donc deux paramètres au niveau phonémique. En termes cliniques, la mesure de l'altération de la communication permet ainsi de s'affranchir de la complétion de quatre questionnaires dont sont issus les 24 items retenus pour la construction du SHC chez des patients souvent fatigables : l'altération de communication pouvant être prédite uniquement au moyen de l'enregistrement et de l'analyse d'un entretien semi-dirigé avec le patient.

Conformément au modèle de Wilson (Wilson & Cleary, 1995), nous avons ensuite inclus dans notre modélisation prédictive les facteurs associés au trouble de parole ou à la communication. L'inclusion de ces facteurs associés aboutit à une corrélation encore plus élevée ($r_s = 0,89$) entre le SHC réel et le SHC prédit au moyen des deux paramètres automatiques précédemment retenus, ainsi qu'aux facteurs retenus par régression LASSO (items relatifs à la constitution des cercles sociaux, l'état anxiodépressif, la fonction d'alerte liée à l'altération de l'odorat et le handicap ressenti quant au trouble de parole). Les performances en termes de généralisabilité sont moindres lorsque les facteurs associés sont inclus mais restent néanmoins bonnes ($r_s = 0,70$).

Enfin, dans le modèle de Wilson (Wilson & Cleary, 1995), l'altération de la communication est elle-même un facteur associé à la qualité de vie. Ainsi, au-delà de l'étude de la dimension fonctionnelle, la démarche que nous avons entreprise concernant la modélisation de l'altération des capacités de communication peut également s'étendre à l'étude de la qualité de vie. Nous avons alors cherché à mener une étude préliminaire sur les données de notre corpus concernant la prédiction de la qualité de vie.

Nous retrouvons que le score de qualité de vie globale n'est que faiblement prédit par le SHC seul ($r_s = 0,21$; $R^2 = 0,05$), mais la corrélation est meilleure quand il est associé aux deux paramètres issus des analyses automatiques sur le signal de parole ($r_s = 0,45$; $R^2 = 0,12$). L'inclusion des facteurs associés n'améliore pas les performances ($r_s = 0,36$; $R^2 = 0,61$).

En revanche, la qualité de vie relative au fonctionnement social est mal prédite par le SHC seul ($r_s = 0,05$; $R^2 < 0,01$) ou associé aux paramètres automatiques ($r_s = 0,11$; $R^2 = 0,03$), mais les facteurs associés améliorent nettement les performances prédictives ($r_s = 0,61$; $R^2 = 0,41$). Les facteurs associés que nous avons ciblés permettent ainsi une meilleure prédiction de la qualité de vie liée au fonctionnement social, mais avec une limite liée à la présence de multicollinéarité dans le modèle avec les paramètres issus des analyses automatiques. Il est donc possible que des indices relatifs à la qualité de vie se trouvent dans des niveaux linguistiques supérieurs au niveau

phonémique, ce que notre système de reconnaissance n’a pas pu mettre en évidence.

Perspectives

Notre étude permet d’ouvrir de multiples perspectives. Un premier axe concerne la taille de l’échantillon. Un deuxième axe s’intéresse à la prise en compte des facteurs associés à la parole ou à la communication dans la détermination des paramètres qui prédisent le mieux l’impact sur la communication et la qualité de vie. Un troisième axe relève des perspectives d’optimisation des analyses automatiques. Un dernier axe porte sur l’applicabilité clinique de notre travail.

Pourquoi élargir la taille de l’échantillon de l’étude ?

Un premier point concerne la nécessité de valider nos résultats sur un échantillon de plus grande taille. Notre étude ayant été menée sur 25 sujets, l’enrichissement du corpus permettrait de préciser et de compléter nos analyses exploratoires. Cette augmentation de la taille de l’échantillon autoriserait alors une **plus grande généralisabilité des résultats**.

Elle permettrait également de réaliser des **analyses complémentaires** en sous-groupes.

En effet, notre revue systématique de littérature a montré que les paramètres acoustiques et automatiques utilisés dans l’étude de la parole étaient différents selon la *localisation de la tumeur* (Balaguer et al., 2020e). Une analyse selon la localisation tumorale (localisation orale *vs.* localisation oropharyngée) mettrait peut-être ainsi en évidence des paramètres prédictifs différents, témoins d’une communication dégradée différemment selon le lieu de l’atteinte.

Une autre analyse stratifiée pourrait être réalisée selon la *taille de la tumeur*, qui est un facteur altérant la qualité de la parole (Balaguer et al., 2019a; Borggreven et al., 2007; Colangelo et al., 2000). D’un côté, des paramètres acoustiques et automatiques différents sont utilisés dans l’étude de la parole selon la taille de la tumeur (voir section 1.3.1). D’un autre côté, l’intelligibilité est plus dégradée après traitement de tumeur de gros volume (T3 ou T4). Les paramètres prédisant l’altération de la communication pourraient être différents selon ce critère, les besoins de communication étant différents selon le volume anatomique atteint :

- Pour les tumeurs de grande taille qui altèrent fortement l'intelligibilité (Balaguer et al., 2019b), il semble pertinent de se demander si l'altération de la communication est principalement liée à des dégradations phonémiques, comme le montre notre étude ;
- Pour les tumeurs de petite taille qui préservent mieux l'intelligibilité, des paramètres différents de plus haut niveau linguistique pourraient alors être mis en évidence. En effet, les systèmes de reconnaissance de la parole présenteraient de meilleures performances en raison d'une plus grande proximité entre les modèles acoustiques (entraînés sur de la parole typique) et la parole dont l'intelligibilité est peu dégradée.

Ainsi, la détermination de groupes selon l'intelligibilité de la parole est une autre piste à explorer. Des études récentes ont montré de bonnes performances de systèmes automatiques basés sur des X-vecteurs dans la prédiction de l'intelligibilité après cancer ORL (Quintas et al., 2020). Une étude sur l'utilisation de telles mesures automatiques dans la détermination de valeurs seuils selon les niveaux d'intelligibilité permettrait de constituer des groupes d'analyses différents. Ils pourraient mettre en évidence des paramètres automatiques différents dans la prédiction de la communication ou de la qualité de vie, selon le niveau d'intelligibilité de la parole. Ils permettraient ainsi de mettre en lumière des besoins de communication différents selon le degré d'atteinte.

Comment mieux prendre en compte les liens entre parole, communication, facteurs associés et qualité de vie ?

Dans tous les cas, la détermination des meilleurs paramètres prédictifs requiert de s'intéresser plus spécifiquement aux liens existants entre trouble de parole, communication, facteurs associés et qualité de vie.

Pour réaliser nos modélisations prédictives, en l'absence d'outil validé de mesure de l'altération de la communication, nous avons construit **un score holistique de communication (ou SHC) à partir d'autoquestionnaires**. Ce score est prédit par deux paramètres au niveau phonémique (le nombre d'occlusives et de sonantes par seconde). Ce lien phonémique / communication peut s'expliquer par le fait que la gêne communicationnelle mesurée par le SHC est autorapportée par les patients, les capacités d'articulation étant la manifestation la plus flagrante du trouble de parole. Or, dans le cadre d'une mesure des capacités de communication, l'interlocuteur humain est particulièrement à considérer car la communication n'est efficace que si le message est certes correctement transmis mais aussi correctement reçu.

L'inclusion au SHC d'une mesure externe des capacités de communication par un auditeur permettrait de prendre en compte d'autres dimensions liées plus globalement à l'impact du trouble sur la compréhensibilité de la parole (par un auditeur) et par effet ricochet sur la communication et la qualité de vie. Elles permettraient aussi de préciser quelles dimensions seraient une aide à la compréhensibilité du message, en contexte de parole dégradée.

Une piste d'exploration des mécanismes de compréhensibilité du message est liée à la communication non-verbale, et au caractère augmentatif qu'elle peut avoir. **L'intégration d'indices non verbaux**, non captés par les analyses acoustiques, à la modélisation de l'altération de la communication donnerait peut-être accès à de nouveaux paramètres complémentaires. Ce type d'étude pourrait être proposé par couplage de l'analyse acoustique du signal de parole avec une analyse du signal visuel lors de la production de parole spontanée (par exemple lors d'un entretien semi-dirigé comme nous l'avons proposé dans notre étude). Des études ont montré que les gestes de pointage (Dhingra et al., 2020) ou les expressions faciales (Siddiqi et al., 2014) peuvent être précisément capturés et analysés au moyen de caméras couleur-profondeur, type RGB-D (Red Blue Green Depth). L'exploration des liens entre ces indices non-verbaux et les facteurs associés à la communication voire à la qualité de vie reste à étudier, et permettrait d'en compléter l'analyse. Par exemple, la résilience étant liée aux capacités de communication (Eadie et al., 2018), les sujets les plus résilients utilisent peut-être différemment les gestes et mimiques.

Mais **d'autres indices pourraient être étudiés plus précisément** dans la modélisation de l'altération de la communication ou de la qualité de vie.

Premièrement, le *délai depuis le traitement* modifie la dynamique fonctionnelle et psychosociale des sujets, car les besoins évoluent au cours du suivi et dans le temps. Notamment, la radiothérapie induit des effets secondaires après de nombreuses années (Kraaijenga et al., 2016), modifiant le ressenti et les besoins de communication à distance du traitement. Ensuite, l'influence de l'olfaction-gustation que nous retrouvons dans notre modélisation est peut-être liée à une dynamique globale, positive ou négative, dans laquelle le patient s'inscrit et qui peut influencer sa manière d'appréhender ses activités quotidiennes (corrélation entre CHI/olfaction-gustation et (i) HAD/anxiété : $r_s = 0,44$; $p = 0,03$; (ii) HAD/dépression : $r_s = 0,35$; $p = 0,09$). D'ailleurs, la corrélation est plus élevée ($r_s = 0,47$; $p = 0,02$) entre le score global d'anxiété et l'item *chiog4* « Vous sentez-vous frustré à cause de vos problèmes de goût ou d'odorat ? ».

Ainsi, il reste encore à explorer de façon plus précise *les liens existants entre les facteurs associés au trouble de parole et à la communication, et à la*

qualité de vie. Si le modèle proposé par Wilson (Wilson & Cleary, 1995) permet de cibler les facteurs en jeu dans la dynamique fonctionnelle et psychosociale de façon globale, il est probable que les liens entre chacun de ces facteurs (trouble de parole, facteurs associés, communication et qualité de vie) ne soient pas aussi linéaires ou directs en cancérologie ORL. Une modélisation théorique spécifique aux patients traités pour un cancer de la cavité buccale ou de l'oropharynx permettrait d'expliquer plus spécifiquement les relations entre chaque niveau, et de mieux expliquer et prédire la qualité de vie.

Comment optimiser la mesure automatique de la parole pathologique ?

Mais la prédiction de l'altération de la communication ou de la qualité de vie est également étroitement liée à la qualité de la mesure du trouble de parole.

La recherche de mesures issues de l'ensemble des niveaux psycholinguistiques nous conduit à nous questionner sur les mesures de bas et haut niveau utilisables dans l'évaluation de la parole pathologique.

Des mesures de bas niveau régulièrement étudiées selon la revue systématique de littérature pourraient être envisagées comme nouvelles variables d'intérêt. Les valeurs formantiques, l'aire du triangle vocalique, ou encore les moments spectraux sur les consonnes pourraient être inclus dans les modélisations en tant que mesures segmentales complémentaires du trouble de parole. Mais les analyses de ces paramètres sur la parole spontanée sont complexes, en raison de contextes de production phonémique non stables. Il serait alors pertinent d'associer dans nos modélisations d'autres paramètres extraits de tâches de parole plus contrainte (comme de la lecture de texte). La recherche de liens entre dégradations phonétiques et phonémiques et déficits analytiques au niveau oral ou oropharyngé – correspondant au niveau le plus antérieur du modèle de Wilson (Wilson & Cleary, 1995) – permettrait de fournir de nouveaux éléments de compréhension des mécanismes en jeu dans la production de parole pathologique, de stratégies de compensation du trouble de parole et de l'impact des déficits de bas niveau sur la dynamique communicationnelle. C'est notamment un des objectifs du projet de recherche DAPADAF-E²³ (2019-2023) dont je suis le coordonnateur scientifique.

²³ <https://www.irit.fr/SAMOVA/site/projects/dapadaf-e/>, consulté le 06/06/2021

Les choix de ces mesures est étroitement lié à la *langue parlée*. Chaque langue présente en effet des spécificités sur la proportion de nasalité (Mayo & Mayo, 2011) ou sur les variations prosodiques au niveau de la fréquence fondamentale (Kuo et al., 2008) notamment. Les mesures de bas niveau doivent donc être adaptées aux caractéristiques de la langue (par exemple, aux langues tonales) pour cibler les paramètres les plus pertinents prédisant l’altération de la communication.

Parmi les analyses automatiques menées sur les **paramètres de plus haut niveau**, nous n’avons pas obtenu de résultat probant quant aux mesures aux niveaux pragma-sémantique et discursif.

L’analyse thématique réalisée de façon automatique est rendue difficile par la compréhension des concepts sous-jacents permettant l’élaboration de champs sémantiques. Ainsi, ces données sont habituellement traitées de façon « manuelle » (Missire, 2013). Toutefois, certaines méthodes issues du traitement automatique des langues pourraient permettre de pallier ces biais. C’est le cas des algorithmes *topic models*, qui peuvent modéliser un texte (ou les sorties d’un système de reconnaissance de la parole) selon des méthodes probabilistes regroupant des extraits de texte en fonction de leur similarité (Xun et al., 2017). En français, des approches par recherche de clusters sémantiques à partir du TLFi (version informatisée du dictionnaire « Trésor de la Langue Française ») ont été tentées. Le regroupement sémantique à l’aide de critères morphologiques et lexico-sémantiques au moyen de cette technique présente des résultats encourageants (Gala et al., 2011).

À l’instar du niveau thématique, *l’analyse de sentiment* que nous avons menée n’a pas permis de dégager de paramètre prédictif de l’altération fonctionnelle ou psychosociale. Pourtant, un état dépressif même non sévère peut avoir des conséquences sur les paramètres acoustiques de la parole (Albuquerque et al., 2021). D’autres approches automatiques basées sur l’analyse des paramètres de polarité mais aussi de la ponctuation ou de l’utilisation d’émoticones dans des commentaires sur des forums communautaires en ligne de soutien après cancer ont présenté des résultats prometteurs, notamment sur les dernières phrases publiées sur ces forums (Biyani et al., 2013). Ainsi, il semble que la localisation temporelle du segment d’étude soit un paramètre à considérer, tout comme la longueur de l’énoncé d’analyse plus réduite que celle que nous avons envisagée. En effet, la longueur moyenne d’un énoncé écrit est de 13,2 mots (Dewaele, 2000). De plus, la prise en compte de la ponctuation dans la détermination de la polarité renvoie aux aspects prosodiques de modalité ou de focus notamment : une approche combinée entre la variation de la fréquence fondamentale et les sorties lexicales du segment à considérer serait à explorer.

Toutefois, si l’optimisation des systèmes d’analyse automatique pragma-sémantique et discursif sont un axe d’amélioration, il semble essentiel de considérer également **l’amélioration des performances du système de reconnaissance automatique**. Ce sont en effet sur les sorties de ce système que se basent les outils d’analyses de plus haut niveau.

Le système de reconnaissance automatique de la parole que nous avons utilisé est entraîné sur de la parole saine. Dans ce cas, il compare des observations acoustiques issues de parole pathologique dégradée à des modèles acoustiques typiques : le système conclut sur un écart entre la parole pathologique et une norme établie sur la parole saine, comme le ferait une oreille humaine standard (mais avec une fiabilité liée au caractère automatique palliant les variabilités perceptives que nous retrouvons section 2.5.2.2 - Tableau 7). Nous pouvons alors nous demander si l’entraînement du système sur de la parole pathologique permettrait d’obtenir des modèles acoustiques plus adaptés et spécifiques, présentant des taux d’erreurs mots plus faibles.

Ce type d’analyse serait rendu possible par une utilisation de corpus de grande taille pour entraîner des modèles acoustiques relativement plus stables et débarrassés du surcroît de la variabilité liée au caractère pathologique de la parole. Aucun corpus de parole cancérologique francophone de grande taille n’existant à ce jour, des techniques d’apprentissage par transfert (ou *transfer learning*) permettent d’adapter des modèles de parole typique à de nouveaux corpus sur relativement peu de données (Gelin et al., 2021; Wang & Zheng, 2015). Concrètement, il serait possible d’adapter le système de reconnaissance de la parole actuel sur les tâches de parole non utilisées de notre corpus (phrases porteuses des consonnes, lecture de texte, lecture de phrases et répétitions de pseudo-mots). La durée totale de ce set d’entraînement est d’environ une heure 30 minutes.

Mais de nouveaux systèmes ont été développés très récemment, et présentent des performances prometteuses en termes de reconnaissance de parole : les architectures LAS (Listen, Attend and Spell) (Chan et al., 2016), ou de type *Transformers* (Karita et al., 2019; Lu et al., 2020). Ces systèmes ont été adaptés à la parole non typique par Gelin (Gelin et al., 2021), en l’occurrence la parole d’enfants. Une étude de ces nouveaux systèmes sur la parole non typique après traitement d’un cancer oral ou oropharyngé serait pertinente pour en étudier les performances.

Ces différentes pistes de réflexion permettraient dès lors d’obtenir des sorties du système plus proches des éléments linguistiques réellement produits. Ceci ouvrirait alors d’autres perspectives complémentaires, en changeant de paradigme. Si l’analyse que nous avons menée sur les éléments reconnus autorise une comparaison à la production standard, l’analyse des éléments

produits serait plus précise en termes d'analyse de contenu. Elle permettrait également de trouver de nouveaux indices à d'autres niveaux linguistiques dans le signal de parole. Enfin, elle contribuerait à l'étude de l'adaptation de la production de parole au déficit : en cas de traitement cancérologique, utilise-t-on davantage certains types de phonèmes ou des enchaînements phonémiques plus faciles à produire, le sentiment se dégageant du discours est-il plus franc... ?

Comment rendre ces résultats applicables en clinique courante ?

De façon plus générale, la méthodologie prédictive que nous avons développée pourrait être **adaptée à d'autres contextes pathologiques** concernés par l'altération de la communication et la qualité de vie qui en est consécutive. C'est le cas des tumeurs laryngées qui dégradent ou abolissent la source vocale : les patients laryngectomisés rapportent une dégradation de leurs capacités de communication et de leur qualité de vie (Sharpe et al., 2019). Toujours en ORL, les patients avec paralysie faciale présentent des modifications acoustiques de leur parole (Robert et al., 2013) et indiquent une dégradation de leurs capacités de communication (Kim et al., 2018). En neurologie également, les patients présentant des apraxies de la parole rapportent une plainte fonctionnelle sur leur communication (Utianski et al., 2020). Dans tous ces cas, une mesure automatique de l'altération de la communication aiderait à mieux adapter les stratégies thérapeutiques des patients, à moyen et long terme.

À cet effet, le **développement d'un dispositif mobile** permettant l'enregistrement et l'analyse de la parole spontanée pourrait être développé pour faciliter l'utilisation de l'analyse automatique de la parole spontanée en pratique clinique courante. L'objectif serait d'obtenir le SHC immédiatement après un entretien mené par un thérapeute auprès d'un patient. Actuellement, une expérimentation de ce type de dispositif est menée par l'IRIT et le CHU de Toulouse. Financé par Toulouse Tech Transfer, ce dispositif permet de prédire en quasi-temps réel l'intelligibilité de la parole d'un patient en utilisant une application développée sur tablette et connectée au réseau du CHU. En l'adaptant à la mesure de la communication, ce type de dispositif fournirait aux cliniciens une nouvelle métrique valide et fiable dès l'évaluation initiale. Une meilleure adaptation précoce des stratégies thérapeutiques prenant en compte la communication, et un meilleur suivi de l'évolution des patients en découleraient.

Bibliographie

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. d., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofo, P. B., Schraub, S., Sneeuw, K., Sullivan, M., & Takeda, F. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology. *JNCI Journal of the National Cancer Institute*, 85(5), 365–376. <https://doi.org/10.1093/jnci/85.5.365>
- Acher, A., Perrier, P., Savariaux, C., & Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. *Clinical Linguistics & Phonetics*, 28(4), 241–256. <https://doi.org/10.3109/02699206.2013.802015>
- Acock, A. C. (2018). *A Gentle Introduction to Stata- 6th Edition*. Stata Press.
- Albuquerque, L., Valente, A. R. S., Teixeira, A., Figueiredo, D., Sa-Couto, P., & Oliveira, C. (2021). Association between acoustic speech features and non-severe levels of anxiety and depression symptoms across lifespan. *PLOS ONE*, 16(4), e0248842. <https://doi.org/10.1371/journal.pone.0248842>
- 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>
- Auzou, P., & Rolland-Monnoury, V. (2019). *BECD : Batterie d'Évaluation Clinique de la Dysarthrie* (Ortho Édit).
- Balaguer, M. (2019). Le Carcinologic Handicap Index (CHI) : questionnaire d'autoévaluation du handicap à destination des patients traités pour cancer des VADS. *1er Congrès Inter-Universitaire CCFUO*.
- Balaguer, M., Boissguérin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019a). Factors influencing intelligibility and severity of chronic speech disorders of patients treated for oral or oropharyngeal cancer. *European Archives of Oto-Rhino-Laryngology*, 276(6), 1767–1774. <https://doi.org/10.1007/s00405-019-05397-6>
- Balaguer, M., Boissguérin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019b). Assessment of impairment of intelligibility and of speech signal after oral cavity and oropharynx cancer. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 136(5), 355–359. <https://doi.org/10.1016/j.anorl.2019.05.012>
- Balaguer, M., Champenois, M., Farinas, J., Pinquier, J., & Woisard, V. (2020a). Validation des dimensions complémentaires du « Carcinologic Handicap Index » et étude de sa capacité à déterminer les besoins des patients. *Revue d'Épidémiologie et de Santé Publique*, 68, S133. <https://doi.org/10.1016/j.respe.2020.03.067>

- Balaguer, M., Champenois, M., Farinas, J., Pinquier, J., & Woisard, V. (2021a). 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*, 278(4), 1159–1169. <https://doi.org/10.1007/s00405-020-06201-6>
- Balaguer, M., Farinas, J., Fichaux-Bourin, P., Puech, M., Pinquier, J., & Woisard, V. (2020b). 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 Phoniatri et Logopaedica*, 72(6), 464–477. <https://doi.org/10.1159/000503448>
- Balaguer, M., Farinas, J., Pinquier, J., & Woisard, V. (2019c). Construction of the automatic Carcinologic Speech Severity Index (C2SI) score. *31st World Congress of the International Association of Logopedics and Phoniatrics (IALP 2019)*. <https://hal.archives-ouvertes.fr/hal-02977799>
- Balaguer, M., Farinas, J., Woisard, V., & Pinquier, J. (2020c). Impact fonctionnel des troubles de la parole évalués par une mesure automatique sur les actes de communication quotidiens chez les patients traités pour un cancer de la cavité buccale ou de l'oropharynx. *École d'été En Orthophonie-Logopédie (EDE)*. <https://hal.archives-ouvertes.fr/hal-02505393>
- 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*. <https://doi.org/10.1016/j.anorl.2017.06.010>
- Balaguer, M., Pinquier, J., Farinas, J., Lepage, B., & Woisard, V. (2021b). Construction d'un index holistique d'impact sur la communication des troubles de la parole chez des patients traités pour un cancer oral ou oropharyngé. *15ème Conférence Francophone d'Épidémiologie CLINique (EPICLIN 2021) et 28èmes Journées Des Statisticiens Des Centres de Lutte Contre Le Cancer*. <https://hal.archives-ouvertes.fr/hal-03258617>
- Balaguer, M., Pinquier, J., Farinas, J., Lepage, B., & Woisard, V. (2021c). Construction d'un index holistique d'impact sur la communication des troubles de la parole chez des patients traités pour un cancer oral ou oropharyngé. *Séminaire AFCP 2021, Thème: Phonétique Clinique*. <http://www.afcp-parole.org/seminaire-afcp-phonetique-clinique-27-mai-2021/>
- Balaguer, M., Pommée, T., Farinas, J., Pinquier, J., & Woisard, V. (2020d). Functional impact of speech disorders in patients treated for oral or oropharyngeal cancer, assessed by perceptual and automatic measurements. *2020 Motor Speech Conference*. <https://doi.org/10.1159/000503448>
- Balaguer, M., Pommée, T., Farinas, J., Pinquier, J., Woisard, V., & Speyer, R. (2020e). 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>

- Barrett, W. L., Gluckman, J. L., Wilson, K. M., & Gleich, L. L. (2004). A comparison of treatments of squamous cell carcinoma of the base of tongue: surgical resection combined with external radiation therapy, external radiation therapy alone, and external radiation therapy combined with interstitial radiation. *Brachytherapy*, 3(4), 240–245. <https://doi.org/10.1016/j.brachy.2004.09.002>
- Baylor, C. R., Yorkston, K. M., Eadie, T. L., Miller, R. M., & Amtmann, D. (2009). Developing the Communicative Participation Item Bank: Rasch Analysis Results From a Spasmodic Dysphonia Sample. *Journal of Speech, Language, and Hearing Research*, 52(5), 1302–1320. [https://doi.org/10.1044/1092-4388\(2009/07-0275\)](https://doi.org/10.1044/1092-4388(2009/07-0275))
- Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The communicative participation item bank (CPIB): Item bank alibration 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))
- Berisha, V., Utianski, R., & Liss, J. (2013). Towards a clinical tool for automatic intelligibility assessment. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2825–2828. <https://doi.org/10.1109/ICASSP.2013.6638172>
- Bidart, C., Degenne, A., & Grosetti, M. (2011). La vie en Réseau. Dynamique des relations sociales. In *Presses Universitaires de France: Vol. "Le Lien S.* <https://doi.org/10.3917/puf.bidar.2011.01>.
- Biyani, P., Caragea, C., Mitra, P., Zhou, C., Yen, J., Greer, G. E., & Portier, K. (2013). Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, December 2014*, 413–417. <https://doi.org/10.1145/2492517.2492606>
- Bjordal, K., Hammerlid, E., Ahlner-Elmqvist, M., de Graeff, A., Boysen, M., Evensen, J. F., Biörklund, A., de Leeuw, J. R. J., Fayers, P. M., Jannert, M., Westin, T., & Kaasa, S. (1999). Quality of Life in Head and Neck Cancer Patients: Validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-H&N35. *Journal of Clinical Oncology*, 17(3), 1008–1008. <https://doi.org/10.1200/JCO.1999.17.3.1008>
- Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with Praat. *Glott International*, 5(9–10), 341–347.
- Böhm, N., Knipfer, C., Maier, A., Bocklet, T., Rohde, M., Neukam, F., Stelzle, F., & Schuster, M. (2016). Sprechqualität und psychische Beeinträchtigung nach der Therapie von Mundhöhlentumoren. *Laryngo-Rhino-Otologie*, 95(09), 610–619. <https://doi.org/10.1055/s-0042-102256>

- Bolt, S., Eadie, T., Yorkston, K., Baylor, C., & Amtmann, D. (2016). Variables associated with communicative participation after head and neck cancer. *JAMA Otolaryngology-Head & Neck Surgery*, *142*(12), 1145–1151. <https://doi.org/10.1001/jamaoto.2016.1198>
- Borggreven, P. A., Verdonck-De Leeuw, I. M., Muller, M. J., Heiligers, M. L. C. H., De Bree, R., Aaronson, N. K., & Leemans, C. R. (2007). Quality of life and functional status in patients with cancer of the oral cavity and oropharynx: Pretreatment values of a prospective study. *European Archives of Oto-Rhino-Laryngology*, *264*(6), 651–657. <https://doi.org/10.1007/s00405-007-0249-5>
- Bruchon-Schweitzer, M., Rascle, N., Gelie, F., Fortier, C., Sifakis, Y., & Constant, A. (2003). Le questionnaire de soutien social de Sarason (SSQ6). Une adaptation française. *Psychologie Française*, *48*(3), 41–53.
- Cai, L. (2012). Latent variable modeling. *Shanghai Archives of Psychiatry*, *24*(2), 118–120. <https://doi.org/10.3969/j.issn.1002-0829.2012.02.010>
- Cameron, C., & Trivedi, P. (2005). *Microeconometrics : Methods And Applications* (Cambridge).
- Carmichael, J. (2007). Introducing Objective Acoustic Metrics for the Frenchay Dysarthria Assessment Procedure. *University of Sheffield*, December. <https://doi.org/10.13140/RG.2.1.4037.8960>
- Carreira-Perpinán, M. (1997). A review of dimension reduction techniques. *Department of Computer Science. University of Sheffield. Tech. Rep. CS-96-09*, 1–69. <http://www.pca.narod.ru/DimensionReductionBrifReview.pdf>
- Chadeau-Hyam, M., Campanella, G., Jombart, T., Bottolo, L., Portengen, L., Vineis, P., Liqueur, B., & Vermeulen, R. C. H. (2013). Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. *Environmental and Molecular Mutagenesis*, *54*(7), 542–557. <https://doi.org/10.1002/em.21797>
- Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2016). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016-May, 4960–4964. <https://doi.org/10.1109/ICASSP.2016.7472621>
- Chen, S.-C., Yu, P.-J., Hong, M.-Y., Chen, M.-H., Chu, P.-Y., Chen, Y.-J., Wang, C.-P., & Lai, Y.-H. (2015). Communication dysfunction, body image, and symptom severity in postoperative head and neck cancer patients: factors associated with the amount of speaking after treatment. *Supportive Care in Cancer*, *23*(8), 2375–2382. <https://doi.org/10.1007/s00520-014-2587-3>
- Chung, E.-J., Lee, D., Kang, H., Park, M., Chung, C., & Rho, Y. (2011). Prospective speech outcome study in patients with soft palate reconstruction in tonsillar cancer. *Oral Oncology*, *47*(10), 988–992. <https://doi.org/10.1016/j.oraloncology.2011.07.022>

- Clapham, R. P., Van Der Molen, L., Van Sonb, R. J. J. H., Van Den Brekelb, M., & Hilgersb, F. J. M. (2012). NKI-CCRT Corpus - Speech intelligibility before and after advanced head and neck cancer treated with concomitant chemoradiotherapy. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 3350–3355.
- Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M. M. (1997). Social ties and susceptibility to the common cold. *JAMA*, 277(24), 1940–1944. <http://www.ncbi.nlm.nih.gov/pubmed/9200634>
- Colangelo, L. A., Logemann, J. A., & Rademaker, A. W. (2000). Tumor Size and Pretreatment Speech and Swallowing in Patients with Resectable Tumors. *Otolaryngology–Head and Neck Surgery*, 122(5), 653–661. [https://doi.org/10.1016/S0194-5998\(00\)70191-4](https://doi.org/10.1016/S0194-5998(00)70191-4)
- Coma, E., Ferran, M., Méndez, L., Iglesias, B., Fina, F., & Medina, M. (2013). Creation of a synthetic indicator of quality of care as a clinical management standard in primary care. *SpringerPlus*, 2(1), 51. <https://doi.org/10.1186/2193-1801-2-51>
- Cunningham, P. (2014). Dimension Reduction. In *Machine Learning Techniques for Multimedia* (pp. 91–112). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75171-7_4
- Danker, H., Wollbrück, D., Singer, S., Fuchs, M., Brähler, E., & Meyer, A. (2010). Social withdrawal after laryngectomy. *European Archives of Oto-Rhino-Laryngology*, 267(4), 593–600. <https://doi.org/10.1007/s00405-009-1087-4>
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62. [https://doi.org/10.1016/S0095-4470\(19\)30776-4](https://doi.org/10.1016/S0095-4470(19)30776-4)
- 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 Bruijn, M., ten Bosch, L., Kuik, D. J., Langendijk, J. A., Leemans, C. R., & Leeuw, I. V. (2011). Artificial neural network analysis to assess hypernasality in patients treated for oral or oropharyngeal cancer. *Logopedics Phoniatrics Vocology*, 36(4), 168–174. <https://doi.org/10.3109/14015439.2011.606227>
- de Carvalho-Teles, V., Sennes, L. U., & Gielow, I. (2008). Speech Evaluation After Palatal Augmentation in Patients Undergoing Glossectomy. *Archives of Otolaryngology–Head & Neck Surgery*, 134(10), 1066. <https://doi.org/10.1001/archotol.134.10.1066>
- de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41(2), 385–390. <https://doi.org/10.3758/BRM.41.2.385>

- Degroote, G., Simon, J., Borel, S., & Crevier-Buchman, L. (2012). The French Version of Speech Handicap Index: Validation and Comparison with the Voice Handicap Index. *Folia Phoniatrica et Logopaedica*, 64(1), 20–25. <https://doi.org/10.1159/000328982>
- DeNittis, A. S., Machtay, M., Rosenthal, D. I., Sanfilippo, N. J., Lee, J. H., Goldfeder, S., Chalian, A. A., Weinstein, G. S., & Weber, R. S. (2001). Advanced oropharyngeal carcinoma treated with surgery and radiotherapy: Oncologic outcome and functional assessment. *American Journal of Otolaryngology*, 22(5), 329–335. <https://doi.org/10.1053/ajot.2001.26492>
- Devevey, A., & Kunz, L. (2018). Les troubles du langage oral. In *Les enfants Dys* (Elsevier M, pp. 85–101). Elsevier Masson. <https://doi.org/10.3917/sh.261.0017>
- Dewaele, J.-M. (2000). Saisir l'insaisissable ? Les mesures de longueur d'énoncés en linguistique appliquée. *IRAL - International Review of Applied Linguistics in Language Teaching*, 38(1), 17–33. <https://doi.org/10.1515/iral.2000.38.1.17>
- Dhingra, N., Valli, E., & Kunz, A. (2020). Recognition and Localisation of Pointing Gestures Using a RGB-D Camera. In *Communications in Computer and Information Science: Vol. 1224 CCIS* (pp. 205–212). https://doi.org/10.1007/978-3-030-50726-8_27
- Doward, L. C., & McKenna, S. P. (2004). Defining Patient-Reported Outcomes. *Value in Health*, 7(SUPPL. 1), S4–S8. <https://doi.org/10.1111/j.1524-4733.2004.7s102.x>
- Dwivedi, R. C., Kazi, R. A., Agrawal, N., Nutting, C. M., Clarke, P. M., Kerawala, C. J., Rhys-Evans, P. H., & Harrington, K. J. (2009). Evaluation of speech outcomes following treatment of oral and oropharyngeal cancers. *Cancer Treatment Reviews*, 35(5), 417–424. <https://doi.org/10.1016/j.ctrv.2009.04.013>
- Dwivedi, R. C., Rose, S. St., Chisholm, E. J., Kerawala, C. J., Clarke, P. M., Nutting, C. M., Rhys-Evans, P. H., Harrington, K. J., & Kazi, R. (2012). Development and validation of first-ever speech-specific perceptual speech evaluation tool for patients with head and neck cancer: The London speech evaluation (LSE) Scale. *Head & Neck*, 34(1), 94–103. <https://doi.org/10.1002/hed.21683>
- 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>

- Dwivedi, R. C., St.Rose, S., Roe, J. W. G., Chisholm, E., Elmiyeh, B., Nutting, C. M., Clarke, P. M., Kerauala, C. J., Rhys-Evans, P. H., Harrington, K. J., & Kazi, R. (2010). First report on the reliability and validity of speech handicap index in native English-speaking patients with head and neck cancer. *Head & Neck*, 36(10), n/a-n/a. <https://doi.org/10.1002/hed.21450>
- Dykstra, A., Hakel, M., & Adams, S. (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>
- Dzebo, S., Mahmutovic, J., Erkocevic, H., & Foco, F. (2017). Frequency of Depression and its Correlation with Quality of Life of Patients with Oral Cavity Cancer. *Materia Socio Medica*, 29(2), 97. <https://doi.org/10.5455/msm.2017.29.97-100>
- Eadie, T., Faust, L., Bolt, S., Kapsner-Smith, M., Pompon, R. H., Baylor, C., Futran, N., & Méndez, E. (2018). Role of Psychosocial Factors on Communicative Participation among Survivors of Head and Neck Cancer. *Otolaryngology–Head and Neck Surgery*, 159(2), 266–273. <https://doi.org/10.1177/0194599818765718>
- Eadie, T. L., Lamvik, K., Baylor, C. R., Yorkston, K. M., Kim, J., & Amtmann, D. (2014). Communicative Participation and Quality of Life in Head and Neck Cancer. *Annals of Otology, Rhinology & Laryngology*, 123(4), 257–264. <https://doi.org/10.1177/0003489414525020>
- Eadie, T. L., Otero, D., Cox, S., Johnson, J., Baylor, C. R., Yorkston, K. M., & Doyle, P. C. (2016). The relationship between communicative participation and postlaryngectomy speech outcomes. *Head & Neck*, 38(S1), E1955–E1961. <https://doi.org/10.1002/hed.24353>
- Ebrahim, S. (1995). Clinical and public health perspectives and applications of health-related quality of life measurement. *Social Science & Medicine*, 41(10), 1383–1394. [https://doi.org/10.1016/0277-9536\(95\)00116-O](https://doi.org/10.1016/0277-9536(95)00116-O)
- Enderby, P. (1980). Frenchay Dysarthria Assessment. *International Journal of Language & Communication Disorders*, 15(3), 165–173. <https://doi.org/10.3109/13682828009112541>
- Fang, C., Li, H., Ma, L., & Zhang, M. (2017). Intelligibility Evaluation of Pathological Speech through Multigranularity Feature Extraction and Optimization. *Computational and Mathematical Methods in Medicine*, 2017, 1–8. <https://doi.org/10.1155/2017/2431573>
- Farinas, J. (2002). *Une modélisation automatique du rythme pour l'identification des langues*. Université Paul Sabatier.
- Feenaughty, L., Tjaden, K., & Sussman, J. (2014). Relationship between acoustic measures and judgments of intelligibility in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics*, 28(11), 857–878. <https://doi.org/10.3109/02699206.2014.921839>

- Felce, D., & Perry, J. (1995). Quality of life: Its definition and measurement. *Research in Developmental Disabilities*, 16(1), 51–74. [https://doi.org/10.1016/0891-4222\(94\)00028-8](https://doi.org/10.1016/0891-4222(94)00028-8)
- Fex, S. (1992). Perceptual evaluation. *Journal of Voice*, 6(2), 155–158.
- Fontan, L. (2012). *De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication*. Université Toulouse 2 Le Mirail.
- François, J., & Nespoulous, J. L. (2011). *L'architecture des processus de production et de réception : aspects (neuro)psycholinguistiques* (Peeters).
- Fratiglioni, L., Wang, H., Ericsson, K., Maytan, M., & Winblad, B. (2000). Influence of social network on occurrence of dementia: a community-based longitudinal study. *The Lancet*, 355(9212), 1315–1319. [https://doi.org/10.1016/S0140-6736\(00\)02113-9](https://doi.org/10.1016/S0140-6736(00)02113-9)
- Furman, W., & Buhrmester, D. (2009). Methods and Measures: The Network of Relationships Inventory: Behavioral Systems Version. *International Journal of Behavioral Development*, 33(5), 470–478. <https://doi.org/10.1177/0165025409342634>
- Gala, N., Hathout, N., Nasr, A., Rey, V., & Seppälä, S. (2011). Création de clusters sémantiques dans des familles morphologiques à partir du TLFi. *Conférence Annuelle Sur Le Traitement Automatique Des Langues Naturelles (TALN2011)*, 1–6. <https://hal.archives-ouvertes.fr/hal-00989473/document>
- 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>
- Gan, H. K., Bernstein, L. J., Brown, J., Ringash, J., Vakilha, M., Wang, L., Goldstein, D., Kim, J., Hope, A., O'Sullivan, B., Waldron, J., Abdul Razak, A. R., Chen, E. X., & Siu, L. L. (2011). Cognitive Functioning After Radiotherapy or Chemoradiotherapy for Head-and-Neck Cancer. *International Journal of Radiation Oncology*Biology*Physics*, 81(1), 126–134. <https://doi.org/10.1016/j.ijrobp.2010.05.004>
- Gatignol, P., & Lannadère, E. (2010). *Bilan informatisé de la Motricité Bucco-linguo-faciales (version adultes)* (Edition Ad).
- Gelin, L., Daniel, M., Pinquier, J., & Pellegrini, T. (2021). *End-to-end acoustic modelling for phone recognition of young readers*. <https://www.lalilo.com/>
- Ghannay, S. (2017). *Étude sur les représentations continues de mots appliquées à la détection automatique des erreurs de reconnaissance de la parole* [Université du Maine]. <https://hal.archives-ouvertes.fr/tel-01661491/>
- Ghio, A., Giovanni, a, Teston, B., Révis, J., Yu, P., Ouaknine, M., Robert, D., & Legou, T. (2008). Bilan et perspectives de quinze ans d'évaluation vocale par méthodes instrumentales et perceptives. *Actes, Journées d'Etude Sur La Parole (JEP); Journées d'Etude Sur La Parole (JEP)*, 309–312. <http://hal.archives-ouvertes.fr/hal-00292401/en/>

- Ghio, A., Giusti, L., Blanc, E., & Pinto, S. (2020). French adaptation of the “Frenchay Dysarthria Assessment 2” speech intelligibility test. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 137(2), 111–116. <https://doi.org/10.1016/j.anorl.2019.10.007>
- Ghio, A., 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*, 589–596. <https://hal.archives-ouvertes.fr/hal-01372037>
- 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. *XXXIIe Journées d’Études Sur La Parole*, 285–293. <https://doi.org/10.21437/JEP.2018-33>
- Gray, C., Baylor, C., Eadie, T., Kendall, D., & Yorkston, K. (2012). The Levels of Speech Usage rating scale: comparison of client self-ratings with speech pathologist ratings. *International Journal of Language & Communication Disorders*, 47(3), 333–344. <https://doi.org/10.1111/j.1460-6984.2011.00112.x>
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire: essai de méthodologie*. Presses Universitaires de France.
- 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
- Ha, J., Sung, I., Son, J., Stone, M., Ord, R., & Cho, Y. (2016). Analysis of speech and tongue motion in normal and post-glossectomy speaker using cine MRI. *Journal of Applied Oral Science*, 24(5), 472–480. <https://doi.org/10.1590/1678-775720150421>
- Haderlein, T., Moers, C., Möbius, B., & Nöth, E. (2012). Automatic Rating of Hoarseness by Text-based Cepstral and Prosodic Evaluation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 7499 LNAI* (Issue Informatik 5, pp. 573–580). https://doi.org/10.1007/978-3-642-32790-2_70
- Halpern, B. M., Son, R. van, Brekel, M. van den, & Scharenborg, O. (2020). Detecting and Analysing Spontaneous Oral Cancer Speech in the Wild. *Interspeech* 2020, 2020-Octob, 4826–4830. <https://doi.org/10.21437/Interspeech.2020-1598>
- Hassanein, K. A.-A. M., Musgrove, B. T., & Bradbury, E. (2001). Functional status of patients with oral cancer and its relation to style of coping, social support and psychological status. *British Journal of Oral and Maxillofacial Surgery*, 39(5), 340–345. <https://doi.org/10.1054/bjom.2001.0652>

- Hays, R., & Reeve, B. (2008). Measurement and Modeling of Health-Related Quality of Life. In J. Killewo, H. K. Heggenhougen, & S. R. Quah (Eds.), *International Encyclopedia of Public Health* (Academic P, pp. 241–252). Elsevier. <https://doi.org/10.1016/B978-012373960-5.00336-1>
- 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 Otolology, Rhinology & Laryngology*, 112(4), 324–333. <https://doi.org/10.1177/000348940311200406>
- Heutte, N., Plisson, L., Lange, M., Prevost, V., & Babin, E. (2014). Les outils de qualité de vie en cancérologie ORL. *Annales Françaises d’Oto-Rhino-Laryngologie et de Pathologie Cervico-Faciale*, 131(1), 24–35. <https://doi.org/10.1016/j.aforl.2013.10.004>
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic Correlates of Breathly Vocal Quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778. <https://doi.org/10.1044/jshr.3704.769>
- 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, 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))
- Jacobi, I., van Rossum, M. A., van der Molen, L., Hilgers, F. J. M., & van den Brekel, M. W. M. (2013). Acoustic Analysis of Changes in Articulation Proficiency in Patients with Advanced Head and Neck Cancer Treated with Chemoradiotherapy. *Annals of Otolology, Rhinology & Laryngology*, 122(12), 754–762. <https://doi.org/10.1177/000348941312201205>
- Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI). *American Journal of Speech-Language Pathology*, 6(3), 66–70. <https://doi.org/10.1044/1058-0360.0603.66>
- Jakobson, R. (1963). Linguistique et poétique. In *Essais de linguistique générale* 1 (pp. 209–248). Minuit.
- Jelinek, F. (1976). Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, 64(4), 532–556. <https://doi.org/10.1109/PROC.1976.10159>
- Johansson, V. (2009). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers in Linguistics*, 53(0), 61–79.
- Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics*, 34(7), 645–649. <https://doi.org/10.1007/s40273-016-0389-9>

- Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., Someki, M., Soplin, N. E. Y., Yamamoto, R., Wang, X., Watanabe, S., Yoshimura, T., & Zhang, W. (2019). A Comparative Study on Transformer vs RNN in Speech Applications. *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 9(4), 449–456. <https://doi.org/10.1109/ASRU46091.2019.9003750>
- Karnell, L. H., Christensen, A. J., Rosenthal, E. L., Magnuson, J. S., & Funk, G. F. (2007). Influence of social support on health-related quality of life outcomes in head and neck cancer. *Head & Neck*, 29(2), 143–146. <https://doi.org/10.1002/hed.20501>
- Karnell, L. H., Funk, G. F., & Hoffman, H. T. (2000). Assessing head and neck cancer patient outcome domains. *Head & Neck*, 22(1), 6–11. [https://doi.org/10.1002/\(sici\)1097-0347\(200001\)22:1<6::aid-hed2>3.3.co;2-g](https://doi.org/10.1002/(sici)1097-0347(200001)22:1<6::aid-hed2>3.3.co;2-g)
- Kawachi, I. (2001). Social Ties and Mental Health. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 78(3), 458–467. <https://doi.org/10.1093/jurban/78.3.458>
- Kazi, R., Prasad, V. M. N., Kanagalingam, J., Georgalas, C., Venkitaraman, R., Nutting, C. M., Clarke, P., Rhys-Evans, P., & Harrington, K. J. (2007). Analysis of formant frequencies in patients with oral or oropharyngeal cancers treated by glossectomy. *International Journal of Language & Communication Disorders*, 42(5), 521–532. <https://doi.org/10.1080/13682820601056566>
- Keintz, C. K., Bunton, K., & Hoit, J. D. (2007). Influence of Visual Information on the Intelligibility of Dysarthric Speech. *American Journal of Speech-Language Pathology*, 16(3), 222–234. [https://doi.org/10.1044/1058-0360\(2007/027\)](https://doi.org/10.1044/1058-0360(2007/027))
- Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward Phonetic Intelligibility Testing in Dysarthria. *Journal of Speech and Hearing Disorders*, 54(4), 482–499. <https://doi.org/10.1044/jshd.5404.482>
- Kim, J. C., Rao, H., & Clements, M. A. (2014). Speech intelligibility estimation using multi-resolution spectral features for speakers undergoing cancer treatment. *The Journal of the Acoustical Society of America*, 136(4), EL315–EL321. <https://doi.org/10.1121/1.4896410>
- Kim, J. H., Fisher, L. M., Reder, L., Hapner, E. R., & Pepper, J.-P. (2018). Speech and Communicative Participation in Patients With Facial Paralysis. *JAMA Otolaryngology–Head & Neck Surgery*, 144(8), 686. <https://doi.org/10.1001/jamaoto.2018.0649>
- 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))

- Kipyatkova, I., & Karpov, A. (2016). DNN-Based Acoustic Modeling for Russian Speech Recognition Using Kaldi. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 9811 LNCS* (Issue August 2016, pp. 246–253). https://doi.org/10.1007/978-3-319-43958-7_29
- Kmet, L. M., Lee, R. C., & Cook, L. S. (2004). *Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields*. <https://doi.org/10.5858/arpa.2020-0217-SA>
- Knuijt, S., Kalf, J. G., van Engelen, B. G. M., de Swart, B. J. M., & Geurts, A. C. H. (2017). The Radboud Dysarthria Assessment: Development and Clinimetric Evaluation. *Folia Phoniatrica et Logopaedica*, 69(4), 143–153. <https://doi.org/10.1159/000484556>
- Knuuttila, H., & Juhani Pukander, T. (1999). Speech Articulation after Subtotal Glossectomy and Reconstruction with a Myocutaneous Flap. *Acta Otolaryngologica*, 119(5), 621–626. <https://doi.org/10.1080/00016489950180892>
- Kraaijenga, S. A. C., Oskam, I. M., van Son, R. J. J. H., Hamming-Vrieze, O., Hilgers, F. J. M., van den Brekel, M. W. M., & van der Molen, L. (2016). Assessment of voice, speech, and related quality of life in advanced head and neck cancer patients 10-years+ after chemoradiotherapy. *Oral Oncology*, 55, 24–30. <https://doi.org/10.1016/j.oraloncology.2016.02.001>
- Kraaijenga, S. A. C., van der Molen, L., Jacobi, I., Hamming-Vrieze, O., Hilgers, F. J. M., & van den Brekel, M. W. M. (2015). Prospective clinical study on long-term swallowing function and voice quality in advanced head and neck cancer patients treated with concurrent chemoradiotherapy and preventive swallowing exercises. *European Archives of Oto-Rhino-Laryngology*, 272(11), 3521–3531. <https://doi.org/10.1007/s00405-014-3379-6>
- Kroenke, C. H. (2018). A conceptual model of social networks and mechanisms of cancer mortality, and potential strategies to improve survival. *Translational Behavioral Medicine*, 8(4), 629–642. <https://doi.org/10.1093/tbm/ibx061>
- Kroenke, C. H., Kwan, M. L., Neugut, A. I., Ergas, I. J., Wright, J. D., Caan, B. J., Hershman, D., & Kushi, L. H. (2013). Social networks, social support mechanisms, and quality of life after breast cancer diagnosis. *Breast Cancer Res Treat.*, 139(2), 515–527. <https://doi.org/10.1016/j.physbeh.2017.03.040>
- Kumar, P., Jain, V., Thakar, A., & Aggarwal, V. (2013). Effect of varying bulb height on articulation and nasalance in maxillectomy patients with hollow bulb obturator. *Journal of Prosthodontic Research*, 57(3), 200–205. <https://doi.org/10.1016/j.jprior.2013.02.002>
- Kuo, C., & Tjaden, K. (2016). Acoustic variation during passage reading for speakers with dysarthria and healthy controls. *Journal of Communication Disorders*, 62, 30–44. <https://doi.org/10.1016/j.jcomdis.2016.05.003>
- Kuo, Y.-C., Rosen, S., & Faulkner, A. (2008). Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, 123(5), 2815–2824. <https://doi.org/10.1121/1.2896755>

- Laaksonen, J.-P., Rieger, J., Harris, J., & Seikaly, H. (2011). A longitudinal acoustic study of the effects of the radial forearm free flap reconstruction on sibilants produced by tongue cancer patients. *Clinical Linguistics & Phonetics*, 25(4), 253–264. <https://doi.org/10.3109/02699206.2010.525681>
- Laaksonen, J., Rieger, J., Happonen, R., Harris, J., & Seikaly, H. (2010). Speech after radial forearm free flap reconstruction of the tongue: A longitudinal acoustic study of vowel and diphthong sounds. *Clinical Linguistics & Phonetics*, 24(1), 41–54. <https://doi.org/10.3109/02699200903340758>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Letanneux, A., Walshe, M., Viallet, F., & Pinto, S. (2013). The Dysarthria Impact Profile: A Preliminary French Experience with Parkinson’s Disease. *Parkinson’s Disease*, 2013, 1–6. <https://doi.org/10.1155/2013/403680>
- Liang, B., Iwnicki, S. D., & Zhao, Y. (2013). Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis. *Mechanical Systems and Signal Processing*, 39(1–2), 342–360. <https://doi.org/10.1016/j.ymssp.2013.02.016>
- Liao, L. J., Hsu, W. L., Lo, W. C., Cheng, P. W., Shueng, P. W., & Hsieh, C. H. (2019). Health-related quality of life and utility in head and neck cancer survivors. *BMC Cancer*, 19(1), 1–10. <https://doi.org/10.1186/s12885-019-5614-4>
- 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. *PLoS Medicine*, 6(7). <https://doi.org/10.1371/journal.pmed.1000100>
- Lindblom, B. (1990). On the Communication Process: Speaker-Listener Interaction and the Development of Speech. *Augmentative and Alternative Communication*, 6(4), 220–230. <https://doi.org/10.1080/07434619012331275504>
- 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>
- Lu, L., Liu, C., Li, J., & Gong, Y. (2020). Exploring Transformers for Large-Scale Speech Recognition. *Interspeech 2020, 2020-Octob*, 5041–5045. <https://doi.org/10.21437/Interspeech.2020-2638>
- Lynn, M. R. (1986). Determination and quantification of content validity. In *Nursing Research* (Vol. 35, Issue 6, pp. 382–386). <http://ijoh.tums.ac.ir/index.php/ijoh/article/view/26>

- Mair, M. D., Nair, S., Nikam, S., Nair, D., Agarwal, J. P., & Chaturvedi, P. (2017). Longitudinal and cross-sectional assessment of quality of life in surgically treated advanced (T4) cancer of the buccal mucosa. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 124(6), 529–536. <https://doi.org/10.1016/j.oooo.2017.09.006>
- Markkanen-Leppänen, M., Isotalo, E., Makitie, A. A., Suominen, E., Asko-Seljavaara, S., & Haapanen, M.-L. (2005). Speech Aerodynamics and Nasalance in Oral Cancer Patients Treated with Microvascular Transfers. *Journal of Craniofacial Surgery*, 16(6), 990–995. <https://doi.org/10.1097/01.scs.0000179753.14037.7a>
- Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice*, 24(5), 540–555. <https://doi.org/10.1016/j.jvoice.2008.12.014>
- Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America*, 126(5), 2619–2634. <https://doi.org/10.1121/1.3224706>
- Matsui, Y., Ohno, K., Yamashita, Y., & Takahashi, K. (2007). Factors influencing postoperative speech function of tongue cancer patients following reconstruction with fasciocutaneous/myocutaneous flaps-a multicenter study. *International Journal of Oral and Maxillofacial Surgery*, 36(7), 601–609. <https://doi.org/10.1016/j.ijom.2007.01.014>
- Mattis, S. (1976). Mental Status Examination for Organic Mental Syndrome in the Elderly Patient. In L. Bellak & T. Karasu (Eds.), *Geriatric Psychiatry: a Handbook for Psychiatrists and primary Care Physicians* (Grune & St, pp. 77–121).
- Mayo, C., & Mayo, R. (2011). Normative nasalance values across languages. *Echo*, 6(11), 22–32. <http://libres.uncg.edu/ir/uncg/listing.aspx?id=9551>
- Mazaux, J.-M., Daviet, J.-C., Darrigrand, B., Stuit, A., Muller, F., Dutheil, S., Joseph, P.-A., & Barat, M. (2006). Difficultés de communication des personnes aphasiques. *Évaluation Des Troubles Neuropsychologiques En Vie Quotidienne*, 73–82. https://doi.org/10.1007/2-287-34365-2_8
- Mendoza, E., Valencia, N., Muñoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice*, 10(1), 59–66. [https://doi.org/10.1016/S0892-1997\(96\)80019-1](https://doi.org/10.1016/S0892-1997(96)80019-1)
- Menin-Sicard, A., & Sicard, E. (2019). *DIADOLAB 3 - Logiciel d'évaluation et de rééducation de la parole*. Gerip.
- Meyer, T. K., Kuhn, J. C., Campbell, B. H., Marbella, A. M., Myers, K. B., & Layde, P. M. (2004). Speech intelligibility and quality of life in head and neck cancer survivors. *Laryngoscope*, 114(11 I), 1977–1981. <https://doi.org/10.1097/01.mlg.0000147932.36885.9e>

- Middag, C. (2013). *Automatische analyse van pathologische spraak Automatic Analysis of Pathological Speech*.
- Middag, C., Clapham, R., Van Son, R., & Martens, J. P. (2014). Robust automatic intelligibility assessment techniques evaluated on speakers treated for head and neck cancer. *Computer Speech and Language*, 28(2), 467–482. <https://doi.org/10.1016/j.csl.2012.10.007>
- Milenkovic, P. H., Wagner, M., Kent, R. D., Story, B. H., & Vorperian, H. K. (2020). Effects of sampling rate and type of anti-aliasing filter on linear-predictive estimates of formant frequencies in men, women, and children. *The Journal of the Acoustical Society of America*, 147(3), EL221–EL227. <https://doi.org/10.1121/10.0000824>
- Min Ang, X., Chi Khang Au, P., Kwok, K., Yeon Park, K., Kujan, O., Frydrych, A. M., & Farah, C. S. (2019). Quality of life in patients with oral leukoplakia. *Journal of Oral Pathology and Medicine*, 48(7), 574–580. <https://doi.org/10.1111/jop.12897>
- Missire, R. (2013). Classes lexicales et activité sémantique. In *Coseriu aujourd'hui, linguistique du texte et philosophie du langage* (pp. 147–160). Lambert-Lucas. <http://formes-symboliques.org/IMG/pdf/doc-230.pdf>
- Mlynarek, A., Rieger, J., Harris, J., O'Connell, D., Al-Qahtani, K., Ansari, K., Chau, J., & Seikaly, H. (2008). Methods of functional outcomes assessment following treatment of oral and oropharyngeal cancer: review of the literature. *Journal of Otolaryngology - Head & Neck Surgery*, 37(1), 2–10. <https://doi.org/10.2310/7070.2008.1001>
- Moerman, M., Vermeersch, H., Van Lierde, K., Fahimi, H., & Van Cauwenberge, P. (2003). Refinement of the free radial forearm flap reconstructive technique after resection of large oropharyngeal malignancies with excellent functional results. *Head & Neck*, 25(9), 772–777. <https://doi.org/10.1002/hed.10295>
- Mohan, K., & Pearl, J. (2014). On the testability of models with missing data. *Journal of Machine Learning Research*, 33(March), 643–650.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, 339(jul21 1), b2535–b2535. <https://doi.org/10.1136/bmj.b2535>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Mowbray, O., Quinn, A., & Cranford, J. A. (2014). Social networks and alcohol use disorders: Findings from a nationally representative sample. *American Journal of Drug and Alcohol Abuse*, 40(3), 181–186. <https://doi.org/10.3109/00952990.2013.860984>

- Murphy, B. A., Ridner, S., Wells, N., & Dietrich, M. (2007). Quality of life research in head and neck cancer: A review of the current state of the science. *Critical Reviews in Oncology/Hematology*, 62(3), 251–267. <https://doi.org/10.1016/j.critrevonc.2006.07.005>
- Nejedlová, D. (2002). Comparative Study on Bigram Language Models for Spoken Czech Recognition. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 2448, Issue 102, pp. 197–204). https://doi.org/10.1007/3-540-46154-X_27
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™. *L'année Psychologique*, 101(3), 447–462. <https://doi.org/10.3406/psy.2001.1341>
- NHMRC. (1999). *A guide to the Development, Evaluation and Implemetation of Clinical Practice Guidelines* (Australian). http://www.ausinfo.gov.au/general/gen_hottobuy.htm
- Nordmann-Dolbeault, S. (2009). *La détresse des patients atteints de cancer : prévalence, facteurs prédictifs , modalités de repérage et de prise en charge*. Université Pierre et Marie Curie - Paris VI.
- Ojo, B., Genden, E. M., Teng, M. S., Milbury, K., Misiukiewicz, K. J., & Badr, H. (2012). A systematic review of head and neck cancer quality of life assessment instruments. *Oral Oncology*, 48(10), 923–937. <https://doi.org/10.1016/j.oraloncology.2012.03.025>
- Payne, N., Palmer Kelly, E., & Pawlik, T. M. (2019). Assessing structure and characteristics of social networks among cancer survivors: impact on general health. *Supportive Care in Cancer*, 27(8), 3045–3051. <https://doi.org/10.1007/s00520-018-4620-4>
- Pearlin, L. I., & Johnson, J. S. (1977). Marital Status, Life-Strains and Depression. *American Sociological Review*, 42(5), 704. <https://doi.org/10.2307/2094860>
- Peddinti, V., Povey, D., & Khudanpur, S. (2015). A time delay neural network architecture for efficient modeling of long temporal contexts. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015-Janua*, 2–6.
- Pellegrino, F. (1998). *Une approche phonétique en identification automatique des langues : la modélisation acoustique des systèmes vocaliques*. Université Paul Sabatier.
- Pépiot, E. (2014). Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *7th International Conference on Speech Prosody 2014*, 305–309. <https://doi.org/10.21437/SpeechProsody.2014-48>

- 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(January), 169–198.
- Picard, D. (1992). De la communication à l'interaction : l'évolution des modèles. *Communication et Langages*, 93(1), 69–83. <https://doi.org/10.3406/colan.1992.2380>
- Plisson, L., Pillot-Loiseau, C., & Crevier-Buchman, L. (2017). Intelligibilité de la parole après le traitement d'un cancer de l'oropharynx : étude descriptive chez sept patients en pré-traitement et en post-traitement précoce. *7èmes Journées de Phonétique Clinique (JPC7)*.
- Pluvy, V. (2018). *Validation de la GEMCO, grille d'évaluation des mouvements orofaciaux à destination des patients atteints de cancer de la cavité orale*. Université Claude Bernard Lyon 1.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- 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, T., Balaguer, M., Pinquier, J., Mauclair, J., Woisard, V., & Speyer, R. (2021b). Relationship between phoneme-level spectral acoustics and speech intelligibility in healthy speech: a systematic review. *Speech, Language and Hearing*, 0(0), 1–28. <https://doi.org/10.1080/2050571x.2021.1913300>
- Pommée, T., Maryn, Y., Finck, C., & Morsomme, D. (2020). Validation of the Acoustic Voice Quality Index, Version 03.01, in French. *Journal of Voice*, 34(4), 646.e11-646.e26. <https://doi.org/10.1016/j.jvoice.2018.12.008>
- Povey, D., Cheng, G., Wang, Y., Li, K., Xu, H., Yarmohammadi, M., & Khudanpur, S. (2018). Semi-Orthogonal Low-Rank Matrix Factorization for Deep Neural Networks. *Interspeech 2018, 2018-Sept(2)*, 3743–3747. <https://doi.org/10.21437/Interspeech.2018-1417>
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hanneman, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.
- Povey, D., Peddinti, V., Galvez, D., Ghahremani, P., Manohar, V., Na, X., Wang, Y., & Khudanpur, S. (2016). Purely Sequence-Trained Neural Networks for ASR Based on Lattice-Free MMI. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 08-12-Sept(September)*, 2751–2755. <https://doi.org/10.21437/Interspeech.2016-595>

- Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075–1091. <https://doi.org/10.1080/02687030444000534>
- Quintas, S., Mauclair, J., Woisard, V., & Pinquier, J. (2020). Automatic prediction of speech intelligibility based on X-vectors in the context of head and neck cancer. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020-Octob*, 4976–4980. <https://doi.org/10.21437/Interspeech.2020-1431>
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. <https://doi.org/10.1109/5.18626>
- Rambeau, A., Beauplet, B., Laviec, H., Licaj, I., Leconte, A., Chatel, C., Le Bon, P., Denhaerynck, J., Clarisse, B., Frenkiel, N., Lange, M., & Joly, F. (2019). Prospective comparison of the Montreal Cognitive Assessment (MoCA) and the Mini Mental State Examination (MMSE) in geriatric oncology. *Journal of Geriatric Oncology*, 10(2), 235–240. <https://doi.org/10.1016/j.jgo.2018.08.003>
- Rebourg, M., Lalain, M., Ghio, A., Fredouille, C., Fakhry, N., & Woisard, V. (2020). Évaluer l’intelligibilité, mots ou pseudo-mots? Comparaison entre deux groupes d’auditeurs. *6e Conférence Conjointe Journées d’Études Sur La Parole (JEP, 33e Édition), Traitement Automatique Des Langues Naturelles (TALN, 27e Édition), Rencontre Des Jeunes Chercheurs En Informatique Pour Le Traitement Automatique Des Langues (RÉCITAL, 22e Édition*, 543–551.
- Regier, N. G., Naik, A. D., Mulligan, E. A., Nasreddine, Z. S., Driver, J. A., Sada, Y. H. F., & Moye, J. (2019). Cancer-related cognitive impairment and associated factors in a sample of older male oral-digestive cancer survivors. *Psycho-Oncology*, 28(7), 1551–1558. <https://doi.org/10.1002/pon.5131>
- Reich, M. (2009). Cancer et image du corps: identité, représentation et symbolique: Le corps retrouvé (French). *Cancer and Body Image: Identity, Representation* (English), 85(3), 247–254. <https://doi.org/10.3917/inpsy.8503.0247>
- Reinert, M. (1990). Une méthode de classification des énoncés d’un corpus présentée à l’aide d’une application. *Les Cahiers de l’analyse Des Données*, 15(1), 21–36.
- Rieke, K., Schmid, K. K., Lydiatt, W., Houfek, J., Boilesen, E., & Watanabe-Galloway, S. (2017). Depression and survival in head and neck cancer patients. *Oral Oncology*, 65, 76–82. <https://doi.org/10.1016/j.oraloncology.2016.12.014>
- 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>

- Robert, M., Mauclair, J., Lannadere, E., Tankéré, F., Lamas, G., & Gatignol, P. (2013). Acoustic analysis of the articulatory disorders of patients affected by peripheral facial paralysis. *Revue de Laryngologie Otologie Rhinologie*, 134(1), 3–11.
- Rogers, S. N., El-Sheikha, J., & Lowe, D. (2009). The development of a Patients Concerns Inventory (PCI) to help reveal patients concerns in the head and neck clinic. *Oral Oncology*, 45(7), 555–561. <https://doi.org/10.1016/j.oraloncology.2008.09.004>
- Roscoe, B. A., Hopke, P. K., Dattner, S. L., & Jenks, J. M. (1982). The Use of Principal Component Factor Analysis to Interpret Particulate Compositional Data Sets. *Journal of the Air Pollution Control Association*, 32(6), 637–642. <https://doi.org/10.1080/00022470.1982.10465439>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Saikia, P. (2017). HMM-DNN Speech Recognition Techniques: a Review. *International Journal of Development Research*, 7(7), 14068–14072. <http://www.journalijdr.com>
- Seikaly, H., Rieger, J., Wolfaardt, J., Moysa, G., Harris, J., & Jha, N. (2003). Functional Outcomes After Primary Oropharyngeal Cancer Resection and Reconstruction With the Radial Forearm Free Flap. *The Laryngoscope*, 113(5), 897–904. <https://doi.org/10.1097/00005537-200305000-00023>
- Shannon, C. E., & Weaver, W. (1964). The mathematical theory of communication. In *International Business*. University of Illinois Press.
- Sharpe, G., Camoes Costa, V., Doubé, W., Sita, J., McCarthy, C., & Carding, P. (2019). Communication changes with laryngectomy and impact on quality of life: a review. *Quality of Life Research*, 28(4), 863–877. <https://doi.org/10.1007/s11136-018-2033-y>
- Siddiqi, M. H., Ali, R., Sattar, A., Khan, A. M., & Lee, S. (2014). Depth Camera-Based Facial Expression Recognition System Using Multilayer Scheme. *IETE Technical Review*, 31(4), 277–286. <https://doi.org/10.1080/02564602.2014.944588>
- St-Jean-Trudel, E., Guay, S., Marchand, A., & O'Connor, K. (2005). Elaboration and validation of a questionnaire measuring social support in situations of anxiety with a population of university students. *Santé Mentale Au Québec*, 30(2), 43–60. <https://doi.org/10.7202/012138ar>
- Stelzle, F., Knipfer, C., Schuster, M., Bocklet, T., Nöth, E., Adler, W., Schempf, L., Vieler, P., Riemann, M., Neukam, F. W., & Nkenke, E. (2013). Factors influencing relative speech intelligibility in patients with oral squamous cell carcinoma: A prospective study using automatic, computer-based speech analysis. *International Journal of Oral and Maxillofacial Surgery*, 42(11), 1377–1384. <https://doi.org/10.1016/j.ijom.2013.05.021>

- Sugiyama, M., Sawai, H., & Waibel, A. H. (1991). Review of TDNN (time delay neural network) architectures for speech recognition. *1991., IEEE International Symposium on Circuits and Systems, 1*, 582–585. <https://doi.org/10.1109/ISCAS.1991.176402>
- Tahani, B., Razavi, S. M., Emami, H., & Alamchi, F. (2017). Assessment of the quality of life of the patients with treated oral cancer in Iran. *Oral and Maxillofacial Surgery, 21*(4), 429–437. <https://doi.org/10.1007/s10006-017-0651-0>
- Takatsu, J., Hanai, N., Suzuki, H., Yoshida, M., Tanaka, Y., Tanaka, S., Hasegawa, Y., & Yamamoto, M. (2017). Phonologic and Acoustic Analysis of Speech Following Glossectomy and the Effect of Rehabilitation on Speech Outcomes. *Journal of Oral and Maxillofacial Surgery, 75*(7), 1530–1541. <https://doi.org/10.1016/j.joms.2016.12.004>
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis, 48*(1), 159–205. <https://doi.org/10.1016/j.csda.2004.03.005>
- Terrell, J. E., Ronis, D. L., Fowler, K. E., Bradford, C. R., Chepeha, D. B., Prince, M. E., Teknos, T. N., Wolf, G. T., & Duffy, S. A. (2005). Comments on Selected Recent Dysphagia Literature. *Dysphagia, 20*(1), 62–67. <https://doi.org/10.1007/s00455-004-0030-5>
- Thomas, L., Jones, T. M., Tandon, S., Carding, P., Lowe, D., & Rogers, S. (2009). Speech and voice outcomes in oropharyngeal cancer and evaluation of the University of Washington Quality of Life speech domain. *Clinical Otolaryngology, 34*(1), 34–42. <https://doi.org/10.1111/j.1749-4486.2008.01830.x>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases, 40*(6), 593–600. [https://doi.org/10.1016/0021-9681\(87\)90019-1](https://doi.org/10.1016/0021-9681(87)90019-1)
- Utianski, R. L., Clark, H. M., Duffy, J. R., Botha, H., Whitwell, J. L., & Josephs, K. A. (2020). Communication Limitations in Patients With Progressive Apraxia of Speech and Aphasia. *American Journal of Speech-Language Pathology, 29*(4), 1976–1986. https://doi.org/10.1044/2020_AJSLP-20-00012
- Vainshtein, J. M., Griffith, K. A., Feng, F. Y., Vineberg, K. A., Chepeha, D. B., & Eisbruch, A. (2014). Patient-reported voice and speech outcomes after whole-neck intensity modulated radiation therapy and chemotherapy for oropharyngeal cancer: Prospective longitudinal study. *International Journal of Radiation Oncology Biology Physics, 89*(5), 973–980. <https://doi.org/10.1016/j.ijrobp.2014.03.013>

- van der Molen, L., van Rossum, M. A., Ackerstaff, A. H., Smeele, L. E., Rasch, C. R., & Hilgers, F. J. (2009). Pretreatment organ function in patients with advanced head and neck cancer: clinical outcome measures and patients' views. *BMC Ear, Nose and Throat Disorders*, 9(1), 10. <https://doi.org/10.1186/1472-6815-9-10>
- Van Nuffelen, G., Middag, C., De Bodt, M., & Martens, J. (2009). Speech technology-based assessment of phoneme intelligibility in dysarthria. *International Journal of Language & Communication Disorders*, 44(5), 716–730. <https://doi.org/10.1080/13682820802342062>
- Venturini, S., & Mehmetoglu, M. (2019). plssem: A Stata Package for Structural Equation Modeling with Partial Least Squares. *Journal of Statistical Software*, 88(8), 0–35. <https://doi.org/10.18637/jss.v088.i08>
- Villalba, J., Chen, N., Snyder, D., Garcia-Romero, D., McCree, A., Sell, G., Borgstrom, J., García-Perera, L. P., Richardson, F., Dehak, R., Torres-Carrasquillo, P. A., & Dehak, N. (2020). State-of-the-art speaker recognition with neural network embeddings in NIST SRE18 and Speakers in the Wild evaluations. *Computer Speech & Language*, 60, 101026. <https://doi.org/10.1016/j.csl.2019.101026>
- Volín, J., & Zimmermann, J. (2011). Spectral Slope Parameters and Detection of Word Stress. *Technical Computing Prague*, 125(January), 5.
- Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1989). Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3), 328–339. <https://doi.org/10.1109/29.21701>
- Wakumoto, M., Ohno, K., Imai, S., Yamashita, Y., Akizuki, H., & Michi, K. -I. (1996). Analysis of the articulation after glossectomy. *Journal of Oral Rehabilitation*, 23(11), 764–770. <https://doi.org/10.1046/j.1365-2842.1996.d01-186.x>
- 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>
- Wang, D., & Zheng, T. F. (2015). Transfer learning for speech and language processing. *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 1225–1237. <https://doi.org/10.1109/APSIPA.2015.7415532>
- Weissman, M. M., Sholomskas, D., & John, K. (1981). The Assessment of Social Adjustment. *Archives of General Psychiatry*, 38(11), 1250. <https://doi.org/10.1001/archpsyc.1981.01780360066006>
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>

- Williams, A. M., Lindholm, J., Cook, D., Siddiqui, F., Ghanem, T. A., & Chang, S. S. (2017a). Association between cognitive function and quality of life in patients with head and neck cancer. *JAMA Otolaryngology - Head and Neck Surgery*, *143*(12), 1228–1235. <https://doi.org/10.1001/jamaoto.2017.2014>
- Williams, A. M., Lindholm, J., Siddiqui, F., Ghanem, T. A., & Chang, S. S. (2017b). Clinical Assessment of Cognitive Function in Patients with Head and Neck Cancer: Prevalence and Correlates. *Otolaryngology - Head and Neck Surgery (United States)*, *157*(5), 808–815. <https://doi.org/10.1177/0194599817709235>
- Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA: The Journal of the American Medical Association*, *273*(1), 59–65. <https://doi.org/10.1001/jama.273.1.59>
- Winance, M. (2008). La notion de handicap et ses transformations à travers les classifications internationales du handicap de l'OMS, 1980 et 2001. *Dynamis*, *28*, 377–406. <https://doi.org/10.4321/s0211-95362008000100016>
- Woisard, V., Astésano, C., Balaguer, M., Farinas, J., Fredouille, C., Gaillard, P., Ghio, A., Giusti, L., Laaridh, I., Lalain, M., Lepage, B., Maclair, J., Nocaudie, O., Piquier, 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, V., Bodin, S., Yardeni, E., & Puech, M. (2007). The Voice Handicap Index: Correlation Between Subjective Patient Response and Quantitative Assessment of Voice. *Journal of Voice*, *21*(5), 623–631. <https://doi.org/10.1016/j.jvoice.2006.04.005>
- Woisard, V., Espesser, R., Ghio, A., & Duez, D. (2013). De l'intelligibilité à la compréhension de la parole, quelles mesures en pratique clinique ? *Revue de Laryngologie, Otologie, Rhinologie*, *134*, 27–33.
- World Health Organization. (2001). *Classification internationale du fonctionnement, du handicap et de la santé : CIF*. Organisation mondiale de la Santé. <https://apps.who.int/iris/handle/10665/42418>
- Wu, F., García-Perera, L. P., Povey, D., & Khudanpur, S. (2019). Advances in Automatic Speech Recognition for Child Speech Using Factored Time Delay Neural Network. *Interspeech* 2019, 1–5. <https://doi.org/10.21437/Interspeech.2019-2980>
- Xu, H., Povey, D., Mangu, L., & Zhu, J. (2011). Minimum Bayes Risk decoding and system combination based on a recursion for edit distance. *Computer Speech & Language*, *25*(4), 802–828. <https://doi.org/10.1016/j.csl.2011.03.001>

- Xun, G., Li, Y., Xin Zhao, W., Gao, J., & Zhang, A. (2017). A correlated topic model using word embeddings. *IJCAI International Joint Conference on Artificial Intelligence*, 4207–4213. <https://www.ijcai.org/Proceedings/2017/0588.pdf>
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5), 971–979. <https://doi.org/10.3758/PBR.15.5.971>
- Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of Dysarthric Speech: Implications for Assessment and Treatment Planning. *American Journal of Speech-Language Pathology*, 5(1), 55–65.
- Yoshida, H., Furuya, Y., Shimodaira, K., Kanazawa, T., Kataoka, R., & Takahashi, K. (2000). Spectral characteristics of hypernasality in maxillectomy patients 1. *Journal of Oral Rehabilitation*, 27(8), 723–730. <https://doi.org/10.1046/j.1365-2842.2000.00537.x>
- Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X. (Andrew), Moore, G., Odell, J., Ollason, D., Pover, D., Ragni, A., Valtchev, V., Woodland, P., & Zhang, C. (2015). *The HTK Book (for HTK Version 3.5, documentation alpha version)* (Issue December). Cambridge University Engineering Department.
- Yu, D., & Deng, L. (2015). Automatic Speech Recognition: A Deep Learning Approach. In *Springer*. <http://research.microsoft.com/pubs/238118/YuDeng2014-book.pdf>
- Zanello, A., Weber Rouget, B., Gex-Fabry, M., Maercker, A., & Guimon, J. (2006). Validation du questionnaire de fonctionnement social (QFS), un autoquestionnaire mesurant la fréquence et la satisfaction des comportements sociaux d'une population adulte psychiatrique. *Encephale*, 32(1), 45–59. <https://doi.org/10.1019/200630005>
- Zigmond, A., & Snalth, R. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67(6), 361–370. <https://doi.org/10.1093/occmed/kqu024>

Annexes

Annexe I Effets des cancers de la cavité buccale et de l'oropharynx sur la parole au moyen d'analyses acoustiques et automatiques – Tableau de résultats (Balaguer et al., 2020e)	168
Annexe II Paramètres acoustiques et automatiques étudiés dans les études incluses	187
Annexe III Grille ECCS	188
Annexe IV Caractéristiques des sujets inclus (total et par lots)	190
Annexe V Éléments de compréhension du fonctionnement des systèmes de reconnaissance automatique de la parole	193
Annexe VI Motifs d'exclusion et conservation des 149 paramètres automatiques	197
Annexe VII Liste des 75 paramètres issus des analyses automatiques de la parole conservés pour les analyses finales	201
Annexe VIII Motifs d'exclusion et conservation des 174 items issus des autoquestionnaires	205
Annexe IX Liste des 40 items retenus pour la construction du score holistique de communication (les codes item en gras indiquent les 24 items finaux retenus)	213

Annexe I Effets des cancers de la cavité buccale et de l’oropharynx sur la parole au moyen d’analyses acoustiques et automatiques – Tableau de résultats (Balaguer et al., 2020e)

Référence	Type d'étude (NHMRC) *	QualSyst †	Participants [‡] , diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Acher et al., 2014)	IV	15/20 75 % (bonne)	<i>Participants :</i> 2 patients (1 M, 1 F) avec hémi- glossectomie et résection cervicale bilatérale, radio- chimiothérapie <i>Taille de la tumeur :</i> non précisée <i>Langue :</i> Français	Patient 1 : M, 28 ans Patient 2 : F, 62 ans	“Moments spectraux” (énergie dans le domaine fréquentiel) : Centre de gravité (COG, 1er moment) : Moyenne de la distribution des fréquences de l'énergie spectrale Asymétrie (3 ^{ème} moment, asymétrie droite-gauche de l'enveloppe spectrale : plus l'asymétrie est élevée, plus l'énergie spectrale est située sur de basses fréquences, comme dans le cas des phonèmes postérieurs Aplatissement (4 ^{ème} moment) du spectre Klatt Voice Onset Time (VOT, paramètre temporel) : transition d'une consonne à une voyelle, étendu des occlusives aux fricatives	<i>Mesure :</i> échelle de Likert à 5 niveaux (1-5 : normal à inintelligible) <i>Juges :</i> 5 orthophonistes <i>Tâches :</i> séquences VCV présentées deux fois dans un ordre aléatoire	24 CVCVC. La consonne médiane est analysée quand elle est entourée par deux voyelles identiques (i-i, u-u, a-a) C : /t/, /d/, /k/, /g/, /s/, /z/, /ʃ/ ou /ʒ/ V : /i/, /u/ ou /a/	COG et asymétrie : les changements sont liés à l'évaluation perceptive pour une large majorité des consonnes analysées dans les contextes i-i et a-a Aplatissement : ne semble pas être un paramètre pertinent dans ce contexte Klatt VOT : congruent avec l'évaluation perceptive quand cette dernière ne pouvait pas être expliquée par les paramètres spectraux

* Niveaux de hiérarchie des preuves (NHMRC) : Niveau 1 Revues systématiques ; Niveau II Essais contrôlés randomisés ; Niveau III-1 Essais contrôlés pseudo-randomisés ; Niveau III-2 Études comparatives avec groupe contrôlé et allocation des sujets non randomisée (études de cohorte), études cas témoins, ou séries chronologiques interrompues avec groupe contrôle ; Niveau III-3 Études comparatives avec cohorte contrôle historique, deux études or plus à bras unique, ou séries chronologiques interrompues sans groupe contrôle ; Niveau IV Études de cas.

† Qualité méthodologique : Forte > 80 % ; Bonne 60-79 % ; Correcte 50-59 % ; Faible < 50 %.

‡ M : Hommes, F : Femmes

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Chung et al., 2011)	III-3	20/22 91 % (forte)	<i>Participants :</i> 53 patients (48 M, 5 F) atteints d'un cancer de l'amygdale traités par chirurgie <i>Taille de la tumeur :</i> T1=10, T2=26, T3=13, T4=4 <i>Langue :</i> Coréen	56,7 ans, (étendue 39-80 ans)	Nasalité : calcul du rapport entre l'énergie acoustique issue des cavités buccale et nasale sur les sons nasaux	<i>Mesure :</i> pourcentage d'identification correcte des consonnes <i>Juge :</i> 1 orthophoniste <i>Tâche :</i> lecture d'un texte oral vs. Lecture d'un texte nasal	Lecture d'un passage ne contenant pas de sons nasaux, et d'un passage contenant de multiples sons nasaux	La technique de reconstruction dénudée ($p < 0,001$), l'étendue de la résection du palais mou ($p = 0,001$), et le stade T ($p < 0,001$) sont significativement associés au score de nasalité évalué de manière instrumentale et perceptive (reconstruction dénudée : $p = < 0,001$, étendue de la résection du palais mou : $p < 0,001$, stade T : $p = 0,006$).

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(de Bruijn et al., 2009)	III-2	21/24 88 % (forte)	<i>Participants :</i> 51 patients (28 M, 23 F) traités par chirurgie (avec reconstruction) pour un cancer avancé oral ou oropharyngé (21 cavité buccale, 30 oropharynx) Inclus 6 mois après le traitement <i>Taille de la tumeur :</i> T2=26, T3-4=25 <i>Langue :</i> Néerlandais	53,8 ans, écart-type 8,7 Étendue : 23 - 73 ans	Formants des voyelles (/a/, /i/, /u/) : F1 : associée à la « hauteur » (degré d'ouverture du conduit vocal) F2 : associé à la position antéro-postérieure de la langue Aire du triangle vocalique : degré de réduction du système vocalique mesuré en Hz Analyse des consonnes vélaires : /k/ : durée de relâchement de la pression d'air (courte période silencieuse de montée en pression + relâchement de la pression). /x/ : pente spectrale	<i>Mesures :</i> corrélation entre paramètres acoustiques et évaluation subjective en aveugle de l'articulation et de la résonance nasale (échelle à 4 niveaux) <i>Juges :</i> 2 orthophonistes <i>Tâche :</i> texte lu	Voyelles cardinales en néerlandais (/a/, /i/, /u/) et consonnes vélaires (/k/, /x/) issues d'un texte lu d'une durée d'environ 60 secondes	Sur les voyelles : F1 /i/ : r = -0,42 avec la résonance nasale F2 /i/ : r = 0,35 avec l'intelligibilité et l'articulation Comparaison entre l'évaluation subjective et l'aire du triangle vocalique : r = 0,39 (p<0,05) avec l'intelligibilité, r = 0,42 (p<0,05) avec l'articulation Sur les consonnes vélaires : /k/ : r>0,40 (significativement différent) pour l'intelligibilité et l'articulation /x/ : r=0,33 (p<0,05) avec la nasalité Paramètres impliqués dans la prédiction de l'intelligibilité : durée du relâchement de la pression d'air sur /k/, aire du triangle vocalique et F1 /i/ (p<0,05)

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(de Bruijn et al., 2011)	III-2	20/22 91 % (forte)	<i>Participants :</i> 51 patients (28 M, 23 F) traités par chirurgie (avec reconstruction) pour un cancer avancé oral ou oropharyngé (21 cavité buccale, 30 oropharynx) Inclus 6 mois après le traitement <i>Taille de la tumeur :</i> T2=26, T3-4=25 <i>Langue :</i> Néerlandais	53,8 ans, écart-type 8,7 Étendue : 23 - 73 ans	Réseaux de neurones (ANN) : représentation des caractéristiques d'un signal de parole d'entrée, contenant un certain nombre de paramètres de modèle (poids des connexions entre les nœuds du réseau) qui déterminent la relation entre l'entrée (dans cet article, contexte d'entrée de 7 MFCC d'entrée) et la sortie (spécifiquement la nasalité dans cette étude : ANN-nasalance).	<i>Mesures :</i> régression linéaire multivariée (pour l'intelligibilité) ou logistique (pour la qualité de l'articulation et l'hypernasalité) pour obtenir un aperçu du rôle des paramètres objectifs dans l'évaluation subjective de la parole (évaluation en aveugle de l'articulation et de la résonance nasale (à l'aide d'une échelle à 4 niveaux) <i>Juges :</i> 2 orthophonistes <i>Tâche :</i> texte lu	Deux réalisations de voyelles cardinales en néerlandais (/ɑ/, /i/, /u/) dans différents contextes phonologiques (consonnes occlusives, liquides et nasales) Voyelles extraites d'un texte de lecture d'une durée approximative de 60 secondes	Prédictions par la quantité sur la nasalance (ANN-nasalance) : Intelligibilité ($R^2=21,3\%$) : 2^{ème} réalisation de /ɑ/ ($p=0,03$) et /i/ ($p=0,023$) Qualité de l'articulation ($R^2=48,7\%$) : 2^{ème} réalisation de /ɑ/ ($p=0,05$) Hypernasalité ($R^2=24,9\%$) : 1^{ère} réalisation de /i/ ($p=0,048$) et /u/ ($p=0,008$) Les analyses sur /ɑ/ prédisent la qualité de l'articulation, sur /i/ et /u/ prédisent l'hypernasalité évaluée perceptivement.

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(de Carvalho-Teles et al., 2008)	III-3	18/22 82 % (forte)	<i>Participants :</i> 36 patients (33 M, 3 F) traités par glossectomie partielle, hémiglossectomie, et utilisant une prothèse d'abaissement du palais stable et adaptée depuis au moins 3 mois <i>Taille de la tumeur :</i> non précisée <i>Langue :</i> Portugais	Étendue 30-80 ans	Évaluation spectrographique des formants (valeurs moyennes de F1, F2 et F3 extraites de la partie la plus stable de chaque voyelle tenue 5 secondes environ) des 7 voyelles du portugais brésilien, avec et sans prothèse	<i>Mesures :</i> formants avec et sans prothèse	Répétition de 18 syllabes avec des sons plosifs, fricatifs, voisés et sourds, qui sont des sons nasaux et liquides avec la voyelle /a/. Voyelles tenues dans la fréquence et l'intensité habituelles : /a/, /e/, /é/, /i/, /o/, /ó/, /u/	Avec et sans prothèse : F1 : différence statistiquement significative pour /a/, /e/, /u/ ($p < 0,001$), et proche du seuil de significativité pour /o/ ($p = 0,09$). F2 : différence significative pour /o/, /ó/, /u/ ($p < 0,001$) et proche du seuil de significativité pour /e/, /i/ ($p = 0,06$ et $p = 0,08$, respectivement). F3 : différence significative pour /a/, /ó/ (analyse de variance, $p < 0,001$).

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Dwivedi et al., 2016)	III-2	24/24 100 % (forte)	<i>Participants :</i> 62 patients (41 M, 21 F) avec cancer oral (21 : 18 langue, 3 plancher buccal) ou oropharyngé (41 : 13 base de langue, 26 amygdale, 2 palais mou), traité par chirurgie <i>Taille de la tumeur :</i> T1=19, T2=31, T3=6, T4=6 <i>Langue :</i> Anglais	Moyenne : 58,9 ans	F0, F1 et F2 analysés sur la partie centrale stable de la voyelle soutenue /i/ (contenant le maximum de périodes possibles - pas moins de 200 millisecondes pour F0). Technique LPC pour évaluer les fréquences des formants F1 et F2	<i>Mesures :</i> comparaison de F1, F2, F0 et évaluation perceptive entre les patients et les contrôles (échelle de Likert à 4 niveaux sur l'intelligibilité, l'articulation, la nasalité, le débit et la faiblesse, note globale) <i>Juges :</i> 3 orthophonistes expérimentés <i>Tâche :</i> extrait d'un texte lu ("The story of Arthur the rat")	Voyelle /i/ tenue à une hauteur et à une intensité confortables (au moins 5 secondes). Pour l'évaluation perceptive : lecture de mots spécifiques (bead, bed, bood) et récitation d'un passage standard à une hauteur et une intensité confortables	Absence de corrélation entre F1 et F2 sur /i/ et les paramètres perceptifs de la parole (grade global, intelligibilité et articulation) ($p>0,15$) Chez les patients : F1 : affectée par l'élévation de la langue, la fermeture de la bouche et la constriction du pharynx. F2 : augmentation avec l'élévation antérieure de la langue ou l'abaissement de la région postérieure de la cavité buccale F0 : augmentation chez les patients masculins atteints d'un cancer de la cavité buccale (surtout de la langue)

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Fang et al., 2017)	III-2	14/20 70 % (bonne)	<i>Corpus 1</i> NKI-CCRT (Clapham, 2012) : 55 patients (45 M, 10 F) avec cancer ORL, traités par chirurgie et chimiothérapie <i>Taille de la tumeur</i> : T1=8, T2=15, T3=21, T4=11 <i>Corpus 2</i> SVD : 2000 individus avec 71 sujets (troubles organiques et/ou fonctionnels, le genre n'est pas précisé) <i>Langue des deux corpus</i> : Allemand	Non précisé, mais selon (Clapham et al., 2012) : moyenne 58 ans, étendue 32-79	MFCC (Mel-Frequency cepstrum coefficients) MSCC (Mel S-transform cepstrum coefficients) : la transformation S est une méthode d'analyse temps-fréquence qui combine les avantages de la transformée en ondelettes avec la transformée de Fourier à court terme (meilleur débruitage, résolution temporelle et localisation temps-fréquence)	<i>Mesures</i> : Sensibilité (Se), spécificité (Sp), aire sous la courbe (UA) and Accuracy calculés sur les MFCC et MSCC dans les deux groupes	Corpus 1 : lecture d'un texte neutre en allemand Corpus 2 : Enregistrements des voyelles /a/, /i/, /u/ produites à une hauteur normale, haute, basse et basse-haute-basse, et enregistrements d'une phrase en allemand	Se : 67,15 % (MSCC), 56,25 % (MFCC) Sp : 62,36 % (MSCC), 46,90 % (MFCC) UA : 64,75 % (MSCC), 51,58 % (MFCC) Accuracy : 63,67 % (MSCC), 50,54 % (MFCC) Les paramètres MSCC améliorent de manière significative le taux de classification entre intelligible et non intelligible par rapport à la MFCC sur les deux corpus.

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Ha et al., 2016)	III-2	20/22 91 % (forte)	<i>Participants :</i> 13 patients (8 M, 5 F) traités par glossectomie latérale post-partielle <i>Taille de la tumeur :</i> T1=7, T2=6 <i>Langue :</i> Anglais américain	Moyenne : 45,3 ans	Les 3 premiers formants du segment médian des voyelles /i/ et /u/, extraits automatiquement par méthode LPC (Linear Prediction Coding)	<i>Mesures :</i> calculs de (F2/F1), (F3/F2), (F3/F1), et comparaisons avec un groupe contrôle de 23 participants	Voyelles /i/ et /u/ à partir de sept répétitions de « a geese » et « a souk ».	Comparativement aux témoins, les patients présentent des rapports F2/F1 significativement plus faibles (F=5,911, p=0,018), et des rapports F3/F1 plus faibles, proches du seuil de significativité (F=3,482, p=0,067). Dans l'analyse des formants, les F2 et F3 des patients sont plus faibles.

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Jacobi et al., 2013)	III-3	19/20 95 % (forte)	<i>Participants :</i> 34 patients (27 M, 7 F) traités par radio-chimiothérapie d'un cancer ORL avancé 3 groupes : BT : 8 base de langue, 1 trigone rétromolaire NT : 6 amygdales, 1 malais mou, 5 autres L : 13 larynx et hypopharynx <i>Taille de la tumeur :</i> T1=6, T2=12, T3=12, T4=4 <i>Langue :</i> Néerlandais	Médiane : 58 ans Étendue : 39 – 77	F1 et F2 sur les voyelles /a/, /i/, /u/ Nasalité sur /a/ : énergie relative de la bande dans la zone des 2^{ème} et 3^{ème} formants. Centre de la fréquence de l'explosion et de l'énergie sur /p/, /t/, /k/ Point de coupure inférieur du spectre de frictions, centre de gravité et énergie de la bande sur /s/, /z/, /x/. Présence du /t/ avant le bruit de friction indiquée acoustiquement par une explosion et une rupture d'énergie sur /t j/ F1, F2 et F3 et leurs amplitudes pour /l/, /r/	<i>Mesures :</i> comparaisons appariées entre avant et après le traitement	Texte standard en néerlandais Diadococinésies (répétition de « pataka ») Liste de mots (DYVA)	Dix semaines après la fin du traitement : Nasalité : diminution par rapport à la ligne de base (p = 0,062 et t = 1,935). Différences significatives pour les productions /t/ (fréquence spectrale plus élevée), /s/ (plus diffuse) et /r/ (augmentation de F2 et F3). 1 an après la fin du traitement, différences significatives (p<0,05) pour /r/, /k/ (plus faibles), /l/ (F3 significativement plus élevé avec intensité plus faible), /x/ (mesures de fréquences plus faibles en position initiale et finale), /s/, /t/, /t j/

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Kazi et al., 2007)	III-2	20/20 100 % (forte)	<i>Participants :</i> 26 patients (19 M, 7 F) traités par glossectomie partielle <i>Taille de la tumeur :</i> non précisée <i>Langue :</i> non précisée	Moyenne : 53.1 ans, écart-type 8.7	F1, F2 et F3 : coefficients LPC moyennés (pour une estimation pour chaque locuteur)	<i>Mesures :</i> corrélation des fréquences des formants entre 31 sujets contrôle et le groupe de patients	Voyelle /i/ tenue, produite à une hauteur et une intensité confortables pendant au moins 5 secondes.	Corrélations de formants significatives : - chez les femmes : entre les patientes et les sujets sains sur F2 ($p=0,04$) et F3 ($p=0,02$) - chez les hommes : entre les patients et les sujets sains sur F1 ($p=0,01$). Pas d'autres corrélations de formants significatives sur les autres comparaisons.

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Kim et al., 2014)	III-3	19/22 86 % (forte)	<i>Participants:</i> corpus NKI-CCRT (Clapham et al., 2012) : 55 patients (45 M, 10 F) traités par chirurgie et radio-chimiothérapie d'un cancer ORL : 5 cavité buccale (2 plancher buccal, 3 langue), 24 oropharynx (10 base de langue, 7 amygdale, 2 palais mou, 5 autres), 26 autres (larynx, nasopharynx) <i>Taille de la tumeur :</i> T1=8, T2=15, T3=21, T4=11 <i>Langue :</i> Néerlandais	Non précisé (mais selon (van der Molen et al., 2009) : moyenne 58 ans, étendue 32-79)	MRSTC : codage par transformée sinusoïdale multirésolution utilisant une analyse de type ondelette (composantes de basse fréquence calculées sur une plus grande longueur de fenêtre d'analyse, composantes de haute fréquence estimées avec une longueur de fenêtre plus courte).	<i>Mesures :</i> Accuracy (ACC) définie comme : (nombre d'occurrences) / (nombre de présences totales) Accuracy non pondérée (Unweighted accuracy, UWA): $UWA = \frac{1}{C} \sum_{c=1}^C \frac{\text{Nb of hits in class } c}{\text{Nb of instances in class } c}$ Comparaison entre le niveau de base SVM (Support Vector Machine) et le SVM de base + MRSTC	Lecture d'un passage de 189 mots tiré d'un conte néerlandais	Sur le classificateur d'intelligibilité dichotomique : ACC : 68,0 (SVM), 72,7 (SVM + MRSTC) UWA : 66.2 (SVM), 71.2 (SVM + MRSTC) Amélioration de l'accuracy et de l'accuracy non pondérée avec l'ajout de MRSTC au SVM

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Knuuttila & Juhani Pukander, 1999)	III-3	16/20 80 % (forte)	<i>Participants :</i> 9 patients (8 M, 1 F) opérés pour un cancer lingual <i>Taille de la tumeur :</i> T1=1, T2=6, T3=2 <i>Langue :</i> Finnois	Étendue : 43 – 75 ans	Premier (F1) et deuxième (F2) formant estimés sur une partie stable de la voyelle tenue (l'analyse par transformation de Fourier rapide a été utilisée pour les analyses spectrales)	<i>Mesures :</i> différences dans les mesures acoustiques des voyelles après et avant la résection de la langue (F1 et F2)	16 voyelles du finnois tenues, produites deux fois	Différence significative uniquement pour F1 de /i/ (p=0,01) et F2 de /ɑ/ (p=0,001). F1 de /i/ plus élevé (moyenne : 32, p=0,05) et F2 de /ɑ/ plus faible (moyenne : -97, p=0,01) après la chirurgie

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Kraaijenga et al., 2015)	III-3	19/22 86 % (forte)	<i>Participants</i> : 18 patients traités par radio-chimiothérapie simultanée en 2 groupes : Groupe NT : 6 tumeurs naso-pharyngées et de l'amygdale Groupe LHBT : 12 tumeurs laryngées, hypo-pharyngées ou de la base de langue <i>Taille de la tumeur</i> : non précisée <i>Langue</i> : non précisée	Moyenne : 63 ans, étendue 45 – 79	Nasalité mesurée avec le logiciel Praat	<i>Mesures</i> : comparaison des antiformants entre une base de référence (post-chirurgie) et à 2 et 6 ans après le traitement <i>Tâches</i> : voyelle /a/ tenue	Voyelle /a/ tenue	Amélioration à 2 ans et détérioration à 6 ans dans le « groupe NT ». Détérioration par rapport à la ligne de base également dans le « groupe LHBT » (test t apparié p = 0,087).
(Kumar et al., 2013)	III-3	16/22 73 % (bonne)	<i>Participants</i> : 10 patients (9 M, 1 F) traités par maxillectomie <i>Taille de la tumeur</i> : non précisée <i>Langue</i> : Hindi	Moyenne : 43 ans, étendue 15 – 75	Nasalité mesurée par le module « nasal view » du logiciel Dr. Speech	<i>Mesures</i> : pourcentage automatique de nasalité à différents moments (préopératoire, lors de la cicatrisation sans obturateur, et 24h et 6 semaines après la pose de l'obturateur)	Production d'une voyelle tenue	Augmentation brusque de la nasalité après la chirurgie : de 20,16 +/- 5,52 à 52,04 +/- 19,25 (p = 0,005).

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Laaksonen et al., 2011)	III-3	21/22 95 % (forte)	<i>Participants :</i> 17 patients (11 M, 6 F) avec cancer de la langue (les 2/3 antérieurs de la langue ont été reconstruits par lambeau libre antébrachial) <i>Taille de la tumeur :</i> T2=14, T3=3 <i>Langue:</i> Anglais canadien	M : 53 ans (moyenne), étendue 27 – 67 F : 59 ans (moyenne), étendue 17 – 72	Moments spectraux du spectre à long terme moyen (LTA) : moyenne (premier moment) et asymétrie (troisième moment). Durée de friction des sons sibilants	<i>Mesures :</i> Comparaison des mesures réalisées avant la résection de la langue, puis 1, 6 et 12 mois après la reconstruction	Lecture de six phrases et du passage « Zoo » : analyse de 892 cibles pour /s,z/, et 88 cibles pour /ʃ/.	Moyenne spectrale : Réduction de la distinctivité acoustique entre /s, z/ et /ʃ/ Diminution d'abord (préopératoire vs. 1 mois postopératoire, $p < 0,0001$, différence moyenne : 1094 Hz) puis a augmenté (1 mois vs. 6 mois postopératoire, $p < 0,0001$, différence moyenne : 752 Hz) sans revenir aux niveaux préopératoires (préopératoire vs. 12 mois postopératoire, $p < 0,0001$, différence moyenne : diminution de 654 Hz). Les femmes sont revenues progressivement au niveau préopératoire en l'espace d'un an (préopératoire vs. 12 mois postopératoire, $p < 0,0001$, différence moyenne : 1142 Hz) Asymétrie : effets à court terme pour les deux groupes Les femmes ont atteint le niveau préopératoire 1 an après l'opération (asymétrie : $p=0,01$, différence moyenne : 0,47). Durée de la friction sur /s, z/ : effets à long terme non trouvés ($p>0,05$)

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Laaksonen et al., 2010)	III-3	17/20 85 % (forte)	<i>Participants :</i> 18 (12 M, 6 F) patients avec cancer lingual (reconstruction des 2/3 antérieurs de la langue) <i>Taille de la tumeur :</i> T2=15, T3=3 <i>Langue :</i> Anglais canadien	Étendue 27 – 72 ans	F1, F2, F0 et la durée ont été analysés. Les fréquences formantiques ont été obtenues en utilisant l'analyse de codage prédictif linéaire (LPC)	<i>Mesures :</i> Comparaison des paramètres avant la résection de la langue et puis 1, 6 et 12 mois après la reconstruction de la langue	Voyelles /i/, /I/, / [^] / and /u/ (choisies en raison de leur spécificité articulatoire, acoustique et perceptive) et diphtongues /aI/, /eI/ d'un ensemble de 6 phrases comprenant des séquences [hVd] (par exemple « heed » ou « who'd »).	Chez les hommes, des effets à long terme ont été observés pour F2 et dans la durée des voyelles. Le F2 a diminué (interaction : F=3,262, p=0,002 ; pré-op. vs 12 mois post-op. (comparaison appariée) p=0,003, différence moyenne 68 Hz). Chez les femmes, aucun changement statistiquement significatif n'a été observé (toutes les valeurs p > 0,05) pour l'une ou l'autre des mesures de résultats.
(Markkanen-Leppänen et al., 2005)	III-3	22/22 100 % (forte)	<i>Participants :</i> 44 patients (29 M, 15 F) opérés d'un cancer oral (OC), oropharyngé (OP) ou hypopharyngé <i>Taille de la tumeur :</i> T2=20, T3=5, T4=11, récurrence tumorale=8 <i>Langue :</i> non précisée	Moyenne : 56.2 ans, étendue 38 – 80	La résonance de la parole est évaluée par la nasalité (c'est-à-dire la modification acoustique objective et spécifique de la nasalité perçue) avec un nasomètre.	<i>Mesures :</i> Comparaison des valeurs avant et après la chirurgie (après 6 semaines, 3 mois, 6 mois, 12 mois)	Phrases constituées de consonnes plosives sourdes ou de consonnes voisées. Chaque phrase est répétée 3 fois	Nasalité normale avant et après chirurgie dans le groupe de patients OC. Chez les patients OP, cependant, la nasalité a augmenté après l'opération, différant significativement des patients OC 6 semaines et 6 mois après l'opération (p<0,05).

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Moerman et al., 2003)	IV	14/22 64 % (bonne)	<i>Participants :</i> 4 patients traités de lésions oropharyngées avec extension vers le palais mou (3 M, 1 F) <i>Taille de la tumeur :</i> T2=2, T3=2 <i>Langue :</i> non précisée	Non précisé	Mesure de la nasalité au moyen d'un nasomètre	<i>Mesures :</i> Comparaison entre les scores moyens de nasalité	Lecture de textes contenant des phonèmes oraux, des phonèmes nasaux, et une combinaison des deux	Scores normaux, quel que soit le texte
(Seikaly et al., 2003)	III-3	17/20 85 % (forte)	<i>Participants :</i> 18 patients (12 M, 6 F) opérés d'un cancer oropharyngé <i>Taille de la tumeur :</i> T1=1, T2=4, T3=10, T4=3 <i>Langue :</i> non précisée	Moyenne : 55.1 ans (45 – 75)	Nasalité mesurée par nasomètre, PERCI-SARS et le Computerized Assessment of Intelligibility of Dysarthric Speech (CAIDS)	Comparaison entre les scores moyens de nasalité à trois moments précis : avant la chirurgie, 1 mois après la chirurgie et avant la radiothérapie (RT), et 6 à 9 mois après la chirurgie et la fin de la radiothérapie.	Non précisé	Les valeurs de nasalité préopératoire différaient significativement des valeurs de temps pré-RT ($p=0,05$) Les scores préopératoires d'intelligibilité des mots différent significativement des valeurs pré-RT ($p<0,01$) et post-RT ($p<0,05$). Aucune différence significative pour aucune de ces variables entre le temps pré-RT et le temps post-RT.

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Takatsu et al., 2017)	III-3	19/20 95 % (forte)	<i>Participants :</i> 62 patients avec cancer lingual Groupe 1 (G1) : 40 (29 M, 11 F) avec glossectomie partielle <i>Taille de la tumeur :</i> T1=6, T2=34 Groupe 2 (G2) : 22 (15 M, 7 F) avec reconstruction <i>Taille de la tumeur :</i> T1=1, T2=4, T3=3, T4=14 <i>Langue :</i> Japonais	G1 : 55.8 ans, étendue 30 – 77 G2 : 47.5 ans, étendue 21-69 ans	Fréquence formantique des voyelles (centre des intervalles de 30 ms des voyelles tenues, en excluant les 25 premières et dernières ms) : F1, F2 et pentes des transitions de formants (formant slopes) entre les diphtongues /ai/ et /ui/. Aire du triangle vocalique : aire intérieure d'un triangle pour chaque patient, obtenue en traçant F1 et F2 pour 3 voyelles (/a/, /i/ et /u/)	<i>Mesures :</i> Comparaison des données sur les caractéristiques acoustiques recueillies pendant les périodes préopératoire et postopératoire, et après la rééducation.	Voyelles tenues : 3 secondes de /a/, /i/, /u/. Tâche de répétition de mots : /taiko/, /tʃisai/, /kaiko/, /suika/ ont été produits 3 fois chacun.	Le F2 de /i/ a significativement diminué après la glossectomie (p<0,001) G1 : la pente du F1 a diminué pendant la période postopératoire (/tai/ : p=0,007 ; /kai/ : p=0,042). La pente F2 a significativement diminuée pour toutes les diphtongues (p<0,001). G2 : Les valeurs de F1 ont augmenté pour /i/ (p=0,064) et /u/ mais ont significativement diminué pour /a/ (p=0,021). Les valeurs de F2 ont diminué pour toutes les voyelles, avec une différence significative pour /i/ (p<0,001) et /u/ (p=0,002).

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Wakumoto et al., 1996)	III-3	18/20 90 % (forte)	<i>Participants :</i> 10 patients traités par glossectomie Groupe 1 (G1) : pas de reconstruction (4 M, 1 F) <i>Taille de la tumeur :</i> T1=1, T2=3, T4=1 Groupe 2 (G2): 5 reconstruits par Lambeau antébrachial (5 M) <i>Taille de la tumeur :</i> T2=2, T3=2, T4=1 <i>Langue :</i> non précisée	G1 : médiane 37 ans, étendue 28 – 67 G2 : médiane 55 ans, étendue 49 – 63	Caractéristiques fréquentielles au niveau de la consonne avec l'enveloppe spectrale extraite par FFT des coefficients LPC : calcul de la fréquence du formant par la méthode du peak picking CPF (Consonant peak energy frequency) : score visant à évaluer quantitativement les caractéristiques fréquentielles au niveau de la consonne. Variance des formants à la partie transitoire : F2-F3	<i>Mesures :</i> Comparaison des scores recueillis avant l'opération, et 1, 6 et 12 mois après l'opération.	Prononciation de la syllabe cible /ta/, sélectionnée parmi les échantillons du test d'intelligibilité de la parole en raison de sa prononciation avec la fermeture antérieure de la cavité buccale à l'aide de la pointe de la langue.	G1 : CPF : certains ont obtenu des scores plus faibles en préopératoire qu'en post-opératoire F2-F3 : certains ont montré des scores proches de la ligne de base 1 mois après l'opération G2 : CPF : certains ont révélé une tendance à la concentration approximativement dans la zone de 3-5 kHz 1 mois après l'opération. F2-F3 : 1 mois après l'opération, certains ont montré des scores proches de la ligne de base.

Référence	Type d'étude (NHMRC)	QualSyst	Participants, diagnostic, langue	Âge des patients	Paramètres acoustiques / automatiques (définitions)	Critère(s) de comparaison	Échantillon(s) de parole	Principale(s) conclusion(s)
(Yoshida et al., 2000)	III-3	19/20 95 % (forte)	<i>Participants :</i> 15 patients (4 M, 11 F) traités pour différents types de cancers palato-maxillaires ou du sinus maxillaire <i>Taille de la tumeur :</i> non précisée <i>Langue :</i> non précisée	Étendue 38 – 78 ans	Analyses spectrales sur des spectres de 1/3 d'octave obtenus à partir de données transformées par analyse FFT (Fast Fourier transform) et filtrées en bande passante dans la gamme de fréquences de 125 Hz à 6,3 kHz	<i>Mesures :</i> évaluation de l'hypernasalité sur une échelle à 5 niveaux <i>Juges :</i> 2 orthophonistes	Voyelle /i/ tenue pendant 0,5 seconde à une hauteur et une intensité confortable pour chaque sujet	Corrélation élevée entre les évaluations perceptives et les valeurs prédites ($r=0,8419$, r^2 ajusté= $0,6872$, $F[2,27]=32,8480$, $P<0,001$) par régression (avec le score perceptif comme variable réponse, et F7 F12, les amplitudes des 7e et 12e multiples dans les spectres normalisés de 1/3 d'octave, comme variables explicatives).

Annexe II Paramètres acoustiques et automatiques étudiés dans les études incluses

	Nasalité		Voyelles				Consonnes				Mesures globales	
	Logiciel dédié ou nasomètres	Rapport de production d'énergie entre cavité buccale et cavité nasale	F1 et F2 (+/- F3)	F1 à F12	Aire du triangle vocalique	Pente	Moments spectraux	Durée de relâchement de l'air	Durée de friction et énergie de la bande	F1, F2, F3 sur /l/ and /r/	Extraction de MFCC, MSCC, MRTSC	Réseaux de neurones - nasalité
(Acher et al., 2014)							✓					
(Chung et al., 2011)		✓										
(de Bruijn et al., 2009)			✓		✓		✓	✓				
(de Bruijn et al., 2011)												✓
(de Carvalho-Teles et al., 2008)			✓									
(Dwivedi et al., 2016)			✓									
(Fang et al., 2017)											✓	
(Ha et al., 2016)			✓									
(Jacobi et al., 2013)		✓					✓	✓	✓	✓		
(Kazi et al., 2007)			✓									
(Kim et al., 2014)											✓	
(Knuuttila & Juhani Pukander, 1999)			✓									
(Kraaijenga et al., 2015)	✓											
(Kumar et al., 2013)	✓											
(Laaksonen et al., 2011)							✓		✓			
(Laaksonen et al., 2010)			✓									
(Markkanen-Leppänen et al., 2005)	✓											
(Moerman et al., 2003)	✓											
(Seikaly et al., 2003)	✓											
(Takatsu et al., 2017)			✓		✓	✓						
(Wakumoto et al., 1996)							✓					
(Yoshida et al., 2000)				✓								

Annexe III Grille ECCS



Institut de Recherche
en Informatique de Toulouse
projet ANR-14-011-0122



UNIVERSITÉ
TOULOUSE III
PAUL SABATIER



ÉVALUATION DE LA CONSTITUTION DES CERCLES SOCIAUX

Pour chaque groupe de personnes, veuillez indiquer vous concernant :

- Le **nombre de personnes** que vous estimez faire partie du groupe concerné (colonne A) ;
- La **fréquence** moyenne à laquelle vous avez des contacts avec ces personnes (colonne B) ;
- Comment vous jugez la **la fréquence** de vos contacts avec ces personnes : de -5 « trop faible » (Je ne les vois pas suffisamment) à +5 « trop élevée » (Je les vois trop souvent à mon goût) (colonne C) ;
- Si vous êtes **satisfait.e de la qualité des échanges** que vous avez avec ces personnes quand vous les voyez : de -5 « très insatisfait » à +5 « très satisfait » (colonne D).

Groupe	A	B	C	D
		Régularité moyenne des contacts (cocher la case)	Satisfaction globale de la fréquence (cocher la case)	Satisfaction globale des échanges (cocher la case)
Privé <i>Personnes avec qui je vis au quotidien.</i>	Avec combien de personnes vivez-vous au quotidien ? ☐ 0 : je vis seul.e ☐ 1 ☐ 2 à 4 ☐ 5 et plus	☐ tous les jours ☐ plusieurs fois par semaine ☐ plusieurs fois par mois ☐ plusieurs fois par an ☐ une fois par an ou moins	☐ -5 : trop faible ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : suffisant ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : trop élevée	☐ -5 : très insatisfait ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : neutre ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : très satisfait
Famille <i>Famille que je ne vois pas forcément tous les jours mais avec qui j'ai des contacts étroits et réguliers.</i>	Combien de personnes y a-t-il dans votre entourage familial ? ☐ 0 ☐ 1 ☐ 2 à 4 ☐ 5 à 9 ☐ 10 et plus	☐ tous les jours ☐ plusieurs fois par semaine ☐ plusieurs fois par mois ☐ plusieurs fois par an ☐ une fois par an ou moins	☐ -5 : trop faible ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : suffisant ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : trop élevée	☐ -5 : très insatisfait ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : neutre ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : très satisfait
Amis <i>Personnes pour qui j'éprouve de la sympathie, que je connais bien et avec qui je me sens bien, avec qui j'ai des liens privilégiés et en qui j'ai confiance.</i>	Combien de personnes y a-t-il dans votre entourage amical ? ☐ 0 ☐ 1 ☐ 2 à 4 ☐ 5 à 9 ☐ 10 et plus	☐ tous les jours ☐ plusieurs fois par semaine ☐ plusieurs fois par mois ☐ plusieurs fois par an ☐ une fois par an ou moins	☐ -5 : trop faible ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : suffisant ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : trop élevée	☐ -5 : très insatisfait ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : neutre ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : très satisfait

Groupe	A	B	C	D
		Régularité moyenne des contacts (cocher la case)	Satisfaction moyenne de la fréquence (cocher la case)	Satisfaction moyenne de la qualité (cocher la case)
Connaissances <i>Personnes que j'ai rencontrées, avec qui j'ai des contacts réguliers mais avec qui je n'ai pas de relation privilégiée (que je connais peu).</i>		☐ tous les jours ☐ plusieurs fois par semaine ☐ plusieurs fois par mois ☐ plusieurs fois par an ☐ une fois par an ou moins	☐ -5 : trop faible ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : suffisant ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : trop élevée	☐ -5 : très insatisfait ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : neutre ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : très satisfait
Aidants – communauté <i>Personnes que je côtoie, à qui je dis « bonjour » et « au revoir », avec qui je partage des intérêts communs sans avoir avec eux de relation privilégiée.</i>		☐ tous les jours ☐ plusieurs fois par semaine ☐ plusieurs fois par mois ☐ plusieurs fois par an ☐ une fois par an ou moins	☐ -5 : trop faible ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : suffisant ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : trop élevée	☐ -5 : très insatisfait ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : neutre ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : très satisfait
Inconnus		☐ tous les jours ☐ plusieurs fois par semaine ☐ plusieurs fois par mois ☐ plusieurs fois par an ☐ une fois par an ou moins	☐ -5 : trop faible ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : suffisant ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : trop élevée	☐ -5 : très insatisfait ☐ -4 ☐ -3 ☐ -2 ☐ -1 ☐ 0 : neutre ☐ +1 ☐ +2 ☐ +3 ☐ +4 ☐ +5 : très satisfait

Annexe IV Caractéristiques des sujets inclus (total et par lots)

	Total		Lot 1		Lot 2		Comparaison inter-lots	
							Valeur p	Test
Effectif total	27		13	48,1 %	14	52,9 %		
Inclus	25	92,6%	12	92,3%	13	92,9%		
Exclus	2	7,4%	1	7,7%	1	7,1%		
Genre	25		12	48,0 %	13	52 %	0,42	<i>Fischer</i>
Homme	15	60,0%	6	50,0%	9	69,2%		
Femme	10	40,0%	6	50,0%	4	30,8%		
Âge	25		12		13		0,79	<i>Mann-Whitney</i>
Minimum	38		53		38			
Maximum	83		83		75			
Moyenne	65,8		66,08		65,54			
Écart-type	9,44		9,08		10,12			
Quartile 1	61		59,75		62			
Médiane	67		65		68			
Quartile 3	73		72,5		73			
Écart interquartile	12		12,75		11			
Département naissance	22		12		12		0,41	χ^2
Occitanie	10	45,5%	6	50,0%	4	33,3%		
Autre	13	59,1%	6	50,0%	8	66,7%		
Département résidence	25		12		13		0,59	<i>Fischer</i>
Occitanie	22	88,0%	10	83,3%	12	92,3%		
Autre	3	12,0%	2	16,7%	1	7,7%		

Lot	Total		Lot 1		Lot 2		Comparaison inter-lots	
							Valeur p	Test
Localisation de la tumeur	25		12		13		0,55	<i>Fischer</i>
Cavité buccale	14	56,0%	6	50,0%	8	61,5%		
Oropharynx	10	40,0%	6	50,0%	4	30,8%		
Cavité buccale et oropharynx	1	4,0%	0	0,0%	1	7,7%		
Détail de la localisation	22	88,0%	11	91,7%	11	84,6%		
Chirurgie	25		12		13		0,10	<i>Fischer</i>
Non	3	12,0%	3	25,0%	0	0,0%		
Oui	22	88,0%	9	75,0%	13	100,0%		
Détail du type de chirurgie	15	60,0%	5	41,7%	10	76,9%		
Radiothérapie	25		12		13		1,00	<i>Fischer</i>
Non	1	4,0%	0	0,0%	1	7,7%		
Oui	24	96,0%	12	100,0%	12	92,3%		
Chimiothérapie	25		12		13		1,00	<i>Fischer</i>
Non	10	40,0%	5	41,7%	5	38,5%		
Oui	15	60,0%	7	58,3%	8	61,5%		
Délai depuis la fin du traitement, en mois	25		12		13		0,76	<i>Mann-Whitney</i>
Minimum	6		6		6			
Maximum	564		564		170			
Moyenne	87,2		111,9		64,38			
Écart-type	121,1		162,1		63,78			
Quartile 1	8		7,5		8			
Médiane	40		40,5		40			
Quartile 3	131		168		131			
Écart interquartile	123		160,5		123			
Classification TNM T	19		9		10		0,32	<i>Fischer</i>
0	0	0,0%	0	0,0%	0	0,0%		
1	2	10,5%	1	11,1%	1	10,0%		
2	6	31,6%	1	11,1%	5	50,0%		
3	3	15,8%	2	22,2%	1	10,0%		
4	8	42,1%	5	55,6%	3	30,0%		

Lot	Total		Lot 1		Lot 2		Comparaison inter-lots	
							Valeur p	Test
Classification TNM N	19		9		10		0,34	<i>Fischer</i>
0	8	42,1%	2	22,2%	6	60,0%		
1	3	15,8%	2	22,2%	1	10,0%		
2	7	36,8%	4	44,4%	3	30,0%		
3	1	5,3%	1	11,1%	0	0,0%		
Classification TNM M	19		9		10		1,00	<i>Fischer</i>
0	17	89,5%	9	100,0%	8	80,0%		
1	1	5,3%	0	0,0%	1	10,0%		
x	1	5,3%	0	0,0%	1	10,0%		
CSP	25		12		13		0,15	<i>Fischer</i>
Agriculteurs exploitants	0	0,0%	0	0,0%	0	0,0%		
Artisans, commerçants et chefs d'entreprise	4	16,0%	1	8,3%	3	23,1%		
Cadres et professions intellectuelles supérieures	2	8,0%	2	16,7%	0	0,0%		
Professions intermédiaires	2	8,0%	2	16,7%	0	0,0%		
Employés	4	16,0%	3	25,0%	1	7,7%		
Ouvriers	1	4,0%	0	0,0%	1	7,7%		
Retraités	12	48,0%	4	33,3%	8	61,5%		
Autres personnes sans activité professionnelle	0	0,0%	0	0,0%	0	0,0%		
Détail profession	21	84,0%		0,0%	9	69,2%		
Statut marital	25		12		13		0,25	<i>Fischer</i>
Célibataire	3	12,0%	3	25,0%	0	0,0%		
Pacsé	2	8,0%	1	8,3%	1	7,7%		
Marié	15	60,0%	6	50,0%	9	69,2%		
Veuf	1	4,0%	1	8,3%	0	0,0%		
Divorcé	4	16,0%	1	8,3%	3	23,1%		
Statut auditif	25		12		13		0,85	<i>Fischer</i>
Pas de notion de plainte	16	64,0%	7	58,3%	9	69,2%		
Plainte par le patient d'un déficit auditif	6	24,0%	3	25,0%	3	23,1%		
Appareillage	3	12,0%	2	16,7%	1	7,7%		

Annexe V Éléments de compréhension du fonctionnement des systèmes de reconnaissance automatique de la parole

Un système de reconnaissance automatique de la parole permet, à partir du signal acoustique, de produire une séquence d'unités (mots ou sons élémentaires appelés phones), notamment au moyen d'une approche probabiliste.

Ce signal acoustique est initialement *prétraité*, pour éliminer les informations non pertinentes (comme le bruit par exemple). Des paramètres spectraux (FBANK par exemple) ou cepstraux (MFCC : Mel-Frequency Cepstral Coefficients, les plus fréquemment utilisés dans la reconnaissance de la parole) sont alors extraits sur des fenêtres temporelles courtes. Ces fenêtres d'une durée de 20 à 50 ms peuvent être extraites avec un chevauchement, pour avoir des observations toutes les 10 ms. Leur durée très courte permet de faire l'hypothèse d'une stabilité du signal acoustique pendant une fenêtre temporelle, et autorise alors le calcul des paramètres acoustiques dans chaque fenêtre (ou observations acoustiques $X = x_1 x_2 \dots x_t$).

Le *système de reconnaissance* permet alors d'obtenir la sortie (phonème ou mot) la plus probable en fonction des observations acoustiques contenues dans le signal selon le principe du maximum de vraisemblance. En d'autres termes, il cherche parmi les différents mots W possibles celui qui maximise la probabilité d'observation en fonction du signal X :

$$W^* = \underset{W}{\operatorname{argmax}} P(W|X)$$

Une transformation par le théorème de Bayes et une simplification permet d'aboutir à la formule de calcul suivante :

$$W^* = \underset{W}{\operatorname{argmax}} \frac{P(X|W) P(W)}{P(X)} = \underset{W}{\operatorname{argmax}} P(X|W) P(W)$$

Un mot pouvant avoir plusieurs prononciations $Pr = (Pr_1 Pr_2 \dots Pr_n)$, en raison des différentes prononciations possibles des phonèmes qui le composent, $P(X|W)$ peut se décomposer ainsi : $P(X|W) = \sum_{Pr} P(X|Pr) P(Pr|W)$

La formule finale détaillée est donc :

$$W^* = \underset{W}{\operatorname{argmax}} P(X|W) P(W) = \underset{W}{\operatorname{argmax}} \sum_{Pr} P(X|Pr) P(Pr|W) P(W)$$

C'est à partir de cette formule qu'un système de reconnaissance automatique va pouvoir reconnaître une séquence d'unités.

De façon plus concrète, l'architecture de base d'un système de reconnaissance est décrite par Jelinek (Jelinek, 1976) (voir Schéma 1).

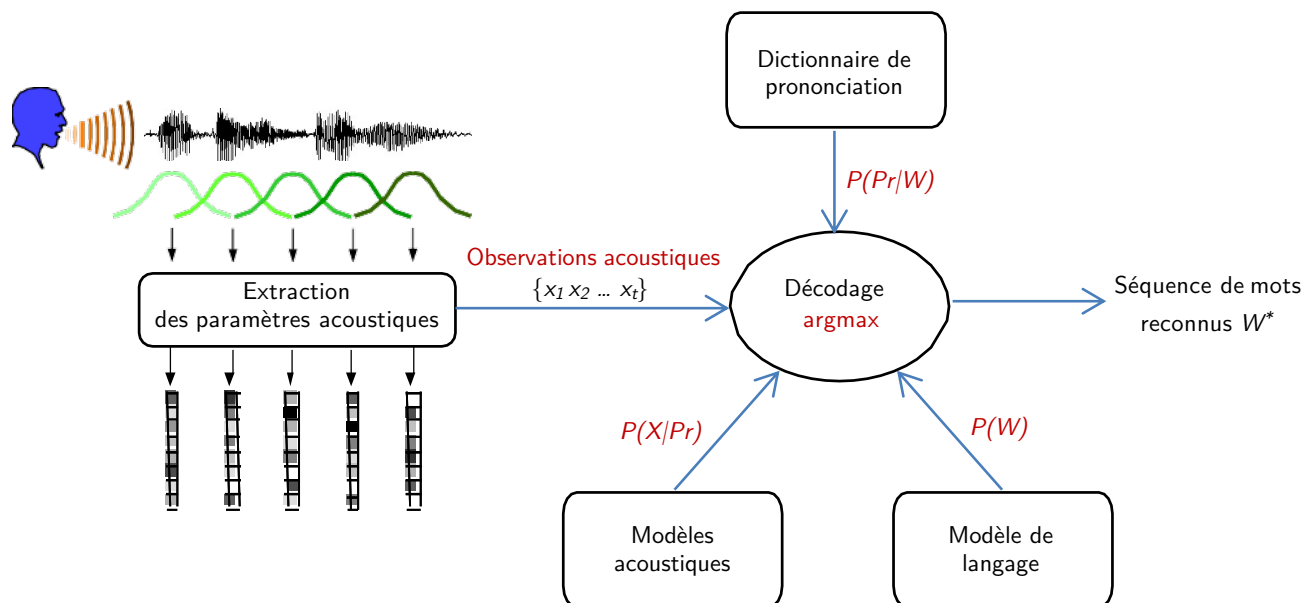


Schéma 1 - Architecture d'un système de reconnaissance de la parole adapté de Ghannay (Ghannay, 2017)

Le *modèle acoustique*, correspondant à $P(X|Pr)$, permet, à partir des paramètres acoustiques X , la reconnaissance des phonèmes les plus probables constituant une des prononciations Pr du mot W . L'ensemble des prononciations possibles d'un mot d'une langue donnée est contenu dans un *dictionnaire de prononciation*, correspondant à $P(Pr|W)$. Il permet donc la correspondance entre la prononciation reconnue par le modèle acoustique et le lexique d'une langue. Un exemple est donné dans le Schéma 2.

```
décelable d e s ^ l a b l
décela d e s ^ l a
décelai d e s ^ l E
décelaient d e s ^ l E
décelais d e s ^ l E
décelait d e s ^ l E
décelant d e s ^ l @
décèle d e s E l
décèle d e s E l
décelé d e s ^ l e
décelé d e s ^ l e
décelée d e s ^ l e
décelées d e s ^ l e
décèlera d e s E l ^ R a
décèlerait d e s E l ^ R E
décélération d e s e l e R a s j N
déceler d E s ^ l e
```

Schéma 2 - Extrait d'un dictionnaire de prononciation

Les modèles acoustiques s'appuient principalement sur des systèmes HMM (Hidden Markov Models) (Rabiner, 1989) : ils permettent de modéliser le phonème le plus probable correspondant à la prononciation Pr reconnue en fonction des

observations acoustiques X . Pour cela, les zones acoustiques stables sont modélisées à l'intérieur de ce modèle de Markov par des distributions issues de mélanges de modèles gaussiens (ou GMM). Pour connaître les distributions des phonèmes et donc établir $P(X|Pr)$, le modèle acoustique doit être préalablement construit à partir de corpus annotés (ou corpus d'apprentissage contenant l'audio et la transcription de l'unité produite) de grande taille.

Le *modèle de langage*, correspondant à $P(W)$ (soit la probabilité d'un mot) est un modèle probabiliste qui permet d'évaluer les contraintes linguistiques pour choisir la séquence de mots les plus probables dans une langue donnée. Pour reconnaître un mot k , il tient compte d'une part de la fréquence d'un mot dans la langue, et d'autre part il considère également les k mots qui le précèdent (le plus souvent, les deux précédents). Par exemple, la probabilité d'obtenir « *prescrit* » après « *le médecin* » est plus élevée que celle d'obtenir « *proscrit* » (et permet de sortir « *prescrit* » plutôt que « *proscrit* »). Le modèle de langage est élaboré à partir de corpus écrits, l'audio n'étant pas nécessaire à son élaboration.

Les premiers systèmes issus de l'apprentissage profond (ou *deep learning*) à avoir été développés étaient des modèles de type DNN-HMM (Deep Neural Network – Hidden Markov Model), dont un exemple d'architecture est fourni par Kipyatkova (Kipyatkova & Karpov, 2016) (voir Schéma 3). Ces modèles utilisent des réseaux de neurones (DNN) à la place des GMM pour aider le HMM au calcul de la probabilité des phonèmes. Ils aboutissent à des résultats plus fiables mais ont besoin de davantage de données pour être correctement entraînés à la reconnaissance de la parole.

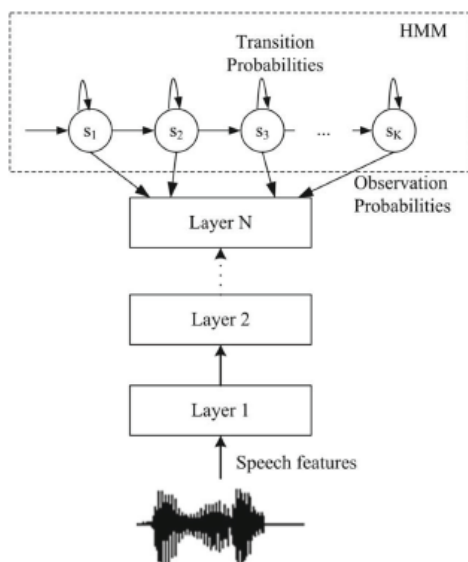


Schéma 3 - Exemple d'architecture d'un DNN-HMM par Yu (Yu & Deng, 2015), cité par Kipyatkova (Kipyatkova & Karpov, 2016)

Les modèles de type TDNN-HMM (Time-Delay Neural Network) (Waibel et al., 1989) sont une évolution des modèles DNN-HMM et ont permis d'optimiser la reconnaissance automatique de la parole. Ces modèles prennent successivement en compte plusieurs fenêtres temporelles, et surtout chaque couche du réseau de neurones considère un nombre plus ou moins grand de fenêtres temporelles (voir Schéma 4). L'intérêt des TDNN-HMM est donc de pouvoir modéliser des éléments différents dans le signal de parole, de bas niveau sur des couches considérant peu de fenêtres temporelles ou de plus haut niveau lors que le nombre de fenêtres considérées augmente.

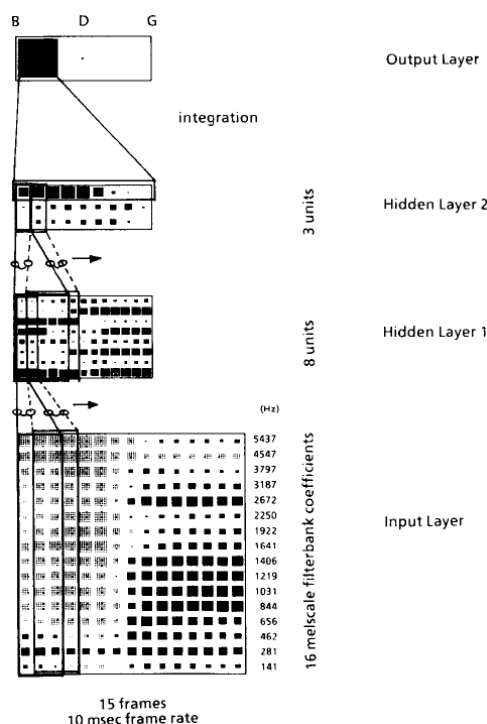


Schéma 4 - Architecture d'un TDNN selon Waibel (Waibel et al., 1989)

Enfin, les TDNNf-HMM (pour TDDN « factorisés ») sont une évolution des TDNN présentés précédemment (Villalba et al., 2020). La factorisation permet de concentrer l'information traitée au niveau des couches (où une couche équivaut à deux couches d'un TDNN classique). Elle améliore ainsi la vitesse de d'entraînement de ces modèles et l'efficacité du décodage en condensant l'information.

Dans ces systèmes de reconnaissance automatique de la parole, le processus de décodage n'est pas séquentiel : les modèles acoustique et de langage et le dictionnaire de prononciation s'influencent mutuellement. Les systèmes les plus récents sont d'ailleurs maintenant entièrement intégrés avec des algorithmes qui gèrent simultanément ces modules, c'est le cas notamment des systèmes end-to-end.

Annexe VI Motifs d'exclusion et conservation des 149 paramètres automatiques

Code	Pertinence				Non redondance	Conservé
	Non normalisés	Variabilité liée au genre du locuteur	Effet absent	Interprétation complexe	Redondant	
<i>cpps</i>						1
<i>slope</i>						1
<i>tilt</i>						1
<i>intrange</i>						1
<i>intsd</i>						1
<i>f0range</i>		1	X	X	X	X
<i>f0sd</i>						1
<i>f0min</i>		1	X	X	X	X
<i>f0max</i>		1	X	X	X	X
<i>f0cent</i>						1
<i>duration</i>					1	X
<i>spkprop</i>					1	X
<i>nsylp</i>						1
<i>npausep</i>						1
<i>nsegw</i>	1	X	X	X	X	X
<i>nsegsecw</i>						1
<i>ndurp</i>						1
<i>ndurw</i>						1
<i>ndursegw</i>						1
<i>phontimep</i>					1	X
<i>phontimew</i>					1	X
<i>silp</i>				1	X	X
<i>spkdurp</i>					1	X
<i>spkdurw</i>						1
<i>spratep</i>						1
<i>artratep</i>						1
<i>asdp</i>					1	X
<i>c_cvps</i>	1	X	X	X	X	X
<i>p_cvps</i>	1	X	X	X	X	X
<i>s_cvps</i>	1	X	X	X	X	X
<i>v_cvps</i>	1	X	X	X	X	X
<i>tot_cvps</i>	1	X	X	X	X	X
<i>cs_cvps</i>				1	X	X
<i>ps_cvps</i>				1	X	X
<i>ss_cvps</i>					1	X
<i>vs_cvps</i>						1
<i>tots_cvps</i>				1	X	X

Code	Pertinence				Non redondance	Conservé
	Non normalisés	Variabilité liée au genre du locuteur	Effet absent	Interprétation complexe	Redondant	
<i>propc_cvps</i>				1	X	X
<i>propp_cvps</i>				1	X	X
<i>props_cvps</i>						1
<i>propv_cvps</i>						1
<i>sumdurc</i>				1	X	X
<i>meandurc</i>				1	X	X
<i>sddurc</i>				1	X	X
<i>sumdurp</i>				1	X	X
<i>meandurp</i>				1	X	X
<i>sddurp</i>				1	X	X
<i>sumdurs</i>				1	X	X
<i>meandurs</i>				1	X	X
<i>sddurs</i>						1
<i>sumdurv</i>					1	X
<i>meandurv</i>					1	X
<i>sddurv</i>						1
<i>sumphon</i>	1	X	X	X	X	X
<i>sumphons</i>						1
<i>difphon</i>						1
<i>csn</i>	1	X	X	X	X	X
<i>csns</i>					1	X
<i>voy</i>	1	X	X	X	X	X
<i>voys</i>					1	X
<i>semics</i>	1	X	X	X	X	X
<i>semicsns</i>						1
<i>txvoy</i>						1
<i>txnasal</i>						1
<i>txvnasal</i>						1
<i>occ</i>	1	X	X	X	X	X
<i>occs</i>						1
<i>fri</i>	1	X	X	X	X	X
<i>fris</i>						1
<i>tx_occcsn</i>						1
<i>tx_fricsn</i>						1
<i>sonant</i>	1	X	X	X	X	X
<i>sonants</i>						1
<i>nonsonant</i>	1	X	X	X	X	X
<i>nonsonants</i>					1	X
<i>tx_sonant</i>						1

Code	Pertinence				Non redondance	Conservé
	Non normalisés	Variabilité liée au genre du locuteur	Effet absent	Interprétation complexe	Redondant	
<i>tx_nsonant</i>					1	X
<i>conf</i>						1
<i>confc</i>						1
<i>confv</i>						1
<i>confs</i>						1
<i>confo</i>						1
<i>conff</i>					1	X
<i>wordwcc</i>	1	X	X	X	X	X
<i>wordwccs</i>					1	X
<i>propwordwcc</i>						1
<i>nword</i>	1	X	X	X	X	X
<i>nwords</i>						1
<i>nworddiff</i>	1	X	X	X	X	X
<i>nworddiffs</i>						1
<i>nwordpond</i>	1	X	X	X	X	X
<i>nwordponds</i>					1	X
<i>nwordconf</i>	1	X	X	X	X	X
<i>nwordconfs</i>					1	X
<i>propnwordconf</i>					1	X
<i>meanconf</i>						1
<i>sdconf</i>						1
<i>meanwordlen</i>						1
<i>sdwordlen</i>						1
<i>mis_values</i>						1
<i>meannbph</i>						1
<i>meannbsyl</i>					1	X
<i>meanpld</i>						1
<i>meanhomoph</i>						1
<i>meanfreqlem</i>						1
<i>meanfreq</i>						1
<i>guiraud</i>						1
<i>totalcgram</i>				1	X	X
<i>cgramwords</i>						1
<i>nom</i>	1	X	X	X	X	X
<i>noms</i>					1	X
<i>adv</i>	1	X	X	X	X	X
<i>advs</i>					1	X
<i>aux</i>	1	X	X	X	X	X
<i>auxs</i>						1

Code	Pertinence				Non redondance	Conservé
	Non normalisés	Variabilité liée au genre du locuteur	Effet absent	Interprétation complexe	Redondant	
<i>conj</i>	1	X	X	X	X	X
<i>conj_s</i>						1
<i>totalia</i>			1	X	X	X
<i>ono</i>	1	X	X	X	X	X
<i>onos</i>						1
<i>prep</i>	1	X	X	X	X	X
<i>preps</i>						1
<i>verb</i>	1	X	X	X	X	X
<i>verbs</i>						1
<i>adj</i>	1	X	X	X	X	X
<i>adjs</i>						1
<i>art</i>	1	X	X	X	X	X
<i>arts</i>						1
<i>pron</i>	1	X	X	X	X	X
<i>prons</i>						1
<i>propnom</i>						1
<i>propadv</i>						1
<i>propaux</i>						1
<i>propcon</i>						1
<i>proplia</i>			1	X	X	X
<i>propono</i>						1
<i>proppre</i>						1
<i>propver</i>						1
<i>propadj</i>						1
<i>propart</i>						1
<i>proppro</i>						1
<i>lex_dens</i>						1
<i>ira_formact</i>	1	X	X	X	X	X
<i>ira_acts</i>					1	X
<i>ira_formsup</i>	1	X	X	X	X	X
<i>ira_sups</i>						1
<i>propactsup</i>					1	X
<i>ira_nbclass</i>						1
<i>sentiment</i>						1
TOTAL	31	3	2	15	23	75

Annexe VII Liste des 75 paramètres issus des analyses automatiques de la parole conservés pour les analyses finales

Niveau	Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
Signal de parole	intrange	Intensité – écart-type	27,27	3,77	26,91	6,56	20,80	34,68
	intsd	Intensité – étendue	3,17	0,24	3,18	0,29	2,71	3,73
	f0sd	F0 – écart-type	29,72	9,53	28,76	15,29	16,72	47,72
	f0cent	F0 cent	2 772,92	426,07	2 955,56	476,00	1 918,15	3 275,13
	cpps	Proéminence du pic cepstral lissé	10,25	1,90	10,24	1,83	6,00	13,52
	slope	Pente spectrale	-30,44	3,63	-31,12	2,85	-34,90	-18,18
	tilt	Inclinaison de la courbe à travers le spectre	-11,02	1,00	-11,06	1,18	-13,51	-8,80
Niveau articulatoire – phonétique articulatoire	phontimew	Durée totale de parole de chaque sujet	521,07	194,54	526,11	263,67	125,34	918,18
	ndurw	Durée globale de production du sujet entre le début de la première prise de parole et la fin de la dernière prise de parole, incluant les pauses	937,36	244,25	930,84	442,26	517,11	1 421,97
	spkdurw	Proportion de durée parlée sur la durée globale de l'échange	0,54	0,12	0,56	0,13	0,23	0,75
	nsegsecw	Proportion de durée parlée sur la durée globale de l'échange, par seconde	0,26	0,04	0,26	0,05	0,19	0,37
	ndursegw	Durée moyenne des segments de parole	2,13	0,53	2,10	0,78	1,06	3,00
	nsylp	Nombre de syllabes / nombre de noyaux syllabiques	1 231,40	506,92	1 280,00	645,00	135,00	2 332,00
	npausep	Nombre de pauses	283,20	87,82	284,00	114,00	124,00	460,00
	spratep	Débit de parole (syll/sec)	1,29	0,40	1,26	0,55	0,25	2,01
	artratep	Débit d'articulation (syll/sec)	3,30	0,46	3,33	0,45	2,52	4,48
	vs_cvps	Nombre de pseudo-voyelles, par seconde	1,75	0,35	1,70	0,59	1,12	2,41
	props_cvps	Proportion de pseudo-silences parmi les pseudo-phonèmes	0,01	0,01	0,01	0,01	0,01	0,04
	propv_cvps	Proportion de pseudo-voyelles parmi les pseudo-phonèmes	0,14	0,02	0,15	0,02	0,10	0,19
	sddurs	Écart-type de la durée des pseudo-silences	0,05	0,02	0,05	0,02	0,03	0,10
	sddurv	Écart-type de la durée des pseudo-voyelles	0,24	0,06	0,23	0,05	0,15	0,40

Niveau	Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
Niveau articulatoire – composante phonémique	sumphons	Nombre de phonèmes reconnus, par seconde	4,55	1,56	4,78	2,40	1,12	7,49
	difphon	Nombre de phonèmes différents reconnus	29,20	5,57	32,00	3,00	5,00	32,00
	semicsns	Nombre de semi-consonnes reconnues, par seconde	0,11	0,08	0,11	0,11	0,00	0,36
	txvoy	Taux de voyelles reconnues	0,51	0,09	0,49	0,04	0,43	0,85
	txnasal	Taux de phonèmes nasaux reconnus	0,19	0,06	0,19	0,06	0,06	0,37
	txvnasal	Taux de voyelles nasales reconnues par les voyelles	0,18	0,10	0,17	0,09	0,05	0,44
	occs	Nombre d'occlusives reconnues, par seconde	0,58	0,41	0,56	0,62	0,00	1,51
	fris	Nombre de fricatives reconnues, par seconde	0,80	0,29	0,85	0,31	0,00	1,15
	tx_occsn	Taux d'occlusives reconnues parmi les consonnes	0,23	0,12	0,27	0,20	0,00	0,37
	tx_fricsn	Taux de fricatives reconnues parmi les consonnes	0,36	0,14	0,34	0,16	0,00	0,78
	sonants	Nombre de sonantes reconnues, par seconde	0,96	0,38	1,00	0,48	0,17	1,62
	tx_sonant	Taux de sonantes reconnues parmi les consonnes	0,46	0,14	0,42	0,11	0,23	1,00
	conf	Score moyen global de confiance	0,84	0,02	0,84	0,03	0,78	0,89
	confc	Score moyen global de confiance des consonnes	0,87	0,04	0,88	0,03	0,76	0,93
	confv	Score moyen global de confiance des voyelles	0,80	0,03	0,80	0,02	0,77	0,91
	confs	Score moyen global de confiance des semi-consonnes	0,76	0,04	0,76	0,04	0,65	0,84
	confo	Score moyen global de confiance des occlusives	0,87	0,07	0,90	0,09	0,72	0,95

Niveau	Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
Niveau lexical (composante morpho-syntaxique)	nwords	Nombre de mots reconnus, par seconde	1,84	0,51	1,79	0,48	1,02	3,33
	nworddiffs	Nombre de mots différents reconnus, par seconde	0,77	0,23	0,78	0,26	0,20	1,27
	meanwordlen	Nombre de caractère moyen par mot	4,18	0,46	4,26	0,19	2,39	4,82
	sdwordlen	Écart-type du nombre de caractère moyen par mot	2,24	0,32	2,31	0,43	1,23	2,64
	mis_values	Nombre de mots reconnus et non référencés dans lexique.org	30,96	18,14	28,00	27,00	0,00	78,00
	meannbph	Nombre moyen de phonèmes par mot	2,77	0,33	2,83	0,25	1,42	3,22
	meanpld	Distance de Levenshtein phonologique moyenne par mot	1,08	0,07	1,09	0,07	0,84	1,19
	meanhomoph	Nombre d'homophones moyen par mot	5,30	0,57	5,13	0,30	4,82	7,90
	meanfreqlem	Fréquence moyenne du lemme, par million	3 577,46	581,90	3 484,37	727,89	2 470,14	4 850,39
	meanfreq	Fréquence moyenne du mot, par million	2 539,02	543,63	2 468,11	521,74	1 561,02	4 504,54
	propwordwcc	Proportion de mots contenant au moins une structure diconsonantique	0,06	0,03	0,06	0,03	0,01	0,13
	meanconf	Score de confiance moyen	0,67	0,04	0,66	0,04	0,61	0,78
	sdconf	Écart-type du score de confiance moyen	0,25	0,01	0,25	0,01	0,22	0,26
	guiraud	Index de Guiraud	12,43	3,35	13,26	2,68	2,12	17,02
	lex_dens	Densité lexicale	0,60	0,06	0,59	0,07	0,45	0,71

Niveau	Code	Paramètre	Moyenne	Écart-type	Médiane	EIQ	Valeur min	Valeur max
Niveau grammatical (composante morpho-syntaxique)	cgramwords	Rapport classes grammaticales reconnus par nombre total de mots	1,85	0,15	1,86	0,16	1,41	2,24
	propnom	Proportion de noms	0,27	0,04	0,28	0,03	0,12	0,31
	propadv	Proportion d'adverbes	0,08	0,02	0,09	0,04	0,04	0,13
	auxs	Nombre d'auxiliaires, par seconde	0,13	0,06	0,13	0,08	0,03	0,25
	propaux	Proportion d'auxiliaires	0,04	0,01	0,04	0,01	0,02	0,09
	conj	Nombre de conjonctions, par seconde	0,08	0,04	0,08	0,04	0,01	0,16
	propcon	Proportion de conjonctions	0,02	0,01	0,02	0,01	0,01	0,04
	onos	Nombre d'onomatopées, par seconde	0,03	0,10	0,01	0,01	0,00	0,50
	propono	Proportion d'onomatopées	0,02	0,07	0,00	0,00	0,00	0,35
	preps	Nombre de prépositions, par seconde	0,17	0,09	0,14	0,08	0,03	0,47
	proppre	Proportion de prépositions	0,05	0,03	0,05	0,01	0,01	0,16
	verbs	Nombre de verbes, par seconde	0,47	0,15	0,46	0,19	0,07	0,83
	propver	Proportion de verbes	0,14	0,03	0,14	0,03	0,05	0,19
	adjs	Nombre d'adjectifs, par seconde	0,52	0,17	0,52	0,24	0,11	0,96
	propadj	Proportion d'adjectifs	0,15	0,02	0,15	0,03	0,08	0,19
	arts	Nombre d'articles, par seconde	0,27	0,10	0,28	0,11	0,10	0,56
	propart	Proportion d'articles	0,08	0,02	0,08	0,02	0,04	0,12
	prons	Nombre de pronoms, par seconde	0,51	0,16	0,49	0,15	0,14	0,93
	proppro	Proportion de pronoms	0,15	0,02	0,15	0,02	0,09	0,20
	ira_sup	Nombre de formes supplémentaires, par seconde	0,20	0,07	0,18	0,09	0,09	0,35
Composante pragma-sémantique et discursive	ira_nbclass	Nombre de classes thématiques par sujet	5,56	1,20	6,00	1,00	3,00	8,00
	sentiment	Score de sentiment	0,50	0,02	0,49	0,03	0,48	0,58

Annexe VIII Motifs d'exclusion et conservation des 174 items issus des autoquestionnaires

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
ecvb1	Etes-vous gêné(e) pour exprimer des choses très simples, par exemple dire que vous désirez boire, manger, aller vous reposer... ? (expression d'un besoin)					1
ecvb2	Etes-vous gêné(e) pour exprimer vos désirs, vos intentions, par exemple faire comprendre que vous voulez aller vous promener (ou) voir une personne en particulier (ou) aller chez le coiffeur ... ? (intention, projet)				1	X
ecvb3	Quand vous allez vous promener, quand vous sortez seul(e), est-ce que vous êtes gêné(e) pour demander votre chemin ? Si la situation ne s'est pas présentée, demander : êtes -vous gêné(e), par exemple, pour expliquer votre destination au chauffeur d'un taxi ?				1	X
ecvb4	Avec un membre de votre famille ou un de vos amis, avez-vous des difficultés pour avoir une conversation sur un sujet courant de la vie quotidienne ? (ex : le temps qu'il fait, votre plat préféré, ce que vous avez mangé à midi...)					1
ecvb4	Avec un membre de votre famille ou un de vos amis, avez-vous des difficultés pour avoir une conversation sur un sujet courant de la vie quotidienne ? (ex : le temps qu'il fait, votre plat préféré, ce que vous avez mangé à midi...)					1
ecvb6	En famille ou avec des amis, est-ce vous qui entamez la conversation ?					1
ecvb7	Avez-vous des difficultés pour exprimer par la parole des sentiments (par exemple la joie, la colère ou la peur) ?					1
ecvb8	Et avec quelqu'un que vous connaissez peu (le facteur ou un chauffeur de taxi par exemple), êtes vous gêné(e) pour avoir une conversation sur des sujets simples ? (le temps ; ce que vous avez fait la veille; les fleurs de votre jardin ...) ?				1	X
ecvb9	Et des sujets plus compliqués, par exemple : expliquer votre situation ou votre maladie à quelqu'un qui ne vous connaît pas, ça vous pose un problème ?					1
ecvb10	Avez-vous des difficultés à prendre la parole lorsque vous êtes avec des gens que vous connaissez peu (lors d'un dîner, d'une sortie, d'une soirée...) ?			1	X	X
ecvb11	Avez-vous des difficultés pour téléphoner à votre famille ?			1	X	X
ecvb12	Avez-vous des difficultés pour téléphoner à vos amis ?					1

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
ecvb13	Quand vous voulez prendre un rendez-vous avec quelqu'un, par exemple votre médecin ou votre orthophoniste, est-ce vous qui téléphonez ?					1
ecvb14	Êtes-vous gêné pour téléphoner à une personne inconnue ? Par exemple appeler un taxi ou appeler le plombier en cas de fuite d'eau ?				1	X
ecvb15	Quand le téléphone sonne et que vous êtes seul, est-ce que vous répondez ?				1	X
ecvb16	Et s'il y a quelqu'un d'autre chez vous (conjoint, enfant, ami...), vous répondez au téléphone ?					1
ecvb17	Avez-vous des difficultés à transmettre à un proche (conjoint, enfant, voisin...) un message téléphonique reçu en son absence ?					1
ecvb18	Depuis votre maladie (ou) notre dernier entretien (ou) le début de la rééducation (adapter selon le cas), êtes-vous allé(e) faire des achats seul(e) ? (au moins un: pain, loto, ticket de bus, journal, café ou autre consommation...).		1	X	X	X
ecvb19	Dans un magasin (en dehors d'une grande surface) est-ce que vous demandez ce que vous désirez au vendeur, si vous ne le trouvez pas ?					1
ecvb20	Quand vous payez, avez-vous des difficultés pour utiliser de l'argent liquide, par exemple : donner la somme exacte ou compter votre monnaie ?		1	X	X	X
ecvb21	Pour payer, utilisez-vous des chèques (ou une carte bancaire) ?	1	X	X	X	X
ecvb22	Lors d'un repas de famille, ou entre amis, avez-vous des difficultés à suivre une conversation rapide, entre plusieurs personnes ?	1	X	X	X	X
ecvb23	Lorsque vous avez besoin d'un renseignement, est-ce que vous adressez la parole à des inconnus ? (ex : vous allez à l'hôpital voir un ami qui s'est cassé une jambe. Vous ne connaissez pas le numéro de sa chambre. Est-ce que vous demanderiez à l'accueil ?).				1	X
ecvb24	Est-ce que vous sortez (chez des amis, au cinéma, au restaurant, au théâtre...) ?		1	X	X	X
ecvb25	Au restaurant / café, avez-vous des difficultés pour passer votre commande vous-même ?					1
ecvb26	Quand vous allez chez un commerçant, par exemple chez l'épicier, le garagiste, le coiffeur, le libraire... est-ce vous qui expliquez ce que vous désirez ?					1
ecvb27	Lisez-vous des journaux, des magazines (éventuellement des livres) ?	1	X	X	X	X
X	Lisez-vous le courrier que vous écrivent vos amis ou votre famille ?	1	X	X	X	X
ecvb29	Avez-vous des difficultés pour lire des papiers administratifs ou une notice d'emploi ?	1	X	X	X	X

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
ecvb30	Avez-vous des difficultés pour lire l'heure ?	1	X	X	X	X
ecvb31	Avez-vous des difficultés pour écrire des mots isolés (liste de choses à faire p. ex.) ?	1	X	X	X	X
ecvb32	Avez-vous des difficultés pour écrire des phrases (p. ex. lettre ou carte pendant vacances, vous, anniversaire...) ?	1	X	X	X	X
ecvb33	Remplissez-vous des feuilles de sécurité sociale ou d'autres papiers administratifs ?	1	X	X	X	X
ecvb34	Est-ce que vous remplissez des chèques ?	1	X	X	X	X
dipa1	Mon problème de parole a eu un effet négatif sur la façon dont je me perçois	1	X	X	X	X
dipa2	Quand je parle, j'ai l'impression d'entendre quelqu'un d'autre que moi	1	X	X	X	X
dipa3	Même si je ne parle pas, j'ai l'impression d'être une personne différente d'avant	1	X	X	X	X
dipa4	Je ne me sens pas mis(e) à l'écart à cause de ma parole		1	X	X	X
dipa5	Je suis autant sûr(e) de moi que lorsque je n'avais pas de problème de parole		1	X	X	X
dipa6	En raison de ma parole, je suis maintenant plus dépendant(e) des autres que je ne l'étais auparavant		1	X	X	X
dipa7	Je ne me sens pas embarrassé(e) par ma parole		1	X	X	X
dipa8	Je ne me sens pas incompetent(e) à cause de ma parole		1	X	X	X
dipa9	Je ne me sens pas idiot(e) lorsque je suis incompris(e)		1	X	X	X
dipa10	Je me sens stupide lorsqu'on me demande de répéter		1	X	X	X
dipa11	Maintenant, j'ai l'impression d'avoir moins de contrôle sur ma vie à cause de ma parole		1	X	X	X
dipa12	Mon problème de parole ne m'a pas fondamentalement changé(e)	1	X	X	X	X
dipb1	Je n'essaye pas de cacher mes difficultés de parole		1	X	X	X
dipb2	Je me fâche quand les gens ne me comprennent pas		1	X	X	X
dipb3	Je ne suis pas satisfait(e) de ma parole en l'état actuel	1	X	X	X	X
dipb4	Je suis susceptible à l'égard de tout ce qui concerne ma parole		1	X	X	X
dipb5	Je ne me soucie pas de ma parole		1	X	X	X
dipb6	Cela ne me dérange pas d'admettre que j'ai un problème de parole		1	X	X	X
dipb7	Je préfère perdre le fil des conversations plutôt que d'admettre avoir un problème de parole		1	X	X	X
dipb8	Mon problème de parole a influencé ma vie plus que tout autre chose		1	X	X	X

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
dipb9	Je ne me mets pas en colère lorsque je ne peux pas me faire comprendre		1	X	X	X
dipb10	J'ai beaucoup d'autres caractéristiques importantes qui me définissent autres que ma parole	1	X	X	X	X
dipc1	Je suis conscient de la réaction des autres vis-à-vis de ma parole		1	X	X	X
dipc2	Ma famille n'est jamais ennuyée quand elle ne comprend pas ce que je dis		1	X	X	X
dipc3	Je n'accorde pas d'importance à ce que les gens pensent de ma parole		1	X	X	X
dipc4	Les autres me traitent comme si j'étais stupide parce qu'ils ne me comprennent pas		1	X	X	X
dipc5	Lorsque je n'arrive pas à me faire comprendre, les gens ne sont pas gênés		1	X	X	X
dipc6	Les personnes qui ne me connaissent pas ne me considèrent pas négativement malgré mon problème de parole		1	X	X	X
dipc7	Les gens sont généralement patients lorsque je parle lentement		1	X	X	X
dipc8	Mes amis font plus d'effort pour me comprendre que ma famille		1	X	X	X
dipc9	Les personnes font semblant de me comprendre alors que je sais que ce n'est pas le cas		1	X	X	X
dipc10	Les gens pensent parfois que j'ai bu de l'alcool alors que ce n'est pas vrai		1	X	X	X
dipc11	La plupart des gens fait un effort pour comprendre ce que je dis		1	X	X	X
dipc12	J'accorde de l'importance à l'opinion des autres à l'égard de ma parole		1	X	X	X
dipc13	Je n'ai jamais le sentiment que les autres rient ou se moquent de mes difficultés de parole		1	X	X	X
dipc14	Les personnes sont plus condescendantes du fait de ma parole		1	X	X	X
dipc15	Les inconnus n'ont pas d'a priori négatif sur moi malgré mon problème de parole		1	X	X	X
dipd1	Malgré mes difficultés à parler, je n'évite pas de communiquer avec les gens que je connais				1	X
dipd2	Ma vie sociale n'a pas changé à la suite à ma dysarthrie					1
dipd3	Lorsque les gens ne me comprennent pas, j'essaye de faire passer mon message par d'autres moyens					1
dipd4	J'évite d'utiliser le téléphone à cause de mes difficultés de parole				1	X
dipd5	J'évite de demander des articles dans des magasins					1
dipd6	Chaque fois que c'est possible, je compte sur les autres pour parler à ma place					1
dipd7	En raison de ma parole, j'écoute plus que je ne prends part aux conversations					1

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
dipd8	Je n'évite pas de parler aux personnes que je ne connais pas				1	X
dipd9	Je me sens à l'aise pour m'exprimer dans la plupart des situations, tant à la maison qu'à l'extérieur					1
dipd10	Les difficultés que j'éprouve avec ma parole limitent ma vie sociale				1	X
dipd11	J'évite de parler seulement lorsque je suis fatigué(e)		1	X	X	X
dipd12	A cause de ma parole, je me suis isolé(e) socialement		1	X	X	X
phif1	La vitesse à laquelle je parle a changé	1	X	X	X	X
phif2	Ma voix rend difficile l'expression de mes émotions				1	X
phif3	J'ai des difficultés à articuler quand je parle		1	X	X	X
phif4	Je dois faire un effort pour parler					1
phif5	Je suis à court de souffle quand je parle	1	X	X	X	X
phic1	J'ai des difficultés à formuler oralement ce dont j'ai besoin (boire, manger, toilette...)					1
phic2	Je suis gêné pour exprimer mes pensées, mes opinions				1	X
phic3	J'ai du mal à communiquer avec des personnes peu familières					1
phic4	On me demande de répéter à cause de mon problème pour parler					1
phic5	J'évite les conversations avec ma famille, les amis, les voisins				1	X
phie1	Je souffre de ma façon de parler		1	X	X	X
phie2	Mes difficultés de parole limitent ma vie personnelle et sociale				1	X
phie3	Je trouve que les autres ne comprennent pas mon problème pour parler		1	X	X	X
phie4	Les gens semblent irrités à cause de mon problème pour parler		1	X	X	X
phie5	Je me sens handicapé à cause de mes difficultés pour parler		1	X	X	X
phisev	Quel degré de sévérité donnez-vous à vos difficultés de parole ?	1	X	X	X	X
phigene	À quel point êtes-vous gêné pour produire une parole compréhensible ?		1	X	X	X
phihcp	À quel point votre problème pour parler vous handicape-t-il dans votre vie quotidienne ?		1	X	X	X
eortc1	Avez-vous des difficultés à faire certains efforts physiques pénibles (sac de courses, valise) ?	1	X	X	X	X
eortc2	Avez-vous des difficultés à faire une longue promenade ?	1	X	X	X	X
eortc3	Avez-vous des difficultés à faire un petit tour dehors ?	1	X	X	X	X

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
eortc4	Êtes-vous obligé de rester au lit ou dans un fauteuil pendant la journée ?	1	X	X	X	X
eortc5	Avez-vous besoin d'aide pour manger, vous habiller, faire votre toilette ou aller aux toilettes ?	1	X	X	X	X
eortc6	Avez-vous été gêné(e) pour faire votre travail ou vos activités de tous les jours ?	1	X	X	X	X
eortc7	Avez-vous été gêné(e) dans vos activités de loisirs ?	1	X	X	X	X
eortc8	Avez-vous eu le souffle court ?	1	X	X	X	X
eortc9	Avez-vous ressenti de la douleur ?	1	X	X	X	X
eortc10	Avez-vous eu besoin de repos ?	1	X	X	X	X
eortc11	Avez-vous des difficultés à dormir ?	1	X	X	X	X
eortc12	Vous êtes-vous senti(e) faible ?	1	X	X	X	X
eortc13	Avez-vous manqué d'appétit ?	1	X	X	X	X
eortc14	Avez-vous eu des nausées (mal au cœur) ?	1	X	X	X	X
eortc15	Avez-vous vomi ?	1	X	X	X	X
eortc16	Avez-vous été constipé(e) ?	1	X	X	X	X
eortc17	Avez-vous eu de la diarrhée ?	1	X	X	X	X
eortc18	Étiez-vous fatigué ?	1	X	X	X	X
eortc19	Des douleurs ont-elles perturbé vos activités quotidiennes ?	1	X	X	X	X
eortc20	Avez-vous eu des difficultés à vous concentrer sur certaines choses (lire journal, TV) ?	1	X	X	X	X
eortc21	Vous êtes-vous senti tendu ?	1	X	X	X	X
eortc22	Vous êtes-vous fait du souci ?	1	X	X	X	X
eortc23	Vous êtes-vous senti irritable ?	1	X	X	X	X
eortc24	Vous êtes-vous senti déprimé ?	1	X	X	X	X
eortc25	Avez-vous eu des difficultés à vous souvenir de certaines choses ?	1	X	X	X	X
eortc26	Votre état physique ou votre traitement médical vous ont-ils gêné dans votre vie familiale ?	1	X	X	X	X
eortc27	Votre état physique ou votre traitement médical vous ont-ils gêné dans vos activités sociales ? (sortie avec amis, cinéma) ?	1	X	X	X	X
eortc28	Votre état physique ou votre traitement médical vous ont-ils causé des problèmes financiers ?	1	X	X	X	X
eortc29	Comment évalueriez-vous votre état de santé au cours de la semaine passée ?	1	X	X	X	X

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
eortc30	Comment évalueriez-vous l'ensemble de votre qualité de vie au cours de la semaine passée ?	1	X	X	X	X
hn31	Avez-vous eu mal dans la bouche ?	1	X	X	X	X
hn32	Avez-vous eu mal à la mâchoire ?	1	X	X	X	X
hn33	Avez-vous eu des douleurs dans la bouche ?	1	X	X	X	X
hn34	Avez-vous eu mal à la gorge ?	1	X	X	X	X
hn35	Avez-vous eu des problèmes en avalant les liquides ?	1	X	X	X	X
hn36	Avez-vous eu des problèmes en avalant des aliments écrasés ?	1	X	X	X	X
hn37	Avez-vous eu des problèmes en avalant des aliments solides ?	1	X	X	X	X
hn38	Avez-vous avalé de travers ?	1	X	X	X	X
hn39	Avez-vous eu des problèmes de dents ?	1	X	X	X	X
hn40	Avez-vous eu des problèmes à ouvrir largement la bouche ?	1	X	X	X	X
hn41	Avez-vous eu la bouche sèche ?	1	X	X	X	X
hn42	Avez-vous eu une salive collante ?	1	X	X	X	X
hn43	Avez-vous eu des problèmes d'odorat ?	1	X	X	X	X
hn44	Avez-vous eu des problèmes de goût ?	1	X	X	X	X
hn45	Avez-vous toussé ?	1	X	X	X	X
hn46	Avez-vous été enrôlé ?	1	X	X	X	X
hn47	Vous êtes-vous senti(e) mal ?	1	X	X	X	X
hn48	Votre apparence vous a-t-elle préoccupé ?	1	X	X	X	X
hn49	Avez-vous eu des difficultés à manger ?	1	X	X	X	X
hn50	Avez-vous eu des difficultés à manger devant votre famille ?	1	X	X	X	X
hn51	Avez-vous eu des difficultés à manger devant d'autres personnes ?	1	X	X	X	X
hn52	Avez-vous eu des difficultés à prendre plaisir aux repas ?	1	X	X	X	X
hn53	Avez-vous des difficultés à parler à d'autres personnes ?					1
hn54	Avez-vous des difficultés à parler au téléphone ?				1	X
hn55	Avez-vous eu des difficultés à avoir un contact social avec votre famille ?		1	X	X	X
hn56	Avez-vous eu des difficultés à avoir un contact social avec vos amis ?		1	X	X	X

code	item	Non conforme à la définition par le comité d'experts	Non retenu par les experts	Redondant / insuffisamment variable	Coefficient de régression nul	Conservé pour le SHC
hn57	Avez-vous des difficultés à sortir en public ?		1	X	X	X
hn58	Avez-vous des difficultés à avoir un contact physique avec votre famille ou vos amis ?	1	X	X	X	X
hn59	Avez-vous éprouvé moins d'intérêt aux relations sexuelles ?	1	X	X	X	X
hn60	Avez-vous éprouvé moins de désir sexuel ?	1	X	X	X	X
hn61	Avez-vous pris des anti-douleurs ?	1	X	X	X	X
hn62	Avez-vous pris des suppléments nutritionnels (à l'exception de vitamines) ?	1	X	X	X	X
hn63	Avez-vous utilisé une sonde d'alimentation ?	1	X	X	X	X
hn64	Avez-vous perdu du poids ?	1	X	X	X	X
hn65	Avez-vous pris du poids ?	1	X	X	X	X
chiph1	Avez-vous des difficultés pour parler ?				1	X
chiph2	Votre entourage a-t-il des difficultés pour vous comprendre ?			1	X	X
chiph3	Parlez-vous moins souvent avec votre famille, vos amis, vos voisins ?			1	X	X
chiph4	Avez-vous des difficultés pour articuler ?		1	X	X	X
chips1	Votre vie personnelle et sociale est-elle limitée par les problèmes liés à votre maladie ?	1	X	X	X	X
chips2	Votre maladie a-t-elle modifiée vos rapports avec autrui ?	1	X	X	X	X
chips3	Êtes-vous contrarié par les problèmes liés à votre maladie ?	1	X	X	X	X
chips4	Vous sentez vous handicapé à cause des suites de votre maladie ?	1	X	X	X	X
	TOTAL	83	47	4	16	24

Annexe IX Liste des 40 items retenus pour la construction du score holistique de communication (les codes item en gras indiquent les 24 items finaux retenus)

code	item	n	Max	Min	Moy	É-T	Q1	Med	Q3	EIQ	Comparabilité des lots (valeurs p, test de Mann-Whitney)
ecvb1	Êtes-vous gêné(e) pour exprimer des choses très simples, par exemple dire que vous désirez boire, manger, aller vous reposer... ? (expression d'un besoin)	25	3	0	2,36	0,81	2	3	3	1	0,16
ecvb2	Êtes-vous gêné(e) pour exprimer vos désirs, vos intentions, par exemple faire comprendre que vous voulez aller vous promener (ou) voir une personne en particulier (ou) aller chez le coiffeur ... ? (intention, projet)	25	3	1	2,56	0,58	2	3	3	1	0,75
ecvb3	Quand vous allez vous promener, quand vous sortez seul(e), est-ce que vous êtes gêné(e) pour demander votre chemin ? Si la situation ne s'est pas présentée, demander : êtes -vous gêné(e), par exemple, pour expliquer votre destination au chauffeur d'un taxi ?	25	3	1	2,32	0,63	2	2	3	1	0,73
ecvb4	Avec un membre de votre famille ou un de vos amis, avez-vous des difficultés pour avoir une conversation sur un sujet courant de la vie quotidienne ? (ex : le temps qu'il fait, votre plat préféré, ce que vous avez mangé à midi...)	25	3	0	2,08	0,95	2	2	3	1	0,42
ecvb5	Avec un membre de votre famille/amis, avez-vous des difficultés pour avoir une conversation sur un sujet plus abstrait ou plus compliqué par exemple : l'actualité, l'éducation, le chômage, la littérature... ?	25	3	0	1,8	0,91	1	2	2	1	0,84
ecvb6	En famille ou avec des amis, est-ce vous qui entamez la conversation ?	25	3	0	1,49	0,83	1	2	2	1	0,31
ecvb7	Avez-vous des difficultés pour exprimer par la parole des sentiments (par exemple la joie, la colère ou la peur) ?	25	3	0	2,35	0,85	2	3	3	1	0,31
ecvb8	Et avec quelqu'un que vous connaissez peu (le facteur ou un chauffeur de taxi par exemple), êtes vous gêné(e) pour avoir une conversation sur des sujets simples ? (le temps ; ce que vous avez fait la veille ; les fleurs de votre jardin ...) ?	25	3	0	2,16	0,85	2	2	3	1	0,84
ecvb9	Et des sujets plus compliqués, par exemple : expliquer votre situation ou votre maladie à quelqu'un qui ne vous connaît pas, ça vous pose un problème ?	25	3	0	1,8	0,96	1	2	3	2	0,43
ecvb12	Avez-vous des difficultés pour téléphoner à vos amis ?	25	3	0	1,68	1,22	1	2	3	2	0,52
ecvb13	Quand vous voulez prendre un rendez-vous avec quelqu'un, par exemple votre médecin ou votre orthophoniste, est-ce vous qui téléphonez ?	25	3	0	1,56	1,36	0	2	3	3	0,75

code	item	n	Max	Min	Moy	É-T	Q1	Med	Q3	EIQ	Comparabilité des lots (valeurs p, test de Mann-Whitney)
ecvb14	Êtes-vous gêné pour téléphoner à une personne inconnue ? Par exemple appeler un taxi ou appeler le plombier en cas de fuite d'eau ?	25	3	0	1,6	1,29	0	1	3	3	0,49
ecvb15	Quand le téléphone sonne et que vous êtes seul, est-ce que vous répondez ?	25	3	0	1,92	1,22	1	2	3	2	0,86
ecvb16	Et s'il y a quelqu'un d'autre chez vous (conjoint, enfant, ami...), vous répondez au téléphone ?	25	3	0	1,72	1,14	1	2	3	2	0,91
ecvb17	Avez-vous des difficultés à transmettre à un proche (conjoint, enfant, voisin...) un message téléphonique reçu en son absence ?	25	3	0	2,33	0,89	2	3	3	1	0,98
ecvb19	Dans un magasin (en dehors d'une grande surface) est-ce que vous demandez ce que vous désirez au vendeur, si vous ne le trouvez pas ?	25	3	0	1,49	1,05	1	1	2	1	0,43
ecvb23	Lorsque vous avez besoin d'un renseignement, est-ce que vous adressez la parole à des inconnus ? (ex : vous allez à l'hôpital voir un ami qui s'est cassé une jambe. Vous ne connaissez pas le numéro de sa chambre. Est-ce que vous demanderiez à l'accueil ?).	25	3	0	2,16	0,99	1	3	3	2	0,53
ecvb25	Au restaurant / café, avez-vous des difficultés pour passer votre commande vous-même ?	25	3	0	1,82	1,08	1	2	3	2	0,26
ecvb26	Quand vous allez chez un commerçant, par exemple chez l'épicier, le garagiste, le coiffeur, le libraire... est-ce vous qui expliquez ce que vous désirez ?	25	3	0	2,28	0,94	2	3	3	1	0,65
dipd1	Malgré mes difficultés à parler, je n'évite pas de communiquer avec les gens que je connais	24	5	1	2,17	1,01	2	2	2	0	0,81
dipd2	Ma vie sociale n'a pas changé à la suite à ma dysarthrie	24	5	1	2,96	1,56	2	2,5	4,25	2,25	0,72
dipd3	Lorsque les gens ne me comprennent pas, j'essaye de faire passer mon message par d'autres moyens	24	5	1	2,25	0,87	2	2	2	0	0,51
dipd4	J'évite d'utiliser le téléphone à cause de mes difficultés de parole	24	5	1	2,67	1,33	2	2	4	2	0,75
dipd5	J'évite de demander des articles dans des magasins	24	5	1	3,29	1,11	2	4	4	2	0,14
dipd6	Chaque fois que c'est possible, je compte sur les autres pour parler à ma place	24	5	1	3,42	1,26	2	4	4	2	0,53
dipd7	En raison de ma parole, j'écoute plus que je ne prends part aux conversations	24	5	1	2,54	1,25	2	2	4	2	0,18
dipd8	Je n'évite pas de parler aux personnes que je ne connais pas	24	5	1	2,28	0,92	2	2	2,7	0,7	0,70
dipd9	Je me sens à l'aise pour m'exprimer dans la plupart des situations, tant à la maison qu'à l'extérieur	24	5	1	2,92	1,34	2	3	4	2	0,37

code	item	n	Max	Min	Moy	É-T	Q1	Med	Q3	EIQ	Comparabilité des lots (valeurs p, test de Mann-Whitney)
dipd10	Les difficultés que j'éprouve avec ma parole limitent ma vie sociale	24	5	1	2,73	1,41	2	2	4	2	0,74
chiph1	Avez-vous des difficultés pour parler ?	23	4	0	1,96	1,63	0,5	2	4	3,5	0,78
hn53	Avez-vous des difficultés à parler à d'autres personnes ?	25	4	1	2,2	1	1	2	3	2	0,89
hn54	Avez-vous des difficultés à parler au téléphone ?	25	4	1	2,36	1,15	1	2	3	2	0,57
phif2	Ma voix rend difficile l'expression de mes émotions	25	4	0	1,96	1,21	1	2	3	2	0,89
phif4	Je dois faire un effort pour parler	25	4	0	2,32	1,52	2	3	4	2	0,80
phic1	J'ai des difficultés à formuler oralement ce dont j'ai besoin (boire, manger, toilette...)	25	4	0	0,99	1,25	0	0	2	2	0,14
phic2	Je suis gêné pour exprimer mes pensées, mes opinions	25	4	0	1,36	1,25	0	2	2	2	0,95
phic3	J'ai du mal à communiquer avec des personnes peu familières	25	4	0	1,76	1,42	0	2	3	3	0,89
phic4	On me demande de répéter à cause de mon problème pour parler	25	4	0	1,68	1,18	1	2	2	1	0,77
phic5	J'évite les conversations avec ma famille, les amis, les voisins	25	4	0	1,24	1,3	0	1	2	2	0,38
phie2	Mes difficultés de parole limitent ma vie personnelle et sociale	25	4	0	1,88	1,62	0	2	3	3	0,98

Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé

Résumé

Les troubles de parole sont une problématique fréquemment rencontrée après traitement d'un cancer de la cavité buccale ou de l'oropharynx. Mais peu d'études s'intéressent à l'heure actuelle aux conséquences de ce trouble sur les capacités de communication des patients ou leur qualité de vie. Or, en contexte clinique, l'optimisation des capacités de communication est un objectif thérapeutique majeur dans le suivi de ces patients. En pratique courante, l'évaluation des troubles de parole donne des scores prédisant mal l'impact de ces troubles sur la communication. L'analyse automatique de la parole, moins variable que l'évaluation perceptive habituellement utilisée, est un axe en plein développement.

Dans cette thèse, nous avons cherché à mesurer l'altération de la communication au moyen d'analyses automatiques de la parole spontanée.

Nous avons étudié trois aspects : la mesure de l'altération de la communication, l'analyse automatique de la parole spontanée, et la prédiction de l'altération de la communication par les paramètres automatiques. Pour ce faire, nous avons constitué un nouveau corpus de parole auprès de 25 sujets traités pour un cancer de la cavité buccale ou de l'oropharynx. Il comprend une tâche de parole spontanée enregistrée au cours d'un entretien semi-dirigé, mais aussi des autoquestionnaires autorisant la mesure de la communication et des facteurs associés à la parole et à la communication.

Concernant le premier aspect, un score de référence mesurant de façon holistique la communication a été construit. Il permet de combler le manque d'outils disponibles en cancérologie ORL pour cette mesure.

Le deuxième aspect concerne l'analyse automatique de la parole. Une revue systématique de littérature nous a conduits à nous intéresser aux outils applicables à l'analyse de la parole spontanée, qui est le contexte de production le plus proche de la communication quotidienne. Cent quarante-neuf paramètres automatiques issus des différents niveaux du modèle psycholinguistique de communication de Caron ont été extraits. Puis, un processus de sélection a abouti à retenir 75 paramètres pertinents et non redondants.

Enfin, pour le troisième aspect, nous avons mené une modélisation prédictive de l'altération de la communication au moyen des paramètres automatiques retenus (corrélation de 0,83 entre score prédit et score réel). La corrélation atteint même 0,89 en incluant à la modélisation des facteurs associés (constitution des cercles sociaux, état anxio-dépressif, déficits associés, auto-perception du handicap lié au trouble de parole).

L'utilisation de l'analyse automatique de la parole permet donc une prédiction fiable de l'altération de communication ressentie par les patients. Cette étude ouvre de nouvelles perspectives quant à l'utilisation et l'optimisation des systèmes de reconnaissance automatique de parole dans l'évaluation clinique d'une part, et la prise en compte des besoins fonctionnels et psychosociaux exprimés par les patients d'autre part.

Mots-clés : Parole, Analyse automatique, Communication, Cancer

Measurement of communication impairment by automatic analyses of spontaneous speech after treatment of oral or oropharyngeal cancer

Abstract

Speech disorders are a frequent problem after treatment of oral or oropharyngeal cancer. However, few studies focus on the consequences of this disorder on the communication abilities of patients or their quality of life. However, in clinical care, the optimization of communication abilities is a major therapeutic objective in the follow-up. In current practice, the evaluation of speech disorders gives scores that poorly predict the impact of speech disorders on communication. Automatic speech analysis, which is less variable, is a growing research field.

In this thesis, we aimed to measure communication impairment using automatic analyses of spontaneous speech.

We studied three aspects: the measurement of communication impairment, the automatic analysis of spontaneous speech, and the prediction of communication impairment by automatic parameters. We built a new speech corpus from 25 subjects treated for oral or oropharyngeal cancer. It includes a spontaneous speech task recorded during a semi-directed interview, but also self-questionnaires allowing the measurement of communication impairment and factors associated with speech and communication.

Regarding the first aspect, a reference score measuring communication in a holistic way has been constructed. It allows to fill the lack of tools available in ENT oncology for this measurement.

The second aspect concerns the automatic analysis of speech. A systematic review of the literature led us to focus on tools applicable to the analysis of spontaneous speech, which is the production context closest to everyday communication. One hundred forty-nine Automatic parameters from the different levels of Caron's psycholinguistic communication model were extracted. Then, a selection process led to retain 75 relevant and non-redundant parameters.

Finally, for the third aspect, we conducted predictive modeling of communication impairment using the selected automatic parameters (correlation of 0.83 between predicted and actual score). The correlation even reached 0.89 when including associated factors (constitution of social circles, anxiety-depression state, associated deficits, self-perception of the handicap linked to the speech disorder) in the modelling.

The use of automatic speech analysis thus allows a reliable prediction of the communication impairment felt by the patients. This study opens new perspectives for the use and optimization of automatic speech recognition systems in clinical evaluation on the one hand, and the consideration of functional and psychosocial needs expressed by patients on the other hand.

Keywords: Speech, Automatic analysis, Communication, Cancer