

## I. Caractéristiques de la parole

La production de la parole est un phénomène complexe. Elle implique de conceptualiser et de programmer un message avant de le produire. Sur le plan physiologique cela nécessite la mobilisation et la coordination de nombreux organes et de plus d'une centaine de muscles.

### A. Physiologie

La phonation est un « acte moteur complexe », qui met en jeu un nombre important de « muscles, de variables physiologiques et un contrôle nerveux » (Auzou et al., 2001). Le système phonatoire est composé de trois sous-systèmes :

- Le système respiratoire (étage sous-glottique). Il permet, grâce au souffle, de produire un son sur l'expiration en évacuant l'air des poumons.
- Le système glottique (étage glottique). Il est constitué du larynx et de ses plis vocaux, les cordes vocales, qui vibrent grâce à la pression sous-glottique, ce qui crée un geste phonatoire.
- Le système de résonance / articulatoire (étage sus-glottique). Il comprend les résonateurs (pharynx, cavités buccales et nasales) qui permettent la modulation du signal sonore et les organes phonatoires (maxillaire, mandibule, lèvres, langue, joues, voile du palais...) qui jouent un rôle dans la précision articulatoire et le dosage du souffle.

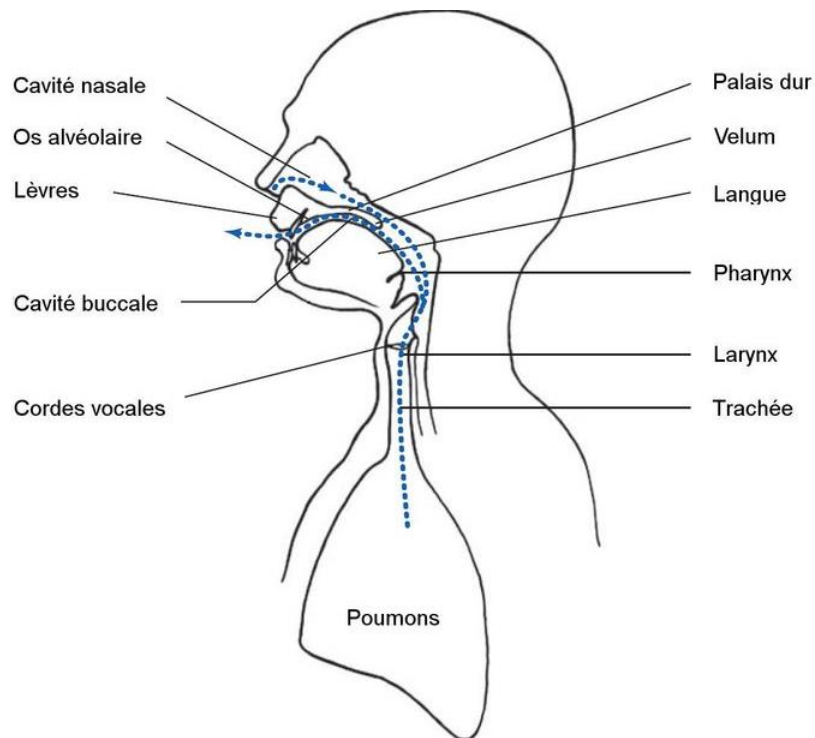


Figure 1 : Organes impliqués dans la production de la parole

La coordination entre ces trois sous-systèmes permet ainsi une bonne production des sons de la parole qui peuvent être examinés à partir d'une analyse phonétique, dans le cadre de la description de la parole normale ou pathologique.

## B. Description

L'évaluation phonétique consiste à évaluer la parole par identification et reconnaissance des traits caractéristiques des phonèmes. On distingue alors les phonèmes vocaliques des phonèmes consonantiques.

L'analyse des voyelles s'effectue selon 4 paramètres (Meunier, 2007) :

- Le **lieu d'articulation**. Il est déterminé par le mouvement de la langue en position avant ou arrière au sein de la cavité buccale, et permet de différencier les voyelles antérieures des voyelles postérieures.
- Le **degré d'aperture**. Il est déterminé par l'ouverture de la mandibule qui permet de différencier les voyelles ouvertes des voyelles fermées.
- L'**arrondissement**. Il est déterminé par la position et la projection des lèvres en avant et permet de distinguer les voyelles arrondies (non étirées), des voyelles non-arrondies (étirées).
- La **nasalité**. Elle dépend de la position du vélum qui permet ou non le passage de l'air vers les fosses nasales et permet de distinguer les voyelles nasales des voyelles orales.

L'analyse des consonnes s'effectue quant à elle selon 3 paramètres (Auzou et al., 2001) :

- Le **voisement**. Il fait référence aux vibrations des cordes vocales et permet de distinguer les consonnes sonores des consonnes sourdes.
- Le **mode d'articulation**. Il est déterminé par la manière dont l'air passe lors de l'articulation et permet de distinguer les consonnes occlusives, nasales, fricatives, glissantes et liquides.
- Le **lieu d'articulation**. Il est lié à la fermeture du conduit phonatoire et permet de distinguer les consonnes bilabiales, labio-dentales, apico-dentales, alvéolaires, vélopalatales...

L'évaluation perceptive, quant à elle, consiste à décrire les caractéristiques de la parole afin de repérer les anomalies et classer les pathologies (Auzou et al., 2001).

En 2001, Auzou explique que « les sons de la parole ont pour origine des phénomènes aérodynamiques et acoustiques ». Ainsi, pour analyser les gestes articulatoires, trois méthodes sont possibles : au niveau du contrôle neuromoteur des muscles, au niveau du mouvement des organes articulateurs, et au niveau des phénomènes que ces mouvements induisent (paramètres aérodynamiques et signal acoustique). Pour ce dernier niveau, les méthodes d'investigation

acoustique sont « fondées sur l'analyse du signal de la parole selon deux aspects » : l'information segmentale (durée des segments phonétiques, timbre et qualité vocale) et l'information suprasegmentale (rythme, débit de la parole, intensité et mélodie). Ces deux types d'information sont accessibles lors de l'analyse spectrale. Cette analyse permet de mieux appréhender les « caractéristiques pathologiques du signal mal ou pas perçues à l'écoute de la voix du patient » (Auzou et al., 2001). L'outil utilisé, le spectrogramme, correspond à une représentation visuelle du signal d'une onde sonore, en trois dimensions. La représentation des voyelles sur le spectrogramme se caractérise par la « présence de zones d'harmoniques renforcées » que sont les formants. Ils sont au nombre de trois et sont déterminés par les paramètres cités précédemment :

**F1** = degré d'aperture et hauteur de la langue

**F2** = position de la langue et position des lèvres

**F3** = position des lèvres

Les mouvements des articulateurs nous renseignent donc sur la configuration formantique des voyelles.

La représentation des consonnes est liée au mode et lieu d'articulation. Les occlusives orales comportent deux séquences temporelles : la tenue et/ou le silence, puis le relâchement. Pour les occlusives orales apériodiques (/p/, /t/, /k/), durant le maintien de l'occlusion, on observe un silence sur le spectrogramme. A la fin de l'occlusion, un bruit impulsif d'explosion apparaît (appelé « burst »), correspondant au relâchement brutal du contact de l'occlusion (avec les articulateurs). Pour les occlusives orales périodiques (/b/, /d/, /g/), on retrouve le même mécanisme, mais une « barre foncée de voisement » correspondant à la vibration des cordes vocales s'ajoute et est située dans les basses fréquences. Les fricatives sont caractérisées par la présence d'un « bruit » durant leur tenue qui correspond au rétrécissement partiel du tractus vocal. A nouveau, les fricatives peuvent être apériodiques (/f/, /s/, /ʃ/), ou périodiques (/v/, /z/, /ʒ/) selon l'aspect sonore et sourd de la consonne. Lors de l'analyse spectrale, le phénomène de co-articulation est à prendre en compte. Cela correspond à un « chevauchement » des articulateurs lors de l'enchaînement des segments phonétiques, une dépendance des gestes articulatoires entraînant une anticipation de la production du futur phonème. Néanmoins, lorsqu'il n'y a pas de pathologie, un phonème est identifiable selon ses caractéristiques propres lors de la lecture sur le spectrogramme (Meunier, 2007). Cette première partie décrit les grandes caractéristiques de la parole « normale », celle qui s'oppose à la parole « pathologique ». Toutefois, il semble important de préciser que la notion de « norme » est difficile à déterminer dans le domaine de la parole car il existe de nombreuses variabilités inter-locuteurs (accents régionaux, habitudes...). Nous considérerons ainsi comme «

parole normale », une parole qui ne représente pas de plainte et de gêne pour le locuteur, qui est compréhensible et n'entrave pas la communication pour l'interlocuteur.

A contrario, certaines pathologies amènent un patient à présenter des perturbations au niveau de sa parole. Les patients atteints de Troubles de la Production de Parole (TPP) sont décrits comme ayant une parole altérée. Nous nous intéressons alors à cette parole pathologique.

## II. Caractéristiques de la parole pathologique

Les troubles de la parole regroupent plusieurs pathologies telles que le bégaiement, l'apraxie, la dysarthrie, la dysprosodie et les troubles de l'articulation. Selon les auteurs, les contraintes articulatoires relevées dans le cadre des paralysies faciales sont nommées « trouble articulatoire » ou « dysarthrie ».

### A. Le trouble articulatoire

Le trouble articulatoire est défini par une « erreur permanente et systématique dans l'exécution du mouvement » lors de l'émission phonétique, quelle que soit sa position dans le mot ou la phrase. Il se caractérise par des positions incorrectes qui peuvent être dues à une constitution anormale des organes. Ces troubles articulatoires peuvent être relatifs à une cause organique, fonctionnelle ou perceptive. Ils peuvent être d'origine développementale (chez les enfants) ou acquise (chez l'adulte). Ils touchent le phonème de manière isolée, indépendamment du contexte phonétique. Le phonème peut ainsi être manquant (par omission), remplacé par un autre phonème (par substitution) ou déformé (par distorsion) (ASHA, 2019).

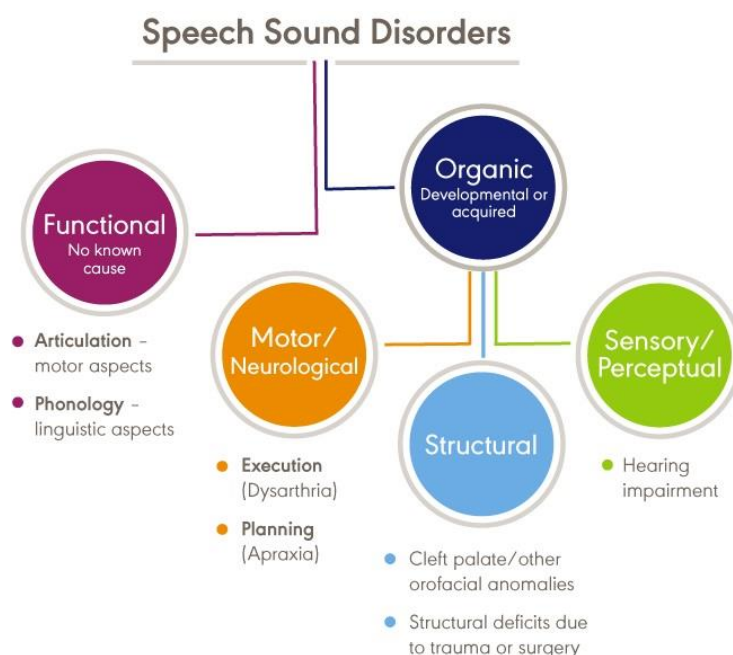


Figure 2 : Causes du trouble articulatoire (ASHA, 2019)

Parmi les troubles articulatoires, la dysarthrie a fait l'objet de nombreux travaux, dont les résultats ont pu contribuer aux réflexions sur la prise en charge des troubles articulatoires de différentes étiologies.

## B. La dysarthrie : Définition, classification et évaluations

En 1975, Darley et collaborateurs définissent la dysarthrie comme « un trouble de la réalisation motrice de la parole » lié à des « lésions du système nerveux, périphérique ou mixte », intégrant diverses perturbations sur la « respiration, la phonation, l'articulation, la résonance et la prosodie ». Auzou en 2001 précise que le terme de dysarthrie concerne également les perturbations restreintes à un seul nerf crânien, telle qu'une « atteinte articulatoire isolée par atteinte du grand hypoglosse (XII) ». La classification la plus utilisée est issue des travaux de Darley et collaborateurs en 1975 et propose six types de dysarthries : la dysarthrie flasque, spastique, ataxique, hypokinétique, hyperkinétique et les dysarthries mixtes. Cette classification a constitué une avancée majeure mais présente néanmoins certaines limites. Ainsi, en clinique, le type de dysarthrie n'est pas toujours précisé et les troubles observables sont davantage liés au site lésionnel qu'à l'étiologie.

Auzou définit 5 grandes approches pour l'évaluation clinique de la dysarthrie : l'analyse perceptive, l'analyse phonétique, l'examen sensori-moteur, l'auto-évaluation et l'analyse acoustique.

- L'analyse perceptive consiste à repérer les distorsions de parole afin d'isoler les caractéristiques essentielles de celle-ci. Elle inclut les concepts de compréhensibilité, d'efficacité et de sévérité perceptive. Elle permet également de mesurer l'intelligibilité qui est définie comme « le degré de précision avec lequel le message est compris par l'auditeur ». Or, la diminution de l'intelligibilité représente la plainte principale des patients et de leurs proches, reflète le handicap majeur et traduit le degré de sévérité de la dysarthrie. Ainsi, l'analyse perceptive est une approche fondamentale et « la plus immédiate en pratique clinique » (Amosse et al., 2004). Pour qu'elle soit la plus complète possible, il est nécessaire de proposer une approche quantitative qui permet de donner un indice global du caractère naturel de la parole (conversion spontanée ou lecture à voix haute) et une approche plus qualitative et analytique afin d'identifier la nature des perturbations (lecture de mots et de phrases). Cependant, étant donné la complexité des mécanismes mis en œuvre dans la parole, l'évaluation perceptive trouve ses limites dans l'appréhension de tous les dysfonctionnements.

- L'analyse phonétique décrit les différentes altérations des phonèmes afin de qualifier le trouble articulaire ainsi que le type de dysarthrie. Elle permet, grâce à la transcription phonétique, une « mesure globale de la justesse articulaire » et une « base phonétique pour la rééducation de ces troubles ». Economique et facile d'utilisation en pratique courante, elle est néanmoins subjective car elle dépend de la perception de la parole qui n'est pas toujours en adéquation totale avec le signal produit (Auzou, 2001).
- L'examen sensori-moteur fournit « une première analyse physiopathologique des dysfonctionnements en cause dans la dysarthrie » en évaluant les diverses fonctions et organes impliqués dans la parole. Il permet notamment une évaluation comparative après remédiation.
- L'auto-évaluation recense quant à elle « le handicap et le retentissement sur la qualité de vie » en recueillant la perception du patient vis-à-vis de son propre trouble.
- L'analyse acoustique représente un « outil privilégié d'étude de la parole » utilisé afin de déterminer « les relations entre les anomalies phonétiques et acoustiques » (Auzou, 2007). Elle permet d'analyser différents paramètres physiologiques de la parole et de la voix (fréquence fondamentale, intensité et hauteur, débits...) afin de fournir une analyse qualitative et descriptive de la parole pathologique. Grâce à l'analyse spectrale, le clinicien a accès à certaines caractéristiques pathologiques du signal peu ou mal perçues à l'oreille (Auzou, 2001). Ces différentes approches sont complémentaires et nécessitent d'être confrontées afin de rechercher les corrélations et de « connaître les liens qui unissent fonctionnement des organes et intelligibilité » (Auzou, 2001). La nécessité d'inclure ces différentes analyses est rappelée par Amosse et collaborateurs en 2004 qui affirment que « le phénomène acoustique n'est pas toujours équivalent à la perception auditive » et qu'il est ainsi nécessaire de proposer une évaluation subjective et objective. En effet, certaines modifications de la parole peuvent ne pas être perçues à l'oreille humaine mais être objectivées de manière acoustique. De plus, les caractéristiques intrinsèques au patient doivent être prises en compte lors de l'évaluation afin que celle-ci soit la plus fonctionnelle possible et réponde au mieux à la demande du patient et de son entourage. La parole est un phénomène complexe et son analyse exige un regard sous différents angles. Dès lors, son altération nécessite une évaluation précise et multiple afin de déterminer les différents retentissements et répondre à la plainte du patient dans le cas des paroles pathologiques.

Les évaluations perceptive et acoustique utilisées pour la dysarthrie, le sont également dans d'autres pathologies présentant des troubles de la parole. Les tâches demandées aux sujets servent à mesurer différents paramètres, comme le degré d'altération des sons ou le degré de

compréhensibilité du message ou les deux à la fois, et ce à différents niveaux linguistiques. Ces évaluations sont au cœur de la pratique du clinicien.

La description du fonctionnement de l'appareil phonatoire, décrit précédemment (cf. I. Caractéristiques de la parole), correspond à un état de fonctionnement "normal". Cependant, une atteinte des organes de cet appareil peut en perturber le fonctionnement, cela peut avoir des conséquences sur la production de parole en perturbant la phonation, la respiration, l'articulation, la coordination entre les différents niveaux impliqués dans la production de la parole (cf. II. Caractéristiques de la parole pathologique).

Parmi les organes impliqués dans le phénomène complexe qu'est la production de parole, les organes constituant les Voies Aéro Digestives Supérieures (VADS) sont les lèvres, la cavité buccale, la langue et les palais dur et mou, le pharynx et l'œsophage qui forment les voies digestives supérieures. Les voies aériennes supérieures sont composées des fosses nasales, des sinus, des cavités de l'oreille moyenne, du pharynx, du larynx et de la trachée. Les Cancers des VADS, ainsi que leurs traitements peuvent perturber ce fonctionnement de la parole. Nous en faisons une présentation dans la partie suivante.

### C. Les cancers des Voies Aéro-Digestives Supérieures : anatomie, généralités, traitement

Les voies aéro-digestives supérieures (VADS) sont constituées de la partie haute des systèmes digestif et respiratoire. Nous les divisons en 3 sites principaux : la cavité orale, le pharynx et le larynx. Elles comprennent ainsi :

- La cavité orale

Elle est constituée : des lèvres rouges, du plancher buccal, de la langue mobile et du palais.

- Le pharynx

Il est formé de trois portions : le nasopharynx (ou cavum), l'oropharynx (base de langue – voile du palais – amygdales palatines) et le pharyngolarynx ou l'hypopharynx (sinus piriformes – paroi pharyngée postérieure – région post cricoïde).

- Le larynx

Il est divisé en trois étages : supra-glottique, glottique et sous-glottique.

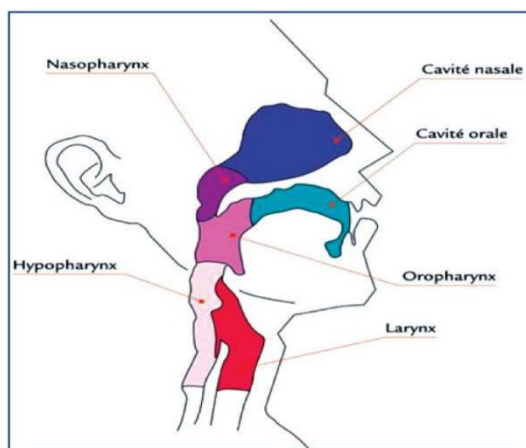


Figure 3 : Représentation des voies aéro-digestives supérieures (Beby-Defaux et al., 2011)

Avec plus de 20 000 nouveaux cas par an en France, les cancers oto-rhino-laryngologiques (ORL) ou cancers des VADS représentent un véritable problème de Santé Publique. En effet, ils constituent la 5<sup>e</sup> cause de cancer en France (Balaguer et al., 2019). L'âge moyen de survenue est de 63 ans. Dans les 3/4 des cas, ces cancers surviennent chez des hommes, et dans 1/4 chez des femmes. Il s'agit dans 90% des cas de carcinomes épidermoïdes, c'est-à-dire des cancers développés à partir de la muqueuse. Il existe également des tumeurs plus rares pouvant correspondre à d'autres histologies. Un tabagisme chronique et/ou une consommation alcoolique sont les deux facteurs de risques principaux. Nous assistons depuis quelques années à une augmentation due à une infection virale à Human Papilloma Virus (HPV), retrouvé dans le cancer du col de l'utérus généralement. Ces cancers HPV sont de meilleur pronostic, et peuvent se déclarer chez des sujets jeunes. Par ailleurs, dans une moindre mesure, l'exposition professionnelle aux poussières de bois est un facteur de risque reconnu des cancers des VADS, en particulier du nasopharynx, des fosses nasales et des sinus. Les symptômes amenant à consulter peuvent être une douleur, une gêne à l'alimentation ou à la parole, des saignements, ou des signes plus généraux comme une fatigue ou une perte de poids.

Suite à cette première consultation, un bilan complet est réalisé. Il comprend des imageries médicales (scanner ou IRM) afin de faire le point sur la lésion, son extension aux tissus adjacents mais aussi pour rechercher des lésions à distances (métastases). Des biopsies vont être réalisées, s'il le faut sous anesthésie générale, ainsi qu'un bilan nutritionnel, parfois psychologique. A l'issue de ce bilan, le dossier du patient est discuté en réunion de concertation pluridisciplinaire, où tous les acteurs de la prise en charge sont présents : ORL, radiologues, chimiothérapeutes, radiothérapeutes, ... La tumeur est classée selon sa taille (T), le nombre de ganglions atteints (N) et la présence de métastases ou pas (M). En fonction de cette classification TNM, de l'état général



du patient, de la localisation de la tumeur et d'autres facteurs encore, le traitement dont va bénéficier le patient va être décidé de manière collégiale.

La chirurgie est, la plupart du temps, le traitement initial d'un cancer des voies aéro-digestives supérieures (La ligue contre le Cancer & Institut National du Cancer, 2012). L'objectif, lorsque cela s'avère possible, est de retirer la totalité de la tumeur. Le chirurgien va aussi enlever les tissus les plus proches de la lésion cancéreuse, afin d'éradiquer un maximum de cellules malignes. En fonction de l'organe atteint par la maladie et les caractéristiques de celle-ci, les conséquences fonctionnelles du geste chirurgical seront plus ou moins importantes. Les chirurgies ORL peuvent engendrer des séquelles extrêmement invalidantes dans la mesure où les organes touchés sont le siège de fonctions essentielles. D'un point de vue esthétique, ces chirurgies peuvent également être particulièrement mutilantes. Dans la mesure du possible, le chirurgien va chercher à préserver au mieux les fonctions de l'organe concerné. Dans certains cas, une reconstruction anatomique est aussi possible, voire nécessaire, afin de limiter l'impact fonctionnel et esthétique lié au traitement chirurgical.

La radiothérapie externe est également un des piliers du traitement de ces cancers. Ces dernières années, la cancérologie des voies aéro-digestives supérieures a été marquée par d'importantes avancées en matière de chirurgie carcinologique et reconstructrice, en matière de chimiothérapie et de thérapies ciblées, mais aussi grâce à l'évolution de la radiothérapie conformationnelle en 3 dimensions.

La radiothérapie externe tient une place prépondérante dans la prise en charge thérapeutique des cancers de la tête et du cou. Elle peut être utilisée à titre exclusif ou en complément de la chirurgie. Son association à une chimiothérapie concomitante permet, dans des indications précises et de plus en plus fréquentes, d'en améliorer l'efficacité. L'évolution et l'amélioration des techniques de radiothérapie externe visent autant à augmenter l'efficacité de la radiothérapie qu'à en diminuer les effets secondaires immédiats et les séquelles tardives. Une radiochimiothérapie est indiquée d'emblée en cas de tumeur localement évoluée et inopérable. Elle peut également être indiquée en postopératoire, dans des situations précises. La chimiothérapie a montré son intérêt dans des situations carcinologiques précises et en cas de marges chirurgicales positives. La chimiothérapie peut également être utilisée seule avant la prise en charge locale, dans une stratégie de préservation laryngée. Cela concerne les grosses tumeurs du larynx et de l'hypopharynx, qui justifieraient une laryngectomie ou une pharyngolaryngectomie totale. Deux ou trois cycles de chimiothérapie sont réalisés et en cas de bonne réponse, la chirurgie est évitée et le traitement consiste en une radiothérapie ou une radiochimiothérapie. La radiothérapie altère la voix si le larynx est inclus dans le champ d'irradiation. La radiothérapie peut entraîner une atrophie du muscle vocal, et une limitation de la mobilité laryngée due à la fibrose : on observe fréquemment un défaut

d'accolement cordal en phonation. La sécheresse de la muqueuse laryngée, un œdème ou un érythème de la muqueuse de la corde vocale diminuent l'amplitude de la vibration cordale en phonation. Les modifications de la voix incluent des modifications de la hauteur. Elle est, le plus souvent, plus grave par défaut de tension des cordes vocales, parfois plus aiguë par excès de tension et rigidité. Les variations de hauteur dans la parole sont diminuées. Le timbre peut être altéré, voilé par défaut d'accolement cordal, parfois éraillée ou serrée. La voix est fatigable, manque d'intensité (Stoicheff et al., 1983). Les paramètres acoustiques et aérodynamiques sont modifiés mais semblent s'améliorer avec le temps tout en restant à un an après la fin de la radiothérapie en dessous des valeurs normales (Bibby et al., 2008).

La parole peut être altérée par la diminution de l'amplitude des mouvements articulatoires due à la perte de force musculaire, à la rigidité et par la perte de sensibilité. Les voyelles sont modifiées dans leur aperture par la limitation de l'ouverture mandibulaire, dans la position de la langue et dans la différenciation orale/nasale. L'articulation des consonnes linguales antérieures peut être perturbée du fait de la diminution de mobilité de la langue (fibrose linguale, neuropathies du nerf XII). On peut observer une rhinolalie, une difficulté à réaliser les consonnes linguales postérieures comme le /R/, dues à un voile fibreux donc peu mobile.

Nous venons de distinguer l'ensemble des séquelles fonctionnelles qu'entraînent ces cancers des VADS. Il est redoutable pour ces patients de voir leurs échanges communicationnels se détériorer. Il convient alors de leur apporter un soutien pluridisciplinaire afin d'améliorer leur qualité de vie. C'est ainsi que la prise en soin par l'orthophoniste et l'évaluation des troubles vont devenir primordiaux.

### III. La prise en charge et l'évaluation des troubles de la parole

Dans le domaine de la cancérologie de la sphère ORL et cervico-faciale, s'il semble naturel que l'orthophoniste ait sa place dans l'équipe médicale de suivi du malade ; celle-ci peut apparaître le plus souvent réduite au rôle de l'apprentissage d'une nouvelle voix, en cas de pharyngo-laryngectomie totale. Or, de plus en plus souvent, les orthophonistes prennent en charge des malades ayant subi des interventions chirurgicales partielles, ou des protocoles radio-chimiothérapeutiques qui entraînent des troubles fonctionnels différents, cela en raison même des progrès médicaux, du dépistage ou des traitements.

#### A. Le rôle de l'orthophoniste

L'orthophoniste, dans le domaine de l'oncologie ORL, travaille principalement autour de trois axes qui sont : la dysphagie, la phonation et l'articulation. Les troubles fonctionnels ne concernent

pas seulement la perte ou la perturbation de la fonction vocale, la sphère oro-bucco-pharyngée comprend 4 fonctions essentielles : l'articulation de la parole, la respiration, la déglutition et la phonation, et chacune de ces fonctions peut être troublée, parfois concomitamment. « Les patients présentant un cancer des VADS ont un besoin important de prise en charge rééducative » (Association Francophone des Soins Oncologiques de Support, 2014).

Cette rééducation doit démarrer précocement, parfois même en amont du traitement. En ce qui concerne les troubles de la mastication et de la déglutition, l'objectif est de permettre au patient de s'alimenter per os, en garantissant des apports liquidiens et caloriques suffisants, et en évitant les fausses routes. Dans le cadre des laryngectomies partielles, la rééducation de la phonation dépend de la chirurgie dont le patient a bénéficié. L'objectif généralement est d'utiliser les structures restantes afin de recréer un sphincter néoglottique fonctionnel. La rééducation de la phonation dans le cas des laryngectomies totales est sensiblement différente. Le patient doit alors apprendre à réaliser une production vocale selon un mécanisme différent de celui qu'il a toujours connu, à partir du sphincter supérieur de l'œsophage. Le travail de l'articulation, généralement nécessaire lors des chirurgies de la cavité buccale et de l'oropharynx, consiste notamment à proposer de nouveaux traits articulatoires au patient et de l'aider à se les approprier puis à les automatiser. Ce sont des mécanismes de compensation. L'orthophoniste a également un rôle de surveillance auprès de son patient. Il doit tenir compte par exemple de plaintes récurrentes à propos de douleurs, de perte de poids, d'une apparition de ganglion suspect, de fatigue chronique, etc. Il a aussi une fonction de prévention, d'éducation thérapeutique auprès du patient et de son entourage, de soutien et d'accompagnement (Ministère de l'Éducation nationale, 2013). Enfin, l'orthophoniste, en oncologie ORL, fait partie intégrante d'une équipe interdisciplinaire. Les troubles qu'il rééduque ne peuvent être traités de façon isolée. Des liens étroits avec les autres acteurs de l'équipe soignante sont nécessaires afin de proposer au patient une prise en charge personnalisée et adaptée à ses problématiques.

En effet, la plupart du temps, lorsque les malades sont suivis en orthophonie, après la décision thérapeutique et à leur sortie d'hospitalisation, et parfois au cours même de leur hospitalisation, l'orthophoniste a un rôle essentiel à jouer auprès du malade : il concourt à l'éducation thérapeutique du malade par l'information – les conséquences fonctionnelles du traitement, « que va-t-on me faire ? », par les conseils d'hygiène et de diététique, il guide le comportement physique – les explications anatomo-physiologiques – « qu'est-ce qui n'est plus pareil ? », « qu'est-ce qu'on m'a fait » ? – il conseille l'entourage du malade dans leur accompagnement au quotidien. Tout cela revêt une importance primordiale pour la réhabilitation du malade ; sans oublier le soutien psychologique dans le travail de deuil de la perte du larynx et de la voix, ou de la langue, par exemple. Dans ce domaine tout à fait particulier, celui de la réhabilitation globale du malade avant

et après ses soins médicaux, l'orthophoniste va pouvoir exercer pleinement son art, qui n'est pas seulement, celui du technicien qui essaie de redonner une voix de substitution à son malade.

Au sein de la prise en soin de ces patients, l'évaluation du trouble de sa parole occupe une place importante puisque le bilan doit permettre l'établissement d'un projet thérapeutique le plus adapté possible aux difficultés du patient.

## B. L'évaluation d'un trouble de la parole

L'évaluation clinique exhaustive d'un trouble de la parole implique six approches distinctes mais interdépendantes (Auzou, 2008), qui permettent d'obtenir une vision complète de la situation du patient, en étudiant les trois composantes du handicap. La Classification Internationale du Handicap (CIH) distingue ainsi les trois composantes qui sont la déficience, l'incapacité et le désavantage :

- L'examen sensori-moteur des organes impliqués dans la parole, l'analyse phonétique et l'analyse de la voix permettent d'estimer l'altération organique et donc la déficience.
- L'analyse perceptive globale des anomalies de la parole et la mesure de la sévérité (intelligibilité, compréhensibilité, efficacité, caractère naturel) rendent compte du degré d'incapacité à accomplir une activité d'une façon normale, qui résulte de la déficience.
- L'auto-évaluation (Letanneux et al., 2013) ou un questionnaire très précis lors de l'anamnèse (Robert, 2009) permettent de connaître le retentissement du trouble sur la qualité de vie du patient. Cette approche reflète donc le désavantage social du patient, c'est-à-dire le handicap à proprement dit, occasionné par la déficience ou l'incapacité et impactant le patient en tant qu'acteur social.

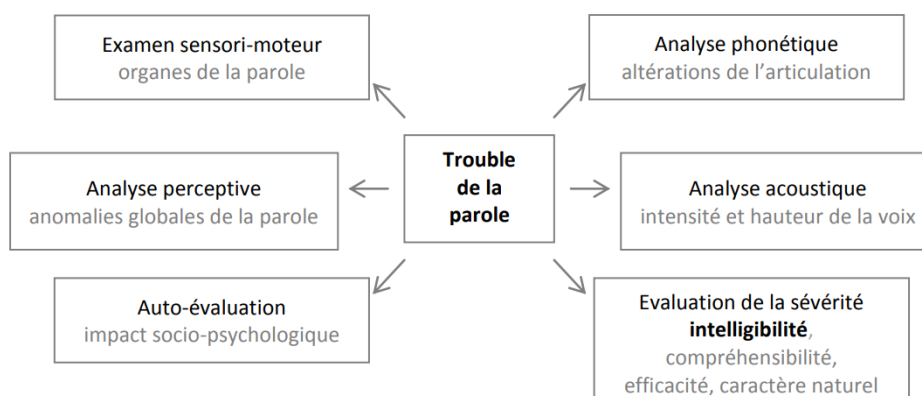


Figure 4 : Evaluation d'un trouble de la parole

Pour que le message d'un locuteur soit correctement reconnu par un récepteur, il faut que ce message soit intelligible. Ainsi, l'intelligibilité peut être mesurée pour les trois niveaux de la communication : la production du message (intelligibilité de la parole), le canal de transmission (l'acoustique d'une salle, par exemple) ou les capacités de réception (l'audition). En pratique clinique, et notamment en cancérologie ORL, le clinicien va ainsi vouloir évaluer le trouble de production de la parole de son patient, et cela est possible par l'évaluation de l'intelligibilité. Dans notre travail, nous traitons de l'intelligibilité de la parole, donc de la production du message. Il existe plusieurs définitions de l'intelligibilité. Certains la décrivent comme le degré de précision avec lequel le message est compris par l'auditeur. Elle s'obtient en déterminant le nombre d'unités de parole reconnues par l'auditeur (Yorkston, 1992). L'intelligibilité de la parole peut être altérée par le dysfonctionnement de l'une ou l'autre des fonctions impliquées dans la réalisation motrice de la parole : phonation, articulation, résonance, prosodie, respiration. Une réduction de l'intelligibilité engendre un trouble du discours et représente souvent la plainte essentielle des patients. Selon la personnalité du patient et ses implications socio-professionnelles, cette perte d'intelligibilité sera ressentie comme un handicap plus ou moins important (Auzou, 2008). Les concepts de perception et de compréhension font également intervenir une autre notion plus étendue qu'il paraît important de préciser, celle de sévérité. C'est une notion complexe qui n'a pas de réelle définition, si ce n'est qu'elle doit être établie sur la base de plusieurs critères de mesures. Les tâches demandées aux sujets servent à mesurer différents facteurs, comme le degré d'altération des sons ou le degré de compréhensibilité du message ou les deux à la fois, et ce à différents niveaux linguistiques. En effet, ce n'est pas parce qu'un son est mal produit que l'ensemble du message est incompréhensible, ce nivellement a un impact sur la mesure. Le professeur Virginie Woisard s'est ainsi intéressée à ce sujet, en faisant bien la distinction entre intelligibilité et compréhensibilité (Rebourg, 2020). L'intelligibilité est la précision avec laquelle un auditeur récupère le signal acoustique produit par le locuteur, alors que la compréhensibilité est la capacité de l'auditeur à interpréter le sens du message oral produit par le locuteur sans tenir compte de la précision ou de la justesse phonétique ou lexicale. Dans notre travail, nous considérons l'intelligibilité comme « la quantité de parole comprise à partir du seul signal acoustique ». Nous pouvons ainsi tracer une relation étroite entre intelligibilité et informations dépendantes du signal (bottom up) liées au processus de décodage des sons. Nous traitons la capacité communicationnelle des patients atteints d'un cancer de la cavité buccale ou de l'oropharynx. Il est donc important pour la recherche et la pratique clinique de prendre en compte la difficulté de l'évaluation clinique de l'intelligibilité afin de l'améliorer.

Plusieurs tests existent afin d'évaluer l'intelligibilité. C'est d'abord en langue anglaise que les premiers tests ont vu le jour, ce qui a ouvert la voie à des recherches sur d'autres langues comme

le français. C'est d'ailleurs dans le cadre des dysarthries que se sont le plus développées ces recherches. Un des premiers tests pour le français est l'évaluation des « Paires Minimales » de Peckels & Rossi (1973). C'est le plus ancien et certainement le plus abouti des tests mais peu usité à l'heure actuelle. En effet, la passation nécessite la production de 216 mots, ce qui est bien trop contraignant pour un patient atteint de troubles de la production de parole. « Le Single Word Intelligibility Test (Kent, 1989) a été adapté et traduit en français par Gentil (1992) puis repris par Auzou (1998) dans l'ECD puis par Crochemore & Vannier (2001) pour devenir le Test Phonétique d'Intelligibilité de la Batterie d'Evaluation Clinique de la Dysarthrie (BECD) réalisé par Auzou et Rolland-Monnoury (2006). » (Ghio, 2016). La BECD attribue un score d'intelligibilité à partir de la lecture de 10 mots et de 10 phrases, or, plusieurs biais sont à mettre en avant. D'abord la liste de mots ne contient que 50 mots, de 1 ou 2 syllabes, avec des mots rares et elle ne contient pas de répartition phonétique pertinente. Ensuite les phrases sont aussi au nombre de 50, avec une structure syntaxique identique et des amorces communes : ainsi, les mots comme les phrases sont facilement mémorisables et prévisibles. Le Frenchay Dysarthria Assessment (FDA), développé par Enderby (1983) est un test anglo-saxon qui consiste à attribuer un score en faisant la somme obtenue sur 8, à trois épreuves : la lecture de 10 mots, la lecture de 10 phrases et en parole spontanée. Une nouvelle version de ce test a été proposée en 2008 par Enderby et Palmer, le FDA-2. Ils ont amélioré la sélection des mots avec une meilleure répartition des phonèmes, ils ont aussi repris les amorces de phrases qui étaient trop répétitives. Ce test présente néanmoins quelques limites, les mots eux-mêmes et leur fréquence d'apparition sont susceptibles de créer des effets d'apprentissage et donc ensuite des effets de restauration pour la personne pratiquant l'évaluation. Il en va de même pour les phrases qui sont en nombres limitées. Ce dernier a également été adapté au français par Emilie Blanc et Laurence Giusti (Mémoire d'orthophonie 2014, et Alain Ghio et al. pour les Journées d'Etudes sur la Parole 2016).

Plusieurs outils existent donc actuellement pour évaluer l'intelligibilité mais aucun ne satisfait pleinement les contraintes cliniques. En effet, l'ensemble de ces tests permet de mettre en avant certains biais méthodologiques comme la lexicalisation, les effets d'apprentissage qu'il convient de prendre en compte afin de les éviter lors de l'élaboration de nouveaux tests d'évaluation de l'intelligibilité. La lexicalisation des items présentés accentue les effets d'apprentissages surtout lorsqu'on a affaire à une liste avec peu d'items. Il y a des effets de mémorisation des mots, ce qui produit des effets de restauration chez les auditeurs, surtout lorsque la liste peut être apprise par cœur : un mot mal prononcé est reconnu par un auditeur car il l'a déjà dans son stock lexical. Ces effets peuvent fausser la perception de l'auditeur. Un corpus de test d'intelligibilité doit comporter un grand nombre d'items pour éviter ces fameux effets d'apprentissage. Il doit prendre en compte les règles phonotactiques de la langue et les fréquences d'apparition des phonèmes. Il faut donc

une représentation de la langue qui soit complète. Il doit réduire au minimum la prédictibilité des items. Donc pour constituer un corpus de test d'intelligibilité il faudrait idéalement, un grand nombre d'items, construits en suivant les règles de la langue cible et qui seraient présentés de manière aléatoire. Ainsi, l'ensemble de ces réflexions ont permis de mettre en exergue les problèmes existants liés à la mesure perceptive de l'intelligibilité, et donc de réfléchir à élaborer le test de mesure de l'intelligibilité par la tâche de Décodage Acoustico Phonétique, tel qu'elle a été conçue dans le projet C2SI. Cela afin d'éviter les effets décrits précédemment.

## IV. La tâche de Décodage Acoustico-Phonétique (DAP)

La perte d'intelligibilité représente la plainte principale des patients atteints de Troubles de la Production de Parole Pathologique (TPPP). Elle affecte la sphère privée et publique en impactant plus ou moins fortement la communication verbale. L'intelligibilité s'évalue du point de vue de la perception, donc par des tests perceptifs. Cependant les patients répondent aussi à des auto-questionnaires qui sont très subjectifs, tels que le Speech Handicap Index (SHI) et le Parole Handicap Index (PHI). La tâche de Décodage Acoustico-Phonétique élaborée dans le cadre du projet se veut plus objectif car il contourne les différents biais des autres tests existants à ce jour, comme la restauration lexicale et les effets d'apprentissages par exemple qui ont été présentés dans les études de Marie Rebourg (Mémoire master 2 et Publication pour les Journées d'Etudes sur la Parole 2020).

### A. Projets C2SI & RUGBI

Le projet de recherche C2SI, pour Carcinologic Speech Severity Index (C2Si) vise au développement d'un indice de sévérité des troubles de la production de parole. Mené de 2014 à 2018, il a obtenu des financements à hauteur de 230,4K euros par l'Institut National du Cancer (INCa). Sous la direction du professeur Virginie Woisard, ce projet avait pour but le développement d'un indice de sévérité des troubles de production de la parole pour la mesure de l'impact des traitements des cancers de la cavité buccale et de l'oropharynx sur la production de la parole, à la fois par des méthodes perceptives et par traitement automatique de la parole (Astesano et al., 2018). V. Woisard, médecin ORL au Centre Hospitalier Universitaire (CHU) de Toulouse dans l'Unité de la voix et de la déglutition, Docteur en Science du Langage, ainsi que les orthophonistes du service ont une expérience clinique dans la médecine et dans la rééducation indispensable au projet. Leur cadre professionnel leur permet d'avoir accès à une population de patients importante ce qui a permis d'envisager et de concrétiser ce projet. ; Elle s'est entourée

d'équipes spécialisées en linguistique et prosodie et en Traitement Automatique des Langues (TAL) pour mener à bien son projet. D'autres partenaires participent à ce projet collaboratif et interdisciplinaire. L'équipe de linguistes est représentée par Alain Ghio, Ingénieur de Recherche au Laboratoire Parole et Langage (LPL, CNRS-AMU) à Aix-en-Provence. D'autres acteurs sont engagés dans cette collaboration tels que Danièle Robert, Praticien Hospitalier, Gilles Pouchoulin, Ingénieur de Recherche, Laurence Guisti, Orthophoniste et Muriel Lalain, Chargée de Recherche CNRS et orthophoniste. Cette équipe travaille en collaboration avec Corine Astésano, Prosodiste et Enseignant Chercheur à l'unité de recherche Octogone-Lordat de l'Université de Toulouse. De plus, deux équipes spécialistes du traitement automatique des langues coopèrent à ce projet. L'Institut de Recherche Informatique de Toulouse (IRIT) à Université Paul Sabatier, représenté par Jérôme Farinas et Julie Mauclair Enseignants Chercheurs, et le Laboratoire d'Informatique d'Avignon (LIA), de l'Université d'Avignon et des pays du Vaucluse, représenté par Corinne Fredouille Enseignant Chercheur et Habilitée à Diriger des Recherches (HDR). Dans le cadre d'un projet interdisciplinaire une collaboration entre tous les membres des équipes est nécessaire.

La première année de ce travail a concerné l'élaboration des corpus et des tâches expérimentales, ainsi que le recueil des données des patients par la passation des tâches. A la suite, un corpus de parole a été constitué pour ce projet et incluait des enregistrements de 85 patients traités pour un cancer de la cavité buccale et/ou de l'oropharynx de taille suffisante pour présenter des troubles de l'articulation de la parole et de 41 sujets contrôles, c'est à dire ne présentant aucun TPP, et appariés en âge et sexe. Les sujets sont exposés à différentes expériences de courte durée. Ils effectuent une tâche de prosodie, et deux tâches de lecture (lecture de phrases et lecture d'un extrait de *La Chèvre de Monsieur Seguin*), mais également des tâches de parole spontanée et de description d'images. Pour ce projet, le Laboratoire Parole et Langage a notamment développé un nouvel outil d'évaluation de l'intelligibilité : la tâche de décodage acoustico-phonétique, ou encore DAP (Ghio et al., 2018). Cette tâche repose sur la production de pseudo-mots, c'est-à-dire des mots dénués de sens mais respectant les règles phonotactiques du français (ex : stoumi, clorton, fiman...). Nous la décrivons ultérieurement.

Le projet RUGBI fait suite au projet C2SI. Débuté en 2018, il est actuellement en cours et a obtenu un financement de l'Agence Nationale de la Recherche (ANR). Il a pour objectif la recherche d'unités linguistiques pertinentes pour améliorer la mesure de l'intelligibilité de la parole altérée par des troubles de production pathologique, pour aboutir à une modélisation automatique du traitement de la parole, utilisable par le clinicien. En d'autres mots, suite au projet C2SI, le projet RUGBI tente d'identifier chez ces patients les segments altérés, ou plus généralement les macro-classes concernées, afin de dégager ceux ou celles qu'il serait pertinent d'évaluer, pour créer une méthode d'évaluation encore plus ciblée et efficace de l'intelligibilité.



Notre travail se base donc sur le corpus de parole du projet C2SI et s'intègre parfaitement dans le cadre du projet RUGBI.

## B. La tâche de Décodage Acoutisco-Phonétique (DAP)

L'objectif de création de la tâche de DAP était de supprimer les limites rencontrées dans les tests classiques d'évaluation de l'intelligibilité, qui mènent à une surévaluation de celle-ci (Lalain et al., 2020). Chaque auditeur est en effet capable de restaurer des séquences distordues lorsqu'il est confronté à un énoncé altéré, grâce aux processus de haut niveau (top-down), c'est-à-dire aux informations non dépendantes du signal de parole mais plutôt au contexte (thème de la discussion par exemple) et aux connaissances de l'auditeur (langue française par exemple). Lorsque ce dernier perçoit un discours altéré, ces processus vont agir comme des mécanismes de compensation afin d'optimiser au maximum la compréhensibilité du message. On retrouve notamment l'effet de la lexicalité, c'est-à-dire le fait qu'une séquence sonore ou orthographique fasse référence à un mot de notre vocabulaire (Ganong, 1980) : un exemple bien connu est celui de la distinction parfois difficile entre /t/ et /d/. Devant la séquence [aʃ], l'auditeur percevra plutôt un /t/ (puisque'il possède dans son lexique le mot « tâche »), alors qu'il le percevra /d/ devant la séquence [isk] (en référence au mot « disque ») (Ghio et al., 2018). La fréquence lexicale intervient aussi puisque les mots fréquents sont plus facilement reconnus. L'auditeur est donc capable d'effectuer une restauration phonémique lorsqu'un énoncé est dégradé. La plupart des tests utilisés pour évaluer l'intelligibilité utilisant des mots connus de tous, ces mécanismes de compensation sont donc rencontrés et altèrent l'évaluation. L'autre problème rencontré est celui de l'effet de mémorisation (Rebourg et al., 2020) : les listes de mots utilisés pour les tests étant courtes, les cliniciens habitués à les utiliser quotidiennement viennent à les connaître presque par cœur, si bien qu'ils peuvent restaurer les séquences altérées par mémorisation des items. Pour lutter contre cela, il a été proposé l'utilisation de pseudo-mots. Ce sont des items respectant les règles phonotactiques du français mais dépourvus de sens (Lalain et al., 2020). Ceux-ci ont été générés au cours du projet C2SI à l'aide d'un module de la base de données Lexique. Une multitude de pseudo-mots a été obtenue, par exemple *stoumo*, *vurtan*, *stolain*, *blanga*, *staji* ... Les auteurs ont obtenu un total de 89 346 pseudo-mots. Suite à un tirage au sort plusieurs listes de 52 pseudo-mots ont été générées, toutes différentes et phonétiquement équilibrées. C'est par ce grand nombre de pseudo-mots et de combinaisons possibles que la neutralisation de l'effet d'apprentissage par l'auditeur est faisable. L'effet de restauration lexicale est quant à lui aboli par l'utilisation de pseudo-mots (Rebourg et al., 2020). Ce test a été utilisé dans les projets C2SI et RUGBI pour évaluer l'intelligibilité. Nous allons voir qu'il est constitué de trois étapes différentes :

1. La phase de production : elle consiste à enregistrer les locuteurs sur la production des listes de 52 pseudo-mots. Au cours d'une de leurs consultations de suivi à l'Institut Universitaire du Cancer à Toulouse, les patients sont installés dans une cabine anéchoïque munie d'un écran, d'un microphone relié à un enregistreur digital et d'un filtre anti-pop. Puis 52 pseudo-mots leur sont présentés successivement. Chaque stimulus (chaque pseudo-mot) est présenté selon une double modalité : visuelle (version orthographiée du pseudo-mot) et auditive (version orale). Chaque locuteur doit alors prononcer ces 52 pseudo-mots et est enregistré.
2. La phase de perception : Une fois les enregistrements effectués, il s'agit ensuite de les faire écouter aux auditeurs afin qu'ils effectuent une transcription orthographique (manuelle) des items perçus. Ces derniers sont français natifs, ne doivent pas présenter de troubles auditifs ou visuels et avoir une bonne maîtrise de l'orthographe. Une étape préalable consiste à segmenter les enregistrements des patients afin de créer un fichier audio par pseudo-mot. Ainsi, un stimulus sonore leur est diffusé, et ils doivent écrire sur le clavier ce qu'ils entendent. Chaque auditeur est muni d'un casque audio et placé devant un écran. Enfin, à l'issue du test un fichier est généré, comprenant les cibles attendues (qui devaient être prononcées par les locuteurs) et leurs transcriptions phonétiques, les réponses orthographiées par les auditeurs et leurs versions phonétiques. Ce sont ces deux transcriptions qui sont ensuite comparées pour le calcul du score.
3. Le calcul du score d'intelligibilité : Une fois les transcriptions recueillies, un certain nombre d'opérations de traitement des données ont été effectuées afin de calculer, pour chaque locuteur, un score d'intelligibilité DPP (PPD) ou score de Déviation Phonologique Perçue. Ce score est calculé sur la base d'une matrice de coût créée par l'équipe du LPL au cours du projet C2SI permettant de mesurer la distance qui existe entre chaque pseudo-mot attendu et le pseudo-mot retranscrit par l'auditeur. Ghio et collègues (Ghio et al., 2018 ; Lalain et al., 2020) ont en effet établi une matrice de coût entre phonèmes, c'est-à-dire un tableau représentant le degré de dissimilitude entre eux. Chaque phonème peut être décomposé en un certain nombre de traits qui permet de les distinguer. En suivant et en complétant la décomposition en traits des voyelles et consonnes par Chomsky et Halle (1968), Ghio a établi de manière binaire qu'un trait était présent ou absent (1 ou 0 respectivement). Il a obtenu ainsi des matrices pour les voyelles et les consonnes, où apparaît le nombre de traits d'écart entre chaque phonème. Grâce à l'utilisation de ces matrices de coût et à un algorithme de Wagner Fisher, nous obtenons une distance locale

pour chaque couple phonème attendu – phonème transcrit, et une distance cumulée pour le couple pseudo-mot attendu – pseudo-mot transcrit.

Référence =>	r	ε	s	t	o	r	ã
r	0	8	9	6	8	0	8
œ	8	1	14	14	2	8	3
s	9	14	0	3	14	9	14
o	8	3	14	14	0	8	3
voc	0	0	5	5	0	0	0
a	8	1	14	14	2	8	3
n	2	8	9	6	8	2	3

distance locale → distance cumulée → =10

Figure 5 : Exemple de calcul de distance entre un mot cible [rɛstorã] et un mot retranscrit [roeso voc an]. La distance locale entre les phonèmes /r/ et /r/ est de 0 trait. Elle est de 1 trait d'écart entre /ε/ et /œ/. La distance cumulée entre la cible et le transcrit est de 10 traits d'écart.

Nous appelons score PPD (pour Perceived Phonologic Deviation, ou Déviation Phonologique Perçue), la distance cumulée entre tous ces phonèmes divisée par le nombre de phonème constituant le pseudo-mot, il s'exprime en nombre de traits altérés par phonème. On considère que l'intelligibilité est altérée à partir du score PPD seuil de 0.6 trait altéré par phonème (Ghio et al., 2020). Le score global d'un patient (ou score PPD global) est obtenu en calculant la moyenne des scores de chacune de ses 52 productions. Ce travail a donc reposé sur deux types de scores : les scores PPD globaux, correspondant à un score général pour chaque patient et étant la moyenne des scores de ses 52 pseudo-mots produits, et les scores PPD locaux reflétant la distance existante entre chaque phonème produit et le phonème transcrit correspondant, permettant une analyse plus précise de l'altération par l'identification des phonèmes les plus (ou les moins) altérés.

## V. Objectif et hypothèses

La problématique de ce travail est double. Nous cherchons d'une part, à vérifier si l'évaluation de l'intelligibilité basée sur la perception de séquences non lexicales, extraites d'un texte lu, est équivalente à une évaluation de décodage acoustico-phonétique de pseudo-mots, tout en dégagant les segments qui sont les plus altérés. En la menant à bien, cette étude contribuera au projet RUGBI en apportant des éléments de réponse quant aux tâches et unités pertinentes pour l'évaluation de l'intelligibilité.

Nous sommes confrontés à un manque d'outil fiable et objectif pour l'évaluation de l'intelligibilité, validé de manière consensuelle dans la communauté scientifique et utilisable de manière routinière par les cliniciens. Avec la création de la tâche de DAP, le projet C2SI tente de pallier ce problème. Ce travail de mémoire va utiliser les comparaisons avec cette tâche dans un cadre expérimental, et participera à l'évaluation de sa pertinence, s'intégrant alors dans la démarche globale du projet RUGBI. L'objectif principal de ce travail sera de mettre en lien les données du test de jugement perceptif du DAP et celles de notre test de jugement perceptif basé sur des séquences délexicalisées extraites de la lecture du texte de *La chèvre de Monsieur Seguin*. Cela afin de déterminer si un auditeur naïf peut devenir expert avec cette tâche d'identification de séquences VCV, de manière globale (grâce aux scores globaux). C'est en répondant à cette question principale que l'on pourra évaluer la pertinence de la tâche des séquences VCV extraites d'un texte lu comme une tâche adaptée à la clinique. Nous soulèverons diverses questions : quelles corrélations pouvons-nous faire avec les évaluations des autres tâches conduites dans le cadre du projet C2SI ? Des segments de parole sont-ils plus altérés que d'autres ? Si oui, sont-ils équivalents avec ceux de la tâche de DAP de pseudo-mots ?

La mise en relation des scores d'intelligibilité avec les autres données précédemment recueillies lors des projets C2SI et RUGBI devrait permettre de prédire la corrélation entre notre tâche de séquences non lexicales d'un texte lu et la tâche de décodage acoustico-phonétique de pseudo-mots. L'analyse plus ciblée des segments de parole devrait quant à elle permettre de confirmer ou d'infirmer nos intuitions et d'affiner le protocole de prise en charge orthophonique.

## VI. Matériel et méthodes

Dans ce travail, nous nous appuyons sur les données du corpus C2SI (Woisard et al., 2020), et plus précisément sur les données de la tâche de décodage acoustico-phonétique en pseudo-mots et sur les enregistrements audio d'une tâche de lecture du texte d'Alphonse Daudet « La chèvre de Monsieur Seguin ». les métadonnées des patients (âge, sexe, localisation de la tumeur). Avant de décrire le traitement et l'analyse de ces données, il est important d'expliquer comment celles-ci ont été recueillies. Enfin, nous détaillerons l'élaboration de notre test de jugement perceptif à partir de séquences non lexicales extraites de la lecture.

### A. Corpus

Le corpus du projet C2SI (Woisard et al., 2020) est constitué au total des enregistrements de 126 locuteurs : 85 patients avec un cancer de la cavité buccale ou de l'oropharynx et 41 sujets sains.

Ils ont été évalués sur plusieurs tâches de production, comme de la lecture, de la description d'images, de la parole spontanée, la tâche de DAP mais aussi des tâches de prosodie. Pour l'inclusion, les patients devaient avoir été traités à l'Institut Universitaire du Cancer à Toulouse pour un cancer de la cavité buccale ou de l'oropharynx, toutes tailles confondues (T1 à T4). Le traitement pouvait être une chirurgie et/ou une radiothérapie et/ou une chimiothérapie. Le délai post-traitement devait être de 6 mois minimum, et le patient devait être en rémission clinique afin que la parole soit la plus stable possible. Les critères d'exclusion étaient une autre atteinte responsable de trouble de la parole (atteinte neurologique type accident vasculaire cérébral, ou encore bégaiement), et des troubles visuels, auditifs ou cognitifs incompatibles avec le bon déroulement de la tâche. D'autres données étaient recueillies : âge, sexe, région atteinte par le cancer (cavité buccale ou oropharynx), localisation précise (langue, plancher buccal, trigone rétromolaire, mandibule pour la cavité buccale ; amygdale, base de langue, voile du palais, « oropharynx » si plusieurs de ces localisations étaient atteintes pour l'oropharynx), stades T et N de la tumeur, délai post-traitement, traitement reçu, type de chirurgie, type de lambeau s'il y avait eu reconstruction, dose de radiothérapie, récurrence ou pas. Tous les patients ont été inclus et enregistrés dans le service d'oncoréhabilitation de l'Oncopole à Toulouse. Les sujets contrôles quant à eux ont été recrutés en respectant les critères d'exclusion valables pour les patients, et devaient avoir un bon niveau en lecture. Ils ont été enregistrés à Toulouse, dans le Vaucluse et au Laboratoire Parole et Langage.

Dans notre travail, nous avons utilisé les enregistrements de parole à partir de la lecture du texte de *La chèvre de Monsieur Seguin* et les données du test perceptif du DAP de pseudo-mots, avec les mêmes patients et sujets contrôles. Au total, la population du corpus était constituée de 85 patients et 41 sujets contrôles : 65 étaient des hommes (51.6%) et 61 des femmes (48.4%). Nous avons 46 hommes et 39 femmes parmi les patients, et 19 hommes et 22 femmes parmi les sujets contrôles. L'âge moyen était de 64.2 ans (écart type 10.7 ans). Parmi les patients, 33 avaient été traités pour un cancer de la cavité buccale, 52 pour un cancer de l'oropharynx.

## B. Evaluer l'intelligibilité avec la tâche de séquences délexicalisées

Pour notre test de perception, notre objectif était de conduire une évaluation de l'intelligibilité à partir de séquences non lexicales extraites d'un corpus de lecture. Le premier travail a consisté en la sélection de ces segments non lexicaux. Il nous a donc fallu élaborer un inventaire des voyelles et consonnes inscrites dans le texte de *La chèvre de Monsieur Seguin* pour pouvoir établir ces labels. Nous avons décidé de sélectionner seulement les consonnes en position intervocalique,

c'est-à-dire entre deux voyelles, car pour pouvoir percevoir une consonne, elle ne peut être isolée. C'est un segment bref, nous avons besoin des segments adjacents, ici nos voyelles.

Ensuite, nous avons réalisé un tableau Excel à partir de l'inventaire pour séquencer le texte et faire un regroupement en archiphonèmes, rassembler, par exemple, les consonnes situées sur un mot de contenu ou un déterminant et les voyelles dont l'accent régional peut varier. Pour ces dernières, cela permet de neutraliser cet accent, par exemple : « un brin d'herbe » et « un ours brun » se prononcent de différentes façons en fonction de la région de naissance du locuteur. Nous avons ainsi regroupé tel que  $\mu$  pour [ẽ /œ̃], Ê pour [e/ɛ] et Û pour [œ/ø/ə]. Nous avons récolté trois archiphonèmes Ê, Û et L. Il nous a donc fallu faire un tri dans notre inventaire avec un choix d'équilibre. Nous avons sélectionné les [e/ɛ] étant indispensables et laissé de côté ceux dont la prononciation est incertaine selon l'accent régional. Les « e » muets ont été quant à eux abandonnés du fait de l'incertitude de prononciation entre les interlocuteurs. Pour ce qui est du L, le problème se posait au niveau des déterminants (le, la, les) et des mots de contenu (loup, liberté, voulant...). Nous avons donc équilibré la sélection en prenant autant de mots porteurs de sens que de déterminants. A la fin, nous avons constitué un inventaire composé de 50 consonnes et 50 voyelles.

Ortho	Phono	Conso	Voye		VCV	Count	LabelVCV	Mot cible	VoyCount	LabelVoy	
Mon	mm eu						-			-	
sieur	ss yy eu		eu				-		1	eu-01	Monsieur
Se	ss eu	ss			euSSeu	1	euSSeu-01	Seguin		-	
guin	gg in	gg	in		euGGin	2	euGGin-02	Seguin	2	in-02	Seguin
n'a	nn aa	nn			inNNaa	3	inNNaa-03	n'avait		-	
vaît	vv ai	vv			aaVVai	4	aaVVai-04	n'avait		-	
ja	jj aa	jj			aiJJaa	5	aiJJaa-05	jamais		-	
mais	mm ei	mm			aaMMei	6	aaMMei-06	jamais		-	
(z) eu	zz uu	zz	uu		eiZZuu	7	eiZZuu-07	jamais	3	uu-03	eu
de	dd eu	dd			uuDDeu	8	uuDDeu-08	de		-	
bo	bb oo	bb	oo		euBBoo	9	euBBoo-09	bonheur	4	oo-04	bonheur
nheur	nn oe rr	nn	oe		ooNNoe	10	ooNNoe-10	bonheur	5	oe-05	bonheur
a	aa		aa				-		6	aa-06	avec
vec	vv ai kk	vv			aaVVai	11	aaVVai-11	avec		-	
ses	ss ai		ai				-		7	ai-07	ses
chè	ch ai	ch			aiCHai	12	aiCHai-12	chèvres		-	
vres.	vv rr						-			-	

Figure 6 : Tableau de données de la transcription orthographique et phonologique du texte, l'inventaire des consonnes et voyelles, ainsi que les segments et leur numérotation nécessaires pour l'extraction des stimuli

Puis, nous nous sommes penchés sur la réalisation pratique et informatique de notre expérience. Il nous a paru nécessaire de réaliser deux tableaux croisés dynamiques pour mieux visualiser nos consonnes et voyelles sélectionnées en parallèle du séquençage du texte. Nous voulons tester le maximum de segments à l'aide de deux manipulations : un premier bloc pour les consonnes et le deuxième pour les voyelles. Pour tester l'ensemble des sujets du corpus C2SI regroupant 85 patients et 41 sujets contrôles, nous avons dû les diviser en deux lots. Pour les constituer, nous

avons classé les locuteurs en fonction de leur score de sévérité clinique puis sélectionné un locuteur sur deux. La réalisation de notre test de perception est effectuée avec une première session de passation avec le Lot1 seulement, il sera nécessaire d'en effectuer une deuxième ultérieurement pour le Lot2. Ce Lot1 est établi avec 60 locuteurs : 20 sujets sains (contrôles) et 40 patients. Chaque locuteur aura produit 50 segments (consonnes et voyelles), ce qui nous donnera 3000 stimuli. Ces stimuli seront écoutés par 3 auditeurs différents, nous obtiendrons alors 9000 écoutes pour chacun des tests.

Il a ensuite été nécessaire d'effectuer les étiquettes de nos stimuli dans le but final de les faire écouter à des auditeurs. Il faut tout d'abord, segmenter les enregistrements des sujets afin de créer un fichier audio par segment. Pour cela, une transcription orthographique de chaque fichier son a été nécessaire, suivie d'une transcription phonétique (cf. figure 7). Cette manipulation a été effectuée par alignement automatique à partir d'indices acoustiques (*LIA-Laboratoire Informatique Avignon*), puis introduit sous le logiciel PRAAT.

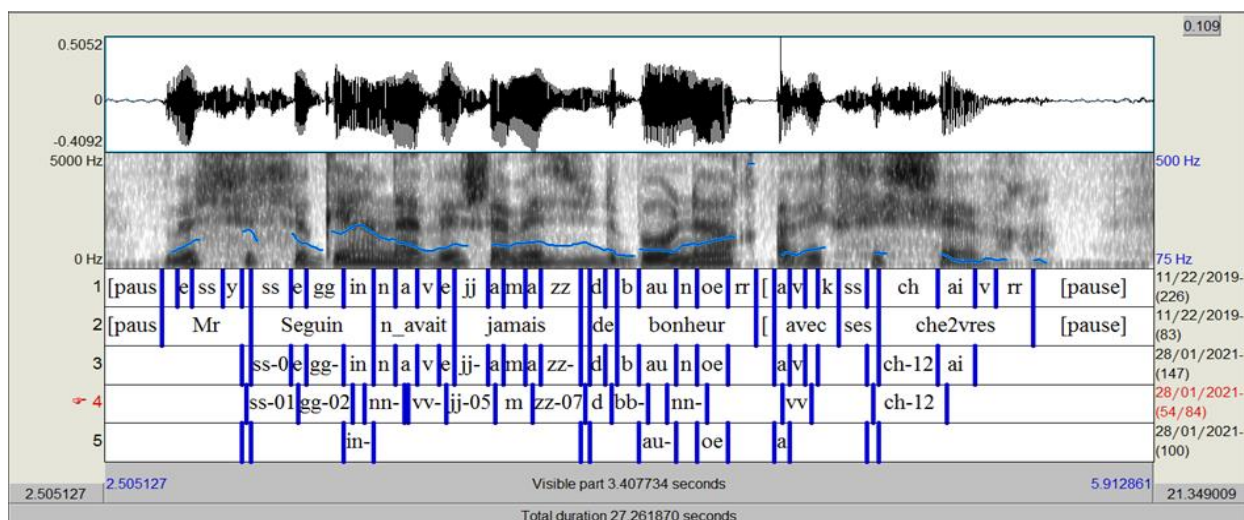


Figure 7 : Capture d'écran d'un fichier audio avec segmentation dans le logiciel PRAAT

Nous avons ensuite extrait nos labels consonantiques suivant un premier model V-C-V, c'est-à-dire en ne conservant que la moitié de la première voyelle, la consonne dans sa globalité et la moitié de la deuxième voyelle. Par exemple (cf. figure 8), pour « jamais », l'item [aaMMai] est extrait. Dans un second temps, ce sont les labels vocaliques qui ont été créés selon le model V simple, conservant la voyelle dans son intégralité. Par exemple pour « Seguin », l'item [in] est prélevé.

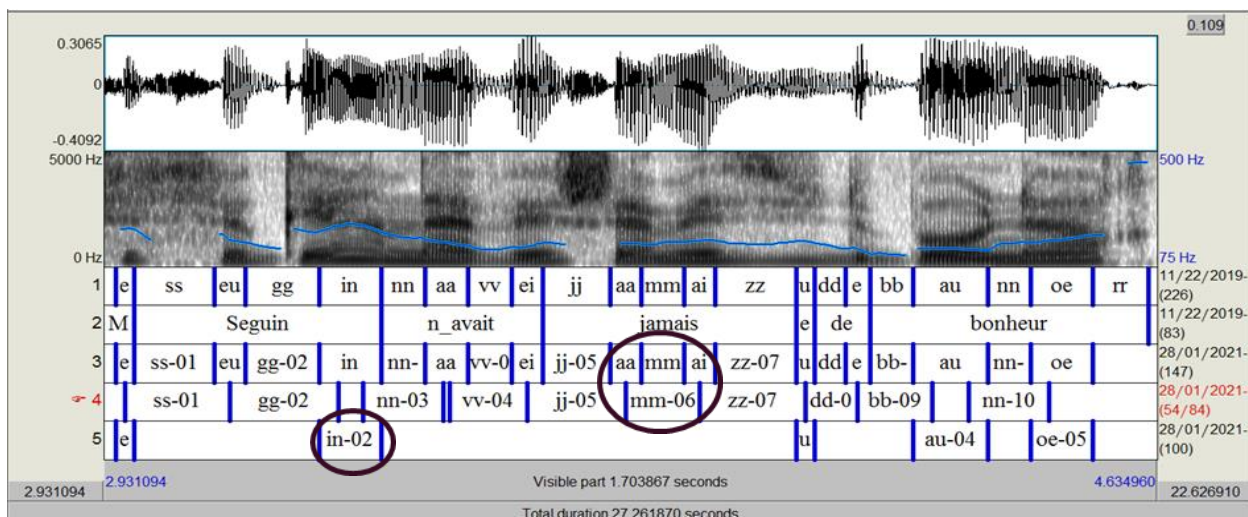


Figure 8 : Zoom capture d'écran d'un fichier audio dans le logiciel PRAAT avec sélection de segments

Une fois nos étiquettes constituées, nous avons procédé à une vérification manuelle. L'analyse de cette notation semi-automatique nous a permis de nous rendre compte de certaines erreurs commises par l'aligneur, par exemple :

- la cible n'était pas détectée lorsqu'un locuteur (atteint) réalisait une pause
- la cible pouvait être décalée lorsqu'une liaison n'était pas réalisée, ou parce que le locuteur commettait une erreur de lecture.

Ainsi, nous avons recueilli un pourcentage d'erreurs sur nos consonnes et voyelles, pouvant aller de 36 à 100% pour certaines. Nous nous sommes rendu compte qu'un trop grand nombre d'étiquettes n'étaient pas détectées du fait de nombreuses pauses exécutées par les locuteurs. Nous avons donc décidé de ne pas en tenir compte et de relancer un étiquetage en ajoutant le model #-C-V (Pause-Consonne-moitié de Voyelle). Par ailleurs, il nous a également fallu remplacer deux étiquettes par un autre segment, du fait de la réalisation non systématique de liaisons. Par exemple pour « s'en allait », le label [anNNaa] a été substitué par « ni les » avec l'item [#NNii].

Ensuite, à partir de ces étiquettes, nous avons récupéré nos stimuli pour qu'ils soient écoutés par des auditeurs afin qu'ils effectuent un choix de transcription orthographique des items perçus. Les stimuli sonores obtenus sont tous randomisés, si bien que chaque auditeur va écouter ces sons brefs produits par plusieurs locuteurs différents. De plus, chaque segment produit par un locuteur est écouté par 3 auditeurs différents.

Le recrutement a été spécifié pour dix-huit participants naïfs francophones natifs de 18 à 60 ans, ne présentant aucun trouble auditif ni visuel. Le test de jugement perceptif, qui s'est déroulé au Centre d'Expérimentation de la Parole (*Fiche\_instrument\_TestsPercept\_2016.pdf (lpl-aix.fr)*) au Laboratoire Parole et Langage à Aix-en-Provence, est réalisé grâce à la plateforme Perceval et son



application Lancelot. Les auditeurs sont munis d'un casque audio, d'un écran et d'une souris d'ordinateur.



Figure 9 : Exemple de déroulement du test de jugement perceptif

La consigne suivante leur est donnée : « Vous allez entendre des extraits de parole assez courts (ex : ati). Vous devrez sélectionner la consonne que vous aurez entendue. Certaines prononciations seront difficiles à identifier mais dans tous les cas, vous devrez proposer une réponse ». Le stimulus sonore leur est diffusé, et ils doivent cliquer sur la consonne qu'ils entendent. Ils peuvent écouter chaque stimulus au maximum 3 fois. Cela représente la première partie du test, dotée de 480 stimuli consonantiques pour la plupart des auditeurs, sauf 455 pour trois d'entre eux, les blocs des stimuli étant plus courts. La deuxième est élaborée de la même manière mais le participant doit choisir la voyelle correspondante cette fois-ci. 501 écoutes sont attendues pour chaque auditeur pour ces labels vocaliques.

Enfin, à l'issue du test un fichier est généré par Perceval sous format Excel (cf. figure 12), comprenant les cibles attendues (prononcées par les locuteurs) et les réponses sélectionnées par les auditeurs. Ces résultats seront par la suite analysés statistiquement.


Sélectionnez la consonne	
	
<input type="radio"/> g (gong)	<input type="radio"/> p (pont)
<input type="radio"/> d (don)	<input type="radio"/> t (ton)
<input type="radio"/> b (bon)	<input type="radio"/> k (con)
<input type="radio"/> m (mon)	<input type="radio"/> r (rond)
<input type="radio"/> n (non)	<input type="radio"/> l (long)
<input type="radio"/> v (vont)	<input type="radio"/> ch (chant)
<input type="radio"/> z (zone)	<input type="radio"/> s (son)
<input type="radio"/> j (jonc)	<input type="radio"/> f (font)
<input type="button" value="Valider"/>	

Figure 10 : Capture d'écran de la tâche VCV de passation du test de perception

Sélectionnez la voyelle

a (anna)  
 eu (heure)  
 é (été)  
 i (iris)  
 au (homme)  
 u (uramus)

an (ange)  
 oe (eux)  
 ê (être)  
 in (inde)  
 o (haut)  
 ou (outre)

on (honte)

Figure 11 : Capture d'écran de la tâche V de passation du test de perception

## VII. Résultats et analyses

Le but de notre travail était de démontrer que la tâche à partir de séquences délexicalisées extraites de la lecture est équivalente à la tâche de décodage acoustico-phonétique (DAP). Premièrement en analysant les scores PPD globaux pour chaque groupe de locuteurs, puis pour chaque locuteur, et enfin en comparant notre tâche VCV avec les autres évaluations du projets C2SI : la tâche de DAP pseudo-mots, la sévérité (en lecture et description d'images), l'intelligibilité subjective (en lecture et description d'images). Enfin, nous avons observé l'altération phonétique, tout en dégagant les phonèmes les plus altérés. Cela est décrit pour la tâche VCV et la tâche V.

### A. Traitement des données Voyelle Consonne Voyelle (VCV)

Avant de procéder à l'analyse des résultats, un pré-traitement est nécessaire pour les données VCV. Une concaténation de l'ensemble des données en un seul fichier Excel a été réalisée. Nous avons en effet recueilli 54 fichiers réponses pour la tâche VCV (18 auditeurs x 3 blocs).

script	listener	block	date	Lot1	trial_id	stim_id	wave	speaker	group_speaker	id_vcv	rank_vcv	duration	expected	response	rtime	compare	PPD
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	141	141 TIO-000020-(TIO-000020	CTRL		bb-17	17	135	bb	bb	2583	b b	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	74	74 TIO-000017-(TIO-000017	CTRL		nn-36	36	100	nn	nn	2606	n n	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	85	85 TIO-000017-(TIO-000017	CTRL		tt-19	19	150	tt	tt	4049	t t	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	133	133 TIO-000019-(TIO-000019	CTRL		tt-39	39	205	tt	dd	5516	t d	1
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	153	153 TIO-000020-(TIO-000020	CTRL		ll-23	23	95	ll	nn	2910	l n	1
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	4	4 TIO-000016-(TIO-000016	CTRL		ch-12	12	200	ch	ch	2521	S S	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	86	86 TIO-000017-(TIO-000017	CTRL		tt-27	27	140	tt	dd	2923	t d	1
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	116	116 TIO-000019-(TIO-000019	CTRL		nn-03	3	180	nn	nn	2493	n n	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	118	118 TIO-000019-(TIO-000019	CTRL		nn-32	32	250	nn	nn	6426	n n	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	7	7 TIO-000016-(TIO-000016	CTRL		dd-24	24	95	dd	nn	2915	d n	2
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	19	19 TIO-000016-(TIO-000016	CTRL		ll-49	49	115	ll	ll	4499	l l	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	10	10 TIO-000016-(TIO-000016	CTRL		gg-02	2	130	gg	vv	3362	g v	3
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	157	157 TIO-000020-(TIO-000020	CTRL		mm-14	14	145	mm	mm	2573	m m	0
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	63	63 TIO-000017-(TIO-000017	CTRL		ll-25	25	90	ll	mm	5360	l m	2
Rugbi-Lec-C2SI-Lot1-VCV	AA	1	2021-3-27	1	Lot1	34	34 TIO-000016-(TIO-000016	CTRL		rr-34	34	120	rr	rr	2879	R R	0

Figure 12 : Tableau de données Rugby-Lecture-C2SI-VCV\_Total

Dans ce tableau général de données sont inscrits : le code alphabétique pour l'anonymisation de l'auditeur, le numéro du bloc réalisé, la date de la passation, l'identification du stimulus, la référence du locuteur qui a produit ce stimulus, la qualité du locuteur (sujet contrôle ou patient), la consonne avec sa position dans le corpus, la durée du stimulus écouté. Ce qui nous intéresse

particulièrement réside dans les dernières colonnes qui nous renseignent sur la consonne qui a été produite par le locuteur (réponse attendue), et la réponse que l’auditeur a fourni (réponse donnée). C’est à partir de ces deux colonnes qu’est calculé le score PPD. Un score PPD est ainsi attribué pour chaque consonne en fonction du nombre de traits articulatoires qui les diffère, allant de 0 à 6.

En ce qui concerne l’évaluation perceptive des segments consonantiques (VCV), nous vérifions dans un premier temps la cohérence des données, en faisant notamment un comptage. Les étiquettes de lignes correspondent aux codes des auditeurs pour leur anonymisation.

Étiquettes de lignes	Nombre de listener	Nombre de PPD
AA	480	480
BB	480	480
CC	480	480
DD	480	480
EE	480	480
FF	455	436
GG	480	480
HH	480	471
II	480	478
JJ	480	480
KK	480	480
LL	455	455
MM	480	480
NN	480	479
OO	480	455
PP	480	479
QQ	480	480
RR	455	455
(vide)		
<b>Total général</b>	<b>8565</b>	<b>8508</b>

Figure 13 : Comptage du nombre d’écoutes et de scores PPD VCV pour chaque auditeur

Nous constatons que pour les 8565 écoutes, nous obtenons 8508 réponses pour l’ensemble des auditeurs. Nous avons donc un total de 57 réponses manquantes (ou non-réponses), ce qui correspond à 0,67 % des données. Ce faible pourcentage indique que la tâche s’est bien déroulée et que nous avons des données en nombre suffisant pour nos analyses.

En amont de notre analyse statistique, nous avons d’abord effectué une transformation de codage pour passer d’un codage à double consonnes (*LIA-Laboratoire Informatique Avignon*) vers un codage lexiphone à une seule consonne, pour permettre les calculs de scores PPD à partir de l’algorithme Wagner-Fischer. Cet algorithme permet d’obtenir une distance locale pour chaque couple phonème attendu – phonème perçu. Ce que nous appelons score PPD (pour Perceived Phonologic Deviation, ou Déviation Phonologique Perçue) est en fait le nombre de traits altérés entre deux phonèmes. Comme attendu (cf. VI. Matériel et méthodes), nous avons bien 480 écoutes pour la plupart des auditeurs, 455 pour trois d’entre eux, certains blocs de stimuli étant plus courts.

Group_speaker	id_vcv	rank_vcv	duration	expected	response	rtime	compare	PPD
CTRL	bb-17	17	135	bb	bb	2583	b b	0
CTRL	nn-36	36	100	nn	nn	2606	n n	0
CTRL	tt-39	39	205	tt	dd	5516	t d	1
PATIENT	mm-18	18	130	mm	mm	2663	m m	0
PATIENT	tt-48	48	215	tt	ss	3123	t s	1

Figure 14 : Tableau des Scores PPD par segment consonantique ligne 1 à 5

Après ces vérifications, et le reformatage des données, nous appliquons l’algorithme de calcul pour obtenir les scores de Déviation Phonologique Perçue ou score d’intelligibilité PPD. Nous obtenons pour chaque consonne produite par chacun des locuteurs contrôles et patients, un score PPD par segment qui correspond au nombre moyen de traits altérés entre la consonne cible (celle qui devait être produite) et la consonne perçue.

Si nous regardons la première ligne de ce tableau, nous pouvons observer que pour le /b/ numéro 17 produit par un sujet contrôle, c’est également un /b/ qui a été perçu par un auditeur, le score PPD est donc nul. A la 5<sup>e</sup> ligne, lorsqu’un patient produit le /t/ position 48, un auditeur a répondu par un /s/, la différence de trait articulatoire applique ainsi un score PPD de 1.

A partir de ces données brutes, nous avons calculé le score PPD moyen pour chaque groupe de locuteurs, puis pour chaque locuteur. Les résultats montrent un score moyen PPD de 0,19 traits d’écart pour le groupe contrôle (CTRL) et de 0,83 traits d’écart pour le groupe patient. La figure suivante montre ainsi que le score PPD des patients est plus élevé que celui des contrôles ce qui traduit comme attendu, une intelligibilité plus altérée pour ce groupe de locuteurs.

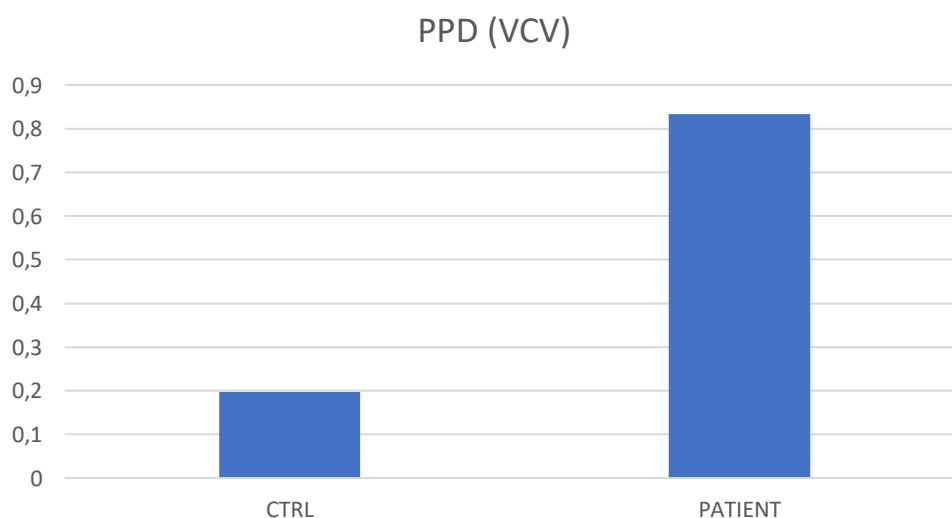


Figure 15 : Score PPD VCV pour chaque groupe de locuteurs

Nous avons ensuite calculé les scores par locuteur, pour chacun des deux groupes. Le graphique suivant présente les scores par locuteur et par groupe, ordonnés du score PPD le plus élevé au moins élevé.

Nous observons les 61 locuteurs sur l'axe des abscisses et le score PPD leur correspondant sur l'axe des ordonnées. Nos deux groupes de locuteurs sont identifiés en bleu pour les patients et en orange pour les sujets contrôles. Nous observons que les scores les plus hauts, c'est-à-dire les plus altérés, correspondent à des patients. La majorité des sujets contrôles ont un score PPD relativement faible, sans toutefois atteindre un score parfait de 0 qui correspondrait à un effet plancher.

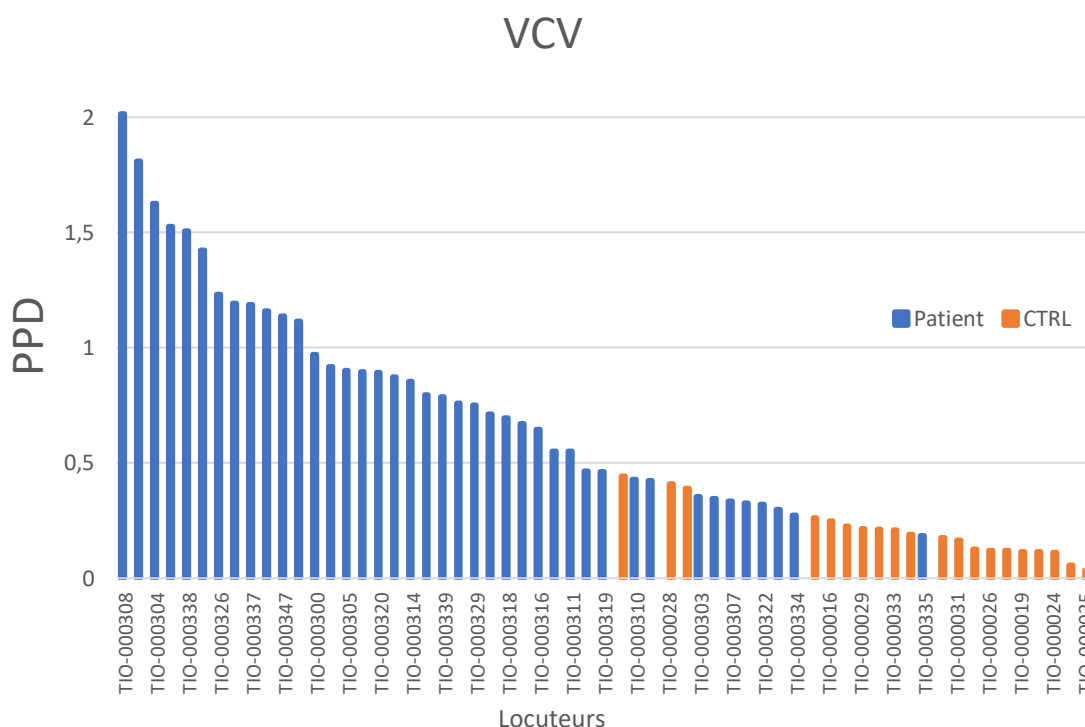


Figure 16 : Scores PPD VCV par groupe et par locuteur

Parallèlement, nous avons validé nos mesures par une méthode de détection utilisée en médecine. Le tracé de la courbe ROC (de l'anglais Receiver Operating Characteristic) permet d'indiquer la puissance de fiabilité d'une évaluation. Elle attribue à chaque patient un positif et à chaque sujet contrôle un négatif. Cela permet ainsi d'obtenir un tracé de faux-positif en fonction de faux-négatif. Plus cette courbe se rapproche de l'angle supérieur gauche, plus la validation des mesures est acceptée, avec 1 en tant qu'indicateur parfait. Notre courbe (fig.17) nous a révélé une aire sous la courbe de 0,96 (AUC pour Area Under the Curve), ce qui est concluant. Nous pouvons également ajouter que le seuil optimal se situe à 0,3 PPD (en rouge). Nous considérons que l'intelligibilité est altérée à partir de ce score PPD seuil de 0,27 trait altéré par phonème.

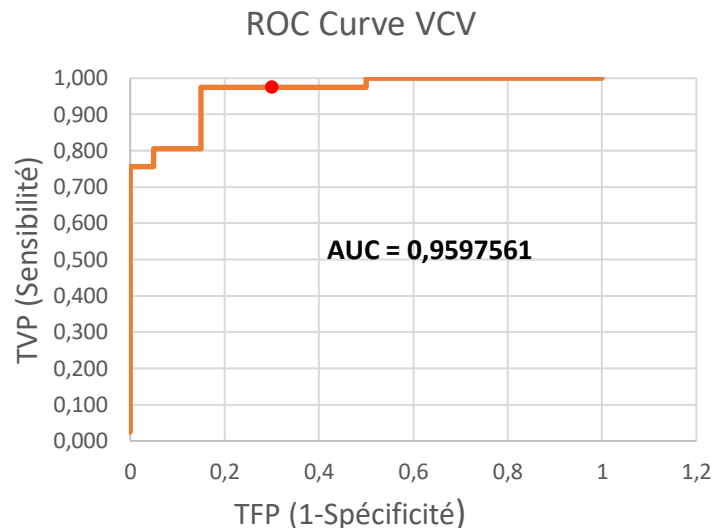


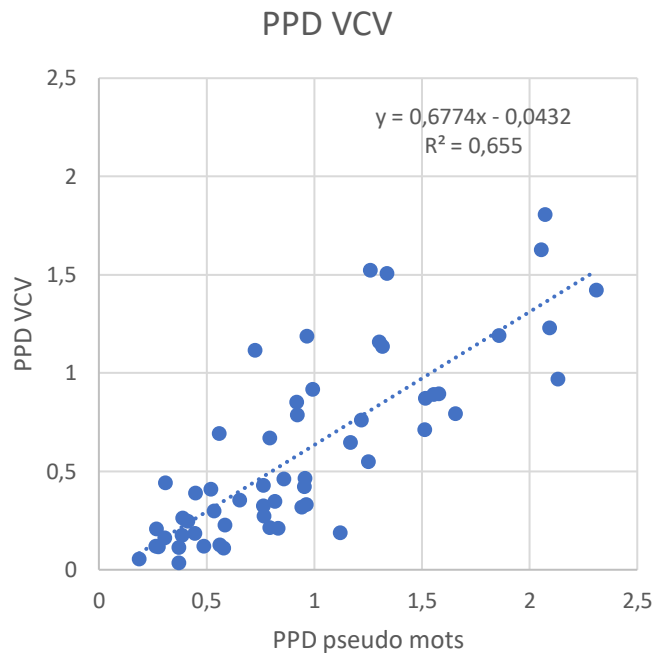
Figure 17 : Courbe ROC VCV

Nous allons maintenant nous attarder sur les résultats VCV par rapport aux autres évaluations du projet C2SI. Rappelons que le but de ce travail est de vérifier si l'évaluation de l'intelligibilité basée sur la perception de séquences non lexicales VCV, extraites d'un texte lu, est équivalente à l'évaluation de décodage acoustico-phonétique de pseudo-mots.

➤ PPD VCV et PPD Pseudo-mots

D'après les données recueillies à partir de l'évaluation de DAP de pseudo-mots, le score PPD des sujets sains est en moyenne à une distance de 0,48 trait par phonème alors que cette distance atteint 1,29 pour les patients.

Si nous comparons nos scores PPD VCV globaux avec les scores PPD Pseudo-mots globaux, nous observons que les scores sont moins élevés pour nos segments consonantiques VCV. En effet, nous retrouvons un score PPD de 0,19 pour nos sujets contrôles, en comparaison à 0,48 avec les pseudo-mots, et un PPD de 0,83 pour les sujets patients, comparé à 1,29 PPD pour l'autre tâche. Ces scores révèlent ainsi une meilleure identification des phonèmes perçus par les auditeurs avec notre tâche VCV. Nous avons ensuite mis en relation ces données PPD globaux de notre tâche avec celles de la tâche de pseudo-mots à l'aide d'un graphique.



*Figure 18 : Scores PPD VCV en fonction des scores PPD pseudo-mots*

Ce graphique nous montre la relation entre ces deux mesures de PPD. Chaque point représente un locuteur. Nous pouvons dégager de cette représentation qu'il existe une bonne corrélation entre les deux mesures : un patient dont l'intelligibilité est évaluée mauvaise avec notre tâche VCV, l'est aussi avec le DAP pseudo-mots.

Le coefficient de corrélation obtenu dans cette analyse est de  $R^2 = 0,66$ . Plus ce coefficient est proche de la valeur 1, plus les données sont liées. Il y a donc une relation entre notre mesure PPD VCV et celle de PPD de pseudo-mots.

Nous savons que l'intelligibilité est surtout évaluée à partir de tests subjectifs. Une partie du corpus a été utilisée dans deux tests subjectifs par l'équipe de Toulouse. Ces deux tests présentent des consignes et des formats de réponses différents, et évaluent indépendamment l'intelligibilité de la parole et la sévérité du trouble de production. Une tâche de lecture de texte et une tâche de description d'images ont été effectuées.

Nous rappelons que notre évaluation à partir de segments non lexicaux est extraite de la lecture. Nous cherchons ainsi à savoir s'il existe une corrélation avec d'autres évaluations.

➤ PPD VCV et Sévérité : lecture et description

Nous avons établi deux graphiques des scores PPD globaux en fonction de la sévérité du trouble dans la tâche de lecture d'une part, et dans la description d'images d'autre part.

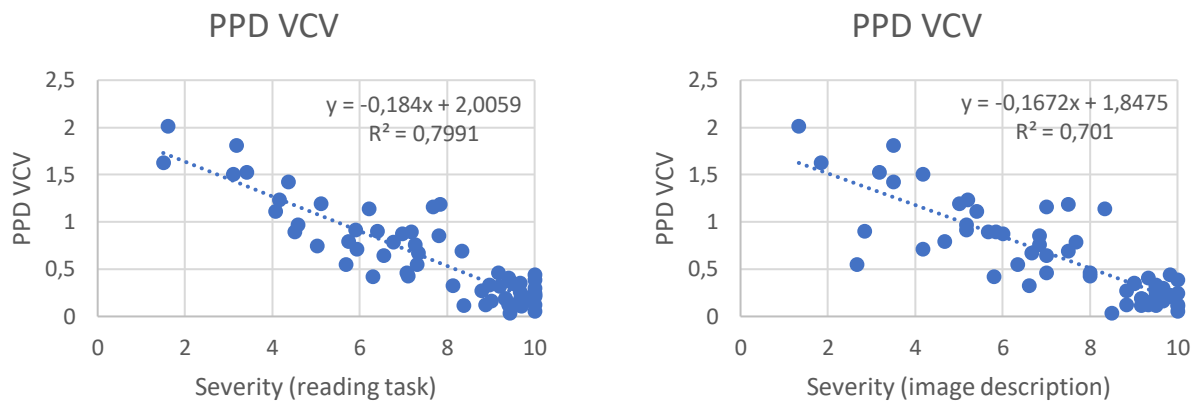


Figure 19 : Scores PPD VCV en fonction de la sévérite en lecture et en description d'images

Nous pouvons observer que les scores PPD VCV et la sévérite du trouble sont corrélés. Plus le score PPD VCV est élevé, plus l'intelligibilité est altérée. Nous observons que les locuteurs dont l'atteinte est jugée sévère (avec un score bas) sont ceux dont l'intelligibilité est jugée altérée (avec un score PPD haut). Ces scores sont d'autant plus corrélés dans la tâche de lecture. Nous le voyons par le nombre de points se rapprochant au plus près de la droite, tandis qu'en description, le nuage de points est dispersé, ce que traduisent les valeurs du coefficient. Nous pouvons dire que le score PPD VCV est corrélé avec le score de sévérite en lecture. Nous avons donc deux mesures différentes sur un même type de matériel linguistique : la lecture.

➤ PPD VCV et Intelligibilité subjective : lecture et description

Nous avons à nouveau élaboré deux graphiques pour comparer les scores PPD VCV, concernant cette fois l'intelligibilité subjective, en lecture et en description d'images. L'évaluation de l'intelligibilité subjective de la parole renvoie au taux d'unités linguistiques (syllabes et mots principalement) dont tous les phonèmes sont correctement identifiés par les auditeurs.

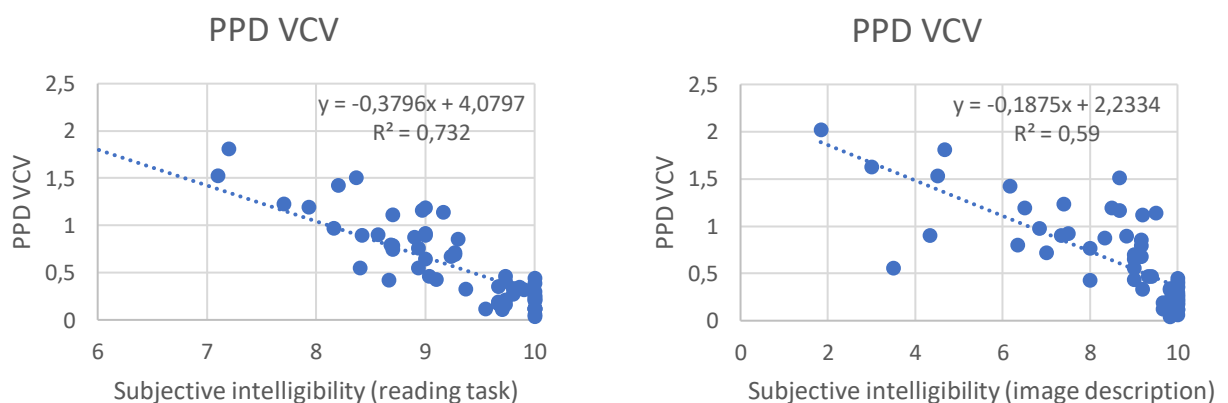


Figure 20 : Scores PPD VCV en fonction de l'intelligibilité subjective en lecture et description d'images



Nous pouvons voir que l'intelligibilité subjective est corrélée avec les scores PPD VCV, à la fois en lecture et en description. Cette corrélation est, encore une fois, d'autant plus forte avec la tâche de lecture. Nous obtenons un nuage de points se rapprochant de la droite pour la tâche de lecture, tandis que les points sont plus dispersés en description. Un locuteur évalué subjectivement avec une intelligibilité altérée (score bas), est également évalué avec un score PPD VCV élevé. Nous pouvons dire que le score PPD VCV est corrélé avec le score d'intelligibilité subjective en lecture.

Pour poursuivre, nous voulions analyser le lien entre notre tâche VCV et l'altération phonétique évaluée perceptivement à l'aide d'une échelle.

➤ PPD VCV et altération phonétique

Nous avons tracé ce graphique pour étudier les scores PPD VCV en fonction de l'altération phonétique.

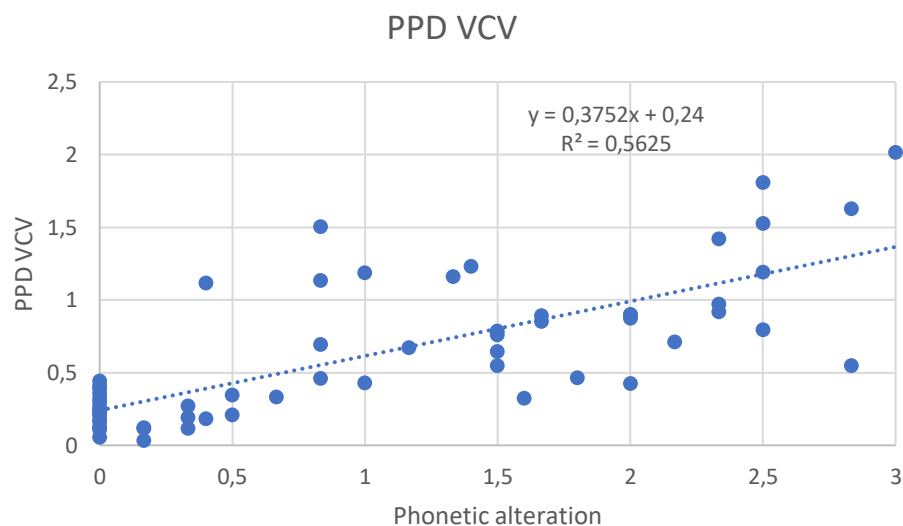


Figure 21 : Scores PPD VCV en fonction de l'altération phonétique

Nous pouvons constater que plus cette altération est élevée plus le score PPD VCV est élevé. Les nuages de points sont plutôt dispersés loin de la droite, nous pouvons ainsi dire que la corrélation existante entre ces deux tâches est plus faible, en comparaison avec les évaluations vues précédemment, comme en témoigne la valeur du coefficient de corrélation.

Nous pouvons classer de sorte que, la tâche VCV est très fortement corrélée avec la tâche de pseudo-mots, qu'elle est bien corrélée avec la tâche de lecture au niveau de la sévérité et de l'intelligibilité subjective, et moins en description d'images.

Puis, nous nous sommes intéressés plus précisément à cette altération phonétique. Nous avons cherché à discerner quels sont les phonèmes les plus dégradés.

➤ Identification des phonèmes les plus altérés (VCV)

Nous avons classé chaque consonne en fonction de leur score PPD respectif et de leur groupe de locuteurs, afin de réaliser un histogramme. Nous observons, pour ce classement des phonèmes (du plus au moins altéré dans le groupe des patients) que les phonèmes /g/, /k/, /d/ sont ceux pour lesquels les scores PPD sont les plus élevés.

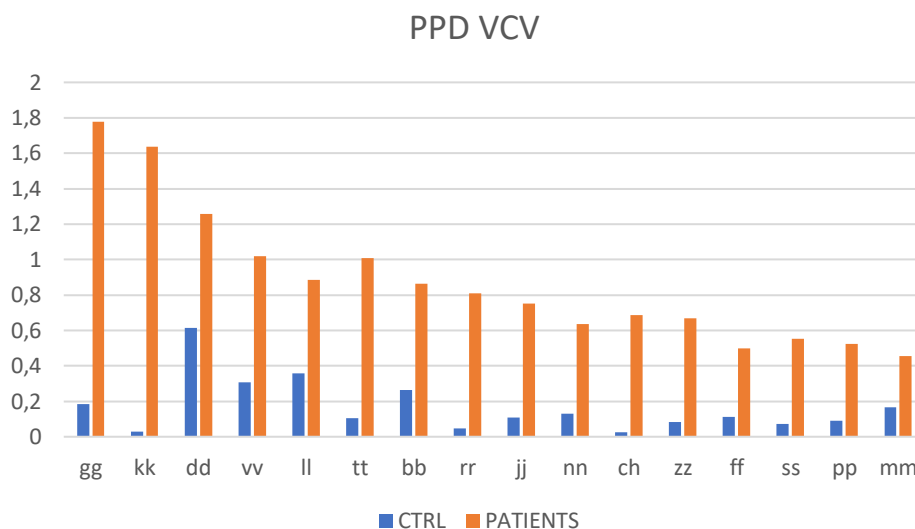


Figure 22 : Histogramme des scores PPD VCV en fonction des phonèmes altérés pour les deux groupes

Cependant, on peut noter que l'écart de score entre les deux groupes est variable en fonction des phonèmes considérés. Cela nous a amené à proposer un nouveau classement qui rende compte de l'écart de score PPD entre les patients et les sujets contrôles. Cette homogénéisation était nécessaire pour neutraliser cet important écart entre les deux groupes du fait de la pathologie (atteinte) qui entrave l'expression du patient. La pathologie étant un frein à son expression, il est évident que l'altération des phonèmes de celui-ci soit plus considérable.

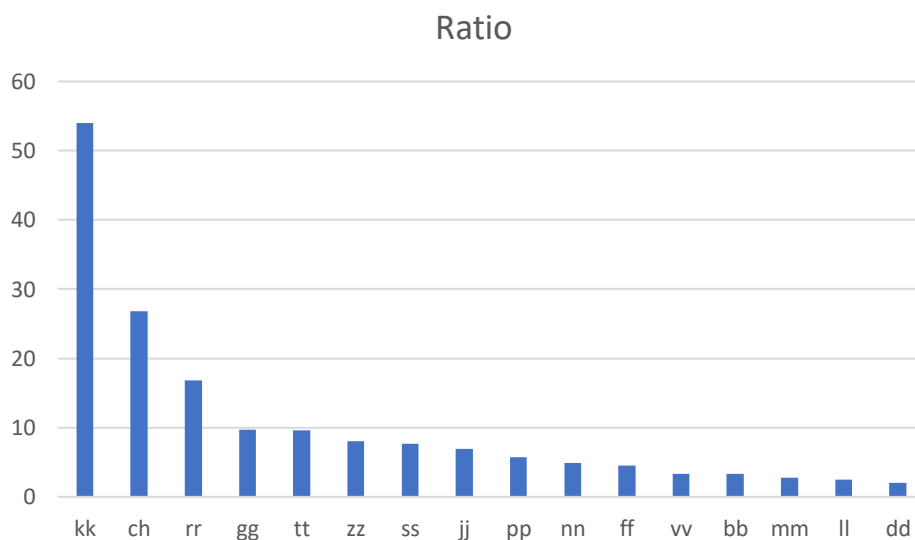


Figure 23 : Ratio entre les patients et les sujets contrôles en fonction des phonèmes altérés

Avec ce nouveau classement, nous retrouvons les phonèmes /k/, /ch/, /r/ et /g/ comme les plus altérés, avec une importance significative pour le /k/.

Pour clore les comparaisons de notre tâche VCV avec les autres évaluations, nous avons aussi voulu vérifier que notre tâche ne comportait pas de biais.

➤ Durée des stimuli

Nous avons cherché à savoir si la durée de nos stimuli avait un impact sur les scores PPD. Autrement dit, est-ce que ce sont les stimuli les plus brefs qui sont les moins bien identifiés, donc qui ont un score PPD élevé. Si tel était le cas, nous aurions un biais : notre tâche n'évaluerait pas l'intelligibilité des stimuli mais leur durée.

La durée des stimuli pouvait varier. Tous nos stimuli concernant les consonnes étaient de 30 ms minimum, le plus court de 60 ms. Il va de soi que plus un stimulus est court, moins bien il est perçu.

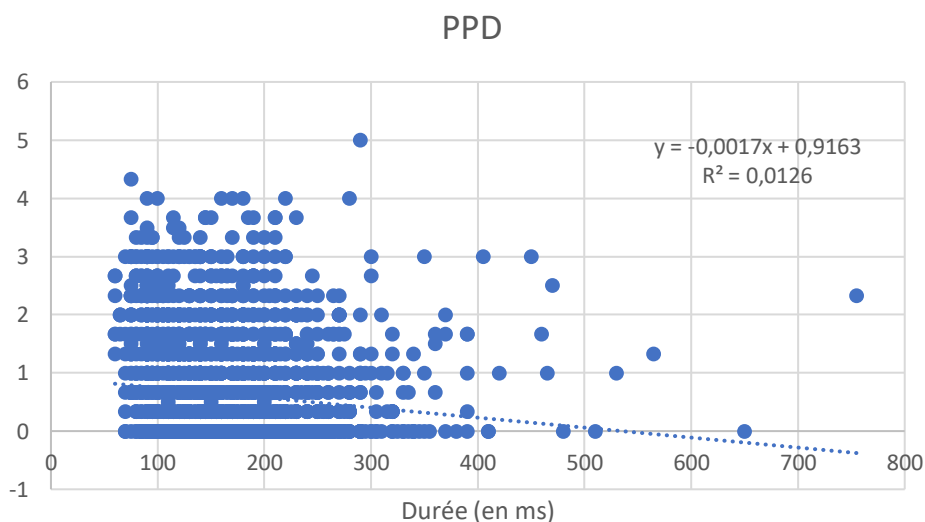


Figure 24 : Graphique des scores PPD VCV en fonction de la durée des stimuli

Ce graphique nous indique alors que nous avons tous les scores PPD pour toutes les durées possibles. Ainsi, les scores PPD ne sont pas en lien avec la durée des stimuli, et sont donc bien en lien avec l'évaluation de l'intelligibilité.

Après avoir observé les données de notre tâche VCV, nous nous sommes concentrés sur les données des voyelles V.

## B. Traitement des données Voyelle (V)

De la même manière que pour les données VCV, un pré-traitement est nécessaire pour les données Voyelle. Une concaténation en un seul classeur Excel a été réalisée à nouveau. Nous avons également recueilli 54 fichiers réponses pour cette tâche V (18 auditeurs x 3 blocs).

script	listener	block	date	Lot1	trial_id	stim_id	wave	speaker	group_speaker	id_voy	rank_vcv	duration	expected	response	rtime	compare	PPD
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		4	2941	TIO-000345-(TIO-000345	PATIENT		in-14	14	150	in	aa	4067	µ a	2
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		83	1382	TIO-000309-(TIO-000309	PATIENT		aa-24	24	40	aa	aa	12587	a a	0
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		135	830	TIO-000034-(TIO-000034	CTRL		ou-44	44	80	ou	ou	2444	u u	0
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		137	2035	TIO-000322-(TIO-000322	PATIENT		an-22	22	60	an	in	3020	@ µ	1
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		49	2552	TIO-000336-(TIO-000336	PATIENT		oe-33	33	60	oe	oe	4383	Ū Ū	0
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		147	2222	TIO-000329-(TIO-000329	PATIENT		uu-34	34	70	uu	oe	4748	ŷ Ū	1
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		16	737	TIO-000033-(TIO-000033	CTRL		aa-24	24	70	aa	aa	5525	a a	0
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		88	2417	TIO-000333-(TIO-000333	PATIENT		uu-03	3	150	uu	ii	2129	ŷ i	1
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		42	2512	TIO-000335-(TIO-000335	PATIENT		uu-03	3	50	uu	ei	4437	ŷ É	2
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		19	1134	TIO-000304-(TIO-000304	PATIENT		aa-24	24	170	aa	aa	2434	a a	0
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		122	898	TIO-000036-(TIO-000036	CTRL		an-45	45	150	an	aa	12905	@ a	1
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		2	1535	TIO-000312-(TIO-000312	PATIENT		ai-07	7	120	ai	oe	8213	É Ū	1
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		105	1948	TIO-000320-(TIO-000320	PATIENT		ei-42	42	80	ei	in	3872	É µ	1
Rugbi-Lec-C2SI-Lot1-VOY	AA		1 2021-3-27; 1 Lot1		64	127	TIO-000019-(TIO-000019	CTRL		ii-47	47	160	ii	ii	2720	ii	0

Figure 25 : Tableau de données Rugbi-Lec-C2SI-VOY\_total

Ce tableau nous indique les mêmes caractéristiques que pour celui des données VCV. La différence réside dans les réponses attendues et les réponses fournies, qui sont cette fois-ci des voyelles. Un score PPD est une nouvelle fois attribué à chaque voyelle en fonction du nombre de traits articulatoires qui les diffèrent.

Rappelons, avant toute chose, que nous avons réalisé une neutralisation au niveau des différences de prononciation en fonction des régions (/o/ ouvert ou fermé, /un/ ou /in/ pour « brun » ou « brin »). Nous les avons donc regroupés en archiphonèmes pour supprimer cette différenciation en fonction des auditeurs (cf. VI. Matériel et méthodes).

Étiquettes de lignes	Nombre de listener	Nombre de PPD
AA	501	501
BB	501	501
CC	501	501
DD	501	501
EE	501	500
FF	501	478
GG	501	500
HH	501	494
II	501	499
JJ	501	501
KK	501	501
LL	501	501
MM	501	501
NN	501	501
OO	501	478
PP	501	501
QQ	501	501
RR	501	501
(vide)		
<b>Total général</b>	<b>9018</b>	<b>8961</b>

Figure 26 : Comptage du nombre d'écoutes et de réponses PPD V pour chaque auditeur

Pour l'évaluation perceptive des segments vocaliques (V), nous vérifions tout d'abord la cohérence des données en faisant le comptage. Nous observons que pour les 9018 écoutes, nous obtenons 8961 réponses pour l'ensemble des auditeurs. Nous avons donc à nouveau un total de 57 réponses manquantes (ou non-réponses), ce qui correspond à 0,67 % des données. Cette tâche s'est aussi bien déroulée et nous avons des données en nombre suffisant pour nos analyses. Comme attendu (cf. VI. Matériel et méthodes), nous avons bien 501 écoutes pour tous nos auditeurs.

Après ces vérifications, et le reformatage des données, nous appliquons l'algorithme de calcul pour obtenir les scores de Déviation Phonologique Perçue ou score d'intelligibilité PPD. Nous obtenons pour chaque voyelle produite par chacun des locuteurs contrôles et patients, un score PPD par segment qui correspond au nombre moyen de traits altérés entre la voyelle cible (celle qui devait être produite) et la voyelle perçue.

Group_speaker	id_voy	rank_voy	duration	expected	response	rtime	compare	PPD
PATIENT	in-14	14	150	in	aa	4067	μ a	2
PATIENT	aa-24	24	40	aa	aa	12587	a a	0
CTRL	ou-44	44	80	ou	ou	2444	u u	0
PATIENT	an-22	22	60	an	in	3020	@ μ	1
PATIENT	oe-33	33	60	oe	oe	4383	û û	0

Figure 27 : Tableau Scores PPD par segment vocalique ligne 1 à 5

Si nous nous attardons sur la première ligne de ce tableau, nous pouvons observer que pour le /in/ numéro 14 produit par un patient, c'est un /a/ qui a été perçu par l'auditeur, le score PPD correspond à deux traits articulatoires d'écart : l'aperture et la nasalité. A contrario, à la deuxième ligne, le /a/ numéro 24 a bien été perçu par l'auditeur, ce qui correspond à un score PPD nul.

A partir de ces données, nous avons une nouvelle fois calculé le score PPD moyen pour chaque groupe de locuteurs. Les résultats montrent un score PPD de 0,56 pour le groupe contrôle (CTRL) et de 0,87 pour le groupe patient. La figure suivante montre ainsi que le score PPD des patients est plus élevé que celui des contrôles ce qui traduit comme attendu, une intelligibilité plus altérée pour ce groupe de locuteurs. Toutefois, le différentiel entre les deux groupes de locuteurs est moins marqué par rapport à celui des données VCV.

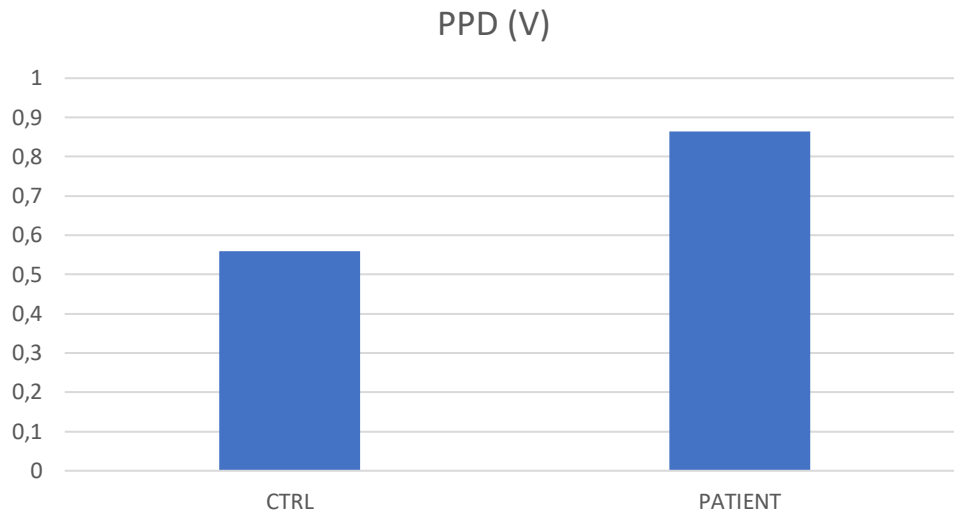


Figure 28 : Scores PPD V pour chaque groupe de locuteurs

Puis, nous avons ordonné chaque score par locuteur, pour nos deux groupes. L'histogramme suivant fournit les scores par locuteur et par groupe, ordonnés du score PPD le plus élevé au moins élevé. Nous avons les 61 locuteurs sur l'axe des abscisses et le score PPD leur correspondant sur l'axe des ordonnées.

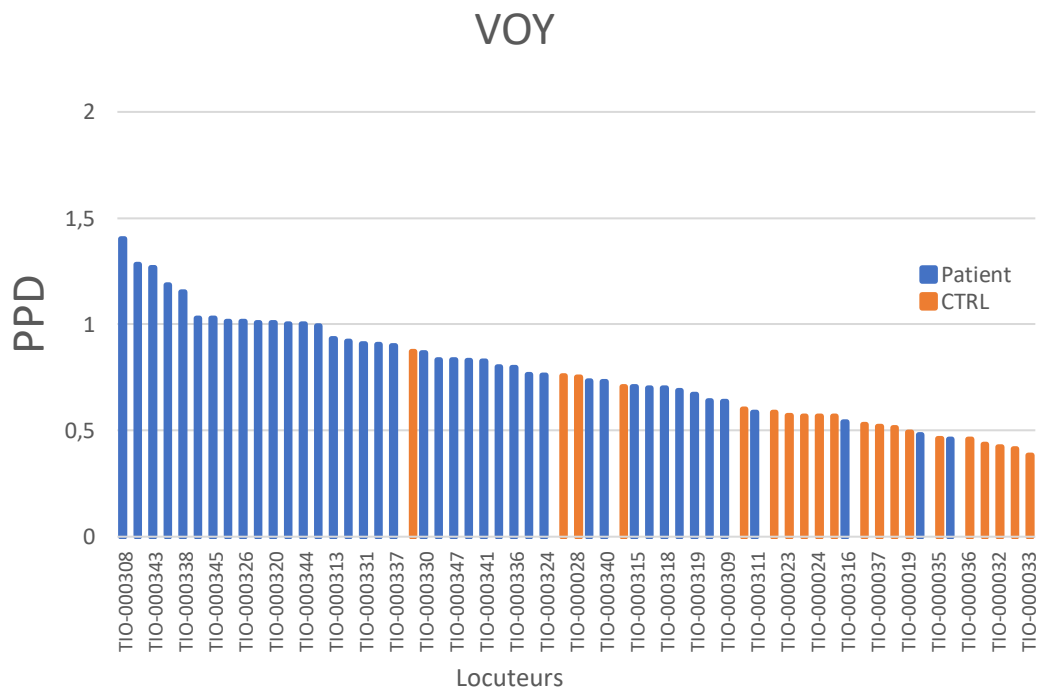


Figure 29 : Scores PPD V par groupe et par locuteur

Les patients sont représentés par les barres bleues et les sujets contrôles par les barres oranges. Nous pouvons formuler que les scores les plus altérés sont des patients.

Nous avons, pareillement, validé nos mesures à l'aide du tracé de la courbe ROC (Receiver Operating Characteristic) afin de nous indiquer la puissance de fiabilité de notre évaluation.

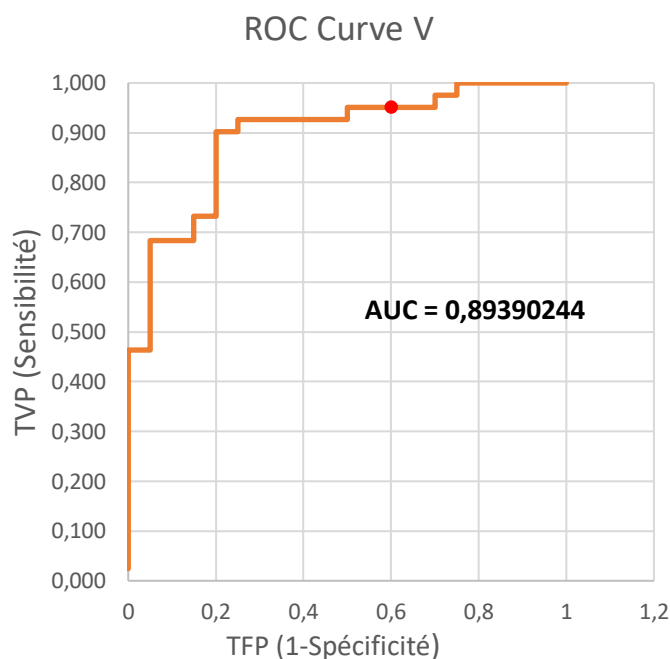


Figure 30 : Courbe ROC V

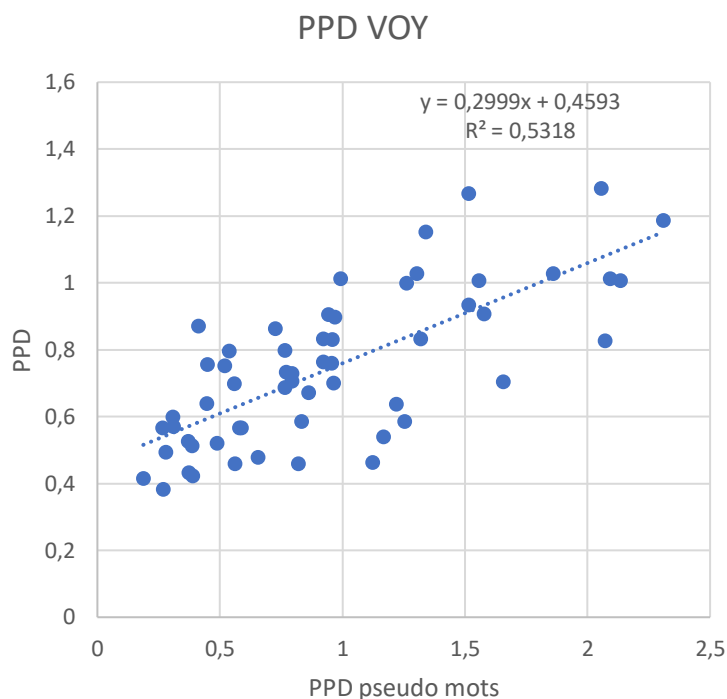
La courbe nous a révélé une aire sous la courbe de 0,89 (AUC pour Area Under the Curve), ce qui est également plutôt proche de 1 donc satisfaisant, même si cela est plus faible que pour les données VCV. Nous pouvons ajouter que le seuil optimal se situe à 0,63 PPD (point rouge). Nous pouvons donc considérer que l'intelligibilité est altérée à partir de ce score PPD seuil de 0,6 trait altéré par phonème.

Poursuivons maintenant en nous attachant sur les résultats V par rapport aux autres évaluations du projet C2SI.

➤ PPD V et PPD Pseudo-mots

Rappelons, tout d'abord, les données recueillies à partir de l'évaluation de DAP de pseudo-mots. Le score PPD des sujets sains est en moyenne à une distance de 0,48 trait par phonème alors que cette distance atteint 1,29 pour les patients.

Si nous comparons nos scores PPD V globaux avec les scores PPD Pseudo-mots globaux, nous observons que les scores sont moins élevés uniquement pour nos segments vocaliques V des patients. En effet, nous retrouvons un score PPD de 0,56 pour nos sujets contrôles, en comparaison à 0,48 avec les pseudo-mots, et un PPD de 0,87 pour les sujets patients, comparé à 1,29 PPD pour l'autre tâche. Ces scores révèlent ainsi une meilleure identification des phonèmes perçus par les auditeurs des locuteurs patients avec notre tâche V.



*Figure 31 : Scores PPD V en fonction des scores PPD pseudo-mots*

Nous avons mis, à nouveau, en relation ces données PPD globaux avec celles de pseudo-mots à l'aide d'un graphique. Il nous montre la relation entre ces deux mesures de PPD. Un point représente un locuteur. Nous pouvons dégager de cette représentation qu'il existe une corrélation entre les deux mesures : un patient dont l'intelligibilité est évaluée mauvaise avec notre tâche V, l'est aussi avec le DAP pseudo-mots.

Le coefficient de corrélation obtenu dans cette analyse est de  $R^2 = 0,53$ . Plus ce coefficient est proche de la valeur 1, plus les données sont liées. Il y a donc une relation entre notre mesure PPD V et celle de PPD de pseudo-mots. Le nuage de points est assez dispersé autour de l'axe, il existe donc une moins bonne corrélation entre ces deux mesures par rapport aux données VCV. Cependant, cela reste pertinent au vu de la représentation homogène de ces points.

Sur le même modèle que pour les comparaisons des données VCV, nous avons cherché aussi à savoir s'il existe une corrélation avec d'autres évaluations. Nous rappelons que notre test perceptif à partir de segments non lexicaux est extrait de la lecture. Nous avons ainsi observé nos résultats en fonction de la sévérité et de l'intelligibilité subjective, au niveau de la lecture et de la description d'images.



➤ PPD V et Sévérité : lecture et description

Ces deux graphiques représentent les scores PPD globaux en fonction de la sévérité du sujet dans la tâche de lecture sur le premier graphique, et dans la description d'images sur le deuxième.

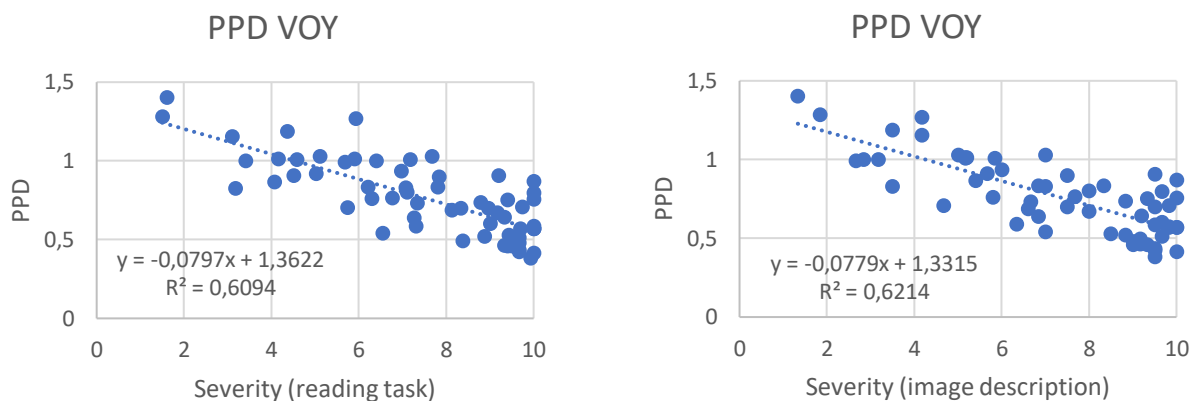


Figure 32 : Scores PPD V en fonction de la sévérité en lecture et description d'images

Nous pouvons observer que les scores PPD V et la sévérité du trouble sont corrélés. Nous retrouvons un score PPD VCV élevé, lorsque l'intelligibilité est altérée aussi. A nouveau, plus la sévérité du trouble est élevée, meilleure est l'intelligibilité. Nous observons que les locuteurs dont l'atteinte est jugée sévère (avec un score bas) sont aussi ceux dont l'intelligibilité est jugée altérée (avec un score PPD haut). Ces scores sont équivalents pour la tâche de description d'images ( $R^2 = 0,62$ ) et celle de lecture ( $R^2 = 0,61$ ).

➤ PPD V et Intelligibilité subjective : lecture et description

Nous avons ensuite établi de nouveau deux graphiques pour confronter les scores PPD V avec l'intelligibilité subjective, en lecture et en description d'images. L'évaluation de l'intelligibilité subjective de la parole repose sur le taux d'unités linguistiques (syllabes et mots principalement) dont tous les phonèmes sont correctement identifiés par les auditeurs.

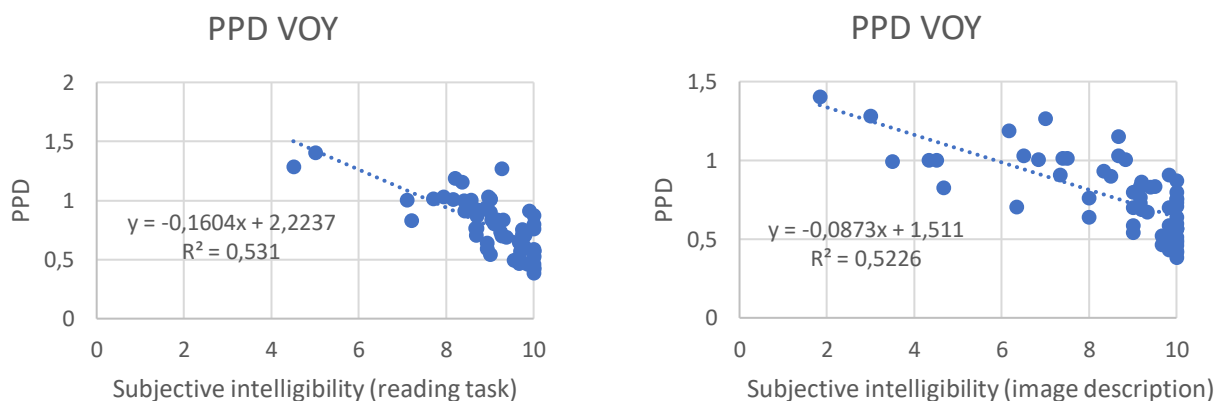


Figure 33 : Scores PPD V en fonction de l'intelligibilité subjective en lecture et description d'images

Nous pouvons voir que l'intelligibilité subjective est moins corrélée avec les scores PPD V, à la fois en lecture ( $R^2 = 0,53$ ) et en description d'images ( $R^2 = 0,52$ ). Nous sommes plutôt loin de 1. Cette corrélation est d'autant plus faible si on la compare avec notre tâche des segments VCV.

Puis, nous voulions observer le lien entre notre tâche V et l'altération phonétique. Cette dernière renvoie aux distorsions des sons du langage dans la chaîne parlée.

#### ➤ PPD V et altération phonétique

Nous avons tracé ce graphique pour étudier les scores PPD V en fonction de l'altération phonétique.

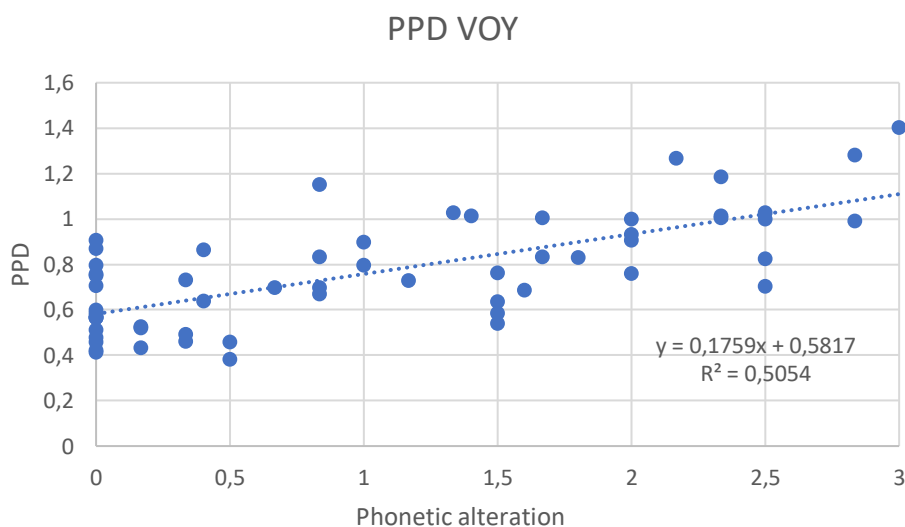


Figure 34 : Scores PPD V en fonction de l'altération phonétique

Nous pouvons constater que plus cette altération est élevée plus le score PPD V est élevé. Les nuages de points sont plutôt dispersés loin de la droite, nous pouvons ainsi dire que la corrélation existante entre ces deux tâches est plus faible, en comparaison avec les évaluations vues précédemment.

Nous pouvons formuler que pour la tâche des voyelles V, la relation est moins nette avec toutes les tâches des autres évaluations.

Nous nous sommes ensuite préoccupés plus précisément de cette altération phonétique. Nous avons cherché à distinguer quels sont les phonèmes les plus détériorés.

#### ➤ Identification des phonèmes les plus altérés (V)

Nous avons classé chaque voyelle en fonction de leur score PPD respectif et de leur groupe de locuteurs, afin de réaliser un histogramme.

Nous avons discerné les phonèmes /an/, /on/, /in/ comme les plus altérés chez les patients avec une légère différence avec les sujets contrôles. Ce sont ces voyelles nasales qui sont le plus difficilement perçues. Toutefois, nous pouvons remarquer que la majorité des voyelles sont distordues, même pour les contrôles, mais que l'écart entre les deux groupes varie en fonction des voyelles considérées.

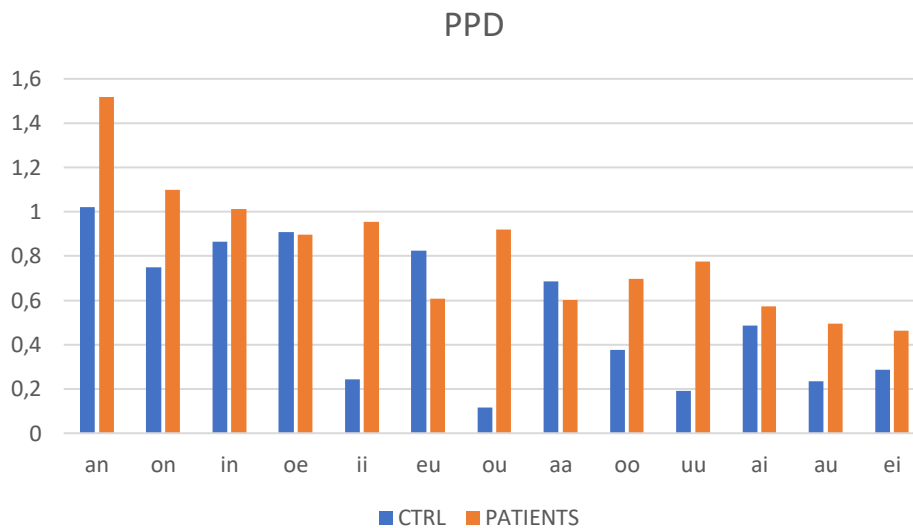


Figure 35 : Histogramme des scores PPD V en fonction des phonèmes altérés pour les deux groupes de locuteurs

Cela nous a amené à proposer un nouveau classement qui rende compte de l'écart de score PPD entre les patients et les sujets contrôles. Cette homogénéisation est nécessaire pour réduire l'écart entre nos deux groupes de locuteurs. La pathologie étant un frein à l'expression du locuteur patient, il est évident que l'altération des voyelles de celui-ci soit plus importante.

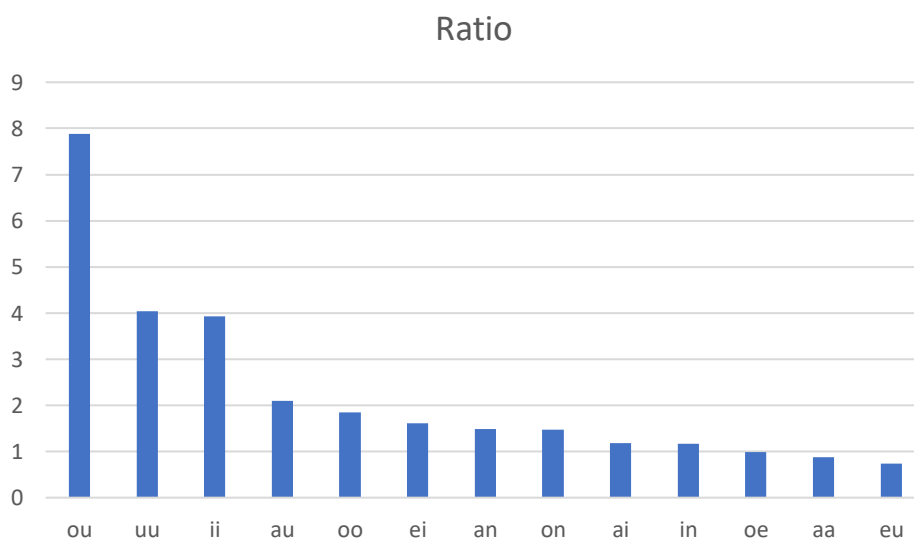


Figure 36 : Ratio entre les patients et les sujets contrôles en fonction des phonèmes altérés

Nous retrouvons les phonèmes /ou/, /u/, /i/ comme les plus altérés, avec une importance significative pour le /ou/. Ce sont finalement les voyelles hautes avec une articulation postérieure qui sont les plus altérées.

Enfin, nous avons également voulu nous assurer que notre tâche V aussi ne comportait pas de biais.

➤ Durée des stimuli

Nous nous sommes à nouveau demandé si la durée de nos stimuli avaient un impact sur les scores PPD : est-ce que ce sont les stimuli les plus brefs qui sont les moins bien identifiés, donc qui ont un score PPD élevé ? Si tel était le cas, nous aurions un biais. Notre tâche n'évaluerait pas l'intelligibilité des stimuli mais leur durée.

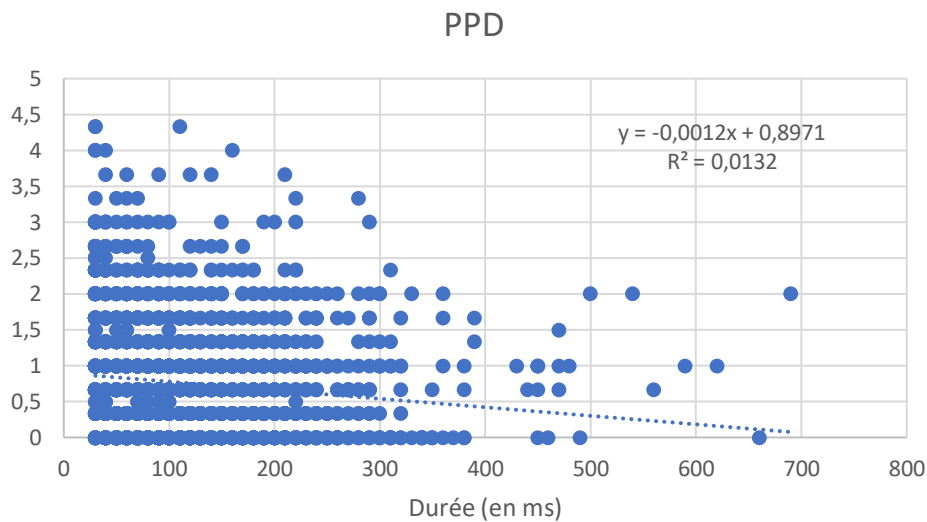


Figure 37 : Graphique des scores PPD V en fonction de la durée des stimuli

La durée des stimuli pouvait varier, avec une durée minimum de 30 ms. Nous rappelons que plus un stimulus est court, moins bien il est perçu. Ce graphique nous indique que nous avons tous les scores PPD pour toutes les durées possibles. Ainsi, les scores PPD ne sont pas en lien avec la durée des stimuli, mais bien en relation avec l'évaluation de l'intelligibilité. Notre tâche V évalue l'intelligibilité et non la durée des stimuli.

## VIII. Discussion

Le but de notre travail était d'analyser la pertinence d'une évaluation de l'intelligibilité basée sur des séquences non-lexicales, extraites de la lecture, par rapport à une évaluation de décodage acoustico-phonétique (DAP) de pseudo-mots.

En amont de nos analyses, nous avons procédé à un certain nombre de vérifications qui nous permettent d'évaluer la qualité de notre protocole. A l'issue du test de perception, nous avons examiné les réponses manquantes lors de notre comptage. Nous nous sommes aperçu que nous avons le même nombre de non-réponses, 57 pour les consonnes et 57 pour les voyelles. Nous sommes ainsi allés examiner plus précisément les réponses des auditeurs et avons constaté que deux auditeurs réunissent à eux deux seuls cette cinquantaine de réponses manquantes pour chacune des tâches VCV et V en fournissant une vingtaine de non-réponses chacun (auditeurs FF et OO). Cela relève d'un pur hasard et nous indique que ce sont les auditeurs qui ont rencontré le plus de difficultés lors de la tâche, toutes proportions gardées, étant donné qu'une vingtaine de non-réponses sur 500 stimuli écoutés ne représente pas une grande proportion. Ensuite, nous avons confirmé que notre tâche évalue la qualité du son, c'est-à-dire l'intelligibilité, et non la durée des stimuli proposés aux auditeurs. Cela est reconnu autant pour la manipulation VCV que V à nouveau. Tous les scores PPD ont été représentés pour toutes les durées. Nous n'avons donc pas de biais dans notre tâche.

Par la suite, nos analyses ont porté sur le fait que la tâche permet l'évaluation de l'intelligibilité avec des séquences non lexicales. Cela est avéré à la fois pour la tâche des segments VCV comme V. Tout d'abord, nous avons pu montrer que la tâche permet de distinguer deux catégories de locuteurs. En effet, nous avons une distinction majeure de score PPD entre le groupe des contrôles et les patients : 0,19 PPD (contrôles) < 0,83 PPD (patients) pour VCV, et 0,56 PPD < 0,87 PPD pour V. Ceci nous conduit à affirmer que les patients ont été évalués, comme attendu, avec une intelligibilité plus altérée que les sujets contrôles. Nous avons ensuite examiné les scores PPD individuels. Ainsi, en plus de la distinction possible patient/contrôle, cette tâche aide à faire une différence entre chaque locuteur. Elle peut coter précisément et de manière analytique chaque locuteur en fonction de son degré d'atteinte et d'en apprécier la sévérité. Ensuite, il a été question de valider nos mesures grâce à la courbe ROC. Celle-ci nous a permis d'identifier les vrais positifs des faux négatifs et nous fournir un seuil correspondant à cette limite entre un patient et un sujet contrôle. Cet ensemble d'analyses ont approuvé que le test basé sur des séquences délexicalisées évalue bien l'intelligibilité.

Puis, nous avons porté notre attention sur l'objectif principal de ce travail, à savoir si notre test de jugement perceptif à partir de séquences délexicalisées extraites de la lecture est équivalente au décodage acoustico-phonétiques (DAP) de pseudo-mots. Nous avons, tout d'abord, une bonne corrélation entre ces deux tâches, nous évaluons l'intelligibilité aussi bien dans une tâche que dans l'autre. En d'autres mots, un patient évalué avec une intelligibilité dégradée en pseudo-mots l'est

aussi lorsqu'il est évalué à partir de séquences délexicalisées, alors même que nous sommes sur deux tâches différentes pour le locuteur. Toutefois, nous avons noté des différences entre les scores PPD VCV/V et ceux de pseudo-mots. Ces différences peuvent s'expliquer en partie par les effets de la tâche du point de vue de l'auditeur. Il faut notifier que notre tâche a été évaluée grâce à un choix forcé pour nos auditeurs. En effet, ils devaient choisir le phonème le plus approprié qu'ils venaient d'écouter parmi un panel de choix devant eux, tandis que lors de la tâche de pseudo-mots, l'auditeur devait retranscrire orthographiquement le pseudo-mot attendu. Cette dernière est source d'erreurs plus nombreuses possibles qu'avec un choix à sélectionner. Ce choix forcé nous a permis d'obtenir peu de non-réponses (ou réponses manquantes). Nous avons donc ici un effet de la tâche non négligeable.

D'autre part, nous avons des résultats différents entre les scores PPD globaux. Ils sont moins élevés pour les segments consonantiques (VCV) par rapport aux scores globaux de DAP pseudo-mots. Il semble donc que l'intelligibilité est jugée moins altérée lorsqu'elle est évaluée à partir des séquences VCV et V.

De plus, nous pouvons rapporter une meilleure corrélation du score PPD VCV avec celui du PPD de pseudo-mots, par rapport au PPD des voyelles avec celui des pseudo-mots. En effet, les scores sont significativement plus faibles lorsque nous utilisons la tâche à partir des voyelles (segments vocaliques). Or, à l'heure d'aujourd'hui, la littérature émet que les voyelles sont plus porteuses de sens, soit d'intelligibilité. D'après l'étude de Kewley et al (2007), examinant la contribution de l'information fournie par les voyelles par rapport aux consonnes pour l'intelligibilité des phrases : les phrases contenant uniquement des voyelles sont significativement plus intelligibles que les phrases comportant des consonnes uniquement.

Contrairement à l'anglais écrit ou mots prononcés isolément, les résultats ont démontré que pour les phrases parlées, les voyelles portent plus d'informations sur l'intelligibilité de la phrase que les consonnes pour les jeunes auditeurs normaux et auditeurs âgés malentendants. Certaines recherches dans ce domaine (comparant les informations sur les voyelles par rapport aux consonnes), tels que les travaux de Fletcher en 1929, ont montré que les consonnes fournissent plus d'informations que les voyelles pour la reconnaissance de mots isolés. Cela a également été récemment confirmé avec Owren et Cardillo (2006). Ainsi, contrairement à la pensée des scientifiques et cliniciens, les auditeurs jeunes et âgés obtiennent de meilleurs résultats lorsque l'information vocalique est conservée dans les phrases par opposition à l'information consonantique. Il n'y a donc pas de réel consensus quant aux unités les plus altérées, entre les consonnes et les voyelles. Dans notre étude, les voyelles sont évaluées comme moins altérées que les consonnes, et il sera intéressant de confronter ce résultat à un test de compréhension.

Enfin, nous avons discerné une cohérence articulatoire pour les consonnes. Nous avons distingué les phonèmes /k/, /ch/, /r/ et /g/ comme les plus altérés, avec une importance significative pour le /k/. D'après l'alphabet phonétique international, nous pouvons classer : /k/ et /g/ comme des occlusives vélares, /ch/ une constrictive post-alvéolaire, et /r/ une constrictive uvulaire. Cela nous permet de dire que les segments consonantiques les plus altérés sont ceux qui présentent une articulation postérieure. Nous pouvons faire le lien avec les travaux Monestier (2020) en ce qui concerne l'identification des phonèmes altérés puisque son mémoire de Master 2 a porté en partie sur ce sujet. Monestier a pareillement rencontré une intelligibilité « significativement plus altérée pour les consonnes palato-vélares /ʒ, ʃ, k, g, R/ donc très postérieures, que pour les consonnes alvéolaires /t, d, s, z, n, l/ ou labiales /p, b, f, v, m/, plus antérieures ». Nous retrouvons ainsi cette notion de postériorité associée à l'altération de l'intelligibilité dans nos deux évaluations.

Au niveau des voyelles, nous observons les phonèmes /an/, /on/, /in/ comme les plus altérés chez les patients avec une légère différence avec les sujets contrôles. Si l'on compare à nouveau avec les données de Monestier (2020), il a aussi été montré que « ce sont les voyelles nasales qui étaient le plus altérées chez les patients, avec des scores locaux plus de deux fois supérieurs aux scores locaux des voyelles orales. ». Nous avons ainsi une nouvelle fois une concordance de nos résultats respectifs. Cependant, lorsque nous regroupons nos deux classes de locuteurs sous un ratio, afin de supprimer la différence d'écart de score PPD et l'homogénéiser, nous nous rendons compte que ce ne sont plus les mêmes voyelles qui sont distordues. Nous avons obtenu les phonèmes /ou/, /u/, /i/ comme les plus altérés, avec une importance pour le /ou/. Ce sont toutes les trois des voyelles fermées, avec une antériorité pour le /u/ et le /i/, et une postériorité pour le /ou/. Lorsque Monestier s'est intéressée « aux voyelles orales /i, u, y/ [...] en ne prenant plus en compte les voyelles nasales. », elle a pu constater « des scores plus élevés pour les voyelles orales hautes ». Ces voyelles orales hautes sont donc moins intelligibles chez nos patients. Nous pouvons alors à nouveau détacher une cohérence articulatoire lorsque nous nous attardons sur les voyelles orales seulement, avec des voyelles hautes dans nos deux tâches. De surcroît, cela montre qu'avec notre tâche de jugement de l'intelligibilité basée sur des séquences délexicalisées extraites de la lecture, nous retrouvons les mêmes « profils » qu'avec la tâche DAP de pseudo-mots, ce qui renforce la corrélation entre ces deux tâches.

A ce stade, notre objectif principal a été validé, même si nos résultats soulèvent de nouvelles questions. En montrant une bonne corrélation entre les évaluations par nos deux tâches, nous avons montré qu'une seule tâche de lecture pourrait permettre d'évaluer la compréhensibilité d'une part et l'intelligibilité segmentale d'autre part. En considérant nos résultats en fonction des autres types d'évaluation clinique, nous avons également mis en évidence de bonnes corrélations, en particulier avec les tâches d'évaluation subjectives opérées à partir de corpus de lecture. Les évaluations de

la sévérité et de l'intelligibilité subjective (c'est-à-dire le degré d'altération des sons produits, et le degré de précision avec lequel le message est compris par l'auditeur) ont montré une corrélation importante dans ces deux types d'évaluation. Cette corrélation est de l'ordre de 0,7 en lecture et de 0,6 en description au niveau de l'intelligibilité subjective. Ces résultats admettent ainsi une bonne corrélation entre l'intelligibilité subjective et la tâche VCV. Cela était attendu puisque nous avons confirmé précédemment que notre tâche VCV évalue bien l'intelligibilité. Si nous comparons ces scores de corrélation avec celui du PPD VCV et pseudo-mots ( $R^2 = 0,66$ ), nous pouvons annoncer qu'ils sont plutôt similaires. Cela était également attendu, étant donné que la tâche de DAP pseudo-mots a montré sa capacité à évaluer, de manière objective et fiable, l'altération de l'intelligibilité.

Néanmoins, ces tests subjectifs ont été réalisés selon deux types de tâches différentes : une lecture et une description d'images. Les scores ont révélé une meilleure corrélation en tâche de lecture, ce qui était attendu puisque nos segments VCV ont été extraits de la lecture du texte de *La chèvre de Monsieur Seguin*. Effectivement, nous avons une condition de parole similaire pour l'intelligibilité subjective et la sévérité lors de la lecture, tandis qu'elle est différente avec de la description d'images.

Au contraire, si nous étudions les résultats des voyelles, ils nous indiquent que les scores PPD V sont corrélés avec la sévérité mais nettement moins bien que pour les scores PPD VCV. Cette corrélation se révèle légèrement plus grande en description d'images d'ailleurs. Il est possible que ces corrélations plus faibles soient en lien avec les scores obtenus dans nos tâches, également plus faibles : Il semble que l'évaluation de l'intelligibilité à partir de segment vocaliques soit moins sensible. Ceci nous conduit à penser que la relation des voyelles avec les autres tâches des tests subjectifs est moins pertinente au regard des segments consonantiques, et ce probablement parce que l'altération se révèle moins marquée. Cela est également vrai au niveau de l'altération phonétique. Les scores PPD V sont moins bien corrélés avec l'altération phonétique, tout comme les scores PPD VCV. Ceci est étonnant car il s'agit d'évaluer un phonème dans ces deux évaluations (tâche VCV/V et altération phonétique). La différence peut résider dans le mode de réponse : un choix forcé dans un cas, et une échelle analogique dans un autre. Cela montre indirectement que la tâche de l'auditeur, et la modalité de réponse peuvent influencer les résultats. Des travaux (Yorkston KM., 1978, Hustad KC., 2008, Miller N., 2013, Landa S., 2014, Stipancic KL., 2016) font écho à notre questionnement en ayant analysé le contrôle des variables de conception, de transcription et d'écoute des auditeurs afin de maximiser la validité et la fiabilité des évaluations de l'intelligibilité. En effet, ils ont fait part que les échelles d'évaluation sont de faibles outils pour donner une information clinique autour de l'intelligibilité.



Au terme de ce travail, l'ensemble des comparaisons et réflexions a permis de reconnaître la pertinence de notre tâche VCV, ainsi que son équivalence avec la tâche de DAP de pseudo-mots. Nous pouvons affirmer que notre tâche d'évaluation de l'intelligibilité basée sur la perception de séquences non lexicales, extraites d'un texte lu, est équivalente à la tâche de décodage acoustico-phonétique de pseudo-mots. Ainsi, lorsqu'un patient est évalué avec une intelligibilité altérée en tâche de séquences VCV, il l'est également avec le DAP. Cette corrélation est d'autant plus vraie lorsque nous utilisons des segments VCV par rapport aux voyelles. Cette étude laisse place maintenant à de nouvelles perspectives. Premièrement, il serait intéressant d'analyser les scores locaux de nos scores PPD VCV et V avec ceux rencontrés dans la tâche de DAP pseudo-mots. Nous avons déjà confronté les résultats de manière globale (scores globaux) mais il serait pertinent de les comparer au niveau segmental pour savoir si ce sont les mêmes segments qui sont reconnus altérés dans les deux tâches. C'est un travail qui pourra être repris à la suite du nôtre par exemple. Enfin, notre test de jugement perceptif s'appuyait sur les données des locuteurs du LOT1 du corpus C2SI. Les autres locuteurs du LOT2 restent donc à évaluer. C'est une autre mission qu'il faudra mettre œuvre afin de tester l'ensemble des locuteurs du projet. Cette opération a d'ailleurs déjà été mise en route. Les passations devraient être terminées début juillet. Je serai intéressée de connaître l'issue de ce travail...