

Evaluation de l'intelligibilité de patients avec traitement du cancer des cavités orales et pharyngales

Alain Ghio¹, Muriel Lalain¹, Marie Rebourg¹,
Corinne Fredouille², Virginie Woisard³

(1) Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France

(2) Laboratoire d'Informatique d'Avignon, Avignon, France

(3) Service ORL, CHU Larrey, URI Octogone-Lordat, Toulouse, France

alain.ghio@lpl-aix.fr

RÉSUMÉ

La perte d'intelligibilité représente une plainte importante des patients atteints de troubles de la parole. Plusieurs batteries de test d'intelligibilité existent mais leurs limitations résident dans la capacité des auditeurs à restaurer les séquences distordues. Nous proposons un nouveau test fondé sur l'utilisation de pseudo-mots en grande quantité afin de complètement neutraliser les effets perceptifs indésirables. Nous avons appliqué ce test à une population de 39 sujets sains et 78 patients post traitement de cancers de la cavité buccale et de l'oropharynx. Chaque locuteur a produit 52 pseudo-mots tirés aléatoirement. 40 auditeurs ont retranscrit ces productions. Les transcriptions orthographiques ont été phonétisées et comparées aux formes phonétiques attendues. Un algorithme fournit un score de déviation phonologique perçue (PPD) fondée sur le nombre de traits différents entre la forme attendue et celle transcrite. Les résultats montrent qu'il existe un seuil PPD de 0.6 traits/phonème au-dessus duquel, la parole produite est dysfonctionnelle. De plus, le score de PPD est bien corrélé au jugement subjectif de la sévérité obtenue auprès d'experts. Ce test semble donc efficace pour mesurer la performance articulatoire des locuteurs.

ABSTRACT

Intelligibility Assessment of patients in the Context of Head and Neck Cancers

The loss of intelligibility is a major complaint from patients with speech impairments. Several intelligibility test batteries exist, but their limitations lie in the ability of listeners to restore distorted sequences. We propose a new test based on the use of pseudo-words in large quantities in order to completely neutralize the unwanted perceptual effects. We applied this test to a population of 39 healthy subjects and 78 post-treatment patients for cancers of the oral cavity and pharynx. Each speaker produced 52 pseudo-words drawn at random. 40 listeners transcribed these productions. The orthographic transcriptions were phoneticized and compared to the expected phonetic forms. An algorithm provides a perceived phonological deviation score (PPD) based on the number of different features between the expected form and the transcribed one. The results show that there is a PPD threshold of 0.6 features / phoneme above which the speech produced is dysfunctional. In addition, the PPD score is well correlated with the subjective judgment of severity obtained from experts. This test therefore seems effective in measuring the articulatory performance of speakers..

MOTS-CLÉS : phonétique clinique, intelligibilité, troubles de la parole, cancer des VADS

KEYWORDS: clinical phonetic, intelligibility, speech disorders, head and neck cancer

1 Une mesure d'intelligibilité privilégiant l'information acoustique

La perception de la parole est une intégration complexe d'informations provenant du signal de parole et d'informations de haut niveau détenues par l'auditeur. L'information extraite du signal de parole est traitée de façon ascendante et correspondant au décodage acoustico-phonétique. L'information de haut niveau est manipulée de manière descendante et correspondant à l'exploitation des connaissances linguistiques, encyclopédiques, situation de communication, contexte psychosocial... Lindblom (1990) définit comme « information dépendante du signal » le résultat de processus ascendant tandis que « l'information indépendante du signal » est liée aux processus descendants. Dans ce cadre, Keintz et al. (2007) définissent l'intelligibilité comme « la quantité de parole comprise à partir du seul signal acoustique ». Inversement, Fontan et al. (2015) définissent la compréhension comme « l'intégration à la fois des informations acoustico-phonétiques et de toutes les informations pertinentes indépendantes du signal afin de comprendre un message parlé dans une situation de communication particulière ». Nous adoptons clairement la définition de Keintz et dans ce cadre, le terme qui devrait être utilisé pour l'intelligibilité est la reconnaissance de la parole et non la compréhension du message. Dans notre définition de l'intelligibilité, la question est la suivante : les auditeurs reconnaissent-ils les sons de la parole qui sont prononcés ? Pour illustrer clairement cette position, nous pouvons rappeler la célèbre phrase composée par Noam Chomsky (1957) « Les idées vertes incolores dorment furieusement » comme exemple syntaxiquement correct, mais sémantiquement absurde. Si cette phrase est clairement produite oralement par un locuteur natif sans trouble de la production de la parole et correctement transcrite par un auditeur natif sans trouble auditif dans un environnement calme, il est raisonnable de dire que l'énoncé est intelligible mais incompréhensible.

Dans son modèle de processus de communication, Lindblom (1990) explique que lorsque les informations dépendant du signal sont précises, l'auditeur est capable de comprendre le message sans information indépendante du signal. En revanche, lorsque les informations dépendant du signal sont insuffisantes, les informations indépendantes du signal deviennent cruciales pour comprendre le message du locuteur. Si un locuteur présentant un trouble de la production de la parole fournit aux auditeurs un signal imprécis, il tentera, avec la complicité de son interlocuteur, de compenser en augmentant les informations indépendantes du signal pour combler les lacunes laissées par des informations dépendantes du signal incomplètes ou compromises. Dans la vie de tous les jours, ces processus sont essentiels et doivent être largement utilisés dans la communication par et avec les patients. Mais dans le cadre d'une évaluation de la gravité d'un trouble de la parole, il est nécessaire de concentrer la mesure sur le locuteur lui-même en minimisant les effets liés à l'auditeur dont les variations peuvent être vues comme un bruit de mesure.

De façon classique, les tests d'intelligibilité sont effectués à partir de phrases ou de mots issus de listes de référence. Les limitations de ce type de test résident dans la capacité des auditeurs à restaurer les séquences distordues (Warren et al., 1970). Cet effet est d'autant plus fort que les auditeurs ont une connaissance forte des mots utilisés dans le test et que ces mots sont peu ambigus et donc fortement prédictibles. C'est généralement le cas des orthophonistes qui peuvent faire un usage si important de ces listes qu'ils/elles finissent par les connaître par cœur. On peut citer par exemple la BECD (Auzou et al., 2006) qui ne comporte que 50 mots. Le biais lié à cette connaissance (Rebourg et al, 2020) et donc à la forte influence des mécanismes perceptifs descendants est un score d'intelligibilité surévalué car la restauration phonémique de l'auditeur rend « transparentes » les distorsions de production.

La solution que nous proposons consiste à utiliser des pseudo-mots, c'est-à-dire des logatomes respectant les structures phonotactiques fréquentes du français, en grande quantité de façon à complètement neutraliser les effets de lexicalité ou d'apprentissage des items par les auditeurs. Au final, les auditeurs sont confrontés à une tâche de décodage acoustico-phonétique suivie d'une transcription écrite. Les détails de la construction du test sont donnés dans (Ghio et al., 2018). Le principe du test est de faire prononcer 52 pseudo-mots tirés aléatoirement d'une liste de 89346 formes possibles, sachant que chaque liste est, par construction, phonétiquement équilibrée. Les pseudo-mots ont été construits avec les formes $C(C)_1V_1C(C)_2V_2$ où $C(C)_i$ est une consonne isolée ou un groupe consonantique. Par exemples: stoumo, vurtant, muja, charou, leba, ranto...

Dans le cadre des troubles de la production de la parole, nous partons de l'hypothèse que l'information dépendante du signal est dégradée en raison de l'imprécision articulatoire et/ou phonatoire et que cette imprécision contribue à diminuer l'intelligibilité. Dans ce contexte, si le locuteur a l'intention de dire quelque chose mais que l'auditeur entend autre chose, nous émettons le postulat qu'il s'agit d'une erreur de production de parole. Nous supposons que le canal de communication est parfait (pièce silencieuse, lecture audio performante) et que l'auditeur est normo entendant.

2 Matériel et méthodes

Le projet C2SI (Carcinologic Speech Severity Index) est un projet financé par l'Institut National du Cancer dont l'objectif est d'obtenir une mesure de l'impact des traitements des cancers de la cavité buccale et du pharynx sur la production de la parole. Cette mesure est explorée à la fois par des méthodes perceptives et par traitement automatique de la parole (Astésano et al., 2018). Dans ce cadre, un certain nombre de patients ont été enregistrés et l'objectif de ce travail est de faire part des résultats de la mesure d'intelligibilité sur cette cohorte de locuteurs.

2.1 Locuteurs

117 locuteurs (39 sujets sains et 78 patients) ont été enregistrés dans le service d'oncoréhabilitation de l'Oncopole à Toulouse. Les patients devaient répondre aux critères d'inclusion suivants:

- avoir un cancer T1 à T4 de la cavité buccale et / ou du pharynx;
- avoir bénéficié d'un traitement par chirurgie et / ou radiothérapie et / ou chimiothérapie;
- être à plus de 6 mois après la fin du traitement pour assurer la stabilité du trouble de la parole, qu'il soit audible ou non.

De même, les critères de non-inclusion étaient de présenter une autre source de troubles de la parole (par exemple le bégaiement) ou de présenter des problèmes cognitifs ou visuels incompatibles avec la conception du protocole d'évaluation. Ces critères de non-inclusion ont également été utilisés pour le recrutement de la population témoin.

2.2 Le corpus

Pour enregistrer le corpus, les locuteurs étaient confortablement installés dans une salle anéchoïque devant un écran d'ordinateur sur lequel était automatiquement affichée la forme orthographique du pseudo-mot à prononcer et une version audio produite en même temps. Cette double modalité, visuelle et auditive, a permis de limiter les erreurs de lecture ou d'éventuelles difficultés auditives et attentionnelles. Les enregistrements ont été effectués avec un microphone à condensateur cardioïde Neumann TLM 102 connecté à un enregistreur numérique FOSTEX. La fréquence d'échantillonnage était de 48 kHz.

Chaque locuteur prononçait une liste différente de pseudo-mots, tirés aléatoirement du dictionnaire des 89346 formes possibles, tout en intégrant des contraintes phonétiques identiques dans chaque liste. En effet, comme décrit dans (Ghio et al., 2016), le même nombre de phonèmes apparaissent dans chaque liste mais avec des combinaisons différentes, ce qui nous conduit à faire l'hypothèse que les listes sont équivalentes. Une fois le locuteur enregistré, le signal de parole était segmenté afin d'obtenir un fichier audio par pseudo-mot. Le corpus était donc composé de 117 locuteurs x 52 items = 6084 stimuli

2.3 Le test de perception

40 auditeurs francophones natifs, sans trouble de l'audition, non spécialistes des troubles de la parole, ont retranscrit ces productions via le logiciel LANCELOT (André et al., 2003). Ils recevaient l'instruction suivante: " Vous allez entendre des non-mots. Un non-mot est une combinaison de sons de la langue française qui n'a pas de signification (ex: gloutu). En respectant les règles de l'orthographe du français, vous devrez transcrire ce que vous entendrez. Certaines prononciations seront difficiles à identifier, mais dans tous les cas, vous devrez fournir une transcription ".

Les différents stimuli étaient répartis dans plusieurs blocs et présentés dans un ordre aléatoire. Chaque stimulus a été transcrit par 3 auditeurs différents, ce qui représente finalement 18 252 réponses (6084 stimuli x 3). Ces tests de perception ont eu lieu au Centre d'Expérimentation sur la Parole (<http://cep.lpl-aix.fr/>) du Laboratoire Parole et Langage à Aix-en-Provence. Chaque auditeur, portant un casque audiophonique Superlux HD 681B a transcrit sur ordinateur environ 456 stimuli en 3 blocs. L'intensité de la présentation a été préréglée par l'auditeur pour être confortable et optimale pour la tâche. Chaque test a commencé avec quatre stimuli d'entraînement. Chaque élément était présenté une fois automatiquement mais l'auditeur pouvait répéter la lecture deux fois.

2.4 Le prétraitement des réponses

Au terme du test de perception, nous avons récupéré 18 252 réponses. Comme détaillé dans (Ghio et al., 2018), ces transcriptions orthographiques ont été phonétisées et comparées aux formes phonétiques attendues des pseudo-mots. Cette comparaison utilise un algorithme de Wagner-Fisher fondé sur un calcul de traits distinctifs déviants entre la forme cible et la forme transcrite. Notre mesure, que nous baptisons PPD (Perceived Phonological Deviation), représente le nombre moyen de traits mal perçus par phonème. Etant donné que le test part du postulat que le canal de communication est parfait (pièce silencieuse, lecture audio performante) et que l'auditeur est normo entendant, l'erreur perçue est directement mise en lien avec une erreur de production.

Puisque chaque pseudo-mot a été transcrit par 3 auditeurs différents, nous obtenons 3 scores pour chaque pseudo-mot. Pour chaque pseudo-mot, nous avons calculé la moyenne des réponses des 3 auditeurs, ainsi que la médiane. Afin d'assurer la cohérence des réponses et l'accord entre les auditeurs, nous avons appliqué une détection des valeurs aberrantes. Nous considérons comme aberrant un résultat qui s'écarte de 2.5 traits de la médiane. Comme +/- 2.5 traits donne une étendue de 5, c'est la moitié de la différence maximale entre 2 phonèmes qui est de 10. À la suite de ce filtrage, nous avons supprimé 1,2% des transcriptions. À cette étape, nous avons un score par pseudo-mot par locuteur.

Dans un deuxième temps, nous avons calculé la moyenne de ces 52 scores par locuteur. Nous obtenons ainsi, pour chacun des 117 locuteurs, un score PPD qui reflète le nombre moyen de traits déviants par phonème.

3 Résultats et discussion

Tous les tests statistiques ont été effectués dans l'environnement logiciel R version 3.4.4 (R Core team, 2017). Nous rappelons que le score PPD (Perceived Phonological Deviation) rend compte d'une dégradation et que plus il est bas, plus nous considérons que l'intelligibilité est bonne.

3.1 Le score PPD en fonction du groupe de locuteur

Le score PPD des sujets sains est en moyenne à une distance de 0,48 trait / phonème (écart-type = 0.23) alors que cette distance atteint 1,29 pour les patients (écart-type = 0.63) (Figure 1). Les données n'étant pas gaussiennes, nous avons effectué une transformation logarithmique du score. Nous avons obtenu des distributions gaussiennes (test de Shapiro, $p > 0,05$) et des variances homogènes (test de Bartlett, $p > 0,05$). L'analyse de la variance (ANOVA) a été réalisée avec le log-score comme variable et le « groupe de locuteurs » comme facteur. La différence entre les deux groupes était significative ($F(1,115) = 127.2$; $p < 0,001$).

On remarque que même les sujets contrôle ont un score PPD non nul, ce qui s'explique par le fait que les auditeurs perçoivent des exemplaires de phonèmes légèrement altérés y compris par des locuteurs non pathologiques. Nous mesurons là les distorsions « normales » qui se produisent dans la production de la parole et qui sont en général rectifiées par l'accès au lexique et au sens, mécanismes inhibés dans notre tâche de décodage acoustico-phonétique.

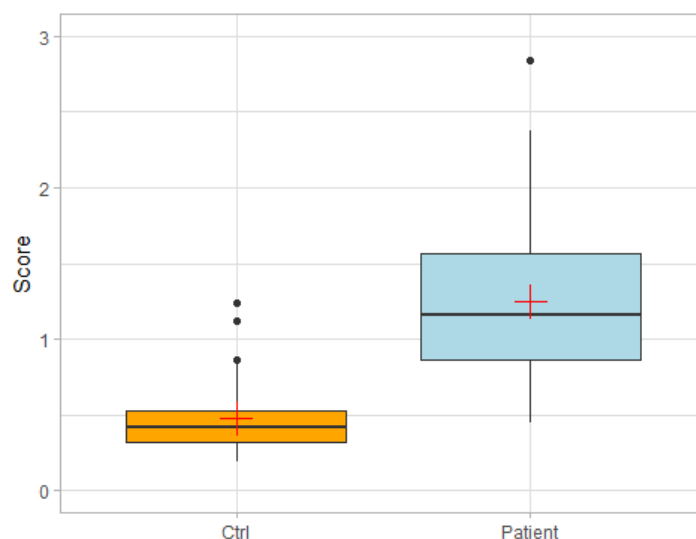


Figure 1 : score PPD pour le groupe contrôle (Ctrl) et pour les groupe des Patients

3.2 Le score PPD par locuteur et son pouvoir classificateur

La Figure 2 illustre la répartition du score PPD par locuteur en fonction du groupe contrôle (CTRL) ou des patients. On remarque une répartition assez distincte entre les sujets sains (faibles scores) et la majorité des patients, ce qui laisse présager un bon pouvoir discriminant de ce test. Certains patients ont un score bas, ce qui peut traduire un faible impact fonctionnel sur la parole du traitement du cancer sur ces sujets. Inversement, quelques locuteurs du groupe contrôle s'illustrent par des scores élevés. En analysant les raisons de ces scores, nous avons observé que ces sujets avaient eu des difficultés ou un manque d'attention à produire les pseudomots. Par exemple, le sujet TTT88 prononce distinctement [mjozo] le pseudomot « minso » ou encore [plɔkso] la séquence « plouco ». Il est donc normal que les auditeurs aient transcrits ces deux items « miozo » ou « plocso », ce qui a généré un PPD important car la cible était « minso » et « plouco ».

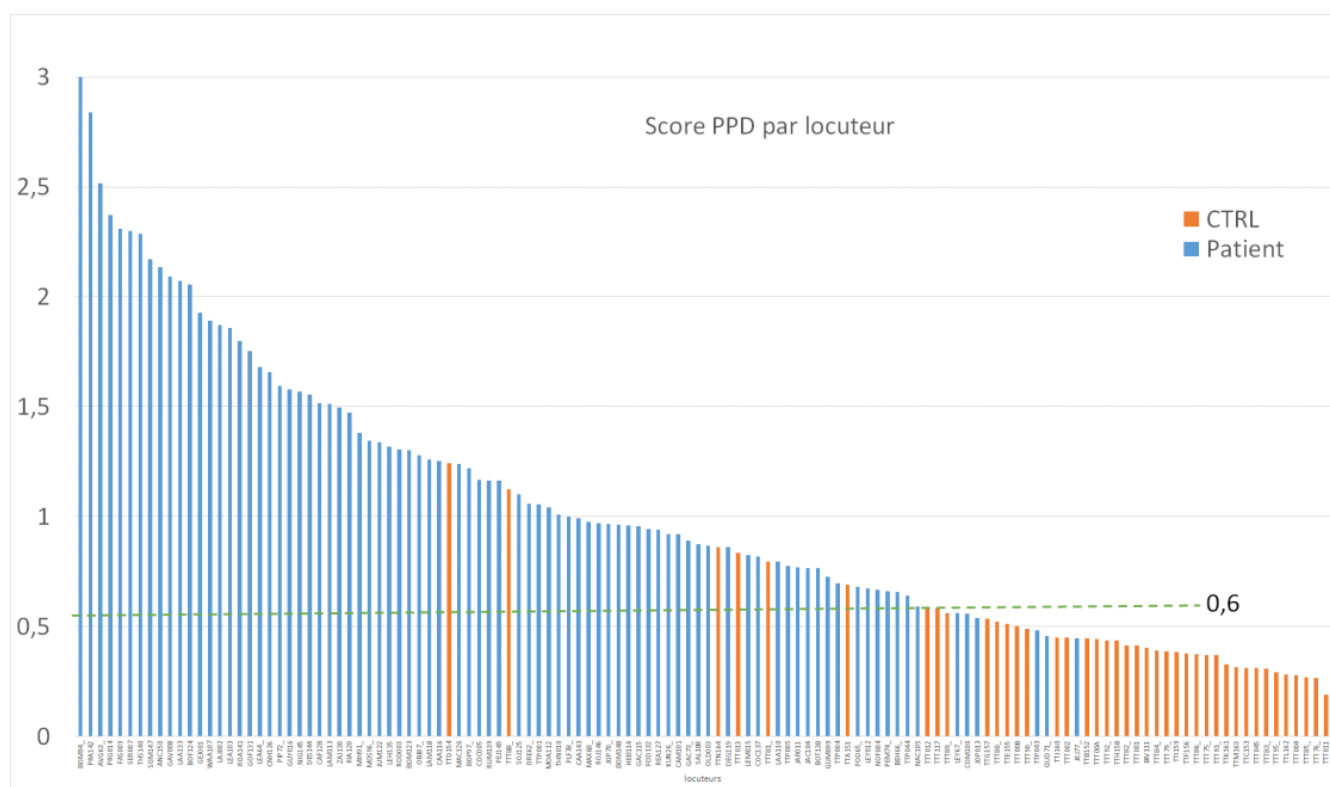


Figure 2 : score PPD par locuteur.

La ligne pointillée horizontale indique le seuil optimal de distinction patient/ contrôle

Afin de mesurer le pouvoir discriminant du score PPD, nous avons établi une courbe de sensibilité/spécificité, qui permet de mesurer la performance du classificateur binaire, c'est-à-dire le pouvoir de catégorisation des deux groupes de locuteurs sur la base du score PPD. La fonction ROC (Receiver Operating Characteristic) se présente sous la forme d'une courbe qui trace le Taux de Vrais Positifs (patients détectés comme patients) en fonction du Taux de Faux Positifs (fraction des sujets sains qui sont incorrectement détectés comme des patients) pour tous les seuils de classification (Delacour et al., 2005).

Cette courbe (Figure 3) est intéressante car elle permet de prédire la performance de classification par la mesure de l'aire sous la courbe (Area Under the Curve $\Leftrightarrow$ AUC). L'AUC, qui mesure l'intégralité de l'aire à deux dimensions située sous l'ensemble de la courbe ROC, indique la probabilité pour que la fonction PPD place un patient devant un sujet contrôle (dans le meilleur des cas, l'AUC vaut 1). Elle permet d'évaluer l'intérêt diagnostique d'un test. Dans notre cas, l'AUC est égal à 0.94, ce qui correspond à une haute précision. Nous pouvons donc affirmer que nous avons obtenu une validité de construit du test fondé sur le score PPD car il permet de distinguer avec une haute précision les locuteurs sains du groupe des patients.

La courbe ROC permet également de déterminer la valeur seuil qui va optimiser le test. Dans notre cas, la question se pose sur la valeur seuil du PPD en dessous duquel se situe la normalité et au-dessus duquel on entre dans le dysfonctionnement. Intuitivement, celle-ci peut être identifiée comme étant le point de la courbe le plus éloigné de la diagonale représentant le test d'apport nul. Ce point correspond également au maximum de l'indice de Youden ($Se + Sp - 1$) » (Delacour et al., 2005). Dans notre cas, le maximum de l'indice de Youden vaut 0.783, qui correspond à un seuil PPD égal à 0.6, indiqué sur la Figure 2 par la ligne pointillée horizontale..

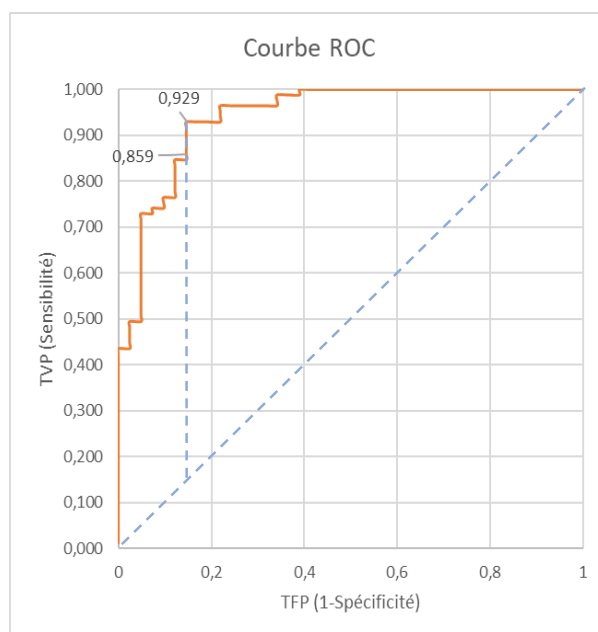


Figure 3 : courbe ROC établie sur le pouvoir classificateur du PPD pour distinguer les sujets sains des patients

3.3 La corrélation avec une mesure clinique de jugement subjectif de la sévérité

Dans le cadre du projet C2SI, les patients et les locuteurs sains ont produit de la parole spontanée obtenue avec une tâche de description d'image (Astesano et al., 2018). La première minute de chaque enregistrement a été utilisée pour une évaluation clinique subjective telle que définie par Kent et al. (1989). Cette évaluation globale consiste en un jugement de la sévérité du trouble sur une échelle ordinaire entre 0 (altération sévère) et 10 (parole normale). Le score relatif à un locuteur est obtenu par la moyenne de six orthophonistes considérés comme experts en troubles de la parole. Afin d'évaluer la fiabilité inter-juges, un coefficient de corrélation inter-classes (ICC) a été calculé. Le degré de concordance entre les notes du jury est bon ($r = 0,77$). Le jury est homogène et fait office de « gold standard » (Balaguer et al., 2019). Afin de vérifier la validité concourante du test PPD, nous avons examiné la corrélation entre le score PPD et le jugement clinique de la sévérité (Figure 4). Ces deux grandeurs sont bien corrélées avec un $R_{\text{spearman}} = -0.85$

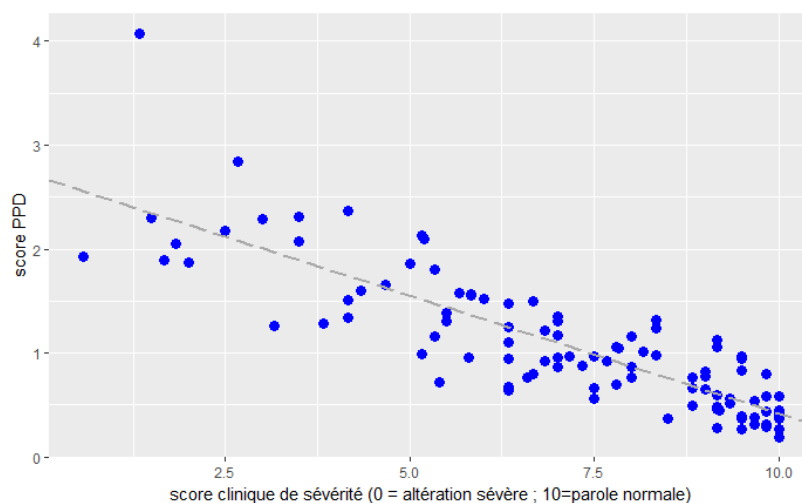


Figure 4 : corrélation entre score PPD et score clinique de sévérité

4 Conclusion

Le test d'intelligibilité que nous proposons est fondé sur des pseudomots ainsi que sur une métrique organisée autour du comptage de traits phonétiques mal identifiés par des auditeurs. Si le test est correctement administré, ces erreurs d'identification peuvent être directement associées aux troubles de la production de la parole. Nous avons conscience de notre définition stricte de l'intelligibilité vue comme « la quantité de parole comprise à partir du seul signal acoustique » Keintz et al. (2007). Un tel test doit être complété par des évaluations de la compréhensibilité comme cela a été fait dans le cadre du projet C2SI par (Nocaudie et al, 2018), pour rendre compte de la réalité du déficit fonctionnel de parole chez ces patients.

Dans ce travail, nous considérons que le score PPD a été validé dans le construit du test dans la mesure où il remplit sa fonction discriminante pour distinguer avec précision les patients des sujets sains. Nous observons aussi une validité concourante puisque cette mesure est corrélée avec le gold standard du jugement subjectif clinique de la sévérité du trouble de la production de la parole. Comparée à cette évaluation subjective de la sévérité, la tâche que nous proposons est une véritable tâche linguistique de décodage et non un processus d'interprétation subjectif. Elle correspond donc plus au processus normal de la communication orale. Grâce au matériel linguistique utilisé (les pseudo-mots), l'évaluation par décodage acoustico-phonétique est moins dépendante des mécanismes top down de la perception et donc moins dépendante de l'auditeur. Cette réduction de la dépendance aux spécificités de l'auditeur pourrait permettre de réduire les phénomènes de variabilité qui fragilisent les résultats des tests subjectifs. De plus, l'utilisation de pseudo-mots présente l'avantage de disposer d'un matériel linguistique bien maîtrisé, standardisé et en très grande quantité. Par conséquent, le score PPD obtenu est moins sujet à un biais d'évaluation. Comme le montrent les résultats, l'évaluation de l'intelligibilité obtenue avec des auditeurs naïfs lors de la tâche de décodage acoustico-phonétique est cohérente avec celle des experts. Bien qu'artificiel dans son matériel, ce test est finalement écologique car il est pertinent avec des auditeurs non experts. Cela représente aussi un avantage supplémentaire, d'un point de vue économique car le recours à des évaluateurs naïfs est moins contraignant que le recrutement de spécialistes.

Au terme de ces travaux, de nouvelles pistes émergent de ces résultats prometteurs. L'obtention de matrices de confusion entre phonèmes permettrait de dépasser la seule valeur scalaire du score PPD. Une analyse plus fine des traits altérés pourrait revêtir une importante valeur d'orientation thérapeutique. Un certain nombre de questions devra être réglé : le tirage aléatoire des listes permet-il une équivalence de résultat ? Autrement dit, un même locuteur confronté à deux listes aura-t-il des résultats identiques avec les 2 listes ? Est-il possible de réduire le nombre d'items (52) de façon à réduire le temps de passation tout en maintenant des résultats robustes ? L'un des axes de recherche prévus dans un avenir proche sera d'appliquer cette méthode sur d'autres pathologies comme par exemple, dans les dysarthries (Projet ANR ANR-18-CE45-0008 Rugby). Nous travaillons également à évaluer la contribution de cette mesure en la comparant à des méthodes automatiques qui pourraient être utilisées comme un auditeur robot dont les caractéristiques déterministes réduiraient un peu plus la variabilité perceptive humaine. Mais l'approbation de ces techniques passera là encore par une validation concourante.

Remerciements

Ce travail a été soutenu par la subvention n ° 2014-135 de l'Institut National pour le Cancer (INCA) projet C2SI et par la subvention ANR-18-CE45-0008 de l'Agence Nationale de la Recherche en 2018 Projet RUGBI

Références

- ANDRÉ C, GHIO A, CAVÉ C, TESTON B . (2003) PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. International Congress of Phonetic Sciences (ICPhS), Barcelona, Spain. pp.1421-1424.
- ASTÉSANO C. , BALAGUER M., FARINAS J., FREDOUILLE C., GAILLARD P., GHIO A., GIUSTI L. et al. (2018), Carcinologic Speech Severity Index Project: A Database of Speech Disorders Productions to Assess Quality of Life Related to Speech After Cancer, LREC, 7-12 May 2018, Miyazaki (Japan)
- AUZOU P, ROLLAND-MONNOURY V. (2006), Batterie d'évaluation de la dysarthrie, 1st ed. Isbergues:
- BALAGUER, M., BOISGUÉRIN, A., GALTIER, A., GAILLARD, N., PUECH, M., & WOISARD, V. (2019). Assessment of impairment of intelligibility and of speech signal after oral cavity and oropharynx cancer. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 136(5), 355–359
- CHOMSKY, N (1957). Syntactic Structures. The Hague/Paris: Mouton. p. 15
- DELACOUR H, SERVONNET A, PERROT A, VIGEZZI JF, RAMIREZ JM (2005) La courbe ROC (receiver operating characteristic) : principes et principales applications en biologie clinique. *Annales de Biologie Clinique*. 2005;63(2):145-154.
- FONTAN L, TARDIEU J, GAILLARD P, WOISARD V, RUIZ R. (2015) Relationship Between Speech Intelligibility and Speech Comprehension in Babble Noise, *Journal of Speech Language and Hearing Research*. 2015 Jun;58(3):977-86. https://doi.org/10.1044/2015_JSLHR-H-13-0335
- GHIO A., GIUSTI L., BLANC E., PINTO S., LALAIN M, ROBERT D., FREDOUILLE C., WOISARD V. (2016) Quels tests d'intelligibilité pour évaluer les troubles de production de la parole ?. *Journées d'Etude sur la Parole*, Paris, France, p.589-596
- GHIO A, LALAIN M, GIUSTI L, POUCHOULIN G, ROBERT D, et al.(2018). Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique. XXXIIe Journées d'Etudes sur la Parole, LPL, 2018, Aix-en-Provence, France. pp.285-293,
- KEINTZ, C. K., BUNTON, K., & HOIT, J. (2007). Influence of visual information on the intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 16, 222–234.
- KENT RD, WEISMER G, KENT JF, ROSENBK JC.(1989) Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*. 1989 Nov;54(4):482-99.
- LAARIDH I, FREDOUILLE C, GHIO A, LALAIN M, , et al (2018). Automatic Evaluation of Speech Intelligibility Based on i-vectors in the Context of Head and Neck Cancers. *Interspeech*, 2943-2947
- LINDBLOM, B. (1990). On the communication process: Speaker listener interaction and the development of speech. *Augmentative and Alternative Communication*, 6, 220–230.
- NOCAUDIE O, ASTÉSANO C, GHIO A, LALAIN M, WOISARD V (2018). Evaluation de la compréhensibilité et conservation des fonctions prosodiques en perception de la parole de patients post traitement de cancers de la cavité buccale et du pharynx. 32e JEP, pp.196-204,
- R CORE TEAM (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- REBOURG M, LALAIN M, GHIO A, FREDOUILLE C, FAKHRY N, WOISARD V, (2020), Évaluer l'intelligibilité, mots ou pseudo-mots ? Comparaison entre deux groupes d'auditeurs, JEP, ce volume
- WARREN, R.M., WARREN, R.P., (1970). Auditory illusions and confusions. *Scientific American*. 223(6), 30–36.