



HAL
open science

Jugement d'altération de l'intelligibilité et de sévérité d'un trouble de la production de la parole séquellaire d'un cancer de la cavité buccale ou de l'oropharynx

Mathieu Balaguer, Aline Boisguérin, Anaïs Galtier, Nadège Gaillard, Michèle Puech, Virginie Woisard

► To cite this version:

Mathieu Balaguer, Aline Boisguérin, Anaïs Galtier, Nadège Gaillard, Michèle Puech, et al.. Jugement d'altération de l'intelligibilité et de sévérité d'un trouble de la production de la parole séquellaire d'un cancer de la cavité buccale ou de l'oropharynx. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 2019, 136 (5), pp.355 - 359. 10.1016/j.anorl.2019.05.012 . hal-02505359

HAL Id: hal-02505359

<https://hal.science/hal-02505359>

Submitted on 13 Mar 2020

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

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

Jugement d'altération de l'intelligibilité et de sévérité d'un trouble de la production de la parole séquellaire d'un cancer de la cavité buccale ou de l'oropharynx

Mathieu Balaguer¹, Aline Boisguérin¹, Anaïs Galtier^{1,2}, Nadège Gaillard¹, Puech Michèle^{1,2}; Virginie Woisard^{1,2,3}

(1) CHU Larrey, Toulouse, France

(2) Oncopole, Toulouse, France

(3) Laboratoire Octogone-Lordat, France

RESUME :

But : le jugement perceptif est une méthode d'évaluation de la sévérité des troubles de la parole dans la pratique clinique. Bien que cette méthode soit d'une fiabilité et d'une reproductibilité limitée, elle reste pour des raisons de facilité d'usage une méthode très utilisée. Ainsi, le choix de la consigne et de l'échantillon de parole support de l'évaluation restent des questions pertinentes.

Objectif : comparer les résultats des jugements perceptifs d'un jury sur deux tâches avec deux consignes différentes.

Matériel et méthode : notre corpus est constitué d'enregistrements de 87 personnes traitées pour un cancer de la cavité buccale et l'oropharynx, jugés par 6 experts sur 2 consignes (altération de l'intelligibilité versus altération du signal de parole) et 2 échantillons de parole différents (parole semi-spontanée versus lue).

Résultats : bien que ces différentes tâches donnent des résultats fortement corrélés ($r > 0,8$), la consigne visant l'altération du signal de parole favorise une distribution des scores offrant une meilleure métrique. La sévérité est en moyenne plus impactée par une localisation orale (5.44, sd 2.47) qu'oropharyngée (6.46, sd 2.24). La parole semi-spontanée tend à réduire l'effet plafond de la mesure de sévérité comparativement à la parole lue (scores moyens à 6.06/10 en description d'image, et 6.51/10 en lecture).

Conclusion : la mesure perceptive de la sévérité du trouble de la parole sur de la parole semi-spontanée semble le score cliniquement le plus pertinent dans l'évaluation des troubles de la parole post-traitement carcinologique de la cavité buccale ou de l'oropharynx.

MOTS-CLES (3 à 5) :

Parole, cancer de la cavité buccale, cancer de l'oropharynx, évaluation

Speech, oral cavity cancer, oropharyngeal cancer, assessment

INTRODUCTION

Les difficultés d'évaluation quantitative des conséquences sur la parole des cancers des voies aérodigestives sont une limite à l'intégration du risque fonctionnel dans les protocoles thérapeutiques. Or le développement récent des outils de télécommunication rend la communication orale essentielle à l'échange inter-individus.

Dans le même temps, le cancer des voies aérodigestives supérieures représente le cinquième cancer le plus fréquent en France (HAS 2009). De par la localisation tumorale, il a fréquemment un impact défavorable sur les capacités de communication orale des patients, en altérant la voix ou l'articulation et grevant sa qualité de vie [1].

L'intelligibilité est l'élément essentiel à l'échange communicationnel, et elle est définie comme le déficit fonctionnel de la parole réduisant la capacité à agir avec autrui [2]. Autrement dit, il s'agit du degré avec lequel le message d'un locuteur peut être compris par un auditeur, soit la « proportion de parole comprise », ou, lors d'une évaluation, le « taux de mots correctement retranscrits » [3,4,5]. Elle dépend donc d'un jugement d'autrui, ce qui induit une difficulté de mesure, en lien avec les stratégies de compensation humaines permettant un meilleur décodage du signal grâce à la structure de la langue et le sens du message [6].

C'est pour cela que la sévérité du déficit de parole est un autre indice utilisé en clinique. Elle fait appel au degré d'altération de l'intelligibilité associé à d'autres éléments du signal de parole tels que la qualité de l'émission du code (code acoustico-phonétique), la mesure du débit de la parole et d'autres paramètres temporels et/ou prosodiques en rapport avec la gêne ressentie [7,8,9]. La sévérité du trouble est une notion plus large englobant l'intelligibilité et les stratégies de compensation.

L'évaluation des troubles de la parole, réalisée par des orthophonistes, repose sur des bilans perceptifs. Un des outils les plus utilisés est la BECD (Batterie d'Évaluation Clinique de la Dysarthrie, [10]). Ces évaluations perceptives peuvent se situer à deux niveaux [11] :

- au niveau de la production du phonème : le phonème perçu par l'auditeur correspond-il à celui attendu. Nous nous situons ici au niveau du décodage acoustico-phonétique, c'est-à-dire à l'intelligibilité ;

- au niveau du discours, lors de la production continue de parole (« running speech intelligibility »), où les mécanismes top-down entrent en jeu.

L'évaluation de la parole doit être double, à la fois qualitative pour déterminer les mécanismes en jeu dans la production de l'altération de la parole, mais également quantitative pour scorer le déficit. Deux stratégies peuvent alors être utilisées : l'utilisation d'une échelle de cotation de type visuelle analogique (allant de la perception d'un trouble majeur à une absence de trouble) ou l'utilisation de données mesurées (par exemple, le taux d'erreurs entre phonèmes attendus et phonèmes perçus).

Toutefois, ces épreuves perceptives présentent un inconvénient : la fiabilité et la reproductibilité sont auditeur-dépendantes [11]. En effet, le degré de familiarité de l'auditeur avec le patient ou avec la tâche proposée peut augmenter la prédictibilité, et les résultats obtenus entre un juré expert ou familier et un juré naïf peuvent alors être différents. D'autre part, leur reproductibilité est là encore sujette à des variations. Malgré ces limites, ce sont bien ces épreuves qui sont les plus utilisées en pratique clinique courante, principalement pour des raisons de facilité d'usage.

Dans cette optique, nous avons cherché à savoir quelles étaient les épreuves, réalisables en pratique clinique courante, les plus pertinentes dans l'évaluation perceptive de la parole pour avoir une donnée de référence permettant le contrôle de l'évolution du trouble.

Notre objectif est de comparer les résultats des tâches de perception de la sévérité du trouble et de l'altération de l'intelligibilité dans deux situations différentes : celle de la lecture d'un texte et celle d'un discours semi-dirigé.

Notre hypothèse est que la mesure de la sévérité réalisée à partir d'un discours semi-dirigé ayant pour support la description d'une image est plus pertinente en clinique qu'une mesure de l'intelligibilité à partir d'un texte lu.

MATÉRIEL ET MÉTHODES

Présentation du corpus

Notre protocole d'étude a été mis en place dans le cadre du projet C2SI (Carcinologic Speech Severity Index) [12]. Son objectif était d'obtenir une mesure de l'impact des traitements des cancers des voies aérodigestives supérieures (cavité

buccale et pharynx) sur la production de parole par traitement automatique de la parole en référence aux méthodes perceptives.

Notre corpus est constitué de patients vus en consultation de suivi post-traitement d'un cancer de la cavité buccale ou de l'oropharynx entre 2015 et 2016 à l'Institut Universitaire du Cancer Oncopole de Toulouse.

Pour être inclus, les patients devaient avoir terminé le protocole thérapeutique depuis 6 mois, être en rémission clinique afin que leur trouble de parole (qu'il soit perceptible à l'oreille ou non pour mesurer les déficits les plus légers) soit le plus stable possible. Ces éléments nous ont permis de nous situer dans l'étude du trouble chronique.

Ont été exclus les patients porteurs d'un trouble de la parole potentiellement lié à une autre pathologie, par exemple après un accident vasculaire cérébral ou atteint d'un trouble de la fluence type bégaiement.

Présentation des épreuves

Nous avons enregistré la parole des patients sur une tâche de lecture (LEC) et une tâche de description d'image (DES).

L'ensemble des enregistrements s'est déroulé à l'Institut Universitaire du Cancer Toulouse – Oncopole, au sein de l'unité d'oncoréhabilitation. Lors d'une consultation de suivi, les patients ont été conduits dans une cabine d'enregistrement insonorisée où les tâches d'évaluation de la parole leur ont été proposées. Les fichiers audios de format WAVE ont été enregistrés sur un enregistreur numérique avec un microphone et un filtre popshield, de façon à obtenir une qualité optimale d'enregistrement et limiter un biais de mesure (mauvaise mesure du trouble en raison de l'enregistrement).

La tâche de lecture (LEC)

Les patients devaient lire à voix haute le premier paragraphe de « La chèvre de Monsieur Seguin » d'Alphonse Daudet : « *Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon. Un beau matin, elles cassaient leur corde, s'en allaient dans la montagne, et là-haut le loup les mangeait. Ni les caresses de leur maître, ni la peur du loup, rien ne les retenait. C'était, paraît-il, des chèvres indépendantes voulant à tout prix le grand air et la liberté* ».

Cet extrait a été choisi en raison de sa durée et de sa constitution phonétique équilibrée, et il est largement utilisé dans les évaluations cliniques en France [6].

La tâche de description d'image (DES)

Les patients devaient tirer au sort une photographie parmi plusieurs, qui représentaient toutes un paysage de bord de mer sur lequel figuraient des bateaux.

Il leur était ensuite demandé de décrire cette image pour qu'un auditeur puisse à distance la dessiner selon leurs instructions. La tâche de description devait durer au moins une minute, de façon à avoir suffisamment de matériel pour une évaluation par le jury d'écoute.

Jurys d'écoute

Une fois le corpus constitué, les enregistrements ont été analysés par un jury d'écoute, constitué de six professionnels orthophonistes qualifiés d'experts, car exerçant en service d'ORL.

Analyse de la tâche DES

L'enregistrement anonymisé de chaque patient a été transmis à chaque juré. Ce dernier devait l'écouter au casque dans un environnement calme.

Les jurés étaient prévenus qu'ils allaient écouter des enregistrements de personnes décrivant une photographie ayant un rapport avec le milieu maritime. Une fois le fichier écouté, ils devaient :

- évaluer l'intelligibilité et la sévérité sur une échelle de 0 à 10 (plus le score est faible, plus le trouble est perçu comme important) au moyen d'un formulaire à compléter permettant de coter par palier de 1 (notation discrète). À noter qu'une définition des deux termes était proposée dans la consigne donnée aux jurés : « *L'intelligibilité est définie comme la compréhensibilité du message véhiculé par le signal, alors que la sévérité est définie comme le degré d'altération globale du signal sonore* ».

- évaluer le degré de difficulté à comprendre les informations données par le locuteur au moyen d'une échelle visuelle analogique et en s'aidant des repères proposés. Les jurés devaient positionner une barre sur une ligne allant de « *je n'ai rien compris* » à « *j'ai tout compris* » (intermédiaires : « *j'ai compris quelques mots mais pas le sens du message* » et « *j'ai compris le sens global mais pas tous les mots* »), ce qui a permis le calcul d'un score quantitatif continu.

Ainsi, pour la tâche DES, nous avons par patient trois scores globaux (un score de sévérité, d'intelligibilité et de degré de difficulté que nous appellerons « intelligibilité graduée »).

Analyse de la tâche LEC

Cette analyse a été réalisée 6 mois après celle de description d'image. Le déroulement de l'analyse perceptive de la tâche LEC est similaire à celle de la tâche DES, à l'exception de l'absence d'évaluation de l'intelligibilité graduée (au moyen de l'échelle visuelle analogique).

Pour la tâche LEC, ont donc été obtenus deux scores (sévérité et intelligibilité) par juré pour chaque patient.

Traitement des données

La description de la distribution, les corrélations entre les différentes mesures et la recherche de différence entre les scores donnent des informations sur la qualité métrique de chaque mesure.

Pour cela, après une description de la population, le résultat de l'évaluation perceptive étant dépendant du jury qui la réalise, nous avons cherché à estimer la fiabilité inter-juges par le calcul du coefficient de corrélation interclasse (ICC) qui exprime la proportion de variance liée à la variabilité inter-patients exempte d'erreurs de mesure. Autrement dit, l'ICC permet d'évaluer la proportion de variance "expliquée" par un facteur. Dans notre cas, le coefficient de corrélation interclasse mesurera la proportion de variabilité du score perceptif liée aux variabilités interjuges par rapport à la variabilité globale du score perceptif : $ICC = \text{(variabilité inter-juges)} / \text{(variabilité inter-juges + erreur aléatoire)}$. Ainsi, plus l'ICC sera élevé, moins l'erreur aléatoire sera importante, et plus les scores inter-juges seront homogènes. Des normes permettant son interprétation sont proposées par différents auteurs, retrouvant une fiabilité modérée lorsque l'ICC est compris entre 0.5 et 0.75, et bonne lorsqu'elle est supérieure à 0.75 [13].

Les résultats des cotations des jurés est présenté par la moyenne des scores attribués par les jurés. La normalité ou non de leur distribution est vérifiée par le test de Shapiro-Wilk, la vérification de l'homoscédasticité des résidus par le test de Breusch-Pagan. La description des résultats aux différentes épreuves a été réalisée à l'aide d'une représentation en boîtes à moustache. La comparaison et les corrélations ont été effectuées au moyen de tests non-paramétriques si nécessaire : test de Wilcoxon et coefficient de corrélation de Spearman respectivement. Une correction de Bonferroni a été appliquée lors de la réalisation des tests multiples pour prendre en compte l'inflation du risque de première espèce.

L'ensemble des analyses a été réalisé au moyen du logiciel Stata 14.2.

3. RÉSULTATS

Population

87 patients ayant été traités pour un cancer de la cavité buccale ou du pharynx ont été enregistrés (51 hommes et 36 femmes), d'âge moyen 65,8 ans (valeurs extrêmes : 36 – 87 ans).

35 patients (40.2%) ont présenté un cancer de la cavité buccale, parmi lesquels 16 étaient localisés au niveau du plancher buccal, 8 au niveau de la langue, 6 au niveau rétromolaire et enfin 5 au niveau mandibulaire. 52 (59.8 %) ont présenté une tumeur oropharyngée, dont la moitié (26) se situait au niveau amygdalien, 13 au niveau basilingual, 9 au niveau oropharyngé (sans précision) et 4 au niveau vélaire.

Concernant la classification tumorale, les tumeurs T2 et T4 étaient les plus représentées (respectivement 33 et 31, soit 38 % et 35 %). Les tumeurs T3 étaient représentées chez 12 patients (14 %), les T1 chez 11 patients (13 %). A noter que plus la valeur T de la classification TNM est élevée, plus la tumeur est volumineuse.

73 patients (83.9%) ont bénéficié d'un traitement chirurgical, associé à un traitement radiothérapique ou chimiothérapique ou à une chirurgie d'exérèse ganglionnaire. 14 patients ont été traités soit par radiothérapie exclusive (pour 5 d'entre eux), soit par une association radio-chimiothérapie pour 9 d'entre eux.

Le délai post-traitement lors des enregistrements était de 64.2 mois en moyenne (valeurs extrêmes : 6 – 239) avec une médiane à 39 mois.

Évaluation de la fiabilité inter-juges

Le résultat de l'évaluation perceptive décrite par le coefficient de corrélation interclasse (ICC) est résumé dans le tableau 1.

Tableau 1 : coefficient de corrélation interclasse pour chaque épreuve

	ICC	Intervalle de confiance à 95 %
Sévérité (DES)	0.69	0.62 – 0.77
Intelligibilité (DES)	0.77	0.70 – 0.83
Intelligibilité graduée (DES)	0.74	0.67 – 0.82
Sévérité (LEC)	0.69	0.61 – 0.77
Intelligibilité (LEC)	0.43	0.32 – 0.55

Le degré de concordance entre les cotations des jurés est donc modéré à bon (>0.69) pour l'ensemble des tâches, à l'exception de la tâche d'intelligibilité en lecture qui est plus faible (0.43).

Le jury constitué par les six jurés orthophonistes experts est donc homogène.

Présentation des résultats obtenus pour chaque score

Les résultats des cotations des jurés se retrouvent dans le tableau 2. Le résultat présenté consiste en la moyenne des scores attribués par les jurés.

Tableau 2 : résultats des cotations des jurés (l'écart-type n'est pas indiqué, aucune distribution ne suivant de distribution normale : p-value (test de Shapiro-Wilk) < 0.01).

	Moyenne	Médiane	Min	Max	Intervalle interquartile
Sévérité (DES)	6.06	6.33	0.58	9.67	4.2 – 8
Intelligibilité (DES)	7.61	8.67	1.2	10	6.83 – 9.5
Intelligibilité graduée (DES)	7.96	9.26	1.16	10	6.7 – 9.9
Sévérité (LEC)	6.51	7.02	1.03	10	5.03 – 8.1
Intelligibilité (LEC)	8.69	9	3.5	10	8.4 – 9.5

Le détail des scores moyens obtenus pour chaque cotation en fonction de la taille de la tumeur et en fonction de la région anatomique atteinte est donné dans le tableau 3.

Tableau 3 : scores moyens issus des différentes cotations des jurés selon la taille de tumeur et la région anatomique atteinte, les écarts-types sont indiqués entre parenthèses (Tx : critère « T » taille de la tumeur selon la classification TNM ; CB = cavité buccale ; OP = oropharynx)

	Critère « T » (classification TNM)				Région anatomique atteinte	
	T1	T2	T3	T4	CB	OP
Sévérité (DES)	6.77 (1.71)	6.95 (2.01)	5.62 (2.83)	5.02 (2.39)	5.44 (2.47)	6.46 (2.24)
Intelligibilité (DES)	8.71 (1.03)	8.55 (1.85)	6.95 (3.14)	6.47 (2.71)	6.83 (2.70)	8.12 (2.23)
Intelligibilité graduée (DES)	9.13 (1.04)	9.05 (1.40)	7.62 (3.29)	6.49 (2.86)	7.35 (2.78)	8.36 (2.31)
Sévérité (LEC)	6.67 (1.79)	7.53 (1.69)	5.88 (2.64)	5.62 (2.16)	6.10 (2.31)	6.77 (2.03)
Intelligibilité (LEC)	8.99 (0.56)	9.23 (0.60)	8.15 (1.56)	8.21 (1.71)	8.40 (1.60)	8.87 (1.10)

Les scores moyens de sévérité sont plus faibles que ceux d'intelligibilité. Il existe également un effet de la taille de la tumeur, avec des scores plus faibles pour les tumeurs évoluées que précoces quelle que soit l'épreuve (les tumeurs évoluées sont les T3 et T4, tandis que les tumeurs précoces sont classées en T1 ou T2 : Dwivedi, 2012). Enfin, une atteinte de l'oropharynx montre des scores plus élevés qu'en cas d'une atteinte de la cavité buccale.

L'analyse de la distribution des valeurs montre un effet plafond au niveau des valeurs les plus importantes (c'est-à-dire avec trouble absent ou limité), effet qui semble nettement plus marqué dans les tâches d'intelligibilité que dans les tâches de sévérité. L'étalement de la mesure est plus étendu pour la sévérité que pour l'intelligibilité (figure 1).

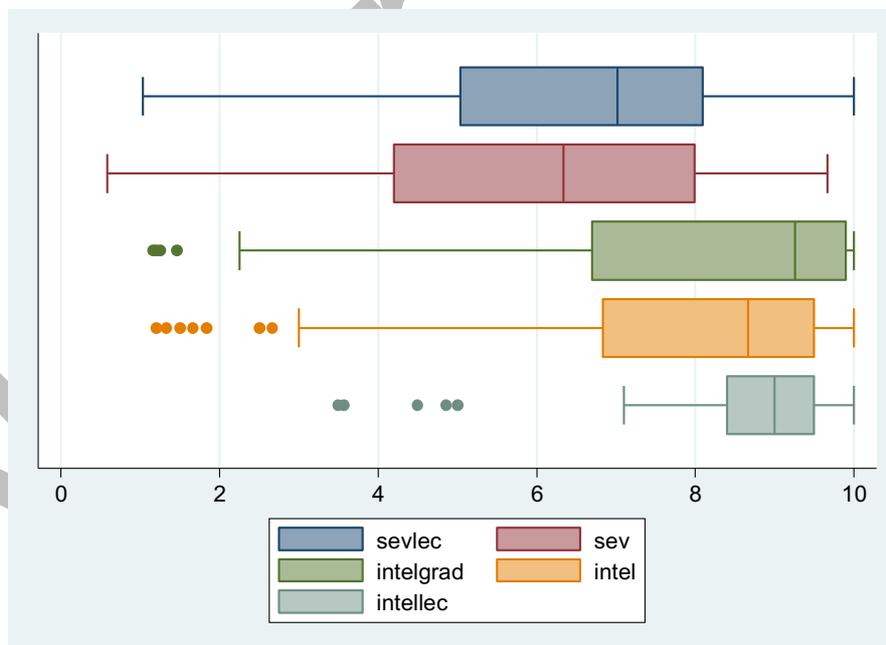


Figure 1 : boîte à moustache des distributions des scores (sevlec = sévérité évaluée en lecture ; sev = sévérité évaluée en description d'image ; intelgrad = intelligibilité évaluée avec l'échelle visuelle analogique en description ; intel = intelligibilité évaluée en description ; intellec = intelligibilité évaluée en lecture)

Les résultats obtenus en intelligibilité graduée (réalisée lors de la tâche de description d'image, valeurs de cotation continues) et en intelligibilité (cotation discrète de 0 à 10) sont proches. La comparaison de ces deux moyennes en séries appariées par un test non paramétrique de Wilcoxon montre une différence significative ($p=0.001$) : la moyenne obtenue

en intelligibilité graduée est significativement plus importante que la moyenne obtenue à la cotation de l'intelligibilité (7.96 vs 7.61).

Corrélations entre chaque score moyen de sévérité et d'intelligibilité

Les corrélations sont très élevées entre les différents scores (coefficients rho de corrélation de Spearman systématiquement supérieurs à 0.72), avec une p-value systématiquement inférieure à 0.001 après correction de Bonferroni.

Sur une même tâche, les corrélations entre sévérité et intelligibilité sont au moins égales à 0.90 (en description : 0.91 ; en lecture : 0.90).

La sévérité en lecture et en description est corrélée à 0.89, tandis que l'intelligibilité dans ces deux mêmes tâches est corrélée à 0.88.

Sévérité et intelligibilité sur des tâches différentes sont corrélées entre 0.80 et 0.87.

Enfin, l'intelligibilité graduée en description d'image est corrélée à 0.81 avec la sévérité en description et à 0.89 avec l'intelligibilité là encore en description. La corrélation est un peu moindre mais toujours forte avec intelligibilité et sévérité en lecture (0.73 à 0.74)

Comparaison des scores entre les tâches de LEC et de DES

Bien que les scores de sévérité soient très bien corrélés entre les 2 tâches (figure 2), les différences de scores sont significativement différentes.

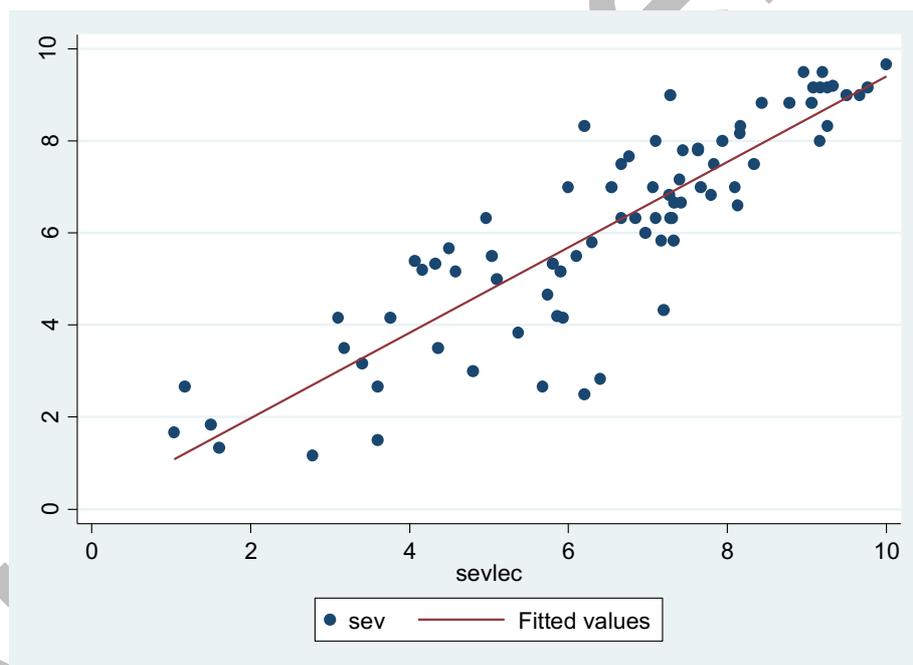


Figure 2 : nuage de points représentant les résultats en sévérité description (sev) par rapport à la sévérité en lecture (sevlec) : $r=0.89$ ($p<0.001$)

Les moyennes obtenues lors de la tâche de LEC sont significativement plus élevées que les moyennes issues de la tâche de description. Ce résultat est valable à la fois pour la sévérité ($p=0.014$) et pour l'intelligibilité ($p<0.001$). La différence est respectivement de 6,51 / 6,06 pour la sévérité et de 8.69 / 7,61 pour l'intelligibilité.

4. DISCUSSION

Peu de travaux rapportent des mesures d'intelligibilité et de sévérité de la parole dans les cancers de la cavité buccale et de l'oropharynx [14,15]. Les résultats de nos échantillons de patients confirment l'impact sur la production de la parole en lien notamment avec la taille de la tumeur initiale et la région anatomique concernée par la pathologie tumorale (une

localisation intra-orale et une tumeur T4 grèvent le plus les scores de sévérité et d'intelligibilité). Par contre, il manque encore des travaux de recherche permettant de mieux prédire les conséquences des traitements et leur impact sur la qualité de vie relative à la parole.

En ce qui concerne la manière de mesurer en pratique clinique la sévérité fonctionnelle sur la production de la parole, l'analyse des résultats montre que l'ensemble des résultats sont corrélés, quelle que soit le type d'épreuve (description d'image ou lecture), et que ce soit sur l'évaluation de la sévérité ou de l'intelligibilité.

Par contre, la fiabilité de l'évaluation inter-juges reste une limite. Dans notre échantillon de population, elle est meilleure sur une tâche de description d'image que sur une tâche de lecture. Elle paraît plus fiable sur la tâche d'intelligibilité que sur la tâche de sévérité pour la description d'image (0.77/0.69).

La distribution plus large des scores moyens sur les épreuves de sévérité comparativement à celles d'intelligibilité (que l'évaluation soit réalisées sur des valeurs discrètes ou graduées) est en faveur de l'utilisation d'une mesure de la sévérité.

L'épreuve de description d'image correspondant à une parole semi-spontanée (car peu contrainte si ce n'est par la composition de l'image et la consigne donnée au patient) montre des scores perceptifs moins soumis à l'effet plafond (vers la valeur extrême d'absence de trouble) que l'épreuve de lecture d'image (parole lue). En effet, la moyenne des scores moyens est moins élevée pour la parole semi-spontanée comme le montre la figure 3, et leur distribution plus établie.

Enfin, l'utilisation d'une échelle graduée, contrairement aux travaux antérieurs réalisés par la même équipe sur une population plus hétérogène en termes de pathologie (incluant des dysarthries neurologiques), n'est pas confirmé [16]. L'effet plafond est plus important avec un moindre étalement de la mesure.

Ainsi, l'ensemble de ces résultats tend à confirmer notre hypothèse : la mesure de la sévérité réalisée à partir d'un discours semi dirigé ayant pour support la description d'une image est plus pertinente en clinique que celle réalisée à partir d'un texte lu.

Bien que les corrélations retrouvées entre les résultats obtenus aux scores perceptifs par les patients atteints de cancer des voies aérodigestives supérieures montrent que les scores moyens des épreuves portent les mêmes informations, la mesure de l'intelligibilité présente un effet plafond, et tend à majorer les résultats de plus de 1.5 point en moyenne par rapport à la mesure de la sévérité. Le fait d'une meilleure fiabilité inter-juges pour l'intelligibilité sur la tâche de description est le seul élément pouvant faire discuter la supériorité clinique d'une mesure de la sévérité par rapport à l'intelligibilité dans une tâche de description d'image. Cependant, le coefficient de corrélation interclasse plus élevé lors des tâches d'intelligibilité peut s'expliquer par l'effet plafond et la concordance plus importante des juges ayant davantage attribué de façon coordonnée la note maximale de 10/10. De plus, la valeur à 0.69 reste suffisante pour pouvoir choisir cette mesure [17].

5. CONCLUSION

L'altération des troubles de la parole est une séquelle fréquente des cancers de la cavité buccale et de l'oropharynx. La problématique de la mesure du déficit de parole reste un challenge pour pouvoir améliorer la prise en charge des patients. Les outils utilisés en clinique courante sont très subjectifs par essence puisqu'ils dépendent de la perception d'un locuteur devant évaluer un défaut de production de la parole par un auditeur. Ce travail est en faveur dans ce contexte de l'utilisation d'une mesure de la sévérité sur un discours semi dirigé, la sévérité étant définie comme le degré d'altération globale du signal sonore et le discours étant produit à partir d'une photographie tirée au hasard parmi un set de 10 photographies représentant toutes un paysage de bord de mer sur lequel figuraient un ou plusieurs bateaux. La fiabilité inter-juges et les caractéristiques métriques de cette mesure sont compatibles avec son utilisation en clinique lors d'une consultation par des personnes ayant l'expérience de ce type de trouble. Mais les limites de ce type de mesure justifient la recherche d'un indice de trouble issu d'un traitement automatisé de la parole permettrait de s'affranchir des variabilités de mesure inter et intra-individuelles.

CONFLITS D'INTERETS

Aucun conflit d'intérêt.

REMERCIEMENTS

Ce travail fait partie du projet C2SI (Carcinologic Speech Severity Index) financé par l'Institut National du Cancer (financement n° n°2014-135) dans le cadre de projets libres de recherche en Sciences Humaines et Sociales, Épidémiologie et Santé Publique. L'investigatrice principale est Virginie Woisard (Hôpital Larrey, CHU de Toulouse).

Nous remercions les équipes médicales et paramédicales de l'Institut Universitaire du Cancer Oncopole de Toulouse pour la réalisation des enregistrements, les jurés d'écoute et les patients qui ont pris part à ce projet, ainsi que l'Institut de Recherche en Informatique de Toulouse (Université Paul Sabatier Toulouse III) qui a pris part à cet article.

Author Version

REFERENCES

- [1] Mlynarek AM, Rieger JM, Harris JR, O'Connell DA, Al-Qahtani KH, Ansari K, Chau J, Seikaly H. Methods of functional outcomes assessment following treatment of oral and oropharyngeal cancer: review of the literature. *J Otolaryngol Head Neck Surg*. 2008 Feb;37(1):2-10.
- [2] Lindblom, B. (1990b). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication*, 6(4):220–230.
- [3] Kent, R. D., Weismer, G., Kent, J. F. et Rosenbeck, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, 54(4):482–499.
- [4] Keintz, C. K., Bunton, K. et Hoit, J. D. (2007). Influence of visual information on the intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 16(3):222–234.
- [5] Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 51(3):562–573.
- [6] Ghio A., Pouchoulin G., Teston B., Pinto S., Fredouille C., De Looze C., Robert D., Viallet F., Giovanni A. (2012), How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? *Speech Communication*, vol. 54. 2012, p. 664-679.
- [7] Kent R. (1992), *Intelligibility in speech disorders*. Amsterdam/Philadelphia. John Benjamins.
- [8] Yorkston KM, Beukelman DR, Tice TR. (1996), *Sentence intelligibility test*. Lincoln, NE: Tice technology Services.
- [9] Auzou P. (2007), Les objectifs du bilan de la dysarthrie. Dans: *Les dysarthries*. Auzou P, Rolland Monnoury V, Pinto S, Öszancak C. Marseille: Solal, 189-195.
- [10] Auzou P, Rolland-Monnoury V. (2006), *Batterie d'évaluation de la dysarthrie*, 1st ed. Isbergues: Ortho Edition.
- [11] Middag, C. (2012). *Automatic analysis of pathological speech*. Ghent University, Department of Electronics and information systems, Ghent, Belgium.
- [12] Astesano 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).
- [13] Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- [14] Plisson L., Pillot-Loiseau C., Crevier-Buchman L. (2017). Intelligibilité de la parole après le traitement d'un cancer de l'oropharynx : étude descriptive chez sept patients en pré-traitement et en post-traitement précoce. 7èmes Journées de phonétique clinique (JPC7), Jun 2017, Paris, France. 2017, <<http://jpc7.ilpqa.fr/>>. <hal-01566225>.
- [15] Breil L., Crépin H., Smadja M., Crevier-Buchman L. (2006). Quality of life after oropharyngectomy. *Rev Laryngol Otol Rhinol (Bord)*. 2006;127(5):305-14.
- [16] Dittner J., Lepage B., Woisard V., Kergadallan M., Boisteaux K., Robert E., Welby-Gieusse M. (2010). Elaboration et validation d'un test quantitatif d'intelligibilité pour les troubles pathologiques de la production de parole. *Rev Laryngol Otol Rhinol*. 2010;131, 1:9-14.
- [17] Elie C., Colombet I. (2011). Méthodes d'estimation de la reproductibilité. *Note méthodologique*. *Sang Thrombose Vaisseaux*. 2011 ; 23, 3 : 138-45.