



**HAL**  
open science

## Évaluer l'intelligibilité, mots ou pseudo-mots ? Comparaison entre deux groupes d'auditeurs

Marie Rebourg, Muriel Lalain, Alain Ghio, Corinne Fredouille, Nicolas Fakhry, Virginie Woisard

### ► To cite this version:

Marie Rebourg, Muriel Lalain, Alain Ghio, Corinne Fredouille, Nicolas Fakhry, et al.. Évaluer l'intelligibilité, mots ou pseudo-mots? Comparaison entre deux groupes d'auditeurs. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1 : Journées d'Études sur la Parole, Jun 2020, Nancy, France. pp.543-551. hal-02798584v3

**HAL Id: hal-02798584**

**<https://hal.science/hal-02798584v3>**

Submitted on 23 Jun 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.

# Évaluer l'intelligibilité, mots ou pseudo-mots ? Comparaison entre deux groupes d'auditeurs

Marie Rebourg<sup>1</sup>, Muriel Lalain<sup>1</sup>, Alain Ghio<sup>1</sup>, Corinne Fredouille<sup>2</sup>, Nicolas Fakhry<sup>1,3</sup>,  
Virginie Woisard<sup>4</sup>

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

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

(3) Service ORL, APHM, La Conception, Marseille, France

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

marie.rebourg@univ-amu.fr, muriel.lalain@univ-amu.fr, alain.ghio@univ-amu.fr,  
corinne.fredouille@univ-avignon.fr, nicolas.fakhry@ap-hm.fr,  
woisard.v@chu-toulouse.fr

## RÉSUMÉ

---

La perte d'intelligibilité représente une plainte importante des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. L'évaluation de l'intelligibilité est essentielle dans le parcours de soin, mais les tests existants ne sont pas satisfaisants. Basés sur la perception de listes de mots par des auditeurs entraînés à restaurer des séquences sonores dégradées, ils conduisent souvent à une sous-évaluation des déficits. Nous avons proposé une nouvelle tâche d'évaluation de l'intelligibilité, la tâche de décodage acoustico phonétique (DAP), basée sur l'utilisation de pseudo-mots (Astésano *et al.*, 2018; Ghio *et al.*, 2018; Ghio *et al.*, soumis, Lalain *et al.*, sous presse). Dans cette étude, nous évaluons la capacité de la tâche DAP à neutraliser les effets de restauration lexicale et d'expertise auditive clinique. Les résultats montrent que contrairement à une évaluation de l'intelligibilité basée sur des mots, une évaluation basée sur des pseudo-mots permet d'obtenir des scores de Déviation Phonologique Perçue (DPP) stables au cours du temps quel que soit le degré d'expertise des auditeurs, naïfs ou cliniciens.

## ABSTRACT

---

### Assess intelligibility, words or pseudo-words? Comparison between two groups of listeners

Loss of intelligibility is a significant complaint from patients treated for Head and Neck cancer. Intelligibility assessment is essential in the treatment process, but the existing tests are not satisfactory. Based on the perception of word lists by listeners trained to restore degraded sound sequences, they often lead to an underestimation of deficits. We have proposed a new task for evaluating intelligibility, the acoustical phonetic decoding (DAP) task, based on the use of pseudo-words (Astésano *et al.*, 2018; Ghio *et al.*, 2018; Ghio *et al.*, soumis). In this study, we assess the ability of the DAP task to neutralize the effects of lexical restoration and clinical auditory expertise. The results show that, unlike a word-based intelligibility assessment, a pseudo-word assessment makes it possible to obtain stable Perceived Phonological Deviation (PPD) scores over time regardless of the level of expertise. listeners, naive or clinicians.

---

**MOTS-CLÉS** : Phonétique Clinique, Intelligibilité, Trouble de la Production de la Parole, Cancer VADS

**KEYWORDS**: Clinical phonetic, Intelligibility, Speech disorders, Head and Neck cancer

---

# 1 L'intelligibilité

La perte d'intelligibilité constitue une plainte récurrente des patients atteints de Troubles de la Production de la Parole (TPP). Dans le cadre de la prise en charge des cancers ORL, les traitements dont bénéficient les patients (chirurgie, chimiothérapie, radiothérapie) entraînent un déficit communicationnel notable. Dans le parcours de soin, l'évaluation de l'intelligibilité est essentielle puisqu'elle permet de mesurer le handicap à la communication en évaluant les composantes linguistiques préservées et/ou dégradées pour établir la prise en charge orthophonique. Elle permet également de mesurer l'effet du traitement préalablement établi dans le parcours de soin.

La définition de l'intelligibilité ne fait pas consensus dans la littérature, elle est généralement définie comme « le degré de précision avec lequel un message est compris par un auditeur » (Yorkston, Dowden et Beukelman, 1992) ; dans le cadre de notre étude nous préférons adopter le point de vue de Hustad, Jones et Dailey (2003) selon lesquels les processus de bas niveaux supportent des informations « dépendantes du signal » alors que les informations « indépendantes du signal » sont liées aux processus de haut niveau. Il apparaît essentiel de préciser le concept de compréhension, défini par Lindblom (1990), comme « L'intégration à la fois des informations acoustico-phonétiques et de toutes les informations pertinentes indépendantes du signal qui permettent de comprendre un message parlé dans une situation de communication particulière ». Cette dernière met en lien le niveau de la compréhension avec les informations de haut niveau (top down) portées par le contexte et les connaissances générales d'un auditeur. En revanche, dans le cadre de ces travaux nous définissons l'intelligibilité, au sens de Keintz, Bunton et Hoit (2007), comme « la quantité de parole comprise à partir du seul signal acoustique ». Ainsi, l'évaluation de l'intelligibilité repose sur la quantité de signal acoustique perçu et correctement identifié par un auditeur. Cette définition établit un lien étroit entre l'intelligibilité et les informations dépendantes du signal liées aux processus de décodage des sons de parole, soit le bas niveau (bottom up).

## 1.1 L'évaluation clinique de l'intelligibilité

La méthode d'évaluation utilisée par les cliniciens repose sur la perception d'un auditeur. Cette évaluation est peu satisfaisante, mais elle constitue le *gold standard* de l'évaluation (Balaguer *et al.*, 2019). La perception de la parole repose sur l'intégration d'informations de haut niveau, issues du contexte, et de bas niveau qui concernent l'identification des sons de parole. L'intégration de ces informations permet à un auditeur de restaurer les séquences sonores dégradées (Warren, 1984; Warren *et al.*, 1997) et ainsi optimise le décodage et la compréhension d'un message. La performance de ces mécanismes est essentielle à la communication dans la vie quotidienne, mais perturbe l'évaluation de l'intelligibilité en contexte clinique en entraînant une sous-évaluation des déficits.

En effet, le contexte spécifique à l'évaluation clinique augmente les effets top down. Le matériel linguistique des tests existants, Batterie d'Evaluation Clinique de la Dysarthrie (BECD) (Auzou et Rolland-Monnoury, 2006), Frenchay Dysarthry Assessment 2 (FDA2) adapté au français, (Blanc *et al.*, 2014) se composent de listes de mots courtes et fermées qui, facilement mémorisables, finissent par intégrer le stock de connaissances linguistiques générales des auditeurs. De plus, les connaissances du clinicien lui permettent de faire des prédictions sur les déficits des patients : la connaissance du dossier médical, du parcours de soin, du patient en tant que personne et de la pathologie influencent son jugement / évaluation. Enfin, une étude (Jarzé *et al.*, 2017) a mis en évidence une grande variabilité dans la perception des troubles de la parole pathologique. L'ensemble de ces facteurs, associés à la pratique et à l'utilisation répétée du test par le clinicien le conduisent à

apprendre le matériel linguistique ce qui entraîne la restauration lexicale des items et donc une sous-évaluation des déficits.

En définitive, les processus de restauration perceptive sont permis par le recours aux informations de haut niveau, accentués par le contexte spécifique de l'évaluation clinique et conduisent à une évaluation subjective. Bien que peu satisfaisante, la méthode d'évaluation de l'intelligibilité repose sur une évaluation perceptive, dont on ne peut s'affranchir, puisque par définition l'intelligibilité implique la production d'un locuteur et la perception d'un auditeur. On peut alors se demander comment rendre l'évaluation plus objective. Comment générer une évaluation qui force le recours aux informations de bas niveau et neutralise les processus de restauration lexicale ?

La solution que nous avons proposée (Ghio *et al.*, 2017) consiste à utiliser des listes de pseudo-mots pour limiter le recours aux informations de haut niveau et forcer l'auditeur à s'appuyer essentiellement sur des informations de bas niveau en le plaçant dans une situation de décodage acoustico-phonétique.

## 1.2 Le recours aux pseudo-mots

Les pseudo-mots sont des séquences qui suivent les règles phonotactiques du français. Dans le cadre du projet de recherche C2SI, plus de 80 000 formes ont été générées et permettent de constituer aléatoirement un très grand nombre de listes respectant des contraintes prédéfinies (Ghio *et al.*, 2018). Le principe consiste à faire produire une liste de pseudo-mots qui sera ensuite perçue et transcrite par un auditeur. Ex de pseudo-mots : *chouvi, granchu, sogu, vosso, zinvo, rouzant, rumo, preglo, coubin*. Nous calculons ensuite un score de Déviation Phonologique Perçue (DPP) en termes de nombre de traits moyen altérés par phonème. Ce score est issu de la comparaison entre une cible qui devait être produite et ce qui a été perçu par un auditeur. Le calcul du score DPP repose sur une matrice de confusion intégrant la distance de Levenshtein à l'algorithme de Wagner Fischer.

Les études conduites jusqu'à présent pour éprouver la robustesse de la tâche de DAP ont montré des résultats positifs (i) pour sa capacité à discriminer deux groupes de locuteurs : patients VS témoins (Ghio *et al.*, 2018), (ii) car quelle que soit la liste utilisée l'évaluation reste stable, les listes de pseudo-mots utilisées sont donc équivalentes (Ghio *et al.*, 2019), (iii) et les scores de Déviation Phonologique Perçues sont corrélés avec les évaluations globales de l'intelligibilité (Lalain, M. *et al.*, sous presse).

Il reste toutefois des questions pour valider les critères de pertinence et d'objectivité de la tâche de DAP. Ainsi, on peut se demander si une évaluation basée sur l'utilisation de pseudo-mots est plus pertinente qu'une évaluation basée sur des mots, autrement dit, quand les auditeurs sont confrontés plusieurs fois aux mêmes stimuli, sont-ils capables de les mémoriser et de les restaurer ? Enfin, l'utilisation de pseudo-mots, si elle permet une évaluation de l'intelligibilité plus objective chez des auditeurs naïfs, permet-elle de neutraliser l'expertise auditive des cliniciens ? Répondre à ces deux questions est l'objectif de l'étude que nous présentons ici.

## 2 Méthodologie

Afin de répondre à ces questions de recherche nous avons constitué un corpus composé des enregistrements de 20 locuteurs. Ce corpus rassemble le matériel linguistique que nous avons utilisé dans le protocole expérimental des tests de jugement perceptif de l'intelligibilité au cours desquels nous collectons les transcriptions de différents groupes d'auditeurs. Ces transcriptions font l'objet d'un traitement spécifique permettant d'obtenir un score de Déviation Phonologique Perçue (DPP). Ces scores (VD) seront mis en lien avec la nature du matériel linguistique (VI : Mots VS Pseudo-mots) et les groupes d'auditeurs (VI : Naïf VS Expert) dans nos analyses statistiques.

## 2.1 Corpus

Pour les besoins de cette étude 10 patients et 10 témoins ont chacun été enregistrés sur la production d'une liste de 50 mots (items de la BECD) et sur la production de 2 listes de 52 pseudo-mots issues du matériel linguistique constitué pour la tâche de DAP. Soit 154 productions par locuteur. Nous avons donc enregistré 154 stimuli x 20 locuteurs, soit 3080 stimuli en tout et respectivement 1000 stimuli pour le corpus BECD et 2080 stimuli pour le corpus DAP. Les locuteurs ont été divisés en deux groupes (A et B) de 10 locuteurs dans lesquels on retrouve 5 patients et 5 témoins, impliquant que les corpus BECD et DAP ont eux aussi été divisés en deux sous corpus composés chacun de 500 stimuli BECD et 1040 stimuli DAP. Ceci afin de limiter le nombre de stimuli transcrit par les auditeurs.

A partir de chaque sous corpus, des listes de stimuli ont été constituées. Nous obtenons alors 6 listes de productions pour le corpus BECD avec 167 stimuli par liste et 12 listes pour le corpus DAP avec 174 stimuli par liste. Les listes 1 à 3 du corpus BECD et les listes 1 à 3 et 7 à 9 du corpus DAP sont produites par le groupe de locuteurs A. Les listes 4 à 6 du corpus BECD et listes 4 à 6 et 10 à 12 du corpus DAP ont été produites par le groupe de locuteurs B. Ces listes ont été utilisées pour effectuer des tests de jugement perceptif de l'intelligibilité auprès de deux populations d'auditeurs, naïfs et cliniciens.

## 2.2 Design expérimental

Dans ces tests, l'auditeur écoute dans un casque les productions de patients (traités pour un cancer des VADS) et de témoins, et transcrit, sur un clavier, ce qu'il entend.

Chaque auditeur écoute et transcrit 3 listes BECD et 6 listes DAP, produites par un groupe de locuteurs (A ou B). Chaque liste est transcrite par trois auditeurs différents et l'ordre de présentation des listes est contrôlé de manière que chaque liste apparaisse dans chaque position possible (T1, T2 et T3) et soit transcrite par trois auditeurs différents. Nous obtenons donc 3 transcriptions, proposées par 3 auditeurs différents, et ce dans chacune des positions (T1, T2 et T3), soit au total 9 transcriptions par liste. L'ordre de passation des listes est résumé dans le tableau de design expérimental suivant :

			T1	T2	T3
Groupe de locuteurs A	3 auditeurs ≠	BECD	1	2	3
		DAP 1	1	2	3
		DAP 2	7	8	9
	3 auditeurs ≠	BECD	2	3	1
		DAP 1	3	1	2
		DAP 2	9	7	8
	3 auditeurs ≠	BECD	3	1	2
		DAP 1	2	3	1
		DAP 2	8	9	7
Groupe de locuteurs B	3 auditeurs ≠	BECD	4	5	6
		DAP 1	4	5	6
		DAP 2	10	11	12

Chaque auditeur transcrit 3 blocs de stimuli (BECD, DAP1 puis DAP2)  
Chaque bloc comprend 3 listes de stimuli

L'ordre de passation des listes au cours de la tâche est contrôlé. Ex : La liste 1 est transcrite en T1, puis, T2 et T3

3 auditeurs ≠	BECD	6	4	5
	DAP 1	5	6	4
	DAP 2	11	12	10
3 auditeurs ≠	BECD	5	6	4
	DAP 1	6	4	5
	DAP 2	12	10	11

TABLE 1 : Tableau récapitulatif du design expérimental du test de jugement perceptif

Ce protocole expérimental a permis de réaliser des tests de jugement perceptif de l'intelligibilité auprès de deux groupes de population.

### 2.3 Test de jugement perceptif

18 auditeurs naïfs, natifs de langue française, sans problème de vue ou d'audition non corrigés, ayant un bon niveau en orthographe et 18 auditeurs experts en écoute de la parole pathologique, tous orthophonistes, ont participé à cette étude. Les auditeurs naïfs ont réalisé la tâche en débutant par la transcription orthographique des mots issus de la BECD puis en transcrivant les listes de pseudo-mots du DAP. Pour des questions d'organisation les auditeurs experts ont réalisé la tâche en débutant par la transcription des listes de pseudo-mots du DAP, puis ont poursuivi l'exercice en transcrivant les mots issus de la BECD. Les auditeurs ont reçu pour consigne de toujours donner une transcription, au plus près de ce qu'ils perçoivent et identifient, en suivant les règles orthographiques du français. Ils ont été dédommagés en ticket Kadeos. Ce design expérimental supporte un double objectif, il permet de comparer les scores d'intelligibilité obtenus par la production et la transcription de deux types de matériel linguistique : des mots et des pseudo-mots ; et de comparer les scores obtenus par les transcriptions de deux catégories d'auditeurs : des auditeurs naïfs et experts. Ce test de jugement perceptif de l'intelligibilité a été mené au CEP (<http://cep.lpl-aix.fr/>), à l'aide de la station de perception PercEval (André *et al.*, 2003).

### 2.4 Traitement des données

Pour réaliser l'analyse statistique des données collectées lors des tests de jugement perceptif de l'intelligibilité, plusieurs étapes sont nécessaires.

Tout d'abord, nous récupérons les données brutes au sortir du logiciel d'expérimentation PercEval (André *et al.*, 2003). Nous réalisons différentes opérations de prétraitement des données visant à leur donner un format compatible avec les outils de calcul du score DPP. En premier lieu nous recodons les transcriptions orthographiques des auditeurs : si un auditeur propose une transcription avec une lettre accentuée (e.g : « é ») elle est encodée par un symbole dans le fichier de réponse (e.g : « %E9 »), nous remplaçons les symboles du fichier réponse par la lettre correspondante transcrite initialement par l'auditeur. Ensuite nous transformons ces transcriptions orthographiques en phonétique, c'est l'étape de phonétisation ((LIA\_Phon, Béchet, 2001) et Lexique.org). Ces transcriptions phonétisées, compatibles avec la matrice de confusion utilisée pour calculer les scores de Déviation Phonologique Perçue (DPP), sont utilisées pour l'analyse des données.

Nous comparons alors les transcriptions phonétiques des réponses des auditeurs aux tests de jugement perceptif de l'intelligibilité avec les transcriptions phonétiques des cibles qui devaient être prononcées par les locuteurs. L'algorithme de Wagner Fischer permet le calcul de distance d'édition entre deux

chaînes de caractères phonétiques. Nous utilisons une matrice de confusion qui attribue un coût en termes de distance basé sur la théorie des traits, développé par A. Ghio dans le cadre de l'analyse des données DAP (Ghio *et al.*, 2018). L'algorithme intègre la distance de Levenshtein, qui considère trois opérations d'éditations élémentaires : la suppression, l'insertion ou la substitution d'un caractère, ce qui permet de traiter les altérations sur les deux axes syntagmatique et paradigmatique. Soit, une distance entre les transcriptions des cibles qui devaient être prononcées par les locuteurs et les transcriptions effectives des réponses des auditeurs.

Les scores de distance cumulée sont obtenus en divisant le score donné par la matrice de confusion pour chaque stimulus par le nombre de caractères de la cible phonétique. Nous obtenons alors des scores de Déviation Phonologique Perçue (DPP) en termes de distance cumulée à la cible, qui représentent le nombre moyen de traits altérés par phonème. Ces scores sont les variables dépendantes (VD) des analyses statistiques. De plus, pour chaque auditeur, un numéro a été assigné à chaque item en fonction de l'ordre de la passation, ainsi, le premier stimulus transcrit est numéroté 1, le second stimulus transcrit est numéroté 2, et ainsi de suite jusqu'à 500 pour les items lexicaux BECD et jusqu'à 1040 pour les items non-lexicaux DAP.

Pour l'analyse statistique, nous comparerons les scores DPP à l'aide d'une analyse de variance appliquée afin de tester les différences significatives des scores d'intelligibilité (VD) en fonction du matériel d'évaluation (Variables Indépendantes (VI) : Mots vs Pseudo-mots) et en fonction du groupe d'auditeurs (VI : Naïfs vs Experts).

### 3 Résultats et discussion

Une analyse de variance (ANOVA) a été conduite, afin de tester les effets simples et les effets d'interaction entre les différents facteurs et ainsi de déterminer les liens entre les variables testées. Nous avons utilisé le score DPP en VD et en VI le groupe d'auditeurs (facteur 2 niveaux : Experts vs Naïfs), le type de matériel linguistique (facteur 2 niveaux : Mots BECD vs Pseudo-mots DAP), et le groupe de locuteurs (Patients et Témoins) en facteur. Les résultats montrent un effet simple des groupes de locuteurs, d'auditeurs et du matériel linguistique ( $p < 0.001$ ). Cela signifie que, dans nos données, les groupes de locuteurs sont distingués, tout comme les groupes d'auditeurs et type de matériel linguistique. Les résultats révèlent également un effet d'interaction significatif entre Locuteur et Auditeur ( $p = 0.027$ ) et entre Matériel linguistique et Auditeur ( $p = 0.047$ ). Cela montre que les locuteurs sont évalués différemment par les deux groupes d'auditeurs et que les groupes d'auditeurs ne proposent pas des scores équivalents avec les deux matériaux linguistiques. Cependant, ils ne montrent pas d'effet d'interaction Locuteur et Matériel linguistique ( $p = 0.44$ ), donc les deux groupes de locuteurs ne sont pas évalués différemment par les matériaux linguistiques. En revanche, ils révèlent également un effet d'interaction Locuteur, Auditeur et Matériel linguistique ( $p < 0.001$ ) ; les groupes d'auditeurs évaluent différemment les groupes de locuteurs en fonction du matériel linguistique utilisé.

Pour établir le sens du lien entre nos variables nous avons établi une matrice de corrélation qui intègre l'ordre de passation des items (VI : Order) à la place du groupe de locuteurs. Les résultats révèlent un effet de l'ordre des items dans la passation pour le matériel linguistique lexical (Mots BECD) et ce pour les deux groupes d'auditeurs, Expert ( $r = -0.197$ ,  $p < 0.001$ ) et Naïfs ( $r = -0.302$ ,  $p < 0.001$ ) (FIGURE 1). En revanche, aucun effet n'est mis en exergue dans une évaluation basée sur l'utilisation de pseudo-mots (DAP) et ce pour les deux groupes d'auditeurs, Experts ( $r = -0.058$ ,  $p = 0.06$ ) et Naïfs ( $r = -0.028$ ,  $p = 0.37$ ) (FIGURE 3).

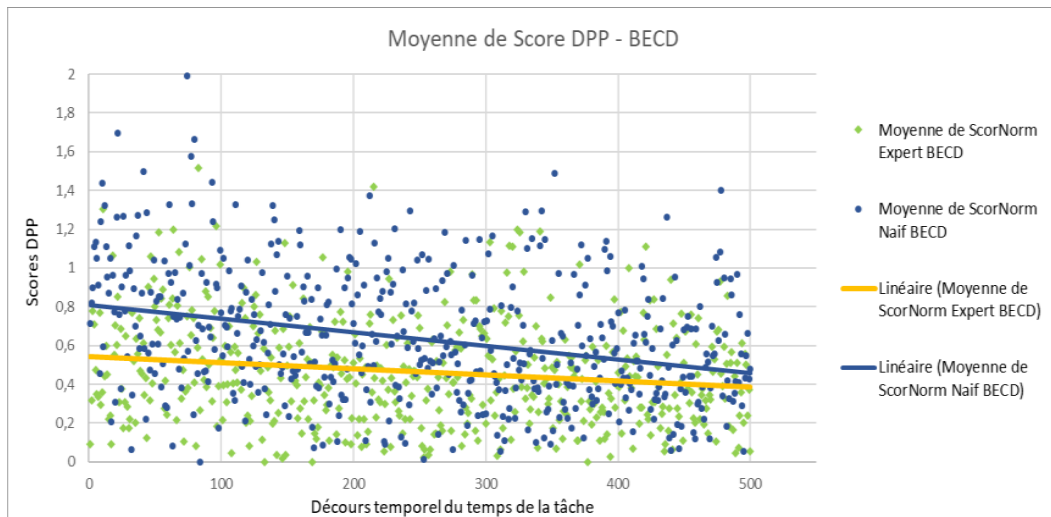


FIGURE 1 : Scores DPP attribués par une évaluation BECD par deux types d'auditeurs

Ceci indique que, dans une évaluation utilisant du matériel linguistique lexical (Mots BECD) il existe un effet d'apprentissage du matériel linguistique au cours de la tâche. L'accès au lexique mental permis par l'emploi de mots favorise la mémorisation des items et leur restauration perceptive lorsque leur production est dégradée. Les listes de mots courtes et fermées classiquement utilisées (Auzou et Rolland-Monnoury, 2006; Blanc *et al.*, 2014) facilitent ces effets. Défavorables dans le cadre de l'évaluation clinique de l'intelligibilité, car conduisant à la sous-évaluation des déficits, ces effets se trouvent neutralisés dans une évaluation s'appuyant sur du matériel linguistique non lexical (Pseudo-mots DAP). En effet, les pseudo-mots DAP permettent d'obtenir une évaluation stable au cours du temps, critère qui valide la fiabilité de la tâche et du matériel linguistique, et de neutraliser l'effet d'apprentissage et de mémorisation, et donc les effets de restauration perceptive liés au matériel linguistique.

Ces résultats confirment l'hypothèse selon laquelle l'emploi de pseudo-mots est pertinent pour évaluer l'intelligibilité (au sens de la définition de Keintz, Bunton et Hoit (2007)) et qu'il permet de neutraliser l'effet d'apprentissage du matériel linguistique, et ce, pour les deux groupes d'auditeurs, naïfs et cliniciens.

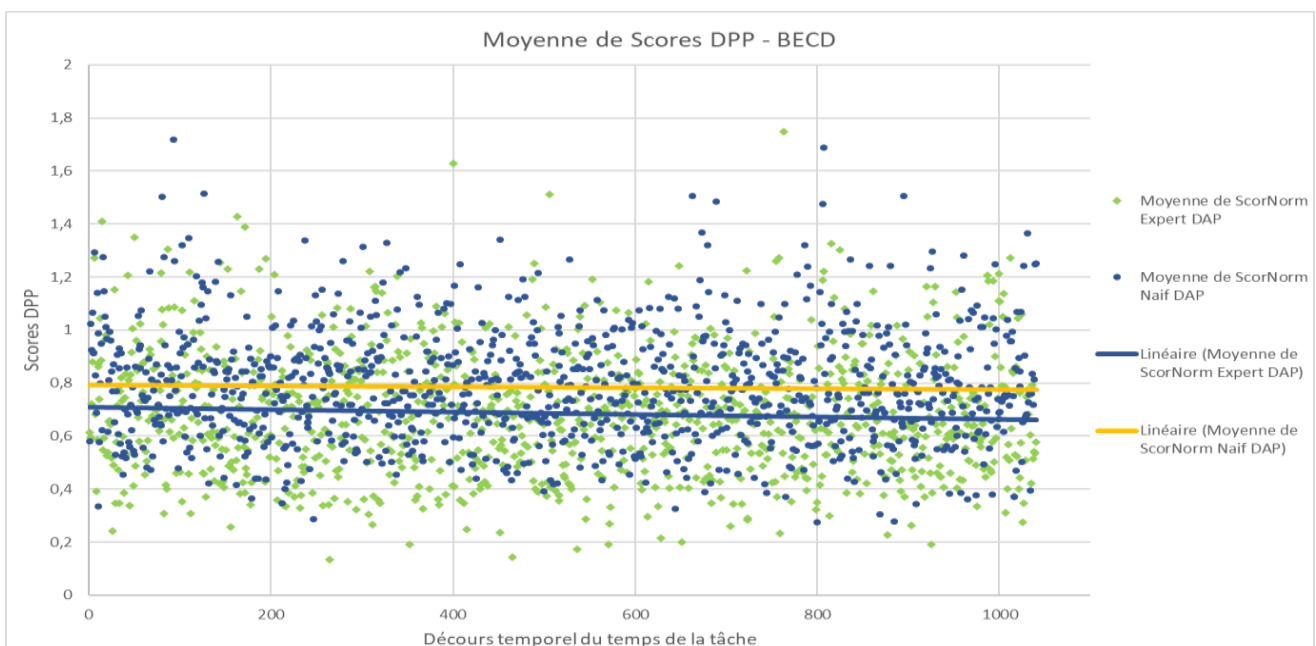


FIGURE 2 : Scores DPP attribués par une évaluation DAP par deux types d'auditeurs



Dans les évaluations BECD la différence de scores moyens entre les deux groupes d'auditeurs, Experts ( $\mu = 0,464$ ) et Naïfs ( $\mu = 0,634$ ) était attendue car les auditeurs Experts ont déjà été confrontés aux items de la BECD au cours de leur pratique clinique. Cependant, cette différence nous interpelle dans les évaluations DAP, Experts ( $\mu = 0,685$ ) et Naïfs ( $\mu = 0,787$ ). Ces scores indiquent que les auditeurs cliniciens sont de meilleurs décodeurs que les auditeurs naïfs, du fait de leur expertise auditive en écoute de la parole pathologique. Ainsi, pour mesurer l'effet d'expertise sur la tâche de DAP, nous avons comparé les scores DPP moyens attribués par les deux groupes d'auditeurs. La différence observée s'est révélée significative ( $p < 0.001$ ). Les experts proposent des scores DPP significativement plus bas que les auditeurs naïfs. Cette différence moyenne correspond à une différence moyenne de 0.1 trait d'écart moyen par phonème, ce qui est très faible. Nous avons donc cherché à évaluer la taille de cet effet en calculant un coefficient  $d$  de Cohen (« cohen.d » in R). Un 'd' aux alentours de 0.2 est considéré comme « faible », à 0.5 « médium » et à 0.8 « fort ». Le  $d$  de Cohen estimé pour le DAP, entre les auditeurs naïfs et experts est ( $d : 0.102$ ) dit négligeable. La différence de scores DPP moyens attribués par les auditeurs naïfs et les auditeurs experts ne perturbe donc pas l'évaluation car sa faiblesse montre que les deux groupes d'auditeurs n'évaluent finalement pas différemment les locuteurs avec le DAP. Cela indique que l'effet d'expertise auditive des cliniciens est neutralisé lors d'une évaluation de l'intelligibilité par la tâche de DAP.

Ces résultats corroborent l'hypothèse selon laquelle l'utilisation de pseudo-mots dans l'évaluation de l'intelligibilité permet de neutraliser l'effet d'expertise auditive des cliniciens.

## 4 Conclusion

Cette étude visait à valider la pertinence de l'utilisation de pseudo-mots, par rapport à des mots pour évaluer l'intelligibilité. Plus précisément, nous avons cherché à montrer qu'une évaluation basée sur l'utilisation de pseudo-mots permettait d'obtenir des scores de Déviation Phonologique Perçue (DPP) plus stables dans le temps qu'une évaluation basée sur des mots, et ce lorsque cette évaluation est conduite par des auditeurs naïfs aussi bien que par des experts. Nos résultats ont confirmé nos hypothèses : la tâche DAP permet d'obtenir un score DPP objectif, en nombre moyen de traits altérés par phonème, qui est représentatif de la sévérité du déficit des patients et du degré de dégradation du signal perçu. La tâche de DAP permet ainsi de neutraliser les effets d'apprentissage et mémorisation du matériel linguistique, ainsi que l'effet d'expertise auditive des cliniciens.

Ces résultats s'inscrivent dans la lignée des études précédemment conduites (Voir Section 1.1) pour évaluer les critères d'objectivité et de pertinence de la tâche DAP. Une partie des données de cette étude ont également permis d'évaluer la stabilité des scores DPP obtenus lorsque ceux-ci sont calculés non plus à partir de 52 pseudo-mots, mais à partir de 16 (Marczyck et al, soumis). Cette réduction, basée sur des critères phonologiques a montré de très bons résultats dans le cadre d'une évaluation par des auditeurs naïfs. Dans la suite de notre étude présentée ici, nous évaluerons la robustesse de l'évaluation de l'intelligibilité basée sur des pseudo-mots, auprès d'auditeurs experts, en utilisant des listes réduites à 16 items.

## Remerciements

Ce travail de recherche a été réalisé dans le cadre d'un contrat doctoral soutenu par la Ligue Contre le Cancer. Il s'inscrit également dans le cadre de projet de recherche 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

- ANDRE, C. ET AL. (2003): «PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception», in *XVth ICPHS. ICPHS*, Barcelone, Espagne, p. 1421-1424.
- ASTESANO, C. ET AL. (2018): «Carcinologic Speech Severity Index Project: A Database of Speech Disorder Productions to Assess Quality of Life Related to Speech After Cancer», in *Language Resources and Evaluation Conference*, Miyazaki.
- AUZOU, P. ET ROLLAND-MONNOURY, V. (2006): *Batterie d'évaluation clinique de la dysarthrie*. Ortho Edition. France: ORTHO.
- BALAGUER, M. ET AL. (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), p. 355-359. doi: [10.1016/j.anorl.2019.05.012](https://doi.org/10.1016/j.anorl.2019.05.012).
- BECHET, F. (2001): «LIA PHON : Un système complet de phonétisation de textes», *Traitement Automatique des Langues, TAL. (TAL - ATALA)*, 42(1), p. 47-67.
- BLANC, E. ET AL. (2014): «Adaptation en français du test d'intelligibilité de la version révisée du « Frenchay Dysarthria Assessment » (FDA-2)», in *Congrès de la Société Française de Phoniatry*. Paris, France. Disponible sur: <https://hal.archives-ouvertes.fr/hal-01615204>.
- GHIÒ, A. ET AL. (2017): «Du décodage acoustico-phonétique pour mesurer l'intelligibilité de locuteurs atteints de troubles de production de la parole», in *Journée de Phonétique Clinique (7ème)*, Paris.
- GHIÒ, A. ET AL. (2018): «Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique», in *XXXIIe Journées d'Études sur la Parole. XXXIIe Journées d'Études sur la Parole*, ISCA, p. 285-293. doi: [10.21437/JEP.2018-33](https://doi.org/10.21437/JEP.2018-33).
- GHIÒ, A. ET AL. (2019): «Application d'un test d'intelligibilité à partir de pseudo- mots dans le cas de patients post traitement de cancers des VADS», in *Journées de Phonétique Clinique*. Mons, Belgium. Disponible sur: <https://hal.archives-ouvertes.fr/hal-02098836>.
- GHIÒ, A. ET AL. (2020), «Evaluation de l'intelligibilité de patients avec traitement du cancer des cavités orales et pharyngales», submitted in this volume.
- HUSTAD, K. C., JONES, T. ET DAILEY, S. (2003): «Implementing Speech Supplementation Strategies: Effects on Intelligibility and Speech Rate of Individuals With Chronic Severe Dysarthria», *Journal of Speech, Language, and Hearing Research*, 46(2), p. 462-474. doi: [10.1044/1092-4388\(2003\)038](https://doi.org/10.1044/1092-4388(2003)038).
- JARZE, S. ET AL. (2017): «Analyse perceptive des voix dysphoniques. Influences et corrélations entre les dimensions G, R et B de l'échelle d'Hirano», in *Journées de Phonétique Clinique*. Paris, France. Disponible sur: <https://hal.archives-ouvertes.fr/hal-01615030>.
- 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), p. 222-234. doi: [10.1044/1058-0360\(2007\)027](https://doi.org/10.1044/1058-0360(2007)027).
- LINDBLOM, B. (1990): «On the communication process: Speaker listener interaction and the development of speech.», in *Augmentative and Alternative Communication*. (6), p. 220-230.
- WARREN, R. M. (1984): «Perceptual restoration of obliterated sounds», *Psychological Bulletin*, 96(2), p. 371-383. doi: [10.1037/0033-2909.96.2.371](https://doi.org/10.1037/0033-2909.96.2.371).
- WARREN, R. M. ET AL. (1997): «Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps», *Perception & Psychophysics*, 59(2), p. 275-283. doi: [10.3758/BF03211895](https://doi.org/10.3758/BF03211895).
- YORKSTON, K. M., DOWDEN, P. A. ET BEUKELMAN, D. R. (1992): «Intelligibility measurement as a tool in the clinical management of dysarthric speakers», in *Intelligibility in speech Disorders : Theory, measurement and management*. Raymond D. Kent. Madison, Wisconsin: John Benjamins Publishing Company, p. 265-286. Disponible sur: <https://benjamins.com/catalog/sspl.1.08yor>