



HAL
open science

Évaluations perceptives et automatiques de l'intelligibilité de la parole dégradée par simulation de la surdité professionnelle

Imed Laaridh, Julien Tardieu, Cynthia Magnen, Pascal Gaillard, Jérôme Farinas,
Julien Pinquier

► To cite this version:

Imed Laaridh, Julien Tardieu, Cynthia Magnen, Pascal Gaillard, Jérôme Farinas, et al.. Évaluations perceptives et automatiques de l'intelligibilité de la parole dégradée par simulation de la surdité professionnelle. XXXIIes Journées d'Études sur la Parole (2018), Association Francophone de la Communication Parlée; CNRS; Laboratoire Parole et Langage, Aix-en-Provence, Jun 2018, Aix-en-Provence, France. pp.392-400, <10.21437/JEP.2018-45>. <hal-01858337>

HAL Id: hal-01858337

<https://hal.science/hal-01858337v1>

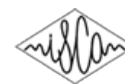
Submitted on 20 Aug 2018

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.



HAL Authorization



Évaluations perceptives et automatiques de l'intelligibilité de la parole dégradée par simulation de la surdité professionnelle

Imed Laaridh¹ Julien Tardieu² Cynthia Magnen² Pascal Gaillard³
Jérôme Farinas¹ Julien Pinquier¹

(1) IRIT, Université de Toulouse, CNRS, Toulouse, France

(2) MSHS-T (USR 3414), Université de Toulouse et CNRS, France

(3) CLLE (UMR 5263) Université de Toulouse et CNRS, France

prenom.nom@irit.fr¹, prenom.nom@univ-tlse2.fr^{2,3}

RÉSUMÉ

Cet article présente une étude comparative entre des mesures perceptives et automatiques pour l'évaluation de l'intelligibilité de la parole dans des conditions dégradées. Il fait suite à une étude précédente qui a permis de proposer une méthodologie pour la simulation et l'évaluation automatique de l'effet de la presbycusis sur l'intelligibilité de la parole. Nous proposons, dans ce travail, d'adapter cette approche à une autre pathologie de l'audition : la surdité par traumatisme en milieu professionnel. Dans ce cadre, un corpus de parole a été constitué afin de refléter différents niveaux de surdité professionnelle et a été soumis à des évaluations perceptives et automatiques. L'évaluation perceptive a démontré des effets similaires des dégradations sur les locuteurs homme, femme et enfant. De plus, les mesures d'intelligibilité automatiques, fondées sur la reconnaissance automatique de la parole, sont très corrélées aux performances humaines.

ABSTRACT

Perceptual and automatic evaluations of the intelligibility of speech degraded by noise induced hearing loss simulation

This study aims at comparing perceptual and automatic intelligibility measures on degraded speech. It follows a previous study that proposed an approach for the simulation and automatic evaluation of the effect of age-related hearing loss (presbycusis) on speech intelligibility. In this work, we propose to adapt this approach to a different hearing disorder : noise induced hearing loss (hearing loss caused by trauma in the professional context). Thus, a speech corpora simulating different levels of noise induced hearing loss is proposed and used to study perceptual and automatic intelligibility measures. The perceptual evaluation showed similar effects of the hearing disorder on the male, female and child speakers intelligibility. Also, the automatic intelligibility measure (based on automatic speech recognition scores) is proved to well represent the effects of the different severity levels of the hearing disorder. Indeed, high correlation measure are computed between the automatic and perceptual intelligibility measures.

MOTS-CLÉS : Mesure de l'intelligibilité de la parole, reconnaissance automatique de la parole, pathologies de l'audition, simulation de la surdité professionnelle.

KEYWORDS: speech intelligibility metric, automatic speech recognition, hearing disorders, noise induced hearing loss simulation.

1 Introduction

Plus de 3 millions de travailleurs en France sont exposés sur leur lieu de travail, d'une manière prolongée, à des niveaux de bruit potentiellement nocifs. En effet, et selon le centre d'information sur le bruit (CidB), environ 1200 cas de surdité professionnelle dus à ce type de traumatisme sonore sont diagnostiqués chaque année. Cette pathologie de l'audition est caractérisée par la présence de zones cochléaires mortes (i.e. détérioration de certaines parties des cellules cillées empêchant le codage tonotopyque). L'apparition de ces zones mortes s'accompagne de phénomènes d'acouphènes et d'hyperacousie résultant en des problèmes de compréhension de la parole à la fois dans les conditions de silence et de bruit.

Plus spécifiquement, cette pathologie se manifeste par une perte neuro-sensorielle dans les hautes fréquences du signal et par une encoche localisée dans la région autour de 4 kHz (McBride & Williams, 2001). La gravité de cette perte et de ce trou auditif dépend de la sévérité du traumatisme et peut évoluer dans le temps affectant d'autres bandes fréquentielles (notamment les bandes conversationnelles) avoisinantes les régions les plus touchées initialement. Ces troubles, lorsqu'ils ne sont pas soignés, peuvent compromettre les capacités de communication des personnes atteintes et engendrer des effets néfastes sur leurs vies personnelle et professionnelle tels que l'isolement (Strawbridge *et al.*, 2000) et la dépression (Gopinath *et al.*, 2009). Une des principales solutions aux pathologies auditives est l'utilisation d'audioprothèses. Ces outils permettent, en amplifiant certaines bandes fréquentielles, de donner une meilleure audibilité aux patients atteints. Cependant, comme rapporté dans (Vestergaard Knudsen *et al.*, 2010), environ 40% des patients équipés de ce type d'outils ne l'utilisent jamais (ou rarement). Ce rejet peut être expliqué, entre autres, par le peu de réglages spécifiques de ces prothèses à chaque utilisateur.

Dans la pratique clinique, l'évaluation des troubles auditifs se fait à l'aide d'audiogrammes tonaux ainsi que des tests vocaux. Les épreuves les plus utilisées sont celles de transcription du contenu linguistique de mots/phrases par les patients. L'aptitude d'un auditeur à reconnaître ce qui a été prononcé peut alors être utilisée comme une mesure de sa capacité auditive. Ces tests peuvent être utilisés dans différentes perspectives ; que ce soit dans la phase du diagnostique ou lors du réglage (et/ou de l'adaptation) des audioprothèses à chaque auditeur afin d'obtenir la meilleure possible intelligibilité de la parole.

Cependant, plusieurs limites peuvent être associées à ces épreuves perceptives. En effet, elles sont souvent très coûteuses en temps, notamment à cause de l'implémentation et du traitement souvent manuels des résultats de l'évaluation. De plus, et afin d'avoir une évaluation assez robuste, l'utilisation d'un jury d'auditeurs et non d'un seul est conseillée. Toutes ces caractéristiques limitent leur application clinique, notamment lors des phases de réglage spécifique de l'audioprothèse pour chaque utilisateur. Par ailleurs, le matériel linguistique utilisé dans ces tests est souvent assez limité (exemple : liste de mots de Fournier (Fournier, 1951)). Les patients sont alors rapidement familiers avec le contenu des tests. Cela, associé au fait que ces évaluations doivent souvent être répétées plusieurs fois afin de bien adapter l'audioprothèse dans différentes conditions sonores, peut ainsi compromettre leurs résultats (Hustad & Cahill, 2003).

Le reste de cet article est organisé comme suit. Dans la section 2, le contexte et les objectifs de ce projet sont expliqués. La section 3 décrit la méthodologie utilisée dans ce travail, notamment le processus de simulation des dégradations liées à la surdité professionnelle ainsi que le système de Reconnaissance Automatique de la Parole (RAP) utilisé. La section 4 présente les différents résultats des évaluations perceptive et automatique alors que la section 5 fournit quelques conclusions et

directions pour de futurs travaux.

2 Contexte et objectifs

Cette étude fait partie d'un projet pluridisciplinaire auquel participent différents acteurs du traitement du langage et de la parole : des informaticiens, des orthophonistes et des ORL. Ce projet fait suite à un travail précédent qui a permis de proposer et valider une méthodologie pour la simulation et l'évaluation automatique des troubles de l'audition liés à la presbycousie sur l'intelligibilité de la parole (Fontan *et al.*, 2017).

Le but de ce projet est d'adapter cette méthodologie à d'autres troubles de l'audition, spécifiquement la surdité professionnelle. L'approche proposée permettra alors d'évaluer automatiquement (sur la base des résultats de la RAP) les effets de la surdité professionnelle sur l'intelligibilité et donc de mieux affiner et faciliter le réglage des audioprothèses proposé aux patients par les experts. Plus largement, le projet visera aussi à adapter cette méthodologie à d'autres langues que le Français (Anglais) et à d'autres conditions de dégradation de la parole (parole dans un milieu bruyé). Dans un autre contexte, la méthodologie proposée pourra être utilisée pour l'évaluation objective de l'intelligibilité de la parole, la rééducation et le suivi clinique de l'évolution de la condition des patients atteints de troubles de la parole : dysarthrie, cancers de la cavité buccale et du pharynx, etc.

3 Méthodologie

La méthodologie proposée pour la simulation et l'évaluation de la surdité professionnelle consiste en 3 phases représentées dans la figure 1.

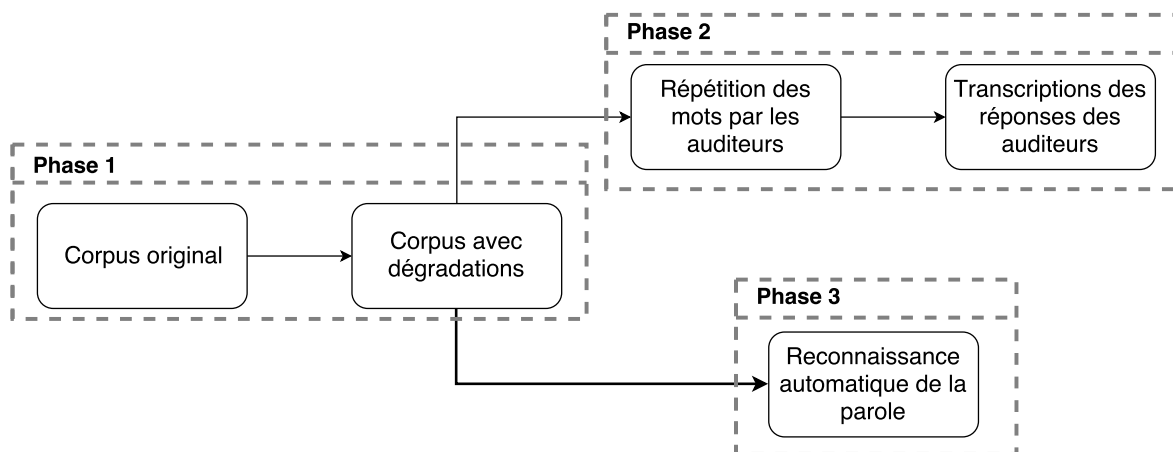


FIGURE 1: Diagramme des différentes phases de la méthodologie proposée.

La première phase décrit la constitution du corpus de données utilisé. Des enregistrements de mots issus de la liste de Fournier sont réalisés par trois locuteurs (un homme, une femme et un enfant).

Ensuite, différents niveaux de dégradation associés à la surdité professionnelle sont simulés sur les enregistrements afin de construire le corpus de parole utilisé lors des évaluations perceptive et automatique. La deuxième phase consiste en une évaluation perceptive de l'intelligibilité des enregistrements de la parole dégradés par 31 auditeurs. Ce protocole est décrit dans la sous-section 3.2. Les mêmes enregistrements sont aussi utilisés dans la troisième phase dans une tâche de RAP afin de mesurer des scores d'intelligibilité automatiques. Le système de RAP utilisé est décrit dans la sous-section 3.3.

3.1 Phase 1 : préparation de données

Le matériel linguistique utilisé dans ce protocole consiste en 60 mots (6 listes de 10 mots) issus des listes de mots proposées par Fournier en 1951 pour la mesure perceptive de l'intelligibilité (Fournier, 1951). Ces listes sont largement utilisées par les audioprothésistes pour l'évaluation de l'audition des patients. Tous les mots débutent avec une consonne et sont proposés sous la forme : article + nom (par exemple : « le parfum »).

Trois locuteurs : un homme (46 ans), une femme (47 ans) et une jeune fille (12 ans) ont participé à la production des enregistrements des mots. Les trois locuteurs avaient le Français comme langue maternelle et ont tous produit tous les mots de la liste. Les enregistrements ont été effectués dans une cabine audiométrique (PETRA¹) à l'aide d'un microphone omnidirectionnel Sennheiser MD46, d'une console de mixage TASCAM DM-3200 et d'un ordinateur MacPro équipé du logiciel Reaper.

Pour simuler la présence d'un bruit ambiant, nous avons choisi d'utiliser un bruit de type « brouhaha » construit à partir des enregistrements dans le silence en suivant la procédure décrite dans (Fontan *et al.*, 2017). Ce fichier de bruit a ensuite été mixé aux mots avec un rapport signal sur bruit de 5 dB. Les stimuli de parole avec et sans bruit ont ensuite été dégradés pour simuler les effets de la surdité professionnelle. La simulation a été effectuée dans MATLAB à partir des algorithmes initialement développés par (Nejime & Moore, 1997). À partir de données issues du terrain², 10 audiogrammes ont été simulés (voir tableau 1) allant d'une audition normale (niveau 1) à une surdité professionnelle grave (niveau 10).

TABLE 1: Audiogrammes utilisés dans la simulation de la surdité professionnelle, exprimés en pertes auditives (dB) pour chaque bande de fréquence (kHz).

Niveau	.125	.25	.5	.75	1	1.5	2	3	4	6	8	10	12	14	16
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	3	5	7	10	20	10	7	5	3	1	0
3	0	0	5	5	5	8	10	15	35	15	10	8	5	5	5
4	0	0	5	6	7	14	18	27	47	30	20	14	7	6	5
5	0	0	5	7	8	19	27	38	58	45	30	19	8	7	5
6	0	0	5	8	10	25	35	50	70	60	40	25	10	8	5
7	4	5	14	19	25	39	49	61	78	68	50	36	20	16	9
8	8	10	23	29	40	53	63	73	85	75	60	48	30	24	13
9	11	15	31	40	55	66	76	84	93	83	70	59	40	32	16
10	15	20	40	50	70	80	90	95	100	90	80	70	50	40	20

1. <http://petra.univ-tlse2.fr>

2. <https://www.uvmt.org/sections.php?op=printpage&artid=568>

3.2 Phase 2 : évaluation perceptive

L'évaluation perceptive consiste en une tâche d'écoute et de répétition des différents stimuli présentés dans les différentes conditions sonores. Après un entraînement sur 10 mots, chaque participant entendait les 60 mots de la liste auxquels étaient appliquées aléatoirement les 60 conditions du plan expérimental : dégradation(10)*locuteur(3)*bruit(2). La tâche du participant consistait à répéter oralement chaque mot entendu avant de passer au suivant. Les participants étaient assis à un mètre des deux haut-parleurs (Focal Solo 6 BE) dans la cabine audiométrique PETRA, et devant le microphone utilisé pour enregistrer les réponses. Le niveau sonore de diffusion des mots non dégradés était de 60 dBA (niveau 1 dans le silence). Les réponses ont ensuite été transcrites par les expérimentateurs afin de calculer les scores d'intelligibilité. Ce protocole d'évaluation a été retenu afin de forcer les auditeurs à fournir une réponse, éliminer les effets potentiels de leurs capacités orthographiques et n'impliquer que les composants de production et de perception de la parole dans l'évaluation.

31 participants (nombre de femmes = 20, âge moyen = 20,5 et écart type = 1,8) ont été sélectionnés pour participer à cette expérience. Les critères d'inclusion sont les suivants : francophones natifs (langue maternelle française, ayant toujours vécu en France), âgés de 18 à 30 ans inclus, étudiants dans des disciplines autres que musique, sciences du langage, langues étrangères ou psychologie, sans problèmes de vue non corrigés par des lentilles ou lunettes. Tous les participants ont été soumis à un audiogramme avec le logiciel AudioConsole afin de contrôler leur niveau d'audition.

3.3 Phase 3 : approche d'évaluation automatique

Dans ce travail, nous avons utilisés un système de RAP basé sur l'outil Sphinx-3 (Seymore *et al.*, 1998) distribué par Carnegie Mellon University (CMU). Il est important de noter qu'à l'inverse des applications classiques de la RAP, le but de ce travail n'est pas de proposer un système capable de reconnaître la parole dans des conditions dégradées et d'améliorer au plus ses performances en terme de taux d'erreur de mots reconnus (Word Error Rate - WER). L'objectif de ce travail est de proposer une approche automatique capable de simuler et de reproduire le comportement de la perception humaine face aux distorsions liées à la surdité professionnelle.

Les modèles acoustiques utilisés dans ce travail sont proposés par le Laboratoire d'Informatique de l'Université du Maine (LIUM) (Deléglise *et al.*, 2005) et appris sur plusieurs heures d'enregistrements radiophoniques issues du corpus ESTER (Galliano *et al.*, 2009). Il s'agit de modèles contextuels composés de 35 phonèmes et 5 types de pauses/silences (résultant en 5725 phonèmes en contexte) dont chaque état est représenté par un mélange de 22 lois gaussiennes (Gaussian Mixture Model - GMM). Chaque enregistrement est échantillonné à 16 kHz et une paramétrisation PLP (Hermansky, 1990) est utilisée. Puisque les modèles ont été appris sur de la parole produite uniquement par des hommes, le comportement du système de RAP n'était pas adapté à la parole produite par les locuteurs femme et enfant. Afin de pallier à ce problème, une adaptation de la longueur théorique du tract vocal (Vocal Tract Length Normalization - VTLN) a été utilisée (Wegmann *et al.*, 1996). Cette approche se fonde sur l'hypothèse de l'existence d'une relation linéaire entre la longueur du conduit vocal d'un locuteur et les fréquences des zones formantiques qui lui sont associées.

Deux modèles de langage ont été utilisés. Un premier modèle trigrammes de base, appelé BM, appris sur le corpus ESTER2 sur la base d'un lexique d'environ 62000 mots. Un second modèle bigrammes reflétant la composition syntaxique particulière de la liste de mots utilisé (article + nom). Ce modèle, appelé ML, a été appris sur des mots dissyllabiques commençant avec une consonne seulement

(environ 15000). Les fréquences utilisées pour les différentes formes du modèle sont celles définies dans (New *et al.*, 2007) et sont basées sur les sous-titres des films de la base de données Lexique 3.8.

4 Résultats

4.1 Évaluation de l'intelligibilité perceptive

Pour chaque mot, le score est égal à 1 si le mot est totalement reconnu (l'article précédent le mot n'est pas pris en compte) ou 0 s'il n'est que partiellement/ou non reconnu. Les scores d'intelligibilité correspondent ensuite aux pourcentages de mots reconnus pour chaque dégradation, chaque locuteur, dans le silence et dans le bruit. Les résultats sont présentés sur la figure 2. Les données ont été analysées en utilisant un modèle linéaire mixte généralisé, et ont révélé un effet significatif de la dégradation ($p < 0,001$) et du bruit ($p < 0,001$), mais un effet non significatif du locuteur (et donc de son âge et genre) ($p = 0,63$). Ces résultats sont cohérents avec ceux obtenus dans nos travaux précédents sur l'effet de la presbyacousie simulée sur l'intelligibilité de mots (Fontan *et al.*, 2017).

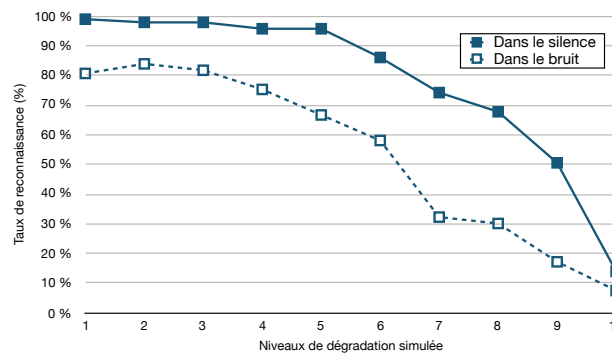


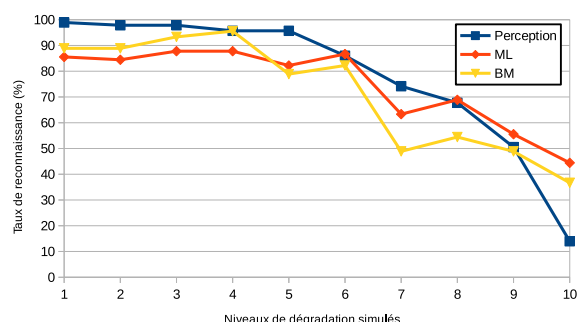
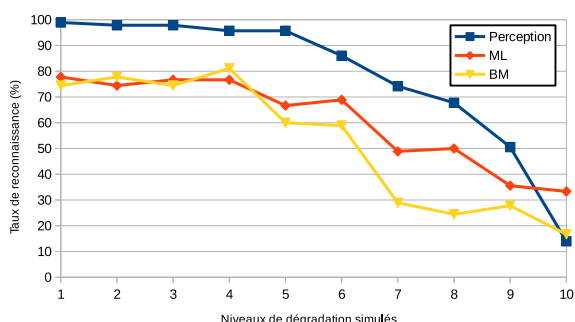
FIGURE 2: Scores d'intelligibilité perceptive dans le silence et dans le bruit.

4.2 Évaluation d'intelligibilité automatique

Nous avons appliqué l'approche de mesure automatique d'intelligibilité de la parole uniquement sur les stimuli en condition de silence. Ce choix vise à valider l'intérêt de l'approche avant de la confronter aux conditions les plus difficiles. Ce choix permet aussi une meilleure comparaison avec les résultats obtenus lors notre étude précédente (Fontan *et al.*, 2017).

La figure 3 rapporte les taux de reconnaissances automatique et perceptive moyens sur les trois locuteurs pour les différents niveaux de dégradation simulés. Deux stratégies de reconnaissance automatique sont retenues ; la première considère qu'un mot est reconnu automatiquement que s'il se retrouve en première position dans les propositions du système de RAP. La deuxième, indique si le mot cible fait partie des 10 mots les plus probables pour le système. Nous observons qu'à l'exception de la condition de dégradation la plus sévère 10, le taux de reconnaissance automatique est inférieur à celui observé chez les auditeurs humains. En effet, le taux de reconnaissance moyen atteint 77,8% pour les auditeurs humains contre 60,9% et 52,4% pour le système automatique en utilisant les modèles de langage ML et BM respectivement et la première stratégie de reconnaissance automatique. La capacité du système automatique, contrairement à la perception humaine, de compenser les dégradations les plus sévères (traduite par un taux de reconnaissance supérieur) nécessite une investigation plus approfondie. Sans surprise, nous observons que le modèle de langage ML, plus adapté aux listes de

mots utilisées dans notre protocole expérimental, atteint des taux de reconnaissance supérieurs à ceux obtenus par le modèle de base BM.

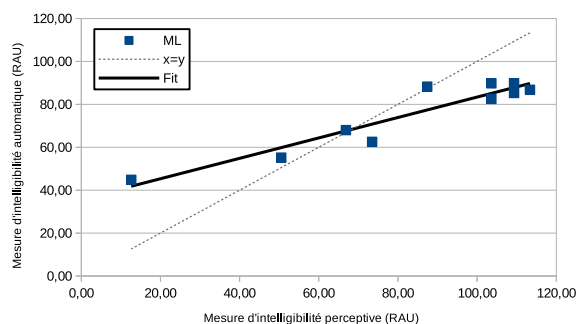
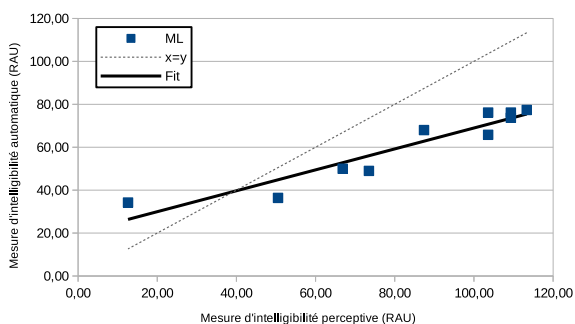


(a) Taux de reconnaissance du mot en première position

(b) Taux de reconnaissance dans les 10 premières positions

FIGURE 3: Taux de reconnaissances automatique (en utilisant deux modèles de langage) et perceptive de différents niveaux simulés de surdité professionnelle.

En utilisant les deux stratégies de reconnaissance automatique, le comportement de l'approche automatique suit la même tendance que l'évaluation perceptive : plus le niveau de dégradation simulée est élevée, plus le système de RAP a du mal à reconnaître les mots prononcés. En observant de plus près l'évolution des taux de reconnaissance selon la sévérité de la dégradation simulée, nous retrouvons que la perception humaine arrive à compenser les distorsions observées sur les 5 premiers niveaux de dégradation. À partir du niveau 5, l'intelligibilité perceptive subit une perte, presque linéaire, avec une importante chute de taux de reconnaissance entre les niveaux 9 (50,5%) et 10 (14%). La mesure d'intelligibilité automatique présente une évolution différente, plus sensible aux faibles dégradations (diminution du taux reconnaissance entre les niveaux 4 et 5) et plus résistante au dégradation sévère du niveau 10.



(a) Taux de reconnaissance du mot en première position.

(b) Taux de reconnaissance dans les 10 premières positions.

FIGURE 4: Distribution de la mesure d'intelligibilité automatique du modèle de langage ML « transformé » (RAU) en fonction de la mesure d'intelligibilité perceptive « transformée » (RAU).

Afin de mesurer la capacité de la mesure d'intelligibilité automatique à produire des mesures comparables à celles d'une évaluation perceptive, nous avons calculé le taux de corrélation de Pearson entre les deux mesures. En vue d'éviter les effets plafonds des scores, une transformation en des unités arc-sinus rationnels (rational-arcsine unit - RAU) (Studebaker, 1985) a été réalisée. La figure 4

montre la relation entre les scores perceptifs et automatiques ainsi transformés. Le tableau 2 rapporte les taux de corrélation entre les mesures d'intelligibilité automatique et perceptive.

TABLE 2: Coefficients de corrélation de Pearson observés en comparant les mesures d'intelligibilité automatique et perceptive ($p < 0,001$).

Modèle	Stratégies de comparaison	
	Mot reconnu en première position	Mot reconnu dans les 10 premières positions
BM	0,91	0,90
ML	0,94	0,94

Les taux de corrélation entre les deux mesures d'intelligibilité atteignent 0,94 dans le cas de l'utilisation du modèle de langage ML. Cette mesure montre la capacité de l'approche automatique à répliquer le comportement de la perception humaine face à de la parole simulant différents niveaux de surdité professionnelle. De plus, ce taux est comparable à celui obtenue lors de notre étude précédente sur la presbyacousie : corrélation de 0,97 obtenue dans (Fontan *et al.*, 2017).

Ce lien entre les mesures d'intelligibilité perceptive et automatique confirme le potentiel de l'approche automatique proposée et son potentiel d'utilisation lors des phases de réglage et d'adaptation des audioprothèses destinées aux patients souffrant de la surdité professionnelle.

5 Conclusions

Ce travail vise à étudier et analyser les effets de la surdité professionnelle sur l'intelligibilité de la parole. Une évaluation perceptive a démontré que ces effets sont indépendants du genre du locuteur (homme ou femme) ainsi que de son âge (adulte ou enfant). Cette évaluation a aussi permis d'établir une référence à laquelle comparer la mesure d'intelligibilité automatique. Le comportement de cette dernière, basée sur un système de RAP, s'est révélé similaire et comparable à la perception humaine (corrélation de Pearson de 0,94 entre les deux mesures). Ce résultat valide l'intérêt de la méthodologie proposée et son potentiel d'utilisation dans le cadre du réglage et de l'adaptation des audioprothèses aux besoins spécifiques de chaque utilisateur.

De futurs travaux permettrons de généraliser cette approche à d'autre langue (Anglais) et d'étudier sa robustesse face à des conditions de dégradation plus difficile (parole en milieu bruité). Finalement, et vue le lien entre la perception de la parole et sa production, il nous semble intéressant d'exploiter cette mesure d'intelligibilité automatique dans le cadre d'évaluation des troubles de production de la parole tels que la dysarthrie.

Remerciements

Cette étude a été réalisée dans le cadre du projet numéro 14054092 « CLE 2015 PHONICS : Intelligent Electronic Device for Measuring Speech Comprehension », financé par la région Midi-Pyrénées. Le projet est porté par la Maison des Sciences de l'Homme et de la Société de Toulouse (MSHS-T) en partenariat avec l'IRIT et la société Archean Technologies.

Références

- DELÉGLISE P., ESTEVE Y., MEIGNIER S. & MERLIN T. (2005). The LIUM speech transcription system : a CMU Sphinx III-based system for french broadcast news. In *Ninth European Conference on Speech Communication and Technology*.
- FONTAN L., FERRANÉ I., FARINAS J., PINQUIER J., TARDIEU J., MAGNEN C., GAILLARD P., AUMONT X. & FÜLLGRABE C. (2017). Automatic speech recognition predicts speech intelligibility and comprehension for listeners with simulated age-related hearing loss. *Journal of Speech, Language, and Hearing Research*, **60**(9), 2394–2405.
- FOURNIER J. E. (1951). Audiométrie vocale : les épreuves d'intelligibilité et leurs applications au diagnostic, à l'expertise et à la correction prothétique des surdités.
- GALLIANO S., GRAVIER G. & CHAUBARD L. (2009). The ESTER 2 evaluation campaign for the rich transcription of french radio broadcasts. In *Tenth Annual Conference of the International Speech Communication Association*.
- GOPINATH B., WANG J. J., SCHNEIDER J., BURLUTSKY G., SNOWDON J., MCMAHON C. M., LEEDER S. R. & MITCHELL P. (2009). Depressive symptoms in older adults with hearing impairments : the blue mountains study. *Journal of the American Geriatrics Society*, **57**(7), 1306–1308.
- HERMANSTADT H. (1990). Perceptual linear predictive (PLP) analysis of speech. *the Journal of the Acoustical Society of America*, **87**(4), 1738–1752.
- HUSTAD K. C. & CAHILL M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, **12**(2), 198–208.
- MCCRIDE D. & WILLIAMS S. (2001). Audiometric notch as a sign of noise induced hearing loss. *Occupational and Environmental Medicine*, **58**(1), 46–51.
- NEJIME Y. & MOORE B. C. (1997). Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*.
- NEW B., BRYBAERT M., VERONIS J. & PALLIER C. (2007). The use of film subtitles to estimate word frequencies. *Applied psycholinguistics*, **28**(4), 661–677.
- SEYMORE K., CHEN S., DOH S., ESKENAZI M., GOUVEA E., RAJ B., RAVISHANKAR M., ROSENFELD R., SIEGLER M., STERN R. *et al.* (1998). The 1997 CMU Sphinx-3 English broadcast news transcription system. In *DARPA Broadcast News Transcription and Understanding Workshop*.
- STRAWBRIDGE W. J., WALLHAGEN M. I., SHEMA S. J. & KAPLAN G. A. (2000). Negative consequences of hearing impairment in old age : a longitudinal analysis. *The Gerontologist*, **40**(3), 320–326.
- STUDEBAKER G. A. (1985). A rationalized arcsine transform. *Journal of Speech, Language, and Hearing Research*, **28**(3), 455–462.
- VESTERGAARD KNUDSEN L., ÖBERG M., NIELSEN C., NAYLOR G. & KRAMER S. E. (2010). Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids : A review of the literature. *Trends in amplification*, **14**(3), 127–154.
- WEGMANN S., MCALLASTER D., ORLOFF J. & PESKIN B. (1996). Speaker normalization on conversational telephone speech. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 1, p. 339–341 : IEEE.