Evaluation automatique de l'intelligibilité de la parole dans le contexte de cancers de la tête et du cou

Imed Laaridh, Corinne Fredouille, Alain Ghio, Muriel Lalain, Virginie Woisard

Evaluation automatique de l'intelligibilité de la parole dans le contexte de cancers de la tête et du cou

Imed Laaridh[1] Corinne Fredouille[1]
Alain Ghio[2] Muriel Lalain[2] Virginie Woisard[3]
(1) LIA, Université d'Avignon
(2) Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France
(3) Service ORL, CHU Larrey, URI Octogone-Lordat, Toulouse, France
corinne.fredouille@univ-avignon.fr

RÉSUMÉ

Dans le contexte de la parole pathologique, l'évaluation perceptive reste la méthode la plus utilisée par les cliniciens et orthophonistes pour mesurer le niveau d'intelligibilité de leurs patients, et ce malgré le caractère subjectif bien connu de ce type d'évaluation. Le travail présenté ici consiste à reproduire la méthode automatique de prédiction de l'intelligibilité, proposée dans une précédente étude, sur un corpus de patients atteints de cancers de la tête et du cou et présentant des troubles de la parole plus ou moins sévères. Ce travail doit permettre (1) de comparer le comportement du système automatique de prédiction du degré d'intelligibilité sur une population de patients différente, (2) de montrer la pertinence, en terme d'application d'outils automatiques, d'un nouveau protocole d'enregistrements des patients, basé sur des logatomes et dédié à l'évaluation de l'intelligibilité. Les résultats expérimentaux obtenus confirment la validité de l'approche automatique (corrélation r=0.84) mais également du protocole utilisé.

ABSTRACT

Automatic evaluation of speech intelligibility in the context of head and neck cancers.

In disordered speech context, and despite its well-known subjectivity, perceptual evaluation has been, and still, the most commonly used method in clinical practice to evaluate the intelligibility level of speech productions of patients. The work presented in this paper consists in reproducing the automatic method of intelligibility prediction, proposed in a previous study, on a corpus of head and neck cancer patients, and presenting more or less severe disordered speech. This work should permit (1) to compare the behavior of the automatic system for predicting the degree of intelligibility in a different patient population, (2) to demonstrate the relevance, in terms of application of automatic tools, of a new protocol of patient recordings, based on logatomes and dedicated to the evaluation of intelligibility. The experimental results obtained confirm the validity of the automatic approach (correlation r = 0.84) but also the one of the protocol used.

MOTS-CLÉS : Traitement automatique de la parole, i-vecteurs, évaluation de l'intelligibilité, troubles de la parole, cancers de la tête et du cou.

KEYWORDS: automatic speech processing, speech disorders, i-vectors, intelligibility assessment, head and neck cancers.

10.21437/JEP.2018-22

1 INTRODUCTION

Les troubles de la parole peuvent affecter, en fonction de leurs origines, différentes composantes de la production de la parole : respiration, phonation, résonance et/ou articulations. Différentes mesures ont été étudiées dans la littérature pour évaluer la qualité de la parole telle que l'intelligibilité, la compréhensibilité et la sévérité de la parole pathologique. Par conséquent, de nombreux protocoles d'évaluation, dédiés à la pratique clinique mais également à la recherche, ont été conçus pour regrouper tout ou partie de mesures. Ces protocoles peuvent être dédiés à des troubles spécifiques de la voix - dysphonie - et/ou de la parole - dysarthrie pour les origines neurologiques ou d'autres troubles de la parole comme dans le cas de cancers de la tête et du cou ou de malformations. Ils visent à aider les cliniciens dans leur connaissance des troubles de la parole et de leur évaluation clinique, essentielle pour suivre la progression de la maladie des patients dans le cas d'un traitement ou l'évolution des altérations dans le cas d'une rééducation de la parole. Dans ce contexte, l'évaluation perceptive reste la méthode la plus utilisée dans la pratique clinique malgré ses limites bien documentées telles que la non reproductibilité et la subjectivité.

La perte d'intelligibilité est l'une des plaintes les plus fréquentes rencontrées chez les patients souffrant de troubles de la parole. Pour faire face aux limitations rapportées ci-dessus, les approches automatiques ont été considérées, très tôt, comme des solutions potentielles en vue d'apporter des outils objectifs pour l'évaluation de l'intelligibilité. Dans la littérature, on peut distinguer deux types principaux d'approches : celles directement basées sur des systèmes automatiques de transcription de parole fournissant un taux d'erreurs de transcription de mots comme score d'intelligibilité (Christensen *et al.*, 2012), et celles utilisant des technologies automatiques capables d'extraire de l'information pertinente, injectée, dans un second temps, dans un système de prédiction automatique du degré d'intelligibilité (Middag *et al.*, 2009; Khan *et al.*, 2014). Parallèlement, d'autres approches automatiques axées sur une analyse plus pointue de la parole dysarthrique, dédiées, par exemple, à la détection d'anomalies, ont également montré des résultats probants dans l'évaluation du degré d'intelligibilité (Laaridh *et al.*, 2015).

Le paradigme des i-vecteurs est une approche état de l'art, appliquée avec succès dans les applications de reconnaissance du locuteur (Dehak *et al.*, 2011). Il est prouvé qu'il représente et capture de manière très pertinente les caractéristiques des locuteurs ciblés(Verma & Das, 2015). Ce paradigme a été utilisé et adapté à plusieurs autres contextes tels que la reconnaissance de la langue et même l'évaluation de la qualité de la parole. Dans (An *et al.*, 2015), cette représentation, combinée à un large ensemble de caractéristiques acoustiques, syllabiques et phonotactiques, a été utilisée pour la prédiction automatique des scores UPDRS (Unified Parkinson's Disease Rate Scale : batterie de tests dédiés à l'évaluation des troubles moteurs de la maladie de Parkinson) de production de parole de patients atteints de la maladie de Parkinson, dans le contexte spécifique du défi ComParE Interspeech 2015. Dans (Wang *et al.*, 2016), le paradigme des i-vecteurs a été utilisé comme une normalisation du locuteur et impliqué dans une approche de classification plus complexe, combinant des caractéristiques acoustiques et articulatoires pour la détection automatique de la Sclérose Latérale Amyotrophique (SLA). Dans (Martínez *et al.*, 2015), les i-vecteurs ont été utilisés pour la représentation de segments de mots produits par 15 locuteurs dysarthriques, permettant d'établir des corrélations importantes entre les mesures d'intelligibilité prédites automatiquement et les mesures d'intelligibilité de référence. Enfin, dans (Garcia *et al.*, 2017), les auteurs ont proposé une approche basée sur une distance cosinus entre la représentation i-vecteur d'une production de parole (test) et deux i-vecteurs de référence représentant individuellement de la parole normale et dysarthrique.

Dans un travail précédent (Laaridh *et al.*, 2017), nous avons proposé une approche basée sur le paradigme des i-vecteurs pour la prédiction automatique de plusieurs métriques d'évaluation de la parole dysarthrique comme l'intelligibilité, la sévérité des troubles de la parole, et plus spécifiquement, le degré d'altération de l'articulation. L'approche proposée a été appliquée sur un corpus de 129 locuteurs dysarthriques et témoins et des mesures de corrélation élevées (entre 0,8 et 0,9) ont été atteintes entre les différentes mesures perceptives d'évaluation de la parole et les prédictions automatiques.

Dans ce travail, soutenu par le projet de recherche C2SI financé par l'Institut National du CAncer (INCA), la même approche automatique basée sur des i-vecteurs est utilisée pour la prédiction automatique de la mesure de l'intelligibilité de la parole, dans le contexte particulier de patients atteints de cancers de la tête et du cou (HNC - Head and Neck Cancer). Comparé au travail précédent, l'objectif de ce travail est double. D'une part, il doit permettre d'observer le comportement de l'approche de prédiction automatique dès lors qu'elle est appliquée sur une population de patients différente. En effet, contrairement à la dysarthrie pour laquelle les troubles de la parole peuvent être diffus, divers et, par conséquent, difficiles à localiser, nous avons ici la connaissance, en fonction du cancer du patient, de la localisation de la déficience (langue, palais, larynx, ...) et une information un peu plus "précise" du trouble de parole attendu. D'autre part, les scores de prédiction automatique sont comparés ici à des mesures d'intelligibilité issues d'un protocole original d'enregistrements et d'évaluation perceptive de productions de parole pathologique basé sur la production de pseudo-mots par des patients et des témoins. Il s'agit par conséquent d'étudier et de valider l'utilisation de ce protocole dans le cadre de l'application d'outils automatiques en vue d'une transposition dans le milieu clinique.

2 DESCRIPTION DU CORPUS

2.1 Population

L'étude actuelle est basée sur le corpus français de parole HNC, enregistré dans le cadre du projet C2SI. Ce corpus comprend des patients atteints de cancers de la tête et du cou (cavité buccale ou oropharyngés) et des locuteurs témoins. Tous les patients ont subi un traitement dédié consistant en une chirurgie, et/ou une radiothérapie, et/ou une chimiothérapie.
Durant le protocole d'enregistrements conçu spécifiquement dans le cadre du projet C2SI, tous les locuteurs ont été invités à enregistrer différentes tâches de production de la parole (voyelles tenues /a/, lecture d'un texte pour l'obtention de parole lue, lecture d'une liste de phrases pour l'étude de la prosodie, description d'images pour l'obtention de parole spontanée, production de pseudo-mots isolés, etc). Le lecteur pourra se référer à (Astesano *et al.*, 2018) pour une description détaillée des enregistrements de ce corpus, des différentes tâches de production de parole réalisées par les locuteurs et de leurs objectifs de recherche.
Dans ce travail, nous concentrons notre attention uniquement sur la tâche de production de pseudo-mots isolés, appelée DAP (pour Décodage Acoustico-Phonétique) dans le reste de l'article. Durant cette tâche, tous les locuteurs devaient prononcer 52 pseudo-mots (incluant 2 essais d'entraînement). Chaque pseudo-mot suivait une structure phonotactique comme suit : $C(C)_1V_1C(C)_2V_2$ où $C(C)_i$ est une consonne isolée ou un groupe consonantique. Les consonnes ont été sélectionnées à partir d'une liste contenant 18 items, les groupes consonantiques à partir de listes contenant 16 ou 32 items et les voyelles à partir de listes contenant 7 ou 8 items en fonction de leur position dans le

pseudo-mot. En utilisant cette méthode combinatoire, environ 95000 pseudo-mots ont été générés. Cet ensemble a ensuite été réduit à environ 90000 éléments après la suppression des pseudo-mots faisant référence à une entrée lexicale de la langue française. Des listes de 52 pseudo-mots prévues pour la tâche de lecture ont ensuite été construites aléatoirement à partir de l'ensemble des 90000 éléments restants tout en respectant un processus de construction commun. Toutes les listes extraites devaient respecter les mêmes règles concernant la distribution des consonnes, des groupes consonantiques et des voyelles. Au total, 85 patients et 41 témoins ont été enregistrés pour cette tâche. Il convient de noter que certains patients n'ont pas terminé la tâche de lecture en raison d'une fatigabilité extrême et ont produit moins de 52 pseudo-mots requis.

2.2 Evaluation perceptive de l'intelligibilité

Tous les pseudo-mots prononcés par les locuteurs du corpus HNC ont été transcrits par 40 auditeurs, suivant le protocole décrit dans (Ghio *et al.*, 2018). Vu l'ampleur de la tâche, chaque pseudo-mot a été évalué par 3 d'entre eux. Le choix des auditeurs naïfs présentant un bon niveau d'orthographe a été délibéré afin d'éviter tous effets d'habituation à la parole pathologique (altérations et phénomènes de compensation) bien connus chez les cliniciens et qui peuvent faciliter leur compréhension. Ces auditeurs ont été confrontés à une tâche qui s'apparente à un décodage acoustico-phonétique (d'où le nom de la tâche de production de la parole - DAP) suivi d'une transcription écrite. Au total, 18360 transcriptions orthographiques des pseudo-mots ont été recueillies. L'annotation a été réalisée en utilisant la plate-forme Lancelot-Perceval (Ghio *et al.*, 2003). Chaque auditeur pouvait écouter chaque pseudo-mot jusqu'à 3 fois avant de fournir sa transcription. Pour chaque pseudo-mot, la distance moyenne de Levenshtein a été utilisée pour comparer la suite de phonèmes attendue et les réponses transcrites, compte-tenu des caractéristiques acoustiques distinctives entre phonèmes. La distance moyenne a ensuite été calculée, pour chaque locuteur, sur l'ensemble des pseudo-mots produits oralement, fournissant ainsi, pour chacun d'eux, une mesure d'(in)intelligibilité (les valeurs élevées correspondent à la plus grande distance entre le pseudo-mot attendu et la réponse transcrite et caractérisent donc les locuteurs les moins intelligibles).

3 METHODOLOGIE

L'approche automatique utilisée ici, et décrite précédemment dans (Laaridh *et al.*, 2017), repose sur deux étapes. La première étape consiste à projeter chaque énoncé de parole dans le sous-espace de variabilité totale de faible dimension et de représenter ainsi chaque enregistrement associé à un locuteur témoin ou à un patient par un i-vector (Dehak *et al.*, 2011).
La deuxième étape est une régression du sous-espace des i-vecteurs vers l'espace d'intelligibilité (à 1 dimension) nécessaire à notre tâche d'évaluation de cette dernière. La régression par Machines à Vecteurs de Support (RVS) sera utilisée compte-tenu du nombre limité de données annotées disponibles pour l'étude. En effet, malgré le grand nombre de patients et de sujets témoins disponibles, la quantité de données disponible reste limitée par rapport à d'autres applications de traitement automatique de la parole «standard».

3.1 Espace de variabilité totale

Le paradigme de l'espace de variabilité totale a d'abord été introduit dans le contexte de la reconnaissance automatique du locuteur. Dans cette approche, un extracteur d'i-vecteurs convertit une séquence de vecteurs acoustiques en un seul vecteur de faible dimension représentant l'ensemble de l'énoncé de parole. Le super-vecteur s dépendant du locuteur et de la session issu de la concaténation des vecteurs de moyennes d'un modèle de mélange gaussien (GMM) est supposé obéir à un modèle linéaire de la forme : $s = m + Tw$
où m est le super-vecteur moyen du modèle de parole générique ou modèle du monde (UBM), T est la matrice de projection de faible rang apprise sur un large ensemble de données par estimation MAP (elle représente le sous-espace de «variabilité totale») et w est une variable latente, appelée "i-vector", ayant une distribution normale $\mathcal{N}(0, I)$. Les algorithmes pour l'estimation de T et l'extraction des i-vecteurs sont décrits dans (Matrouf *et al.*, 2007).

3.2 Extraction des i-vecteurs

Une étape de paramétrisation des signaux de parole, nécessaire à tout processus automatique, repose ici sur l'extraction de 19 coefficients cepstraux (LFCC), leurs 19 dérivées premières (Δ) et leurs 11 dérivées secondes ($\Delta\Delta$). Une normalisation de la moyenne et de la variance (MVN) est ensuite appliquée aux paramètres LFCC, valeurs estimées sur les portions de parole de chaque enregistrement détectées à l'aide d'un alignement phonétique automatique contraint par le texte. Un UBM de 512 composantes dépendant du genre et une matrice de variabilité totale T de rang 400 estimée à partir des corpus de parole français Ester 1 & 2, REPERE et ETAPE (7690 sessions de 2906 locuteurs) (Ajili *et al.*, 2016) sont utilisées pour extraire un i-vecteur par enregistrement de parole. Le package LIA_SpkDet de la boîte à outils open source ALIZE (Larcher *et al.*, 2013) est utilisé pour les différents traitements en lien avec les i-vecteurs décrits ci-dessus.

3.3 Régression par Machines à Vecteurs de Support (SVR)

Dans ϵ-SVR, l'idée de base est de trouver une fonction qui a, au plus, ϵ d'écart par rapport aux valeurs de référence cibles pour toutes les données d'entraînement. Quand une telle tâche n'est pas réalisable, des variables de compromis et de relâchement sont introduites pour faire face au problème d'optimisation (Smola & Schölkopf, 2004).
Pour chaque vecteur de test, et compte-tenu des vecteurs d'apprentissage $x_i \in R^{400}$, $i = 1, ..., n$, la fonction de décision est alors :

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{1}$$

où α_i et α_i^* sont des multiplicateurs de Lagrange, K est la fonction du noyau et b est le biais. Dans ce travail, les noyaux RBF ont été utilisés.
En outre, et compte-tenu des excellentes performances obtenues par la représentation à base d'i-vecteurs dans le domaine de la reconnaissance du locuteur, un processus de validation croisée sur 10 sous-ensembles a été mis en place pour éviter les biais liés à la présence des mêmes locuteurs dans les phases d'apprentissage et de test.

4 RESULTATS ET DISCUSSIONS

Pour évaluer les performances de l'approche automatique sur le corpus HNC, les mesures de corrélation de Pearson (r) et d'erreur quadratique moyenne (RMSE) ont été calculées entre les scores d'intelligibilité prédits automatiquement et les mesures d'intelligibilité issues de l'évaluation perceptive dans le cadre du DAP (section 2.2).

4.1 Prédiction de l'intelligibilité au niveau locuteur

La première expérience que nous avons réalisée consistait en la prédiction automatique de l'intelligibilité de chaque locuteur en utilisant l'ensemble des enregistrements de parole, dans lequel il prononçait les 52 pseudo-mots. Cela signifie que nous disposions d'une quantité importante de données pour estimer le i-vecteur représentant la production acoustique de chaque locuteur incluant les altérations de parole. La figure 1 fournit la mesure d'intelligibilité prédite automatiquement par rapport à l'évaluation perceptive de référence issue du protocole DAP par locuteur. Nous observons que l'approche automatique est capable d'effectuer une bonne séparation entre les patients et les groupes de locuteurs témoins. La pente de régression confirme la capacité du système à détecter et représenter la perte d'intelligibilité mesurée perceptivement par les auditeurs. En effet, les mesures r et RMSE atteignent respectivement 0.84 et 2.339. Ce taux de corrélation est tout à fait cohérent avec les résultats précédents observés sur la tâche de lecture de texte produite par des patients dysarthriques (Laaridh *et al.*, 2017). De plus, la mesure RMSE obtenue est assez faible considérant que l'intervalle de la mesure de référence est caractérisé par une large variation (mesures de référence appartenant à l'intervalle [0,22] pour ce corpus) et sa sensibilité extrême (l'évaluation perceptive est mesurée sur un intervalle discret ; la moindre différence de caractéristiques sur n'importe quel phonème entre la référence et la transcription des pseudo-mots aboutit, au minimum, à une distance de 1 point).

4.2 Prédiction de l'intelligibilité au niveau de sous-listes de mots

La pertinence de l'approche automatique, en terme de prédiction de l'intelligibilité sur les pseudo-mots étant confirmée dans la section précédente, nous proposons ici d'étudier la quantité de parole requise pour effectuer une prédiction fiable. En effet, comme mentionné dans la section 2, la fatigabilité des patients peut les contraindre à abandonner un tâche de production de parole dans le cadre d'un protocole d'évaluation. S'assurer que l'ensemble du protocole d'enregistrement est nécessaire au bon fonctionnement des outils automatiques est un aspect majeur. Vérifier que ce dernier pourrait être allégé en est un autre, tout aussi crucial, notamment dans une perspective de pratique clinique. Considérant que chaque locuteur a produit au moins une liste de 52 pseudo-mots, nous avons divisé la liste de ces derniers en 5 sous-listes d'environ 10 mots chacune. Chaque sous-liste représentait environ 7 secondes seulement de parole, ce qui est évidemment très court dans le cadre d'un traitement automatique suivant la tâche visée. En dehors de leur taille, la distribution des pseudo-mots parmi les sous-listes reste totalement aléatoire et ne tient pas compte de leur structure ou de leurs phonèmes de composition. Chaque sous-liste a été assignée à une mesure d'intelligibilité perceptive en faisant la moyenne des distances mesurées sur les pseudo-mots (section 2.2) la composant. Au total, 623 sous-listes ont été extraites représentant les 126 locuteurs. De manière identique à l'expérimentation précédente, une validation croisée de 10 sous-ensembles a été mise en œuvre au niveau du locuteur pour éviter les biais résultant de l'utilisation de sous-listes produites par le même locuteur dans les

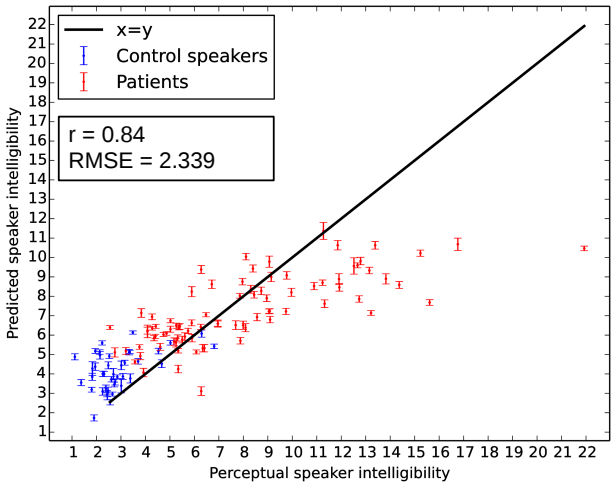

FIGURE 1 – Mesures d'intelligibilité par locuteur, issues de la prédiction automatique sur l'ensemble des 52 pseudo-mots et des mesures perceptives (DAP). Droite de pente 1 (noir) donnée pour indication.

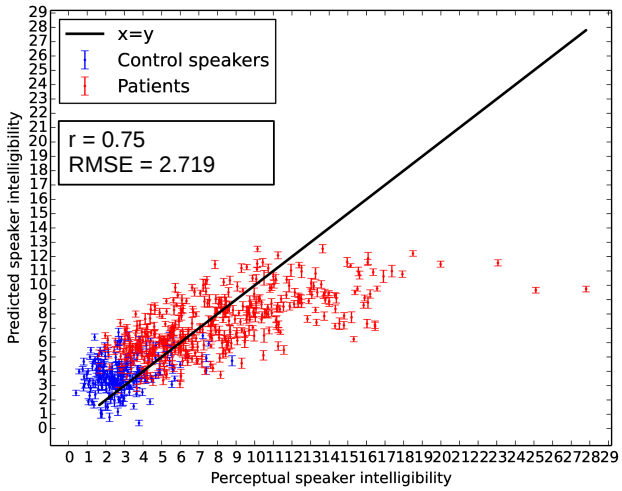

FIGURE 2 – Mesures d'intelligibilité par sous-liste (incluant ∼10 pseudo-mots), issues de la prédiction automatique et des mesures perceptives (DAP). Droite de pente 1 (noir) donnée pour indication.

phases d'apprentissage et de test.

La figure 2 fournit la mesure d'intelligibilité prédite automatiquement par rapport à l'évaluation perceptive de référence issue du protocole DAP pour chacune des sous-listes produites par les locuteurs. Nous observons que le caractère discriminant de l'approche automatique entre patients et locuteurs de contrôle reste conservé et que les 10 sous-listes de pseudo-mots utilisées étaient de taille suffisante pour l'approche automatique de détection de la perte d'intelligibilité pour les patients.

4.3 Discussions

Une hypothèse clé faite dans l'expérience précédente était que les différentes sous-listes, extraites aléatoirement, étaient équivalentes et portaient la même information du point de vue de l'intelligibilité. Cependant, l'observation des mesures de corrélation entre les prédictions automatiques et

TABLE 1 – Mesures de corrélation de Pearson (r) et d'erreur quadratique moyenne (RMSE) entre scores d'intelligibilité prédits et perceptifs en fonction des meilleures et des pires sous-listes.

	r	RMSE
Sélection des meilleures sous-listes	0.87	1.685
Sélection des pires sous-listes	0.58	4.278

les évaluations perceptives calculées sur les 5 sous-listes extraites par locuteur montre des valeurs variant de 0,72 à 0,80. Cette variabilité signifie que les sous-listes, et donc les pseudo-mots, ont été considérées différemment par le système de prédiction automatique et l'approche à base d'i-vecteurs. Partant de cette observation, le choix des sous-listes à utiliser pour la tâche de prédiction semble avoir un impact non négligeable.

Afin de mettre en évidence ce comportement, la table 1 présente les meilleures (pire) mesures r et RMSE qui pourraient être obtenues si nous considérions seulement les sous-listes qui minimisent (maximisent) les erreurs de prédiction. Nous observons que la mesure de corrélation pourrait atteindre jusqu'à 0,87 et la RMSE seulement 1,685 si les sous-listes utilisées pour l'évaluation sont bien choisies. En revanche, la mesure r descend à 0,58 si les sous-listes utilisées contiennent des mots «moins significatifs». Même si, dans ce cas, la perte de performance est importante, cette valeur peut être considérée comme un seuil en termes de mauvaises prédictions pour l'approche automatique. Comme reporté plus haut, le potentiel discriminant d'un pseudo-mot en termes d'intelligibilité peut varier ostensiblement. Par conséquent, en plus de la méthode sophistiquée de conception de la liste initiale de 90000 pseudo-mots (voir section 2.1), nous pouvons facilement supposer qu'une méthodologie plus réfléchie pour composer les sous-listes de pseudo-mots pourrait entraîner des gains encore plus importants en terme de précision de la prédiction automatique.

Cette observation peut être très utile lors de la mise en œuvre de nouveaux protocoles pour l'évaluation automatique / perceptive des troubles de la parole. En effet, la plupart des tests d'évaluation sont substantiels et nécessitent beaucoup d'efforts de la part du patient (et par conséquent de la fatigue potentielle) et de l'auditeur évaluateur. Une sélection plus pertinente des unités linguistiques pourrait s'avérer extrêmement utile pour réduire de manière drastique les efforts à fournir.

5 CONCLUSIONS ET PERSPECTIVES

Cet article étudie une approche automatique pour la prédiction de l'intelligibilité de la parole basée sur le paradigme des i-vecteurs et des modèles de régression à base de machines à support de vecteurs. Cette approche a été appliquée sur un protocole de lecture dédié de pseudo-mots produits par des locuteurs souffrant de cancers de la tête ou du cou. Une corrélation élevée (r=0,84) a été obtenue entre les mesures d'intelligibilité prédites automatiquement sur chaque locuteur et celles issues de l'évaluation perceptive basée sur le DAP dès lors que les 52 pseudos-mots sont considérés dans un seul enregistrement de parole. Par ailleurs, l'approche s'est avérée stable et robuste au manque de données puisque r = 0,75 a été atteint en utilisant seulement 20% environ de la production de parole de chaque locuteur (pseudo-mots $\sim$10). Les résultats de cette dernière expérimentation a permis de montrer l'effet des sous-listes utilisées pour l'évaluation, et par conséquent de la pertinence des pseudo-mots qui les composent en termes de prédiction de l'intelligibilité. Les travaux futurs étudieront les informations portées par chaque pseudo-mot et l'impact de son contenu phonétique sur l'évaluation de l'intelligibilité, du point de vue perceptif et automatique.

Références

AJILI M., BONASTRE J.-F., BEN KHEDER W., ROSSATO S. & KAHN J. (2016). Phonetic content impact on forensic voice comparison. In *Spoken Language Technology Workshop (SLT), 2016 IEEE*, p. 210–217 : IEEE.

AN G., BRIZAN D. G., MA M., MORALES M., SYED A. R. & ROSENBERG A. (2015). Automatic recognition of unified parkinson's disease rating from speech with acoustic, i-vector and phonotactic features. In *Proceedings of Interspeech'15*, Dresden, Allemagne.

ASTESANO C., BALAGUER M., FARINAS J., FREDOUILLE C., GAILLARD P., GHIO A., GIUSTI L., LAARIDH I., LALAIN M., LEPAGE B., MAUCLAIR J., NOCAUDIE O., PINQUIER J., PONT O., POUCHOULIN G., PUECH M., ROBERT D., SICARD E. & WOISARD V. (2018). Carcinologic speech severity index project : A database of speech disorders productions to assess quality of life related to speech after cancer. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2018)*, Myazaki, Japan.

CHRISTENSEN H., CUNNINGHAM S., FOX C., GREEN P. & HAIN T. (2012). A comparative study of adaptive, automatic recognition of disordered speech. In *Proceedings of Interspeech'12*, Portland, USA.

DEHAK N., KENNY P. J., DEHAK R., DUMOUCHEL P. & OUELLET P. (2011). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(4), 788–798.

GARCIA N., OROZCO-ARROYAVE J. R., D'HARO L., DEHAK N. & NÖTH E. (2017). Evaluation of the neurological state of people with parkinson's disease using i-vectors. In *Proceedings of Interspeech'17*.

GHIO A., ANDRÉ C., TESTON B. & CAVÉ C. (2003). Perceval : une station automatisée de tests de perception et d'évaluation auditive et visuelle. *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA)*, **22**, 115–133.

GHIO A., LALAIN M., GIUSTI L., POUCHOULIN G., ROBERT D., FREDOUILLE C., LAARIDH I. & WOISARD V. (2018). Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique. In *Journées d'Etude sur la Parole (JEP), Aix-en-Provence, France, Juin 2018*.

KHAN T., WESTIN J. & DOUGHERTY M. (2014). Classification of speech intelligibility in parkinson's disease. *Biocybernetics and Biomedical Engineering*, **34(1)**, 35–45.

LAARIDH I., BEN KHEDER W., FREDOUILLE C. & MEUNIER C. (2017). Automatic prediction of speech evaluation metrics for dysarthric speech. In *Proceedings of Interspeech'17*, p. 1834–1838.

LAARIDH I., FREDOUILLE C. & MEUNIER C. (2015). Automatic detection of phone-based anomalies in dysarthric speech. *ACM Transactions on accessible computing*, **6**(3), 9 :1–9 :24.

LARCHER A., BONASTRE J.-F., FAUVE B. G., LEE K.-A., LÉVY C., LI H., MASON J. S. & PARFAIT J.-Y. (2013). Alize 3.0-open source toolkit for state-of-the-art speaker recognition. In *Proceedings of Interspeech'13*, p. 2768–2772.

MARTÍNEZ D., LLEIDA E., GREEN P., CHRISTENSEN H., ORTEGA A. & MIGUEL A. (2015). Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace. *ACM Transactions on Accessible Computing (TACCESS)*, **6**(3), 10.

MATROUF D., SCHEFFER N., FAUVE B. G. & BONASTRE J.-F. (2007). A straightforward and efficient implementation of the factor analysis model for speaker verification. p. 1242–1245.

MIDDAG C., MARTENS J.-P., VAN NUFFELEN G. & DE BODT M. (2009). Automated intelligibility assessment of pathological speech using phonological features. *EURASIP Journal on Advances in Signal Processing*, **2009**(1), 1–9.

SMOLA A. J. & SCHÖLKOPF B. (2004). A tutorial on support vector regression. *Statistics and computing*, **14**(3), 199–222.

VERMA P. & DAS P. K. (2015). i-vectors in speech processing applications : a survey. *International Journal of Speech Technology*, **18**(4), 529–546.

WANG J., KOTHALKAR P. V., CAO B. & HEITZMAN D. (2016). Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples. In *Proceedings of Interspeech'16*.