

Détection automatique de zones de déviance dans la parole dysarthrique : étude des bandes de fréquences

Corinne Fredouille¹, Gilles Pouchoulin²

(1) LIA, CERI, Université d'Avignon et des Pays de Vaucluse, Avignon, France

(2) LPL, CNRS, Université d'Aix-Marseille, Aix-en-Provence, France

corinne.fredouille@univ-avignon.fr, gilles.pouchoulin@lpl-aix.fr

RÉSUMÉ

Cet article propose d'associer à un système de détection automatique de zones anormales/déviantes sur des productions de parole altérée, une analyse en sous-bandes de fréquences. Ce travail vise à montrer que les portions anormales de parole peuvent être détectées différemment selon les bandes de fréquences. La complémentarité des sous-bandes fréquentielles pourraient ensuite être utilisées afin d'améliorer la robustesse de la détection automatique.

Les résultats expérimentaux, établis sur un groupe de patients de parité homme-femme souffrant de dysarthrie, mettent en évidence un comportement très intéressant des sous-bandes de fréquences moyennes et élevées, comportement différent selon le sexe des patients et maintenu à travers une analyse phonétique par classe (voyelles, consonnes, ...). Les observations relevées permettent d'entrevoir de larges perspectives d'investigation concernant l'analyse des gains apportés par les sous-bandes individuelles par rapport à la bande de fréquence totale, ainsi que le potentiel de la stratégie de combinaison par sous-bande.

ABSTRACT

Abnormal Zone Detection in Dysarthric Speech Utterances according to Frequency Bands

This paper proposes to join a speech processing-based system devoted to the automatic detection of abnormal zones in impaired speech utterances with an analysis in frequency subbands. This work aims to demonstrate that abnormal zones could be detected differently according to the frequency bands. The complementarity of the frequency subbands could be used afterwards to improve the robustness of the automatic detection.

Experimental results, reported for a set of gender-balanced patients suffering from dysarthria, highlight a very interesting behavior of medium and high frequency subbands, different from male and female patients and supported by a comparison between vowel and consonant classes. The related observations open a large set of investigation perspectives, regarding the analysis of gains brought by individual subbands compared with the full frequency band, but also regarding potential subband combination strategy.

MOTS-CLÉS : Troubles de la parole, dysarthrie, détection automatique, zones de déviance.

KEYWORDS: Speech disorders, dysarthria, automatic detection, deviant speech zones.

1 Introduction

Depuis de nombreuses années, cliniciens, orthophonistes, phonéticiens et chercheurs en sciences du langage et en traitement de l'information et de la communication, démontrent un réel intérêt à mieux comprendre les troubles de la parole dans le contexte de dysarthrie. Pourtant, la classification des différents types de dysarthrie proposée par (Darley *et al.*, 1969b,a, 1975), basée sur des dimensions perceptives, est par exemple toujours d'actualité aujourd'hui. Dans la pratique clinique, les troubles de la parole sont évalués de manière perceptive suivant différentes échelles ou grilles d'évaluation telles que "the Frenchay dysarthria assessment" proposée par (Enderby, 1983) et adaptée au français par (Auzou *et al.*, 2000a) ou l'item 18 de l'échelle UPDRS ("Unified Parkinson's Disease Rating Scale") (Weismer, 1984) dédiée à la maladie de Parkinson. Néanmoins, plusieurs études ont démontré les limites de ces évaluations perceptives, au regard notamment de la variabilité intra- et inter-jugement qui peut être observée et du manque de reproductibilité, même dans le cas de protocoles standardisés ou de passations réalisées par des experts du domaine (Auzou *et al.*, 2000b). De ce constat, il en ressort que la mise en place d'approches ou de méthodologies plus objectives est nécessaire pour mieux cibler et caractériser les effets de la dysarthrie sur la production de parole, permettant ainsi de venir compléter les évaluations perceptives.

Différentes études portant sur la caractérisation de la dysarthrie à partir d'analyses acoustiques du signal de parole ont été proposées dans la littérature (Weismer, 1984; Kent *et al.*, 1999; Kent et Kim, 2003). Elles ont permis de mettre en évidence des ensembles de paramètres majeurs dans la distinction de patients prototypiques atteints de différents types de dysarthrie. Néanmoins, la très grande diversité pouvant être observée dans les échantillons de parole de patients dysarthriques, démontre que ces études doivent être encore approfondies et que l'axe phonétique doit venir compléter l'axe purement acoustique.

Le travail présenté ici vise à aider les phonéticiens dans leur analyse manuelle d'échantillons de parole dégradée que l'on sait coûteuse en temps et en ressource. L'objectif de l'approche proposée est de guider les chercheurs vers des zones du signal potentiellement attractives en terme de "déviance" (d'un point de vue de la phonétique clinique) de manière à permettre une analyse plus fine des zones concernées. Cette approche repose sur une détection préalable des zones de parole considérées comme "déviantes" ou anormales. Cette détection reprend une méthodologie déjà présentée dans (Fredouille et Pouchoulin, 2011) à laquelle se greffe une analyse en bandes fréquentielles. Cette dernière repose sur l'hypothèse que certaines altérations du signal de parole dues à la dysarthrie pourront être mieux détectées en ciblant des bandes de fréquences bien spécifiques.

L'article est organisé de la manière suivante : la section 2 rappellera les principales étapes de l'approche de détection automatique des zones de déviance dans un signal de parole tandis que la section 3 montrera son application dans l'analyse des bandes de fréquences. La section 4 sera dédiée au protocole expérimental mis en oeuvre dans ce travail, ainsi qu'à la description du corpus de parole dysarthrique et des mesures d'évaluation sur lesquels les expériences seront réalisées. Cette section reportera également les résultats obtenus. Ces derniers seront discutés en section 5, apportant différentes perspectives à ce travail.

2 Détection automatique des zones de déviance

La détection automatique de zones de déviance, décrite dans (Fredouille et Pouchoulin, 2011) repose sur 3 étapes principales : (1) un alignement automatique en phonèmes ; (2) le calcul de score de normalité pour chaque phonème et (3) la production d'une cartographie mettant en évidence les zones de déviance.

2.1 Alignement automatique en phonèmes

Comme énoncé précédemment, la première étape de la détection automatique consiste à segmenter le signal de parole en unités minimales qui seront par la suite analysées en vue de les étiqueter comme zones normales ou anormales. Dans les travaux conduits jusqu'à présent, le phonème a été choisi comme unité de segmentation pour deux raisons principales : (1) la durée des phonèmes est considérée comme suffisante pour fournir un score de normalité (voir section suivante), comparé notamment à des unités telles que les trames de signal de 20ms utilisées généralement par les outils de traitement de la parole ; (2) les phonèmes peuvent subir des distorsions du point de vue acoustique en présence de troubles de la voix et/ou de la parole. Cette segmentation en phonèmes est fournie par un outil automatique d'alignement contraint par le texte développé par le Laboratoire Informatique d'Avignon (LIA). Cet outil prend en entrée le signal de parole accompagné d'une transcription orthographique du contenu linguistique et un lexique phonétisé, et fournit en sortie une liste de frontières (début et fin) pour chaque phonème rencontré dans la transcription.

2.2 Mesure des scores acoustiques

A partir de la séquence de phonèmes et de leurs frontières fournies par l'étape précédente, cette seconde étape consiste à calculer un score acoustique normalisé pour chacun des phonèmes. Ce score sera ensuite utilisé pour déterminer le degré de normalité de chaque phonème. Dans ce travail, le score acoustique normalisé est défini comme suit :

$$L_p^{norm}(y_p) = \log\left(\frac{L_p^{Constraint}(y_p)}{L_{p'}^{Nonconstraint}(y_p)}\right) \quad (1)$$

où $L_p^{norm}(y_p)$ est le score acoustique normalisé obtenu pour chaque phonème p et calculé sur le segment de parole y_p . $L_p^{Constraint}(y_p)$ est le score acoustique assigné au phonème p pendant le processus d'alignement automatique contraint par le texte. $L_{p'}^{Nonconstraint}(y_p)$ est le score acoustique obtenu sur le segment y_p à partir d'un alignement en phonèmes non contraint par le texte. Le phonème p' (ou la séquence de phonèmes) pouvant être potentiellement différent du phonème attendu p , le score normalisé obtenu permettra ainsi d'avoir une première mesure du degré de distorsion du phonème p .

2.3 Cartographie

La dernière étape de la détection automatique consiste à exploiter les scores normalisés assignés à chaque phonème individuellement, en déterminant de manière automatique si le phonème doit être considéré comme normal ou anormal du point de vue acoustique. Cette décision est établie grâce à un indice de normalité attribué à chaque phonème par projection de leur score acoustique normalisé sur une échelle de référence.

Cette échelle est construite à partir d'une population de sujets sains produisant de la parole considérée comme normale. Des scores acoustiques normalisés sont calculés sur les signaux de parole produits par cette population en suivant les deux étapes décrites précédemment. A partir de ces scores, des valeurs de scores minimums, maximums et médians sont estimées et utilisées pour définir l'échelle de référence. La projection des scores acoustiques normalisés issus des productions de parole d'un patient dysarthrique sur l'échelle de référence permet finalement de définir, la position d'un phonème sur l'échelle de référence - à l'intérieur, il est considéré comme normal, à l'extérieur, considéré comme anormal. Sa position permet en outre de déterminer l'indice de normalité associé.

En vue de faciliter la lecture des résultats issus de cette projection, une cartographie est produite, permettant de représenter graphiquement grâce à une échelle de couleurs associée à l'échelle de référence, l'ensemble des phonèmes et leur indice de normalité. Cette représentation permet ainsi de visualiser très rapidement les zones de déviance (accumulation de plusieurs phonèmes d'indice de normalité faible) et de comparer des cartographies établies sur des productions de parole différentes issues d'un même patient par exemple.

3 Détection et domaine fréquentiel

Comme mentionné en introduction, l'objectif de ce travail est de déterminer si la méthodologie proposée pour la détection automatique de zones de déviance peut être plus pertinente dès lors qu'elle est appliquée sur des bandes de fréquences limitées comparé à la bande fréquentielle totale [0-8kHz]. Il est attendu que certaines bandes de fréquences soient plus pertinentes que d'autres pour cibler des anomalies sur des phonèmes bien précis. Dans ce sens, il a été choisi de manière ad-hoc d'analyser le signal de parole suivant 6 bandes de fréquences de 1kHz, chacune répartie également sur la bande [0-6kHz], à comparer à l'utilisation de la bande fréquentielle totale [0-8kHz]. La méthodologie décrite dans la section précédente est par conséquent appliquée en considérant chacune des 6 sous-bandes individuellement et la bande totale.

4 Protocole expérimental et résultats

Les expériences présentées dans cette section sont conduites sur un corpus de parole dysarthrique enregistré à l'hôpital La Pitié-Salpêtrière de Paris. Ce corpus comprend des enregistrements de 7 locuteurs contrôles et de 8 patients dysarthriques. Ces patients souffrent de maladies génétiques rares (maladies lysosomales) et présentent des degrés de sévérité de la dysarthrie très variables dus notamment à une progression différente de leur maladie. Tous les locuteurs ont été enregistrés à plusieurs reprises avec des périodes, entre deux enregistrements, d'une semaine pour les sujets contrôles et pouvant aller jusqu'à six mois pour les sujets dysarthriques. De 3 à 5

enregistrements sont ainsi disponibles par locuteur. Tous les locuteurs ont été enregistrés dans les mêmes conditions sur une tâche de lecture de texte ("Le cordonnier"). La durée des productions de parole varie de 48s à 196s, avec une moyenne de 60s environ pour les locuteurs contrôles et 85s pour les patients.

Pour finir, les productions de parole des patients ont été analysées par un expert humain en vue d'annoter les zones de parole considérées comme anormales/déviantes. Pour chaque enregistrement, cette analyse a été réalisée à partir d'une écoute du signal de parole, des analyses/indicateurs fournis par le logiciel Praat et de la segmentation automatique en phonèmes fournie par l'outil du LIA. Le résultat attendu de la part de l'expert et de son analyse était, pour chaque phonème, une étiquette "normal" ou "anormal" et dans ce dernier cas, des indications sur la nature observée de la déviance (information non utilisée dans ce travail) comme par exemple bruit, dévoisement, distorsion spectrale, etc.

4.1 Protocole d'évaluation

En vue d'évaluer la détection automatique des zones de déviance, les sorties de cette dernière sont comparées aux annotations fournies par l'expert humain sur le corpus dysarthrique, considérées ici comme référence. Cette comparaison est quantifiée selon deux mesures issues du domaine de la recherche d'information :

- mesure de **Rappel** de la classe "phonème déviant" donnée par le rapport entre le nombre de phonèmes déviants correctement détectés par le système automatique (vis-à-vis de la référence) et le nombre de phonèmes étiquetés comme déviants par l'expert humain. Ce rapport mesure les performances du système automatique dans sa tâche de détection des phonèmes déviants : plus le rapport est proche de 1, plus le système est performant ;
- mesure de **Précision** de la classe "phonème déviant" donnée par le rapport entre le nombre de phonèmes déviants correctement détectés par le système automatique et le nombre de phonèmes détectés comme déviants par le système automatique. Ce rapport mesure le taux inverse de faux positifs produits par le système automatique dans sa tâche de détection : plus le rapport est proche de 1, plus le système est précis.

4.2 Résultats

Les résultats de la détection automatique des zones de déviance couplée à une analyse en bandes de fréquences de 1kHz ou appliquée sur la bande de fréquences totale [0-8kHz] sont reportés en figure 2. Exprimés en termes de rappel et de précision, ces résultats sont fournis par patient et représentent la moyenne des valeurs de rappel et de précision obtenues individuellement sur chacun de leurs enregistrements.

La comparaison des résultats entre les différentes bandes de fréquences montre que les mesures de précision sont similaires ou légèrement supérieures sur la bande totale que sur les bandes de 1kHz dans la grande majorité des cas. Une plus grande variabilité est observable sur les mesures de rappel entre la bande totale et les bandes de 1kHz, mais également entre les bandes elles-mêmes. En considérant le sujet "Homme1" qui présente les plus grandes différences dans les valeurs de rappel, la bande totale est associée à une valeur de rappel de 0.3 contre 0.52, 0.61, 0.76 et 0.62 pour les bandes [0-1], [2-3], [3-4] et [4-5]kHz respectivement alors que les mesures de précision restent plutôt stables. Même si les différences entre valeurs ne sont pas si notables chez les autres sujets masculins, il est à noter que les bandes [2-3], [3-4] et [4-5]kHz affichent

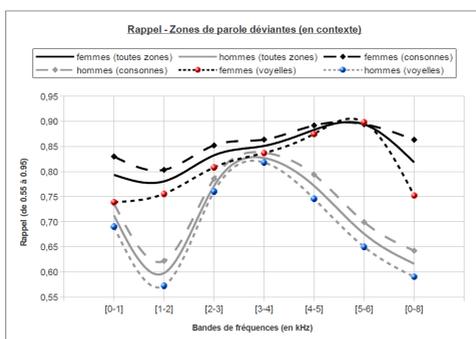


FIGURE 1: Mesures de rappel obtenues par bande de fréquence de 1kHz et sur la bande totale. Résultats donnés en fonction du genre des patients et de la classe de phonèmes observée.

systématiquement des mesures de rappel meilleures que la bande totale et que ces différences sont statistiquement significatives (excepté pour le patient “Homme3”). Cette observation est intéressante étant donné que les patients présentent des degrés de sévérité de la dysarthrie très différents.

Les patients féminins montrent un comportement différent. En premier lieu, les différences entre bandes de 1kHz sont moins marquées en termes de mesures de rappel. Néanmoins, on peut remarquer que les bandes [3-4], [4-5] et [5-6]kHz se détachent des autres, obtenant des mesures de rappel supérieures à celles de la bande totale (et des différences statistiquement significatives entre la bande [5-6]kHz et la bande totale, excepté pour la patiente “Femme2”). Les mesures de précision sont quant à elles plus faibles sur les bandes de 1kHz comparées à la bande totale mais les différences ne sont pas significatives dans ces derniers cas.

5 Discussion

Comme première analyse des résultats présentés en section 4.2, les mesures de rappel par bande de fréquences et par genre des patients sont reportées sur la figure 1 en distinguant 3 classes de phonèmes : l'ensemble des phonèmes présents dans les productions de parole des patients (“all zones”), l'ensemble des voyelles et celui des consonnes. On peut ainsi observer que la courbe de variation des mesures de rappel obtenues sur les différentes bandes de fréquences est tout à fait similaire si l'on compare les trois classes de phonèmes considérées. La classe des consonnes est associée aux meilleures mesures de rappel quelle que soit la bande de fréquences observée. En outre, considérant les classes des voyelles et des consonnes, les bandes [2-3], [3-4] et [4-5]kHz restent celles pour lesquelles les valeurs de mesures de rappel sont les plus élevées pour les patients masculins et les bandes [3-4], [4-5] et [5-6]kHz pour les patients féminins, similairement à ce qui a pu être observé sur l'ensemble des phonèmes. Si le comportement des consonnes sur ces bandes et notamment les bandes de hautes fréquences, peut être expliqué par une analyse plus ciblée des altérations liées aux consonnes fricatives ou occlusives (meilleure

prise en compte de l'amplification du bruit de friction), les résultats observés sur les voyelles sont plus surprenants, demandant une étude plus approfondie. Ces différents résultats et observations ouvrent différentes voix d'investigation :

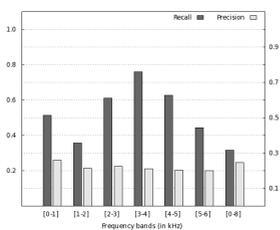
- les gains en termes de mesures de rappel observés sur certaines bandes de fréquences doivent faire l'objet d'une étude basée sur l'analyse des zones de déviations détectées par chacune des bandes et sur le taux de recouvrement de ces zones entre bandes. En d'autres termes, "est-ce que l'analyse par bande de fréquences permet simplement d'augmenter le nombre des zones de déviance correctement détectées ou est-ce qu'elle révèle de nouvelles zones de déviance non détectées par l'analyse en bande totale ?"
- l'analyse phonétique doit être approfondie sur des classes plus fines (fricatives, occlusives, ...) afin de mieux cibler les zones de déviance et mieux comprendre le comportement de la détection sur certaines classes (cas des voyelles par exemple) ;
- le caractère complémentaire des bandes doit être étudié en vue de proposer le cas échéant un paradigme de fusion des informations ciblées par chacune d'elles permettant d'en tirer bénéfice pour améliorer la détection des zones de déviance ;
- finalement, cette étude doit être élargie à un corpus de parole dysarthrique plus conséquent, présentant davantage de patients et des types de dysarthrie différents.

Remerciements

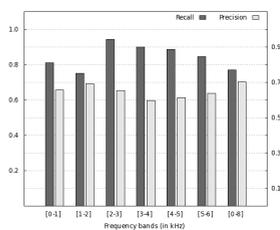
Ce travail est financé par l'Agence Nationale de la Recherche (ANR-08-BLAN-0125). Nous remercions Nathalie Lévêque et Frédéric Sedel de nous avoir fourni leur corpus de parole dysarthrique, Olavo Panseri pour ses annotations manuelles et Cécile Fougeron pour son aide.

Références

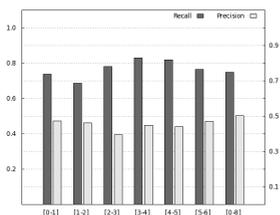
- AUZOU, P., OZSANCAK, C., JAN, M., LEONARDON, S., MENARD, J. F., GAILLARD, M. J., EUSTACHE, F. et HANNEQUIN, D. (2000a). Evaluation clinique de la dysarthrie : Présentation et validation d'une méthode. *Revue de neurologie*, 154 (6-7).
- AUZOU, P., OZSANCAK, C., MORRIS, J. R., JAN, M., EUSTACHE, F. et HANNEQUIN, D. (2000b). Voice Onset Time in aphasia, apraxia of speech and dysarthria : a review. *Clinical Linguistics and Phonetics*, 14 (2).
- DARLEY, F. L., ARONSON, A. E. et BROWN, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, 12:462–496.
- DARLEY, F. L., ARONSON, A. E. et BROWN, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12:246–269.
- DARLEY, F. L., ARONSON, A. E. et BROWN, J. R. (1975). *Motor speech disorders*. Philadelphia.
- ENDERBY, P. (1983). Frenchay dysarthric assessment. *Pro-Ed, Texas*.
- FREDOUILLE, C. et POUCHOULIN, G. (2011). Automatic detection of abnormal zones in pathological speech. *In Intl Congress of Phonetic Sciences (ICPhS'11)*, Hong Kong.
- KENT, R. D. et KIM, Y. J. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics and Phonetics*, 17 :6:427–445.
- KENT, R. D., WEISMER, G., KENT, J. F., VORPERIAN, H. K. et DUFFY, J. R. (1999). Acoustic studies of dysarthric speech : Methods, progress, and potential. *The Journal of Communication Disorders*, 32 :3:141–186.
- WEISMER, G. (1984). Acoustic description of dysarthric speech : Perception correlates and physiological inferences. *Rosenbeck, C. J. (ed), Seminar in speech and language, Thieme Stratton, New York*.



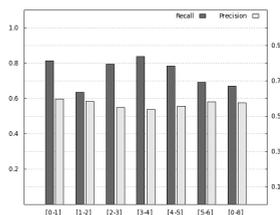
(a) Homme1



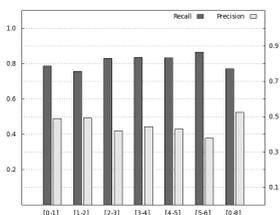
(b) Homme2



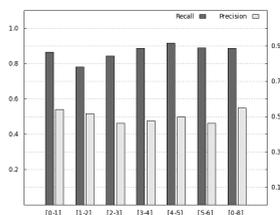
(c) Homme3



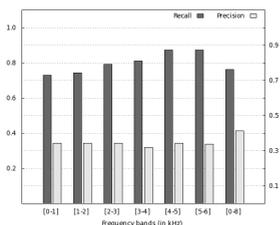
(d) Homme4



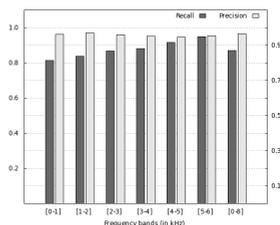
(e) Femme1



(f) Femme2



(g) Femme3



(h) Femme4

FIGURE 2: Performance de la détection automatique des zones de déviance sur les 8 patients dysarthriques. Les performances sont données en termes de rappel et précision suivant les bandes de fréquences de 1kHz et la bande totale [0-8]kHz.