

Sur l'utilisation de la reconnaissance automatique de la parole pour l'aide au diagnostic différentiel entre la maladie de Parkinson et l'AMS

Imed Laaridh¹ Julie Mauclair¹

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

imed.laaridh@irit.fr, julie.mauclair@irit.fr

RÉSUMÉ

Cet article présente une étude concernant l'apport du traitement automatique de la parole dans le cadre du diagnostic différentiel entre la maladie de Parkinson et l'AMS (Atrophie Multi-Systématisée). Nous proposons des outils de reconnaissance automatique de la parole pour évaluer le potentiel d'indicateurs de la parole dysarthrique caractérisant ces deux pathologies. Dans ce cadre, un corpus de parole pathologique (projet ANR Voice4PD-MSA) a été enregistré au sein des Centres Hospitaliers Universitaires (CHU) de Toulouse et Bordeaux. Les locuteurs sont des patients atteints de stades précoces de la maladie de Parkinson et d'AMS ainsi que des locuteurs témoins.

Des mesures automatiques caractérisant la qualité de la reconnaissance automatique de la parole ainsi que la prosodie des patients ont montré un intérêt pour la caractérisation des pathologies étudiées et peuvent être considérées comme un outil potentiel pour l'aide à leur diagnostic différentiel.

ABSTRACT

On using automatic speech recognition for the differential diagnosis of Parkinson's Disease and MSA

This article presents a study regarding the contribution of automatic speech processing in the differential diagnosis between Parkinson's disease and MSA (Multi-System Atrophies). We propose to adapt automatic speech recognition tools for the evaluation and study of the dysarthric speech of these two pathologies. A corpus of pathological speech (Voice4PD-MSA) has been recorded at both Toulouse and Bordeaux university hospitals. Speakers are patients with early cases of Parkinson's disease and MSA as well as healthy controls.

Automatic measures characterizing the quality of the automatic speech recognition as well as the prosody of the patients speech have shown an interest for the characterization of the studied pathologies and can be considered as a potential tool for automatic assistance in their differential diagnosis.

MOTS-CLÉS : Dysarthrie, traitement automatique de la parole, parole pathologique, diagnostic différentiel, aide au diagnostique.

KEYWORDS: Dysarthria, automatic speech processing, pathological speech, differential diagnosis, Assistive diagnosis.

1 Introduction

La Maladie de Parkinson (MP) est la maladie neuro-dégénérative la plus courante après la maladie d'Alzheimer, touchant environ 1,5% de la population de plus de 65 ans et environ 170 000 Français (Tison *et al.*, 1994). Son diagnostic clinique se base sur la présence d'une lenteur de mouvement associée à une manifestation motrice de rigidité, tremblement au repos et instabilité posturale (Hughes *et al.*, 1992). Il est souvent confirmé par une réponse positive à la thérapie de remplacement de la dopamine et à l'absence de signes qui suggèrent la présence d'un autre trouble parkinsonien tel que l'Atrophie Multi-Systématisée (AMS). Cependant, le diagnostic définitif ne peut être confirmé que par la découverte de certaines protéines suite à une autopsie (*post-mortem*). L'apparition d'une dysarthrie de type hypokinétique (parole monotone avec des difficultés à initier l'articulation, un débit variable et une voix rauque et soufflée), est souvent associée à la MP (Darley *et al.*, 1975). Ces troubles de la parole apparaissent au début de la MP et même pendant la période pré-symptomatique ; cette observation peut aider à un diagnostic précoce, comme l'a démontré (Harel *et al.*, 2004) dans une étude analysant rétrospectivement des enregistrements de patients atteints de la MP.

L'AMS est une maladie neuro-dégénérative rare d'étiologie inconnue. Elle se caractérise par une combinaison variable de parkinsonisme, atteinte cérébelleuse, troubles dysautonomiques et syndromes pyramidaux. Les premiers symptômes de l'AMS apparaissent généralement à partir de 60 ans (Tison *et al.*, 2000) et le pronostic vital des patients est rapidement engagé. La survie médiane des patients se situe entre 5,8 et 9,5 ans après le diagnostic (Schrag *et al.*, 2008). Le diagnostic clinique de l'AMS présente un degré de confiance limité et il nécessite une confirmation *post-mortem* (Gilman *et al.*, 2008). Comme pour la MP, les patients atteints d'AMS souffrent assez tôt d'une dysarthrie de type mixte présentant des composantes hypokinétiques, spastiques (parole lente et laborieuse avec une articulation imprécise) et ataxiques (parole brusque, ralentie, irrégulière, explosive et scandée) (Darley *et al.*, 1975).

Dans le domaine médical, plusieurs méthodes ont été étudiées pour le diagnostic de l'AMS. L'imagerie par résonance magnétique (IRM) du cerveau aide le clinicien en révélant la présence d'anomalies distinctes chez les patients atteints d'AMS (Barbagallo *et al.*, 2016). Cependant, l'IRM du cerveau peut être normale, en particulier chez les patients dont le diagnostic différentiel entre la MP et l'AMS est difficile. D'autres techniques d'imagerie telles que la tomographie permettent d'identifier des schémas métaboliques distincts entre la MP et l'AMS (Niethammer & Eidelberg, 2012). Mais cette technique est très coûteuse et n'est pas disponible en routine clinique. Au-delà de l'imagerie, plusieurs études sont en cours pour comparer les niveaux de certains marqueurs de dégénérescence axonale dans le plasma des patients atteints de la MP et d'AMS.

L'analyse de voix et de la parole est souvent utilisée pour aider le diagnostic médical des troubles affectant les patients atteints de MP ou d'AMS. L'analyse acoustique automatique, méthode non invasive par excellence, a fait l'objet de plusieurs études afin de fournir une évaluation plus objective de la pathologie et de la voix dysphonique aux orthophonistes. Ont ainsi été étudiés les prononciations de voyelles soutenues (Dibazar *et al.*, 2002), de mots (Fredouille *et al.*, 2019) ou du texte (Laaridh *et al.*, 2015) au travers de mesures telles que le rapport harmonique/bruit, l'énergie (Shama *et al.*, 2006), des mesures aérodynamiques (J. Holmes *et al.*, 2000), la durée des phonèmes, l'étranglement de la voix et l'altération de la prosodie (Ma *et al.*, 2010; Whitehill, 2010).

Peu de travaux se concentrent sur l'identification de la pathologie et la majorité porte sur l'évaluation de la qualité de la parole et la mesure automatique de l'intelligibilité. Dans (Rusz *et al.*, 2015), la comparaison entre les patients atteints de MP et d'AMS a révélé des différences significatives

dans les dimensions ataxiques, hypokinétiques et spastiques de la dysarthrie. Ces dimensions offrent une évaluation objective possible des symptômes de la dysarthrie à l'aide de plusieurs mesures acoustiques, telles que le tremblement vocal, la durée des voyelles, les variations d'intensité, le rapport harmonique/bruit. Cependant, seules les voyelles soutenues et la répétition des syllabes /pataka/ ont été utilisées comme tâches de production et l'étude ne ciblait pas particulièrement le diagnostic différentiel précoce des pathologies.

Contexte et Objectifs En stade précoce, les symptômes de la MP et de l'AMS sont très similaires, en particulier chez les patients atteints d'AMS-P où le parkinsonisme prédomine. Par conséquent, le diagnostic différentiel entre AMS et MP est souvent très difficile à établir aux premiers stades de la maladie, alors que la certitude d'un diagnostic précoce est importante pour le patient en raison du pronostic divergent et de la gravité du pronostic de l'AMS. Aucun marqueur objectif validé n'est actuellement disponible pour guider le clinicien dans ce diagnostic et le besoin de tels marqueurs est donc très élevé dans la communauté neurologique. Comme la dysarthrie est un symptôme commun et précoce dans les deux maladies, mais d'origine différente, notre approche consiste à rechercher comment la caractériser, par le biais du traitement automatique de la parole et du signal, et rechercher des différences entre les patients atteints de MP et d'AMS aux premiers stades des maladies.

Cette étude s'intègre dans un projet pluridisciplinaire financé par l'ANR (ANR VOICE4PD-MSA), auquel participent des professionnels de la médecine, de l'orthophonie (CHU de Bordeaux et Toulouse), de l'informatique et la statistique (INRIA Bordeaux, Université Toulouse 3). Le but de ce projet est de proposer des marqueurs extraits à partir de la parole pour l'assistance dans le diagnostic différentiel entre la MP et l'AMS.

Cette publication donne un état d'avancement de l'étude. L'article est organisé comme suit. La section 2 décrit le corpus construit dans le cadre de ce projet. La section 3 décrit la méthodologie utilisée dans ce travail, basée sur un alignement contraint par le texte ainsi qu'une reconnaissance automatique des phonèmes. Dans la section 4, différentes mesures sont analysées en fonction des populations étudiées et de leurs intérêts pour la caractérisation de chaque pathologie alors que la section 5 fournit quelques conclusions et directions pour de futurs travaux.

2 Corpus

Afin de répondre aux objectifs du projet ANR Voice4PD-MSA, un corpus de parole pathologique est en cours de construction. Les patients enregistrés dans ce corpus souffrent de stades précoces d'une des deux pathologies affectant la parole : la MP et l'AMS. À ce stade du projet, ont été enregistrés 39 patients (20 atteints de la MP, 14 de l'AMS) et 5 témoins sains avec comme but de réunir à terme les enregistrements de 60 patients et 30 locuteurs témoins.

Tous les enregistrements ont été réalisés dans les CHU de Toulouse et de Bordeaux sous le contrôle des phoniâtres en respectant le même protocole expérimental.

Les sessions d'enregistrements ont été réalisées dans les salles de consultation caractérisées par un environnement sonore silencieux. Un enregistreur numérique de haute qualité (de marque Zoom H4n) a été utilisé avec deux microphones, un microphone-casque et un microphone Rode NT1 cardioïde à environ 10 centimètres du locuteur et un microphone AKG C1000S cardioïde branché sur une deuxième station ; les deux stations pour l'enregistrement et l'analyse de la parole pathologique sont

des stations "état de l'art" commercialisées par la société SQ-lab. Tous les microphones réalisent leurs enregistrements en parallèle afin d'éviter au patient de répéter les mêmes tâches plusieurs fois.

Durant la session d'enregistrement, tous les locuteurs ont réalisé les mêmes tâches de production de parole : lecture de texte, parole spontanée, voyelle /a/ tenue, répétition de syllabes /pataka/ et /badaga/, lecture de logatomes. La durée d'une session d'enregistrement est d'environ 15 minutes par locuteur.

Dans ce travail, nous nous concentrons sur la tâche de lecture du texte "La chèvre de monsieur Seguin" d'environ 70 mots. Les résultats présentés dans ce travail se basent sur les enregistrements réalisés par le microphone-casque. Le tableau 1 regroupe les informations liées à ce corpus.

Population	# de locuteurs	Durée moyenne $\pm$ écart-type (sec)
Maladie de Parkinson (MP)	20	22.22 $\pm$ 3.56
Atrophie Multi-Systématisée (AMS)	14	23.86 $\pm$ 4.51
Témoins	5	22.0 $\pm$ 1.82

TABLE 1: Informations liées au corpus Voice4PD-MSA incluant le nombre de locuteurs par type de population et la durée des enregistrements en seconde par locuteur.

3 Méthodologie basée sur le traitement automatique

Afin de pouvoir identifier le potentiel et la pertinence d'indices acoustiques de manière la plus objective possible, nous avons privilégié leur extraction à partir de deux traitements automatiques :

- Le premier est un alignement forcé du signal de parole sur la suite phonétique contrainte par le texte prononcé.
- Le deuxième est une tâche de reconnaissance automatique du signal en une suite de phonèmes, sans connaissance a priori du texte prononcé et sans lexique.

Les indices acoustiques étudiés sont extraits des sorties de chacun de ces traitements ; ils peuvent ensuite être analysés indépendamment du traitement ou comparés.

Dans cette publication, nous proposons des indicateurs temporels basés sur la durée des voyelles et la vitesse d'élocution, ainsi qu'un indicateur spectral issu de la reconnaissance des phones.

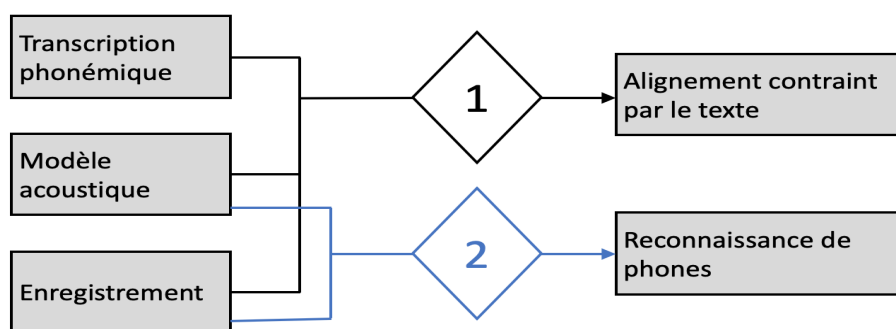


FIGURE 1: Diagramme des deux traitements automatiques utilisés, l'alignement contraint par le texte et la reconnaissance de phones.

Les deux traitements automatiques utilisés sont réalisés grâce à la boîte à outils Kaldi (Povey *et al.*, 2011). Les lexiques phonétisés des deux traitements utilisent le même jeu de phonèmes, à savoir les 37 phonèmes de la langue française. Les modèles acoustiques des phonèmes, des HMM indépendants

du contexte, sont appris sur environ 200 heures d'enregistrements radiophoniques du corpus ESTER (Galliano *et al.*, 2005). La différence entre ces deux traitements réside dans le modèle de langage utilisé comme nous le verrons par la suite.

Alignement contraint par le texte Pour réaliser l'alignement forcé des enregistrements de parole sur la suite de phonèmes a priori prononcée, le système prend en entrée la séquence de mots prononcés lors de chaque enregistrement ainsi que le lexique phonétisé. La séquence de mots est le résultat d'une transcription manuelle, après écoute, où sont introduits les ajouts, suppressions et substitutions de mots ainsi que les pauses, réalisés par les patients par rapport au texte de référence. Des variantes de prononciation pour chaque mot sont introduites dans le lexique, conformément aux règles de prononciation standard et aux règles de liaison potentielles.

L'alignement automatique repose sur une mise en correspondance à l'aide de l'algorithme de Viterbi du signal de parole avec les modèles statistiques associés. De ce processus d'alignement résultent deux segmentations temporelles des enregistrements : l'une correspond à la localisation des phonèmes avec, pour chacun, ses frontières de début et de fin dans le signal ; l'autre permet de localiser les mots et les pauses.

Reconnaissance automatique de phones Le système de reconnaissance automatique de phones utilise comme élément de base le phonème et non le traditionnel "mot". La parole produite par les locuteurs est transcrite en une suite de phonèmes. Pour le modèle de langage, il est supposé que tout phonème peut suivre n'importe quel phonème. En dehors du fait qu'il s'agit de l'hypothèse la plus simple, le choix d'un modèle unigramme équiprobable s'explique par deux raisons principales :

- Utiliser l'information acoustique uniquement lors de la reconnaissance de phones sans correction éventuelle introduite par le lexique de mots et/ou le modèle de langage associé.
- Pouvoir utiliser le même modèle sur la tâche de production de logatomes (pseudo-mots) où lexique et modèle de langage n'ont pas de sens.

Il résulte de ce processus une localisation temporelle des phones, indépendante de la transcription du texte lu, ainsi que leur labellisation.

4 Étude des indicateurs acoustiques et temporels

Comme annoncé, les différentes segmentations permettent d'extraire un grand nombre d'indices, mesures acoustiques (d'ordre spectral) ou temporelles sur les différentes populations du corpus. Pour cette première étude, nous nous sommes attardés sur les mesures suivantes :

- la qualité acoustique comparant la réalisation acoustique d'un son à celle attendue, en prenant en compte les segmentations issues de l'alignement forcé et de la reconnaissance phonémique ;
- la durée des voyelles observées ;
- les mesures globale et locale de la vitesse d'élocution.

4.1 Qualité acoustique au niveau phonème

L'alignement forcé sur le texte lu tente de localiser tous les phones au risque de pénaliser par endroit la qualité acoustique. À l'inverse, le système de reconnaissance phonétique est basé entièrement sur la reconnaissance d'une information acoustique indépendante du locuteur. Une approche pour mesurer la qualité de la réalisation acoustique de chaque enregistrement consiste à aligner temporellement les sorties de l'alignement contraint par le texte et de la reconnaissance automatique phonétique, et à comparer les phones ainsi associés trame à trame (toutes les 10 ms). Le pseudo taux de reconnaissance

phonémique, à savoir le ratio de trames bien reconnues par la reconnaissance automatique, s'apparente à un indicateur de qualité TR :

$$TR = 100 * \frac{\# \text{ trames bien reconnues par la reconnaissance automatique}}{\# \text{ de trames de référence}} \quad (1)$$

Les mesures de l'indicateur TR pour chaque individu des trois populations du corpus d'étude sont rapportées sur la figure 2 : un point représente un individu. Les trois populations sont isolées les unes des autres afin de pouvoir visualiser la moyenne et la variance pour chacune d'elles. Plus ce taux est important, mieux la parole a été reconnue par rapport à une réalisation acoustique standard.

Pour la population "témoin", le taux de reconnaissance phonémique est de 62%, un taux normal pour un système indépendant du locuteur, ce qui valide la qualité des modèles acoustiques appris et justifie l'examen des résultats sur les deux autres populations.

Pour les patients atteints d'AMS, le système a beaucoup plus de difficulté à reconnaître les phonèmes prononcés : seulement 38% de trames sont bien reconnus sur l'ensemble des enregistrements contrairement aux patients atteints de la MP pour qui le TR moyen est de 57%, proche du groupe "témoin". Une différence significative entre ces deux pathologies est confirmée statistiquement ($p < 0.001$) (Anova à un facteur). Ce comportement reflète une information importante que nous retrouvons dans la littérature caractérisant ces deux pathologies : la dysarthrie liée à l'AMS est souvent plus sévère que celle associée à la MP. Cette sévérité est reflétée dans la qualité de la reconnaissance automatique des phonèmes et l'indicateur TR présente un potentiel pour son utilisation pour la différenciation entre les deux pathologies.

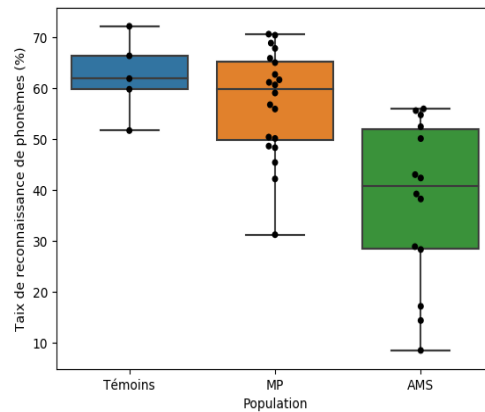


FIGURE 2: Valeur de l'indicateur TR (%) par individu et par population.

4.2 Durée moyenne des voyelles

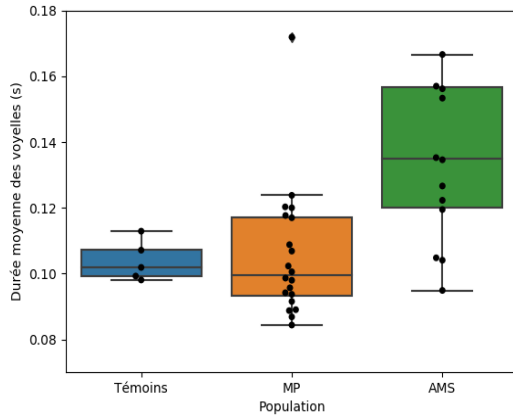
L'allongement des phonèmes, plus particulièrement celui des voyelles, est souvent associé à la dysarthrie ataxique (Darley *et al.*, 1975). Vu que la dysarthrie mixte associée à l'AMS comporte une composante ataxique, absente de la dysarthrie associée à la MP, l'étude de la durée moyenne des voyelles chez les patients des deux populations est pertinente. Ainsi, l'indicateur Dur_{voy} est obtenu en moyennant sur chaque enregistrement, les longueurs des segments reconnus comme une voyelle :

$$Dur_{voy} = \frac{\text{Durée totale des voyelles}}{\# \text{ de voyelles}} \quad (2)$$

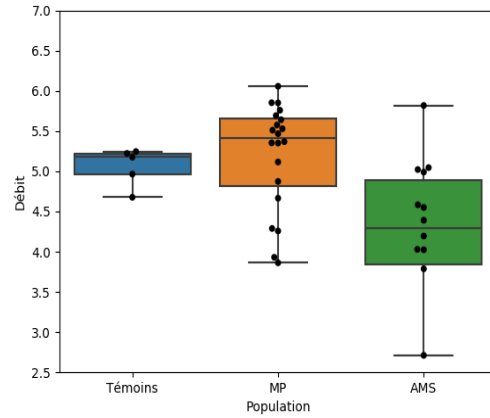
Cet indicateur Dur_{voy} est calculé sur chaque enregistrement à l'issue de la reconnaissance automatique de phonèmes, des travaux antérieurs de segmentation automatique du signal de parole ayant

montré que les frontières détectées par les systèmes de reconnaissance phonétique sont fiables, même si l'identification phonémique en elle même peut être erronée.

Les résultats sont rassemblés sur la figure 3. Les patients atteints d'AMS présentent un comportement distinctif, avec des prononciations de voyelles significativement plus longues que la normale ($p < 0.001$). Ce résultat confirme la pertinence de la mesure Dur_{voy} qui reflète l'allongement des voyelles associé à la dysarthrie ataxique, comportement décrit dans la littérature et basé sur l'analyse perceptive de la parole.



(a) Durée moyenne de voyelle Dur_{voy} (s) par population



(b) Débit de parole $Débit$ par population

FIGURE 3: Dur_{voy} et $Débit$ par population à partir de la reconnaissance automatique de phonème.

4.3 Débit de parole

L'allongement de phonème observé chez les patients atteints de l'AMS agit a fortiori sur leur débit de parole. La dysarthrie ataxique dont souffrent ces patients est caractérisée par un ralentissement du débit de la parole, contrairement à la dysarthrie hypokinétique associée à la MP caractérisée par des irrégularités (voire des accélérations) du débit.

Afin d'estimer automatiquement le débit de parole, nous proposons de calculer le nombre de voyelles prononcées par seconde, noté $Débit$; c'est un estimateur proche du nombre de syllabes par seconde et de la vitesse d'élocution, par ailleurs couramment utilisé (Rouas *et al.*, 2004).

$$Débit = \frac{\# \text{ de voyelles}}{Durée \text{ totale de lecture} - (Durée \text{ pause} + Durée \text{ respiration})} \quad (3)$$

Le paramètre $Débit$ est extrait à partir des sorties de la reconnaissance automatique de phonèmes. Pour éviter tout biais, les pauses et respirations ne sont pas prises en compte pour le calcul.

Les résultats par type de population sont rassemblés sur la figure 3 (b). Comme pour l'étude de la durée des voyelles, des résultats prometteurs s'observent : les patients atteints d'AMS présentent un débit de parole significativement plus lent que les locuteurs témoins ou atteints de la MP ($p < 0.01$). La variance élevée de ce paramètre sur chaque population peut s'interpréter de deux manières différentes : soit les individus au sein d'une même population présentent des comportements très différents l'un de l'autre (plus que probable), soit chaque individu a un comportement très irrégulier au cours du temps et le paramètre, $Débit$, moyenne sur l'ensemble du texte lu, n'est pas significatif, compte tenu de l'instabilité éventuelle d'un débit "local" .

La suite naturelle de cette étude est d'examiner le comportement d'un débit "local" de parole, calculé sur une fenêtre glissante courte (de l'ordre de 2 secondes) au cours du temps.

5 Conclusion

Ce travail vise l'étude et la recherche de marqueurs issus d'un traitement automatique de la parole, afin d'aider un diagnostic différentiel entre les deux maladies, la MP et l'AMS. Les premiers résultats mettent en évidence trois indicateurs issus du pré traitement basé sur la simple reconnaissance automatique de phonèmes, à savoir la qualité de la reconnaissance phonétique, la durée des voyelles reconnues et le débit de parole.

Des différences significatives dans le calcul de ces paramètres sont en adéquation avec les observations effectuées par le corps médical sur les problèmes de prononciation dus à l'AMS, à savoir une qualité acoustique moindre, des voyelles plus longues et un débit de parole plus lent. Ces indicateurs ne conduisent pas à une différenciation entre patients atteints de la MP et témoins non atteints.

A noter également que le calcul de ces paramètres ne demande aucune connaissance a priori qu'une transcription phonétique du texte produit par les patients ; il ne nécessite pas de travail supplémentaire d'annotation ou de transcription sur le signal.

Les dernières expériences ouvrent de nouvelles pistes : une étude du suivi de ces paramètres au cours du temps, sur un texte de longueur raisonnable, comme celui utilisé dans ce protocole, devrait conduire à préciser le comportement des malades en termes de régularité ou non, et sans doute à différencier les patients atteints de la MP des témoins. D'autres corpus devront être étudiés et il est prévu d'augmenter le nombre de patients enregistrés pour consolider ces premières observations.

Remerciements Ce travail fait partie du projet VOICE4PD-MSA (ANR-16-CE19-0010) financé par l'ANR. Nous remercions le Pr. Wassilios Meissner, le Dr. Solange Milhé de Saint Victor, le Dr. Anne Pavy Le Traon, le Pr. Virginie Woisard et les équipes des CHU de Bordeaux et de Toulouse, pour le recueil des données du corpus. Les auteurs remercient également Régine André-Obrecht pour son travail d'analyse des résultats et de relecture ainsi que Khalid Daoudi, porteur du projet.

Références

- BARBAGALLO G., SIERRA-PEÑA M., NEMMI F., TRAON A. P.-L., MEISSNER W. G., RASCOL O. & PÉRAN P. (2016). Multimodal mri assessment of nigro-striatal pathway in multiple system atrophy and parkinson disease. *Movement Disorders*, **31**(3), 325–334.
- DARLEY F. L., ARONSON A. E. & BROWN J. R. (1975). *Motor speech disorders*. Philadelphia : W. B. Saunders and Co.
- DIBAZAR A. A., NARAYANAN S. & BERGER T. W. (2002). Feature analysis for automatic detection of pathological speech. In *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society* [Engineering in Medicine and Biology, volume 1, p. 182–183 : IEEE.
- FREDOUILLE C., GHIO A., LAARIDH I., LALAIN M. & WOISARD V. (2019). Acoustic-phonetic decoding for speech intelligibility evaluation in the context of head and neck cancers. In *Intl Congress of Phonetic Sciences (ICPHs'19)*, p. 3051–3055, Melbourne, Australia.

- GALLIANO S., GEOFFROIS E., MOSTEFA D., CHOUKRI K., BONASTRE J.-F. & GRAVIER G. (2005). ESTER phase II evaluation campaign for the rich transcription of French broadcast news. In *Proceedings of Interspeech'05*, p. 1149–1152, Lisbon, Portugal.
- GILMAN S., WENNING G., LOW P. A., BROOKS D., MATHIAS C., TROJANOWSKI J., WOOD N. W., COLOSIMO C., DÜRR A., FOWLER C. *et al.* (2008). Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*, **71**(9), 670–676.
- HAREL B., CANNIZZARO M. & SNYDER P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient parkinson's disease : A longitudinal case study. *Brain and cognition*, **56**(1), 24–29.
- HUGHES A. J., DANIEL S. E., KILFORD L. & LEES A. J. (1992). Accuracy of clinical diagnosis of idiopathic parkinson's disease : a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, **55**(3), 181–184.
- J. HOLMES R., M. OATES J., J. PHYLAND D. & J. HUGHES A. (2000). Voice characteristics in the progression of parkinson's disease. *International Journal of Language & Communication Disorders*, **35**(3), 407–418.
- 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.
- MA J. K.-Y., WHITEHILL T. L. & SO S. Y.-S. (2010). Intonation contrast in cantonese speakers with hypokinetic dysarthria associated with parkinson's disease. *Journal of Speech, Language, and Hearing Research*.
- NIETHAMMER M. & EIDELBERG D. (2012). Metabolic brain networks in translational neurology : concepts and applications. *Annals of neurology*, **72**(5), 635–647.
- POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P. *et al.* (2011). The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on automatic speech recognition and understanding : IEEE Signal Processing Society*.
- ROUAS J.-L., FARINAS J. & PELLEGRINO F. (2004). Evaluation automatique du débit de la parole sur des données multilingues spontanées. *XXVe Journées d'Etude sur la Parole (JEP 2004)*, Fes, Maroc, p. 437–440.
- RUSZ J., BONNET C., KLEMPÍŘ J., TYKALOVÁ T., BABOROVÁ E., NOVOTNÝ M., RULSEH A. & RŽIČKA E. (2015). Speech disorders reflect differing pathophysiology in parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Journal of neurology*, **262**(4), 992–1001.
- SCHRAG A., WENNING G. K., QUINN N. & BEN-SHLOMO Y. (2008). Survival in multiple system atrophy. *Movement Disorders*, **23**(2), 294–296.
- SHAMA K., KRISHNA A. & CHOLAYYA N. U. (2006). Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP Journal on advances in signal processing*, **2007**, 1–9.
- TISON F., DARTIGUES J., DUBES L., ZUBER M., ALPEROVITCH A. & HENRY P. (1994). Prevalence of parkinson's disease in the elderly : a population study in gironde, france. *Acta neurologica scandinavica*, **90**(2), 111–115.
- TISON F., YEKHLEF F., CHRYSOSTOME V. & SOURGEN C. (2000). Prevalence of multiple system atrophy. *The Lancet*, **355**(9202), 495–496.
- WHITEHILL T. L. (2010). Studies of chinese speakers with dysarthria : informing theoretical models. *Folia Phoniatrica et Logopaedica*, **62**(3), 92–96.