

# Méthodes objectives issues du traitement automatique de la parole pour la recherche de zones “déviantes” dans la parole dysarthrique

Pierre CLÉMENT, Corinne FREDOUILLE

Laboratoire Informatique d’Avignon - CERILIA - Université d’Avignon, France  
Courriel : pierre.clement@etd.uapv.fr, corinne.fredouille@univ-avignon.fr

## ABSTRACT

A deficiency or a malfunction of an enzyme present in the lysosomes causes the lysosomal diseases. One of the symptoms due to these diseases is the dysarthria. Dysarthria is defined as a speech disorder caused by a damage of the nervous system. At the present time, the severity degree of dysarthria is measured perceptively by clinicians. Even if various perceptual and visual criterias have been defined to evaluate dysarthria, this evaluation depends directly on the clinician, leading to subjectivity factor. Because of that, the evaluation of this pathology should be done objectively. This article describes the methodologies proposed to find “deviant” part of dysarthric speech. The study of its characteristics with automatic speech processing tools will be the basis of the analysis of the dysarthric speech.

## 1. INTRODUCTION

Une déficience ou un dysfonctionnement d’une enzyme présente dans les lysosomes est à l’origine des maladies dites lysosomales. Cet ensemble de maladies génétiques touche essentiellement les enfants, environ un sur 7500. Parmi les nombreux symptômes pouvant être liés à ces maladies, ces jeunes patients peuvent être atteints de dysarthrie. Ce symptôme, présent aussi pour d’autres maladies (telles que la maladie de Parkinson), est un trouble de l’élocution qui, suivant sa sévérité, peut traduire une incapacité totale à communiquer de manière intelligible.

A l’heure actuelle, le degré de sévérité de la dysarthrie est évalué de façon perceptive par les cliniciens. Bien qu’un certain nombre de critères sur lesquels les cliniciens s’appuient pour cette évaluation soient définis ([Dar69a], [Dar69b], [Dar75], [Duf05] et [Auz06]), il n’en reste pas moins vrai que cette dernière ne peut-être considérée comme totalement fiable, de par son caractère subjectif.

Il existe deux grands types de variabilité pour cette évaluation : la variabilité inter-cliniciens et intra-clinicien. Face aux limites de l’évaluation actuelle de la parole dysarthrique, liées à ces deux variabilités, la mise en place d’un nouveau protocole d’évaluation plus objectif devient une nécessité. Dans l’idéal, ce nouveau protocole s’intégrerait dans l’examen clinique du patient en vue d’apporter une aide au diagnostic et au suivi thérapeutique. Les solutions envisagées dans ce cadre peuvent reposer soit sur des technologies instrumentales (à base de capteurs de mesures acoustiques, aérodynamiques et physiologiques) soit sur des systèmes de traitement automatique de la parole.

Néanmoins, un tel système d’évaluation de la parole dysarthrique nécessite avant tout de meilleures connaissances de ce type de parole. Cet article s’inscrit dans ce cadre et décrit les méthodologies suivies pour repérer des zones déviantes de la parole dysarthrique avec des outils de traitement automatique de la parole. Dans un premier temps, nous expliquerons ce que sont la dysarthrie et les maladies lysosomales. Nous présenterons ensuite les méthodologies proposées. Enfin, nous présenterons nos conclusions sur les analyses préliminaires que nous avons effectuées.

## 2. PATHOLOGIE

### 2.1. Dysarthrie

La dysarthrie est un trouble de l’élocution ayant pour origine un trouble neurologique. Les flux nerveux contrôlant les articulateurs nécessaires à l’élocution proviennent de différentes zones du cortex cérébral. La dysarthrie peut être le fruit de la lésion d’un seul des neurones par lesquels transitent les informations de parole. Cette lésion entraîne une altération des mouvements de la langue, de la gorge, des lèvres, ...

Les symptômes de la dysarthrie sont :

- parole marmonnée
- débit de parole altéré, rythme saccadé
- parole plus nasale
- dysphonie
- voix rauque
- voix monotone

Il se peut aussi que le patient éprouve des difficultés à avaler et qu’il souffre de dysphagie<sup>1</sup>.

Parmi les pathologies à l’origine de ces lésions, on peut notamment trouver la maladie de Parkinson, une complication de la syphilis entraînant une paralysie totale, la sclérose en plaques, la sclérose latérale amyotrophique (SLA), les maladies lysosomales telles que Niemann-Pick ou encore la maladie de Tay Sachs (appelée aussi gangliosidose GM2).

En raison de sa grande variabilité, il est devenu nécessaire de proposer une classification de la dysarthrie, pour la pratique clinique et les travaux de recherche. En 1957, Grewel ([Gre57]) s’inspire du travail de Peacher<sup>2</sup> ([Pea50]) afin de diviser les dysarthries en quatorze classes. Ces classes

<sup>1</sup>La dysphagie est une sensation de gêne ressentie lors du passage d’aliments dans la bouche, le pharynx ou l’œsophage.

<sup>2</sup>Peacher a effectué en 1950 une première classification des maladies dysarthriques.

reposent sur la localisation des lésions cérébrales, avec parfois une information supplémentaire concernant la pathologie associée ou l'étiologie<sup>3</sup>. Le fait que cette classification repose sur des critères différents (anatomiques, étiologiques ou physiopathologiques<sup>4</sup>) en est son principal inconvénient.

L'équipe de Darley et al. de la Mayo Clinic a proposé une classification de la dysarthrie ([Dar69a], [Dar69b] et [Dar75]) en six classes reposant uniquement sur les troubles neurologiques qui ont provoqué la dysarthrie. Cette classification, complétée en 2005 par Duffy ([Duf05]) qui ajouta deux nouvelles classes, est la classification actuellement utilisée.

## 2.2. Maladies lysosomales

Les lysosomes sont des structures contenues au cœur de nos cellules. Ils ont pour rôle de recycler les métabolites<sup>5</sup>. Les lysosomes remplissent leur fonction grâce à trois types d'enzymes qu'ils contiennent : des lipases, des protéases et des osidases. L'altération du fonctionnement d'une de ces trois enzymes, due à une déficience génétique, est à l'origine des "maladies lysosomales".

Les enzymes sont créées et codées par les gènes. Si le gène code mal ou ne crée pas l'enzyme, alors le bon fonctionnement de celle-ci est compromis. Chaque enzyme intervient sur le recyclage d'un métabolite. Par conséquent, si l'une des trois fonctionne mal ou ne fonctionne pas, le métabolite concerné ne sera pas traité ou pas correctement par le lysosome. Avec le temps, les métabolites s'accumulent, les cellules s'engorgent et entraînent un défaut de fonctionnement des différents organes.

Ainsi, on englobe sous le terme de "maladies lysosomales" une cinquantaine d'affections handicapantes de l'enfant à l'adulte. Nous nous intéressons ici à 2 maladies particulières :

- la maladie de Niemann-Pick type C (accumulation de cholestérol dans les cellules)
- la maladie de Tay Sachs (accumulation de gangliosides GM2 dans les cellules)

Il semblerait que la dysarthrie résultante de ces pathologies soit liée à diverses atteintes dans le système nerveux, cette dysarthrie peut donc être classée comme dysarthrie mixte ([Dar69a], [Dar69b] et [Dar75]).

## 3. MÉTHODOLOGIES

### 3.1. Corpus

Nous avons à notre disposition un corpus constitué à l'hôpital de la Pitié-Salpêtrière où les patients ont été enregistrés par une orthophoniste.

Ce corpus comprend des enregistrements de quatre patients féminins et quatre patients masculins. Ces huit patients atteints de deux maladies lysosomales différentes :

- *Tay Sachs* : deux femmes et un homme
  - *Niemann-Pick type C* : deux femmes et trois hommes
- Tous ces patients suivent un traitement thérapeutique expérimental basé sur la molécule de miglustat. Ce traitement doit, normalement, apporter une stabilisation des symptômes des patients (dont la dysarthrie).

Les enregistrements, effectués sur une période de deux ans, à plusieurs mois d'intervalle, contiennent plusieurs types de parole. Lors de nos expériences, nous n'avons utilisé que les enregistrements de la lecture d'un texte (le cordonnier qui contient 19 phrases), pour les raisons suivantes :

- il nous fallait une quantité de parole relativement importante, homogène et présentant une variabilité phonétique élevée.
- il nous fallait de la parole commune à chacun, afin de pouvoir comparer les résultats entre eux. La parole spontanée est donc à exclure.

La parole lue semble par conséquent la mieux adaptée même si cela entraîne la perte d'autres informations (comme la spontanéité par exemple).

Nous avons aussi, pour chaque patient, son "contrôle". Il s'agit d'enregistrements d'une personne, sans pathologie particulière (saine), associés aux enregistrements d'un patient. L'appariement patients/contrôles a été effectué par des orthophonistes sur des critères physiques de similarité (sexe et tranche d'âge).

Le corpus est constitué d'environ 5h45 d'enregistrement au total, tout type de parole, de maladie, de genre etc. confondus et les "contrôles" inclus. La durée totale des enregistrements de parole lue, donc utilisée lors de nos expériences, représente 23% du corpus, soit environ 1h20 d'enregistrement.

### 3.2. Présentation des méthodologies

La figure 1 décrit les deux grandes méthodologies que nous avons suivies afin de rechercher des zones "déviantes" dans la parole dysarthrique. En utilisant les outils de traitement automatique de la parole, nous souhaitons mettre en évidence des entités linguistiques (phonèmes, classes de phonèmes etc...) pertinentes pour distinguer la parole dysarthrique de la parole normale par rapport à cet objectif.

Dans un premiers temps, les patients et les contrôles seront comparés à un modèle générique de parole estimé sur une grande population de locuteurs normophoniques (corpus "Bref", [Lam91]). Ce modèle générique de parole reposera sur les techniques de modélisation statistique à base de GMM (Gaussian Mixture Model) éprouvées en reconnaissance automatique de la langue (RAL, [Bim04]). Nous calculerons des scores de vraisemblances associés aux signaux de parole par rapport à ce modèle. A partir de ces scores, nous produirons une cartographie des zones de déviance et des histogrammes (expliqués en section 3.4) afin d'observer si la parole des patients est très éloignée de notre modèle générique.

Dans un deuxième temps, les signaux de parole seront comparés à des modèles spécifiques de parole. En effet, l'alignement automatique nous fournira des scores de confiance par phonème à partir desquels seront produits à nouveau une cartographie des zones de déviance et des his-

<sup>3</sup>L'étiologie est l'étude des causes et des facteurs d'une pathologie.

<sup>4</sup>La physiopathologie est la discipline biologique qui traite des dérèglements de la physiologie, c'est-à-dire du mode de fonctionnement normal des éléments constitutifs du corps humain, d'un animal ou d'un végétal.

<sup>5</sup>Les métabolites sont des matières issues du fonctionnement cellulaire.

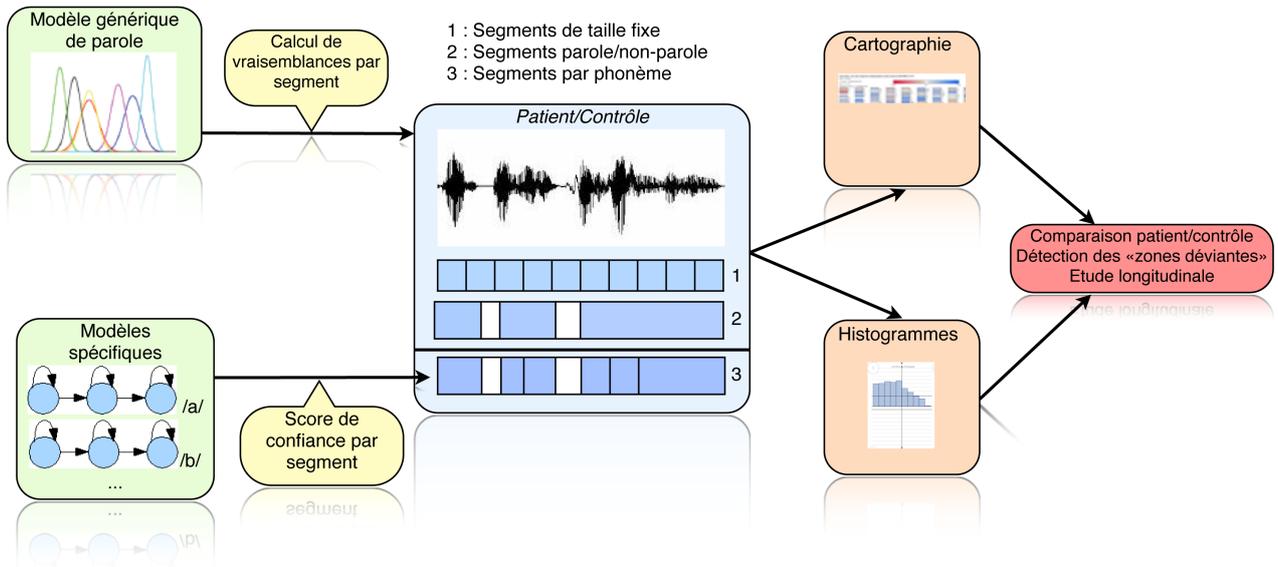


FIG. 1: Méthodologies

togrammes.

### 3.3. Calcul des scores

Suivant les deux méthodologies décrites précédemment, deux types de score seront produits :

- score de vraisemblance dans le cadre du modèle générique de la parole
- score de confiance dans le cadre des modèles spécifiques

**Vraisemblance** Lors de la phase de reconnaissance, un signal de parole est comparé au modèle statistique généré lors de la phase d'apprentissage sur un corpus de parole normophonique ("Bref"). Cette comparaison repose sur la mesure d'une similarité entre les vecteurs acoustiques  $y_t$  extraits du signal et le modèle  $X$ . Cette similarité, appelée vraisemblance, peut-être définie par :

$$L(y_t|X) = \sum_{i=1}^M p_i L_i(y_t) \quad (1)$$

$$\text{où } L_i(y_t) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(y_t - \bar{x}_i)^T (\Sigma_i)^{-1} (y_t - \bar{x}_i)}$$

La vraisemblance calculée à partir de l'équation [1] correspond à l'estimation de la représentativité du vecteur de test  $y_t$  par le mélange de gaussiennes  $X$ .

**Score de confiance** L'alignement automatique contraint par le texte ([Bru93]) est une tâche du domaine de la reconnaissance automatique de la parole (RAP). Il consiste, à partir d'un signal de parole et de sa transcription en mots (texte), à détecter automatiquement les frontières dans le signal de parole de chacun des phonèmes présents dans le texte prononcé (illustré par la figure 2). Cet alignement automatique repose sur un décodage du signal de parole par l'algorithme Viterbi basé sur des modèles statistiques (HMM<sup>6</sup>) associés à chacun des phonèmes recherchés. De

cet alignement automatique résulte une segmentation temporelle en phonèmes comprenant, pour chaque phonème, ses frontières de début et de fin au sein du signal de parole ainsi qu'un score de vraisemblance (équation 1) entre le segment de parole associé et le modèle statistique correspondant.

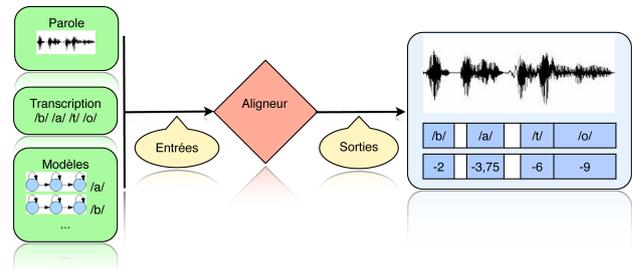


FIG. 2: Alignement contraint par le texte

Ces scores de vraisemblance, notés  $L_i(y_t)$  n'étant pas normalisés, il est difficile de savoir si un "mauvais" résultat est dû à la qualité du signal ou à une mauvaise prononciation du phonème. Pour pallier ce problème, les résultats sont normalisés par les vraisemblances résultantes du décodage acoustico-phonétique (DAP).

Le DAP est en fait un alignement non-contraint par le texte (tâche également issue de la RAP, illustré par la figure 3). N'ayant pas accès à la transcription en mots, le système n'a aucun a priori sur la séquence de phonèmes prononcés par le locuteur. Pour un segment de parole donné, un système DAP compare, par le biais d'une mesure de vraisemblance, les paramètres de ce dernier avec le modèle statistique de tous les phonèmes connus du système. Le phonème associé à la plus grande vraisemblance est alors désigné comme étant celui réalisé dans le segment de parole traité.

<sup>6</sup>HMM : Hidden Markov Model, Modèle de Markov Caché. Ce modèle est un automate à états finis ([Bak90] et [Jel76]). Les HMM permettent de modéliser des unités acoustiques indépendantes plus longues

(phonèmes par exemple) que celles modélisées par des GMM. A noter qu'un GMM est en fait un HMM particulier comportant un seul état, à comparer, par exemple, aux HMM à trois états, généralement utilisés, pour la modélisation des phonèmes..

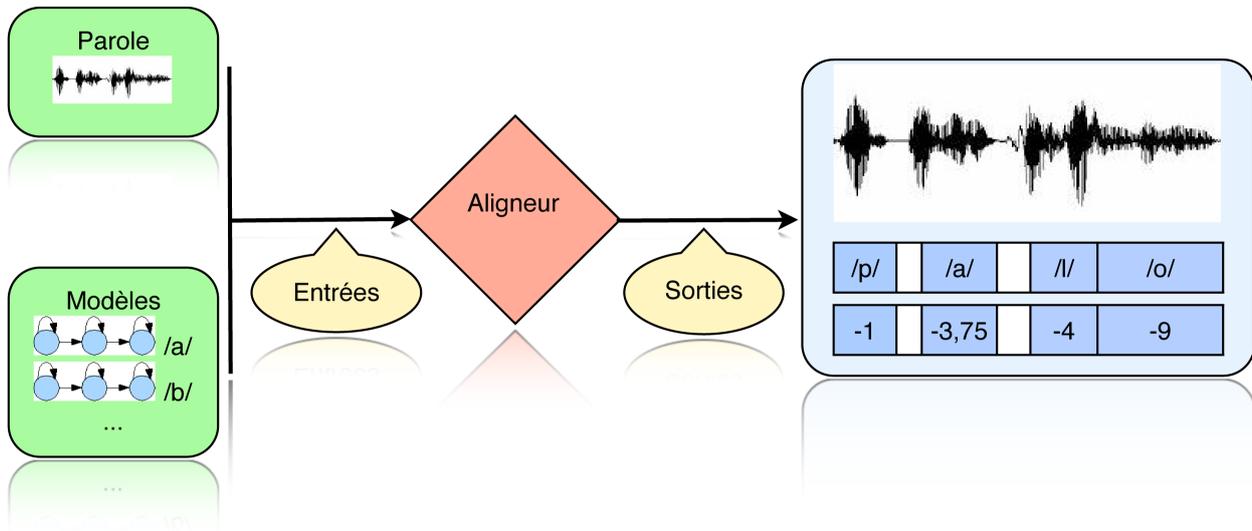


FIG. 3: Décodage acoustico-phonétique (DAP)

Finalement, la normalisation des scores de vraisemblance issus de l'alignement contraint par le texte repose sur l'équation suivante :

$$L_i^{norm}(y_t) = \log \frac{L_i^C(y_t)}{L_i^{NC}(y_t)} \quad (2)$$

où  $L_i^{norm}(y_t)$  est la vraisemblance issue de la normalisation,  $L_i^C(y_t)$  la vraisemblance issue de l'alignement contraint par le texte et  $L_i^{NC}(y_t)$  la vraisemblance issue du DAP.

Ces scores normalisés seront considérés comme des scores de confiance et associés à chaque segment phonémique.

### 3.4. Présentation des résultats

Afin de faciliter l'analyse des scores issus des deux méthodologies mises en place en vue de la détection des zones de déviance de la parole dysarthrique ou de la comparaison entre enregistrements d'un même patient (étude longitudinale), deux grands types de représentation sont proposés.

**Cartes** Le premier repose sur la mise en place d'une cartographie des enregistrements des patients/contrôles. Chaque enregistrement est associé à une carte, illustrant par segment de parole (par exemple un phonème), l'écart du score obtenu par rapport à une référence donnée. Cette référence, qui résulte d'un seul ou plusieurs enregistrements, nous fournit l'étendue des scores de comparaison en fonction de trois points clés que sont le minimum, la médiane et le maximum. Ces valeurs nous permettent d'établir une échelle de référence graduée (note : un dégradé de couleurs, allant du minimum au maximum, sera utilisé pour la visualisation de cette échelle sur les cartes) sur laquelle sera positionné chacun des scores des segments associé à un enregistrement à analyser.

Le choix de la référence est important et très dépendant de la tâche visée. En effet, s'il s'agit de comparer les enregistrements d'un même patient dans le cadre d'une étude longitudinale (plusieurs temps d'enregistrement pour étudier l'évolution de la maladie), l'enregistrement à  $T_0$  peut être considéré comme la référence

contre laquelle les enregistrements à  $T+n$  seront comparés. Dans notre étude, différentes références peuvent être choisies comme illustré par le tableau 1. Suivant l'analyse visée (détection de zones déviantes, étude longitudinale par patient, ...), la référence adéquate sera donc utilisée. La figure 4 est un exemple de carte, la référence étant prise sur l'enregistrement courant du patient.

Références possibles pour l'établissement du minimum et du maximum	Comparaisons possibles
Enregistrement courant	Aucune
Tous les enregistrements d'un même patient	Comparaisons entre les enregistrements
Tous les enregistrements de tous les contrôles	Comparaisons entre tous les fichiers
Tous les enregistrements de tous les patients et contrôles	Comparaisons entre tous les fichiers

TAB. 1: Choix de la référence pour les cartes et incidences sur les comparaisons.

**Histogrammes** Le deuxième type de représentation des scores repose sur l'établissement d'histogrammes. Les histogrammes permettent de synthétiser les écarts des scores, par rapport à notre référence pour chacun des enregistrements des contrôles et des patients. Cette synthèse peut être fournie pour un phonème donné (considérant l'ensemble des productions de ce même phonème dans le texte), pour une classe de phonème, pour une phrase... La figure 5 montre l'histogramme du phonème [l] d'un patient. En partie gauche de l'axe des ordonnées on retrouve les moyennes des écarts des scores pour tous les contrôles sur le phonème étudié et en partie droite celles du patient. Chaque barre de l'histogramme représente un enregistrement. Dans ce cadre, nous avons défini deux seuils, le premier se situant à la moyenne des écarts des scores sur les enregistrements des contrôles moins leur écart-type

Min: -41.2945

Average: -21.4662885360186

Max: -15.3314

Standard deviation: 2.7100886879169

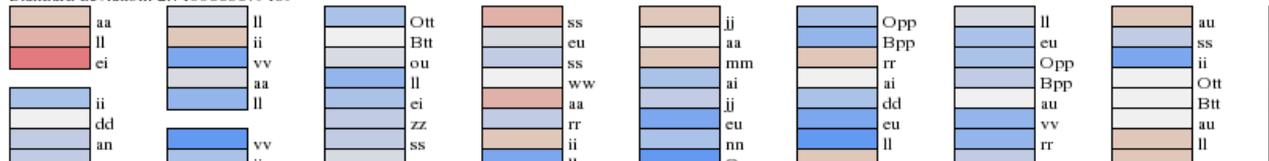


FIG. 4: Exemple d'une carte

et le second à la moyenne des écarts des scores moins deux fois leur écart-type. Ces seuils représentent respectivement 98% et 84% des données, en supposant que la répartition des données soit gaussienne. Ils nous permettront par la suite de faire les observations suivantes : si la moyenne des valeurs, pour un phonème d'un patient est au dessus de ces seuils, alors il ne sera pas considéré comme déviant, il sera dans la "normalité" telle que nous l'avons définie.

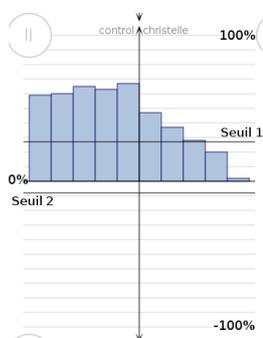


FIG. 5: Exemple d'histogramme de répartition des moyenne des écarts des scores par rapport à la référence

## 4. ANALYSES

### 4.1. Analyses par segment

Afin d'éviter de fausser les résultats avec des statistiques sur les pauses, nous avons segmenté les fichiers en parole/non-parole, ceci nous permettant de nous focaliser uniquement sur les temps de parole. Cette segmentation est réalisée de manière automatique : un détecteur basé sur le coefficient énergie fournit la segmentation d'un fichier paramétrisé en parole/non-parole.

Nous avons également utilisé la segmentation phonémique fournie par le système d'alignement contraint par le texte. A partir de ces deux types de segmentation, la comparaison entre les signaux des patients/contrôles et notre modèle générique a permis de fournir un score de vraisemblance moyen par segment, à partir desquels cartes et histogrammes sont produits.

L'analyse des cartes et histogrammes alors générés ne nous a pas permis de mettre en évidence des phrases, mots, classes de phonèmes ou phonèmes déviant. En effet, aucune comparaison aux contrôles ou étude longitudinale (comparaison entre les enregistrements d'un patient) ne met en évidence des portions de parole nettement déviantes et communes aux différents patients et contrôles.

Le fait que rien ne ressorte de ces analyses peut s'expliquer de la façon suivante : il semblerait que le corpus dysarthrique ne soit pas aussi éloigné du modèle générique du point de vue de la machine. Notre système arrive, avec plus ou moins de succès, à associer ces segments de parole dysarthrique à de la parole normale. De cette observation on peut émettre l'hypothèse suivante : les sons produits par les patients dysarthriques sont connus de nos modèles. Cette hypothèse, très importante car non-attendue avant les expériences, semble aller dans le sens des évaluations perceptives préliminaires réalisées sur ce même corpus. En effet, les auditeurs semblent reconnaître les événements acoustiques perçus mais soulignent le fait que leur réalisation dans le signal de parole est anormalement positionnée. Par exemple, si un phonème [p] est remplacé par un [b] car le patient n'arrive pas à prononcer le [p] de "petit", le système fournira une vraisemblance correcte, cet événement ([b]) étant bien présent dans le modèle générique de parole, bien que le phonème attendu soit [p].

### 4.2. Analyses des modèles spécifiques

Comme vu précédemment, la cartographie et les histogrammes reflètent ici les scores de confiance fournis par l'aligneur contraint par le texte, c'est-à-dire issus des modèles spécifiques de parole.

Les analyses ont été effectuées à deux échelles :

- par phonème
- par classe de phonèmes

Ceci dans le but d'éviter qu'une particularité de prononciation chez un patient non-lié à la pathologie fausse les résultats si on ne considère que les phonèmes. Les classes de phonèmes nous permettent en effet d'obtenir des moyennes plus représentatives car calculées sur plus d'observations. En effet, si on prend par exemple le phonème /b/, il n'est prononcé que deux fois. Si jamais ce phonème réagit bien au système (c'est à dire que son score moyen est en dessous des seuils définis), on ne pourrait malheureusement rien déduire car la moyenne résultante ne serait pas représentative du cas présent.

Une première analyse préliminaire des cartes et histogrammes a permis de mettre en évidence différents axes d'observation à croiser ultérieurement avec une analyse perceptive et phonétique :

- contrairement à la méthodologie précédente (basée sur un modèle générique), les cartographies des différents patients montrent des zones de déviance comparées à leur contrôle. L'utilisation des modèles spécifiques semble donc pertinente.
- on peut observer que le système est moins réactif sur les

hommes que sur les femmes, notamment en observant la répartition des classes de phonèmes par rapport à nos deux seuils. En effet, très peu de classes de phonèmes, chez les hommes, se retrouvent en dessous du premier seuil de “normalité”.

- de même, on peut constater, chez les femmes, que les patientes atteintes de la maladie de Tay Sachs présentent moins de variabilité dans la répartition des classes de phonèmes par rapport à nos seuils que les patientes atteintes de la maladie de Niemann-Pick type C.
- nous avons également observé que les consonnes voisées et les voyelles nasales sont les classes de phonèmes qui montrent le plus de variabilité vis à vis de nos deux seuils de “normalité”. Par exemple, les consonnes voisées sont la seule classe à être en dessous du seuil haut pour la patiente jugée perceptivement comme la moins dysarthrique. Pour les hommes, ce sont principalement les voyelles nasales qui varient.

Ces quelques observations préliminaires semblent donc montrer l’intérêt de cette méthodologie. Il reste maintenant à comparer les zones de déviance détectée automatiquement par le système aux résultats d’analyse de ces mêmes signaux d’un point de vue phonétique (en cours de réalisation).

## 5. CONCLUSION ET PERSPECTIVES

Nous avons pu, grâce à ce travail, mettre en place deux grandes méthodologies nous permettant de représenter un signal de parole dysarthrique sous forme de cartes et d’histogrammes. Ces données seront, dans un avenir proche, analysées en détails. Elles seront notamment comparées aux analyses perceptives réalisées par des experts en phonétique afin de voir si nos cartes permettent de mettre en évidence des déformations de la parole relevées de façon manuelle par ces derniers.

Sur le travail effectué, la taille du corpus ne nous permet malheureusement pas d’affirmer ou d’infirmes les hypothèses, observations et autres remarques que nous avons effectuées. L’exemple le plus marquant est probablement l’observation émise à partir des dernières analyses : on ne peut pas confirmer que le système est plus réactif avec les patientes atteintes de la maladie de Niemann-Pick type C qu’avec les patientes atteintes de la maladie de Tay Sachs, par manque de cas d’études.

Par ailleurs, l’un des inconvénients actuels des cartes est qu’elles ne se basent que sur la segmentation propre à chaque patient/contrôle. Or, ces segmentations pouvant être différentes d’un patient à l’autre (ajout de mots possibles, omission de phonèmes, ...), les cartes ne sont pas, actuellement, directement comparables. Une solution à mettre en place serait l’utilisation d’un “masque” qui ne contiendrait que la parole commune à tous les enregistrements. Tout ce qui serait en dehors de ce masque ne serait pas analysé. A noter que cette étape ne pourra pas se faire de façon automatique, l’intervention humaine sera nécessaire.

Par ailleurs, les méthodologies mises en place seront testées suivant différentes configurations :

- de nouveaux modèles spécifiques, plus complets et plus précis seront expérimentés (issus des campagnes d’évaluation ESTER)
- des modèles dépendants du genre seront également étudiés

Enfin ce travail sera appliqué à un corpus plus conséquent comprenant d’autres types de dysarthrie telle que celle associée à la maladie de Parkinson et à la SLA.

## 6. REMERCIEMENTS

Cette étude a reçu le soutien de l’association “Vaincre les Maladies Lysosomales (VML)” (soutient financier) et de l’hôpital de la Pitié-Salpêtrière (partage des enregistrements).

## RÉFÉRENCES

- [Auz06] Auzou, P. and Rolland-Monnoury, V. (2006), *Batterie d’évaluation clinique de la dysarthrie*, ORTHO. 1
- [Bak90] Baker, J.K. (1990), “Stochastic modeling for automatic speech understanding”, , pp. 297–307. 3
- [Bim04] Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska, D. and Reynolds, D.A. (2004), “A tutorial on text-independent speaker verification”, *EURASIP Journal on Applied Signal Processing*, vol. 39, pp. 430–451. 2
- [Bru93] Brugnara, F., Falavigna, D. and Omologo, M. (1993), “Automatic segmentation and labeling of speech based on hidden markov models”, *Speech Communication*, vol. 12, pp. 357–370. 3
- [Dar69a] Darley, F.L., Aronson, A.E. and Brown, J.R. (1969), “Clusters of deviant speech dimensions in the dysarthrias.”, *J Speech Hear Res*, vol. 12, pp. 462–496. 1, 2
- [Dar69b] Darley, F.L., Aronson, A.E. and Brown, J.R. (1969), “Differential diagnostic patterns of dysarthria.”, *J Speech Hear Res*, vol. 12, pp. 246–269. 1, 2
- [Dar75] Darley, F.L., Aronson, A.E. and Brown, J.R. (1975), *Motor speech disorders*, Philadelphia. 1, 2
- [Duf05] Duffy, J.R. (2005), *Motor speech disorders : substrates, differential diagnosis, and management*, Mosby. 1, 2
- [Gre57] Grewel, F. (1957), “Classification of dysarthrias.”, *Acta Psychiatr Neurol Scand*, vol. 32, pp. 325–337. 1
- [Jel76] Jelinek, F. (1976), “Continuous speech recognition by statistical methods”, vol. 64, pp. 532–556. 3
- [Lam91] Lamel, L.F., Gauvain, J.L. and Eskénazi, M. (1991), “BREF, a large vocabulary spoken corpus for french”, Genoa, Italy, pp. 505–508. 2
- [Pea50] Peacher, W.G. (1950), “The etiology and differential diagnosis of dysarthria.”, *J Speech Disord*, vol. 15, pp. 252–265. 1