Évaluation de la parole dysarthrique : Apport du traitement automatique de la parole face à l'expertise humaine

Imed Laaridh

ACADÉMIE D'AIX-MARSEILLE
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

THÈSE

présentée à l'Université d'Avignon et des Pays de Vaucluse
pour obtenir le diplôme de DOCTORAT

SPÉCIALITÉ : Informatique

École Doctorale 536 « Sciences et Agrosciences »
Laboratoire d'Informatique (EA 4128)

Évaluation de la parole dysarthrique : Apport du traitement automatique de la parole face à l'expertise humaine

par
Imed LAARIDH

Soutenue publiquement le - devant un jury composé de :

M.	Francis GRENEZ	Professeur, BEAMS/ULB, Bruxelles	Rapporteur
M.	Philippe BOULA DE MAREÜIL	DR, CNRS/LIMSI, Paris	Rapporteur
M.	Philippe BLACHE	DR, LPL/CNRS, BLRI, Aix-en-Provence	Examinateur
M.	Thomas PELLEGRINI	MCF, IRIT, Toulouse	Examinateur
M^{me}.	Danièle ROBERT	Praticien Hospitalier, LPL/CNRS, Aix-en-Provence	Examinateur
M.	Jean François BONASTRE	Professeur, LIA, Avignon	Directeur
M^{me}.	Corinne FREDOUILLE	MCF (HDR), LIA, Avignon	Co-Directeur
M^{me}.	Christine MEUNIER	CR (HDR), LPL/CNRS, Aix-en-Provence	Co-Encadrante

Laboratoire Informatique d'Avignon

Résumé

La dysarthrie est un trouble de la parole affectant la réalisation motrice de la parole causée par des lésions du système nerveux central ou périphérique. Elle peut être liée à différentes pathologies : la maladie de Parkinson, la Sclérose Latérale Amyotrophique (SLA), un Accident Vasculaire Cérébral (AVC), etc. Plusieurs travaux de recherche ont porté sur la caractérisation des altérations liées à chaque pathologie afin de les regrouper dans des classes de dysarthrie. La classification la plus répandue est celle établie par F. L. Darley comportant 6 classes en 1969, (complétée par deux classes supplémentaires en 2005)

Actuellement, l'évaluation perceptive (à l'oreille) reste le standard utilisé dans la pratique clinique pour le diagnostique et le suivi thérapeutique des patients. Cette approche est néanmoins reconnue comme étant subjective, non reproductible et coûteuse en temps. Ces limites la rendent inadaptée à l'évaluation de larges corpus (dans le cadre d'études phonétiques par exemple) ou pour le suivi longitudinal de l'évolution des patients dysarthriques.

Face à ces limites, les professionnels expriment constamment leur besoin de méthodes objectives d'évaluation de la parole dysarthrique. Les outils de Traitement Automatique de la Parole (TAP) ont été rapidement considérés comme des solutions potentielles pour répondre à cette demande.

Le travail présenté dans ce rapport s'inscrit dans ce cadre et étudie l'apport que peuvent avoir ces outils dans l'évaluation de la parole dysarthrique, et plus généralement pathologique.

Dans ce travail, une approche pour la détection automatique des phonèmes anormaux dans la parole dysarthrique est proposée et son comportement est analysé sur différents corpus comportant différentes pathologies, classes dysarthriques, niveaux de sévérité de la maladie et styles de parole. Contrairement à la majorité des approches proposées dans la littérature permettant des évaluations de la qualité globale de la parole (évaluation de la sévérité, intelligibilité, etc.), l'approche proposée se focalise sur le niveau phonème dans le but d'atteindre une meilleure caractérisation de la dysarthrie et de permettre un feed-back plus précis et utile pour l'utilisateur (clinicien, phonéticien, patient). L'approche s'articule autours de deux phases essentielles : (1) une première phase d'alignement automatique de la parole au niveau phonème (2) une classification de ces phonèmes en deux classes : phonèmes normaux et anormaux.

L'évaluation de l'annotation réalisée par le système par rapport à une évaluation perceptive d'un expert humain considérée comme "référence" montre des résultats très encourageants et confirme la capacité de l'approche à detecter les anomalies au niveau phonème. L'approche s'est aussi révélée capable de capter l'évolution de la sévérité de la dysarthrie suggérant une potentielle application lors du suivi longitudinal des patients ou pour la prédiction automatique de la sévérité de leur dysarthrie.
Aussi, l'analyse du comportement de l'outil d'alignement automatique de la parole face à la parole dysarthrique a révélé des comportements dépendants des pathologies et des classes dysarthriques ainsi que des différences entre les catégories phonétiques. De plus, un effet important du style de parole (parole lue et spontanée) a été constaté sur les comportements de l'outil d'alignement de la parole et de l'approche de détection automatique d'anomalies.

Finalement, les résultats d'une campagne d'évaluation de l'approche de détection d'anomalies par un jury d'experts sont présentés et discutés permettant une mise en avant des points forts et des limites du système.

Mots clés : Dysarthrie, troubles de parole, Traitement Automatique de la Parole, détection d'anomalies, parole lue, parole spontanée, alignement automatique de la parole

Abstract

Dysarthria is a speech disorder resulting from neurological impairments of the speech motor control. It can be caused by different pathologies (Parkinson's disease, Amyotrophic Lateral Sclerosis - ALS, etc.) and affects different levels of speech production (respiratory, laryngeal and supra-laryngeal). The majority of research work dedicated to the study of dysarthric speech relies on perceptual analyses. The most known study, by F. L. Darley in 1969, led to the organization and the classification of dysarthria within 6 classes (completed with 2 additional classes in 2005).

Nowadays, perceptual evaluation is still the most used method in clinical practice for the diagnosis and the therapeutic monitoring of patients. However, this method is known to be subjective, non reproductive and time-consuming. These limitations make it inadequate for the evaluation of large corpora (in case of phonetic studies) or for the follow-up of the progression of the condition of dysarthric patients. In order to overcome these limitations, professionals have been expressing their need of objective methods for the evaluation of disordered speech and automatic speech processing has been early seen as a potential solution.

The work presented in this document falls within this framework and studies the contributions that these tools can have in the evaluation of dysarthric, and more generally pathological speech.

In this work, an automatic approach for the detection of abnormal phones in dysarthric speech is proposed and its behavior is analyzed on different speech corpora containing different pathologies, dysarthric classes, dysarthria severity levels and speech styles (read and spontaneous speech). Unlike the majority of the automatic methods proposed in the literature that provide a global evaluation of the speech on general items such as dysarthria severity, intelligibility, etc., our proposed method focuses on the phone level aiming to achieve a better characterization of dysarthria effects and to provide a precise and useful feedback to the potential users (clinicians, phoneticians, patients). This method consists on two essential phases : (1) an automatic phone alignment of the speech (2) an automatic classification of the resulting phones in two classes : normal and abnormal phones.

When compared to an annotation of phone anomalies provided by a human expert considered to be the "gold standard", the approach showed encouraging results and proved to be able to detect anomalies on the phone level. The approach was also able to

capture the evolution of the severity of the dysarthria suggesting a potential relevance and use in the longitudinal follow-up of dysarthric patients or for the automatic prediction of their intelligibility or the severity of their dysarthria.
Also, the automatic phone alignment precision was found to be dependent on the severity, the pathology, the class of the dysarthria and the phonetic category of each phone. Furthermore, the speech style was found to have an interesting effect on the behaviors of both automatic phone alignment and anomaly detection.

Finally, the results of an evaluation campaign conducted by a jury of experts on the annotations provided by the proposed approach are presented and discussed in order to draw a panel of the strengths and limitations of the system.

Key words : Dysarthria, speech disorders, automatic speech processing, anomaly detection, read speech, spontaneous speech, automatic phone alignment

Table des matières

Chapitre 1

Introduction

Dans notre société, la communication est une faculté fondamentale de l'être humain et représente une clé pour la réussite sociale et professionnelle. Elle peut être réalisée sous la forme de différentes modalités : orale, écrite, gestuelle, etc. La communication parlée reste malgré tout le centre de la communication humaine en permettant une communication simple et efficace pouvant contenir différents messages (informer, demander, exprimer un avis/sentiment, etc.). Elle permet aussi grâce au riche vocabulaire et style de parole de véhiculer et traduire plusieurs nuances parfois difficiles à exprimer autrement. Outre un moyen de communication, ses caractéristiques sont le reflet et la représentation de notre identité et personnalité humaine. Malgré la révolution numérique des dernières décennies qui a permis l'introduction et la démocratisation de nouveaux outils de communication (messagerie électronique, messagerie instantanée, réseaux sociaux, etc.), la parole n'a pas perdu son statut de moyen principal et incontournable de communication. Paradoxalement, on peut même avancer que cette évolution a confirmé une nouvelle fois la parole comme centre de la communication humaine grâce à sa facilitation d'échanges instantanés indépendamment de la localisation géographique des interlocuteurs. De plus, l'introduction plus récente d'assistants automatiques axés sur la communication parlée (Siri (Apple), Cortona (Windows), google home (Google), Alexa (Amazon), etc.) a prouvé encore une fois l'importance de la parole dans la communication humaine.

C'est la raison pour laquelle tout trouble de la parole peut avoir des conséquences importantes sur la vie quotidienne des personnes concernées allant jusqu'à engendrer des comportements de repli sur soi et d'isolement de la société. Ces comportements peuvent être liés à des causes physiques où la personne atteinte voit ses capacités de parole diminuer limitant ses aptitudes professionnelles. Aussi, les troubles de la parole, entraînant des productions marquées et symptomatiques, peuvent décourager les personnes concernées à parler en public affectant aussi bien leur vie professionnelle que sociale.

Dans ce travail, nous nous intéressons à un trouble particulier de la parole : la dysarthrie. La définition de la dysarthrie a bien évolué au cours du temps. Après avoir été décrite comme des troubles de parole purement articulatoires, Peacher a introduit,

dans (Peacher, 1950), l'implication d'autres facteurs notamment psychologique et neurophysiologique. La dysarthrie est maintenant définie comme un trouble de la réalisation motrice de la parole dû à des lésions du système nerveux central ou périphérique. Elle peut être le résultat de différentes lésions neurologiques causées par différents accidents ou pathologies tels qu'un Accident Vasculaire Cérébral (AVC), un traumatisme crânien, la Sclérose Latérale Amyotrophique (SLA), la maladie de Parkinson, etc. Les travaux de référence sur la parole dysarthrique restent ceux de Darley (Darley et al., 1969b,a, 1975) qui ont permis la classification des dysarthries en 6 classes distinctes (classification complétée par deux nouvelles classes dans (Duffy, 2005)) sur la base des anomalies perçues par un jury d'experts dans un corpus de 212 patients dysarthriques. Cependant, des études plus récentes ont mis en lumière certaines limites de cette classification et ont montré la nécessité d'une meilleure caractérisation de la dysarthrie et des altérations qui y sont liées.

Le moyen le plus utilisé dans la pratique clinique pour l'évaluation de la dysarthrie est l'analyse perceptive de la parole. Cette analyse repose sur le principe de l'association entre la parole et sa perception et s'appuie essentiellement sur l'appareil auditif humain. Lors de ce type d'analyse, un expert humain décidera, suite à l'écoute de la parole d'un patient, de la présence ou non d'une dysarthrie et jugera de son état d'avancement et de sa sévérité. Cette évaluation peut prendre plusieurs formes et reposer sur différents critères perceptifs tels que l'intelligibilité, la compréhensibilité, le débit de parole, l'articulation, etc. De plus, ces évaluation sont très importantes puisqu'elles permettent aux cliniciens de définir la prise en charge thérapeutique du patient.

Ce type d'évaluation présente tout de même des limites bien documentées et reconnues même par les professionnels (Zyski et Weisiger, 1987; Özsancak et Devos, 2007). La critique la plus souvent adressée à l'évaluation perceptive de la parole est son caractère subjectif. En effet, elle est très dépendante de l'auditeur qui la réalise, de son expérience, âge, langue maternelle et même de certains facteurs socio-culturels. Elle peut aussi dépendre des conditions dans lesquelles elle est réalisée ce qui la rend non reproductible et donc inadaptée aux études longitudinales de l'évolution de la dysarthrie chez les patients. Aussi, et face à la large variabilité des classes dysarthriques et des pathologies qui y sont liées ainsi qu'aux différentes stratégies de compensation que peut développer chaque patient, une "juste" évaluation de la parole devient compliquée même pour des professionnels expérimentés. Finalement, et vu son important coût en temps et en ressources, l'évaluation perceptive de la parole dysarthrique se montre inadaptée dans le cadre de grands corpus de données nécessaires aux travaux de recherches. Pour pallier ces limites, les cliniciens ont exprimé leurs besoins de moyens et d'outils permettant une évaluation plus objective de la parole dysarthrique. De tels outils devraient permettre à la fois d'assister et d'objectiver l'évaluation de la parole et d'assurer, grâce à leur reproductibilité, un meilleur suivi thérapeutique des patients.

Les méthodes issues du domaine du Traitement Automatique de la Parole (TAP) ont été considérées comme des candidats potentiels pour l'élaboration de moyens d'évaluation objectifs de la parole pathologique, plus spécifiquement de la parole dysarthrique. Les premières études visant à automatiser les grilles d'évaluation perceptives de la parole ont vu le jour dès les années 90 (Ferrier et al., 1992). Depuis, et suite à l'avancée

et l'amélioration des performances des outils de reconnaissance automatique de la parole, plusieurs méthodes reposant sur ces outils et visant soit à automatiser des tests perceptifs, soit créer de nouvelles mesures permettant une évaluation plus objective de la dysarthrie (Fredouille et Pouchoulin, 2011) ont vu le jour. Les performances de ces méthodes sont néanmoins en deçà des espérances, ce qui explique en partie leur absence dans la pratique clinique.

Cette thèse s'inscrit dans le même cadre général que ces travaux visant à adapter les outils de traitement automatique de la parole pour l'évaluation objective de la parole dysarthrique. Néanmoins, et contrairement à la majorité des travaux existants qui seront détaillés dans le chapitre suivant et qui visent à évaluer la parole sur des critères globaux tels que l'intelligibilité ou la sévérité, l'approche d'évaluation automatique de la parole dysarthrique proposée ici se concentre sur le niveau phonème. Nous estimons que ce niveau de granularité, bien que plus difficile à évaluer, peut permettre un feedback plus précis aux cliniciens et aux patients et les assister dans leur suivi thérapeutique et travail de rééducation. De plus, nous estimons que ce type d'approches basées sur des observations de durées courtes (phonème, syllabe) peut être extrapolé dans un deuxième temps pour l'évaluation de l'intelligibilité et de la sévérité de la dysarthrie. Dans le cadre de ce travail, nous essaierons de répondre à plusieurs questions :

— **Comportement des outils de TAP :** *Comment se comportent les outils de traitement automatique de la parole face à la variabilité de la parole dysarthrique ?* En effet, la dysarthrie peut être causée par différentes pathologies résultant en plusieurs types d'altérations de la parole. De plus, les troubles qui y sont liés évoluent avec le temps et peuvent dépendre du style de parole (parole lue et parole spontanée). Cette large variabilité peut alors induire des comportements différents des outils automatiques dépendants de la pathologie, la classe et la sévérité de la dysarthrie.

— **Capacité de l'approche :** *Un système automatique est-il capable de détecter des altérations du signal de parole au niveau phonème dans la parole dysarthrique ?* La caractérisation des phonèmes anormaux est une tâche difficile due à la large variabilité observée dans la parole dysarthrique. La détection automatique de ces phonèmes, difficile perceptivement, est un vrai défi auquel nous faisons face. Répondre à cette question permettra d'évaluer la capacité de l'approche proposée à caractériser ces phonèmes.

— **Machine et perception humaine :** *Les anomalies détectées par un tel système sont-elles les mêmes que celles détectées perceptivement par un expert humain ?* La seule capacité du système à détecter les anomalies n'est pas suffisante pour évaluer sa performance. Une comparaison entre le comportement de "la machine" et "l'oreille humaine" permettra d'étudier s'ils détectent les mêmes anomalies ou non.

— **Pertinence :** *Les anomalies détectées automatiquement par l'approche proposée sont-elles pertinentes pour les cliniciens et les phonéticiens ?* En plus d'assister les cliniciens dans leur pratique, les anomalies détectées automatiquement par l'approche peuvent être utilisées par les phonéticiens pour une meilleure caractérisation de la parole dysarthrique.

Ce document a été organisé comme suit :

— la première partie du document permettra de mieux contextualiser ce travail et ses motivations. Le chapitre 2 consiste en un état de l'art multidisciplinaire. Dans un premier temps, nous y présenterons le processus de production de la parole et les différents organes et mécanismes qui y interviennent. Ensuite, nous définirons la dysarthrie, les troubles qui y sont liés et les principales grilles d'évaluation perceptive utilisées dans la pratique clinique. Les différentes approches automatiques d'évaluation de la parole dysarthrique visant à contourner les limites d'évaluation perceptive seront aussi présentées. Le chapitre 3 permettra de présenter le contexte général de ce travail, les projets qui y sont liés ainsi que les différents corpus de parole utilisés dans cette étude ;

— la deuxième partie de ce document a vocation à présenter les outils de traitement automatique de parole utilisés et proposés pour l'évaluation de la parole dysarthrique. Le chapitre 4 présentera un outil d'alignement automatique de la parole en phonèmes et son comportement face à la parole dysarthrique. Ensuite, le chapitre 5 détaillera l'approche proposée pour la détection des phonèmes anormaux dans la parole dysarthrique et son comportement face à la variabilité dysarthrique (pathologie, classe, sévérité) observée dans les corpus de données dont nous disposons. Le chapitre 6 détaillera les résultats d'une campagne d'évaluation de l'approche proposée par un jury d'experts.
Finalement, ce travail se clôturera par des conclusions et des réflexions sur l'approche proposée ainsi que la tâche d'évaluation automatique de la parole dysarthrique. Ces observations permettront la proposition de perspectives sur l'utilisation des outils de traitement automatique de la parole pour l'évaluation de la parole dysarthrique (et dans un cadre plus général pathologique).

Première partie

État de l'art et contexte général

Chapitre 2

Parole pathologique et traitement automatique de la parole

Sommaire

Ce chapitre est consacré à l'introduction du cadre de travail de notre sujet de recherche. Il permettra dans un premier temps de présenter l'appareil vocal humain et son fonctionnement, de l'élaboration à la réalisation de la parole. Nous présenterons ensuite la parole dysarthrique à laquelle est consacré ce travail. Nous définirons la dysarthrie, ses différents types et classes ainsi que les différentes grilles et échelles utilisées pour son évaluation.
Dans un deuxième temps, nous présenterons les différentes études portant sur l'utilisation des outils de traitement automatique de la parole dans le cadre de la parole pathologique, et plus généralement de la parole atypique. Nous nous concentrerons sur les études portant sur l'évaluation de la qualité de la parole ainsi que celles orientées vers l'assistance et l'aide aux patients dans leur vie quotidienne.

2.1 La production de la parole

La production de la parole est un acte dont la complexité et la multitude des organes qui y prennent part sont souvent masquées par son caractère naturel et facile. Outre l'acte moteur, la parole engage des opérations linguistiques complexes, telles que la sémantique, la syntaxe, la phonologie, etc. Dans les travaux de (Ferrand, 2001), les modèles psycholinguistiques définissent 3 étapes dans la production de la parole : (1) la conceptualisation du message (définition des idées à exprimer) (2) la lexicalisation du message (choix des mots à utiliser) (3) l'articulation du message. Nous nous focaliserons ici sur la troisième étape de ce modèle : l'articulation du message. En effet, la production de la parole n'est pas un acte involontaire, mais plutôt un acte moteur planifié et particulièrement complexe qui nécessite une importante coordination ainsi qu'une succession de mouvements faisant intervenir aussi bien le système anatomique périphérique que le système neurologique central. *La parole est une fonction complexe, nécessitant de l'attention, une certaine motivation, et une coordination motrice importante* (Pinto, 2007).

Chaque acte moteur volontaire s'articule autours de 3 étapes :

— la planification et l'élaboration du mouvement : il s'agit de la définition d'une stratégie pour sélectionner les mouvements adaptés parmi notre répertoire de possibilités ;
— la préparation du mouvement : correspond à l'établissement de la séquence de contractions musculaires nécessaires pour les mouvements définis ;
— l'exécution du mouvement durant laquelle le cortex moteur primaire, le tronc cérébral et la moelle épinière génèrent et conduisent l'information correspondant à la séquence définie dans l'étape précédente aux organes concernés.

Avant même sa réalisation, la production de la parole est programmée dans notre système nerveux et sa production fait intervenir à la fois le système nerveux central et le système nerveux périphérique.

Le système nerveux central est composé de l'encéphale (cerveau, tronc cérébral et cervelet) et de la moelle épinière. Le système nerveux périphérique comporte les nerfs permettant le transfert de l'information du système nerveux central aux différents organes. Il renferme 2 types de nerfs : les nerfs crâniens (liés à l'encéphale) et les nerfs spinaux ou rachidiens (liés à la moelle épinière).

2.1.1 La parole : acte moteur volontaire

Comme mentionné précédemment, la parole est un acte moteur volontaire. Sa production suit alors les 3 étapes caractérisant tout acte de ce type (Pinto, 2007).

Élaboration de la parole

Les processus impliquant la génération de la parole et du langage sont souvent liés. En effet, la production de la parole évoque généralement une progression du processus du langage vers celui de la parole. Du coup, l'activation des aires du langage et

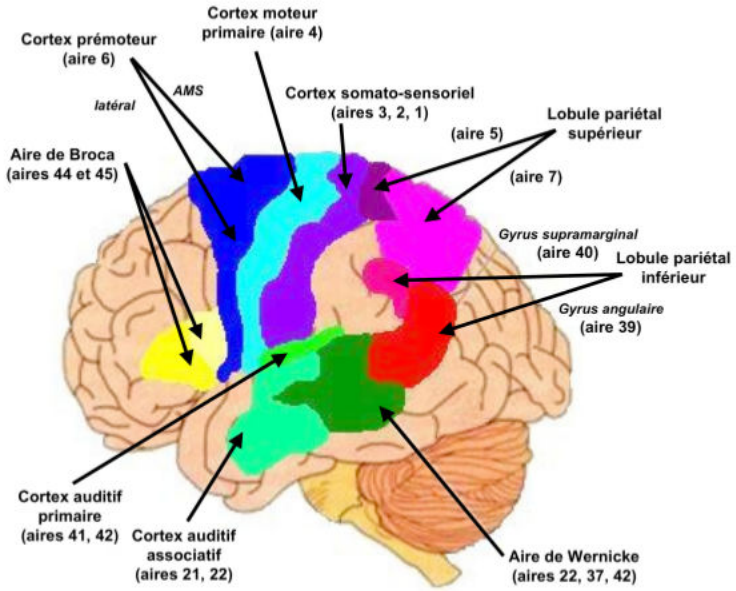


FIGURE 2.1 – *Représentation synthétique des différentes aires corticales cérébrales intervenant dans l'élaboration, la préparation et l'exécution de la parole. (Pinto, 2007)*

leurs interactions est antérieure à la sollicitation du processus de production de la parole. Plusieurs aires corticales cérébrales interviennent dans cette phase et différentes hypothèses ont été émises sur leurs rôles dans l'élaboration de la parole. Nous nous appuierons sur la délimitation des aires du cortex cérébral dite de Brodmann [1] présentée dans la figure 2.1 pour les identifier.

L'aire de Broca (chevauchant les aires 44 et 45 de Brodmann) a été longtemps associée au contrôle de la parole et considérée comme "l'aire motrice du langage". Cependant, de nouvelles études basées sur l'imagerie médicale lui associent plutôt une fonction de planification de la parole (préparation du mouvement des organes de la parole, conceptualisation anticipatoire d'une action).

Les aires auditives primaires (aire 41), associatives (aires 21 et 22) et d'intégration (aire 42) interviennent aussi dans cette phase. Des études ont montré que l'aire 41 s'active lors de la production de la parole suite au feed-back auditif lié à la production. De plus, l'implication des régions 21 et 22 dans le traitement phonémique et sémantique du son reçu a été mise en évidence. Aussi, l'aire de Wernicke localisée entre le cortex auditif primaire et le lobule pariétal inférieur le long de l'aire 22 [2] joue un rôle important à la fois dans la perception et la production de la parole. En effet, des travaux suggèrent que la même "représentation auditive" des mots serait utilisée dans leur perception et leur production. Une participation de l'insula dans la planification motrice

1. Des zones du cortex cérébral définies par le neurologue Korbinian Brodmann. Dans cette délimitation, chaque région du cortex ayant une même organisation cellulaire est associée à un nombre allant de 1 à 52.
2. Certains auteurs y incluent aussi des parties des aires 37 et 42.

des mouvements articulatoires a aussi été mise en évidence.

Finalement, la participation des lobules pariétaux supérieurs (aires 5 et 7) et inférieurs (aires 39 et 40) dans l'élaboration de la parole est moins documentée mais certaines études suggèrent leurs rôles dans l'intégration d'informations sensori-motrices, visuo-spatiales, ou encore relevant du traitement phonologique et articulatoire des mots.

Génération de la parole

Suite à l'élaboration de la parole, la génération de la séquence des mouvements à réaliser est nécessaire pour la production de la parole. Lors de cette étape, l'aire 6 de Brodmann (cortex pré-moteur latéral et l'Aire Motrice Supplémentaire AMS) ainsi que le gyrus cingulaire sont impliqués. L'AMS est fortement impliquée dans la planification et la conceptualisation des mouvements. Ces aires interagissent également avec le cervelet et les noyaux gris centraux lors de cette phase. Le cervelet est composé de trois parties :

— l'archéocervelet qui participe principalement à l'équilibre et à l'orientation spatiale ;
— le paléocervelet qui intervient dans la régulation de l'activité musculaire par adaptation du tonus musculaire ;
— le néocervelet qui est impliqué dans la coordination et la précision des mouvements volontaires par action sur les muscles antagonistes.

Les noyaux gris centraux, ou ganglions de la base, sont un ensemble de 4 paires de noyaux interconnectés plus ou moins volumineux localisés au centre du cerveau. Cet ensemble est composé de : (1) le striatum composé de deux sous-structures le noyau caudé et le putamen (2) le pallidum qui se décompose en deux sous-structures, le pallidum interne et externe (3) la substance noire (aussi appelée locus niger) qui se décompose en substance noire pars compacta et pars reticula (4) le noyau sous-thalamique (aussi appelé corps de Luys).

Exécution motrice de la parole

Comme nous le verrons plus tard, la production implique différents organes nécessaires aux mécanismes de phonation et de l'articulation supra-laryngée. Suite à la réception des informations de la part de l'aire 6 et du cortex somato-sensoriel, le cortex moteur primaire transmet ces informations aux différents moto-neurones liés à la contraction des muscles impliqués dans le mouvement des organes nécessaires à la production de parole. Ces informations sont transmises par le biais des voies pyramidales et extra-pyramidales jusqu'aux nerfs crâniens concernés dans le tronc cérébral et jusqu'aux nerfs spinaux dans la moelle épinière.

La réalisation de la parole fait appel à un ensemble de structures pour réguler et contrôler les mouvements des différents organes. Ces fonctions, appelées boucles motrices de régulation, impliquent les noyaux gris centraux et le cervelet : (1) la boucle "cortico-cérébello-corticale" qui permet au cervelet de recevoir des informations sur l'intention du mouvement (afférences corticales) en provenance des aires associées à la motricité et de participer à la programmation des mouvements (2) la boucle "cortico-triato-subthalamo-pallidale-corticale" qui permet aux noyaux gris centraux de recevoir

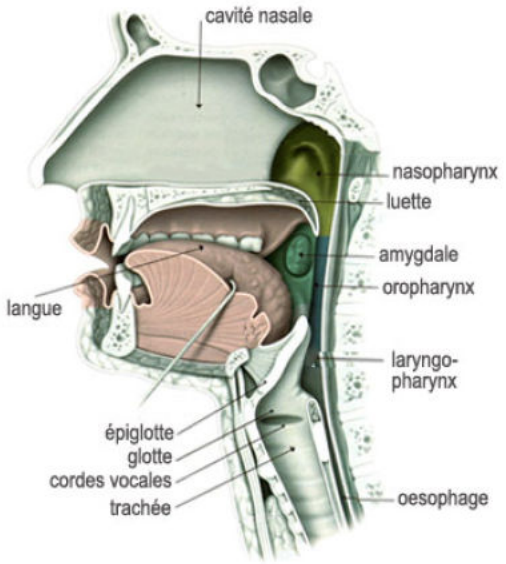


FIGURE 2.2 – *L'appareil phonatoire (source : http ://lecerveau.mcgill.ca/).*

les afférences corticales provenant du cortex cérébral. Ces informations permettent de définir les mouvements et d'ajuster les paramètres de force, de direction et d'amplitude qui leurs sont associés.

Les différents organes impliqués dans l'exécution motrice de la parole seront détaillés dans la section suivante.

2.1.2 Les organes de production de la parole

La production de la parole fait intervenir, outre les composants du système nerveux central et périphérique, divers organes (le larynx, le pharynx, les poumons, la langue, les lèvres, les cavités buccale et nasale, etc.). Cette section décrira ces différents organes intervenant dans ce processus.

La phonation se définit par la production de sons lors de la parole ou du chant. Elle n'est pas exclusive à l'être humain et peut être observée chez tous les êtres disposant d'un appareil phonatoire (mammifère, oiseau, etc.).

L'appareil phonatoire est l'ensemble des organes permettant la production du son. Il comprend l'appareil respiratoire, le larynx et notamment les cordes vocales et les cavités supra-glottiques (la cavité pharyngale, la cavité buccale et les fosses nasales) (figure 2.2) :

— l'appareil respiratoire (poumons et voies aériennes supérieures). Il permet lors de l'expiration de faire vibrer les cordes vocales situées dans le larynx ;

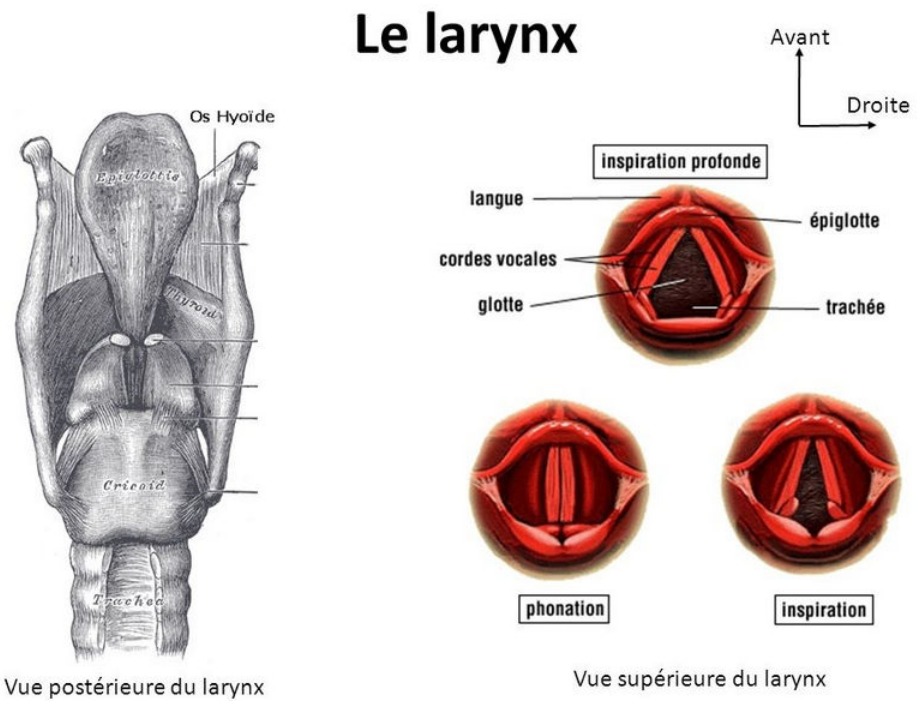


FIGURE 2.3 – *Vue postérieure du larynx et des cordes vocales dans différents états (source : http ://reannecy.org/).*

- le larynx se situe dans la partie intermédiaire du cou. Il s'agit du canal permettant de conduire l'air entre le pharynx et les poumons ;
- les cordes vocales sont des plis souples au niveau du larynx qui peuvent prendre différents états. Elles peuvent être ouvertes, fermées ou rapprochées ;
- le pharynx est un conduit musculo-membraneux où se croisent les voies respiratoires et digestives. Il joue avec les différentes cavités avec lesquelles il communique le rôle de résonateur des sons émis par les cordes vocales ;
- les cavités supra-glottiques forment le tractus ou conduit vocal. Il s'agit des cavités pharyngale, orale et nasale.

La figure 2.3 présente une vue postérieure du larynx et des cordes vocales dans différents états.
En effet, suite à l'inspiration, l'air quitte les poumons à travers le larynx. C'est à ce niveau que l'air passe par 2 tissus appelés cordes vocales. Ces cordes divisent le passage en 2 étages : cavité supra-glottique (au dessus des cordes vocales) et cavité sous-glottique (en dessous des cordes vocales). La partie où se trouvent les cordes vocales est appelée cavité glottique.

Plusieurs théories ont été émises pour expliquer et détailler le fonctionnement des cordes vocales (Crevier-Buchman, 2007). Tous ces travaux s'accordent sur la complexité du fonctionnement des cordes vocales et différents facteurs ont été mis en évidence : (1) la structure hétérogène des cordes vocales (fibres musculaires et muqueuse) (2) la

masse, l'élasticité, la longueur, la tension, la raideur et l'inertie des cordes vocales (3) le caractère ondulatoire de la vibration des cordes vocales.

Il faut distinguer entre deux différentes fonctions que peut avoir l'appareil phonatoire : la respiration et la phonation. Lors de la respiration, les cordes vocales sont ouvertes et séparées ce qui laisse la cavité glottale libre pour le passage de l'air. Par contre, durant la phonation, la glotte se referme permettant l'augmentation de la pression et ainsi le mouvement (vibration) conjoint de ces cordes vocales pour la diminuer. Lors de la réalisation de la parole, ce cycle se répète à une fréquence qu'on appelle fréquence fondamentale ou F0.

2.1.3 Les sons du Français

Dans cette partie, nous décrirons les différents sons résultants des mécanismes articulatoires produit par l'être humain dans la parole. Ces différents sons font intervenir les différents composants du conduit vocal (pharynx, le voile du palais, la langue, la mâchoire inférieure et les lèvres). Le mouvement de ces composants peut conduire à la réalisation d'environ 150 sons différents, qui font la base de toutes les langues du monde (Teston, 2007).

Deux termes sont généralement utilisés pour la caractérisation de chaque son, **le mode** définissant le type d'obstruction du conduit phonatoire (c'est la façon avec laquelle l'obstruction est réalisée : totale avec les occlusives, très resserrée avec les fricatives, etc.) et le **lieu d'articulation** qui représente le point de rapprochement ou de contact de l'articulateur avec les parties fixes du conduit vocal. La figure 2.5 détaille la répartition des différents sons des langues du monde selon les différents modes et lieux d'articulation existants.

Par ailleurs, la modalité de la source glottique permet la production de deux types de sons : les sons sonores (voisés) caractérisés par une vibration des cordes vocales et les sons sourds (non voisés) pour lesquels les cordes vocales sont ouvertes comme lors de la respiration. Pour ces sons, il n'y a pas de vibration, et donc pas de phonation lors du passage de l'air par le larynx.

En Français, les phonèmes sont repartis en deux catégories : les voyelles et les consonnes. D'un point de vue articulatoire, la différence entre ces deux catégories est liée à la circulation de l'air dans le conduit vocal. En effet, contrairement aux voyelles pour lesquelles la circulation de l'air est relativement libre à travers le tractus vocal, la production des consonnes est accompagnée d'un rétrécissement du tractus vocal qui entraîne une interruption/perturbation dans la circulation de l'air (Meunier, 2007). Cependant, la distinction la plus marquante entre ces deux catégories est plus d'ordre linguistique qu'articulatoire ; les voyelles sont un élément indispensable pour une syllabe et y sont généralement le noyau ou le centre ce qui n'est généralement pas le cas des consonnes.

Les voyelles

Toutes les voyelles du Français sont voisées i.e. elles résultent d'une vibration des cordes vocales. Quatre dimensions sont souvent mises en valeur pour caractériser les

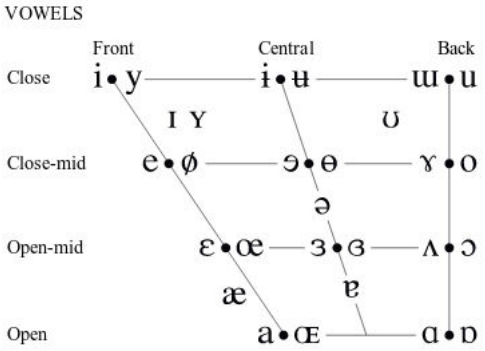


FIGURE 2.4 – *Répartition des voyelles orales des langues du monde sous la forme d'un trapèze vocalique suivant les critères articulatoires de l'aperture de la mandibule, la position de la langue, et la position des lèvres (source : http ://internationalphoneticassociation.org/).*

voyelles :

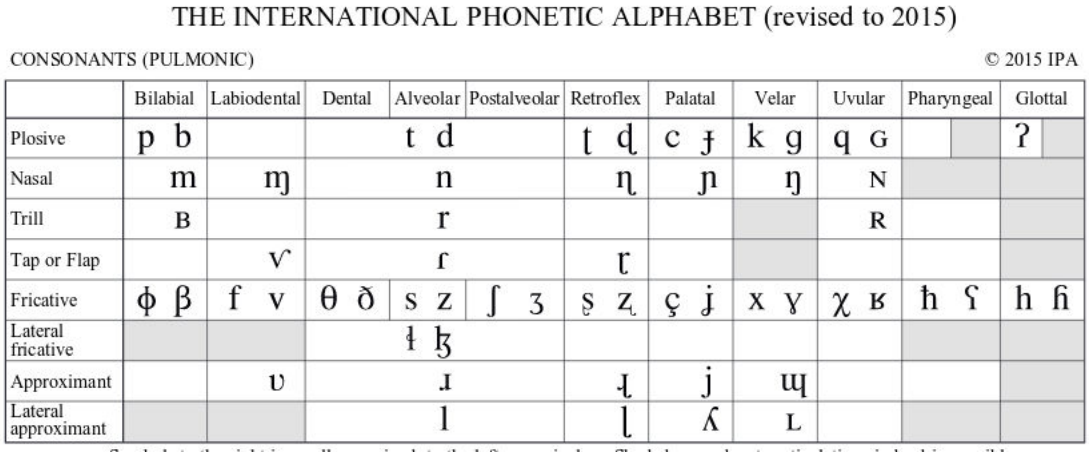
- l'aperture : elle caractérise l'ouverture de la bouche et permet d'opposer les voyelles ouvertes aux voyelles fermées. Le premier formant[3] F1 est déterminé par ce degré d'ouverture : la mesure de F1 est élevée pour les voyelles ouvertes et basse pour les voyelles fermées ;
- la position de la langue (avant ou arrière) modifie la taille de la cavité buccale. Elle permet d'opposer les voyelles antérieures et les voyelles postérieures. Le formant F2 est affecté par cette dimension et sa mesure est élevée pour les voyelles antérieures et basse pour les voyelles postérieures ;
- la position des lèvres (arrondies ou non) : il s'agit d'un allongement de la cavité buccale résultant de la projection des lèvres en avant. On parle d'arrondissement des lèvres et cette configuration affecte les formants F2 et F3. Les valeurs de F2 et F3 sont plus élevées pour les voyelles non arrondies que pour les voyelles arrondies ;
- la nasalité : le voile du palais s'abaisse lors de la production de certaines voyelles laissant les fosses nasales en communication avec le conduit vocal. On parle dans ce cas des voyelles nasales.

Généralement, les voyelles orales sont présentées sous la forme d'un trapèze, dit vocalique, prenant en compte les dimensions de l'ouverture de la bouche et la hauteur de la langue (F1) et de la position de la langue et des lèvres (F2). La figure 2.4 représente les voyelles orales des langues du monde sous ce format.

Les consonnes

Contrairement aux voyelles, l'ensemble des sons constituant les consonnes est souvent décrit comme hétérogène. De plus, en contraste avec les voyelles qui sont toutes sonores, les consonnes peuvent être sonores ou sourdes. L'ensemble des consonnes est alors souvent décomposé en des classes de sons jugées plus homogènes. Ces clas-

3. les formants sont des zones d'harmoniques renforcées de l'onde glottique lorsque celle-ci traverse le conduit vocal, considéré comme une caisse de résonance. Le premier harmonique présentant la plus forte amplification est le premier formant (F1), le second harmonique est le deuxième formant (F2) et le troisième harmonique est le troisième formant (F3) etc..



THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

CONSONANTS (PULMONIC) © 2015 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

FIGURE 2.5 – *Répartition des différents sons selon leur mode et leur lieu d'articulation (source : http ://internationalphoneticassociation.org/).*

sifications peuvent différer selon les critères considérés (purement articulatoires ou acoustico-articulatoires). Néanmoins, deux macro-classes principales émergent dans toutes les classifications : les fricatives et les occlusives.

Les occlusives

La réalisation des occlusives se fait sur deux temps : une tenue pendant laquelle le flux d'air est bloqué dans le conduit vocal suivie d'une explosion et du relâchement de l'occlusion permettant la libération du flux d'air bloqué. Selon la présence ou non d'une périodicité liée aux mouvements des cordes vocales, nous pouvons distinguer deux types d'occlusives : les occlusives sourdes (/p/, /t/ et /k/) et les occlusives sonores (/b/, /d/ et /g/).

Les occlusives sonores sont généralement caractérisées par une durée de tenue plus courte que les occlusives sourdes ainsi qu'un bruit d'explosion de moindre intensité. Les occlusives peuvent aussi être classées selon leurs lieux d'articulation : (1) occlusives bilabiales (/p/ et /b/) (2) occlusives alvéolaires (/t/ et /d/) (3) occlusives vélaires (/k/ et /g/).

Les fricatives

Les fricatives se distinguent par la présence d'un bruit durant toutes leurs tenues. Tout comme les occlusives, on peut distinguer deux types de fricatives en fonction du mouvement ou non des cordes vocales : les fricatives sourdes (/f/, /s/ et /ʃ/) et les fricatives sonores (/v/, /z/ et /ʒ/).

Nous pouvons aussi différencier les fricatives en fonction de leurs lieux d'articulation : (1) les fricatives labiodentales (/f/ et /v/) (2) les fricatives alvéolaires (/s/ et /z/) (3) les fricatives palatales (/ʃ/ et /ʒ/).

Les consonnes vocaliques

Cette classe regroupe des sons assez différents dont la caractéristique commune est

de présenter une structure de formants sur le signal acoustique. On y inclut les :

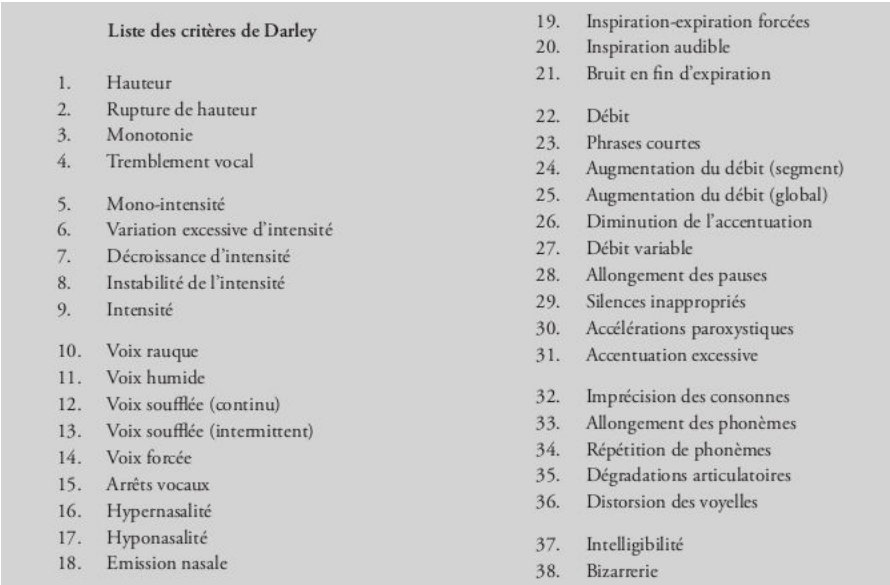
— consonnes nasales : ces consonnes sont souvent considérées comme des occlusives d'un point de vue articulatoire étant donnée l'obstruction complète du conduit vocal lors de leur réalisation. Le Français comprend deux consonnes nasales : une bilabiale (/m/) et une alvéolaire (/n/) ;
— la consonne approximante latérale /l/ dont le lieu d'articulation est situé en avant du palais. Elle présente un rétrécissement du conduit vocal moins important que celui observé sur les fricatives ;
— la consonne /r/ a un statut tout particulier dans le Français en raison des différentes formes (mode et lieu d'articulation) qu'elle peut prendre. Sa réalisation la plus répandue est uvulaire. Cependant, elle peut être vibrante, fricative ou se rapprochant d'une approximante selon le contexte et le phonème précédent ;
— Les glissantes (/j/, /w/ et /ɥ/) présentent une réalisation assez atypique et particulière entre les voyelles et les consonnes ce qui leur vaut les noms de "semi-voyelles" ou de "semi-consonnes". L'articulation de /j/, /w/ et /ɥ/ peut être comparée aux trois voyelles fermées du Français, /i/, /u/ et /y/ respectivement.

2.2 La dysarthrie

Les troubles de la communication sont définis par The American Speech and Hearing Association (ASHA) de la manière suivante : "*An impairment in the ability to receive, send, process, and comprehend concepts or verbal, nonverbal and graphic symbol systems. A communication disorder may be evident in the processes of hearing, language, and/or speech. A communication disorder may range in severity from mild to profound. It may be developmental or acquired. Individuals may demonstrate one or any combination of the three aspects of communication disorders. A communication disorder may result in a primary disability or it may be secondary to other disabilities*" (ASHA, 1993). Sur la base de cette définition, les troubles de communications englobent toute altération de la voix, du langage, de l'audition ou de la parole. La dysarthrie, étant un trouble de la parole, est alors un trouble de communication.

Initialement, les dysarthries ont été décrites comme des troubles purement articulatoires. Peacher, dans (Peacher, 1950), était l'un des premiers à évoquer la possibilité de l'implication de facteurs autre que l'articulation relevant de la neurophysiologie, la psychologie, la phonétique instrumentale et de la pathologie de la parole (Auzou et al., 2000).

En 1957, (Grewel, 1957) a proposé le terme de dysarthro-pneumo-phonie afin de rendre compte des atteintes non articulatoires de la dysarthrie. Cependant, ce terme bien qu'assez descriptif des différents niveaux d'atteintes dans la plupart des pathologies liées à la dysarthrie, ne s'est pas répandu dans la pratique courante. C'est en 1975, que Darley définit la dysarthrie comme un trouble de la réalisation motrice de la parole, secondaire à des lésions du système nerveux central et/ou périphérique (Darley et al., 1975). Actuellement, le terme "dysarthrie" englobe les troubles moteurs de la parole



Liste des critères de Darley

1. Hauteur
2. Rupture de hauteur
3. Monotonie
4. Tremblement vocal
5. Mono-intensité
6. Variation excessive d'intensité
7. Décroissance d'intensité
8. Instabilité de l'intensité
9. Intensité
10. Voix rauque
11. Voix humide
12. Voix soufflée (continu)
13. Voix soufflée (intermittent)
14. Voix forcée
15. Arrêts vocaux
16. Hypernasalité
17. Hyponasalité
18. Emission nasale
19. Inspiration-expiration forcées
20. Inspiration audible
21. Bruit en fin d'expiration
22. Débit
23. Phrases courtes
24. Augmentation du débit (segment)
25. Augmentation du débit (global)
26. Diminution de l'accentuation
27. Débit variable
28. Allongement des pauses
29. Silences inappropriés
30. Accélérations paroxystiques
31. Accentuation excessive
32. Imprécision des consonnes
33. Allongement des phonèmes
34. Répétition de phonèmes
35. Dégradations articulatoires
36. Distorsion des voyelles
37. Intelligibilité
38. Bizarrerie

FIGURE 2.6 – *Liste des critères perceptifs utilisés dans la classification de Darley (Auzou, 2007a).*

d'origine neurologique acquis et non développementaux (à l'exception de l'apraxie) (Auzou, 2007a). Cette définition se limite aux troubles d'origine neurogène et exclue les troubles mécaniques (fractures mandibulaires, fentes palatines, etc.) qui peuvent aussi affecter la parole.

Les dysarthries sont multiples et résultent en plusieurs altérations perturbations (certaines générales et d'autres propres à chacune des dysarthries). Cette multitude de troubles a été le sujet de différents travaux de recherche ce qui a conduit à plusieurs classifications des dysarthries.

2.2.1 Classifications des dysarthries

Il existent plusieurs classifications des dysarthries qui reposent sur des considérations neurologiques, physiopathologiques, cliniques ou même des combinaisons des trois. La plus utilisée est celle de Darley (Darley et al., 1969b,a) qui se base sur les caractéristiques perceptives observées dans la parole dysarthrique. Cette classification est le résultat d'une étude perceptive réalisée sur 212 patients sur une tâche de lecture de texte.

Les jurés ont côté, sur une échelle de 7 points, chacun des patients sur 38 paramètres regroupés en 7 catégories : hauteur, intensité, qualité vocale, respiration, prosodie, articulation ainsi qu'une évaluation globale de la parole (intelligibilité, bizarrerie). Les critères sont détaillés dans la figure 2.6.

L'étude des relations entre les critères les plus déviants a permis de dégager des mécanismes physio-pathologiques. Lorsque la corrélation entre deux critères est à la fois significative et physiologiquement pertinente, les auteurs regroupaient ces critères

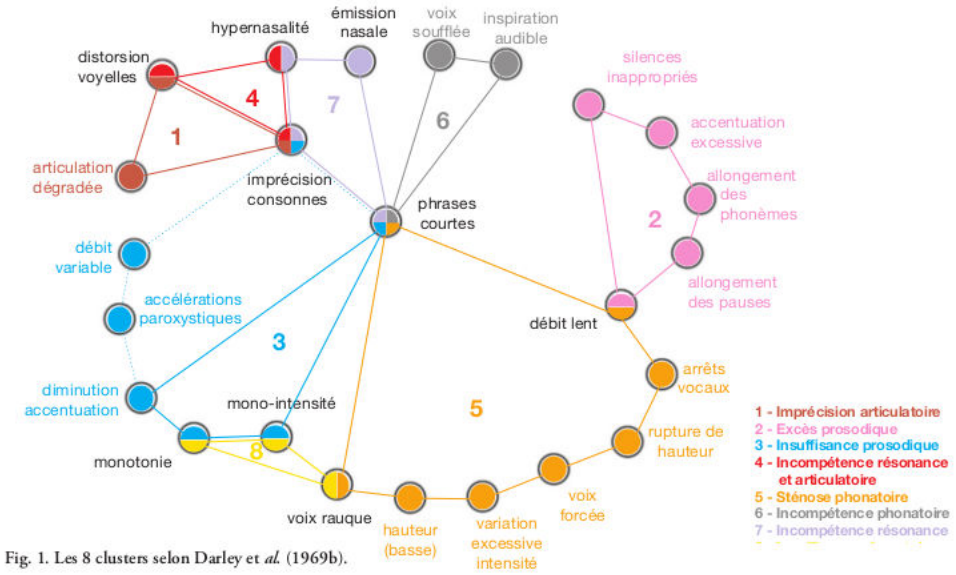


FIGURE 2.7 – *Les 8 clusters dysarthriques de Darley (Darley et al., 1969b). (Auzou, 2007a)*

dans un même ensemble, nommé "cluster". 8 clusters ont été identifiés et nommés en fonction de la physiopathologie sous-jacente supposée. La figure 2.7 illustre les 8 clusters dysarthriques de la classification de Darley. Ensuite, chaque type de dysarthrie a été décrit par ses clusters constitutifs. Cette description selon les clusters est rapportée dans la tableau 2.8.

Six classes de dysarthrie ont été définies :

— dysarthrie flasque : causée par l'atteinte des moto-neurones périphériques (situées au niveau de la moelle épinière ou du bulbe rachidien), de la jonction neuromusculaire ou des muscles impliqués dans la production de la parole. Ces atteintes peuvent aussi se situer au niveau des nerfs crâniens et des nerfs spinaux innervant les muscles intervenant dans la production de parole. Dans l'étude de (Darley et al., 1969b), cette classe était représentée par des patients souffrant d'atteinte bulbaire ;

— dysarthrie spastique : causée par des lésions bilatérales des voies reliant les structures hémisphériques aux noyaux du tronc cérébral contrôlant les effecteurs de la parole. Dans l'étude de (Darley et al., 1969b), cette classe était représentée par des patients souffrant du syndrome pseudo-bulbaire ;

— dysarthrie ataxique : causée par une atteinte du cervelet ou des voies cérébelleuses d'origine dégénérative, vasculaire, démyélinisant, traumatique, néoplasique, inflammatoire, toxique ou métabolique. Dans l'étude de (Darley et al., 1969b), la présence de symptômes de dysfonctionnement cérébelleux était le critère d'inclusion à cette classe ;

— dysarthrie hypokinétique : causée par une atteinte des noyaux gris centraux du système nerveux. La cause la plus typique est la maladie de Parkinson. Dans l'étude de (Darley et al., 1969b), cette classe était représentée par des patients

	Flasque	Spastique	Ataxique	Hypokinétique	Hyperkinétique (chorée)	Hyperkinétique (dystonie)	Mixte
Hauteur		X		X			X
Rupture de hauteur		X					
Monotonie	X	X	X	X	X	X	X
Mono Intensité	X	X	X	X	X		X
Variation excessive d'intensité					X	X	
Voix rauque	X	X	X	X	X	X	X
Voix soufflée	X Continue	X Intermittente		X Continue			X Continue
Voix forcée		X			X	X	X
Arrêts vocaux						X	
Hypernasalité	X	X			X		X
Emission nasale	X						X
Inspiration audible	X						X
Débit		X	X			X	X
Phrases courtes	X	X			X	X	X
Diminution de l'accentuation		X		X	X	X	X
Débit variable				X	X		
Allongement des pauses			X		X	X	X
Silences inappropriés				X	X	X	X
Accélérations paroxystiques				X			
Accentuation excessive		X	X		X		X
Imprécision des consonnes	X	X	X	X	X	X	X
Allongement des phonèmes			X		X	X	X
Dégradations articulatoires			X		X	X	
Distorsion des voyelles		X	X		X	X	X

FIGURE 2.8 – *Critères déviants en fonction de la classe de la dysarthrie (Auzou, 2007a)*

atteints de la maladie de Parkinson ;

— dysarthrie hyperkinétique : causée par une atteinte des noyaux gris centraux et concerne des pathologies liées à des mouvements involontaires comme la maladie de Huntington. Dans l'étude de (Darley et al., 1969b), cette classe était représentée par des patients choréiques ou dystoniques ;

— dysarthries mixtes : dans l'étude de (Darley et al., 1969b), cette classe était représentée par les patients atteints de Sclérose Latérale Amyotrophique (SLA) qui cause une dysarthrie associant une composante centrale (spastique) et une autre périphérique (flasque).

En 2005, dans (Duffy, 2005), deux autres classes ont été rajoutées à cette classification ramenant le total à 8 :

— les dysarthries liées à une atteinte unilatérale du premier neurone moteur ;

— les dysarthries d'étiologie indéterminée.

Limites de la classification de Darley

Comme nous l'avons indiqué au début, la classification de Darley reste toujours le standard utilisé en pratique clinique et dans les travaux de recherche portant sur la parole dysarthrique. Cependant, certaines limites peuvent lui être reprochées (Kent, 1996) :

— les caractéristiques cliniques des patients sont insuffisamment décrites. Cela est en partie dû à l'absence de certains moyens de diagnostique au moment de la réalisation des travaux ;

— les conditions d'écoute : l'écoute des enregistrements n'a pas été faite à l'aveugle et la cotation a été réalisée par groupe pathologique ce qui pourrait biaiser le

jugement du jury. Une écoute à l'aveugle selon un ordre aléatoire aurait dû être employée ;
— la constitution des clusters : les clusters ont été construits grâce à des corrélations des critères déviants. Cependant, plusieurs corrélations ont été éliminées et considérées comme des corrélations interclusters quand elles ont été jugées moins pertinentes : un choix assez arbitraire ;
— validité des clusters : les auteurs n'ont pas fourni de preuves pour lier les anomalies perceptives utilisées lors de la classification et les dysfonctionnements moteurs dans la production de la parole ;
— la non reproductibilité des résultats : dans les travaux de (Zyski et Weisiger, 1987) qui ont repris les mêmes enregistrements que Darley, la capacité de prédiction de la classe dysarthrique sur la base de l'analyse et la cotation perceptive s'est avérée très limitée.

2.2.2 Pathologies liées à la dysarthrie

Différentes pathologies peuvent être associées à des dysarthries. Nous nous limiterons, dans ce document, à la présentation de 4 pathologies différentes associées à 3 classes dysarthriques. Il s'agit des pathologies pour lesquelles nous disposons de patients dans nos corpus et sur lesquelles nos travaux ont été réalisés.

Les ataxies cérébelleuses

Les étiologies des ataxies cérébelleuses sont multiples et peuvent être de nature acquise ou héréditaire. Le syndrome cérébelleux est causé par une lésion du cervelet ou de ses voies efférentes (à partir du cervelet) ou afférentes (vers le cervelet). Le cervelet contrôle l'équilibre et la coordination des mouvements ce qui lui donne un rôle majeur dans le contrôle de la motricité volontaire (Özsancak et Devos, 2007). Les ataxies cérébelleuses peuvent être caractérisées par plusieurs signes cliniques aussi bien moteurs que cognitifs :

— L'ataxie est une désorganisation spatio-temporelle du mouvement. Elle peut affecter les mouvements liés à la posture générale (on parle alors d'ataxie statique) et ceux liés à la motricité des membres (on parle alors d'ataxie dynamique). L'ataxie statique concerne l'état stationnaire du patient qui devient difficile à maintenir accompagné de brusques oscillations irrégulières. La marche peut être aussi affectée et le malade a même tendance à écarter ses bras pour plus d'équilibre. On parle de démarche "pseudo-ébrieuse" ou "festonnante". L'ataxie dynamique affecte la motricité des membres, la parole, l'écriture et l'oculomotricité. Elle peut être caractérisée par :
  - l'hypermétrie : problème de réglage de l'amplitude et de la localisation du mouvement ;
  - la dyschronométrie : le mouvement du patient est ralenti et son initiation est retardée ;
  - l'asynergie : un manque d'harmonie du mouvement causé par la perte des programmes moteurs automatiques ;

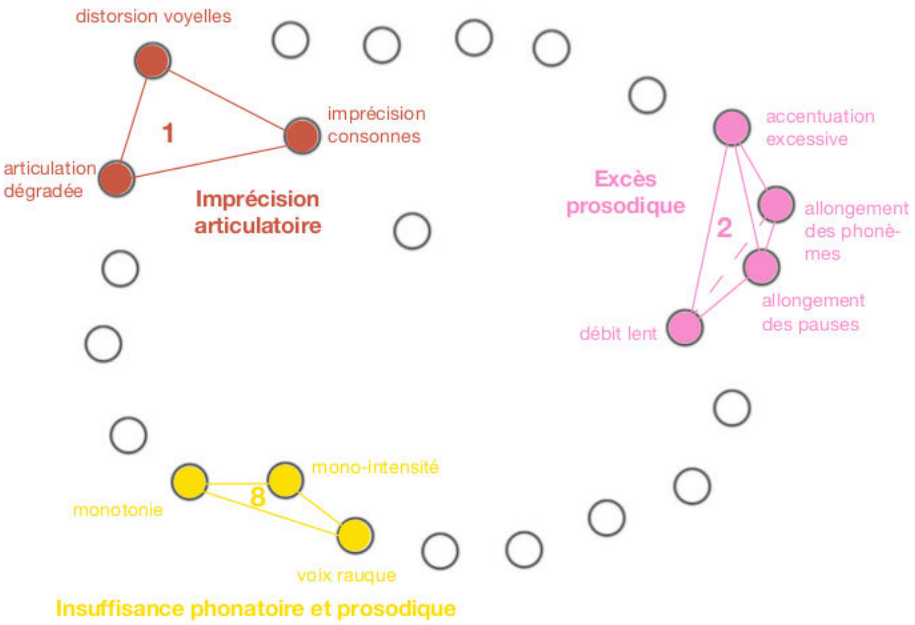


FIGURE 2.9 – *Les clusters liés à la dysarthrie ataxique (Auzou, 2007a).*

- l'adiadococinésie : difficulté à enchaîner rapidement des mouvements volontaires, successifs et alternatifs ;

— l'hypotonie : une hyperlordose lombaire (accentuation de la courbure du bas du dos, là où sont situées les 5 vertèbres lombaires) ;
— le tremblement : un tremblement de grande amplitude plus marqué au début du mouvement qu'à sa fin. Il n'est observé que dans une minorité des cas ;
— atteinte cognitive : syndrome dysexécutif, distractibilité, troubles attentionnels, troubles de la mémoire visuospatiale, dysprosodie et changements de personnalité.

Dans le cas de l'ataxie dynamique, l'écriture est affectée par l'hypermétrie, la dyschronométrie, l'adiadococinésie et l'asynergie alors que la parole est affectée par le manque de coordination entre les différents muscles effecteurs de la production. De plus, elle devient ralentie et retardée (dyschronométrie).

La dysarthrie dans les ataxies cérébelleuses

Tous les troubles décrits précédemment affectent la production de la parole. Elle comporte alors des erreurs dans l'organisation temporelle, l'amplitude, la force et la direction des mouvements. Ces troubles perturbent aussi les fonctions articulatoires, laryngées et respiratoires des patients (Schalling, 2007). Les dysarthries ataxiques présentent généralement une détérioration de l'articulation des consonnes et des voyelles, une altération du rythme de la parole ainsi qu'une dégradation de la qualité vocale.

Dans l'étude de (Darley et al., 1969b), 3 clusters de perturbation cliniques sont liés à cette dysarthrie (figure 2.9) :

— les troubles articulatoires : imprécision des consonnes, dégradation intermittente

de l'articulation et la distorsion des voyelles ;
— les excès prosodiques : accentuation excessive, allongement des phonèmes, allongement des pauses et débit lent ;
— Une insuffisance phonatoire et prosodique : voix rauque, monotonie et mono-intensité.

Le tableau 2.1 reprend les 10 critères perceptifs les plus déviants dans la dysarthrie cérébelleuse comme rapportés dans les travaux de (Schalling, 2007).

TABLE 2.1 – *Les 10 critères perceptifs les plus déviants dans la dysarthrie cérébelleuse rapportés dans les travaux de (Schalling, 2007)*

Imprécision des consonnes
Accentuation excessive
Dégradation intermittente de l'articulation
Distorsion des voyelles
Voix rauque
Allongement des phonèmes
Allongement des pauses
Monotonie
Mono-intensité
Débit lent

La maladie de Parkinson

La maladie de Parkinson est l'une des maladies neuro-dégénératives les plus fréquentes (deuxième après la maladie d'Alzheimer). Elle est liée à un dysfonctionnement du système nerveux central, caractérisé par une dénervation dopaminergique nigro-striatale progressive. Cela cause un dysfonctionnement chronique du système des noyaux gris centraux, dont le rôle est essentiel dans le contrôle de l'exécution des plans moteurs appris (Viallet et Teston, 2007). Il faut tout de même la distinguer des syndromes parkinsoniens plus sévères dans leur évolution. La cause de cette maladie est encore inconnue et elle est vraisemblablement multifactorielle incluant des facteurs environnementaux, des facteurs génétiques ainsi que le style de vie des patients (Defebvre, 2007).

La maladie de Parkinson est beaucoup plus fréquente chez les personnes âgées. En effet, l'âge moyen du début de la maladie est entre 58 et 62 ans mais elle peut aussi débuter avant 40 ans (environ 10% des cas). Bien que les études ne soient pas unanimes sur le sujet, les hommes sont plus souvent atteints que les femmes. Les signes inaugurales de la maladie de Parkinson sont :

— un tremblement de repos (observé dans environ 70% des cas) : il débute au niveau des membres supérieurs, mais peut parfois concerner les pieds, les lèvres, la mâchoire et la langue ;
— une hypertonie musculaire (rigidité dans les muscles) ;
— une akinésie des mouvements et une perturbation des mimiques.

D'autres symptômes peuvent apparaître au cours de l'évolution de la maladie (troubles

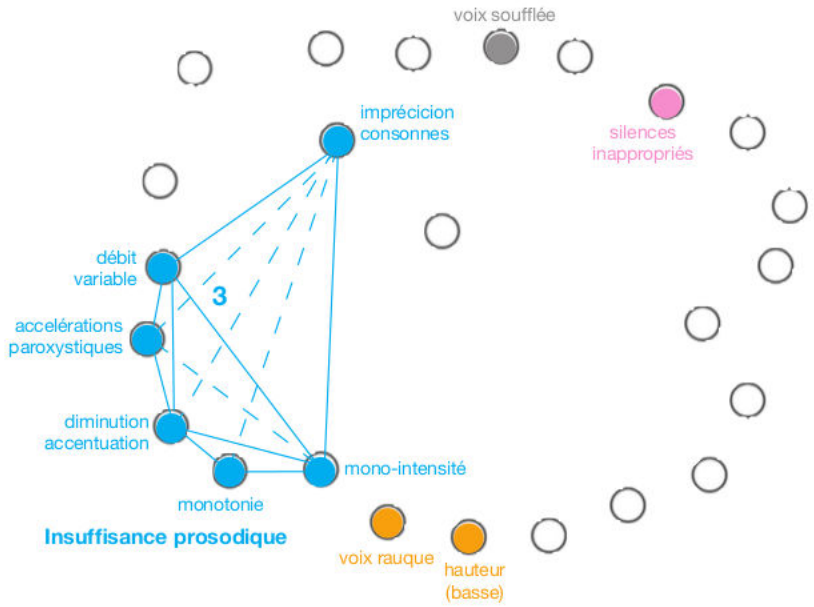


FIGURE 2.10 – *Les clusters liés à la dysarthrie parkinsonienne (Auzou, 2007a).*

digestifs, hypertension artérielle, crampes, troubles du sommeil, démence, déficits cognitifs dans le traitement des informations visuo-spatiales, dépression, etc.).

Le traitement le plus efficace sur la symptomatologie parkinsonienne est la L-Dopa. Cependant, ce traitement présente quelques complications liées essentiellement à des fluctuations d'efficacité. En effet, dans un cas sur deux, des signes et symptômes parkinsoniens réapparaissent après quelques années de traitement (Defebvre, 2004).

La dysarthrie dans la maladie de Parkinson

La production de la parole est une des activités motrices affectées par la maladie de Parkinson. Dans la pratique clinique, l'évaluation de la parole parkinsonienne se fait le plus souvent sur l'échelle UPDRS (Unified Parkinson's Disease Rating Scale) qui note l'évolution de la sévérité sur une échelle de 5 points :

— 0 = parole normale ;
— 1 = baisse légère de l'intonation et du volume ;
— 2 = parole monotone, brouillée mais compréhensible, nettement perturbée ;
— 3 = perturbation marquée de la parole, difficile à comprendre ;
— 4 = parole inintelligible.

D'autres échelles d'évaluation perceptive de la parole dysarthrique seront présentées dans la section 2.2.3.

Plusieurs travaux ont porté sur le moment de l'apparition de la dysarthrie chez les patients parkinsoniens. Dans (Müller et al., 2001), le délai moyen de la survenue de la dysarthrie a été de 84 mois après le début de la maladie de Parkinson. Cependant, dans les travaux de (Harel et al., 2004), les auteurs ont démontré que les troubles de parole

apparaissent dès le début de la maladie voire même dans la période présymptomatique. Ces résultats ont mis en avant la prise en compte insuffisante de la dysarthrie chez les patients parkinsoniens.

Dans (Logeman et al., 1978), et sur 200 patients, l'analyse perceptive de la parole a mis en avant une prédominance de la dysphonie, qui apparaît chez 89% des patients, par rapport aux troubles de l'articulation et du débit qui apparaissent chez 45% et 20% des patients successivement. Ces résultats ont été retrouvés dans les travaux de (Ho et al., 1998) où les auteurs ont confirmé le caractère précoce de la dysphonie chez les patients atteints de la maladie de Parkinson. En effet, les troubles de l'articulation et du débit prenaient progressivement une importance croissante avec l'évolution de la maladie.

Dans l'étude de (Darley et al., 1969b), un seul cluster principal de troubles de parole a été retrouvé dans la parole dysarthrique parkinsonienne (2.10).

Le tableau 2.2 détaille les 10 critères perceptifs les plus déviants dans la dysarthrie parkinsonienne.

TABLE 2.2 – *Les 10 critères perceptifs les plus déviants dans la dysarthrie parkinsonienne.*

Monotonie
Diminution de l'accentuation
Mono-intensité
Imprécision des consonnes
Pauses inappropriées
Accélérations brèves (paroxystique)
Voix rauque
Voix soufflée
Hauteur moyenne (abaissement)
Débit variable (accéléré)

La Sclérose Latérale Amyotrophique (SLA)

La SLA, connue aussi sous le nom de la maladie de Charcot, est une maladie neurologique dégénérative primitive (non provoquée par un agent extérieur). Elle affecte à la fois les neurones des voies motrices centrales (cortico-spinales et cortico-bulbaires) et ceux des voies motrices périphériques qui leur font suite. Cela cause la perte progressive de la force motrice chez le patient (Danel-Brunaud, 2007). La SLA est la maladie des motoneurones la plus fréquente et la plus grave sur le pronostic vital. En effet, il n'existe pas de traitement curatif pour la SLA et la durée médiane de survie des patients est de 36 mois seulement. 58% des patients sont atteints avant leurs 60 ans.

Les signes initiaux de la SLA sont la perte de performance sportive, la maladresse, la difficulté de port de tête, l'instabilité posturale, des troubles de la marche, la faiblesse au port de charges lourdes, des difficultés d'élocution, de phonation ou de déglutition. Cliniquement, une atrophie des muscles est constatée et le patient en perd peu à peu le contrôle volontaire. La maladie gagne progressivement l'ensemble de la musculature

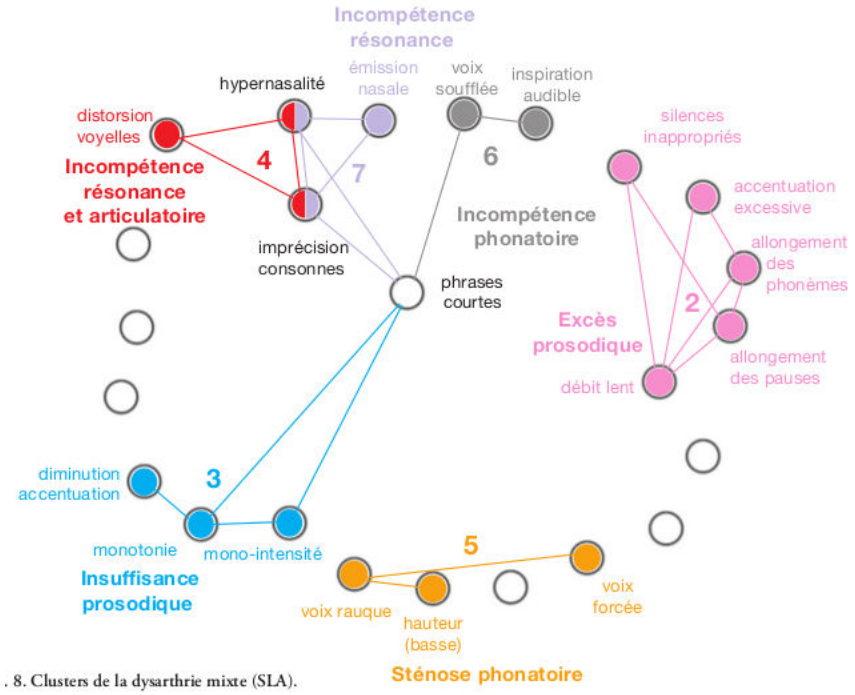


FIGURE 2.11 – *Les clusters liés à la dysarthrie mixte de la SLA (Auzou, 2007a).*

et le patient perd la capacité de parler et de déglutir les solides et liquides. Le pronostic vital des patients est engagé et le décès est souvent causé par l'atteinte progressive des muscles respiratoires.

La SLA peut prendre plusieurs formes cliniques :

— les formes spinales : elles débutent soit aux membres supérieurs (observées dans environ 37 % des cas) soit aux membres inférieurs du patient (observées dans 31% des cas) ;

— les formes bulbaires : elles débutent par des troubles de l'élocution, de la phonation et/ou de la déglutition et sont observées dans 30% des cas. C'est la forme présentant le pronostic vital le plus sévère.

La dysarthrie dans la SLA

La SLA se caractérise par une paralysie progressive de l'ensemble des organes de la réalisation de la parole (appareil respiratoire, vibrateur laryngé, organe articulatoire) (Robert, 2007). Selon la classification de Darley, la dysarthrie liée à la SLA est une dysarthrie mixte variable selon le type de l'atteinte :

— dysarthrie liée au **syndrome bulbaire**. Ce syndrome correspond à une dégénérescence progressive des motonorones périphériques du tronc cérébral. Elle entraînent des troubles de la parole, de la voix et de la déglutition. Les principaux signes sont l'atrophie musculaire, les fasciculations de la langue et parfois des joues et du menton (contractions musculaires non volontaires et localisées) et la perte de force et de vitesse musculaire ;

TABLE 2.3 – *Les critères perceptifs les plus déviants dans la dysarthrie résultant de la SLA*

Imprécision des consonnes
Hypernasalité
Voix rauque
Débit (lent)
Monotonie
Phrases courtes
Distorsion des voyelles
Hauteur (basse)
Mono-intensité
Accentuation excessive
Allongement des pauses
Diminution accentuation
Allongement des phonèmes
Voix forcée
Voix soufflée (continue)
Inspiration audible
Silences inappropriés
Émission nasale

— dysarthrie liée au **syndrome pseudo-bulbaire**. Il s'agit d'une atteinte des motoneurones centraux. Elle est caractérisée par l'absence d'amyotrophie des muscles et une dissociation automatico-volontaire plus ou moins marquée (déficit de la commande volontaire et non lors des mouvements automatiques).

Il n'est pas toujours facile d'identifier le syndrome pseudo-bulbaire du syndrome bulbaire. De plus, les deux syndromes peuvent coexister. La dysarthrie liée à la SLA est variable selon le stade d'évolution de la maladie. Elle comporte des troubles de l'articulation, des troubles prosodiques ainsi qu'une dysphonie. Un des critères distinctifs de la dysarthrie mixte est la nasalisation marquée des consonnes et voyelles orales qui apparaît assez tôt dans l'évolution de la maladie. Les troubles prosodiques correspondent à une apparition rapide d'un ralentissement du débit de parole et une diminution du nombre et de la longueur des pauses. Plus tardivement, la mélodie se dégrade avec une perte des contours intonatifs. Les formes pseudo-bulbaires s'accompagnent parfois de troubles de l'initiation de la parole. La dysphonie apparaît tôt dans la dysarthrie mixte et peut modifier le timbre vocal de différentes façons (voilé, éraillé, serré, tremblé, mouillé, ou grésillant selon les cas). De plus, la hauteur de voix devient plus aggravée (plus remarquable chez les femmes) et une perte de l'intensité de la voix est généralement observée. La figure 2.11 et le tableau 2.3 présentent les clusters des troubles de la parole et les critères perceptifs les plus déviants dans la dysarthrie mixte résultant de la SLA dans l'étude de (Darley et al., 1969b).

Les maladies lysosomales

Le terme "maladies lysosomales" regroupe plusieurs troubles qui peuvent toucher

l'enfant et l'adulte. Elles tiennent leurs nom de leurs effets au niveau des lysosomes. Les lysosomes sont des entités présentes dans chacune de nos cellules. Elles ont comme rôle de recycler les matières issues du fonctionnement cellulaire. 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, pour une raison génétique, cause une maladie lysosomale. Progressivement, et dû à l'altération du travail du lysosome, les métabolites s'accumulent dans les cellules et dans le tissu du corps et perturbent leurs fonctionnements. Cela résulte en l'apparition de lésions au niveau des organes avec des conséquences graves et irréversibles. Les maladies lysosomales ne présentent pas de signes révélateurs à la naissance, les troubles n'apparaissent qu'après une période d'évolution pouvant aller de quelques mois à plusieurs années. Une dysarthrie mixte est souvent associée aux maladies lysosomales.

Nous présenterons ici deux troubles lysosomaux correspondant aux troubles dont souffrent les patients présents dans nos corpus de données :

— la maladie de Tay-Sachs de forme tardive qui se caractérise, entre autre, par un retard psychomoteur, une macrocéphalie, une perte de la vision, une ataxie locomotrice et une détérioration intellectuelle ;
— la maladie de Niemann-Pick K (accumulation de cholestérol dans les cellules) dont les signes neurologiques typiques sont une dysarthrie, une dystonie, une paralysie des muscles oculaires, une épilepsie, et souvent une démence progressive.

2.2.3 Évaluation perceptive de la dysarthrie

Le moyen d'évaluation de la parole dysarthrique le plus utilisé dans la pratique clinique est l'évaluation perceptive (Duffy, 2005), une évaluation à l'oreille de la parole du patient.
Le principe de cette évaluation peut paraître assez simple et repose sur l'indissociabilité entre l'appareil auditif (l'oreille) et la parole elle-même. Les buts d'une évaluation perceptive de la parole pathologique, et dans notre cas dysarthrique, sont :

— d'identifier si la parole est effectivement pathologique ou non ;
— d'aider à définir les objectifs de prise en charge thérapeutique de la parole ;
— d'aider à mesurer l'évolution de la parole lors de prises en charge longitudinales des patients.

Ce bilan clinique de la dysarthrie doit alors, dans l'idéal, permettre une évaluation qualitative et même quantitative de la parole. Cette évaluation doit permettre de quantifier la sévérité de la dysarthrie, les principales anomalies la caractérisant, les organes effecteurs concernés dans ces anomalies ainsi que l'auto-perception de la dégradation de la parole par le patient lui-même.

Un des critères importants à évaluer lors de bilans cliniques est la sévérité de la dysarthrie afin de pouvoir définir les objectifs thérapeutiques et pouvoir évaluer son évolution. Cependant, cette sévérité peut être vue comme un critère d'évaluation à part ou comme une agrégation de plusieurs autres paramètres perceptifs tels que l'intelligibilité, la compréhensibilité et l'efficacité (Auzou, 2007b; Hustad, 2008; Lowit et Kent,

2010).
L'intelligibilité peut être définie par la précision avec laquelle le message émis par le locuteur est décodé par l'auditeur. Plusieurs méthodes ont été établies pour mesurer cette intelligibilité en se basant généralement sur le taux d'unités (mots, phonèmes, syllabes) correctement reconnues par l'auditeur (Barreto et Ortiz, 2008; Fontan, 2012; Hustad, 2008). La compréhensibilité est souvent décrite comme une forme particulière de l'intelligibilité prenant en compte les informations contextuelles (connaissance du patient, indices sémantiques, indices visuels, etc.) lors de l'évaluation de la parole. Un troisième critère d'évaluation est l'efficacité. Elle se mesure par la quantité de messages intelligibles transmis par le locuteur par unité de temps. Sa mesure peut refléter une altération du débit de parole ou de l'intelligibilité. Ces évaluations d'intelligibilité et de sévérité présentent plusieurs avantages surtout au niveau de leurs implémentation qui est assez simple, naturelle et réalisable selon le cadre du travail par un clinicien (suivi thérapeutique des patients) ou un jury d'écoute (travaux de recherche sur la dysarthrie).

Historiquement, l'analyse perceptive de la parole a été fortement structurée par les travaux de la Mayo Clinic aux États-Unis (Darley et al., 1969b,a, 1975) portant sur la classification des dysarthries sur la base de leurs critères perceptifs. Progressivement, plusieurs échelles et grilles d'évaluation de la parole ont été proposées. Nous présenterons quelques exemples des plus importantes dans la partie suivante.

Selon les types d'échelles constituant les grilles proposées, l'évaluation de la parole peut être quantitative ou qualitative. Les échelles qualitatives telles que les échelles bipolaires sémantiques (Revis, 2004) permettent de noter la présence ou non d'un critère donnée ; il s'agit d'une question de type oui/non. Ce type d'échelle permet d'évaluer la présence ou non de la dysarthrie ou d'une altération particulière sur un segment mais ne donne aucune indication sur la sévérité de cette altération. Par contre, les évaluations quantitatives utilisent des échelles de classes (échelle à points équidistants) ou des échelles visuelles. Elles permettent donc de noter la sévérité estimée de l'évaluation.

Exemples d'échelles d'évaluation

Échelle "Dysarthria Profile" (Robertson et Thomson 1982) : Elle permet une évaluation de la respiration, la phonation, la musculature faciale, les diadococinésies, les réflexes, l'articulation, l'intelligibilité et la prosodie.
Échelle "Frenchay Dysarthria Assessment" (Enderby, 1983) : C'est une échelle composée de 28 épreuves reparties sur 8 catégories. Les 7 premières catégories constituent une évaluation fonctionnelle des organes et de leurs fonctionnement (réflexes, respiration, lèvres, mâchoires, voile du palais et langue).
La grille d'Hartelin et Svensson (Enderby, 1983) : Cette grille comporte 54 items et permet d'évaluer la respiration, la phonation, la motricité oro-faciale, l'articulation, la prosodie et l'intelligibilité. Un test du temps maximum de phonation d'une voyelle et d'une fricative fait aussi partie du bilan.
Échelle "Unified Parkinson's Disease Rating Scale" - UPDRS : Elle est organisée en six sections comprenant chacun un certain nombre d'items. L'évaluation de la parole est l'item 18 appartenant à la troisième section (l'examen moteur). Il s'agit d'évaluer la

parole sur une échelle de 0 à 4 : 0= parole normale ; 1= baisse légère de l'intonation et du volume ; 2= parole monotone, brouillée mais compréhensible, nettement perturbée ; 3= perturbation marquée de la parole, difficile à comprendre et 4= parole inintelligible.
L'évaluation clinique de la dysarthrie (Auzou et al., 2000) : Cette évaluation est faite sur 4 étapes : (1) une conversation avec le patient donne au clinicien une première impression de l'intelligibilité globale, la prosodie, le nasonnement, etc. (2) une évaluation basée sur la grille "Frenchay Dysarthria Assessment" (3) une étude de la production des phonèmes et de mots permet de dresser l'état articulatoire du patient (4) une tâche de lecture d'un texte permet d'étudier la prosodie et les éventuelles fluctuations de la parole sur de longues durées.
La Batterie d'Évaluation Clinique de la Dysarthrie - BECD (Auzou et Rolland-Monnoury, 2006) : Il s'agit d'une version approfondie et enrichie de l'échelle précédente. Elle comprend une évaluation de la sévérité (score perceptif, intelligibilité, TPI), une analyse perceptive, une analyse phonétique, un examen moteur, une auto-évaluation et une analyse acoustique. La BECD regroupe 32 critères perceptifs : 12 sur la qualité vocale, 6 sur la réalisation phonétique, 12 sur la prosodie, un sur l'intelligibilité et un sur la naturalité de la parole. Tous les items de la BECD sont notés sur une échelle de 5 points allant de 0 (absence d'anomalie) à 4 (anomalie sévère).
La Grille d'Évaluation Perceptive de la Dysarthrie - GEPD (Lhoussaine, 2012) : Cette échelle se base sur la BECD mais considère que cette dernière contient un nombre trop important de critères à évaluer et peut donc devenir longue et non adaptée dans la pratique clinique. Elle permet d'évaluer la parole sur 9 critères perceptifs. Les patients des différents corpus utilisés dans cette étude ont été évalués suivant cette échelle. Plus de détails sur cette dernière sont présentés dans la section 3.2.2.

Limites des échelles d'évaluation perceptive

Bien qu'elle reste le "gold standard" dans la pratique clinique pour l'évaluation de la parole dysarthrique, l'évaluation perceptive de la parole présente plusieurs limites. Ces limites ont été soulevées dans les travaux de (Özsancak et Devos, 2007) et (Zyski et Weisiger, 1987) où la classification de Darley était difficilement réplicable en se reposant uniquement sur l'analyse perceptive de la parole. Ces travaux ont mis en évidence une des limites souvent attribuées à ce type d'évaluation : la non reproductibilité.

Dans (Hirano, 1989), l'auteur note les différences qui persistent dans la définition même des critères perceptifs utilisés dans la caractérisation de la voix et de la parole et réclame le besoin d'un standard et de plus de précision non seulement dans les échelles d'évaluation mais aussi dans la terminologie utilisée.

Néanmoins, la critique la plus souvent adressée à l'évaluation perceptive de la parole dysarthrique est son caractère subjectif. En effet, cette évaluation est très dépendante de l'auditeur qui la réalise, et même dans le cas d'experts de la parole dysarthrique, des différences de jugements peuvent subsister. Ces différences sont conséquentes du fait que chaque auditeur possède une représentation de la normalité qui lui est propre et qui dépend de son expérience, âge, langue et même de certains facteurs socio-culturels. C'est ce que Fex a appelé le "référent interne" de chaque auditeur (Fex,

1992). On peut même argumenter qu'il n'y a pas de standard de parole normale sur laquelle tous les auditeurs peuvent se baser lors des évaluations. C'est ce caractère subjectif qui rend cette évaluation non reproductible, parfois même par le même auditeur.

Afin de pallier ces limites, on a souvent recours à des jurys d'écoute dans le cadre des travaux de recherche afin d'obtenir une évaluation plus robuste et de diluer l'effet du référent interne de chaque auditeur. La fiabilité de l'évaluation dépendra alors de la variabilité inter-juge observée. Cependant, la réunion de ce type de jurys très coûteux matériellement et temporellement n'est pas adaptée au contexte d'évaluation clinique.

En raisons de toutes ces limites, les cliniciens ont exprimé leur besoin de méthodes d'évaluation de parole plus objectives et robustes. Plusieurs méthodes ont été proposées, certaines reposant sur l'analyse instrumentale de la parole d'un point de vue acoustique, d'autres ont étudié la possibilité de l'utilisation des outils de traitement automatique de la parole (TAP) dans le cadre de l'évaluation de la parole pathologique.

2.3 Traitement automatique de la parole pathologique

Comme pour toute parole "atypique" (enfants, apprenant d'une deuxième langue, etc.), les outils de Reconnaissance Automatique de la Parole (RAP) ont présenté des limites et des performances non consistantes lors de leurs applications à la parole pathologique. Deux visions ont émergé : (1) la première tente d'utiliser ces difficultés et les erreurs commises par les outils de RAP sur la parole pathologique pour l'évaluer et mesurer son intelligibilité (2) la deuxième voit dans ces outils un moyen pour faciliter et assister les patients dans leurs vies quotidiennes et se concentre dans l'amélioration des performances de ces outils face à la parole pathologique. On parle alors de systèmes de communication alternative et augmentée (Augmentive and Alternative Communication - AAC)

2.3.1 TAP pour l'évaluation de la parole

Deux écoles essentielles ont émergé dans le cadre de l'utilisation de la RAP pour l'évaluation de la parole (Martinez et al., 2013). Dans la première, la RAP est utilisée pour fournir une transcription automatique de la parole dont la qualité estimée au travers du taux de reconnaissance de mots peut être corrélée et interprétée comme une mesure d'intelligibilité (Doyle et al., 1997; Sharma et al., 2009; Christensen et al., 2012). La deuxième approche utilise les outils de TAP pour extraire des informations pertinentes permettant la caractérisation de la parole et pouvant être utilisées dans des systèmes de prédiction automatique d'intelligibilité (Carmichael, 2007; Middag et al., 2009; Nuffelen et al., 2009; Khan et al., 2014). Cependant, l'éventuelle utilité de tels systèmes ne fait pas l'unanimité et des réserves sur leurs intérêts ont été émises dans les travaux de (Griffin et al., 2000). Ces réserves portaient essentiellement sur le type de parole à utiliser lors de l'apprentissage des systèmes (parole normale ou parole dysarthrique) et son effet sur le feed-back fournit par ces système à l'utilisateur.

Les premières utilisations des outils de TAP pour l'évaluation de la parole dysarthrique remontent aux années 90 (Shriberg et al., 1990; Parsons, 1997). Dans (Ferrier et al., 1992), le système de RAP DragonDictate est utilisé pour transcrire des mots lus par des locuteurs dysarthriques et des locuteurs contrôles. Le feed-back fourni par le système a permis une amélioration dans l'articulation des mots par les participants.

Dans (Carmichael, 2007) une version informatisée du test Frenchay Dysarthria Assessment a été proposée. Le système évalue l'intelligibilité de la parole au niveau mot et phrase en utilisant les mêmes données que le test perceptif original. Contrairement à la majorité des travaux qui reposaient sur le taux des mots correctement reconnus pour l'évaluation de l'intelligibilité, cette approche utilise un alignement automatique contraint par le texte des mots prononcés par le patient avec des modèles appris sur de la parole normale. L'intelligibilité des locuteurs est alors calculée en utilisant les scores de vraisemblance mesurés au niveau phonème.

Le même schéma a été utilisé dans les travaux de (Fredouille et Pouchoulin, 2011) sur de la parole lue. Par contre, le but ici n'était pas de mesurer l'intelligibilité du locuteur mais plutôt de détecter les déviances au niveau phonème dans la parole produite. Les scores de vraisemblance au niveau phonème sont utilisés pour décider de la présence ou non d'une déviance ainsi que pour fournir un degré de normalité du phonème sur une échelle allant de -100 à 100.

Dans (Hahm et al., 2015), les auteurs ont combiné 6373 paramètres acoustiques extraits en utilisant l'outil openSMILE (Eyben et al., 2010, 2013) avec des paramètres articulatoires pour la prédiction de l'état neurologique des patients atteints de la maladie de Parkinson (score UPDRS). Deux méthodes ont été utilisées pour la prédiction de ces scores, une régression "Support Vector Regression - SVR" et un réseau de neurones profonds (DNN). Dans le cadre de la même tâche de prédiction, les travaux de (An et al., 2015) ont étudié l'apport des paramètres reflétant le débit de parole (durée des syllabes, durée des silences, nombre de syllabes par seconde, etc.), les mesures de formants F1 et F2 et les paramètres phonotactiques (durée des phonèmes et distribution des monophones, biphones et triphones) dans la prédiction des scores UPDRS des patients. Les gains observés étaient légers pour les différents paramètres étudiés à l'exception des paramètres phonotactiques qui n'ont introduit aucune amélioration des résultats. Finalement, les travaux de (Orozco-Arroyave et al., 2016) se sont basés sur les mesures d'énergie réalisées sur des phases de transitions entre phonèmes non voisés-voisés et voisés-non voisés afin d'étudier les phases d'initiation et d'arrêt du mouvements des cordes vocales. Ce travail a démontré l'intérêt de l'utilisation de ce type de paramètres dans la prédiction des scores UPDRS pour les patients parkinsoniens sévèrement dysarthriques.

Malgré tout l'intérêt qu'elles ont suscité, les approches automatiques pour l'évaluation de l'intelligibilité ne sont pas traduites dans des outils utilisés dans la pratique clinique. Une des contraintes limitant leur généralisation est le manque de précision dans les évaluations produites automatiquement par de tels outils (Hamidi et Baljko, 2013). Afin d'être utilisables par des patients, ces outils doivent être précis et approcher un taux de faux positifs (réalisation normale jugée comme déviante) quasi nul. En ef-

fet, même si des études ont montré que ces approches automatiques encouragent les patients à pratiquer et à fournir plus d'efforts dans les cas de rééducation et dans les séances de thérapie (Parsons, 1997; Shriberg et al., 1990), des fausses décisions prises par ces outils où, par exemple, de bonnes productions par les patients sont jugées négativement par le système peuvent avoir un effet désastreux sur les patients et leur volonté de suivre et continuer le traitement thérapeutique. Une des solutions proposées afin de limiter ce type d'erreurs est l'utilisation de modèles acoustiques dépendants du locuteur (patient dans notre cas) pouvant s'adapter un peu plus à sa prononciation. Cependant, de tels modèles nécessitent des données généralement non disponibles du patient lui même (avant sa maladie) et peuvent résulter en un effet inverse ou des productions très à l'écart de la normalité sont acceptées et reconnues comme normales par les outils de TAP.

Il est intéressant de noter que la majorité des travaux portant sur l'évaluation automatique de la parole ont visé l'adaptation et l'utilisation des outils de TAP pour la prédiction de l'intelligibilité. Néanmoins, et malgré le fait que l'intelligibilité est l'un des items les plus utilisés dans les grilles d'évaluation perceptive de la parole et reflète son intérêt communicatif, elle ne permet pas un retour précis pour le clinicien ou le patient sur la qualité de sa production et articulation des sons courts tels que les phonèmes.

Dans (Hamidi et Baljko, 2013), les auteurs ont enquêté auprès de spécialistes de la parole et du langage sur leurs besoins au niveau du feed-back fourni par les outils de RAP utilisés pour l'évaluation de la parole pathologique. Deux types d'évaluation qui font écho aux échelles utilisées dans les grilles d'évaluation perceptive sont soulignés :

— une décision sur la normalité ou pas du segment (phonème, syllabe, mot) de parole produit par le patient (échelle bipolaire) ;
— une évaluation de la normalité ou du degré de déviance de cette production (échelle de classe).

2.3.2 TAP dans les technologies de communication alternative et augmentée

Comme mentionné précédemment, la majorité des approches automatiques à base de RAP pour l'évaluation de la parole pathologique (intelligibilité, déviance au niveau phonème) exploite et met en profit les difficultés qu'éprouvent ces outils face à la parole atypique, et dans notre contexte pathologique. Suite à la généralisation des applications à base de RAP dans la vie quotidienne, ces difficultés sont également devenues des contraintes à l'utilisation normale de ces applications par les locuteurs atteints de troubles de la parole.

De plus, et dans plusieurs cas, la dysarthrie s'accompagne de plusieurs handicaps physiques qui limitent le champs d'activité et de manœuvre du patient et ses capacités aussi bien communicatives que sociales. Ces difficultés peuvent se manifester dans des tâches simples de l'ordre du contrôle des outils et d'appareils électroniques dans une maison (TV, ordinateur, téléphone, etc.) ainsi que des équipements normaux (porte, fenêtre, etc.). En plus des troubles de parole dus à la dysarthrie résultant en la diminution

de l'intelligibilité ou même sa disparition, le risque d'isolement et de retrait de la vie sociale des patients augmente.

Afin de répondre à ces besoins, un autre champs d'application des outils de la RAP a vu le jour : les systèmes de communication alternative et augmentée. Ces outils développés pour les personnes souffrant de troubles de la communication ou de parole permettent au patients de remplacer ou de compléter la parole et l'écriture pour mieux communiquer et gérer les différents outils à leur disposition dans leurs environnements. En effet, et même pour les patients dysarthriques, la parole reste le vecteur de communication le plus naturel, le plus performant et parfois le plus facile à réaliser dans le cas de sévères handicaps liés à la maladie. C'est dans ce cadre qu'a émergé le besoin de systèmes de RAP capables de reconnaître la parole des patients les plus dysarthriques soit pour la commande des dispositifs ou pour la génération d'une parole synthétique (à partir de la transcription de la parole ou du signal lui même) plus intelligible et donc plus communicatif pour le patient (Griffin et al., 2000; Hawley et al., 2005).

Les outils de RAP font face à plusieurs difficultés dans le cas de la parole dysarthrique :

— les différentes altérations et troubles de l'articulation caractérisant la dysarthrie : imprécision des consonnes conduisant les systèmes à des erreurs de substitution (Rudzicz, 2010), rallongement des voyelles dû à un débit de parole faible poussant les systèmes à décoder des mots à deux syllabes au lieu d'une seule (Morales et Cox, 2009) ;
— la grande variabilité observée chez les patients dysarthriques par rapport à la parole normale ;
— la grande variabilité inter-patient observée dans la parole dysarthrique. En effet, la parole est dépendante à la fois de la pathologie du patient, de la sévérité de la dysarthrie ainsi que du locuteur lui même. En effet, les travaux de (Hawley et al., 2005; Young et Mihailidis, 2010) ont montré que les patients atteints de dysarthrie sévères posent plus de difficultés que ceux atteints d'une dysarthrie légère.

La solution la plus intuitive pour tous ces problèmes est l'apprentissage de nouveaux modèles acoustiques sur de la parole dysarthrique contrairement aux modèles appris sur la parole normale généralement utilisée. Cependant, de tels modèles nécessitent une grande quantité de données indisponible dans le cas de la parole dysarthrique. Dans l'idéal, et afin de refléter la grande variabilité observée dans la parole dysarthrique, de telles bases doivent renfermer des patients atteints de différentes pathologies à des niveaux de progression différents. De tels corpus ne sont généralement pas disponibles dû à la difficulté d'enregistrer un nombre conséquent de patients dysarthriques sur de longues durées.
Nous pouvons citer comme exemple le corpus Nemours (Menendez-Pidal et al., 1996) qui contient les enregistrements de 11 patients sur des tâches de lecture de paragraphes et de mots et le corpus UA-speech (Universal Access speech corpus) qui contient les enregistrements de 19 patients sur une tâche de lecture de mots. Cependant, seulement 5 des 19 locuteurs de cette base sont des femmes. La base de données TORGO (Rudzicz

et al., 2012), encore en construction, comprend pour le moment les enregistrements de 7 patients sur des tâches de lecture de non-mots, mots courts, phrases et description des images. Cette base, visant à répondre à la fois aux besoins des systèmes de RAP mais également d'évaluation automatique de la parole pathologique fournit, pour chaque patient, son évaluation suivant la Frenchay Dysarthria Assessment (2.2.3). La particularité de ce corpus est d'inclure aussi bien les données acoustiques que les données articulatoires des locuteurs. Ces données ont été relevées grâce au système 3D d'articulographie électromagnétique EMA (Electro-Magnetic Articulograph) AG500 et à un système d'enregistrements vidéo de marqueurs faciaux phosphorescents. Ces systèmes permettent l'enregistrement des mouvements articulatoires intérieurs et extérieurs du conduit vocal en synchronie avec la parole.
Ces corpus sont généralement enregistrés dans des conditions très contrôlées à la fois du point de vue de la nature des données (mots isolés, phrases, paragraphes) et des conditions d'enregistrement (chambres sourdes ou du moins silencieuses, etc.). De telles conditions sont souvent nécessaires dans le cadre d'évaluation de la parole mais sont inadaptées dans le cadre applicatif où les outils de RAP feront face à de la parole dysarthrique dans des espaces ouverts où subsistent des bruits, des problèmes liés aux canaux de transmission et où la parole peut être plus spontanée.

2.3.3 Adaptation des modèles à la parole dysarthrique

Comme indiqué précédemment, un des problèmes concernant l'utilisation des outils de TAP sur la parole dysarthrique est le manque de données nécessaires pour l'apprentissage des modèles ainsi que la large variabilité observée dans la parole dysarthrique. Deux principales stratégies ont été mises en place pour contourner le problème de manque de données d'apprentissage : (1) l'adaptation de modèles acoustiques appris sur la parole normale aux locuteurs dysarthriques avec des techniques telles que MAP (Gauvain et Lee, 1994) ou MLLR (Leggetter et Woodland, 1995) (2) l'utilisation des modèles appris seulement sur la parole dysarthrique (malgré les contraintes de quantité).

Dans (Morales et Cox, 2009), des systèmes à base de modèles appris sur la parole normale et adaptés à la parole dysarthrique ont donné des gains absolus en terme de WER (Word Error Rate) entre 15% et 40% par rapport aux modèles non adaptés. Parallèlement, les travaux de (Rudzicz, 2007) ont montré que les systèmes de RAP à base de modèles acoustiques appris sur de la parole normale et adaptés aux locuteurs dysarthriques amélioraient les taux de reconnaissance pour les patients légèrement et moyennement dysarthriques par rapport aux modèles appris exclusivement sur la parole dysarthrique. Les patients souffrant de dysarthrie sévère ont montré des résultats comparables pour les deux modèles. Des résultats plus nuancés ont été retrouvés par (Sharma et Hasegawa-Johnson, 2010) où les WER de 5 des 7 patients incluant les 2 locuteurs les moins intelligibles étaient plus faibles pour les modèles appris sur la parole normale et adaptés à la parole dysarthrique (MAP) par rapport aux modèles appris uniquement sur la parole dysarthrique. Une étude menée par (Christensen et al., 2014) a montré l'intérêt de n'utiliser qu'un sous ensemble de locuteurs dans la phase d'appren-

tissage du modèle acoustique, choisis sur la base de leurs ressemblances acoustiques à la parole du patient étudiée.

Il n'existe toujours pas de réponse à la question du meilleur paradigme d'apprentissage des modèles de parole à utiliser pour la reconnaissance de parole dysarthrique mais la multitude des travaux existant n'excluent pas que le type de modèle à adopter pourra dépendre de la pathologie, du type de dysarthrie et surtout de la sévérité et du degré d'intelligibilité du locuteur.

2.3.4 TAP pour la parole "atypique"

Bien que notre travail porte sur l'étude du comportement des outils de RAP face à la parole pathologique, un contexte plus général est l'apport de ces outils dans le cadre de la parole "atypique". Un des contextes comparables à l'évaluation automatique de la parole pathologique est celui d'apprentissage de langue assisté automatiquement (Computer Assisted Language-Learning - CALL). Le but de ces technologies est l'amélioration de la prononciation des locuteurs apprenant une deuxième langue.

Dans l'étude (Eskenazi, 2009), l'auteur spécifie que ces approches ont comme tâche généralement d'assigner un score à chaque réalisation produite par le locuteur. Dans (Bachman, 1990), les auteurs soulignent le fait que dire à l'utilisateur qu'il a tort alors qu'il avait raison est plus négatif que de lui dire le contraire. Bien que cela soit comparable aux études réalisées dans le cadre du feed-back donné au patient par les outils de TAP appliqué à l'évaluation de la parole pathologique, nous estimons que les enjeux des deux approches et le besoin de précision ne sont pas semblables. En effet, l'effort fourni par les patients dysarthriques dans le cadre de rééducation est très conséquent et la moindre démotivation ou perte de capacités communicatives peut résulter dans l'isolement et le retrait des patients de la vie sociale.

Deux approches principales ont vu le jour dans l'apprentissage assisté par un ordinateur : (1) la première porte sur les erreurs individuelles au niveau phonème, syllabe ou mot (2) la deuxième porte sur l'aspect général de la parole et sa normalité.

Dans les premières approches, l'apport de la détection à un niveau plus fin (phonème/ syllabe/ mot) est la précision du système dans son feed-back à l'utilisateur. Sans cette précision, l'utilisateur ne pourra pas cibler l'erreur détectée pour la corriger. L'importance de ce retour fourni par les approches automatiques dans l'apprentissage de L2 a été mis en évidence dans les travaux de (Precoda et al., 2000; Neri et al., 2006). À la différence des outils visant la parole pathologique où leur utilisation sera plutôt supervisée par des professionnels (orthophonistes, cliniciens, etc.), les outils de CALL ciblent plutôt les apprenants eux-même, d'où le besoin d'un feed-back pertinent pour des utilisateurs naïfs.

Les premiers travaux de (Eskenazi, 1996) et (Neumeyer et al., 1996) ont utilisé soit des méthodes d'alignement contraint par le texte de la parole soit les résultats de la RAP sur les réalisations du locuteur. Dans (Sevenster et al., 1998), les auteurs ont utilisé les scores issus d'un modèle de parole appris sur la parole native. Dans ce cadre,

la connaissance préalable de la langue native de l'utilisateur peut s'avérer utile afin d'adapter le système et viser les erreurs fréquentes que font ce groupe d'utilisateurs. Ces connaissances peuvent aussi être utilisées pour personnaliser le feed-back de l'approche sur la base des connaissances sur les sons de la langue native du locuteur.

Dans le cas du deuxième type d'approches, les systèmes tentent de répliquer l'impression générale que peut avoir un auditeur face à la parole réalisée par un locuteur apprenant d'une L2. Ces systèmes peuvent reposer sur les mêmes méthodes que celles visant une évaluation plus localisée de la nativité de production, mais elles n'offrent pas ce type de feed-back au locuteur. Les mesures locales sont regroupées pour le calcul de mesures d'ordre global comparables à l'intelligibilité ou à la sévérité dans le contexte pathologique. De plus, de telles méthodes peuvent reposer sur des indicateurs prosodiques (débit de parole, durée et fréquence des pauses, etc.) pour l'évaluation de la qualité globale de la production.

Dans un autre contexte, les outils de TAP ont aussi été utilisés pour l'identification et la caractérisation des accents régionaux ou étrangers. Dans (Boula de Mareüil et al., 2008), des mesures de voisement des consonnes, formants des voyelles ainsi que des indices prosodiques sont extraits sur des phonèmes issus d'un alignement automatique de la parole et utilisés pour l'identification des accents régionaux et étrangers dans le Français

Dans (Woehrling et al., 2009), des mesures similaires sont extraites et utilisées pour classer des variétés régionales du Français. Un taux d'identification de 82% est atteint pour la classification d'un corpus d'environ 170 locuteurs répartis sur 5 classes d'accents.

2.3.5 Motivations

La section précédente nous a permis de présenter plusieurs travaux portant sur l'évaluation automatique à travers les outils de TAP de la parole dysarthrique. Cependant, et malgré les nombreux travaux étudiant l'utilisabilité de telles approches, peu de systèmes ont vu le jour et leur utilisation dans la pratique clinique reste toujours rare. Cela peut s'expliquer par le manque de précision dont souffrent encore ces approches et la nécessité d'un rendement quasi idéal pour pouvoir les utiliser dans l'évaluation des patients.

Nous remarquons aussi que la majorité des travaux existant se concentre pour l'évaluation de la parole sur des critères assez globaux tels que l'intelligibilité et la sévérité de la dysarthrie. Peu de méthodes ont visé une évaluation plus précise et une granularité plus faible telle que la syllabe ou le phonème (Fredouille et Pouchoulin, 2011).

Bien qu'il s'agisse d'une tâche plus difficile, où le risque d'erreur est plus grand, nous estimons que de telles évaluations sont très pertinentes puisqu'elles peuvent fournir aux cliniciens et aux patients un feed-back plus précis sur les ajustements et le suivi thérapeutique nécessaire pour le patient. De plus, ces évaluations peuvent être exploitées dans un second temps pour la récupération d'indices plus globaux tels que la sévé-

rité de la dysarthrie et l'intelligibilité. Ces approches peuvent être formulées comme des tâches de classification en deux classes : normal et anormal (on parle alors d'anomalie).

Dans (Chandola et al., 2009), la détection d'anomalie est présentée comme un problème nécessitant de trouver des patterns de données qui ne sont pas conformes au comportement attendu. Plusieurs champs d'application de ces approches existent tels que la détection d'intrusion, la détection de fraudes, etc. Cette tâche reste tout de même complexe comme le souligne (Chandola et al., 2009) pour plusieurs raisons : (1) la frontière entre les comportements normaux et anormaux n'est souvent pas précise (2) si le comportement anormal est le résultat d'actions préméditées, des stratégies d'adaptation peuvent être employées afin de cacher ces comportements (3) les données nécessaires pour la modélisation des deux comportements ne sont pas toujours disponibles (4) les données peuvent contenir du bruit qui peut être faussement interprété comme des comportement anormaux. Nous estimons que le problème de l'évaluation de la parole dysarthrique au niveau local (phonème) peut être formulé comme un problème de détection automatique d'anomalies : ici l'anomalie sera la réalisation pathologique i.e. atypique d'un phonème.

2.4 Conclusion

Ce chapitre a permis dans un premier temps de présenter la complexité du processus de production de la parole nécessitant différents mécanismes neurologiques, phonatoires et articulatoires et l'intervention et la coordination de plusieurs organes de l'appareil phonatoire. Nous avons ensuite défini la dysarthrie, les propositions de classification ainsi que les principaux travaux portant sur son évaluation souvent perceptive. Les limites de ce type d'évaluation, représentant le standard dans la pratique clinique, ont motivé l'utilisation des outils de TAP afin de proposer des méthodes plus objectives pour l'évaluation de la parole dysarthrique. C'est dans le but de répondre à ce besoin, souvent exprimé par les cliniciens eux-même, que les travaux présentés tout au long de ce manuscrit s'inscrivent.

Le prochain chapitre permettra de présenter le contexte expérimental général de nos travaux, les projets dans lesquels nous sommes intervenus ainsi que les corpus de parole dysarthrique utilisés.

Chapitre 3

Contexte Expérimental

Sommaire

Ce chapitre est consacré au contexte général de la réalisation de cette thèse. Nous y introduisons les deux projets, *DesPhoAPady* et *TYPALOC*, dans lesquels ce travail s'inscrit. Nous décrirons ensuite les corpus de données utilisés tout au long de cette étude ainsi que le protocole expérimental établi pour étudier et évaluer le comportement et la performance des approches et outils automatiques d'évaluation de la parole dysarthrique que nous avons proposés.

3.1 Projets

3.1.1 *DesPhoAPady*

Le projet *DesPhoAPady* [1], pour Description Phonético-Acoustique de la Parole Dysarthrique, est un projet financé par l'Agence Nationale de la Recherche (ANR). Il vise à

1. Projet ANR-08-BLAN-0125, 2009-2012

explorer l'étendue de la variabilité de la parole par le biais de la description des caractéristiques phonético-acoustiques de la parole dysarthrique.
Ce projet a permis de réunir une large équipe multidisciplinaire réunissant phonéticiens, phonologues, cliniciens, ingénieurs en informatique, traitement du signal et traitement automatique de la parole et de faire une synthèse de tout le savoir faire qu'apporte chaque discipline dans l'étude de la parole pathologique, et plus particulièrement de la parole dysarthrique. Les équipes intervenant sur ce projet sont issus de 3 partenaires : le Laboratoire de Phonétique et Phonologie (LPP - Paris), le Laboratoire Parole et Langage (LPL - Aix-en-Provence) et le Laboratoire Informatique d'Avignon (LIA - Avignon).
Les objectifs de ce projet sont :

— l'identification et la quantification des caractéristiques phonético-acoustiques des dysarthries à travers une approche combinant procédures d'analyses manuelles détaillées et procédures issues du TAP ;
— l'évaluation de la validité de ces critères sur la base de leur potentiel à distinguer parole dysarthrique et parole normale, différents types de dysarthries et à caractériser la dégradation d'une dysarthrie dans son évolution longitudinale ; styles de parole (lecture de phrases, histoire en image, parole spontanée).

Ce projet a aussi permis dans sa phase préliminaire la mise en place et la structuration d'une base de données informatisée de la parole pathologique. Ce corpus, dénommé *DesPhoAPady* dans le reste de ce manuscrit, a été alimenté par des patients dysarthriques collectés par différentes équipes et issus de 2 principaux ensemble d'enregistrements. Il sera détaillé dans la section 3.2.2.

3.1.2 *TYPALOC*

Le projet *TYPALOC* [2], pour Variations normales et anormales de la parole : Typologie, Adaptation, Localisation, est aussi un projet financé par l'ANR. Ce projet constitue une continuité du projet *DesPhoAPady* et vise à mieux comprendre l'étendue des variations de la parole chez des populations saines et affectées d'une pathologie dans des conditions de parole différentes.
Ce projet a comme objectifs de :

— dresser un inventaire typologique des variations de parole normale et "anormale" ;
— tester l'adaptabilité de locuteurs sains et pathologiques face à des conditions de production différentes (des situations plus ou moins contrôlées) ;
— tester l'effet de diverses contraintes linguistiques et de communication sur les variations observées au niveau de la parole selon les différentes populations (saines et pathologiques) ;
— étudier et améliorer le comportement des systèmes de TAP face à de la parole "atypique".

Ce projet est porté par les mêmes partenaires que le projet *DesPhoAPady* (LPL, LPP et LIA).

2. Projet ANR-12-BSH2-0003, 2012-2015

Ce projet a permis de définir un corpus plus compact de locuteurs dysarthriques issus du corpus *DesPhoAPady* et choisis sur des critères de qualité de signal, d'utilisabilité et d'intérêt pour l'étude phonétique et automatique. De plus, de nouveaux locuteurs sains ont été enregistrés et intégrés à ce corpus. Le corpus sera détaillé dans la section 3.2.3.

3.2 Corpus

3.2.1 Le corpus *VML*

Ce corpus a été construit dans le cadre d'un partenariat avec l'association "Vaincre les Maladies Lysosomales". L'objectif du projet qui réunissait des compétences multi-disciplinaires était la mise en place d'une méthode d'évaluation objective de la parole dysarthrique. Le corpus, enregistré à l'hôpital la Pitié-Salpétrière à Paris, est composé de 8 patients et de 8 locuteurs contrôles. Deux enregistrements de locuteurs contrôles ont été retirés du corpus et seulement 6 contrôles ont été retenus. Les patients souffrent de maladies lysosomales (section 2.2.2) et montrent divers états de progression de leur maladie. Tous les patients ont été enregistrés par une orthophoniste. Les 8 patients sont atteints de deux maladies lysosomales différentes :

— Tay Sachs : deux femmes et un homme ;
— Niemann-Pick type C : deux femmes et trois hommes.

La dysarthrie résultante de ces pathologies est liée à diverses atteintes dans le système nerveux est peut donc être classée comme une dysarthrie mixte (section 2.2.2).

Tous les patients suivaient un traitement thérapeutique expérimental basé sur la molécule de miglustat. Ce traitement devait, normalement, stabiliser les symptômes des patients dont notamment la dysarthrie.

Nous disposons dans ce corpus pour chaque patient de son "contrôle" associé. Il s'agit d'enregistrements d'une personne, sans pathologie particulière (saine), associés aux enregistrements d'un patient. L'appariement patient/contrôle a été effectué par des orthophonistes sur des critères physiques de similarité (sexe et tranche d'âge). Cependant, 2 des sujets contrôles ont été éliminés de cette étude suite à des anomalies remarquées sur leurs enregistrements. Le corpus (contrôles inclus) est constitué d'environ 5h45 d'enregistrements au total, tout type de parole, de maladie et de genre confondus. La durée totale des enregistrements de parole lue, utilisée dans ce travail, représente 23% du corpus, soit environ 1h20 d'enregistrements.

De plus, tous les patients ont été enregistrés longitudinalement sur 2 ans avec des intervalles approximatifs de 6 mois résultant en 3 à 5 enregistrements par locuteur. Les contrôles ont été enregistrés sur un mois avec un intervalle d'une semaine entre les différentes sessions. Tous les enregistrements ont été faits dans des conditions similaires et les locuteurs ont tous lu le même texte, une fable pour enfants intitulée "le cordonnier" ou "tic tac" issue du protocole CCM (figure 3.1). La durée des enregistrements varie de 48s à 196s avec une moyenne de 60s pour les contrôles et 85s pour les patients.

FIGURE 3.1 – *Le texte "Le cordonnier"*

Dans un petit village de la montagne, il y a un pauvre cordonnier, tout vieux et tout cassé. Les villageois lui apportent des chaussures à réparer. Mais il ne travaille pas vite. Tous les soirs, il mange tout seul, bien tristement.
Ce soir, il a devant lui, un gros tas de souliers et de guêtres à recoudre.
-"Jamais je ne pourrai les réparer. Je suis trop âgé et trop malade."
Près de lui, la grosse horloge fait: tic tac, tic tac. Le pauvre vieux, tout découragé, s'endort.
Aussitôt, l'horloge s'ouvre, et deux petits lutins sautent sur le plancher. L'un s'appelle Tic, l'autre s'appelle Tac.
- "Rangeons les étagères, réparons les souliers, recousons le linge", dit Tic.
- "Préparons un gâteau, mettons du gui au plafond, changeons ces vieux rideaux", ajoute Tac.
Minuit sonne! Les deux vaillants petits lutins rentrent dans la pendule. Le lendemain, le pauvre cordonnier s'éveille:
- "O joie! Qui a préparé ce bon gâteau? Qui donc a rangé la maison?"
- "Tic tac! Tic tac!", dit la vieille horloge.

Pour chaque enregistrement, une transcription orthographique a été réalisée par un auditeur humain. Un guide/protocole de transcription a été mis en place spécifiquement au sein du projet *DesPhoAPady* afin de prendre en compte le codage de quelques cas particuliers tels que les distorsions d'un phonème, l'élision ou l'insertion de mots dans le texte initial :

- les supressions de phonèmes ou de mots sont annotées entre parenthèses. Par exemple : l'annotation "pauv(r)e" marque l'élision du phonème "r ", l'annotation "dans un (petit) village" transcrit la suppression du mot "petit" de la phrase ;
- les distorsions ou substitutions de phonèmes ou de mots sont transcrites entre deux balises "[su]" avec la transcription du mot original ainsi que la prononciation effective codée en code SAMPA. Par exemple l'annotation "[su=grosse] gRa∼de [su]" transcrit la substitution du mot "grosse" par le mot "grande".
- les répétitions et les faux départs sont annotés en utilisant un trait d'union suivant la répétition. Par exemple : "le- le- le pauvre cordonnier" transcrit un mauvais départ au niveau du "le".
- l'insertion de mots par rapport au texte original est notée entre étoiles. Par exemple "pauvre *petit* cordonnier" transcrit l'insertion du mot "petit" vis-à-vis du texte initial.

Cependant, au vu du traitement automatique prévu dont notamment la détection automatique des anomalies, la transcription orthographique réalisée n'avait pas à être la plus fidèle possible de la production réalisée par le locuteur. La figure 3.2 donne un exemple d'une transcription d'un enregistrement selon le protocole décrit précédemment.

FIGURE 3.2 – *Exemple d'annotation pour un enregistrement d'un patient*

<s> dans un petit village de la montagne il y a un pauvre cordonnier tout vieux et tout casse1 </s> <s> les villageois lui apportent des chaussures a2 re1parer mais il ne travaille pas vite </s> <s> tous les soirs il mange tout seul bien tristement </s> <s> ce soir il y a devant lui un gros tas de souliers et de gue3tres a2 recoudre </s> <s> jamais je ne pourrai les re1parer </s> <s> je suis trop a3ge1 et trop malade *sevRE* </s> <s> pre2s de lui la grosse horloge fait *s* tic tac tic tac </s> <s> le pauvre vieux tout de1courage1 s_endort </s> <s> aussito3t l_horloge s_ouvre et deux petits lutins sautent sur le plancher </s> <s> l_un s_appelle tic l_autre s_appelle tac </s> <s> rangeons les e1tage2res re1parons les souliers recousons le linge *zi* dit tic </s> <s> pre1parons un ga3teau mettons du gui au plafond changeons ces vieux rideaux ajoute tac </s> <s> minuit sonne </s> <s> les deux *ve* vaillants petits lutins rentrent dans la pendule </s> <s> le lendemain le pauvre cordonnier s_e1veille </s> <s> o3 joie </s> <s> qui a *pa* pre1pare1 ce beau ga3teau </s> <s> qui donc a range1 la maison </s> <s> tic tac tic tac dit la vieille horloge [ext] *SS* *tRE* *bje~* [ext] </s>

Tous les enregistrements des patients ont été analysés par un expert humain afin d'annoter au niveau phonème les anomalies observées dans la parole. L'annotateur avait comme tâche de signaler tout phonème issu d'un alignement automatique de la parole comme normal ou anormal et dans ce cas d'indiquer le type d'anomalies observées (bruit, voisement, distorsion spectrale, etc.). Aidé par le logiciel Praat (Boersma et al., 2002), l'expert avait aussi pour mission de corriger l'alignement automatique afin d'avoir une référence pouvant être utilisée pour l'évaluation de la qualité de l'alignement automatique. Le tableau 3.1 regroupe les différentes informations liées à chaque locuteur de ce corpus.

3.2.2 Le corpus *DesPhoAPady*

Ce corpus a été constitué dans le cadre du projet *DesPhoAPady* décrit précédemment. Deux ensembles d'enregistrements ont été utilisés pour sa construction : le corpus *CCM* et le corpus des hôpitaux d'Aix.

Corpus *CCM*

Durant plus de 30 ans, Dr Claude Chevrie-Muller et son équipe du laboratoire d'étude de la voix et de la parole INSERM U3 ont enregistré des patients qui lui ont

TABLE 3.1 – *Informations sur le corpus VML indiquant le nombre d'enregistrements, le nombre de phonèmes prononcés et le nombre de phonèmes annotés comme anormaux par l'expert humain. (Les valeurs sont des moyennes calculées sur les différents enregistrements disponibles par locuteur).*

	Locuteurs	# d'enregistrements	# moyen de phonèmes	# moyen de phonèmes anormaux	% de phonèmes anormaux
Locuteurs dysarthriques	Homme 1	4	532	50	9,5
	Homme 2	3	116	43	37,5
	Homme 3	5	549	84	15,3
	Homme 4	6	268	80	30.2
	Femme 1	5	529	87	16,5
	Femme 2	5	100	77	76,3
	Femme 3	4	540	82	15,2
	Femme 4	3	306	102	33,5
Locuteurs contrôles	C Homme 1	4	561	-	-
	C Homme 2	3	554	-	-
	C Femme 1	4	559	-	-
	C Femme 2	5	558	-	-
	C Femme 3	4	553	-	-
	C Femme 4	3	557	-	-

étaient adressés par différents neurologistes pour l'évaluation de leurs troubles de parole. Ce travail a permis la collection d'un unique et important corpus de parole française liée à des troubles neurologiques connus sous le nom de "Pathologie de la voix et de la parole en neurologie" ou le corpus *CCM*. Ce corpus contient environ 1000 heures de parole pathologique produites par 5000 patients (adultes et enfants) souffrant souvent de dysphonie et de dysarthrie, mais aussi d'aphasie, de bégaiement et de troubles psychiatriques. Le corpus *CCM* contient des patients atteints de plusieurs pathologies causant des dysarthries. Un sous groupe de ces locuteurs souffrant de SLA, de maladie de Parkinson et d'ataxie cérébelleuse a été extrait et associé à des locuteurs sains afin de former le corpus *DesPhoAPady*. Tous les enregistrements ont été effectués dans une chambre sourde ou silencieuse et les signaux audio et electroglottographique ont été enregistrés sur deux canaux différents d'une cassette Revox. Un travail important de numérisation de ces enregistrements a été nécessaire et réalisé dans le cadre du projet *DesPhoAPady*.

Corpus des hôpitaux d'Aix - service de neurologie

Ce corpus comporte des patients enregistrés sur la station EVA (Teston et al., 1999) et comporte 990 patients majoritairement atteints de la maladie de Parkinson et de 160 locuteurs contrôles. Des locuteurs atteints de la maladie de Parkinson issus de ce corpus ont aussi été intégrés au corpus *DesPhoAPady*.

Corpus *DesPhoAPady*

Tous les locuteurs retenus dans le corpus *DesPhoAPady* ont lu le même texte, "le cordonnier" et tous les enregistrements ont été transcrits en respectant le même protocole utilisé pour le corpus *VML*.
Le corpus *DesPhoAPady* est composé de :

— 27 patients atteints de SLA (dysarthrie mixte) ;
— 31 patients atteints de la maladie de Parkinson (dysarthrie hypokinétique) ;
— 21 patients atteints d'ataxie cérébelleuse (dysarthrie ataxique) ;
— 29 locuteurs contrôles.

Pour ce corpus, la réalisation d'une annotation en anomalies au niveau phonème par un expert humain a été jugée trop coûteuse. Cependant, et afin de permettre l'exploitation du corpus, une autre évaluation perceptive a été effectuée sur tous les enregistrements selon la grille d'évaluation GEPD (section 2.2.3). Cette évaluation représente "une synthèse des différentes échelles d'évaluation perceptive standardisées" (Lhoussaine, 2012) de la parole dysarthrique. Elle décrit 9 critères de la parole tous cotés sur une échelle quantitative de type échelle de classes :

— le grade de la dysarthrie, coté sur une échelle à 4 points : de 0 pour l'absence de la dysarthrie à 3 pour une dysarthrie sévère ;
— l'irrégularité globale de la voix et/ou de la parole, cotée sur une échelle à 4 points : de 0 pour l'absence d'irrégularités à 3 pour la présence d'irrégularités sévères ;
— la mélodie, cotée sur une échelle à 7 points : de -3 pour l'absence de mélodie, 0 pour une mélodie normale et 3 pour une parole très mélodique ;
— la vitesse de parole (débit), cotée sur une échelle à 7 points : de -3 pour un débit très faible, 0 pour un débit normal et 3 pour un débit de parole très rapide ;
— le nasonnement, coté sur une échelle à 4 points : de 0 pour l'absence de nasonnement à 3 pour un nasonnement sévère ;
— les palilalies [3], coté sur une échelle à 4 points : de 0 pour l'absence d'anomalie à 3 pour des altérations sévères ;
— la réalisation articulatoire, cotée sur une échelle à 4 points : de 0 pour une articulation normale à 3 pour une articulation sévèrement altérée ;
— l'irrégularité de la vitesse de parole (du débit), cotée sur une échelle à 4 points : de 0 pour l'absence d'irrégularités à 3 pour des irrégularités sévères ;
— l'intelligibilité, cotée sur une échelle à 4 points : de 0 pour une intelligibilité normale à 3 pour une parole inintelligible.

3. Il s'agit d'un trouble de l'initiation de la parole résultant en une répétition de syllabes. Ce symptôme doit être distingué du bégaiement.

FIGURE 3.3 – *Évaluation perceptive du corpus DesPhoAPady suivant la grille d'évaluation perceptive GEPD.*

Population	Sexe	Nb de locuteurs	Grade	Irrégularité globale	Nasonnement	Palilalie	Articulation	Irrégularité du débit	Intelligibilité	Mélodie		Débit	
										< 0	> 0	< 0	> 0
Ataxie cérébelleuse	F	8	1,51	1,14	0,85	0,43	1,34	1,25	1,00	-0,88	0,73	-1,27	0,76
	H	13	1,46	1,04	0,78	0,42	1,33	1,20	0,93	-0,98	0,45	-1.31	0.18
Maladie de Parkinson	F	8	0,65	0,61	0,31	0,33	0,49	0,69	0,43	-1,15	0,18	-	0,81
	H	23	1,04	0,96	0,32	0,52	0,85	1,11	0,75	-1,17	-	-0.95	1.09
SLA	F	23	1,91	0,96	1,63	0,11	1,66	0,72	1,25	-1,14	0,44	-1,46	0,00
	H	14	1,48	1,33	0,88	0,94	1,31	1,11	0,12	-1,15	0,09	-1.39	0.82
Contrôles	F	14	0,10	0,12	0,05	0,08	0,03	0,12	0,05	-0,40	0,14	-0,30	0,48
	H	15	0,16	0,13	0,21	0,03	0,12	0,18	0,03	-0,45	0,27	-0.25	0.31
Maladie lysosomale	F	4	2,15	1,68	1,58	0,88	1,90	1,93	1,60	-1,30	0,20	-1.40	2.80
	H	4	1,90	1,68	0,88	0,88	1,70	1,93	1,53	-0,85	-	-2.10	0.80

Cette évaluation perceptive a été réalisée par deux jurys distincts : un premier composé d'experts et un deuxième de naïfs. Elle ne prenait en considération que la première minute de chaque enregistrement. Le premier enregistrement de chaque patient issu du corpus *VML* a été également inclus dans cette évaluation.

Dans notre travail, nous nous sommes restreints à l'évaluation faite par le jury d'experts jugée plus pertinente. Ce jury contenait 11 juges experts : 10 orthophonistes et un médecin ORL-phoniatre disposant chacun d'une expérience dans l'écoute de la parole dysarthrique variant de 7 à 26 ans.

Le tableau 3.3 regroupe quelques informations portant sur le corpus notamment le nombre de locuteurs de chaque population (pathologies et contrôles) ainsi que les moyennes des évaluations du jury pour chaque critère.

3.2.3 Le corpus *TypALoc*

Le corpus *TypALoc* a été construit dans le cadre du projet ANR *TYPALOC* décrit précédemment. Ce corpus reprend des enregistrements de patients issus du corpus *DesPhoAPady* et y rajoute de nouveaux locuteurs contrôles (Meunier et al., 2016).

Dans ce corpus, un choix portant sur la diversification au niveau de l'âge et d'accents régionaux de la parole contrôle a été fait. Cela visait à répondre à la variabilité observée dans les profils dysarthriques rencontrés. En effet, deux groupes de locuteurs sains ont été choisis, une population (3 hommes et 3 femmes âgés de 63 à 82 ans) comportant des personnes âgées de 6 locuteurs. La plupart de ces locuteurs sont issus du nord de la France. Les autres locuteurs contrôles sont plutôt jeunes (entre 29 et 47 ans) et leurs enregistrements sont issus d'un large corpus de parole conversationnelle, le corpus *CID* (Bertrand et al., 2008). La plupart de ces locuteurs sont issus du sud de la France. Tous les locuteurs contrôles et dysarthriques ont été enregistrés dans des chambres sourdes ou dans des cabines de son équipées de microphones de haute qualité. Pour tous les locuteurs, des enregistrements de parole produite dans deux styles de parole différents sont fournis. La tâche de lecture reste identique à celle des corpus déjà décrits avec la lecture du texte "Le cordonnier". Le corpus *TypALoc* dispose aussi d'enregistrements de parole spontanée pour tous les locuteurs (patients et contrôles). La méthode de collecte de cette parole dépendait de la population. Pour les patients et les locuteurs sains âgés, la parole spontanée a été produite dans le cadre d'une interview conduite par un chercheur comprenant des extraits de parole produite par le chercheur (monologue virtuel). Les locuteurs devaient parler de leur vie quotidienne, leurs histoires personnelles, leurs travaux ainsi que quelques événements personnels. Par contre, les locuteurs sains jeunes devaient parler d'événements ou situations particulières (séquences narratives) lors d'une conversation relaxée et interactive avec un seul interlocuteur.

Le choix des patients, issus du corpus DesPhoAPady, à inclure dans chaque population dysarthrique a été fait en respectant trois critères essentiels :

— la qualité de la parole des enregistrements audio ;

— la qualité et quantité de parole spontanée disponible. En effet, les enregistrements des patients sont fréquemment courts (moins de 2 minutes) dû à des effets de fatigue attendus. Les patients ayant les enregistrements les plus longs ont été retenus ;
— la sévérité de la dysarthrie. Afin d'être capable à la fois d'analyser la parole phonétiquement et en utilisant des outils de TAP, des cas de patients trop sévèrement atteints ainsi que des enregistrements contenant une parole non intelligible et fortement distordue ont été exclus du corpus.

Trois populations dysarthriques ont été incluses dans ce corpus, toutes les trois sont associées à des maladies neurologiques dégénératives et ont été choisies afin d'illustrer les troubles observés sur trois systèmes neurologiques différents :

— 8 patients atteints de la maladie de Parkinson ;
— 12 patients atteints de SLA ;
— 8 patients atteints d'ataxie cérébelleuse.

Chaque enregistrement a été segmenté manuellement en des unités inter-pausales (UIP). Les UIP sont des séquences de parole séparées l'une de l'autre par des pauses silencieuses d'au moins 250ms. De plus, tous les bruits de types rires, respirations, intervention de l'interlocuteur, etc. ont été annotés et isolés de la parole du locuteur participant à l'étude. Pour chaque locuteur, une transcription orthographique réalisée au niveau de chaque UIP a été réalisée. Deux guides/protocoles de transcription différents ont été retenus afin de prendre en compte les deux styles de parole étudiés. Le protocole d'annotation de la parole lue est identique à celui utilisé précédemment. Celui correspondant à la parole spontanée a repris les mêmes instructions dans le cas de distorsions d'un phonème, d'élision ou de codage des nouveaux mots et s'est inspiré du guide de transcription du corpus *CID* pour inclure des codes pour quelques séquences inattendues telles que les pauses remplies (euh, mmh, be hein, hum), les noms propres (codés en SAMPA) et les onomatopées (ah, oh, eh, ouh, aie, paf, boum, etc.).

Puisque les enregistrements de la parole lue de ce corpus sont issus du corpus *DesPhoAPady*, nous disposons déjà de l'évaluation de la qualité de cette parole selon la grille d'évaluation GEPD par un jury d'experts. Une évaluation des enregistrements de la parole spontanée des mêmes patients par le même jury suivant la même grille a été réalisée. Les tableaux 3.4 et 3.5 détaillent les résultats de cette évaluation pour **la parole lue** et **la parole spontanée** respectivement du corpus *TypALoc*.

Finalement, tous les enregistrements des patients et des sujets contrôles de ce corpus contenant de la parole lue ou spontanée ont été alignés manuellement par des experts humains au niveau phonème. Cet alignement permet la délimitation des frontières de début et de fin de chaque phonème dans les enregistrements et nous sera utile pour évaluer le comportement des outils de TAP utilisés dans ce travail. Afin de réaliser cette segmentation, les experts humains corrigeaient les frontières des phonèmes générées par un alignement automatique de la parole (détaillée dans le chapitre 4). Cet alignement qu'on nommera manuel nous sera utile par la suite pour l'évaluation de la qualité de l'alignement automatique.

FIGURE 3.4 – *Résultats de l'évaluation perceptive du jury d'experts sur* ***la parole lue*** *produite par les patients du corpus TypALoc*

			Ataxie cérébelleuse									
Sexe	Age	Locuteur	Grade	Irrégularité globale	Mélodie	Débit	Nasonnement	Palilalie	Articulation	Irrégularité du débit	Intelligibilité
F	59	CCM-002710-01	2.1	1.3	-0.4	-1.3	1.5	0.0	1.6	1.4	1.2
H	32	CCM-003094-01	1.5	1.7	0.3	-2.2	0.7	0.4	1.4	1.7	0.8
F	69	CCM-003110-01	1.3	1.2	-0.6	-0.8	0.4	0.1	0.5	1.0	0.6
F	35	CCM-003493-01	0.9	0.8	-1.5	-1.5	0.2	0.5	0.7	0.6	0.5
H	77	CCM-003998-01	1.5	1.2	-1.5	-1.9	0.5	0.8	0.7	1.4	1.1
H	45	CCM-004523-01	0.8	0.3	-0.5	-0.9	0.3	0.1	0.7	1.0	0.3
F	68	CCM-004538-01	1.0	0.9	-0.5	0.8	0.7	0.6	1.1	1.2	0.6
H	55	CCM-004773-01	1.2	0.9	-1.5	-1.3	0.5	0.5	1.1	1.1	0.7
			Maladie de Parkinson								
H	64	CCM-001773-01	0.4	0.3	-0.1	1.7	0.1	0.0	0.4	0.6	0.4
H	62	CCM-003130-01	0.8	0.9	-1.5	0.0	0.0	0.5	0.8	0.6	0.6
F	60	CCM-003148-01	1.3	1.4	-1.4	1.5	0.5	0.8	1.2	1.8	0.9
F	81	CCM-003346-01	0.5	0.6	-0.7	0.6	0.1	0.5	0.2	0.7	0.4
H	48	CCM-003557-01	0.6	0.3	-1.6	-0.5	0.2	0.6	0.1	1.3	0.4
H	61	CCM-003733-01	1.4	1.7	-1.6	-0.5	0.4	0.5	1.1	2.2	1.2
H	58	CCM-003734-01	0.5	0.5	-0.7	-0.1	0.2	0.2	0.4	0.7	0.1
H	77	CCM-003848-01	1.3	1.1	-1.2	1.5	0.5	0.3	0.6	1.2	0.6
			SLA								
F	70	PHO-000024-01	1.5	0.1	1.2	-0.2	2.6	0.0	1.0	0.5	1.0
F	81	PHO-000566-01	2.9	1.3	-2.0	-2.7	2.3	0.0	2.3	0.8	1.7
F	68	PHO-000814-01	2.3	1.0	-1.0	-2.0	2.2	0.0	1.8	0.8	1.3
F	63	PHO-001070-01	2.4	1.1	0.1	-1.1	2.1	1.0	2.2	1.0	1.6
F	50	PHO-001329-01	0.9	0.7	0.0	-0.5	0.5	0.0	1.1	0.5	0.5
H	67	PHO-001473-01	2.1	0.9	0.1	-2.2	1.1	0.0	1.5	0.8	1.0
F	62	PHO-001499-01	1.8	1.0	0.9	-1.4	1.6	0.1	1.5	1.1	1.0
F	-	PHO-001522-01	2.6	0.8	-2.3	-2.5	1.2	0.0	1.8	0.6	1.4
H	-	PHO-001594-01	2.2	1.5	-1.5	-1.6	1.4	0.4	2.3	1.0	1.6
H	71	PHO-001670-01	1.9	1.1	-1.1	-1.2	1.2	0.0	1.9	0.7	1.5
H	74	PHO-001836-01	2.5	1.1	-1.0	-1.5	2.7	0.4	2.5	0.9	2.1
H	56	PHO-307175-01	1.4	1.2	-0.6	1.5	2.0	0.5	1.2	1.7	1.0

FIGURE 3.5 – *Résultats de l'évaluation perceptive du jury d'experts sur* ***la parole spontanée*** *produite par les patients du corpus TypALoc*

						Ataxie cérébelleuse				
CCM-002710-01	F	1,8	0,5	-1,2	0,9	1,2	0,8	1,6	1,2
CCM-003094-01	H	1,9	-0,1	-1,4	0,9	0,9	0,7	1,6	1,1
CCM-003110-01	F	1,2	-0,6	-0,6	0,5	0,5	0,3	0,9	0,6
CCM-003493-01	F	0,8	-1,1	-0,7	0,5	0,3	0,1	0,5	0,5
CCM-003998-01	H	1,2	-1,0	-1,5	0,0	0,5	0,3	0,5	0,5
CCM-004523-01	H	0,8	-0,7	-0,5	0,5	0,5	1,1	0,8	0,8
CCM-004538-01	F	0,6	-0,1	0,0	0,7	0,5	0,0	0,8	0,4
CCM-004773-01	H	1,5	-1,1	-0,9	0,6	0,7	0,9	1,1	0,9
						Maladie de Parkinson				
CCM-001773-01	H	1,5	-1,4	1,7	1,6	0,4	0,8	1,3	1,1
CCM-003130-01	H	1,2	-1,6	-0,7	0,7	0,4	0,1	1,3	0,8
CCM-003148-01	F	0,5	-0,3	-0,3	0,4	0,4	0,9	0,7	0,5
CCM-003346-01	F	0,4	-0,2	-0,3	0,3	0,3	0,3	0,3	0,3
CCM-003557-01	H	0,8	-1,1	-0,5	0,5	0,4	0,1	0,6	0,7
CCM-003733-01	H	1,6	-1,8	-1,2	0,8	0,5	0,4	1,4	1,3
CCM-003734-01	H	0,6	-0,5	-0,5	0,5	0,5	0,1	0,7	0,7
CCM-003848-01	H	1,3	-1,2	0,6	0,5	0,6	0,4	0,5	0,8
						SLA				
PHO-000024-01	F	1,2	0,0	0,5	0,5	2,0	0,5	1,0	0,5
PHO-000566-01	F	2,7	-1,5	-2,6	0,2	1,6	0,2	2,1	1,7
PHO-000814-01	F	2,5	-0,3	-1,8	0,6	2,1	0,2	2,0	1,6
PHO-001070-01	F	2,0	-0,2	-1,6	0,5	1,5	0,5	1,9	1,3
PHO-001329-01	F	1,3	-0,2	0,0	0,6	1,3	0,3	1,3	0,9
PHO-001473-01	H	2,1	-0,4	-2,2	0,4	0,9	0,2	1,6	1,2
PHO-001499-01	F	1,8	0,3	-1,5	0,4	1,7	0,0	1,2	0,6
PHO-001522-01	F	2,5	-2,1	-2,3	0,2	0,8	0,5	1,9	1,4
PHO-001594-01	H	1,7	-1,4	-1,3	0,2	0,9	0,3	1,7	1,5
PHO-001670-01	H	2,1	-0,8	-0,5	0,7	1,5	0,0	1,9	1,5
PHO-001836-01	H	2,7	0,3	-1,8	0,4	2,3	0,1	2,3	2,2
PHO-307175-01	H	1,6	-0,5	1,1	1,6	1,7	1,7	1,0	1,0

3.2.4 Le corpus *BREF*

Le corpus *BREF* a été construit et développé au sein du Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) en 1991 (Lamel et al., 1991). Plusieurs unités et composantes ont contribué dans le financement de ce projet : le GDR-PRC Communication Homme/Machine, la CEE (projet ESPRIT Polyglot) et l'Aupelf-Uref. L'élaboration de ce corpus était dans le cadre de travaux portant sur le développement de machines de dictée. Ce corpus fournissait des données pour le développement de ces machines ainsi que l'évaluation des systèmes de reconnaissance de la parole.

Le corpus BREF contient plus de 100 heures de parole lue produites par 120 locuteurs (65 femmes et 55 hommes). Tous les locuteurs vivaient dans la région parisienne et sont tous français sauf 6 (4 marocains, 2 luxembourgeois). Sur les 120 locuteurs, 50 locuteurs ont approximativement produit 10000 mots chacun. Le reste des locuteurs ont lu des textes contenant environ 5000 mots chacun. Tous les enregistrements ont été fais en stéréo dans une chambre sourde. De plus, un contrôle sur la correspondance de la parole produite avec le texte lu a été réalisée. Les textes lus sont issus de 3 mois du journal français "Le Monde". En total, 11 002 textes ont été sélectionnés sur un critère de maximisation de contexte phonémique et du nombre de différents mots prononcés tout en restant assez facile à lire à haute voix.

3.2.5 Le corpus *Ester*

Le corpus Ester (Galliano et al., 2005) a été élaboré dans le cadre d'une campagne d'évaluation réalisée entre 2003 et 2005. Cette campagne portait sur les systèmes de transcription automatique d'enregistrements radiophoniques français.

Le corpus est composé d'environ 100 heures d'enregistrements manuellement transcrits. Ces enregistrements proviennent de plusieurs sources radio telles que : France Inter, France Info, Radio France Internationale et Radio Télévision Marocaine et Radio classique. Ces programmes ont été enregistrés en 1998, 2000, 2003 et 2004. Le tableau 3.2 donne quelques informations sur les enregistrements radiophoniques du corpus Ester.

TABLE 3.2 – *Information sur le corpus Ester*

Radio	Durée (heures)
France Inter	35
France Info	10
Radio France International	25
Radio Télévision Marocaine	20
Radio Classique	10

3.3 Mesures d'évaluation

Dans le cadre de ce travail, une approche de détection automatique des anomalies au niveau phonème dans la parole dysarthrique sera préposée. Cette approche repose, entre autre, sur une première phase d'alignement automatique de la parole en phonèmes. Différentes mesures d'évaluation ont alors été utilisées pour l'évaluation du comportement des différents outils de TAP utilisés. Ces mesures permettront l'évaluation de la qualité de l'alignement automatique de la parole (par rapport à l'alignement manuel des experts) ainsi que la capacité et la fiabilité de l'approche proposée dans la détection de phonèmes anormaux.

3.3.1 Évaluation de la qualité de l'alignement automatique

Comme décrit dans la section précédente, les corpus *VML* et *TypALoc* sont des corpus pour lesquels nous disposons d'alignements manuels réalisés par des experts en utilisant l'outil Praat. Afin de mesurer la qualité de l'alignement automatique réalisé sur ces corpus, nous utilisons ces alignement manuels comme référence. Trois mesures ont été proposées par segment (phonème) :

— le décalage au début du phonème DD : cette mesure est donnée par la différence entre la frontière de début du phonème issue de l'alignement automatique et celle issue de l'alignement manuel ;
— le décalage au milieu du phonème DM : cette mesure est donnée par la différence entre le point intermédiaire du phonème issu de l'alignement automatique et celui issu de l'alignement manuel ;
— le différence de durées de phonème $DDur$: cette mesure est donnée par la différence entre la durée du phonème issu de l'alignement automatique et celle issue de l'alignement manuel.

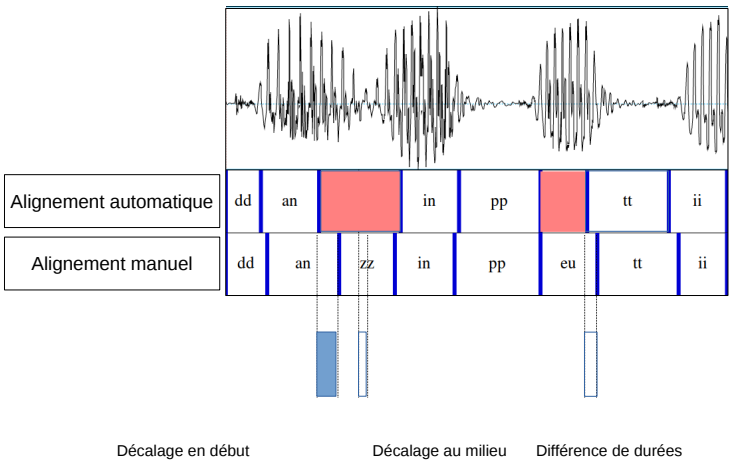
Ces différentes mesures sont calculées pour chaque phonème présent dans la séquence définie par l'alignement automatique. Des mesures globales par patient/catégorie phonétique/pathologie y sont tirées. De plus et afin de pouvoir valider la qualité globale de l'alignement automatique de la parole, le taux des phonèmes bien alignés est calculé pour chaque locuteur. Un phonème est considéré comme bien aligné si le décalage entre ses frontières issues de l'alignement automatique et de l'alignement manuel ne dépasse pas les 20ms. Il s'agit du seuil utilisé dans la majorité des études portant sur la qualité d'alignement automatique de la parole (Audibert et al., 2010; Goldman, 2011).

La figure 3.6 rapporte un exemple des mesures calculées pour chaque phonème.

3.3.2 Évaluation de la détection d'anomalies

Différentes mesures ont été utilisées afin d'évaluer la qualité de l'approche proposée de détection automatique d'anomalies dans la parole dysarthrique.

FIGURE 3.6 – *Mesures d'évaluation de la qualité de l'alignement automatique*



Dans le cas du corpus *VML*, la fiabilité de l'approche est étudiée en comparant les annotations réalisées automatiquement avec celles issues de l'annotation par l'expert humain. En effet, cette annotation experte sera considérée tout au long de l'étude comme la référence standard pour ces enregistrements.

Les contraintes qui résultent de cette hypothèse sont celles liées à toute annotation humaine de parole pathologique telles que la subjectivité et la non reproductibilité. En effet, la difficulté de la tâche et le fait que nous ne disposons que des annotations d'un seul expert limitent le degré de certitude concernant cette référence. En effet, comme décrit précédemment, ce type d'évaluation est souvent considéré comme subjectif et non reproductible. Une des solutions permettant de contourner ces limites est la multiplication des experts lors de l'évaluation afin de lui fournir plus de robustesse. Cependant, la tâche d'annotation au niveau phonème étant trop coûteuse en ressources humaines et en temps, cette solution n'a pu être appliquée dans notre cas.

Néanmoins, et dans la pratique clinique, cette évaluation perceptive de la parole dysarthrique est souvent réalisée par un seul expert. De plus, l'approche proposée dans ce travail vise dans un premier temps à assister les professionnels dans le cadre de leur travail et non à les substituer et n'a pas donc obligation à remplacer l'évaluation pouvant être fournie par un jury d'experts. Elle peut par contre être utile à l'expert pour faciliter son annotation et l'objectiver en lui fournissant des avis concordants ou opposés aux siens. Nous avons alors décidé d'utiliser cette annotation fournie par l'expert humain comme référence.

Pour cette évaluation, nous proposons deux mesures issues du domaine de l'extraction de l'information (Makhoul et al., 1999) qui se concentrent sur la détection d'anomalies.

— la mesure du rappel de détection d'anomalies, mesurée entre 0 et 1, nommée *AnRappel*, et qui est donnée par le ratio entre le nombre de segments (pho-

nèmes) correctement détectés comme anomalies par l'approche automatique et le nombre de segments annotés comme anomalies par l'expert dans la référence. Ce ratio mesure la performance de l'outil et son aptitude à détecter les anomalies dans les enregistrements utilisés. Plus cette mesure se rapproche de 1, plus notre système est performant dans la tâche de détection automatique d'anomalies ;
— la mesure de précision de détection d'anomalies, mesurée entre 0 et 1, nommée *AnPrec*, et qui est donnée par le ratio entre le nombre de segments (phonèmes) correctement détectés comme anomalies par l'approche automatique et le nombre de segments annotés comme anomalies par l'approche (correctement ou faussement). Ce ratio mesure la précision de l'outil dans la détection des anomalies et correspond à l'inverse du taux de faux positifs détectés par l'approche. Plus cette mesure se rapproche de 1, plus notre système est précis dans sa détection des vraies anomalies.

Ces deux mesures se concentrent seulement sur la détection de phonèmes anormaux. Des mesures similaires peuvent être proposées pour l'annotation des phonèmes normaux mais il ne s'agit pas de la tâche principale. De plus, le nombre d'anomalies détectées sur les enregistrements des locuteurs contrôles permettra d'évaluer le comportement de l'approche sur les phonèmes normaux. De plus, il est important de noter que les *AnRappel* et *AnPrec* sont deux mesures complémentaires dans l'évaluation de l'approche automatique. En effet, si le *AnRappel* est nécessaire pour la mesure de l'efficience du système, le *AnPrec* est important afin de mesurer l'utilisabilité des anomalies détectées automatiquement dans le cadre d'applications cliniques ou phonétiques et leur pertinence.

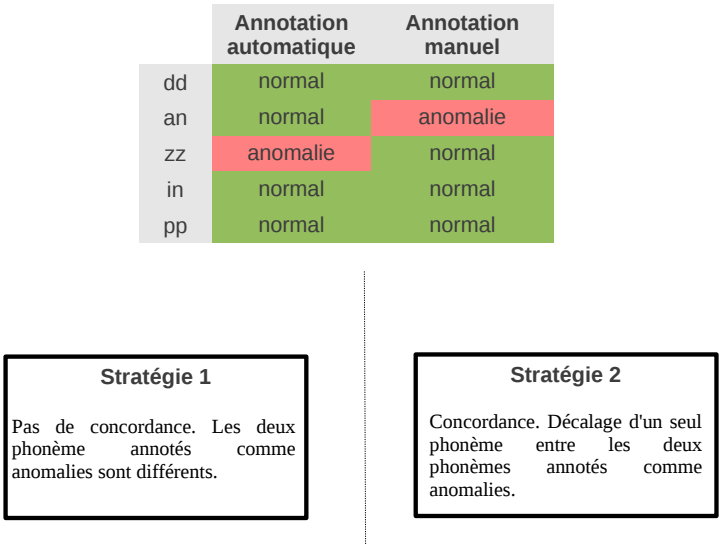
De plus, la comparaison entre les annotations automatiques et manuelles est réalisée selon deux stratégies distinctes :

— la comparaison est effectuée entre les deux annotations, un phonème à la fois, sans considération du contexte local. Dans ce cas, on considère une concordance entre les annotations du système et de l'expert uniquement dans le cas où les deux ont donné le même label (normal/anomalie) pour le même phonème ;
— on compare les annotations effectuées sur chaque phonème ainsi que ces deux phonèmes adjacents (contexte local). Dans ce cas, si l'expert humain considère un phonème comme anomal alors que le système automatique détecte le phonème précédent ou suivant comme anormal et non le phonème ciblé, une bonne concordance est tout de même notée. Cette stratégie vise la prise en compte d'éventuels décalages d'un phonème lors de la phase d'alignement automatique. Cette stratégie peut être considérée moins sévère que la précédente. En effet, elle permet d'évaluer davantage la capacité de l'approche à retrouver les anomalies que sa précision dans la tâche. Nous pouvons alors s'attendre à ce que les mesures de *AnRappel* et *AnPrec* selon cette stratégie soient plus élevées que celles calculées selon la stratégie 1.

Dans le cadre des patients issus des corpus *DesPhoAPady* et *TypALoc* ainsi que les sujets contrôles pour lesquels aucune annotation humaine au niveau phonème n'est disponible, le calcul des mesures *AnRappel* et *AnPrec* est impossible.

Le taux de phonèmes annotés par le système comme anormaux par rapport au

FIGURE 3.7 – *Stratégies de comparaison des annotations d'anomalies de l'expert et de l'approche automatique*



nombre total de phonèmes sera alors calculé pour chaque enregistrement. L'étude de ces taux permettra de comparer le comportement du système entre les différentes pathologies. De plus, et sur la base des mesures perceptives globales de la qualité de parole décrites dans les parties 3.2.2 et 3.2.3 et produites par le jury de 11 experts, l'évolution du comportement de l'approche face à des dysarthries de sévérités différentes sera mesurée. Ces taux d'anomalies permettront aussi d'étudier les différences éventuelles dans le comportement de l'approche face à la parole lue et spontanée. Pour les locuteurs contrôles, nous pouvons supposer l'absence d'anomalies dans leurs enregistrements. Cette hypothèse ne peut être confirmée sans l'évaluation perceptive de ces enregistrements par des experts humains ; la présence éventuelle de quelques distorsions, réductions ou disfluences dans leurs productions pouvant être considérées comme des anomalies par le système automatique n'est pas exclue. Cependant, on peut affirmer la non présence d'anomalies à caractères pathologiques dans ces productions puisqu'il s'agissait d'un des critères essentiels lors de la sélection des locuteurs contrôles.

3.4 Conclusion

Tout au long de ce travail, les différents corpus détaillés précédemment ont été utilisés. Cette multitude de corpus peut parfois rendre la comparaison des résultats et des performances de l'approche proposée difficile dû aux différents types d'annotations fournis (annotation d'anomalies au niveau phonème sur le corpus *VML*, annotation globale sur les corpus *DesPhaAPady* et *TypALoc*). Cependant, cela souligne aussi la dif-

ficulté à laquelle font face les travaux portant sur l'évaluation automatique de la parole pathologique. En effet, peu de corpus de données existe, leur taille est souvent très limitée et des différences dans les protocoles d'enregistrement et d'annotation peuvent subsister ce qui peut restreindre leur utilisabilité.

Néanmoins, nous soutenons que ce contexte porte aussi un intérêt indéniable. En effet, et face à la large variabilité dans la parole dysarthrique, l'étude du comportement des différents outils de TAP face à différentes classes dysarthriques, différents niveaux de sévérité, différents styles de parole et différentes formes d'annotation d'experts humains peut soutenir l'hypothèse d'une généralisation des résultats obtenus et observations relevées.

Deuxième partie

Apport des outils de TAP face à la parole dysarthrique

Chapitre 4

Alignement automatique de la parole

Sommaire

Dans ce chapitre, nous décrivons l'outil d'alignement automatique de la parole en phonème utilisé dans ce travail. Cet alignement automatique est la première phase dans notre approche de détection automatique d'anomalies au niveau phonème dans la parole dysarthrique qui sera présentée dans le chapitre 5.
Une analyse du comportement de cet outil face à la parole dysarthrique sera détaillée dans la deuxième partie du chapitre. Elle mettra en évidence l'effet de la grande variabilité observée dans la parole dysarthrique (pathologie/classe dysarthrique/sévérité) sur la précision de l'alignement. Les différences entre les catégories phonétiques seront aussi étudiées en fonction de leurs liens avec les caractéristiques des différentes pathologies/classes dysarthriques.

4.1 Alignement automatique de la parole

L'alignement automatique de la parole consiste à trouver les frontières de début et de fin pour chaque unité acoustique (phonème) produite dans le signal de parole. Ce processus nécessite des phases de paramétrisation du signal ainsi qu'une modélisation des unités acoustiques à utiliser lors de la segmentation.

4.1.1 Paramétrisation du signal

Le signal de parole est un signal très redondant ce qui le rend non adapté à une utilisation directe dans les systèmes de RAP. Une première phase de paramétrisation est alors nécessaire afin d'y extraire des informations permettant de le caractériser. Ces paramètres doivent être assez robustes face à des altérations de bruit ou de canal de transmission tout en restant assez discriminants afin de conserver les informations propres à chaque signal et son (phonème) étudiés. Cette étape permet de transformer le signal de parole de nature continue en un ensemble discret de vecteurs de paramètres.

Prétraitement du signal

Une phase de prétraitement est souvent nécessaire avant la paramétrisation du signal. Dans notre travail, tous les enregistrements utilisés sont échantillonnés à 16 kHz. Le signal est ensuite décomposé en des segments courts (trames) sur lesquels le signal peut être considéré comme stationnaire. Ces trames sont extraites toutes les 10ms. Cependant, ce type de segmentation résulte en une altération du signal sur les bords de chaque trame générée (il s'agit des effets de bords). Plusieurs fenêtres de pondération peuvent alors être appliquées sur les trames afin de limiter ces effets (fenêtre de Hamming, Hanning, Blackman, etc.). Une fenêtre de Hamming a été utilisée dans notre outil d'alignement. Sa pondération est définie par l'équation suivante :

$$Hamming(i) \;=\; 0.54 - 0.46\, cos(\frac{2\pi i}{N}) \quad avec \;\; i \;\in\; [0, N-1] \tag{4.1}$$

où N est la taille de la fenêtre en nombre d'échantillons du signal.

Suite à cette décomposition, différentes techniques de paramétrisation du signal peuvent être utilisées. Les plus fréquentes sont les MFCC (Mel Frequency Cepstral Coefficients, (Davis et Mermelstein, 1980)), PLP (Perceptual Linear Prediction, (Hermansky, 1990; Hermansky et Cox, 1991)), LPCC (Linear Prediction Cepstral Coefficients, (Markel et Gray, 1976)) et RASTA-PLP (Relative Spectral PLP, (Hermansky et al., 1991)). Le contexte de notre travail étant fortement lié à l'évaluation perceptive humaine de la parole, la paramétrisation PLP a été choisie de fait de son intégration de concepts reflétant la perception humaine de la parole.

Paramétrisation par prédiction linéaire perceptuelle - PLP

Comme son nom l'indique, cette paramétrisation repose sur un modèle de la perception humaine de la parole. Elle est fondée sur le même principe que l'analyse prédictive tout en intégrant trois concepts de perception de la parole dans le but d'extraire une estimation plus précise du spectre auditif (Hermansky, 1990) :

— intégration des bandes critiques : la prédiction linéaire classique produit la même approximation de l'enveloppe spectrale pour toute la zone de fréquences utiles. Cette représentation est contradictoire avec le fonctionnement réel de l'appareil perceptif humain. En effet, l'oreille humaine a la faculté d'intégrer certaines zones de fréquences en bandes appelées "bandes critiques". Ces dernières sont reparties selon l'échelle de Bark [1]. Le passage de Bark en Hertz est obtenu par la transformation suivante :

$$F_{Hz} = 600 \ \sinh(\frac{F_{Bark}}{6}) \tag{4.2}$$

où $sinh()$ est le sinus hyperbolique. La nouvelle densité spectrale est alors échantillonnée selon cette nouvelle échelle ce qui augmente la résolution dans les basses fréquences ;

— la pré-accentuation du signal selon une courbe d'isosonie : les expériences psycho-acoustiques dans (Fletcher et Munson, 1933) ont montré que l'oreille humaine possède des caractéristiques non linéaires. Afin de simuler ce phénomène dans la paramétrisation PLP, la densité spectrale résultante de l'étape précédente est multipliée par une fonction de pondération approximant cette sensibilité ;

— compression en racine cubique : les deux précédentes transformations ne suffisent pas pour faire correspondre l'intensité mesurée à celle subjective perçue par l'humain (la sonie). La loi de Stevens présente la relation entre ces deux mesures comme suit :

$$Sonie = (Intensite)^{\frac{1}{3}} \tag{4.3}$$

Sur chacune des trames issues de la segmentation, 12 paramètres PLP et une mesure de l'énergie sont extraits. De plus, et afin de mieux refléter les phénomènes de co-articulation et les caractéristiques du locuteur, des paramètres dynamiques sont aussi extraits (Furui, 1986). Ces paramètres sont représentés par les dérivés temporelles d'ordre 1 (Δ) et 2 ($\Delta\Delta$) des coefficients statiques calculés initialement. Chaque trame est alors représentée par un vecteur de 39 paramètres.

4.1.2 Modélisation acoustique de la parole : Modèles de Markov Cachés

Afin de modéliser acoustiquement le signal de la parole, un ensemble réduit d'unités acoustiques est considéré. Ces unités peuvent être considérées comme des sons élémentaires de la langue. Généralement, l'unité choisie dans la modélisation d'une langue est le phonème. Un phonème peut être considéré seul ou dans le cadre de son contexte, on parle alors de modèle contextuel de parole. D'autre modélisations utilisent des unités plus larges comme la syllabe. Cependant, plus la taille de ces unités augmente, plus on a besoin d'unités pour couvrir toute la parole, ce qui n'est pas toujours possible dans les contextes expérimentaux.

1. Échelle psycho-acoustique mesurant la sonie.

Dans le cadre de ce travail, l'unité acoustique de modélisation utilisée est le phonème et les modèles sont non contextuels. La modélisation revient alors à représenter les séquences des vecteurs acoustiques issus de la phase de paramétrisation au niveau de chaque phonème.

La solution utilisée dans l'état de l'art pour cette modélisation repose sur les Modèles de Markov Cachés (Hidden Markov Models - HMM) gauche droite à trois états. Dans ces modèles, chaque phonème est représenté par 3 états et chaque état peut représenter au moins un des vecteurs acoustiques issus de la phase de paramétrisation.
Le tableau 4.1 détaille tous les phonèmes du français et leurs codes API (Alphabet Phonétique International), SAMPA (Speech Assessment Methods Phonetic Alphabet) ainsi que celui utilisé dans les systèmes de TAP au sein du LIA. Ce dernier codage sera utilisé par la suite.

Structure d'un HMM

Un HMM est un automate probabiliste à états finis pouvant générer une séquence d'observations selon un processus stochastique Markovien. Ce modèle est contrôlé par deux processus stochastiques : (1) un premier processus interne au HMM et donc caché à l'observateur qui débute sur l'état initial puis se déplace d'état en état en respectant la topologie du HMM (2) un second processus stochastique qui génère les observations (phonèmes) correspondant à chaque état parcouru par le premier processus.
Un HMM est caractérisé par l'ensemble de paramètres suivant :

- son ensemble d'états, $S = \{s_1, s_2, ..., s_N\}$ avec N le nombre d'états du modèle ;
- une matrice A de transitions entre les états où a_{ij} correspond à la probabilité de transition de l'état s_i vers l'état s_j. Les modèles utilisés en reconnaissance de la parole sont d'ordre 1, c'est-à-dire que la probabilité de passer dans un état dépend uniquement de l'état courant ;
- une matrice Π donnant la distribution initiale des états où π_i est la probabilité d'être dans l'état s_i au départ ;
- une matrice B des densités de probabilité des observations associées à chaque état s_i du modèle, avec $b_i(o_n) = P(x_n|s_i)$ représentant la probabilité d'émettre l'observation o_n étant dans l'état s_i. Cette probabilité est généralement modélisée par une mixture de gaussienne (Gaussian Mixture Model - GMM).

Un GMM est une somme pondérée de M distributions gaussiennes multi-dimensionnelles. Chacune de ces lois gaussiennes $N_i(X)$ est caractérisée par son poids w_i, sa moyenne μ_i et sa matrice de covariance Σ_i. La fonction de probabilité de l'état s_i s'écrit alors sous la forme :

$$p(X|s_i) = \sum_{i=1}^{M} w_i N_i(X) = \sum_{i=1}^{M} w_i \frac{1}{(2\pi|\Sigma_i|)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)} \tag{4.4}$$

Tout HMM peut alors être modélisé par l'ensemble des paramètres détaillés précédemment $\theta = \{N, A, \Pi, B\}$. Ces paramètres sont souvent estimés empiriquement sur de larges corpus de données.

TABLE 4.1 – *Les phonèmes du Français et leurs codes API (Alphabet Phonétique International), SAMPA (Speech Assessment Methods Phonetic Alphabet) ainsi que celui utilisé dans les systèmes de TAP au sein du LIA.*

API	SAMPA	LIA	Exemple
[p]	p	pp	patte
[b]	b	bb	bois
[t]	t	tt	toit
[d]	d	dd	dans
[k]	k	kk	quand
[g]	g	gg	gare
[f]	f	ff	fois
[v]	v	vv	vase
[s]	s	ss	tasse
[z]	z	zz	rose
[ʃ]	S	ch	choix
[ʒ]	Z	jj	gens
[j]	j	yy	paille
[m]	m	mm	main
[n]	n	nn	nom
[l]	l	ll	loix
[R]	R	rr	roi
[w]	w	ww	quoi, oui
[ɥ]	H	uy	juin, nuit
[i]	i	ii	vie
[e]	e	ei	ses, pêcher
[ɛ]	E	ai	seize, céder
[a]	a	aa	patte, plat, papa
[ɑ]	A	aa	pâte
[ɔ]	O	oo	comme, porte, mort
[o]	o	au	gros, mot, lot
[u]	u	ou	doux, coup, fou
[y]	y	uu	du, rue, bue
[∅]	2	eu	deux
[œ]	9	oe	neuf
[ə]	@	eu	justement
[ɛ̃]	e∼	in	vin, pain
[ã]	a∼	an	vent
[õ]	o∼	on	bon, pompe
[œ̃]	9∼	un	brun

Dans ce travail, trois modèles HMM différents ont été appris. Un premier modèle, HMM-Bref, indépendant du genre de dimension 128 a été appris sur le corpus *BREF*. Deux modèles dépendants du genre, HMM-Ester-F et HMM-Ester-H, de dimensions 256 ont été appris sur le corpus *Ester*. La figure 4.1 représente la structure générale des modèles HMM utilisés.

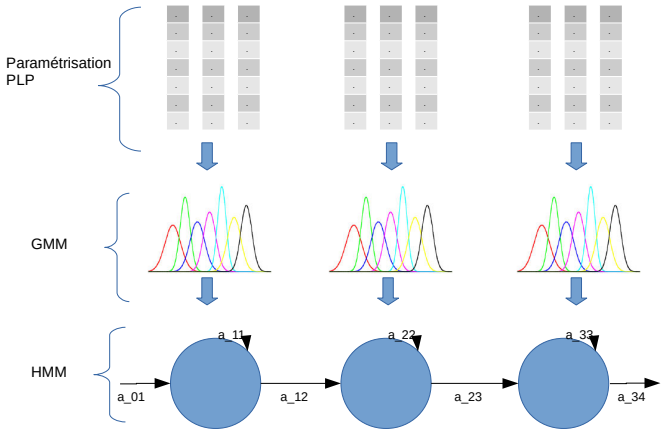


FIGURE 4.1 – *Structure d'un HMM.*

4.1.3 Alignement automatique de la parole

Les modèles décrits précédemment sont par la suite utilisés pour un alignement contraint par le texte des enregistrements de parole des corpus *VML*, *DesPhoAPady* et *TypALoc*.
Un alignement contraint par le texte consiste à trouver pour chaque phonème (issu de la transcription du texte produit par le locuteur) ses frontières de début et de fin dans le signal de parole. La figure 4.2 résume les étapes de ce processus.

L'outil d'alignement prend comme entrées :

— la transcription du texte prononcé par le locuteur (dans le cas de parole lue ou spontanée). Cette transcription est réalisée en respectant le protocole décrit dans la section 3.2 pour l'annotation des substitutions, délétions et insertions ;
— un lexique phonétisé contenant tous les mots présents dans le texte avec différentes variantes phonologiques, basées sur l'ensemble des 37 phonèmes de la langue française ;
— les modèles HMM de parole appris sur les corpus d'apprentissage *Ester* et *Bref* ;
— les vecteurs de paramétrisation PLP du signal à aligner.

Cependant, et compte tenu de la large variabilité de la parole dysarthrique et les productions atypiques qui y sont liées, les modèles appris sur de la parole normale feront face à d'importantes difficultés dans la reconnaissance et la délimitation des frontières

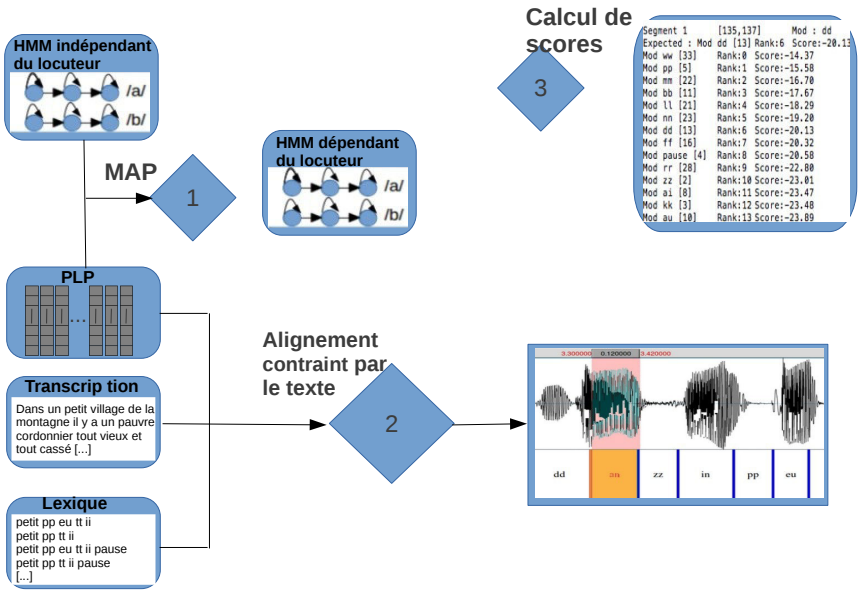


FIGURE 4.2 – *Alignement automatique de la parole contraint par le texte*

des phonèmes. Afin de remédier à ce problème et avoir le meilleur alignement possible, ces modèles seront adaptés à chaque locuteur à travers une adaptation de type Maximum à Posteriori (MAP) (Gauvain et Lee, 1994) à trois itérations résultant en des modèles dépendants du locuteur. Cette adaptation permet de modifier les paramètres du HMM pour le rapprocher de la production spécifique de chaque locuteur.

Le processus d'alignement lui même est basé sur un algorithme de décodage Viterbi (Viterbi, 1967). Cet algorithme permet de déterminer l'alignement optimal de la séquence de phonèmes donnée en entrée par rapport aux modèles HMM.

Suite à cet alignement, chaque segment sera associer à des scores caractérisant sa normalité et son rapprochement du modèle du phonème auquel il est associé. Ces scores seront utilisés par la suite dans le chapitre 5 dans le cadre de l'approche de détection automatique de phonèmes anormaux dans la parole dysarthrique. En effet, une fois les frontières de chaque phonème retrouvées, et en utilisant les modèles HMM initiaux indépendant du locuteur HHM-Ester-H et HHM-Ester-F, des mesures de vraisemblances sont calculées pour chaque segment w_p associé au phonème p avec l'ensemble des modèles HMM, pris indépendamment les uns des autres. Ces scores sont donnés par :

$$L_p^{p'} = \frac{log(P(w_p|p'))}{T_{w_p}} \tag{4.5}$$

où T_{w_p} est la taille en trames du segment w_p et p' est un phonème du français représenté dans notre modèle.

Tous ces scores de vraisemblance sont ensuite ordonnés afin d'indiquer pour chaque

segment les phonèmes les plus et les moins probables d'y être associés. La figure 4.3 donne l'exemple des différents scores de vraisemblance mesurés sur deux segments. Dans un des cas, le phonème associé au segment dans la transcription a eu le meilleur score de vraisemblance (figure à gauche), dans le deuxième, un autre phonème a obtenue le meilleur score (on parle alors de confusion dans l'alignement).

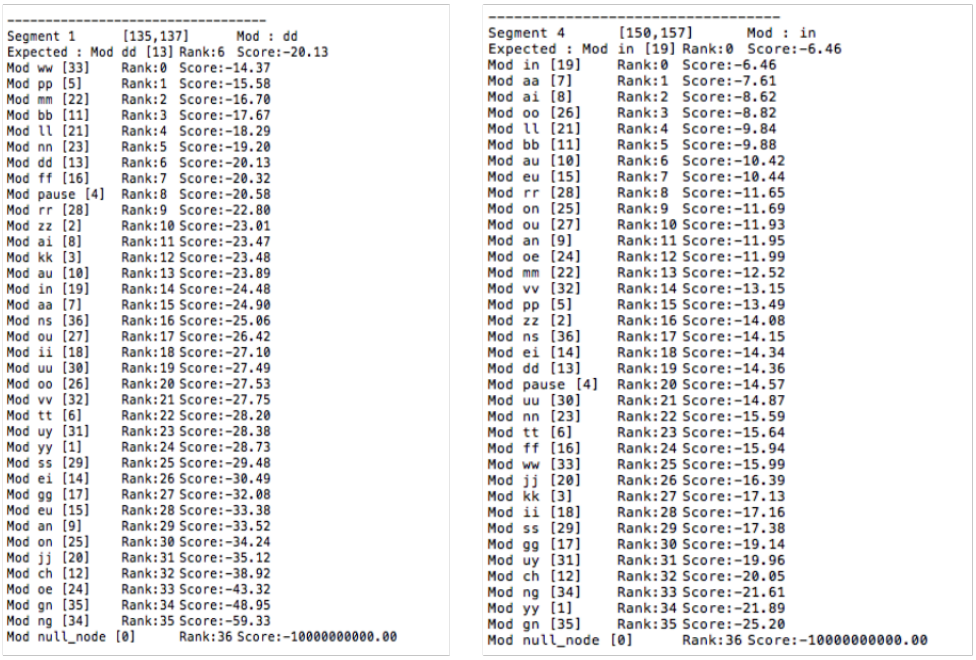


FIGURE 4.3 – *Scores de vraisemblance issues de l'alignement automatique de la parole*

4.2 Étude du comportement du système d'alignement face à la parole dysarthrique

Cette section permettra la description et l'analyse du comportement de l'outil d'alignement automatique de la parole face aux différentes pathologies et styles de parole disponibles dans nos corpus. Afin de pouvoir analyser et évaluer la qualité de l'alignement, les différentes mesures d'évaluation décrites dans la partie 3.3.1 seront utilisées (décalage au début (DD), décalage au milieu (DM) et différence de durées ($DDur$)). Puisque ces mesures sont calculées par rapport à un alignement manuel de référence réalisé par un expert humain, ces analyses ne porteront que sur les corpus *VML* et *TypALoc*, les seuls pour lesquels nous disposons de telles références.

4.2.1 Parole lue

Nos premières analyses de l'alignement automatique ont porté sur la parole lue. Ce choix repose sur l'hypothèse que ce style de parole, se réalisant dans un contexte plus contrôlé, posera moins de difficulté à l'outil d'alignement. De plus, puisque tous les locuteurs patients et contrôles ont lu le même texte, cette étude permettra de comparer, dans des conditions semblables, les comportements des différentes populations et pathologies présentes dans nos corpus.

Parole normale et parole dysarthrique

Le tableau 4.2 détaille les taux de phonèmes avec des mesures de décalage en début (DD), au milieu (DM) et de différence de durées ($DDur$) en dehors de l'intervalle $\pm$20ms pour les contrôles ainsi que les patients issus des corpus *VML* et *TypALoc* regroupés par grade de sévérité [2] de la dysarthrie pour notre outil d'alignement ainsi que l'outil d'alignement EasyAlign (Goldman, 2011).

TABLE 4.2 – *Taux moyen et écart-type (σ) de phonèmes avec des mesures de décalage en début (DD), au milieu (DM) et de différences de durées ($DDur$) $\notin \pm$20ms pour les contrôles ainsi que les patients issus des corpus VML et TypALoc regroupés par grade de sévérité de la dysarthrie en utilisant notre outil d'alignement (LIA) et l'outil d'alignement EasyAlign.*

	EasyAlign			LIA		
	DD	DM	$DDur$	DD	DM	$DDur$
Contôles	17.6 (4.0)	18.0 (3.4)	27.5 (4.0)	15.3 (1.6)	12.9 (1.1)	25.5 (3.0)
Grade de sévérité 1	31.2 (7.6)	32.0 (8.7)	42.5 (9.6)	23.5 (6.0)	21.5 (6.6)	35.2 (8.3)
Grade de sévérité 2	48.8 (10.4)	52.7 (12.0)	65.8 (10.3)	36.0 (8.1)	38.6 (9.8)	53.4 (8.6)
Grade de sévérité 3	73.9 (10.4)	78.2 (11.7)	86.2 (6.4)	48.8 (9.2)	57.9 (11.3)	72.5 (8.3)

La comparaison des taux d'erreurs d'alignement des deux systèmes montre que l'outil d'alignement que nous utilisons (LIA) présente un meilleur alignement que EasyAlign sur toutes les populations étudiées. En effet, sur les locuteurs contrôles, les taux d'erreurs d'alignement selon DD sont de 15.3% pour LIA contre 17.6% pour EasyAlign. La différence est d'autant plus importante sur les patients atteints de dysarthrie sévère (grade 3) où le taux d'erreurs atteint 74% pour EasyAlign contre "seulement" 49% pour notre outil. Ces différences peuvent être expliquées par l'importante quantité de données utilisées lors l'apprentissage de nos modèles (corpus *Bref* et *Ester* d'environ 200h) contre l'utilisation de seulement 30min de parole lors de l'apprentissage du modèle de EasyAlign. De plus, des modèles dépendants du genre ont été utilisés au sein de notre outil contre un modèle unique dans le cas de EasyAlign. Il est important de noter que cette comparaison n'est pas, en soi, le but de notre analyse mais vise seulement à appuyer la robustesse et la pertinence des observations tirées du comportement de notre

2. Les grades de sévérité ont été établis en concertation avec un phonéticien : (1) les patients caractérisés par un degré de sévérité perceptif $\leq$1.5 sont associés au grade 1 (2) ceux caractérisés par un degré de sévérité perceptif $\leq$2.5 sont associés au grade 2 (3) ceux caractérisé par un degré de sévérité $>$ 2.5 sont associé au grade 3.

outil d'alignement automatique de la parole sur la parole dysarthrique et présentées par la suite.

Considérons seulement l'outil d'alignement LIA, nous observons déjà qu'un meilleur alignement est réalisé sur les locuteurs contôles par rapport aux patients dysarthriques. En effet, seulement 15% et 13% des phonèmes produits par les locuteurs contôles présentent des mesures de DD et DM en dehors de $\pm 20ms$ respectivement. Ce premier comportement est tout de même attendu puisque la production de ces locuteurs se rapproche le plus de la parole "standard" sur laquelle les modèles HMM ont été appris.

À l'inverse, les patients présentent beaucoup plus d'erreurs d'alignement (décalages $\notin \pm 20$ms) avec des taux de DD de 24%, 36% et 49% pour les grades de sévérité 1, 2 et 3 respectivement. Une première tendance émerge de ces mesures : plus la dysarthrie est sévère, plus la qualité de l'alignement se dégrade atteignant presque un phonème correctement aligné sur deux pour les patients atteints de dysarthrie sévère. La même tendance est observée sur les taux de DM et $DDur \notin \pm 20$ms qui augmentent chez les patients atteints de dysarthries sévères atteignant 58% et 73% respectivement contre 13% et 26% pour les locuteurs contôles. Il est aussi intéressant de relever que les écart-types augmentent au sein du groupe de grade de sévérité élevé par rapport au grade inférieur et aux contôles (sauf pour la durée où ils sont plus stables). Cela reflète la large variabilité et altérations existantes dans la parole dysarthrique par rapport à la parole normale.

Variabilité inter-pathologique dans la parole dysarthrique

Le tableau 4.3 détaille les taux d'erreurs d'alignement selon les décalages DD, DM et $DDur$ pour les différentes populations des corpus *VML* et *TypALoc*. Des différences importantes peuvent être observées entre les pathologies dysarthriques étudiées. En effet, dans l'ensemble, les patients atteints de dysarthrie parkinsonienne sont ceux présentant les taux d'erreurs d'alignement les plus faibles. Cela peut être attendu puisqu'il s'agit de la population la moins dysarthrique (degré de sévérité perceptive moyen de 0.9). Les taux d'erreurs d'alignement, sur la base de la mesure DD, sont plus élevés sur les autres populations atteignant 29%, 33% et 35% pour les patients atteints respectivement d'ataxie cérébelleuse, de maladies lysosomales et de SLA. Ces taux d'erreurs bien qu'attendus pour la SLA, population la plus dysarthrique (degré de sévérité moyen de 2.0), le sont moins pour l'ataxie cérébelleuse (degré de sévérité moyen 1.3). Afin d'être

TABLE 4.3 – *Taux moyen et écart-type (σ) d'erreurs d'alignement selon les décalages DD, DM et DDur calculés sur les différentes populations des corpus VML et TypALoc.*

Population	DD (σ)	DM (σ)	$DDur$ (σ)
Contôles	15.3 (1.6)	12.9 (1.1)	25.5 (3.0)
Park.	20.1 (5.7)	17.8 (6.1)	29.2 (6.9)
AC	28.9 (4.2)	27.8 (5.4)	43.6 (4.4)
SLA	34.9 (10.4)	37.1 (12.8)	52.2 (12.6)
Lysos.	32.5 (12.9)	34.7 (18.5)	48.5(17.8)

en mesure de mieux comparer les comportements de ces populations, le tableau 4.4 présente les taux d'erreurs d'alignement calculés sur les patients appartenant au grade de sévérité perceptive dysarthrique 1 seulement. Cet ensemble renferme tous les patients atteints de la maladie de Parkinson, 7 patients atteints d'ataxie cérébelleuse, 3 patients atteints de SLA et 3 patients atteints de maladies lysosomales.

TABLE 4.4 – *Degré de sévérité, débit de parole et taux moyen d'erreurs (écart-type) d'alignement selon les décalages DD, DM et DDur pour les contôles et les patients de garde de sévérité 1 regroupés par pathologie.*

Population	Contôles	Park.	AC	SLA	Lysos.
Sévérité (σ)	-	0.9 (0.4)	1.2 (0.4)	1.3 (0.3)	1.5 (0.0)
Débit (σ)	-	0.5 (0.5)	-1.1 (0.6)	0.3 (0.8)	1.5 (1.1)
DD (σ)	15.3 (1.6)	20.1 (5.7)	27.8 (3.3)	25.9 (5.3)	20.2 (4.8)
DM (σ)	12.9 (1.1)	17.8 (6.1)	26.2 (3.5)	22.8 (5.3)	18.8 (7.3)
$DDur$ (σ)	25.5 (3.0)	29.2 (6.9)	42.7 (3.9)	37.0 (6.7)	31.7 (5.7)

En observant les mesures rapportées dans ce tableau, la dysarthrie ataxique se démarque comme étant celle présentant le plus d'erreurs pour les différentes mesures atteignant des taux d'erreurs d'alignement de 28%, 26% et 43% sur DD, DM et $DDur$ respectivement. Ce comportement est d'autant plus intéressant que cette population ne présente pas le degré de sévérité le plus élevé (1.2 contre 1.3 et 1.5 pour les SLA et les maladies lysosomales respectivement). Les patients atteints de maladies lysosomales présentent quant à eux un faible taux d'erreurs comparable à celui observé sur la maladie de Parkinson (20% et 19% sur DD et DM respectivement) malgré la différence importante de sévérité de dysarthrie entre les deux populations.

TABLE 4.5 – *Degré de sévérité, débit de parole et durée moyenne des phonèmes issues de l'alignement manuelle (Durée-M) et de l'alignement automatique (Durée-A) pour les contôles et les patients de garde de sévérité 1 regroupés par pathologie.*

Population	Contôles	Park.	AC	SLA	Lysos.
Sévérité (σ)	-	0.9 (0.4)	1.2 (0.4)	1.3 (0.3)	1.5 (0.0)
Débit (σ)	-	0.5 (0.5)	-1.1 (0.6)	0.3 (0.8)	1.5 (1.1)
Durée-M (σ)	90.4 7.6)	89.0 (12.4)	118.9 (13.5)	98.6 (9.8)	93.2 (11.7)
Durée-A (σ)	86.8 (6.5)	85.7 (11.8)	112.5 (13.4)	95.1 (7.1)	87.5 (8.6)

Une des hypothèses avancées pour expliquer ces tendances est le lien entre la qualité de l'alignement et le débit de la parole. Le tableau 4.5 donne pour les patients appartenant au grade de sévérité perceptive dysarthrique 1 la durée moyenne des phonèmes issus de l'alignement manuel (Durée-M) et de l'alignement automatique (Durée-A) ainsi que l'évaluation perceptive de sévérité et du débit de parole. En effet, les deux populations dysarthriques présentant les taux d'erreurs les moins importants sont celles caractérisées par un débit de parole rapide (0.5 et 1.5 pour Park. et Lysos. respectivement). Ce débit rapide se traduit par une durée moyenne des phonèmes plus courte pour ces deux populations (89ms et 93ms respectivement). Inversement, la dysarthrie ataxique, caractérisée par un débit de parole plus lent et une durée de phonèmes plus longue

(118.9ms), présente le taux d'erreurs d'alignement le plus élevé. La figure 4.4 donne la distribution des erreurs d'alignement exprimées en décalage de début (DD) et décalage au milieu (DM) pour les différentes populations des corpus *VML* et *TypALoc*.

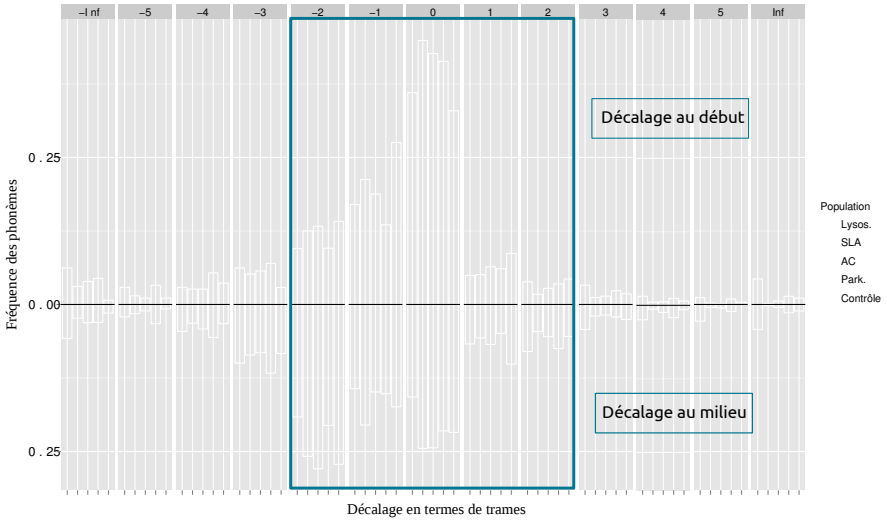


FIGURE 4.4 – *Distribution des erreurs d'alignement exprimées en décalage de début et décalage au milieu pour les différentes populations des corpus VML et TypALoc.*

Ces hypothèses sont confirmées par le calcul de la corrélation de Pearson entre les taux d'erreurs d'alignement et l'évaluation perceptive de la sévérité et du débit de la parole des patients affichés dans le tableau 4.6. Effectivement, ces mesures varient entre 0.82 et 0.85 pour la sévérité et -0.73 et -0.82 pour le débit de parole. La qualité de l'alignement est donc dépendante à la fois de la sévérité (plus la dysarthrie est sévère moins l'alignement est précis) et du débit de la parole (plus la parole est lente moins l'alignement est précis).

TABLE 4.6 – *Corrélation entre les taux d'erreurs d'alignement et les mesures de sévérité et de débit de parole de tous le patients des corpus VML et TypALoc.*

	DD	DM	$DDur$
Degré de sévérité	0.82	0.85	0.85
Débit de parole	-0.73	-0.74	-0.82

Variabilité phonétique

Nous avons aussi étudié les différences observées dans l'alignement automatique entre les différentes catégories phonétiques. Le tableau 4.7 donne pour les contôles et les patients de grade de sévérité 1 les taux moyens d'erreurs d'alignement par catégorie phonétique.

Observant les taux d'erreurs d'alignement pour les contôle, nous remarquons que

TABLE 4.7 – *Taux moyens d'erreurs d'alignement (écart-type) par catégorie phonétique pour les contôles et les patients de grade de sévérité 1 des corpus VML et TypALoc.*

Catégorie	Mesures	Contrôles	Park.	CA	SLA	Lysos.
Occlusives sourdes	DD	16.5 (6.5)	23.2 (6.4)	27.2 (4.5)	23.2 (4.5)	20.9 (5.9)
	DM	22.8 (6.6)	21.7 (6.8)	30.9 (6.0)	20.0 (7.0)	18.0 (4.7)
	$DDur$	32.8 (6.3)	29.9 (8.8)	44.6 (5.5)	34.2 (4.0)	33.1 (7.1)
Occlusives sonores	DD	11.3 (5.5)	17.2 (10.1)	26.4 (6.9)	29.5 (14.0)	27.2 (13.0)
	DM	7.5 (5.0)	9.2 (7.6)	16.9 (7.5)	24.1 (10.0)	21.5 (11.6)
	$DDur$	18.0 (4.6)	22.2 (13.5)	30.2 (11.5)	35.4 (9.7)	25.6 (10.9)
Voyelles orales	DD	14.3 (2.7)	15.3 (6.1)	24.2 (4.2)	19.0 (2.4)	15.6 (4.4)
	DM	10.1 (1.2)	14.4 (6.8)	22.8 (3.1)	20.5 (4.0)	17.0 (9.3)
	$DDur$	23.8 (3.2)	28.0 (7.5)	39.8 (4.7)	38.5 (3.5)	28.2 (4.6)
Voyelles nasales	DD	4.9 (4.4)	13.4 (10.8)	19.6 (4.7)	14.5 (2.2)	11.3 (6.7)
	DM	9.2 (6.6)	12.4 (7.8)	21.6 (6.1)	19.3 (9.7)	10.4 (3.1)
	$DDur$	26.3 (9.5)	27.1 (12.2)	40.8 (9.6)	37.8 (10.2)	32.0 (8.1)
Fricatives sourdes	DD	20.4 (8.7)	28.6 (13.4)	47.8 (14.5)	34.9 (6.7)	43.2 (12.1)
	DM	28.9 (7.1)	34.0 (9.2)	44.7 (15.6)	31.3 (6.8)	36.3 (14.0)
	$DDur$	25.9 (7.0)	44.1 (10.1)	71.1 (16.8)	34.9 (7.2)	53.7 (20.8)
Fricatives sonores	DD	17.2 (3.0)	24.0 (6.8)	29.8 (5.3)	36.7 (11.0)	24.5 (6.0)
	DM	9.0 (2.3)	21.4 (9.0)	28.7 (5.6)	26.8 (7.0)	22.5 (8.2)
	$DDur$	24.6 (5.0)	28.3 (6.5)	42.1 (6.2)	36.3 (12.3)	32.6 (5.2)
Semi-consonnes	DD	21.7 (5.1)	34.6 (11.2)	43.0 (10.1)	42.1 (6.3)	0.0 (0.0)
	DM	18.8 (6.5)	19.6 (6.0)	30.6 (9.1)	28.8 (8.7)	0.0 (0.0)
	$DDur$	30.3 (6.4)	34.1 (13.4)	50.5 (8.8)	45.2 (8.8)	0.0 (0.0)
Consonnes nasales	DD	18.6 (6.3)	22.6 (12.5)	29.9 (16.5)	24.0 (12.9)	5.6 (7.9)
	DM	11.3 (5.5)	10.6 (11.5)	19.3 (7.1)	19.0 (11.7)	2.8 (3.9)
	$DDur$	20.6 (9.1)	32.4 (17.5)	47.7 (16.5)	32.6 (11.3)	24.7 (5.9)

les voyelles présentent généralement les alignements les plus précis (5% et 14% d'erreur sur DD pour les voyelles nasales et orales respectivement). On remarque aussi que les occlusives et les fricatives sourdes présentent plus d'erreurs que les occlusives et les fricatives sonores. Plus précisément, les fricatives sourdes sont la catégorie posant le plus de difficulté au système avec des taux d'erreurs de 20% et 29% respectivement sur DD et DM. Il s'agit, avec la catégorie des semi-consonnes, des phonèmes les moins bien alignés par le système. Il est tout de même intéressant de noter que les phonéticiens ont aussi exprimé plus de difficultés lors de la segmentation manuelle de ces consonnes sourdes. Le comportement de l'outil face à cette dernière catégorie peut être expliqué par ses caractéristiques acoustiques assez particulières et sa fréquence d'apparition moins importante dans le français ce qui implique que les modèles de ces phonèmes ont été appris sur peu de données.

Observant maintenant le comportement du système sur la parole dysarthrique, nous remarquons que la dysarthrie ataxique se caractérise par un taux d'erreurs d'alignement plus important que les autres populations sur les voyelles atteignant 24% sur les

voyelles orales selon la mesure DD. Ces erreurs peuvent être le résultat des distorsions des voyelles caractérisant ce type de dysarthrie (tableau 2.1). Par ailleurs, cela peut être lié une nouvelle fois à l'effet du débit lent caractérisant cette dysarthrie et affectant plus les voyelles que les autres phonèmes.

Contrairement aux autres populations, les patients atteints de SLA et de maladies lysosomales (dysarthrie mixte) présentent moins d'erreurs d'alignement sur les occlusives sourdes que sur les occlusives sonores. Ce comportement peut être le résultat du phénomène de voisement/dévoisement subi par ces phonèmes et résultant en une amélioration de l'alignement des occlusives sourdes et une détérioration de l'alignement des occlusives sonores. Ce comportement apparaît aussi (de manière plus nuancée) sur les fricatives chez la population SLA où l'écart entre les taux d'erreurs d'alignements sur les fricatives sourdes et sonores est plus faible que celui observé au niveau de la dysarthrie ataxique. En comparant ces deux populations, on remarque qu'elles affichent des tendances opposées. En effet, les occlusives et fricatives sourdes présentent moins d'erreurs d'alignement chez les SLA que chez les patients atteints d'ataxie cérébelleuse alors que l'inverse est observé sur les occlusives et fricatives sonores.

Nous remarquons aussi que les consonnes nasales présentent un comportement particulier chez les patients atteints de maladies lysosomales avec des taux d'erreurs d'alignement plus faible que ceux observés sur les autres pathologies ainsi que sur les contôles (6% contre 19% respectivement sur DD). Une tendance similaire peut être observée sur les voyelles nasales où moins d'erreurs sont réalisées sur ces patients que sur toutes les autres populations dysarthriques malgré un degré de sévérité moyen plus important. Ce comportement peut être lié au caractère d'hypernasalité associé à la dysarthrie mixte résultante des maladies lysosomales. Ce comportement n'est cependant pas visible sur les voyelles nasales des patients atteints de SLA associée elle aussi à la dysarthrie mixte.

Cette section nous a permis d'analyser le comportement de l'outil d'alignement automatique sur de la parole dysarthrique lue et son lien avec les caractéristiques des différents types de dysarthries. En effet, un nombre plus important d'erreurs d'alignement ont été observées sur les patients atteints d'ataxies cérébelleuses. Ce comportement peut être lié à la durée plus longue des phonèmes observée chez ces patients, résultant du faible débit de parole les caractérisant. En outre, sur la même population, un important taux d'erreurs d'alignement a été observé sur les voyelles pouvant refléter le critère de distorsion des voyelles caractérisant la dysarthrie ataxique. Par ailleurs, un faible taux d'erreurs d'alignement a été observé sur les voyelles et consonnes nasales chez les patients atteints de maladies lysosomales, ce comportement pouvant être lié à l'hypernasalité souvent associée à la dysarthrie mixte. Finalement, au sein de chaque population, un effet de la sévérité de la dysarthrie a été relevé amenant à davantage d'erreurs d'alignement observées sur les patients atteints de dysarthrie sévère.

4.2.2 Parole spontanée

Dans un deuxième temps, nous nous sommes proposés d'étudier le comportement de l'outil d'alignement face à la parole dysarthrique spontanée. En effet, et compte tenu de la plus grande variabilité ainsi que des conditions moins contrôlées liées à ce style de parole, nous nous attendions à ce que plus d'erreurs d'alignement soient réalisées sur cette parole. Cette partie de l'étude porte uniquement sur le corpus *TypALoc* puisqu'il s'agit du seul corpus comportant de la parole spontanée alignée manuellement

Le tableau 4.8 présente les mesures de corrélation de Pearson entre les taux d'erreurs d'alignement et l'évaluation perceptive de la sévérité et du débit de la parole des patients. Nous remarquons que le comportement de l'outil suit la même tendance observée sur la parole lue. Les taux d'erreurs d'alignement selon les mesures DD, DM, et $DDur$ sont toujours corrélés à la fois à la sévérité de la dysarthrie (entre 0.72 et 0.75) et au débit de la parole (entre -0.74 et -0.82). On note néanmoins une corrélation mois importante sur la sévérité de la dysarthrie sur la parole spontanée par rapport à la lecture.

TABLE 4.8 – *Corrélation entre les taux d'erreurs d'alignement et les mesures de sévérité et de débit de parole spontanée des patients du corpus TypALoc.*

	DD	DM	$DDur$
Degré de sévérité	0.72	0.77	0.75
Débit de parole	-0.74	-0.75	-0.82

Afin de pouvoir comparer au mieux les différentes pathologies, le tableau 4.9 détaille les taux d'erreurs d'alignement calculés sur les patients appartenant au grade de sévérité perceptive de la dysarthrie 1 seulement. Les durées moyennes des phonèmes issus de l'alignement manuel et de l'alignement automatique ainsi que l'évaluation perceptive de sévérité et du débit de parole sont indiqués.

TABLE 4.9 – *Degré de sévérité, débit de parole, durée moyenne des phonèmes issus de l'alignement manuel (Durée-M) et de l'alignement automatique (Durée-A) et taux moyen (écart-type) d'erreurs d'alignement selon les décalages DD, DM et DDur pour la parole spontanée des contôles et des patients de grade de sévérité 1 du corpus TypALoc.*

Population	Contôles	Park.	AC	SLA
Sévérité(σ)	-	1.0 (0.4)	1.1 (0.4)	1.4 (0.2)
Débit (σ)	-	-0.1 (0.9)	-0.8 (0.5)	0.5 (0.4)
Durée-M (σ)	84.1 (9.8)	84.4 (13.5)	116.5 (17.3)	85.3 (12.2)
Durée-A (σ)	81.3 (10.1)	82.1 (14.1)	111.8 (17.5)	85.3 (13.0)
DD (σ)	14.9 (2.6)	17.3 (6.4)	24.2 (4.7)	14.8 (3.0)
DM (σ)	13.3 (2.7)	16.7 (6.9)	22.7 (4.8)	15.8 (3.6)
$DDur$ (σ)	21.8 (3.0)	24.9 (8.3)	35.0 (7.7)	21.3 (4.0)

Nous observons que les contôles sont toujours la population présentant le moins d'erreurs d'alignement (15% et 13% sur DD et DM respectivement). Cependant, les patients atteints de SLA présentent des taux d'erreurs d'alignement similaires à ceux observés chez les contrôles et inférieurs à ceux observés chez les autres pathologies

dysarthriques malgré le degré de sévérité plus important sur ces patients. Bien que ce comportement peut être associé au débit de parole supérieur (0.5) sur cette population par rapport aux patients atteints de maladie de Parkinson (-0.1) et d'ataxie cérébelleuse (-0.8), cela n'explique pas le fait qu'ils sont au même niveau que ceux mesurés sur la parole normale (contôles).

4.2.3 Parole lue et parole spontanée

Le tableau 4.10 reprend les taux moyens d'erreurs d'alignement sur les différentes populations du corpus *TypALoc*. En observant les taux d'erreurs sur les deux styles de parole, des comportements inattendus apparaissent. En effet, moins d'erreurs d'alignement sont retrouvées sur la parole spontanée que sur la parole lue pour la majorité des populations étudiées. Ceci est contraire à notre hypothèse de départ qui s'appuyait sur la plus large variabilité généralement observée sur ce style de parole (faux départ, hésitations, réductions, productions non standards de quelques phonèmes dans des contextes particuliers, etc.). Une hypothèse permettant d'expliquer ces tendances est la nature du corpus de données *Ester* utilisé pour l'apprentissage des modèles HMMs issu d'enregistrements radiophoniques et s'approchant donc plus de la parole spontanée que de la lecture.

TABLE *4.10* – *Taux d'erreurs d'alignement en termes de DD et DM pour les différentes populations et grades de sévérité de la dysarthrie du corpus TypALoc.*

Population	Parole spontanée		Parole lue	
	DD	*DM*	*DD*	*DM*
Contôles	14.9 (2.6)	13.3 (2.7)	15.3 (1.6)	12.9 (1.1)
Park.	17.3 (6.4)	16.7 (6.9)	20.1 (5.7)	17.8 (6.1)
AC	25.9 (6.2)	24.8 (7.2)	28.9 (4.2)	27.8 (5.4)
SLA	29.5 (13.4)	31.5 (14.2)	34.9 (10.4)	37.1 (12.8)
Park. - Grade 1	17.3 (6.4)	16.7 (6.9)	20.1 (5.7)	17.8 (6.1)
AC- Grade 1	24.2 (4.8)	22.7 (4.8)	27.8 (3.3)	26.2 (3.5)
SLA - Grade 1	14.8 (3.0)	15.8 (3.6)	25.9 (5.3)	22.8 (5.3)
AC - Grade 2	37.4 (-)	39.7 (-)	36.5 (-)	39.0 (-)
SLA - Grade 2	31.1 (11.3)	33.5 (11.6)	35.5 (8.3)	38.9 (9.2)
SLA - Grade 3	45.9 (5.1)	47.9 (7.9)	46.4 (10.6)	52.2 (9.7)

Il est aussi intéressant de remarquer que les différences entre les deux styles de parole sont plus visibles chez les locuteurs dysarthriques que chez les locuteurs sains. L'hypothèse que nous émettons pour expliquer ces différences observées entre les deux styles de parole repose sur la plus grande liberté qu'offre la parole spontanée aux patients. En effet, elle leur permet de mieux contrôler les segments à produire en appliquant d'éventuelles stratégies d'évitement ou de compensation de contextes particuliers jugés plus "difficiles" à produire.

En observant l'évolution des taux d'erreurs d'alignement selon les grades de sévérité de dysarthrie, nous retrouvons que la différence entre la parole lue et la parole

spontanée s'atténue pour les dysarthries plus sévères. Cette tendance est surtout visible sur la population SLA où la différence absolue des taux d'erreurs selon *DD* est de 11.1%, 4.2% et 0.5% pour les grades 1, 2 et 3 respectivement.

4.2.4 Confusion phonémique dans l'alignement automatique de la parole lue

Comme nous l'avons décrit dans la section 4.1, des scores de vraisemblance peuvent être calculés entre un segment associé à un phonème et l'ensemble des modèles HMM représentant les 37 phonèmes du français (et utilisés dans le cadre de l'alignement contraint par le texte (figure 4.3). Ces scores et leur classement traduisent le fait qu'une production d'un phonème par un locuteur dysarthrique peut s'approcher plus du modèle d'un autre phonème du français. Cette section présente une étude préliminaire des confusions phonémiques observées sur nos populations de patients dans le but d'y déceler des liens avec les différentes classes dysarthriques étudiées.

On définit une confusion phonémique quand le meilleur score de vraisemblance mesuré sur un segment correspond à un phonème différent de celui avec lequel il est associé (par exemple, un segment associé au phonème *pp* et dont le meilleur score de vraisemblance est obtenue avec le phonème *bb*).

Le tableau 4.11 fournit les taux de reconnaissance par population mesurés sur les phonèmes bien alignés de la parole lue du corpus *TypALoc*. Considérant les différentes populations, on observe que les contrôles présentent le plus grand nombre de phonèmes bien reconnus (sans confusion) comparés aux patients (81% pour les contrôles contre 67% pour les patients atteints de maladie de Parkinson et d'ataxie cérébelleuse et 58% pour ceux atteints de SLA). Cette variabilité inter-pathologique est vraisemblablement due à la différence de grade de sévérité de la dysarthrie plus important chez les patients atteints de SLA par rapport aux autres pathologies.

TABLE *4.11* – *Taux de reconnaissance pour les phonèmes bien alignés de la parole lue du corpus TypALoc*

Population	Taux de reconnaissance (%)
Contôles	81
Maladie de Parkinson	67
Ataxie cérébelleuse	67
SLA	58

Le tableau 4.12 détaille les taux de reconnaissance et de confusion par population et catégorie phonétique mesurés sur les phonèmes bien alignés de la parole lue du corpus *TypALoc*. En comparant occlusives et fricatives, on trouve que contrairement aux contrôles et SLA pour lesquelles les deux catégories présentent des taux de reconnaissance globaux comparables, l'outil d'alignement réalise beaucoup plus de confusions sur les fricatives que sur les occlusives pour la maladie de Parkinson et l'ataxie cérébelleuse. Les taux de reconnaissance sur les occlusives et fricatives atteignant (70% ;45%)

TABLE 4.12 – *Confusion phonémique (%) par classe phonétique pour les phonèmes bien alignés de la parole lue du corpus TypALoc*

		Taux de confusion					
	Taux de reconnaissance	Occlusives	Fricatives	Consonnes Nasales	Voyelles orales	Voyelles nasales	Autre
Contôles							
Occlusives	84	12	0	0	0	0	3
Fricatives	82	4	12	0	0	0	1
Consonnes nasales	82	3	0	8	0	0	7
Voyelles orales	81	0	0	0	14	3	2
Voyelles nasales	77	0	0	0	8	13	1
Maladie de Parkinson							
Occlusives	70	19	2	1	3	0	6
Fricatives	45	19	16	1	4	0	16
Consonnes nasales	71	6	1	8	6	1	8
Voyelles orales	71	0	0	0	20	6	2
Voyelles nasales	64	1	0	0	17	18	1
Ataxie cérébelleuse							
Occlusives	65	23	1	1	2	0	9
Fricatives	38	22	29	1	1	1	9
Consonnes nasales	83	8	0	1	0	0	7
Voyelles orales	76	0	0	0	18	4	2
Voyelles nasales	58	0	0	1	17	23	1
SLA							
Occlusives	62	14	7	5	3	0	9
Fricatives	61	6	23	2	4	1	3
Consonnes nasales	81	2	0	8	3	1	4
Voyelles orales	61	0	0	1	22	14	2
Voyelles nasales	53	0	0	0	24	20	2

et (65% ;38%) respectivement. Il est aussi intéressant de constater qu'une large majorité des confusions observées sur les fricatives pour les contrôles et même la SLA sont avec d'autres fricatives contrairement à la maladie de Parkinson où il y a plus de confusions de fricatives avec des occlusives qu'avec d'autres fricatives.

Considérant les voyelles, et à l'exception de la SLA, les voyelles orales sont généralement bien reconnues pour toutes les populations atteignant des taux de reconnaissance de 81%, 71% et 76% pour les locuteurs contôles, parkinsoniens et atteints de dysarthrie ataxique respectivement. De plus, nous remarquons que le taux de confusion de ces voyelles avec les voyelles nasales ne dépassent pas 6% des phonèmes pour toutes ces populations.

Par contre, dans le cas des patients atteints de SLA, le taux de reconnaissance sur les voyelles orales est de 61% seulement avec 14% de ces voyelles confondues avec des voyelles nasales. Cela peut s'expliquer par le phénomène d'hypernasalisation caractérisant la dysarthrie mixte liée à cette pathologie (Auzou, 2007a). Il est intéressant tout de même de noter que l'effet de l'hypernasalisation sur cette population était moins important dans l'étude portant sur les taux d'erreurs d'alignement (tableau 4.7) incluant

tous les phonèmes et non seulement ceux bien alignés.

Cette tendance est confirmée en observant les confusions faites par le système sur les occlusives sonores pour les différentes populations. En effet, 19% des phonèmes de cette classe sont confondues avec des consonnes nasales pour les SLA alors que ce taux est nul pour les contrôles et ne dépasse pas 4% pour les autres dysarthries.

4.3 Conclusion

Cette section a été consacrée à la présentation de l'outil d'alignement automatique de la parole ainsi que son comportement face à la parole dysarthrique. Cette étude a permis d'observer l'effet de la sévérité de la dysarthrie sur la qualité de l'alignement. Plus d'erreurs d'alignement ont été observées chez les patients les plus dysarthriques. Une corrélation importante entre la qualité de l'alignement et le débit de la parole a également été relevée. De plus, des différences entre les classes dysarthriques ont aussi pu être observées : plus d'erreurs ont été faites sur les patients atteints d'ataxie cérébelleuse et de SLA par rapport aux patients parkinsoniens. La répartition de ces erreurs d'alignement selon les différentes catégories phonétiques a révélé des comportements propres à chaque dysarthrie. Plus d'erreurs ont été observées sur les voyelles pour la dysarthrie ataxique et sur les fricatives pour la dysarthrie parkinsonienne.

Finalement, un effet intéressant du style de la parole a pu être observé. En effet, des taux d'erreurs d'alignement plus faibles ont été observés sur la parole spontanée par rapport à la parole lue. Ce comportement, observé sur toutes les populations dysarthriques, était plus marqué chez les patients atteints de SLA de sévérité légère. Ce comportement peut être lié à la liberté qu'offre le style spontané aux patients pour choisir les mots et contextes à produire et compenser ainsi leurs difficultés à produire de la parole.

Dans le chapitre suivant, nous proposons une approche pour la détection automatique de phonèmes anormaux dans la parole dysarthrique. Cette approche reposera sur l'outil automatique d'alignement présenté ici pour la caractérisation de chaque phonème.

Chapitre 5

Détection automatique d'anomalies au niveau phonème

Sommaire

Dans ce chapitre, nous décrivons notre approche pour la détection automatique des anomalies dans la parole dysarthrique au niveau phonème. Cette approche repose sur deux phases essentielles. La première est l'alignement automatique de la parole décrit dans le chapitre précédent. Nous présentons dans ce chapitre la deuxième phase de l'approche qui consiste en une classification des phonèmes en deux catégories : normal et anormal (anomalie). Cette approche a été appliquée sur les corpus *VML*, *DesPhoAPaDy* et *TypALoc* décrits dans le chapitre 3. Nous fournirons une analyse du comportement du système sur ces corpus ainsi qu'une comparaison entre son comportement face à la parole lue et la parole spontanée. De plus, une étude sur la détection des anomalies en fonction de la qualité et la précision de l'alignement automatique sera présentée.

Finalement, les résultats d'une étude préliminaire sur la localisation des anomalies dans les mots bisyllabiques seront également évoqués.

5.1 Approche de détection automatique d'anomalies

Un des buts de ce travail est la proposition d'une approche de détection automatique de phonèmes anormaux dans la parole pathologique. L'approche proposée repose sur une classification binaire de chaque phonème en normal ou anormal. Pour effectuer cette tâche, un modèle représentant la normalité et un second représentant les anomalies seront estimés. La classification permettra de labelliser et d'affecter chaque phonème à une des deux classes.

Avant la classification, une première phase d'extraction de paramètres caractérisant chaque phonème est nécessaire.

5.1.1 Extraction de paramètres

Les paramètres utilisés dans ce processus sont issus des deux alignements successifs, contraint et non contraint par le texte, réalisés sur la parole. Pour chaque phonème, ces paramètres reflètent sa ressemblance à son modèle (dans notre cas, un modèle HMM), ainsi qu'aux modèles des autres phonèmes (Laaridh et al., 2015). Pour chaque segment y_p associé au phonème p, les paramètres suivants sont extraits :

- la durée du phonème, exprimée en termes de nombre de trames de 10ms ;
- le nombre de trames pour lesquelles le meilleur état du HMM reconnu correspond au phonème p ;
- le score de vraisemblance avec p ;
- le rang de p par rapport à tous les autres phonèmes lors du deuxième alignement non contraint par le texte ;
- le score de vraisemblance avec le phonème p', reconnu comme le meilleur correspondant au segment y_p lors du deuxième alignement non contraint par le texte (si p est le meilleur phonème reconnu, le score du second phonème est considéré) ;
- la catégorie phonétique de p' ;
- le score de vraisemblance avec le phonème p'', reconnu comme le second meilleur phonème correspondant au segment y_p lors du deuxième alignement non contraint par le texte (si p est l'un des deux meilleurs phonèmes reconnus, le score du troisième phonème est considéré) ;
- la catégorie phonétique de p''.

Ces paramètres seront extraits sur tous les enregistrements du corpus *VML* et seront utilisés lors de la phase d'apprentissage de nos modèles de phonèmes normaux et anormaux.

5.1.2 Classification

Lors de la phase de modélisation des deux classes, nous avions le choix entre plusieurs méthodes d'apprentissage et classification supervisées. Nous avons opté pour la modélisation à base de SVM (Support Vector Machine). Cette approche d'apprentissage

automatique a été largement appliquée dans le cadre des problèmes de reconnaissance de formes. Elle est basée sur la recherche de la surface qui permet la meilleure séparation d'un ensemble de données dans différentes classes de façon à ce que les marges entre ces dernières soient maximales (Vapnik, 1995; Scholkopf et Smola, 2001).

Dans notre cadre, on dispose de deux classes : les phonèmes normaux et les anomalies. Cette approche est alors appliquée pour classer tout phonème dans une de ces deux classes. Puisque le corpus *VML* est le seul corpus pour lequel nous disposons d'une annotation par un expert des anomalies, il sera utilisé lors de la phase d'apprentissage des modèles. Ces anomalies sont utilisées pour modéliser la classe des phonèmes anormaux, et les phonèmes produits par les locuteurs contrôles pour modéliser la classe de normalité.

L'idée de base des classifieurs SVM est de ramener le problème de classification et de discrimination à un problème de recherche d'un hyperplan optimal. Dans le cadre d'une discrimination binaire, le but est de trouver une fonction de décision associant à chaque observation (phonème dans notre cas) sa classe (normal ou anomalie). Cette fonction séparera l'espace de représentation des données en deux parties : les observations au dessus du plan de séparation appartiennent à une classe, celle en dessous à une autre.

Soit $X = \{x_1, x_2, ..., x_n\}$ l'ensemble des observations existant (phonèmes) et $Y = \{y_1, y_2, ..., y_n\}$ leurs labels correspondant, avec $y_i \in \{1, -1\}$ (1 pour les anomalies, -1 pour les phonèmes normaux). L'hyperplan séparateur est représenté par l'équation suivante :

$$f(x) = wx_i + b \tag{5.1}$$

Si $f(x_i) > 0$, l'observation x_i est classée dans la classe 1, sinon, elle est assignée à la classe -1.

Ce type de problème est dit linéairement séparable lorsqu'il existe une fonction de décision linéaire permettant une juste classification de toutes les données. Néanmoins, plusieurs hyperplans séparateurs peuvent exister (figure (a) 5.1). L'apprentissage de SVM définit alors l'hyperplan optimal comme celui dont la distance avec les observations des deux classes les plus proches est maximale. Ces observations sont notamment appelées vecteurs supports. Cette distance, ou marge, est de l'ordre de $\frac{1}{||w||^2}$. L'hyperplan optimal est alors associé à la solution unique du système d'équations suivant (figure (b) 5.1) :

$$\begin{gathered} \hat{w} = \underset{w}{\operatorname{argmin}}(||w||^2) \\ avec \quad y_i(wx_i + b) \geq 1 \quad pour \quad i \in \{1, .., n\} \end{gathered} \tag{5.2}$$

Cependant, dans la majorité des cas, les données ne sont pas linéairement séparables (figure (c) 5.1). C'est pour prendre en compte cette nature que les fonctions noyaux ont

été introduites. Ces fonctions permettent de projeter les observations d'entrées dans un nouvel espace de grande dimension où la recherche d'un plan optimal peut s'avérer plus fructueuse. Il existe plusieurs fonctions noyaux possibles : linéaire, gaussien, sigmoïde, polynomiale, etc.. Le choix de cette fonction dépend de la nature des données, et quand une visualisation des données est impossible (dimension ≥ 4), le choix de la fonction noyaux est fait empiriquement.

Avec l'introduction des fonctions noyaux, la résolution du système 5.2 définie l'hyperplan optimal comme suivant :

$$f(x) = \sum_{i=1}^{n} \lambda_i y_i K(x, x_i) + b \tag{5.3}$$

où K est la fonction noyau, y_i est la valeur classe cible, λ_i est le multiplieur de Lagrange, n le cardinal des données d'entraînement, b la valeur du biais et x_i le vecteur support.

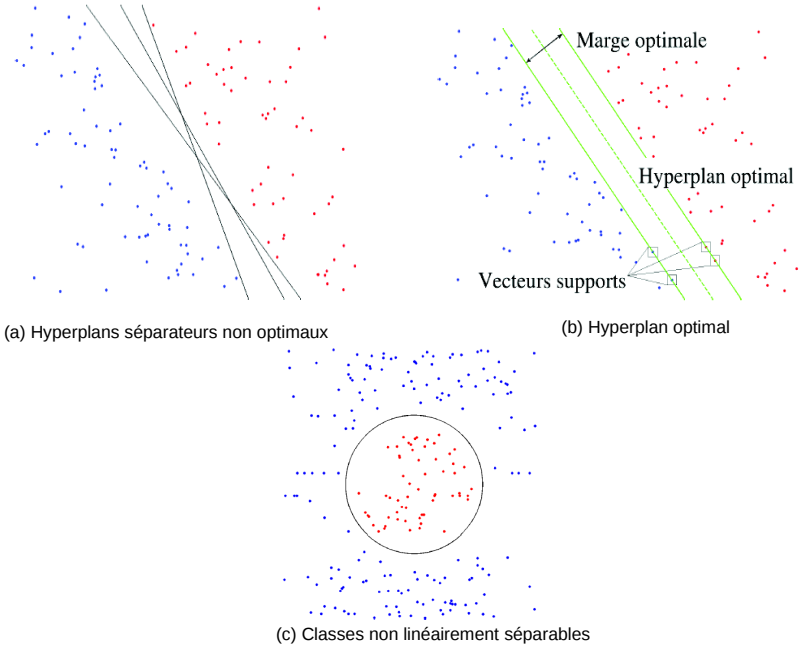


FIGURE 5.1 – *Schéma représentatif du principe d'un SVM, de l'hyperplan optimal et d'un problème non linéairement séparable.*

Dans le cadre de ce travail, tous les SVMs utilisés ont des fonctions noyaux polynomiales. Et puisque les modèles HMM utilisés lors de l'alignement automatique de la parole sont dépendants du genre, deux modèles différents, l'un pour les femmes et l'autre pour les hommes, ont été appris et utilisés pour la classification. De plus, et afin de capturer la large variabilité observée dans la production phonémique, quatre modèles différents ont été appris par genre, reflétant 4 catégories phonétiques essentielles :

les consonnes sourdes, les consonne sonores, les voyelles orales et les voyelles nasales. Ce choix est motivé par la volonté d'investiguer des modèles phonétiques plus fins prenant en compte les spécificités de chaque catégorie phonétique (Ladefoged, 1975). Les différents sous-modèles SVM ont été implémentés en utilisant l'outil SVMlight (Joachims, 1999).

Cependant, un résultat collatéral de ce choix est la quantité limitée de données labellisées comme anomalies disponibles pour chaque sous-modèle appris. Afin d'équilibrer la phase d'apprentissage des modèles, la même quantité de phonèmes normaux et anormaux a été utilisée pour chaque sous-modèle (catégorie phonétique).

Étant donné le nombre limité d'enregistrements disponibles, la technique de validation croisée "leave-one-out" a été utilisée pour chaque enregistrement. Cette technique consiste à retirer à chaque fois un enregistrement et à utiliser tous les autres pour l'apprentissage des modèles. Ces modèles seront par la suite utilisés pour la classification des phonèmes de l'enregistrement exclus lors de la phase d'apprentissage afin d'éviter de biaiser les résultats. Cette technique permet d'apprendre les modèles de classification sur plus de données et d'exploiter en même temps tous les enregistrements annotés par l'expert dans la phase d'évaluation de l'approche.

5.2 Évaluation de l'approche automatique de détection d'anomalies au niveau phonème

Nous présentons dans cette section les résultats de l'application de l'approche proposée sur les 3 corpus *VML*, *DesPhoAPaDy* et *TypALoc*. La méthode d'évaluation de l'approche sur chacun de ces corpus dépendra de la nature des annotations disponibles (annotation d'anomalies au niveau phonème, mesure d'intelligibilité, etc.)

5.2.1 Application sur un corpus annoté au niveau phonème *VML* :

L'évaluation de la qualité de détection automatique des anomalies est faite en comparant les sorties du classifieur avec les annotations de l'expert humain disponible pour ce corpus. Cette comparaison est réalisée selon les deux stratégies décrites dans la section 3.3. Cela permet de calculer les mesures d'évaluation, $AnRappel$ et $AnPrec$ décrites dans la même section. Le tableau 5.1 détaille ces mesures pour le corpus *VML*. Les résultats fournis par locuteur correspondent à la moyenne des valeurs obtenues sur les différents enregistrements longitudinaux disponibles.

À titre de comparaison, les résultats obtenus sur le même corpus par une autre approche automatique de détection d'anomalies reposant sur le même outil d'alignement automatique de parole sont donnés dans le tableau 5.2. Cette approche, nommée par la suite "baseline" s'appuie uniquement sur les scores de vraisemblance calculés sur un modèle de parole normale uniquement pour labelliser chaque phonème comme étant normal ou anormal (Fredouille et Pouchoulin, 2011).

TABLE 5.1 – *Performances de l'approche proposée sur les locuteurs dysarthriques du corpus VML, exprimées en termes de AnRappel et AnPrec, selon les deux stratégies de comparaison utilisées.*

	Stratégie 1		Stratégie 2	
Locuteurs dysarthriques	*AnRappel*	*AnPrec*	*AnRappel*	*AnPrec*
H1	0.15	0.23	0.37	0.52
H2	0.47	0.24	0.90	0.68
H3	0.43	0.32	0.72	0.57
H4	0.54	0.33	0.89	0.65
Moyenne - Hommes	0.40	0.28	0.72	0.61
F1	0.59	0.23	0.85	0.42
F2	0.83	0.68	1.00	0.98
F3	0.44	0.29	0.80	0.58
F4	0.60	0.34	0.90	0.60
Moyenne - Femmes	0.62	0.39	0.89	0.65
Moyenne	0.51	0.34	0.81	0.63

Observant les mesures rapportées dans le tableau 5.1, l'approche proposée arrive à détecter en moyenne 80% des anomalies sur tous les locuteurs en considèrant la deuxième stratégie d'évaluation. Sa précision reste tout de même moyenne avec des taux de *AnPrec* de 0.61 et 0.63 pour les hommes et les femmes respectivement.

Il faut aussi noter que la deuxième stratégie, basée sur la prise en compte d'un éventuel décalage d'un phonème dans l'alignement, présente de meilleures mesures de *AnRappel* et surtout de *AnPrec* que la première stratégie. Ce résultat, même si attendu compte tenu de la nature des mesures utilisées, peut aussi refléter une meilleure considération donnée aux anomalies produites au niveau de la transition d'un phonème à un autre. En effet, même pour les experts humains, la localisation de l'anomalie au niveau phonème reste une tâche critique et difficile à réaliser. De plus, les anomalies produites en début de phonèmes (par exemple lors de l'occlusion pour les occlusives) ou à leurs fins peuvent être, à cause des décalages observés dans l'alignement automatique, détectées au niveau du phonème adjacent.

Il faut aussi noter que mis à part H1, les taux de *AnRappel* dépassent 0.72 pour tous les autres patients hommes et femmes.

En comparant les scores obtenus à ceux de la "baseline", on observe que l'approche proposée reposant sur la classification à base de SVM obtient de meilleures performances de *AnRappel* atteignant des moyennes de 0.89 et 0.72 pour les locutrices et les locuteurs dysarthriques respectivement. Cela représente un gain absolu de 6% (7.2% relatif) et 7% (10.7% relatif) respectivement. Cependant, un gain de précision est enregistré seulement sur les patients hommes atteignant une mesure de *AnPrec* de 0.61 (gain absolu de 5%). La précision au niveau des patientes femmes stagne à 0.65 pour les deux approches.

Des comportements similaires peuvent être observés sur le taux d'anomalies obtenu par les deux approches sur les locuteurs contrôles du corpus *VML* présentés dans

TABLE 5.2 – *Performances du système "baseline" sur les locuteurs dysarthriques du corpus VML, exprimées en termes de AnRappel et AnPrec, selon les deux stratégies de comparaison utilisées.*

	Stratégie 1		Stratégie 2	
Locuteur dysarthrique	*AnRappel*	*AnPrec*	*AnRappel*	*AnPrec*
H1	0.16	0.12	0.36	0.30
H2	0.44	0.38	0.77	0.77
H3	0.48	0.28	0.76	0.53
H4	0.44	0.37	0.73	0.64
Moyenne - Homme	0.38	0.29	0.65	0.56
F1	0.48	0.22	0.79	0.45
F2	0.43	0.66	0.87	0.98
F3	0.50	0.31	0.79	0.55
F4	0.60	0.36	0.88	0.63
Moyenne - Femme	0.50	0.39	0.83	0.65
Moyenne	0.44	0.34	0.74	0.61

le tableau 5.3. En effet, une dimunition des taux d'anomalies est observée sur les locuteurs contrôles hommes alors que ceux mesurés sur les contrôles femmes sont semblables pour les deux approches. Seulement 7% de phonèmes des locuteurs contrôles sont considérés comme anomalies par l'approche proposées (9,1% pour les femmes et 2,4% pour les hommes). Bien que nous considérons ces anomalies comme des faux positifs étant détectées sur les contrôles, d'autres hypothèses peuvent être émises.

En effet, même si ces locuteurs sont sains et non dysarthriques, leurs enregistrements n'ont pas été évalués par des experts humains afin de vérifier la qualité de leurs productions au niveau phonème. De plus, la parole saine peut elle aussi contenir des productions qu'on peut juger "atypiques" qui peuvent être liées à des accents régionaux ou des réductions de nature non pathologique.

TABLE 5.3 – *Performances de l'approche proposée et du système "baseline" sur les locuteurs contrôles du corpus VML, exprimées en termes de taux d'anomalie (%).*

Locuteurs dysarthriques	Système "baseline"	Système proposée (SVM)
Moyenne contrôles - Homme	7.2	2.4
Moyenne contrôles - Femme	9.0	9.1
Moyenne contrôles	8.4	6.9

Finalement, le tableau 5.4 détaille les taux de *AnRappel* et *AnPrec* mesurés pour chaque catégorie phonétique utilisée lors de la phase d'apprentissage des sous-modèles SVM. En comparant ces résultats, on observe que l'approche automatique présente beaucoup de difficultés à détecter les anomalies sur les voyelles en général et les voyelles nasales plus particulièrement. La mesure de *AnRappel* calculée sur cette catégorie en utilisant la stratégie 2 de comparaison n'atteint que 0.77, alors qu'elle est de 0.83, 0.84 et 0.86 pour les voyelles orales, consonnes sonores et consonnes sourdes respectivement. Cela peut être dû au nombre limité de phonèmes utilisés lors de l'apprentissage du

TABLE 5.4 – *Performances de l'approche proposée sur les patients du corpus VML, exprimées en termes de mesures de AnRappel et AnPrec sur les différentes catégories phonétiques utilisées dans l'apprentissage de chaque classifieur.*

	Stratégie 1		Stratégie 2	
Catégorie phonétique	*AnRappel*	*AnPrec*	*AnRappel*	*AnPrec*
Consonnes sourdes	0.68	0.41	0.86	0.61
Consonnes sonores	0.60	0.31	0.84	0.57
Voyelles orales	0.42	0.44	0.83	0.75
Voyelles nasales	0.35	0.47	0.77	0.77

modèle pour cette catégorie par rapport aux autres catégories phonétiques.

Cependant, ce taux de *AnRappel* relativement faible coïncide avec le meilleur taux de *AnPrec* atteignant 0.77 sur cette catégorie phonétique. Cela reste tout de même comparable aux mesures enregistrées sur les voyelles orales (0.75) mais bien supérieur au mesures obtenues sur les consonnes qui n'atteignent que 0.57 et 0.61 pour les consonnes sonores et sourdes respectivement.

En se basant sur ces premiers résultats, les observations suivantes peuvent êtres retenues :

— le comportement de l'approche automatique proposée est assez stable si on compare les performances mesurées sur les locuteurs hommes et femmes. Cela est confirmé sur les différents enregistrements disponibles pour chaque locuteur (à l'exception du locuteur H1) ;
— l'approche présente de meilleurs résultats sur les locuteurs atteints de dysarthrie sévère. Pour les locuteurs F2, F4 et M2 qui sont considérés comme les locuteurs les plus dysarthriques dans ce corpus, l'approche automatique atteint les meilleurs taux de *AnRappel* et *AnPrec* (1 et 0.98 respectivement pour F2). Les locuteurs M1, F1 et F3, les moins dysarthriques dans ce corpus, présentent les plus faibles performances pour les deux mesures d'évaluation ;
— l'ensemble des paramètres utilisés pour la caractérisation des phonèmes se montre pertinent dans la tâche de détection d'anomalies dans la parole dysarthrique résultant en une amélioration des taux de *AnRappel* et *AnPrec* par rapport à l'approche "baseline". Cependant, et malgré les gains observés, la précision de l'approche reste plutôt faible et montre une tendance à détecter plus d'anomalies qu'il y en a, et donc à être plus sévère que l'expert humain dans le jugement des phonèmes ;
— les anomalies détectées automatiquement sur les locuteurs contrôles peuvent être liées à des bruits ou des productions "atypiques" non pathologiques ;
— même si les mesures de *AnRappel* sont assez proches et comparables pour toutes les catégories, des différences notables sont observées pour les *AnPrec* où plus de faux positifs sont détectés sur les consonnes par rapport aux voyelles.
— la détection de plus d'anomalies au niveau des consonnes peut aussi résulter de la dégradation systématique et importante de ces phonèmes par la dysarthrie alors que les effets subis par les voyelles sont moins fréquents et marqués (Darley

et al., 1969b).

5.2.2 Application sur un corpus non annoté *DesPhoAPaDy* :

Le corpus *VML* utilisé dans la première évaluation de l'approche présente quelques inconvénients : il ne comporte que 8 patients dysarthriques atteints tous de la même pathologie (maladies lysosomales) liée à la même classe dysarthrique (dysarthrie mixte). De plus, l'annotation utilisée comme référence pour l'évaluation du système a été effectuée par un seul expert ce qui la rend moins robuste. Afin de généraliser les observations annoncées précédemment sur ce corpus, l'évaluation expérimentale de l'approche proposée a été élargie au corpus *DesPhoAPaDy*, beaucoup plus large et contenant plus de pathologies/classes de dysarthrie que le corpus *VML*.

Cependant, et puisqu'aucune annotation des anomalies au niveau phonème n'est disponible pour ce corpus, l'étude a porté sur le lien entre le taux de phonèmes annotés comme anomalies par l'approche automatique et les critères perceptifs globaux présents dans l'évaluation de la parole réalisée par le jury d'experts et détaillée dans le tableau 3.3.

Les figures 5.2 et 5.3 montrent, pour chaque genre, le taux d'anomalies détectées automatiquement pour chaque locuteur en fonction de son grade de sévérité de dysarthrie. Ces figures montrent une importante relation entre les deux mesures, confirmée par une mesure de corrélation de Pearson de 0.89 et 0.86 pour les locuteurs dysarthriques femmes et hommes respectivement. Cette observation, même si non concluante, supporte le comportement de l'approche observé précédemment sur le corpus *VML*, pour lequel les mesures de $AnRappel$ et $AnPrec$ sont meilleurs pour les locuteurs sévèrement dysarthriques (degré de sévérité de la dysarthrie important).

Ces corrélations, donc, mêmes si elles ne prouvent pas une précision de l'approche dans la détection des anomalies au niveau phonème, confirment le potentiel de la méthode face aux différentes altérations acoustiques de la parole et sa capacité à prendre en compte l'évolution de la dysarthrie. De plus, ces figures présentent des pentes de régression plus importantes pour les degrés de sévérité de dysarthrie importants. Cette tendance, ainsi que le comportement observé sur les patients sévèrement dysarthriques du corpus *VML* (F2, F4 et M2), suggèrent que l'approche automatique est plus performante quand les locuteurs sont atteints de dysarthrie moyennes ou sévères. Un tel comportement peut être utilisé dans le cadre du suivi longitudinal de l'évolution de la dysarthrie chez les patients.

Outre le degré de sévérité de la dysarthrie, l'évaluation perceptive réalisée par le jury d'experts jugeait la parole sur d'autres critères de qualité de parole. Le tableau 5.5 présente ses différentes mesures de corrélation par genre et par pathologie. Ces valeurs montrent des différences intéressantes entres les différentes pathologies selon les critères :

— des taux de corrélation importants (supérieurs à 0.87) entre le taux d'anomalies détectées automatiquement et le degré de sévérité globale de la dysarthrie pour

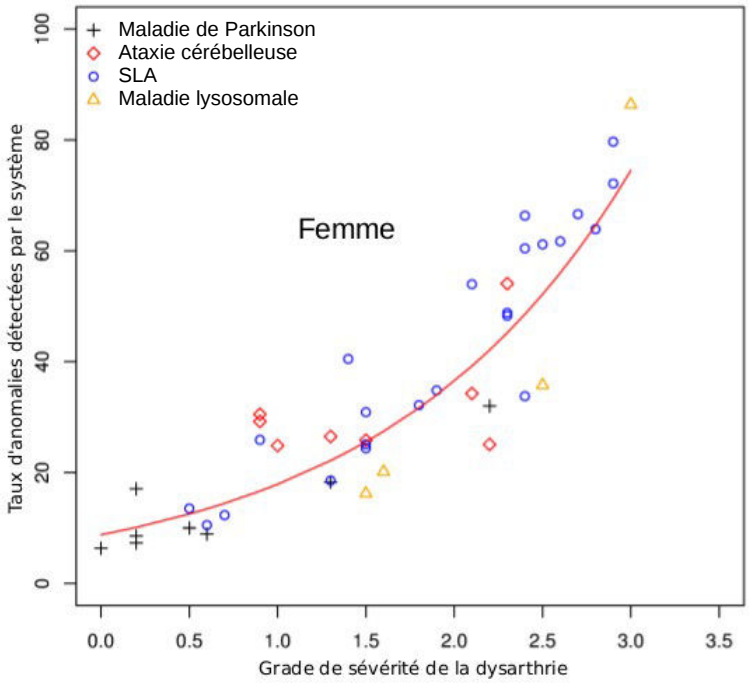


FIGURE 5.2 – *Taux d'anomalies détectées automatiquement par l'approche proposée en fonction du degré de sévérité de la dysarthrie pour les locuteurs dysarthriques femmes du corpus DesPhoAPaDy.*

toutes les pathologies sauf les patients atteints d'ataxies cérébelleuses (surtout les femmes, 0.52) ;

— des comportements similaires sont observés pour les troubles de l'articulation et l'intelligibilité avec des mesures de corrélation avec les taux d'anomalies détectées par l'approche $\in [0.8; 0.9]$ pour toutes les pathologies sauf les ataxies cérébelleuses (femmes) ;

— les corrélations avec le débit de parole, bien que supérieures à 0.5, sont plus hétérogènes. En effet, les troubles de l'articulation et d'intelligibilité sont plus observables au niveau phonème que le débit de parole. En effet, la seule dimension temporelle associée à l'approche proposée est la durée des phonèmes qui est, bien que dépendante du débit, plus locale et liée à la catégorie du phonème lui même ;

Il est intéressant aussi de relever des taux de corrélation similaires observés sur les patients atteints de SLA, maladie de Parkinson et maladies lysosomales. Ce comportement confirme l'intérêt de l'approche proposée dans la détection d'anomalies au niveau phonème et soutient sa portabilité sur d'autres pathologies et classes dysarthriques. Le comportement particulier de l'approche face à la dysarthrie ataxique (surtout les femmes) peut être lié à son effet distingué sur le débit de parole. En effet, les patients atteints d'ataxies cérébelleuses peuvent, tout en présentant des dysarthries de

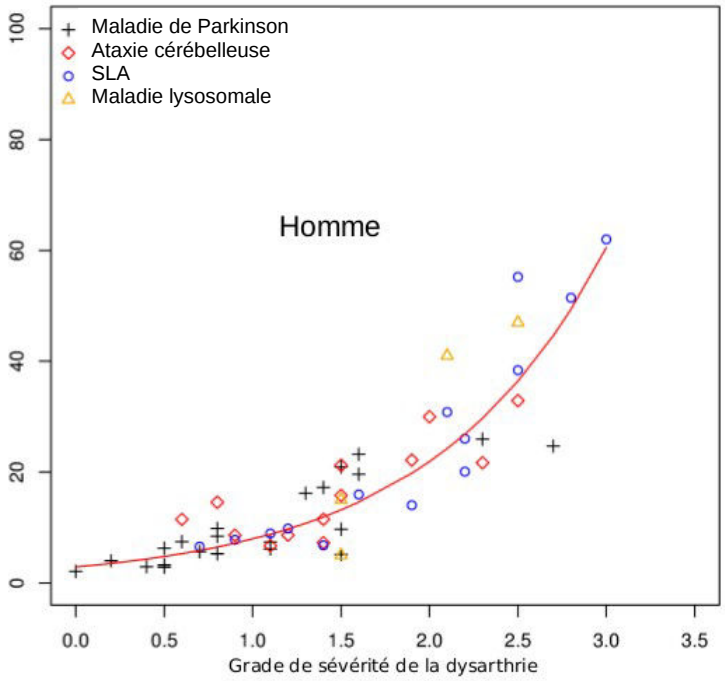


FIGURE 5.3 – *Taux d'anomalies détectées automatiquement par l'approche proposée en fonction du degré de sévérité de la dysarthrie pour les locuteurs dysarthriques hommes du corpus DesPhoAPaDy.*

sévérité moyenne voire légère, parler avec un débit très faible qui induit l'outil de détection d'anomalies à l'erreur. Nous remarquons alors que la mesure de corrélation entre le taux d'anomalies automatiques et les items perceptifs la plus importante sur les patientes atteintes d'ataxie cérébelleuse est avec le débit de la parole atteignant 0.64 contre des corrélations de 0.52 et 0.50 avec la sévérité de la dysarthrie et l'intelligibilité respectivement.

Le tableau 5.6 donne les taux d'anomalies de l'approche sur les locuteurs contrôles du corpus *DesPhoAPaDy*. L'approche automatique détecte environ 10% et 3% des phonèmes comme anomalies sur les contrôles femmes et hommes respectivement. Ces résultats sont consistants avec ceux observés sur les contrôles issus du corpus *VML* pour lesquels environ 9% et 4% d'anomalies ont été détectées sur les contrôles femmes et hommes respectivement.

En conclusion, la stabilité du comportement de l'approche sur les locuteurs contrôles et patients issus des corpus *VML* et *DesPhoAPaDy* semble confirmer son intérêt et sa possible utilisation pour différents types de dysarthrie et différentes pathologies.

TABLE 5.5 – *Corrélation entre le taux d'anomalies détectées automatiquement et les évaluations perceptives (moyennées sur les différents experts) des patients des corpus DesPhoAPaDy et VML regroupés par population dysarthrique.*

Pathologie - sexe	degré de sévérité	Troubles de l'articulation	Intelligibilité	Débit
Maladie de Parkinson - F	0.89	0.86	0.89	0.81
Ataxies cérébelleuses - F	0.52	0.50	0.38	0.64
SLA - F	0.91	0.86	0.83	0.92
Maladie lysosomale - F	0.90	0.81	0.87	0.43
Maladie de Parkinson - H	0.87	0.80	0.84	0.60
Ataxies cérébelleuses - H	0.81	0.65	0.83	0.59
SLA - H	0.91	0.82	0.86	0.67
Maladie lysosomales - H	0.96	0.88	0.68	0.69

TABLE 5.6 – *Performances de l'approche proposée sur les locuteurs contrôles du corpus DesPhoAPaDy, exprimées en termes de taux d'anomalies (%).*

Locuteurs	degré de sévérité de la dysarthrie	Taux d'accord
Moyenne contrôles - Homme	0.16	2.9
Moyenne contrôles - Femme	0.10	10.4
Moyenne contrôles	0.13	6.6

5.3 Discussion du comportement de l'approche de détection d'anomalies

Les premières analyses décrites précédemment ont confirmé l'intérêt général de la méthode et sa capacité à capter l'évolution de la sévérité de la dysarthrie dans les différentes pathologies. Cependant, toutes ces évaluations ont été réalisées sur de la parole lue, qui est certes toujours utilisée dans la pratique clinique pour l'évaluation de la dysarthrie, mais qui correspond à des conditions de production très contrôlées qui ne peuvent être toujours adaptées à une utilisation plus généralisée de ce type d'approche automatique par les professionnels et potentiellement les patients eux-même.

5.3.1 Comportement face à la parole lue et spontanée

Cette section décrit et discute le comportement de l'approche de détection automatique d'anomalies face à la parole lue et spontanée issue du corpus *TypALoc* (Laaridh et al., 2016a).

Analyse par population

La figure 5.4 illustre les taux d'anomalies automatiquement détectées par rapport aux degré de sévérité de la dysarthrie pour les deux styles de parole.

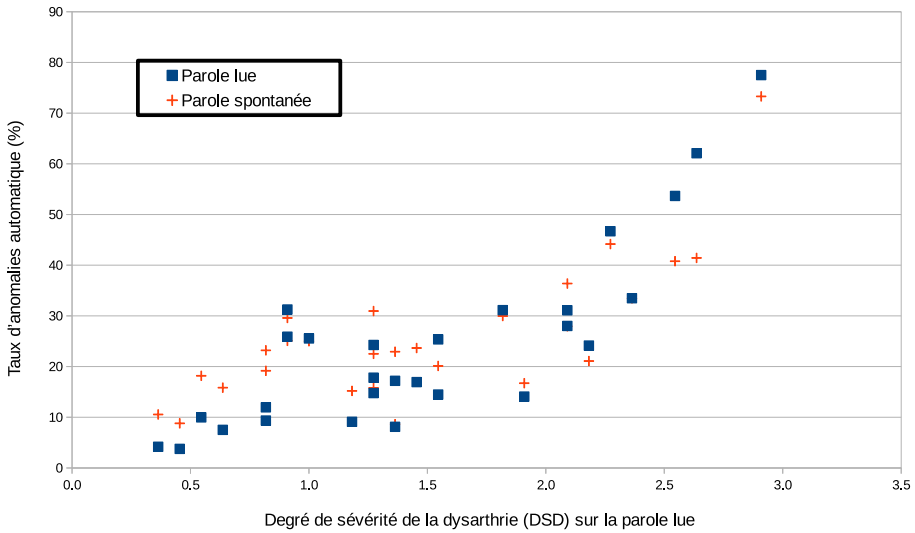


FIGURE 5.4 – *Taux d'anomalies détectées automatiquement en fonction du degré de sévérité de la dysarthrie et du style de parole pour le corpus TypALoc.*

À l'image des résultats retrouvés dans le section précédente sur la parole lue, nous retrouvons que l'approche détecte plus d'anomalies chez les patients atteints de dysarthrie plus sévère. Ce comportement, bien que moins tranché, semble se conserver sur la parole spontanée. En effet, la corrélation de Pearson entre ces taux et le degré de sévérité de la dysarthrie atteint 0.81 et 0.60 pour la parole lue et spontanée respectivement. Cette observation soutient l'aptitude du système à capturer l'évolution de la dysarthrie indépendamment de la tâche réalisée par le patient.

Le tableau 5.7 présente les taux d'anomalies relevés sur la parole lue et spontanée regroupés par population du corpus *TypALoc*. On remarque que pour les contrôles, l'approche détecte plus d'anomalies sur la parole spontanée comparée à la parole lue. Cela peut être lié au fait que les modèles de phonèmes normaux et anormaux ont été appris sur de la parole lue exclusivement. En effet, la parole spontanée peut présenter plus de variabilité acoustique qui peut être considérée comme atypique comparée à la parole lue mais non pathologique. Ces variations peuvent être liées au débit de parole plus rapide en spontanée, aux faux départs, aux hésitations ainsi qu'à des phénomènes de réduction plus fréquents souvent observés dans le cadre de parole spontanée.

La figure 5.5 illustre la relation entre la différence numérique des taux d'anomalies entre les deux styles de parole (en abscisse x= taux d'anomalies sur la parole spontanée $-$ taux d'anomalies sur la parole lue) et l'évaluation du degré de sévérité de la parole lue. Chaque point correspond à un locuteur (contrôle ou patient). Observant le tableau 5.7 et la figure 5.5, nous remarquons que, comme pour les contrôles, l'approche

TABLE 5.7 – *Taux moyen d'anomalies détectées (%) par pathologie et style de parole sur le corpus TypALoc.*

Population	Parole lue	Parole spontanée
Maladie de Parkinson	10.6	17.8
Ataxies cérébelleuse	20.6	24.4
SLA	35.8	31.9
Contrôles	5.4	13.7

détecte plus d'anomalies sur la parole spontanée pour les patients atteints de la maladies de Parkinson et d'ataxies cérébelleuses. Cette tendance est conforme aux résultats retrouvés dans (Van Lancker Sidtis et al., 2012; Kempler et Van Lancker, 2002) sur des patients atteints de la maladie de Parkinson où la parole spontanée était moins intelligible et contenait plus de dysfluences que la parole lue.

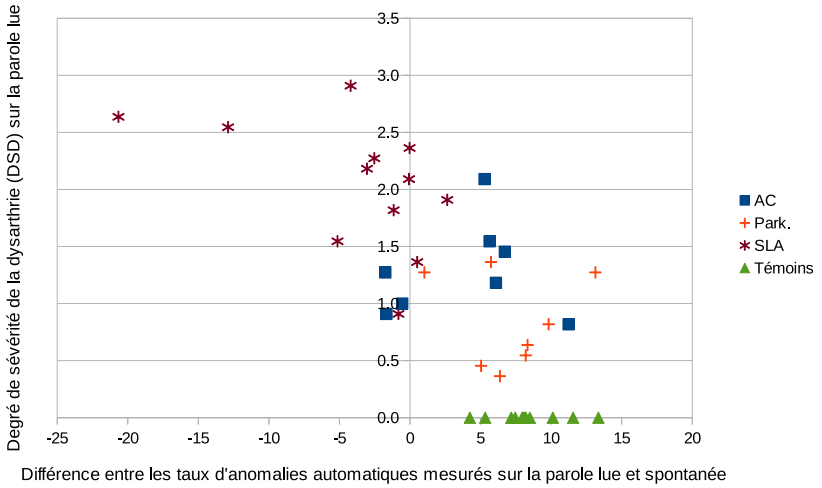


FIGURE 5.5 – *Distribution des différences entre les taux d'anomalies sur les paroles spontanée et lue selon le degré de sévérité de la parole lue.*

Par contre, les patients atteints de SLA présentent des taux d'anomalies similaires, voire même inférieurs, sur la parole spontanée par rapport à la lecture (32% et 36% respectivement). Dans notre corpus, ces patients souffrent des dysarthries les plus sévères (degrés de sévérité élevés). Une analyse plus approfondie de cette tendance est nécessaire pour vérifier si elle est liée aux caractéristiques intrinsèques de la SLA qui affecterait plus la tâche de lecture que celle de la parole spontanée. La deuxième hypothèse émise serait que ce phénomène est davantage lié au degré de sévérité élevé de la dysarthrie des patients SLA qu'à leur pathologie. En effet, la tâche de production de parole spontanée offre aux patients plus de "liberté" pour contrôler leur fatigue, débit de parole ainsi que les phonèmes et contextes à produire, ce qui pourrait donner lieu à

moins d'anomalies au niveau phonème.

D'un point de vue plus général, le tableau 5.7 montre une augmentation plus importante des taux d'anomalies sur la parole spontanée chez les contrôles que chez les patients dysarthriques : une augmentation relative de 154% pour les contrôles contre 68%, 18% et -11% pour les patients atteints de la maladie de Parkinson, ataxies cérébelleuses et SLA respectivement. Ces nombres suggèrent que la différence des taux d'anomalies entre la parole spontanée et lue est inversement proportionnelle à la sévérité de la dysarthrie. Cela supposerait que les contrôles changent davantage leurs productions selon le style de parole (ce qui résulte dans notre cas en plus d'anomalies détectées en spontanée vu la nature des modèles appris sur la lecture) alors que les patients (surtout les plus dysarthriques) perdent cette capacité à s'adapter aux différents styles et ont tendance à uniformiser leurs productions indépendamment de la tâche. Cette hypothèse, remarquable et très intéressante, devra nécessiter une étude plus approfondie pour être confirmée.

Analyse par classes phonétiques

Le tableau 5.8 détaille les taux d'anomalies détectées (%) par pathologie, catégorie phonétique et style de parole. Pour les contrôles, toutes les catégories phonétiques présentent des taux d'anomalies plus importants sur la parole spontanée comparée à la parole lue. Les fricatives enregistrent cependant une hausse nette des taux d'anomalies passant de 7% sur la parole lue à 23% sur la parole spontanée.

TABLE 5.8 – *Taux moyen d'anomalies détectées (%) par pathologie, catégorie phonétique et style de parole.*

	Parole lue				Parole spontanée			
Catégorie phonétique	Contrôles	Park.	AC	SLA	Contrôles	Park.	AC	SLA
Consonnes occlusives	7	12	22	37	11	16	24	35
Consonnes fricatives	7	26	48	37	23	55	53	38
Consonnes nasales	7	12	21	31	19	10	19	35
Consonnes liquides	7	9	23	41	15	9	27	44
Voyelles orales	2	6	10	33	11	10	13	29
Voyelles nasales	4	8	20	44	8	13	13	26
Autres	10	16	26	46	19	13	26	37

Chez les patients atteints de la maladie de Parkinson, on trouve que les fricatives présentent également des taux d'anomalies les plus importants pour les deux styles de parole. Ce comportement rappelle l'important taux de confusion phonémique observé sur cette catégorie sur la parole lue et rapporté dans la section 4.2.4. En effet, le rang du phonème associé à chaque segment lors l'alignement non contraint par le texte (figure 4.3 est un des paramètres utilisés lors de la caractérisation et la classification des phonèmes comme normaux et anormaux (section). La présence de confusions

phonémiques peut alors favoriser la détection d'anomalies sur cette classe. Il s'agirait donc davantage de l'amplification du comportement déjà observé sur cette catégorie sur la parole lue et lié à la dysarthrie parkinsonienne que de l'apparition d'un nouveau phénomène. Cependant, il est intéressant de constater que, contrairement aux autres catégories phonétiques où la hausse des taux d'anomalies sur la parole spontanée est légère, les fricatives présentent une augmentation absolue de 29% (112% relative).
Considérant les patients atteints de SLA, bien que les taux d'anomalies soient plutôt stables entre les deux styles de parole, on remarque tout de même que les voyelles présentent moins d'anomalies en spontanée qu'en lecture (-4% et -18% pour les voyelles orales et nasales respectivement). Cela peut être lié à la durée des phonèmes moins courtes observée sur la parole spontanée par rapport à la parole lue (durées moyennes de 85.3ms (tableau 4.5) et 95.1ms (tableau 4.9) respectivement). Cette différence de durées bien qu'observée aussi au niveau des consonnes ne semble pas impacter la détection d'anomalies. Cela implique que la durée est plus importante dans la caractérisation des phonèmes normaux et anormaux sur les voyelles que sur les consonnes. De plus, il est intéressant de relever que le taux d'anomalies sur les voyelles nasales pour cette population passe de 44% sur la parole lue à 26% sur la parole spontanée. Cette différence peut indiquer une hypernasalité moins importante sur cette parole. En effet, lors de l'évaluation perceptive du corpus *TypALoc* sur l'item caractérisant le nasonnement de la parole, les jurés ont coté en moyenne la parole lue des patients atteints de SLA à 1.7 et la parole spontanée à 1.5.

La figure 5.6 illustre les taux d'anomalies pour chaque population et style de parole dans le corpus *TypALoc*. Pour chaque population, une ANOVA à un facteur a été réalisée afin d'étudier l'effet du style de parole (2 niveaux : parole lue, parole spontanée).

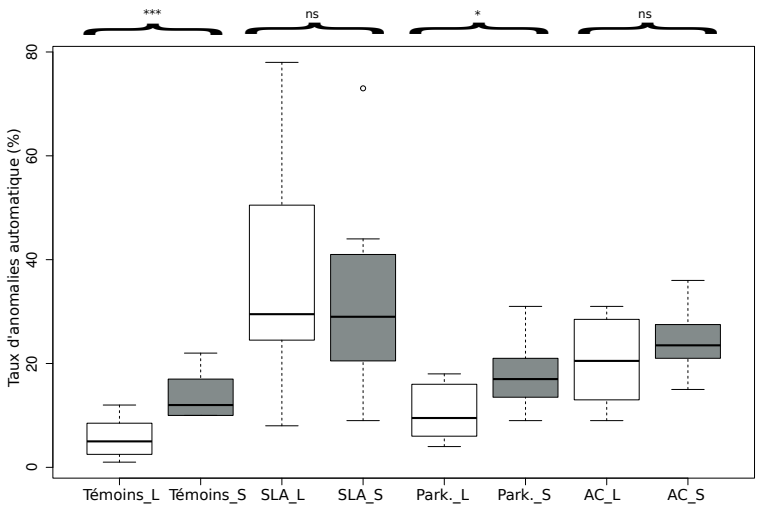


FIGURE 5.6 – *Taux d'anomalies par population et style de parole (parole lue (blanc) et spontanée (gris)) dans le corpus TypALoc.*

Chez les sujets contrôles ainsi que ceux souffrant de dysarthrie parkinsonienne, une différence significative est trouvée entre la lecture et le spontané ($(p<0.001, F(1,22)=28)$ et $(p<0.05, F(1,14)=5.4)$ respectivement). Ces différences sont encore plus visibles quand on se concentre seulement sur les fricatives ($(p<0.001, F(1,22)=19)$ et $(p<0.01, F(1,14)=14)$; La différence entre les deux styles de parole est moins flagrante pour les patients souffrant d'ataxie cérébelleuse et de SLA. Plus particulièrement pour les SLA, l'effet lié aux styles de parole peut être masqué par l'importante variabilité intra-pathologique observée sur cette population au niveau des taux d'anomalies automatiques. De plus, cette variabilité est également observable au niveau des degrés de sévérité des patients SLA présentant des dysarthrie légères, moyennes et sévères contrairement aux patients parkinsoniens atteints tous pour l'ensemble d'une dysarthrie légère.

Dans cette section, nous avons pu observer et comparer le comportement de l'approche de détection automatique d'anomalies au niveau phonème sur la parole dysarthrique lue et spontanée. Un effet important de la tâche et du style de parole a été observé. Les patients atteints de SLA, contrairement à toutes les autres populations, présentent plus d'anomalies sur la parole lue que la parole spontanée. En outre, sur l'ensemble des patients, plus la dysarthrie est sévère moins il y a de différences entre les deux styles de parole. Une hypothèse émise sur la base de ces observations est que les contrôles adaptent leur production selon le style de parole (ce qui résulte en plus d'anomalies détectées sur la parole spontanée que sur la parole lue) alors que les patients dysarthriques perdent graduellement cette capacité à s'adapter aux différents styles de parole.

5.3.2 Détection d'anomalies et alignement de la parole

L'approche proposée repose sur deux phases, un alignement contraint par le texte de la parole suivi d'une classification des phonèmes alignés. La précision de l'alignement étudiée dans le chapitre 4, est alors primordial pour le bon fonctionnement du système. Nous proposons dans cette section d'étudier le comportement de l'approche de détection des phonèmes anormaux selon la précision de l'alignement automatique (Laaridh et al., 2016b).

Le corpus *VML*

La première partie de cette étude porte sur le corpus *VML* puisqu'il comporte à la fois un alignement et une annotation d'anomalies de référence au niveau phonème.

Les résultats présentés dans la section précédente ont montré que l'approche proposée détecte plus d'anomalies et peut être par conséquent considérée comme plus sévère que l'expert humain dans l'évaluation des phonèmes. Le tableau 5.9 présente la distribution de tous les phonèmes issus de l'alignement automatique, phonèmes labellisés comme anomalies par l'approche automatique, phonèmes labellisés comme anomalies par l'expert humain et phonèmes correctement détectés comme anomalies par l'approche automatique (vrais positifs) selon les valeurs DD.

TABLE 5.9 – *Distribution des phonèmes, anomalies détectées automatiquement, anomalies annotées manuellement et anomalies correctement détectées (vrais positifs) selon les mesures du décalage entre les alignements automatique et manuel en début de phonème (DD) pour les patients du corpus VML.*

DD	$\geq$60ms	=50ms	=40ms	=30ms	=20ms	=10ms	= 0ms
phonèmes	1263	259	410	696	1392	2614	2465
Anomalies automatiques	620	92	100	127	199	400	595
Anomalies manuelles	363	61	70	81	132	228	310
Vrais positifs	222	32	32	29	59	120	194

Cette distribution montre que 71% des phonèmes sont alignés dans l'intervalle $\pm 20ms$. Ce taux est satisfaisant considérant les degrés de sévérité de dysarthrie observés sur les patients de ce corpus.

Deux comportements peuvent être soulignés ; le premier concerne les phonèmes pour lesquels les mesures de DD sont en dehors de l'intervalle $\pm 20ms$. Ici, moins de 50% des anomalies annotées automatiquement (44%) et manuellement (46%) sont présentes. De plus, on remarque que la probabilité qu'un phonème soit labellisé comme une anomalie par l'approche automatique augmente pour les valeurs extrêmes de DD. Ce comportement est attendu dû à la nature des paramètres utilisés lors de la phase de classification et issus essentiellement de la phase d'alignement de la parole. Il est intéressant aussi de remarquer que la probabilité qu'un phonème soit labellisé manuellement comme anomalie augmente avec les valeurs de DD. Cette observation confirme une des hypothèses émises lors de cette étude : une erreur d'alignement sur un phonème tend à engendrer la détection automatique de ce même phonème comme anormal lors de la classification. Cependant, une partie des erreurs d'alignement résulte de la présence d'altérations dans la parole. Dans ce cas, l'alignement contraint par le texte fait face à plus de difficultés pour délimiter les frontières du phonème. Cela résulte en des valeurs importantes de DD sur ces phonèmes, ce qui augmente la probabilité qu'ils soient détectés comme des anomalies par l'approche.

Nous remarquons aussi que seulement 29% des anomalies de référence ont des mesures de $DD \geq 60ms$. (61% de ces anomalies sont détectées par l'approche proposée). Beaucoup de faux positifs, dus aux erreurs d'alignements, sont rapportés sur ces phonèmes. Une réflexion est nécessaire sur les causes de ces grands décalages d'alignement sur des segments jugés perceptivement comme normaux par l'expert. L'éventuelle présence de fluctuations du débit de parole où d'un débit de parole atypique mais global, non annoté par conséquent par l'expert humain sur chaque phonème peut notamment expliquée de telles erreurs.

Le deuxième comportement observé dans le tableau 5.9 concerne les phonèmes avec des mesures de DD dans l'intervalle $\pm 20ms$. Ici, plus de la moitié des anomalies manuelles sont représentées. Pour ces phonèmes, et malgré les patterns acoustiques anormaux annotés par l'expert humain, ces irrégularités n'ont pas un effet visible sur alignement automatique. Observant seulement les vraies anomalies détectées par l'approche automatique, environ 54% ne sont donc pas liées à des erreurs d'alignement.

Cela montre la capacité de l'approche dans la détection des phonèmes anormaux sur la seule base de leurs irrégularités acoustiques. La mesure de *AnRappel* sur ces phonèmes est de 0.56 (calculée selon la première stratégie d'évaluation ne prenant en considération que le phonème lui même dans la comparaison). Cependant, elle présente toujours un comportement plus sévère que l'expert humain dans son annotation (*AnPrec* de 0.31). Une caractérisation de ces anomalies annotées automatiquement sera nécessaire afin de relever le type et l'amplitude de distorsion que l'approche est capable de relever.

Le corpus *TypALoc*

Les figures 5.7 et 5.8 représentent les distributions des phonèmes et des anomalies détectées respectivement en fonction des valeurs de *DD* pour les populations du corpus *TypALoc*. Afin de pouvoir comparer le comportement observé sur ce corpus avec celui détaillé précédemment, les taux mesurés sur le corpus *VML* sont aussi rapportés dans les figures.

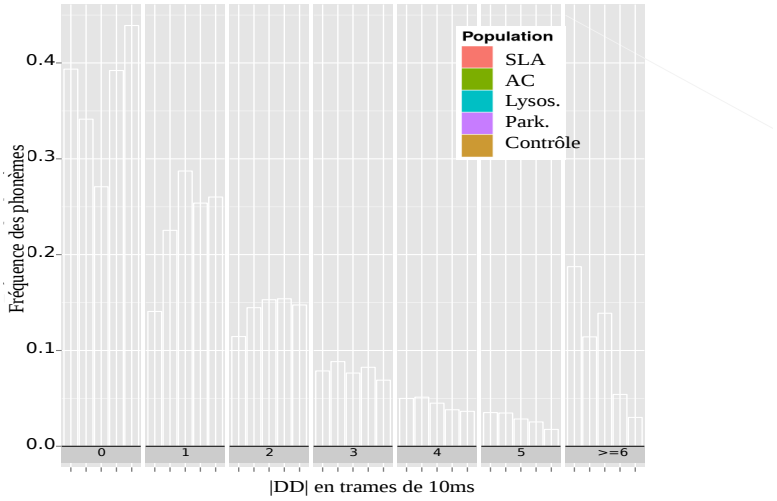


FIGURE 5.7 – *Distribution des phonèmes selon les valeurs de DD pour les populations du corpus TypALoc et les patients du corpus VML (Lysos.). Chaque tranche représente un décalage d'une trame (10ms).*

Analyse par population

Regardant la distribution des phonèmes dans la figure 5.7, 85%, 64%, 71% et 79% des phonèmes présentent des mesures de *DD* dans l'intervalle $\pm 20ms$ pour les contrôles, et les patients atteints de SLA, ataxie cérébelleuse et maladie de Parkinson respectivement. Ces taux ne sont pas si éloignés de ceux observés sur le corpus *VML* (66%, chapitre 4), et la variabilité inter-pathologique observée peut être expliquée par les différences de degrés de sévérité de dysarthrie et des débits de parole observés sur ces

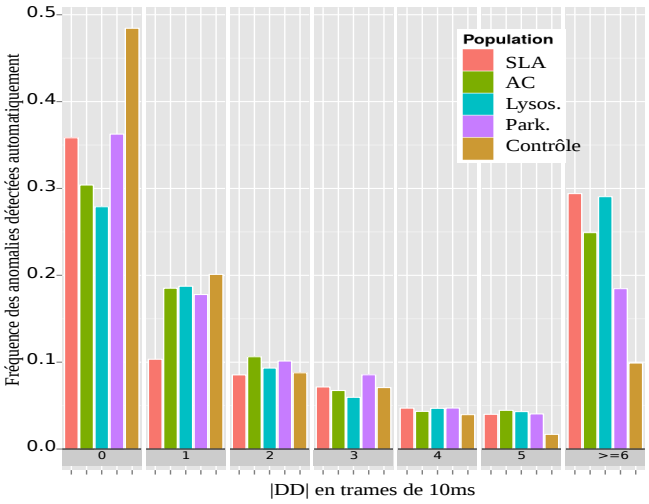


FIGURE 5.8 – *Distribution des anomalies détectées selon les valeurs de DD pour les populations du corpus TypALoc et les patients du corpus VML . Chaque tranche représente un décalage d'une trame (10ms).*

populations.

Observant la figure 5.8, des différences entre les taux d'anomalies détectées sur toutes les populations au niveau de chaque tranche de DD peuvent être observées. Cependant, il est important de noter que les deux comportements observés précédemment sur le corpus *VML* sont conservés et même accentués pour les phonèmes avec des mesures de DD dans l'intervalle $\pm 20ms$. En effet, 77%, 55%, 60% et 64% des anomalies détectées sur les contrôles et les patients atteints de SLA, d'ataxie cérébelleuse et de maladie de Parkinson respectivement sont désormais dans cet intervalle, contre 56% pour les patients atteints de maladies lysosomales.

Sur l'intervalle ($DD \geq$ 60ms), nous pouvons constater que les patients atteints de SLA ont un comportement similaire aux VML, avec des taux d'anomalies détectées automatiquement élevés, tous deux associées à une dysarthrie mixte. Les taux observés chez les patients atteints d'ataxies cérébelleuses et de la maladie de Parkinson sont conformes aux degrés de sévérité de la dysarthrie relevés chez ces populations.

Analyse par classe phonétique

Le tableau 5.10 rapporte les taux d'anomalies détectées par population et catégorie phonétique pour les phonèmes avec des mesures de $DD \in \pm 20ms$ et les phonèmes avec $DD \notin \pm 20ms$.

On remarque que les consonnes montrent les taux d'anomalies les plus importants

pour toutes les populations indépendamment de la valeurs de DD observée. Ce comportement est consistent avec l'impact systématique de la dysarthrie sur la production des consonnes chez toutes les pathologies. Dans l'ensemble, des taux relatifs d'anomalies plus importants sont observés sur les phonèmes liés à des mesures de DD en dehors de l'intervalle $\pm 20ms$. Cela n'indique pas qu'il y a plus d'anomalies détectées liées à des erreurs d'alignement qu'à des traits acoustiques anormaux relevés par l'approche mais plutôt que la probabilité de détecter des phonèmes comme étant anormaux augmente pour les mesures de DD importants. Autrement dit, il plus facile de détecter les anomalies affectant de manière flagrante la précision de l'alignement que celles plus subtiles indifférentes au processus d'alignement.

TABLE 5.10 – *Taux d'anomalies détectées (%) par population et catégorie phonétique mesurés pour les phonèmes avec $DD \in \pm 20ms$ et ceux avec $DD \notin \pm 20ms$. (corpus TypALoc et patients du corpus VML).*

Catégorie phonétique	$DD \in \pm 20ms$					$DD \notin \pm 20ms$				
	Contrôles	AC	Park.	SLA	Lysos.	Contrôles	AC	Park.	SLA	Lysos.
Consonnes	7	23	11	32	25	10	35	21	46	40
Voyelles	3	10	5	29	12	4	16	13	47	28

5.4 Localisation des anomalies sur les mots bisyllabiques

En français, les mots bisyllabiques sont caractérisés par une structure court-long très robuste aussi bien lors de la production que de la perception. Il s'agit du pattern iambique. Ce caractère permet de renforcer la deuxième syllabe de ces mots. Une étude récente a montré que ce caractère rythmique semble robuste même dans la parole dysarthrique bien qu'il soit un peu affecté quand la sévérité de la dysarthrie augmente (Georgeton et Meunier, 2016).
Nous avons alors étudié le comportement de l'approche sur ces seuls mots. L'hypothèse émise dans le cadre de cette étude est que, dû au renforcement observé sur la deuxième syllabe, le taux d'anomalies détectées sur ces dernières sera moins important que celui des anomalies détectées sur la première syllabe.

En se basant sur l'alignement automatique contraint par le texte (la suite de phonèmes générée) ainsi que la transcription de la parole produite, une syllabification automatique est réalisée en utilisant l'outil SPPAS (Bigi et Hirst, 2012). Ce processus repose sur 2 principes (1) une syllabe contient une seule voyelle (2) une pause est une frontière de syllabes.

La figure 5.9 montre effectivement que le pattern iambique observé chez les locuteurs contrôles (les durées moyennes des syllabes 1 et 2 sont de 178ms et 226ms respectivement) est conservé chez les patients atteints de maladie de Parkinson et d'ataxies cérébelleuses respectivement. Le pattern est moins visible chez les patients atteints de SLA. Cela est lié à deux facteurs importants propres à ce groupe : (1) il s'agit de la po-

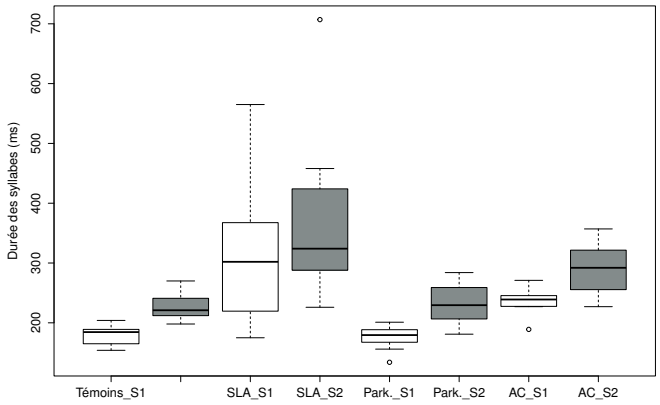


FIGURE 5.9 – *Durées de la première (S1) et de la deuxième (S2) syllabe pour les population du corpus TypALoc.*

pulation présentant les degrés dysarthriques les plus sévères (2) elle présente le plus de variabilité en termes de sévérité des patients.

Suite à la syllabification, la décision de normalité prise par notre approche de détection automatique d'anomalies au niveau phonème est rapportée au niveau syllabe. En effet, si une anomalie est observée sur un phonème dans la première syllabe (deuxième syllabe respectivement), cette syllabe est considérée comme anormale. Le but étant d'étudier l'effet du pattern iambique sur la détection automatique d'anomalies.

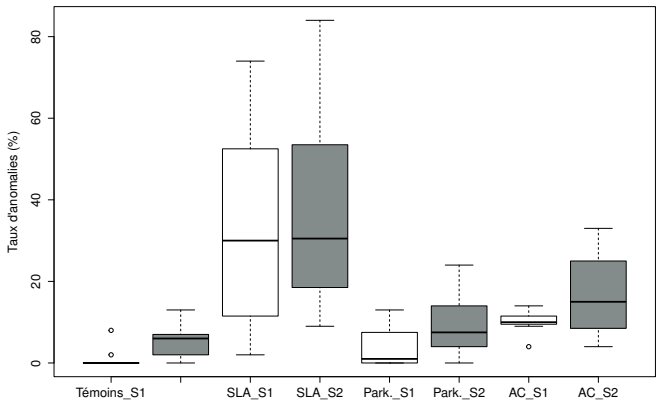


FIGURE 5.10 – *Taux d'anomalies sur la première (S1) et la deuxième (S2) syllabe pour les populations du corpus TypALoc.*

En observant la figure 5.10, une différence nette des taux d'anomalies détectées apparaît entre les deux syllabes. En effet, beaucoup plus d'anomalies sont retrouvées sur les deuxièmes syllabes des locuteurs contrôles et des patients atteints de la maladie de

Parkinson et d'ataxie cérébelleuse. Par contre, la population SLA montre un comportement différent, identique à celui observé au niveau des durées des syllabes (figure 5.9). De plus, cette tendance est généralisée sur toutes les syllabes indépendamment de leurs noyaux (la voyelle de la syllabe). La figure 5.11 trace les taux d'anomalies pour les différentes populations et les types de syllabes du corpus *TypALoc*.

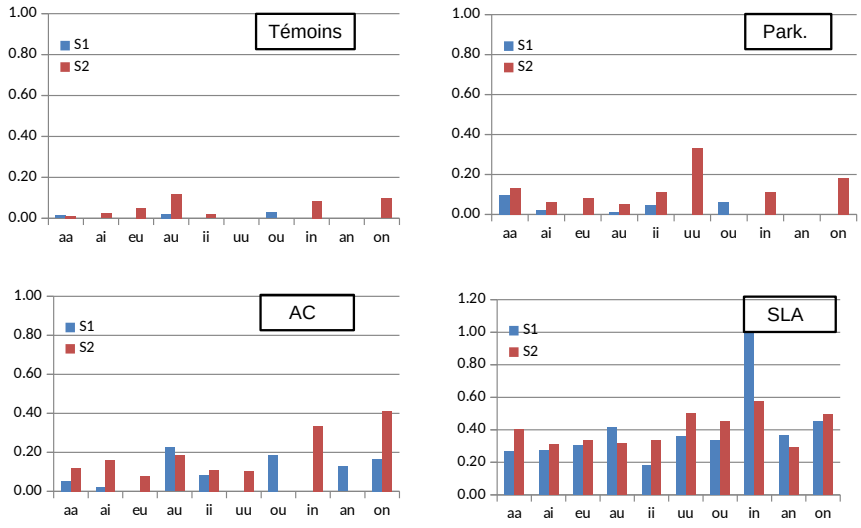


FIGURE 5.11 – *Taux d'anomalies détectées par localisation et type de syllabes pour les différentes populations du corpus TypALoc.*

Cela nous pousse à conclure que l'approche automatique a tendance à détecter plus d'anomalies quand le pattern iambique est respecté. Cela peut être expliqué notamment par la plus longue durée des phonèmes de la deuxième syllabe vu par le système comme un comportement caractérisant plus les phonèmes anormaux que les phonèmes normaux. De plus, aucun contrôle n'a été effectué sur la localisation des phonèmes (première ou deuxième syllabe) utilisés dans la phase d'apprentissage de modèles normal et anormal du classifieur. Une deuxième hypothèse peut être avancée pour expliquer ce comportement. En effet, bien que notre hypothèse de départ s'appuyait sur le rallongement de la deuxième syllabe des mots bisyllabiques comme caractère de renforcement, les travaux de Cécile Fougeron dans (Fougeron, 1998) rapportent une tendance contraire. Dans ces travaux, un renforcement articulatoire local des consonnes en position initiale a été mis en évidence sur des mots polysyllabiques. Un tel renforcement peut expliquer la détection de moins d'anomalies sur la première syllabe que sur la deuxième syllabe par l'approche proposée. Ces résultats très préliminaires devront faire l'objet d'une étude plus approfondie permettant de mieux comprendre l'effet du pattern iambique sur l'approche de détection automatique d'anomalies.

5.5 Conclusion

Ce chapitre nous a permis dans un premier temps de présenter l'approche proposée pour l'annotation des anomalies au niveau phonème dans la parole dysarthrique. L'observation du comportement du système sur le corpus *VML* a montré la capacité de l'approche à détecter les phonèmes déviants vis-à-vis d'une parole normale. Sa précision reste néanmoins à améliorer : elle est plus sévère que l'expert humain dans son jugement. Les faibles taux d'anomalies détectées sur les locuteurs contrôles valident aussi la pertinence des paramètres utilisés pour la caractérisation des phonèmes lors de la classification. L'évaluation de l'approche sur le corpus *DesPhoAPaDy* contenant trois pathologies et classes dysarthriques différentes a confirmé la robustesse de son comportement face à des pathologies non utilisées dans l'apprentissage du classifieur. Aussi, la méthode s'est révélée capable de capter l'évolution de la sévérité de la dysarthrie et peut donc être envisagée dans le cadre des suivis longitudinaux des patients. En effet, les mesures de corrélation élevées (entre 0.8 et 0.9) relevées entre les taux de phonèmes annotés comme déviants et les évaluations perceptives du degré de sévérité, d'intelligibilité et de troubles de l'articulation des patients par le jury d'experts valident la possibilité d'interpoler ces annotations locales d'anomalies pour fournir une estimation automatique de ces mesures.

Une comparaison entre la parole lue et spontanée a été réalisée sur le corpus *TypALoc*. L'observation du comportement de l'approche a montré un effet de la tâche et du style de parole lié aux pathologies des patients. La SLA, population la plus dysarthrique, présente, contrairement à toutes les autres populations, plus d'anomalies sur la parole lue que sur la parole spontanée. De plus, les contrôles ont montré la progression la plus importante des taux d'anomalies entre les deux styles de parole. L'hypothèse avancée est que les contrôles adaptent leurs productions selon le style de parole, ce qui résulte en des productions considérées comme atypiques (mais non pathologiques) par le système automatique dans la parole spontanée. Par contre, les patients les plus dysarthriques perdent progressivement cette capacité et montrent une "uniformisation" de leurs productions à travers les différents styles de parole.

Nous avons aussi étudié le processus de détection d'anomalies en fonction de la précision de l'alignement automatique. Deux comportements ont été observés : (1) une erreur d'alignement importante sur un phonème favorise sa labellisation comme anormal par l'approche. Certains de ces décalages sont notamment dus à de vraies anomalies annotées aussi par l'expert humain (2) 54% des anomalies correctement détectées par l'approche ne sont pas liées à des erreurs d'alignement ce qui confirme sa capacité à caractériser et relever les distorsions acoustiques dans la parole dysarthrique.

Une dernière étude préliminaire portant sur la localisation des anomalies dans des mots bisyllabiques a été proposée. Nous avons observé qu'à l'exception des SLA, toutes les populations présentent plus d'anomalies sur la deuxième syllabe indépendamment de son noyau (sa voyelle correspondante). Ce résultat contradictoire avec notre hypothèse de départ nécessite une étude plus approfondie afin de mieux comprendre le comportement de l'approche automatique sur ces mots.

Des interrogations persistent sur la relation entre l'approche proposée et le système auditif humain. L'approche proposée est elle capable de détecter tout type d'anomalie ? Et inversement, l'appareil auditif humain présente-t-il le même comportement que le système face à la variabilité de la parole dysarthrique ?

Pour tenter de répondre à cette question, le chapitre suivant détaillera les résultats d'une campagne d'évaluation perceptive par des jurys d'experts des sorties de l'approche automatique réalisée dans le cadre d'un mémoire d'orthophonie co-encadré par le LPL et le LIA auquel nous avons activement participé.

Chapitre 6

Évaluation perceptive de l'approche de détection automatique d'anomalies dans la parole dysarthrique

Sommaire

Dans ce chapitre, nous présenterons les résultats d'une campagne d'évaluation perceptive de la fiabilité de l'approche de détection automatique d'anomalies dans la parole dysarthrique. Ce travail a été réalisé dans le cadre d'un mémoire d'orthophonie (Pianelli et Restivo, 2016) sur lequel nous sommes activement intervenus.

6.1 Protocole

La première question posée lors de ce travail était le niveau de granularité des séquences de parole utilisées lors de l'évaluation. En effet, les résultats obtenues dans (Piro et Ziamni, 2014) dans le cadre d'une évaluation perceptive conduite au niveau phonème ont souligné un certain nombre de contraintes et limites de ce type de protocole d'évaluation. En effet, dans cette évaluation, les jurés ont exprimé beaucoup de difficultés dans la localisation des phonèmes anormaux. Ces problèmes étaient notamment liés à l'effet de la coarticulation et la tendance de l'appareil auditif humain à réaliser une évaluation plus globale de la parole. Cette réflexion nous a amené à choisir un

niveau de granularité plus important que celui des phonèmes. Le niveau "mot" a alors été retenu pour ce deuxième protocole d'évaluation perceptive.

L'annotation automatique des anomalies par l'approche proposée étant réalisée au niveau phonème (comme vu dans le chapitre 5), une extrapolation de cette annotation au niveau mot a alors été nécessaire pour la suite de l'évaluation. Lors de cette transition, la normalité ou non d'un mot dépendra à la fois du nombre de syllabes le constituant et du nombre de phonèmes "anormaux" détectés par l'approche automatique. Ainsi, un mot est considéré "anormal' :

— s'il est monosyllabique et au moins un de ses phonèmes a été annoté comme anomalie par l'approche automatique ;
— s'il contient deux syllabes ou plus et au moins deux de ses phonèmes ont été annotés comme anomalies par l'approche automatique.

Dans ce cadre et pour faciliter le choix des patients et des segments à inclure dans l'étude, une représentation sous forme de cartographies de l'annotation automatique au niveau mots a été générée. Dans ces cartographies, le texte lu par les patients est segmenté en phrases (colonne) et chaque phrase est segmentée en mots (cases). La couleur de chaque case (représentant un mot) reflète son annotation : (1) blanc= mot normal (2) jaune= un mot de deux syllabes ou plus où un seul phonème est anormal (3) rouge= mot anormal.

Des cartographies plus globales reflétant les mots anormaux de chaque population ont aussi été générées pour faciliter le choix des locuteurs et des phrases à utiliser dans cette évaluation. Ces cartographies reflètent, pour chaque population étudiée, le nombre de fois où chaque mot a été annoté comme anormal par l'approche automatique.
Les figures 6.1 et 6.2 présentent des exemples de ces cartographies.

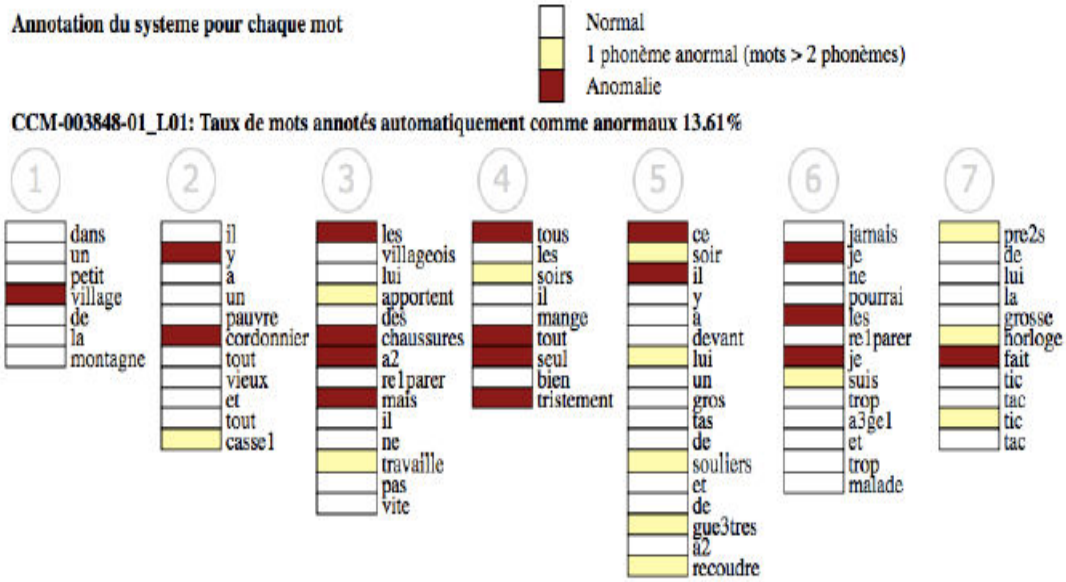


FIGURE 6.1 – *Exemple de cartographie d'anomalies au niveau mot.*

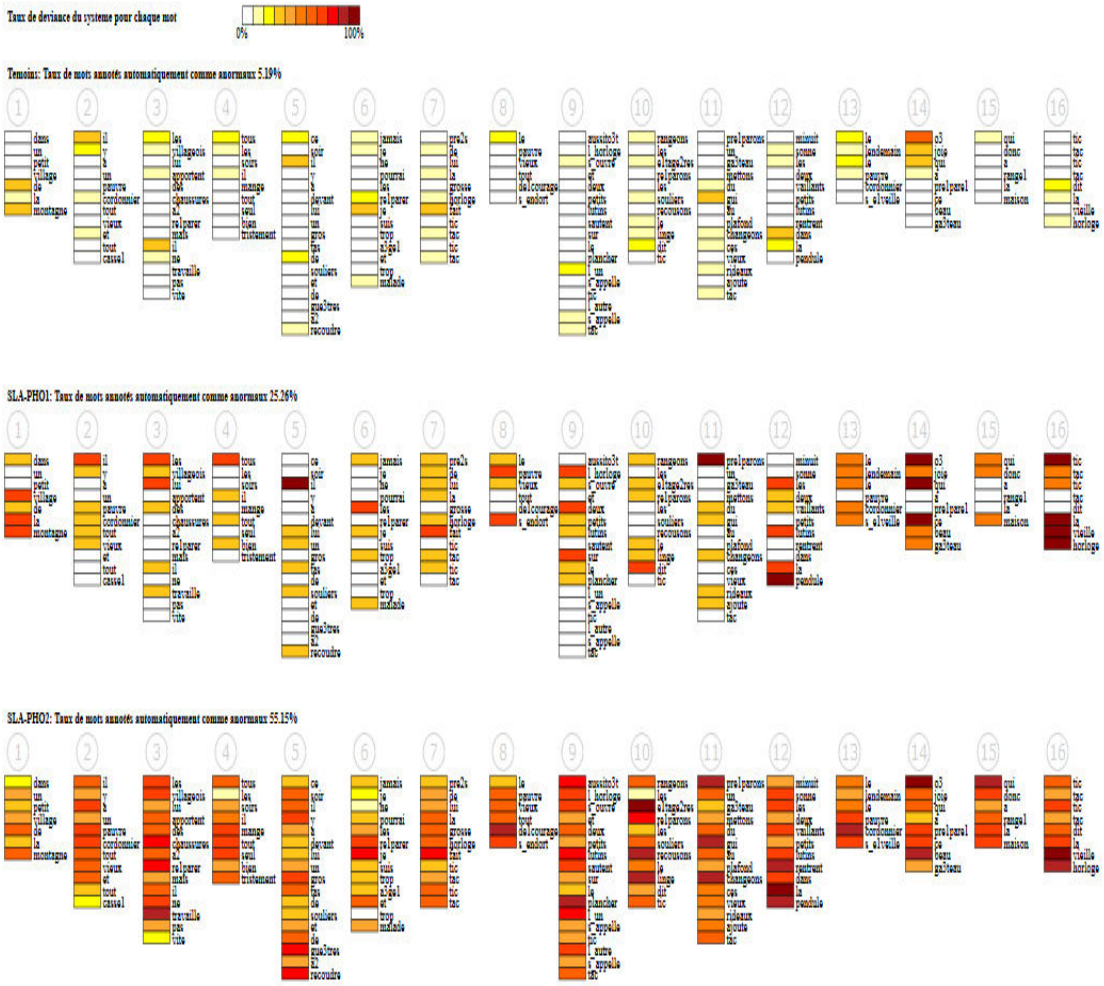


FIGURE 6.2 – *Exemple de cartographie d'anomalies au niveau mot pour la population SLA. Les trois niveaux correspondent aux contrôles et aux deux premiers grades de sévérité.*

6.1.1 Corpus

Lors de la sélection des patients à inclure dans cette étude, un choix d'inclusion de plusieurs pathologies dysarthriques a été effectué afin d'apporter le plus de généralisation et de robustesse aux résultats trouvés. De ce fait, des patients souffrant des différentes pathologies présentes dans nos corpus ont été inclus dans cette évaluation. De plus, et afin de contourner le phénomène d'habituation à la parole souvent observé lors de ce type d'évaluation, des locuteurs contrôles ont aussi été inclus à l'étude. Dans l'ensemble, 41 locuteurs issus des corpus *TypALoc* et *VML* ont été retenus.

À partir des différents enregistrements utilisés, plusieurs séquences de parole ont été sélectionnés. Chacune de ces séquences comporte un ou plusieurs mots cibles pour

l'analyse. Par exemple, dans le passage "des **chaussures** à réparer", "chaussures" est le mot ciblé par l'évaluation. Le tableau 6.1 détaille quelques informations par rapport à ces locuteurs notamment le nombre de séquences de parole utilisées pour chacun.

Le choix des mots cibles a été fait en deux étapes. Premièrement, en se basant sur les cartographies représentant chaque population, une présélection des mots les plus fréquemment annotés comme anormaux pour chaque population a été faite. Ensuite, sur la base de concordances ou non de l'annotation du système avec celle des deux élèves orthophonistes, 3 catégories de mots ont été établies :

— les concordances (représentant 50% des séquences retenues) : cette catégorie correspond aux échantillons où les deux annotations (automatique et perceptive) sont d'accord sur la présence ou non d'une anomalie. Cette catégorie est elle même divisée en deux sous catégories (1) les concordances évidentes (25% des cas) (2) les concordances douteuses (75% des cas) où l'accord est plus discutable et les deux orthophonistes n'étaient pas d'accord sur l'évaluation à donner ;
— les faux positifs (représentant 25% des séquences retenues) : cette catégorie correspond au cas où un mot est détecté comme anormal par le système mais pas par les orthophonistes ;
— les faux négatifs (représentant 25% des séquences retenues) : cette catégorie correspond au cas où une anomalie est perçue par les orthophonistes mais non détectée par l'approche automatique.

Aussi, des critères de variabilité de nature (mots grammaticaux/mots lexicaux), de longueur (long/court) et de position dans la phrase (début/milieu/fin) ont été considérés lors du choix des séquences à utiliser dans l'évaluation. Dans l'ensemble, 98 séquences ont été sélectionnés.

L'évaluation des différentes séquences a été faite par un jury d'experts. Les membres de ce jury devaient avoir le français comme langue maternelle et ne présenter aucun problème d'audition ou de l'apprentissage. Le jury a été composé de 32 membres : 18 étudiants en dernière année d'orthophonie, 13 orthophonistes et 1 ORL/Phoniatre. Chaque juré avait la possibilité d'écouter chaque séquence jusqu'à trois fois. L'évaluation a été réalisée avec le logiciel Perceval (Ghio et al., 2003).

6.2 Résultats et discussions

L'évaluation du système a reposé sur le calcul de deux taux de concordances entre ses annotations automatiques et celles du jury d'experts. Le premier taux mesure l'accord entre les deux annotations sur les mots labellisés comme anormaux par le système, le deuxième sur ceux jugés normaux.

La figure 6.3 détaille les taux d'accord système-jury sur les mots détectés comme déviants. Cette distribution montre une large hétérogénéité dépendant de la catégorie de l'anomalie. Ce taux s'élève à 81% pour les concordances évidentes, 53% pour les concordances douteuses, 35% pour les faux négatifs et 11% pour les faux positifs.

TABLE 6.1 – *Les locuteurs utilisés dans l'évaluation perceptive.*

Populations	Locuteurs	Nombre de séquences de parole
Maladie de Parkinson (*TypALoc*)	CCM-001773-01	3
	CCM-003130-01	4
	CCM-003148-01	2
	CCM-003346-01	2
	CCM-003733-01	2
	CCM-003848-01	2
Ataxie cérébelleuse (*TypALoc*)	CCM-002710-01	3
	CCM-003094-01	3
	CCM-003110-01	2
	CCM-003493-01	1
	CCM-003998-01	4
	CCM-004523-01	4
	CCM-004538-01	1
	CCM-004773-01	4
SLA (*TypALoc*)	PHO-000024-01	2
	PHO-000566-01	1
	PHO-000814-01	2
	PHO-001070-01	1
	PHO-001329-01	5
	PHO-001473-01	2
	PHO-001499-01	4
	PHO-001522-01	1
	PHO-001594-01	5
	PHO-001670-01	4
	PHO-307175-01	1
Maladies lysosomales (*VML*)	PSN-GALE00-01	2
	PSN-GHEL00-01	3
	PSN-GSAN00-01	3
	PSN-NCHR00-01	4
	PSN-NDIA00-01	1
	PSN-NGHI00-01	2
	PSN-NRON00-01	1
	PSN-NVAL00-01	2
Contrôles (*TypALoc*)	AEX-CSR000-01	1
	BEX-CEB000-01	5
	BEX-CDB000-01	2
	BEX-CHE000-01	2
	BEX-CKN000-01	2
	BEX-CMB000-01	2
	BEX-CNKN00-01	1

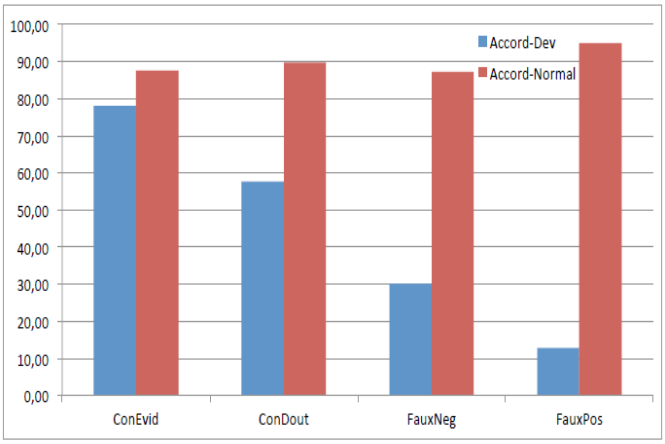


FIGURE 6.3 – *Taux d'accord système-jury pour les mots annotés déviants (Accord-Dev) et ceux annotés normaux (Accord-Normal)*

Le taux d'accord élevé sur les concordances évidentes reflète une nouvelle fois la capacité du système à détecter les segments où les déviances apparaissent clairement. Par contre, le faible taux d'accord observé sur les faux positifs révèle les limites de l'approche et son comportement, parfois arbitraire, dans la recherche d'anomalies plus subtiles. Ce résultat appelle à une analyse acoustique approfondie de ces échantillons afin de mieux comprendre le comportement du système automatique. Néanmoins, l'accord système-jury sur les mots jugés normaux qui s'élève à 88% sur l'ensemble des échantillons est plutôt rassurant sur le comportement du système face aux faux positifs qui reste marginal et non généralisé sur toutes les réalisations normales de parole.

Les taux d'accord mesurés sur les deux autres catégories mettent l'accent une fois de plus sur la difficulté de la tâche d'évaluation perceptive de la parole dysarthrique. Ici, une réponse sur deux du jury et d'accord avec le système sur les anomalies douteuses. Aussi, un tiers des réponses portant sur la classe des faux négatifs sont en accord avec le jugement du système sur la présence d'une anomalie.

Une analyse des résultats de l'évaluation par population montre que les jurés, et malgré leur niveau d'expertise dans l'évaluation de la parole pathologique, sont influencés par les traits acoustiques et la qualité globale de la parole. Ceci est important à souligner compte tenu du fait que les consignes de l'évaluation mentionnaient explicitement la restriction de l'évaluation à la seule réalisation articulatoire des locuteurs. Cette tendance est particulièrement claire au niveau des locuteurs atteints de SLA où les jurés ont annoté le plus d'anomalies par rapport à toutes les autres populations et où le taux d'accord système-jury sur la catégorie des faux positifs atteint 20%. En effet, la dysarthrie liée à cette pathologie est caractérisée par une hypernasalité, une raucité et une lenteur générale de la parole. Cela a résulté en une tendance des jurés à annoter plus d'anomalies qu'attendu sur cette population. Un comportement opposé peut être observé sur les populations contrôles ainsi que les patients atteints de la maladie de Parkinson. Sur ces populations, la qualité globale de la parole décourage les jurés à annoter des anomalies ce qui résulte simultanément en des scores d'accord système-jury

faibles sur les anomalies (37% et 41% respectivement) et élevés sur les mots normaux (99% et 93% respectivement).

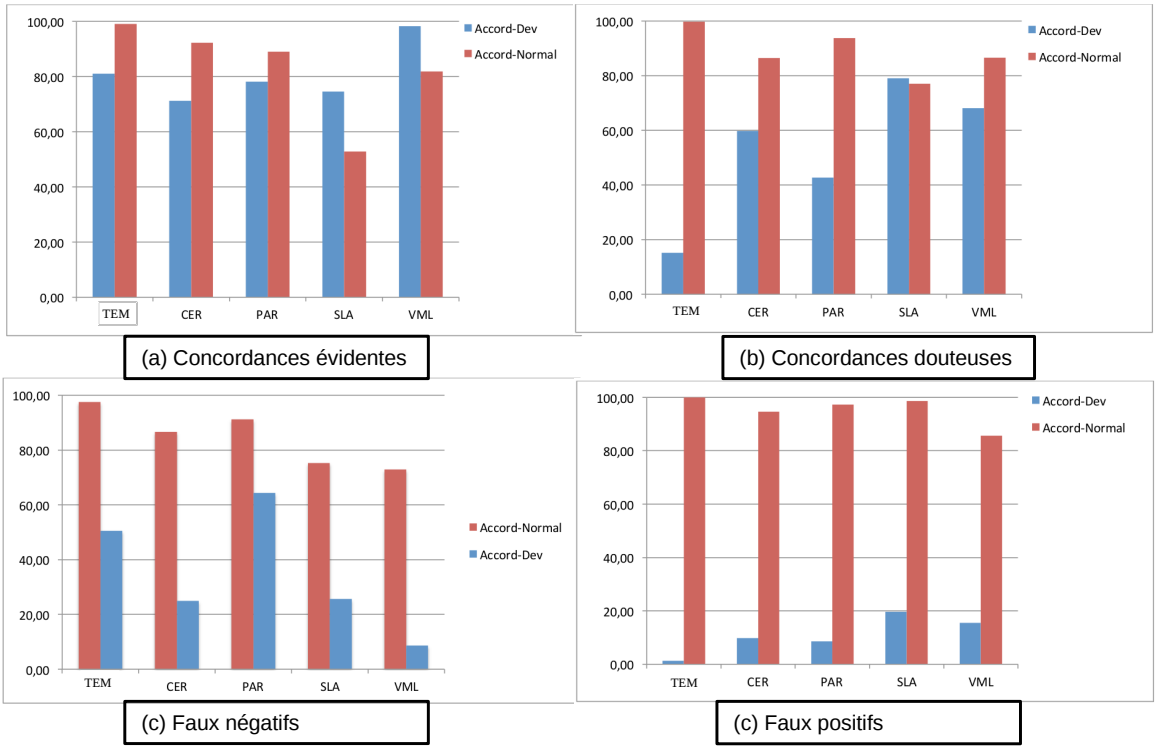


FIGURE 6.4 – *Taux d'accord système-jury pour les mots annotés déviants (Accord-Dev) et ceux annotés normaux (Accord-Normal) par catégorie.*

Nous pouvons aussi noter que le meilleur taux d'accord système-jury sur les anomalies est relevé sur les patients atteints de maladies lysosomales (68%) atteignant même 98% sur la catégorie des concordances évidentes. Il s'agit de la population utilisée lors de l'apprentissage de notre classifieur de phonèmes normaux et anormaux (5.1.2). Ce comportement rejoint les observations émises précédemment sur la performance de l'approche sur ces patients et souligne l'importance de la phase d'apprentissage du système. L'élargissement des données d'apprentissage à d'autres pathologies et classes dysarthriques devrait permettre d'améliorer les résultats sur ces populations bien qu'ils soient déjà très prometteurs au vu des résultats et comparaisons rapportés dans les chapitres précédents.

Une dernière analyse de ces résultats a porté sur la variabilité inter-jury. Ici, la variabilité a été évaluée en termes de disparité de taux de mots annotés comme anormaux par chaque juré (allant de 7.5% à 40%). Ce comportement souligne une nouvelle fois la difficulté de la tâche d'évaluation perceptive de la parole dysarthrique même lorsqu'elle est réalisée par des experts sur des segments assez longs (mots). Cette varia-

bilité soutient une nouvelle fois le besoin qui a motivé le travail présenté ici visant à automatiser et objectiver l'évaluation de la parole dysarthrique.

Nous pouvons aussi noter qu'une sélection d'un sous-groupe de 7 jurés dont les annotations ont été plus homogènes a permis d'atteindre des taux de concordances système-jury plus élevés sur les anomalies issues des catégories de concordances évidentes et concordances douteuses (95% et 64% respectivement).

6.3 Conclusion

Ce chapitre a permis de présenter les résultats d'une évaluation perceptive des sorties du système de détection automatique des anomalies. Cette évaluation a permis de mettre en avant les limites de cette approche, notamment au niveau des faux positifs détectés par le système Néanmoins, elle a permis de montrer également que l'approche présente des taux d'accord élevés avec le jury d'experts pour les anomalies importantes (concordances évidentes). De plus, et même sur les anomalies plus nuancées (concordances douteuses), le jury d'experts a été d'accord une fois sur deux avec l'approche automatique. Ces tendances réaffirment une nouvelle fois la capacité de l'approche à détecter les anomalies (surtout les plus importantes) et le caractère sévère de son annotation.

Des questions relatives à la détection des faux positifs par l'approche demeurent ouvertes. Plusieurs hypothèses sur les causes de ce comportement peuvent être avancées : (1) des altérations dans la qualité des enregistrements (2) la présence de vraies anomalies non détectées par l'appareil auditif humain (3) la présence de données erronées dans le corpus d'apprentissage (erreur dans l'annotation humaine utilisée comme référence). Ces hypothèses sont motivées aussi bien par les limites reconnues dans la littérature de l'évaluation perceptive de la parole dysarthrique que par la large variabilité inter-jury observée lors de la campagne d'évaluation décrite précédemment. La présence d'erreurs d'évaluation à la fois dans l'annotation de l'expert utilisée pour l'apprentissage du système et dans l'annotation du jury lors de l'évaluation devient alors fortement probable.

Un travail de caractérisation des anomalies est encore nécessaire afin de mieux les différencier des productions normales. Aussi, une réflexion sur la granularité de l'évaluation est nécessaire. Bien que nous soutenons toujours l'intérêt et l'utilité d'une évaluation plus locale et précise que les mesures d'intelligibilité et de sévérité globale de la dysarthrie, l'évaluation au niveau phonème s'est révélée être une tâche très complexe et parfois inadaptée à la fois pour l'approche automatique et lors des évaluations perceptives.

Chapitre 7

Conclusions et perspectives

Tout au long de ce document, nous avons présenté notre travail portant sur l'apport des outils du traitement automatique de la parole dans le cadre de la parole dysarthrique, le but final du travail étant de proposer une approche de détection automatique de phonèmes anormaux dans la parole dysarthrique.

La première partie de ce document a permis de présenter le processus de production de la parole et les différents composants qui y interviennent. Nous avons par la suite défini la dysarthrie, les différents types de classification proposés dans la littérature et les différentes altérations de la parole qui en résultent. Cette présentation visait à mettre en avant la large variabilité observée dans la parole dysarthrique liée à la fois au type et à la sévérité de la dysarthrie. Cette variabilité inter-pathologique, conjuguée à la variabilité intra et inter-locuteur indépendante de la dysarthrie, rend la tâche de caractérisation de la parole dysarthrique et son traitement automatique très complexe.

La deuxième partie du document a été consacrée à la présentation des outils de traitement automatique de la parole proposés et utilisés dans ce travail. Dans le chapitre 4, nous avons présenté l'outil d'alignement automatique de la parole au niveau phonème et étudié son comportement face à la parole dysarthrique. Cette étude a permis d'observer l'effet de la sévérité de la dysarthrie sur la qualité de l'alignement automatique réalisé. Sur la parole lue, la corrélation entre le taux d'erreurs d'alignement et l'évaluation perceptive de la sévérité de la dysarthrie a atteint 0.82 sur les corpus *VML* et *TypALoc*. Une corrélation importante entre le taux d'erreurs d'alignement et l'évaluation perceptive du débit de la parole a aussi été relevée atteignant -0.73. Une analyse plus fine de ces erreurs d'alignements au niveau phonétique a aussi révélé des différences importantes entre les populations : les patients atteints d'ataxie cérébelleuse (dysarthrie ataxique) et de SLA (dysarthrie mixte) sont sujets à plus d'erreurs d'alignement sur les voyelles alors que ceux atteints de maladie de Parkinson (dysarthrie hypokinétique) montraient un important taux d'erreurs sur les consonnes fricatives.

En se basant sur le corpus *TypALoc* contenant des enregistrements de parole lue et spontanée, un effet intéressant du style de la parole sur la qualité de l'alignement a pu être observé. En effet, moins d'erreurs d'alignement ont été observées sur la parole

spontanée par rapport à la parole lue. Ce comportement est plus visible chez les patients atteints de dysarthries légères que chez les patients sévèrement dysarthriques. Cette observation est très intéressante puisqu'elle peut refléter le développement de stratégies de compensation et d'évitement par les patients dysarthriques appliquée au niveau de la parole spontanée et résultant en un meilleur alignement de cette parole.

Dans le chapitre 5, cet outil d'alignement automatique de la parole a été mis au service d'une approche de détection automatique de phonèmes anormaux dans la parole dysarthrique. Cette approche s'appuie sur les scores de vraisemblance issus des deux phases d'alignement de parole ainsi que sur des informations liées à la catégorie phonétique des phonèmes pour la détection des phonèmes anormaux. L'observation du comportement de cette approche sur les corpus *VML*, *DesPhoAPady* et *TypALoc* a révélé différentes tendances :

— la capacité de l'approche proposée à détecter les phonèmes déviants dans la parole dysarthrique (mesure de *AnRappel* moyen de 0.74 sur le corpus *VML*). La précision de l'approche reste tout de même en deçà des attentes (*AnPrec* moyen de 0.61 sur le corpus *VML*), ce taux reflète un comportement "sévère" de l'approche qui détecte plus d'anomalies que l'expert. Différentes hypothèses peuvent être avancées pour expliquer ce comportement (bruit dans le signal, altération globale de la parole, erreurs d'alignements, erreurs dans les données d'apprentissage, etc.). Néanmoins, le faible taux d'anomalies détectées sur les locuteurs témoins (6.9% et 6.6% sur les témoins issus des corpus *VML* et *DesPhoAPady* respectivement) nous conforte sur la pertinence des paramètres utilisés pour la caractérisation des phonèmes lors de la classification ;

— la capacité de l'approche à capter l'évolution de la sévérité de la dysarthrie. En effet, des corrélations importantes ont été mesurées entre les taux d'anomalies automatiquement détectées et les évaluations perceptives de sévérité, d'intelligibilité et de troubles de l'articulation des différents patients du corpus *DesPhoAPady* (mesures entre 0.8 et 0.9). Ce comportement de l'approche valide sa pertinence face à des pathologies et types de dysarthrie non utilisées lors de la phase d'apprentissage des modèles ainsi que son éventuel apport dans le cadre du suivi longitudinal des patients. Aussi, ces mesures de corrélation suggèrent l'utilisation de l'approche pour la prédiction et l'évaluation de la sévérité et de l'intelligibilité de la parole dysarthrique.

L'étude de la détection automatique d'anomalies en fonction de la qualité de l'alignement automatique de la parole a permis de dégager deux comportements distincts en fonction de la présence ou non d'une erreur d'alignement. En effet, les erreurs d'alignement importantes favorisent la détection d'anomalies sur les phonèmes concernés. Certaines de ces erreurs sont tout de même liées à de vraies anomalies annotées aussi par l'expert. Par contre, 54% des anomalies correctement détectées par l'approche sont sur des phonèmes bien alignés. Cela confirme une nouvelle fois la capacité de l'approche à capter différentes formes de distorsions acoustiques dans la parole dysarthrique.

Une dernière étude proposée dans le chapitre 5 a porté sur l'effet du style de parole sur l'approche de détection d'anomalies. Ici, les patients atteints de SLA se sont distin-

gués des autres populations avec moins d'anomalies détectées sur la parole spontanée par rapport à la lecture. Par contre, les locuteurs témoins ont été ceux montrant la différence la plus importante entre les deux styles de parole (8.3% d'anomalies de plus en moyenne sur la parole spontanée par rapport à la parole lue). Une hypothèse avancée pour expliquer ces différences suppose que les témoins adaptent leurs productions selon le style de parole, ce qui résulte en des productions considérées comme atypiques (mais non pathologiques) par le système automatique dans la parole spontanée. Par contre, les patients les plus dysarthriques perdent progressivement cette capacité et subissent une "uniformisation" de leurs productions à travers les différents styles de parole.

Finalement, nous avons proposé une étude préliminaire portant sur l'effet du pattern iambique des mots bisyllabiques sur l'approche de détection automatique d'anomalies. Ce caractère de renforcement de la deuxième syllabe dans ces mots (résultant en un allongement de la deuxième syllabe) a eu un effet inattendu sur l'approche résultant en la détection de plus d'anomalies sur la deuxième que sur la première syllabe. Cette tendance est généralisée sur tous les phonèmes et les populations étudiées (à l'exception de la population SLA). Ce comportement, pouvant être lié à la longueur de ces syllabes, nécessite une étude plus approfondie pour l'expliquer.

Dans le chapitre 6, les résultats d'une campagne d'évaluation de l'approche de détection automatique d'anomalie réalisée dans le cadre d'un mémoire d'orthophonie auquel nous avons participé sont rapportés. Cette campagne consistait à faire évaluer les annotations issues de l'approche automatique, interpolées au niveau mot, par un jury de 32 experts. Des résultats mitigés ont été observés. En effet, bien que la capacité de l'approche à détecter les anomalies importantes a été validée avec un taux d'accord système-jury sur ces anomalies de 81%, son comportement face aux anomalies jugées "douteuses" est moins fiable avec un taux d'accord système-jury de seulement 53%. Néanmoins, et sur des anomalies décrites par des experts comme douteuses, le jugement perceptif du jury est forcément douteux. Il devient alors pertinent de s'interroger, dans ce contexte, de la véracité des décisions émises par l'approche automatique et le jury. Aussi, et sur les mots annotés comme normaux par l'approche mais jugés comme déviants par les deux élèves orthophonistes, le jury était d'accord avec l'approche dans 35% des cas. Ces taux reflètent bien les limites de l'approche automatique proposée mais également celles de l'évaluation perceptive des experts. Bien que les consignes de l'évaluation demandaient au jury de se limiter aux altérations acoustiques dans la production, les résultats semblent montrer que les juges se sont laissés influencer par la qualité globale de la parole et le type de la dysarthrie

Le faible taux d'accord système-jury sur les déviances jugées comme "faux positives" (11%) est tout de même sans équivoque et nécessite plus d'analyses sur ces segments afin d'y déceler les causes de leur détection comme anomalies par l'approche. Ces fausses anomalies sont-elles liées à des erreurs de l'approche ou à la présence de vraies altérations acoustiques non détectées par l'appareil auditif humain ? Un important travail de caractérisation des anomalies est encore nécessaire afin de répondre à cette question.

En reprenant les problématiques émises en introduction de ce travail et de ce document, nous pouvons conclure que :

— **Comportement des outils de TAP :** le comportement de l'outil d'alignement automatique de la parole et de l'approche proposée pour la détection des phonèmes anormaux a révélé une variabilité dépendante de la sévérité et de la classe dysarthrique des patients. Plus d'erreurs d'alignement et d'anomalies ont été détectées sur les patients atteints de dysarthrie sévère. Aussi, plus d'anomalies ont été détectées sur les patients atteints de SLA et un comportement intéressant a été observé sur les patients atteints de dysarthrie ataxique pouvant refléter un effet caractéristique du débit. Des différences entre les styles de parole ont aussi été relevées. Plus d'erreurs d'alignement ont été commises sur la parole lue. De plus, à l'exception des patients atteints de SLA, plus d'anomalies ont été détectées sur la parole lue que sur la parole spontanée.
— **Capacité de l'approche :** l'approche automatique proposée s'est révélée capable de déceler automatiquement les phonèmes anormaux dans la parole dysarthrique. Cette capacité a été mise en évidence en étudiant les taux de $AnRappel$ mesurés sur le corpus *VML* ainsi que lors de l'évaluation perceptive de l'approche automatique proposée par un jury d'experts présentée dans le chapitre 6.
— **Système et perception :** le comportement de l'approche automatique diverge de celui des experts. En effet, la campagne d'évaluation perceptive menée à la fin de ce travail a montré que les deux méthodes présentent des comportements similaires sur les phonèmes dont les déviances sont importantes. Cependant, des divergences sont observées sur les anomalies qualifiées de douteuses et l'approche automatique a détecté plus d'anomalies que l'expert humain sur le corpus *VML*. Ces différences, à notre avis, ne mettent pas en doute la capacité et l'intérêt de l'approche proposée mais reflètent la difficulté de la tâche d'annotation des anomalies et l'enrichissement que peut être une annotation automatique. Nous pouvons même avancer que l'approche n'a pas forcément à répliquer le comportement des experts humains dont les limites et la subjectivité ont été observées lors de cette étude.
— **Pertinence :** les phonèmes détectés comme anormaux par l'approche automatique présentent un intérêt indéniable pour les phonéticiens. Dû à la difficulté de cette tâche d'annotation au niveau phonème pour les phonéticiens, une annotation automatique permettra d'avoir une première référence pour l'étude des similitudes et des différences entre les comportements du système et des experts humain pour une meilleure caractérisation des altérations liées à la dysarthrie au niveau phonème. Aussi, une analyse approfondie des segments détectés automatiquement et non par les experts (considérés comme des faux positifs dans ce travail) peut être utile pour proposer et développer des nouvelles grilles d'évaluation de la parole dysarthrique.

Une des difficultés auxquelles nous étions confrontés dans ce travail est l'annotation humaine au niveau phonème utilisée comme référence. L'implication de plusieurs annotateurs aurait pu apporter plus d'objectivité à ces annotations utilisées dans l'apprentissage des modèles de parole. Cette solution n'était cependant pas possible dans le cadre de notre travail.

Un dernier point nécessitant plus de réflexion est le choix de la granularité de l'unité utilisée dans l'évaluation automatique de la parole. En effet, une de nos motivations au début de ce travail était le besoin d'une évaluation sur des unités courtes permettant un feed-back précis pour l'utilisateur. Nous avons aussi estimé que ce type d'annotation pourra être utilisée dans un deuxième temps pour la prédiction de la sévérité de la dysarthrie et de l'intelligibilité de la parole. Et bien que nous soutenons toujours cette idée surtout que l'approche automatique proposée a montré sa capacité à capter l'évolution de la sévérité de la dysarthrie, le niveau phonème n'est pas nécessairement le mieux adapté pour cette évaluation. Ce constat a été émis dans (Piro et Ziamni, 2014) où les jurés ont exprimé des difficultés à évaluer la parole dysarthrique au niveau phonème. C'est la raison pour laquelle le niveau mot a été retenu dans la campagne d'évaluation décrite dans le chapitre 6. De plus, et malgré le fait que nous avons montré que les anomalies détectées par l'approche ne sont pas liées exclusivement à des erreurs d'alignement, nous estimons que le passage à une granularité plus importante (la syllabe par exemple) pourrait limiter ces erreurs d'alignement (et donc les fausses anomalies qui y résultent) et mieux refléter les anomalies liées à la coarticulation.

Perspectives

Le premier axe de travail que nous proposons est d'affiner les types d'anomalies que nous cherchons à détecter automatiquement. En effet, dans son état actuel, l'approche s'est reposée sur l'annotation des anomalies réalisée par un expert sur un corpus très limité de parole lue (35 enregistrements) associée à une seule pathologie. Toutes les anomalies annotées par l'expert ont été utilisées sans différenciation dans l'apprentissage des modèles. Il peut être intéressant de se limiter à quelques anomalies bien documentés et plus facilement détectables automatiquement (voisement, dévoisement, etc.). L'utilisation de ce type d'approche permettra l'annotation automatique de nouveaux corpus et leurs utilisation pour l'apprentissage de meilleurs modèles reflétant ces anomalies.

Aussi, l'étude de l'approche de détection automatique d'anomalies proposée a révélé un meilleur comportement sur les patients atteints de dysarthries sévères par rapport à ceux souffrant de dysarthries légères. Il peut alors être intéressant d'adapter les modèles de parole normale et anormale utilisés à la sévérité des patients à évaluer. Cette sévérité peut être estimée automatiquement dans un premier temps (sur la base du taux d'anomalies détectées automatiquement par exemple). Cette information relative à la sévérité de la dysarthrie permettra de contrôler la sévérité (le taux global d'anomalies) souhaitée de l'approche dans la deuxième phase de classification. Cependant, une telle approche peut mettre en péril l'éventuelle utilisation du système dans le cadre d'un suivi longitudinal d'un patient. En effet, les modèles utilisés peuvent changer suite à l'évolution de la sévérité dysarthrique d'un patient ce qui pourra causer la perte de la référence d'annotation automatique liée au début du suivi (la première évaluation automatique).

Un autre axe à étudier est l'utilisation des résultats de l'alignement automatique

de la parole différemment pour la détection des phonèmes anormaux. En effet, nous avons montré dans ce travail que les taux de confusion phonémique lors d'un alignement non contraint par le texte dépendent à la fois de la pathologie et de la sévérité de la dysarthrie. Une étude plus fine de ces confusions et leur mise en lien avec les traits acoustiques caractérisant chaque phonème pourra être utile pour la détection des phonèmes déviants et la différentiation entre les anomalies de nature pathologique liées à la dysarthrie et celles propres à l'alignement automatique.

De plus, l'étude du comportement de l'approche proposée ici face à d'autre troubles de parole peut aussi être pertinent pour mesurer sa portabilité et sa possible généralisation face à d'autres types d'altérations de parole. Il sera intéressante dans ce cadre d'élargir l'ensemble des paramètres utilisés afin d'y inclure des paramètres prosodiques (F0) et articulatoires par exemple.

Finalement, il nous parait indispensable de proposer une nouvelle version de cette approche plus ergonomique et dont la portabilité lui permet d'être utilisée par les professionnels dans la pratique clinique. En effet, et malgré les différentes solutions proposées dans la littérature, peu d'outils d'évaluation automatique de la parole pathologique sont utilisés dans la pratique clinique. Cette absence d'utilisation dans des conditions réelles limite la portée des évaluations de ces solutions et le retour d'expérience nécessaire pour leur amélioration et adaptation aux besoins cliniques. Le développement d'une solution facile à utiliser pourrait permettre à la fois une meilleure évaluation des performances des approches proposées et la constitution de larges corpus de parole pathologique (annoté ou non) nécessaires pour l'amélioration de ces approches.

Liste des illustrations

Liste des tableaux

Bibliographie

(An et al., 2015) G. An, D. G. Brizan, M. Ma, M. Morales, A. R. Syed, & A. Rosenberg, 2015. Automatic recognition of unified parkinson's disease rating from speech with acoustic, i-vector and phonotactic features. *Proceedings of Interspeech'15*, Dresden, Allemagne.

(Audibert et al., 2010) N. Audibert, C. Fougeron, C. Fredouille, C. Meunier, & O. Panseri, 2010. Evaluation d'un alignement automatique sur la parole dysarthrique. *Journées d'Etudes sur la Parole*, Mons, Belgique, 1–4.

(Auzou, 2007a) P. Auzou, 2007a. Définition et classifications des dysarthries. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(31)*, 308–323.

(Auzou, 2007b) P. Auzou, 2007b. L'intelligibilité. *Les dysarthries, édition Solal Evaluation, Part II(18)*, 204–209.

(Auzou et al., 2000) P. Auzou, C. Ozsancak, M. Jan, S. Leonardon, J. F. Menard, M. J. Gaillard, F. Eustache, & D. Hannequin, 2000. Evaluation clinique de la dysarthrie : Présentation et validation d'une méthode. *Revue de neurologie 154 (6-7)*.

(Auzou et Rolland-Monnoury, 2006) P. Auzou & V. Rolland-Monnoury, 2006. *Batterie d'évaluation clinique de la dysarthrie*. Édition Ortho.

(Bachman, 1990) L. F. Bachman, 1990. *Fundamental considerations in language testing*. Oxford University Press.

(Barreto et Ortiz, 2008) S. d. S. Barreto & K. Z. Ortiz, 2008. Intelligibility measurements in speech disorders : a critical review of the literature. *Pró-Fono Revista de Atualização Científica 20*(3), 201–206.

(Bertrand et al., 2008) R. Bertrand, P. Blache, R. Espesser, G. Ferré, C. Meunier, B. Priego-Valverde, & S. Rauzy, 2008. Le cid-corpus of interactional data-annotation et exploitation multimodale de parole conversationnelle. *Traitement automatique des langues 49*(3), 1–30.

(Bigi et Hirst, 2012) B. Bigi & D. Hirst, 2012. Speech phonetization alignment and syllabification (sppas) : a tool for the automatic analysis of speech prosody. *Speech Prosody*, 1–4.

(Boersma et al., 2002) P. Boersma et al., 2002. Praat, a system for doing phonetics by computer. *Glot international 5*(9/10), 341–345.

(Boula de Mareüil et al., 2008) P. Boula de Mareüil, B. Vieru-Dimulescu, C. Woehrling, & M. Adda-Decker, 2008. Accents étrangers et régionaux en français. *Traitement Automatique des Langues 49*(3), 135–163.

(Carmichael, 2007) J. Carmichael, 2007. *Introducing objective acoustic metrics for the Frenchay Dysarthria Assessment procedure*. Ph.d. dissertation, University of Sheffield.

(Chandola et al., 2009) V. Chandola, A. Banerjee, & V. Kumar, 2009. Anomaly detection : A survey. *ACM computing surveys (CSUR) 41*(3), 15.

(Christensen et al., 2014) H. Christensen, I. Casanueva, S. Cunningham, P. Green, & T. Hain, 2014. Automatic selection of speakers for improved acoustic modelling : Recognition of disordered speech with sparse data. *Spoken Language Technology Workshop (SLT), 2014*, 254–259.

(Christensen et al., 2012) H. Christensen, S. Cunningham, C. Fox, P. Green, & T. Hain, 2012. A comparative study of adaptive, automatic recognition of disordered speech. *Proceedings of Interspeech'12*, Portland, USA.

(Crevier-Buchman, 2007) L. Crevier-Buchman, 2007. Modélisation du fonctionnement laryngé. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(7)*, 91–99.

(Danel-Brunaud, 2007) V. Danel-Brunaud, 2007. La sclérose latérale amyotrophique et les autres maladies du neurone moteur. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(35)*, 439–447.

(Darley et al., 1969a) F. L. Darley, A. E. Aronson, & J. R. Brown, 1969a. Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research 12*, 462–496.

(Darley et al., 1969b) F. L. Darley, A. E. Aronson, & J. R. Brown, 1969b. Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research 12*, 246–269.

(Darley et al., 1975) F. L. Darley, A. E. Aronson, & J. R. Brown, 1975. *Motor speech disorders*. Philadelphia : W. B. Saunders and Co.

(Davis et Mermelstein, 1980) S. Davis & P. Mermelstein, 1980. Comparison of parametric representation for monosyllabic. word recognition in continuous spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Volume 4, 357–366.

(Defebvre, 2004) L. Defebvre, 2004. Les complications motrices de la dopathérapie chez le malade parkinsonien : sémiologie clinique et modalités d'évaluation. *Thérapie 59*(1), 93–96.

(Defebvre, 2007) L. Defebvre, 2007. La maladie de parkinson et les syndromes parkinsoniens. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(36)*, 364–374.

(Doyle et al., 1997) P. C. Doyle, H. Leeper, A.-L. Kotler, N. Thomas-Stonell, C. O'Neill, M.-C. Dylke, & K. Rolls, 1997. Dysarthric speech : a comparison of computerized speech recognition and listener intelligibility. *Journal of rehabilitation research and development 34(3)*, 309–316.

(Duffy, 2005) J. R. Duffy, 2005. *Motor speech disorders : substrates, differential diagnosis and management*. Motsby- Yearbook, St Louis, 2nd edition.

(Enderby, 1983) P. Enderby, 1983. Frenchay dysarthric assessment. *Pro-Ed, Texas*.

(Eskenazi, 1996) M. Eskenazi, 1996. Detection of foreign speakers' pronunciation errors for second language training-preliminary results. *Proceedings of the Fourth International Conference on Spoken Language, ICSLP 96.*, Volume 3, 1465–1468.

(Eskenazi, 2009) M. Eskenazi, 2009. An overview of spoken language technology for education. *Speech Communication 51*(10), 832–844.

(Eyben et al., 2013) F. Eyben, F. Weninger, F. Gross, & B. Schuller, 2013. Recent developments in opensmile, the Munich open-source multimedia feature extractor. *Proceedings of the 21st ACM international conference on Multimedia*, Barcelone, Espagne, 835–838. ACM.

(Eyben et al., 2010) F. Eyben, M. Wöllmer, & B. Schuller, 2010. Opensmile : the Munich versatile and fast open-source audio feature extractor. *Proceedings of the 18th ACM international conference on Multimedia*, 1459–1462. ACM.

(Ferrand, 2001) L. Ferrand, 2001. La production du langage : Une vue d'ensemble. *Psychologie Francaise 46*, 3–15.

(Ferrier et al., 1992) L. J. Ferrier, N. Jarrell, T. Carpenter, & H. C. Shane, 1992. A case study of a dysarthric speaker using the Dragon Dictate voice recognition system. *Journal for Computer Users in Speech and Hearing 8(1)*, 33–52.

(Fex, 1992) S. Fex, 1992. Perceptual evaluation. *Journal of voice 6*(2), 155–158.

(Fletcher et Munson, 1933) H. Fletcher & W. A. Munson, 1933. Loudness, its definition, measurement and calculation. *Bell System Technical Journal 12*(4), 377–430.

(Fontan, 2012) L. Fontan, 2012. *De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication*. Thèse de Doctorat, University of Toulouse 2, Le Mirail, France (in French).

(Fougeron, 1998) C. Fougeron, 1998. *Variations articulatoires en début de constituants prosodiques de différents niveaux en français*. Thèse de Doctorat, Université Paris.

(Fredouille et Pouchoulin, 2011) C. Fredouille & G. Pouchoulin, 2011. Automatic detection of abnormal zones in pathological speech. *Intl Congress of Phonetic Sciences (ICPHs'11)*, Hong Kong.

(Furui, 1986) S. Furui, 1986. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP'86)*, Volume 34, 52–59.

(Galliano et al., 2005) S. Galliano, E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre, & G. Gravier, 2005. ESTER phase II evaluation campaign for the rich transcription of French broadcast news. *Proceedings of Interspeech'05*, 1149–1152.

(Gauvain et Lee, 1994) J. L. Gauvain & C. H. Lee, 1994. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transactions on Speech and Audio Processing 22*, 291–298.

(Georgeton et Meunier, 2016) L. Georgeton & C. Meunier, 2016. Préservation du pattern syllabique iambique dans la production des locuteurs dysarthriques. *Proceedings des JEP-TALN-RECITAL 2016*, Paris, France.

(Ghio et al., 2003) A. Ghio, C. André, B. Teston, & C. Cavé, 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.

(Goldman, 2011) J.-P. Goldman, 2011. EasyAlign : an automatic phonetic alignment tool under praat. *Proceedings of Interspeech'11*, Florence, Italie.

(Grewel, 1957) F. Grewel, 1957. Classification of dysarthrias. *Acta Psychiatrica Scandinavica 32*(3), 325–337.

(Griffin et al., 2000) S. Griffin, L. Wilson, & E. Clark, 2000. Speech pathology applications of automatic speech recognition technology. *8th Australian International Conference on Speech Science and Technology (SST)*, Canberra, Australy.

(Hahm et al., 2015) S. Hahm, D. Heitzman, & J. Wang, 2015. Recognizing dysarthric speech due to amyotrophic lateral sclerosis with across-speaker articulatory normalization. *6th Workshop on Speech and Language Processing for Assistive Technologies*, 47–54.

(Hamidi et Baljko, 2013) F. Hamidi & M. Baljko, 2013. Automatic speech recognition : a shifted role in early speech intervention. *4th workshop on Speech and Language processing for assistive technologies (SLPAT)*, Grenoble, France.

(Harel et al., 2004) B. Harel, M. Cannizzaro, & P. J. Snyder, 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.

(Hawley et al., 2005) M. S. Hawley, P. Green, P. Enderby, S. Cunnigham, & R. K. Moore, 2005. Speech technology for e-inclusion of people with physical disabilities and disordered speech. *Proceedings of Interspeech'05*, Lisbon, Portugal.

(Hermansky, 1990) H. Hermansky, 1990. Perceptual linear predictive (plp) analysis of speech. *Journal of the Acoustical Society of America 87*, 1738–1752.

(Hermansky et Cox, 1991) H. Hermansky & L. Cox, 1991. Perceptual linear predictive (plp) analysis-resynthesis technique. *Applications of Signal Processing to Audio and Acoustics, 1991. Final Program and Paper Summaries., 1991 IEEE ASSP Workshop on*, 0_37–0_38. IEEE.

(Hermansky et al., 1991) H. Hermansky, N. Morgan, A. Bayya, & P. Kohn, 1991. Rasta-plp speech analysis. *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, Volume 1, 121–124. Citeseer.

(Hirano, 1989) M. Hirano, 1989. Objective evaluation of the human voice : Clinical aspects. *Clinical Linguistics and Phonetics 41*, 89–144.

(Ho et al., 1998) A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, & S. Gates, 1998. Speech impairment in a large sample of patients with parkinson's disease. *Journal of behavioural neurology 11*, 131–137.

(Hustad, 2008) K. C. Hustad, 2008. The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research 51*(3), 562–573.

(Joachims, 1999) T. Joachims, 1999. Making large-scale SVM learning practical. B. Schölkopf, C. Burges, & A. Smola (Eds.), *Advances in Kernel Methods - Support Vector Learning*, Chapter 11, 169–184. Cambridge, MA : MIT Press.

(Kempler et Van Lancker, 2002) D. Kempler & D. Van Lancker, 2002. Effect of speech task on intelligibility in dysarthria : a case study of parkinson's disease. *Brain and language 80*(3), 449–464.

(Kent, 1996) R. D. Kent, 1996. Hearing and believing : some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology 5*(3), 7–23.

(Khan et al., 2014) T. Khan, J. Westin, & M. Dougherty, 2014. Classification of speech intelligibility in parkinson's disease. *Biocybernetics and Biomedical Engineering 34(1)*, 35–45.

(Laaridh et al., 2015) I. Laaridh, C. Fredouille, & C. Meunier, 2015. Automatic detection of phone-based anomalies in dysarthric speech. *ACM Transactions on accessible computing 6*(3), 9 :1–9 :24.

(Laaridh et al., 2016a) I. Laaridh, C. Fredouille, & C. Meunier, 2016a. Automatic anomaly detection for dysarthria across two speech styles : Read vs spontaneous speech. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia.

(Laaridh et al., 2016b) I. Laaridh, C. Fredouille, & C. Meunier, 2016b. Evaluation of a phone-based anomaly detection approach for dysarthric speech. *Proceedings of Interspeech'16*, San Francisco, USA.

(Ladefoged, 1975) P. Ladefoged, 1975. A course in phonetics. *Orlando : Harcourt Brace. 2nd ed 1982, 3rd ed. 1993, 4th ed. 2001, 5th ed. Boston : Thomson/Wadsworth 2006, 6th ed. 2011*.

(Lamel et al., 1991) L. F. Lamel, J. L. Gauvain, & M. Eskénazi, 1991. BREF, a large vocabulary spoken corpus for french. *Proceedings of European Conference on Speech Communication and Technology (Eurospeech'91)*, Genoa, Italy, 505–508.

(Leggetter et Woodland, 1995) C. J. Leggetter & P. C. Woodland, 1995. Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models. *Computer Speech & Language 9(2)*, 171–185.

(Lhoussaine, 2012) L. Lhoussaine, 2012. *Première validation de la Grille d'Évaluation Perceptive de la Dysarthrie (G.E.P.D.) : effet du niveau d'expertise du jury et différenciation entre types de dysarthrie*. Mémoire d'orthophonie, Speech therapist thesis, University of Paris VI, Pierre et Marie Curie (in French).

(Logeman et al., 1978) J. A. Logeman, H. B. Fisher, B. Boshes, & E. R. Blonsky, 1978. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients. *Journal of Speech and Hearing Disorders 43*, 47–57.

(Lowit et Kent, 2010) A. Lowit & R. D. Kent, 2010. *Assessment of motor speech disorders*, Volume 1. Plural publishing.

(Makhoul et al., 1999) J. Makhoul, F. Kubala, R. Schwartz, & R. Weischedel, 1999. Performance measures for information extraction. *Proceedings of DARPA Broadcast News Workshop*.

(Markel et Gray, 1976) J. D. Markel & A. H. Gray, 1976. Linear prediction of speech. *Springer Verlag*.

(Martinez et al., 2013) D. Martinez, P. Green, & H. C. and, 2013. Dysarthria intelligibility assessment in a factor analysis total variability space. *Proceedings of Interspeech'13*, Lyon, France.

(Menendez-Pidal et al., 1996) X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, & H. T. Bunnell, 1996. The nemours database of dysarthric speech. *Proceedings of the Fourth International Conference on Spoken Language, ICSLP*, Volume 3, 1962–1965.

(Meunier, 2007) C. Meunier, 2007. Phonétique acoustique. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(13)*, 164–173.

(Meunier et al., 2016) C. Meunier, C. Fougeron, C. Fredouille, B. Bigi, L. Crevier-Buchman, E. Delais-Roussarie, L. Georgeton, A. Ghio, I. Laaridh, T. Legou, C. Pillot-Loiseau, & G. Pouchoulin, 2016. The TYPALOC corpus : A collection of various dysarthric speech recordings in read and spontaneous styles. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia.

(Middag et al., 2009) C. Middag, J.-P. Martens, G. Van Nuffelen, & M. De Bodt, 2009. Automated intelligibility assessment of pathological speech using phonological features. *EURASIP Journal on Advances in Signal Processing 2009*(1), 1–9.

(Morales et Cox, 2009) S. O. C. Morales & S. J. Cox, 2009. Modelling errors in automatic speech recognition for dysarthric speakers. *EURASIP Journal on Advances in Signal Processing 2009*(1), 1–14.

(Müller et al., 2001) J. Müller, G. K. Wenning, M. Verny, A. McKee, K. R. Chaudhuri, K. Jellinger, W. Poewe, & I. Litvan, 2001. Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders. *Archives of neurology 58*(2), 259–264.

(Neri et al., 2006) A. Neri, C. Cucchiarini, & H. Strik, 2006. ASR-based corrective feedback on pronunciation : does it really work ? *INTERSPEECH-06*, Pennsylvanie, USA.

(Neumeyer et al., 1996) L. Neumeyer, H. Franco, M. Weintraub, & P. Price, 1996. Automatic text-independent pronunciation scoring of foreign language student speech. *Proceedings of the Fourth International Conference on Spoken Language, ICSLP 96*, Volume 3, 1457–1460.

(Nuffelen et al., 2009) G. V. Nuffelen, C. Middag, M. D. Bodt, & J.-P. Martens, 2009. Speech technology-based assessment of phoneme intelligibility in dysarthria. *International journal of language and communication disorders 44(5)*, 716–730.

(Orozco-Arroyave et al., 2016) J. Orozco-Arroyave, J. Vdsquez-Correa, J. Arias-Londo, J. Vargas-Bonilla, S. Skodda, J. Rusz, E. Noth, et al., 2016. Towards an automatic monitoring of the neurological state of Parkinson's patients from speech. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6490–6494.

(Parsons, 1997) C. L. Parsons, 1997. Communication with computers : the use of communication technology in speech-language pathology. *Australian communication quarterly, Spring*, 9–15.

(Peacher, 1950) W. G. Peacher, 1950. The etiology and differential diagnosis of dysarthria. *The Journal of speech disorders 15*(3), 252.

(Pianelli et Restivo, 2016) L. Pianelli & L. Restivo, 2016. *Evaluation d'un système de détection de déviances dans la réalisation articulatoire dans la dysarthrie*. Mémoire d'orthophonie, Speech therapist thesis, Aix-Marseille University (AMU), Marseille (in French).

(Pinto, 2007) S. Pinto, 2007. De l'élaboration à la production de la parole. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(1)*, 1–12.

(Piro et Ziamni, 2014) L. Piro & L. Ziamni, 2014. *Elaboration d'un protocole d'analyse perceptive des zones déviantes de la parole dysarthrique*. Mémoire d'orthophonie, Speech therapist thesis, Aix-Marseille University (AMU), Marseille (in French).

(Precoda et al., 2000) K. Precoda, C. A. Halverson, & H. Franco, 2000. Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability. *Proceedings of STILL 2000*, 102–105.

(Revis, 2004) J. Revis, 2004. *L'analyse perceptive des dysphonies : approche phonétique de l'évaluation vocale*. Thèse de Doctorat, Université de la Méditerranée.

(Robert, 2007) D. Robert, 2007. La dysarthrie dans la sclérose latérale amyotrophique. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(45)*, 448–455.

(Rudzicz, 2007) F. Rudzicz, 2007. Comparing speaker-dependent and speaker adaptive acoustic models for recognizing dysarthric speech. *Proceedings of the Ninth International ACM SIGACCESS Conference on Computers and Accessibility*, Tempe, USA.

(Rudzicz, 2010) F. Rudzicz, 2010. Towards a noisy-channel model of dysarthria in speech recognition. *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, Los Angeles, USA, 80–88.

(Rudzicz et al., 2012) F. Rudzicz, A. K. Namasivayam, & T. Wolff, 2012. The torgo database of acoustic and articulatory speech from speakers with dysarthria. *Proceedings of the International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turquie, 523–541.

(Schalling, 2007) E. Schalling, 2007. La dysarthrie dans les pathologies cérébelleuses. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(35)*, 349–356.

(Scholkopf et Smola, 2001) B. Scholkopf & A. J. Smola, 2001. *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA : MIT Press.

(Sevenster et al., 1998) B. Sevenster, G. de Krom, & G. Bloothooft, 1998. Evaluation and training of second-language learners' pronunciation using phoneme-based HMMs. *Proc. STiLL*, 91–94.

(Sharma et Hasegawa-Johnson, 2010) H. V. Sharma & M. Hasegawa-Johnson, 2010. State-transition interpolation and map adaptation for hmm-based dysarthric speech recognition. *HLT/NAALC Workshop on speech and language processing for assistive technology (SPLAT)*.

(Sharma et al., 2009) H. V. Sharma, M. Hasegawa-Johnson, J. Gunderson, & A. Perlman, 2009. Universal access : preliminary experiments in dysarthric speech recognition. *Proceedings of Interspeech'09*, Brighton, United Kingdom.

(Shriberg et al., 1990) L. D. Shriberg, J. Kwiatowski, & T. Synder, 1990. Tabletop versus microcomputer-assister speech management : response evocation phase. *Journal of speech and hearing disorders 55*, 635–655.

(Teston, 2007) B. Teston, 2007. L'étude instrumentale des gestes dans la production de la parole : Importance de l'aérophonométrie. *Les Dysarthries*, 115–117.

(Teston et al., 1999) B. Teston, A. Ghio, & B. Galindo, 1999. A multisensor data acquisition and processing system for speech production investigation. *International Congress of Phonetic Sciences (ICPhS)*, 2251–2254.

(Van Lancker Sidtis et al., 2012) D. Van Lancker Sidtis, K. Cameron, & J. J. Sidtis, 2012. Dramatic effects of speech task on motor and linguistic planning in severely dysfluent parkinsonian speech. *Clinical linguistics & phonetics 26*(8), 695–711.

(Vapnik, 1995) V. Vapnik, 1995. *The Nature of Statistical Learning Theory*. New York, NY, USA : Springer-Verlag New York, Inc.

(Viallet et Teston, 2007) F. Viallet & B. Teston, 2007. La dysarthrie dans la maladie de Parkinson. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(37)*, 375–382.

(Viterbi, 1967) A. J. Viterbi, 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory 13(2)*, 260–269.

(Woehrling et al., 2009) C. Woehrling, P. B. de Mareüil, & M. Adda-Decker, 2009. Linguistically-motivated automatic classification of regional french varieties. *Proceedings of Interspeech'09*, Brighton, Royaume-Uni, 2183–2186.

(Young et Mihailidis, 2010) V. Young & A. Mihailidis, 2010. Difficulties in automatic speech recognition of dysarthric speakers and the implications for speech-based applications used by the elderly : a literature review. *Assistive technology, RESNA Journal 22*, 99–112.

(Zyski et Weisiger, 1987) B. J. Zyski & B. E. Weisiger, 1987. Identification of dysarthria types based on perceptual analysis. *Journal of Communication Disorders 20*(5), 367–378.

(Özsancak et Devos, 2007) C. Özsancak & D. Devos, 2007. Les ataxies cérébelleuses. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part III(35)*, 337–348.

Bibliographie personnelle

Revues internationales avec comité de sélection

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Automatic Detection of Phone-Based Anomalies in Dysarthric Speech »*dans Transactions on Accessible Computing TACCESS* , 2015

Conférences d'audience internationale avec comité de sélection

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Evaluation of a Phone-Based Anomaly Detection Approach for Dysarthric Speech » *dans INTERSPEECH*, 2016

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Automatic Anomaly Detection for Dysarthria across Two Speech Styles: Read vs Spontaneous Speech » *dans LREC*, 2016

MEUNIER C., FOUGERON C., FREDOUILLE C., BIGI B., CREVIER-BUCHMAN L., DELAIS-ROUSSARIE E., GEORGETON L., GHIO A., LAARIDH I., LEGOU T., PILLOT-LOISEAU C. ET POUCHOULIN G. « The TYPALOC Corpus: A Collection of Various Dysarthric Speech Recordings in Read and Spontaneous Styles » *dans LREC*, 2016

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Automatic speech processing for dysarthria: A study of Inter-pathology variability » *dans ICPHS*, 2015

Conférences d'audience nationale avec comité de sélection

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Détection automatique d'anomalies sur deux styles de parole dysarthrique: parole lue vs spontanée » *dans JEP*, 2016

Workshop d'audience nationale avec comité de sélection

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Traitement automatique de la parole dysarthrique: Étude de la variabilité inter-pathologique » *dans JPC*, 2015

LAARIDH I., FREDOUILLE C. ET MEUNIER C. « Détection automatique d'anomalies dans la parole dysarthrique » *dans JPC*, 2015

Annexes

Annexes A

Corpus *DesPhoAPady*

TABLE *A.1* – *Patients atteints d'ataxie cérébelleuse du corpus DesPhoAPady.*

Genre	Locuteur	Age	Grade	Irrégularité globale	Mélodie	Débit	Nasonnement	Palilalies	Articulation	Irrégularité du débit	Intelligibilité
F	CCM-002595-01_L01	34	0.9	0.5	-1.2	-1	0.6	0.1	0.7	0.5	0.5
F	CCM-002710-01_L01	59	2.1	1.3	-0.4	-1.3	1.5	0	1.6	1.4	1.2
F	CCM-003110-01_L01	69	1.3	1.2	-0.6	-0.8	0.4	0.1	0.5	1	0.6
F	CCM-003493-01_L01	35	0.9	0.8	-1.5	-1.5	0.2	0.5	0.7	0.6	0.5
F	CCM-004538-01_L01	68	1	0.9	-0.5	0.8	0.7	0.6	1.1	1.2	0.6
F	MTO-00KNSM-01_L01	79	1.8	1.4	-1.5	0.3	0.8	0.5	1.9	2	1.8
F	MTO-00LIAF-01_L01	33	2.2	1.5	-1.2	1.4	1.3	1	2.2	2.2	1.8
F	MTO-00SNJF-01_L01	56	2.3	1.7	0.8	-1.7	1.5	0.6	2.4	1.6	1.6
F	MTO-00WUCF-01_L01	49	1.5	1.2	0.6	0.1	0.6	0.5	1.5	1.5	1.1
H	CCM-002616-01_L01	49	1.1	0.5	-0.6	0	0.8	0.2	1.2	0.7	0.5
H	CCM-003094-01_L01	32	1.5	1.7	0.3	-2.2	0.7	0.4	1.4	1.7	0.8
H	CCM-003998-01_L01	77	1.5	1.2	-1.5	-1.9	0.5	0.8	0.7	1.4	1.1
H	CCM-004523-01_L01	45	0.8	0.3	-0.5	-0.9	0.3	0.1	0.7	1	0.3
H	CCM-004529-01_L01	70	0.6	0.7	-1.1	0.4	0.4	0.5	0.5	0.7	0.4
H	CCM-004773-01_L01	55	1.2	0.9	-1.5	-1.3	0.5	0.5	1.1	1.1	0.7
H	MTO-00ADRM-01_L01	86	2	1.5	0.8	-1.2	0.7	0.2	1.8	1.5	1.6
H	MTO-00CAGM-01_L01	51	1.4	0.6	-0.3	-1	1.1	0.7	1.4	0.6	1
H	MTO-00COSM-01_L01	33	2.5	1.6	-1.3	-2	1.5	0.7	2.2	2.4	1.5
H	MTO-00KRLM-01_L01	52	1.9	1.5	0.6	-1	0.7	0.5	2.1	1.5	1.4
H	MTO-00LAMM-01-L02	54	2.3	1.5	-1	-1.4	1.5	0.5	1.9	1.2	1.3
H	MTO-00MYMM-01_L01	68	1.4	0.8	0.2	-1	0.9	0.5	1.4	1.1	0.8
H	MTO-00PEPM-01_L01	72	0.9	0.5	0.4	-0.6	0.5	0	0.9	0.8	0.6

TABLE *A.2* – *Patients atteints de la maladie de Parkinson du corpus DesPhoAPady.*

Genre	Locuteur	Age	Grade	Irrégularité globale	Mélodie	Débit	Nasonnement	Palilalies	Articulation	Irrégularité du débit	Intelligibilité
F	AHN-001209-01_L04	61	0	0.1	-0.8	0.6	0.1	0	0.2	0.1	0
F	AHN-001233-01_L04	73	0.6	0.3	0.3	0.8	0.5	0.1	0.3	0.8	0.1
F	AHN-001273-01_L04	62	0.2	0.5	0.3	0	0.1	0	0.2	0.1	0
F	AHN-001283-01_L04	62	0.2	0.1	0	0	0.2	0.3	0.4	0.2	0.1
F	AHN-001286-01_L04	66	0.2	0.3	-0.6	0.3	0	0.3	0	0.4	0.2
F	CCM-003148-01_L01	60	1.3	1.4	-1.4	1.5	0.5	0.8	1.2	1.8	0.9
F	CCM-003346-01_L01	81	0.5	0.6	-0.7	0.6	0.1	0.5	0.2	0.7	0.4
F	CCM-003541-01_L01	72	2.2	1.7	-2.2	2.6	1	0.6	1.5	1.5	1.8
H	AHN-000938-05_L04	51	0.2	0.4	-0.7	0.7	0.2	0.3	0.3	0.3	0
H	AHN-000969-02_L04	70	0	0.2	-0.2	0.2	0	0.2	0.1	0.2	0
H	AHN-001211-01_L04	81	0.7	0.9	-0.5	0.5	0.5	0.5	0.7	0.9	0.4
H	AHN-001214-01_L04	85	0.5	0.3	-1	1.2	0	0.2	0.5	0.3	0.2
H	AHN-001220-01_L03	74	1.5	2.1	-1.9	-0.9	0.5	0.8	1.2	2.2	1.3
H	AHN-001228-01_L04	67	1.1	1.1	-0.5	0.1	0.2	0.1	0.7	0.8	0.4
H	AHN-001251-01_L04	72	0.8	1	-0.8	0.5	0.2	1.5	0.8	1.1	0.7
H	AHN-001260-01_L04	66	1.1	1	-1.3	1.5	0.2	1.2	1.2	1.7	0.9
H	AHN-001263-01_L04	71	0.5	0.5	-0.4	0.4	0.5	0.3	0.4	0.2	0.1
H	AHN-001276-01_L04	70	0.8	0.6	-0.9	0.8	0.6	0.2	1	0.7	0.4
H	CCM-001773-01_L01	64	0.4	0.3	-0.1	1.7	0.1	0	0.4	0.6	0.4
H	CCM-002631-01_L01	59	2.3	1.5	-2.2	2.2	0.4	1.4	2.1	1.7	2.2
H	CCM-003106-01_L01	63	1.5	1.3	-0.6	1.5	0.2	0	0.9	1.5	0.8
H	CCM-003130-01_L01	62	0.8	0.9	-1.5	0	0	0.5	0.8	0.6	0.6
H	CCM-003557-01_L01	48	0.6	0.3	-1.6	-0.5	0.2	0.6	0.1	1.3	0.4
H	CCM-003558-01_L01	63	2.7	1.7	-2.4	2.8	1.5	1.5	2.5	1.9	2.6
H	CCM-003722-01_L01	56	0.5	0.5	-1.1	0.4	0.2	0.3	0.2	1	0.3
H	CCM-003733-01_L01	61	1.4	1.7	-1.6	-0.5	0.4	0.5	1.1	2.2	1.2
H	CCM-003733-02_L01		1.6	1.3	-2.3	-1.8	0.5	0.4	1.4	1.7	1
H	CCM-003734-01_L01	58	0.5	0.5	-0.7	-0.1	0.2	0.2	0.4	0.7	0.1
H	CCM-003810-01_L01	60	1.6	1.4	-1.9	1.7	0.5	0.5	1	1.1	1.3
H	CCM-003848-01_L01	77	1.3	1.1	-1.2	1.5	0.5	0.3	0.6	1.2	0.6
H	CCM-004760-01_L01	57	1.5	1.5	-1.4	2.5	0.3	0.7	1.3	1.5	1.5

TABLE *A.3* – *Patients atteints de SLA du corpus DesPhoAPady.*

Genre	Locuteur	Age	Grade	Irrégularité globale	Mélodie	Débit	Nasonnement	Palilalies	Articulation	Irrégularité du débit	Intelligibilité
F	PHO-000005-01_L01	66	1.5	0.8	-0.6	-1.3	1.9	0	1.2	0.6	0.6
F	PHO-000006-01_L01	68	2.1	1.1	-0.9	-1.5	1.9	0	2.2	0.6	1.8
F	PHO-000008-01_L01	89	2.7	1.3	-0.4	-2.2	1.9	0	2.2	0.8	1.8
F	PHO-000011-01_L01	71	0.6	0.5	0	-0.1	0.6	0	0.6	0.4	0.1
F	PHO-000024-01_L01	70	1.5	0.1	1.2	-0.2	2.6	0	1	0.5	1
F	PHO-000566-01_L01	81	2.9	1.3	-2	-2.7	2.3	0	2.3	0.8	1.7
F	PHO-000814-01_L01	68	2.3	1	-1	-2	2.2	0	1.8	0.8	1.3
F	PHO-001003-01_L01	70	2.5	1.5	-1.5	-1.2	1.3	0.1	2.2	0.8	2
F	PHO-001038-02_L01	70	1.4	0.7	-1.1	-1.5	0.7	0	1.1	0.5	0.6
F	PHO-001055-01_L01	74	2.4	0.8	-1.7	-2.2	2.5	0	2	0.7	1.6
F	PHO-001058-01_L01	59	2.3	0.9	-1.1	-2.2	2.1	0	2	0.7	1.4
F	PHO-001066-01_L01	69	2.9	1.7	-1.9	-2.5	2.9	0.1	2.7	1.3	2.6
F	PHO-001070-01_L01	63	2.4	1.1	0.1	-1.1	2.1	1	2.2	1	1.6
F	PHO-001277-01_L01	55	2.4	1.3	-1.6	-2.5	1.7	0.1	2.3	0.6	1.7
F	PHO-001329-01_L01	50	0.9	0.7	0	-0.5	0.5	0	1.1	0.5	0.5
F	PHO-001484-01_L01	59	0.7	0.6	-0.4	-0.5	0.1	0.6	0.5	0.3	0.3
F	PHO-001486-01_L01	70	1.3	0.8	-0.5	-1.2	1.3	0.3	0.8	0.4	0.6
F	PHO-001499-01_L01	62	1.8	1	0.9	-1.4	1.6	0.1	1.5	1.1	1
F	PHO-001522-01_L01		2.6	0.8	-2.3	-2.5	1.2	0	1.8	0.6	1.4
F	PHO-001568-01_L01		1.5	1	-0.4	-1.4	1.5	0	1.5	1.1	0.6
F	PHO-001940-01_L01	73	2.8	1.5	-1.7	-2	2.5	0.3	2.6	1.1	2.5
F	PHO-115371-01_L01	74	1.9	0.8	-1	-1	1.7	0	2	0.8	1.5
F	PHO-123576-01_L01	69	0.5	0.7	-0.5	0	0.5	0	0.5	0.4	0.3
H	PHO-000007-01_L01	70	2.8	2	-1.2	-1.6	2.1	0.1	2.9	1.2	2.5
H	PHO-001133-01_L01	54	2.2	1.9	-1.8	-0.3	1.4	0.6	1.6	1.5	1.7
H	PHO-001269-01_L01	52	2.5	0.9	-1.5	-2.3	2.5	0.3	2.1	0.6	1.6
H	PHO-001348-01_L01	44	1.1	0.7	-1	0.2	0.6	0.3	1.5	1	0.8
H	PHO-001436-01_L01	61	0.7	0.3	-0.9	-0.8	0.8	0.1	0.1	0.2	0.1
H	PHO-001442-01_L01	64	0.9	0.5	-0.5	-0.4	1.4	0.2	0.5	0.4	0.1
H	PHO-001473-01_L01	67	2.1	0.9	0.1	-2.2	1.1	0	1.5	0.8	1
H	PHO-001581-01_L01	49	3	1.5	-1.6	-2.5	2.8	0	2.7	1.3	2.5
H	PHO-001594-01_L01		2.2	1.5	-1.5	-1.6	1.4	0.4	2.3	1	1.6
H	PHO-001670-01_L01	71	1.9	1.1	-1.1	-1.2	1.2	0	1.9	0.7	1.5
H	PHO-001750-01_L01	59	1.2	0.9	-0.5	-0.5	0.4	0	1.5	0.6	0.7
H	PHO-001804-01_L01	61	1.6	0.8	-1.5	-1.8	1.5	0.1	1.4	1	1
H	PHO-001836-01_L01	74	2.5	1.1	-1	-1.5	2.7	0.4	2.5	0.9	2.1
H	PHO-307175-01_L01	56	1.4	1.2	-0.6	1.5	2	0.5	1.2	1.7	1

TABLE *A.4* – *Locuteurs contrôles du corpus DesPhoAPady.*

Genre	Locuteur	Age	Grade	Irrégularité globale	Mélodie	Débit	Nasonnement	Palilalies	Articulation	Irrégularité du débit	Intelligibilité
F	CCM-001762-01_L01	45	0.1	0.1	-0.1	-0.1	0.1	0	0	0	0
F	CCM-001764-01_L01	39	0.1	0.2	-0.7	-0.2	0	0.4	0.1	0.4	0.1
F	CCM-002715-01_L01	43	0	0.1	0.1	0	0	0	0	0.1	0
F	CCM-002717-01_L01	35	0.1	0.1	-0.4	0.1	0.1	0.2	0	0	0
F	CCM-002757-01_L01	48	0.2	0.3	0.1	1.9	0	0.3	0.3	0.4	0.2
F	CCM-002758-01_L01	43	0	0.1	0.1	0.5	0	0	0	0.2	0
F	CCM-002759-01_L01	35	0	0.1	-0.2	0.5	0	0.1	0	0	0.1
F	CCM-002763-01_L01	37	0	0	0.2	-0.2	0	0	0	0	0
F	CCM-002902-01_L01	33	0.1	0.1	0.2	0.3	0	0	0	0	0.1
F	CCM-003231-02_L01	37	0	0	0.4	0.1	0	0	0	0	0
F	CCM-004426-01_L01	61	0	0	0	0.4	0.1	0.2	0	0	0
F	CCM-004427-01_L01	58	0.1	0.2	-0.5	0.7	0.1	0	0	0.1	0.1
F	CCM-004429-01_L01	76	0.6	0.5	-0.7	-0.7	0.3	0	0.1	0.3	0.1
F	CCM-004497-01_L01	59	0.1	0.1	-0.3	0.3	0	0	0	0.4	0
H	CCM-000632-02_L01	54	0.2	0.2	-0.4	0.4	0	0	0.3	0.2	0
H	CCM-001765-01_L01	32	0.3	0.1	-0.1	0.2	0.5	0	0.2	0.2	0.1
H	CCM-002718-01_L01	37	0.5	0.2	-0.3	0.1	0.8	0.1	0.2	0.2	0.1
H	CCM-002720-01_L01	35	0	0	0.2	0	0.1	0	0	0	0
H	CCM-002723-01_L01	32	0.1	0	-0.5	0.5	0.2	0.1	0	0	0
H	CCM-002729-01_L01	33	0	0	-0.8	0.7	0.2	0	0.1	0	0.1
H	CCM-002901-01_L01	34	0	0.1	-0.1	0.2	0	0	0	0.1	0
H	CCM-003214-01_L01	47	0	0.1	-0.1	-0.1	0	0	0	0.1	0
H	CCM-004371-03_L01	62	0	0	0.5	0.5	0	0	0	0.1	0
H	CCM-004413-01_L01	59	0.4	0.2	-0.5	0.7	0.4	0.2	0.6	0.3	0
H	CCM-004414-01_L01	54	0	0.1	-0.2	0	0.1	0	0	0.3	0
H	CCM-004415-01_L01	60	0.5	0.4	0.1	-0.5	0.5	0	0.1	0.5	0.1
H	CCM-004420-01_L01	57	0.6	0.5	-1.5	0.1	0.4	0.1	0.3	0.5	0.1
H	CCM-004425-01_L01	60	0	0	-0.5	-0.2	0.2	0	0	0	0
H	CCM-004428-01_L01	61	0	0.2	0.4	-0.2	0	0	0	0.3	0

Annexes B

Consignes de l'évaluation perceptive

Vous allez entendre des enregistrements de textes lus dans lesquels des séquences de un ou plusieurs mots ont été extraites. La parole produite au cours de cette lecture peut éventuellement présenter des déviances pathologiques.

Nous vous demandons de juger si les mots de cette séquence sont déviants ou non, en sachant que chacune des séquences peut être altérée en partie, totalement, ou pas du tout.

Vous avez la possibilité d'écouter au maximum trois fois la même séquence en cliquant sur le petit haut-parleur. Toutefois, si une seule écoute est suffisante pour donner votre réponse vous pouvez passer directement à la séquence suivante toujours en cliquant sur la case "suivant". Par défaut, chaque mot apparaît comme "normal", si l'un (ou plusieurs) d'entre eux vous semblent déviants, cocher dans la ligne "déviant", la (les) case(s) située(s) sous ce(s) mot(s).

Attention, les séquences ont été découpées dans un flux de parole continue :
Les débuts et/ou fins peuvent parfois être abruptes, il ne faut pas en tenir compte.

L'expérience devrait durer entre 30 et 40min.

Pour vous familiariser avec la tâche, un entraînement va vous être présenté. Il n'y a pas de bonnes ou de mauvaises réponses ce qui nous intéresse c'est votre jugement.