Traitement Automatique et Troubles de la Voix et de la Parole : champs d'application, contraintes et limites

Corinne Fredouille

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

HABILITATION À DIRIGER DES RECHERCHES

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

SPÉCIALITÉ : Informatique

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

TRAITEMENT AUTOMATIQUE ET TROUBLES DE LA VOIX ET DE LA PAROLE : CHAMPS D'APPLICATION, CONTRAINTES ET LIMITES

par

Corinne FREDOUILLE

Soutenue publiquement le 29 novembre 2016 devant un jury composé de :

M^{me}	Martine ADDA-DECKER	Directeur de Recherche, CNRS, Laboratoire de Phonétique et Phonologie, Paris	Rapporteur
M.	Laurent BESACIER	Professeur, Laboratoire Informatique de Grenoble	Rapporteur
M.	Philippe BLACHE	Directeur de Recherche, CNRS, Laboratoire Parole et Langage, Aix-en-Provence	Rapporteur
M^{me}	Lise CREVIER-BUCHMAN	Chargé de Recherche, HDR, CNRS, Docteur en médecine, Laboratoire de Phonétique et Phonologie, Paris	Examinateur
M.	Jean-François BONASTRE	Professeur, Laboratoire Informatique d'Avignon	Examinateur

---

Laboratoire Informatique d'Avignon

A mon père.

Table des matières

Introduction

"Quand des paroles sortent, s'envolent en l'air,
vivent un instant et meurent,
c'est ce qui s'appelle parler." (Paul Auster)

Bien que les modes de communication aient largement évolué depuis une vingtaine d'années avec l'ère du numérique, pouvant se substituer dans de nombreuses situations à la parole - messagerie électronique, réseaux sociaux, blogs, messageries instantanées... celle-ci reste indispensable pour une bonne intégration dans notre société. La communication orale permet de partager avec les autres, que ce soit dans le monde professionnel ou la vie de tous les jours : partager pour exprimer une demande, pour informer, pour conseiller, pour débattre, pour convaincre, pour enseigner, pour divertir, pour pervertir, pour exprimer un sentiment, ... pour vivre en société. La démocratisation de la téléphonie mobile, des réseaux de communication, des logiciels de visioconférences et de l'accès à Internet fait que nous pouvons communiquer oralement partout avec tous nos contacts professionnels ou personnels, quelle que soit leur localisation géographique. Aussi, au vu de l'importance de la communication orale, perdre la parole ou le langage peut être ressenti comme une perte d'humanité.

Les troubles de la communication sont définis par The American Speech and Hearing Association 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 la communication englobent toute altération au niveau de la voix, de la parole, du langage et de l'audition nuisant à la communication. Une enquête réalisée en 2012 aux Etats-Unis estime à 10% le nombre de personnes disant souffrir d'un trouble de la communication (seulement 2% se déclarant sur la base d'un diagnostic clinique) (Morris et al., 2016). En Australie, une étude menée auprès d'une association professionnelle regroupant 70% des orthophonistes du pays évalue à 5% ce même chiffre, jugé comme sous-estimé.

Nous nous intéresserons, dans ce document, aux troubles de la parole et de la voix. De manière plus précise, nous nous focaliserons, d'une part, sur la dysphonie, définie

comme une altération de la voix, support de la parole. Cette altération pourra affecter les différents éléments acoustiques qui caractérisent la voix comme le timbre, la hauteur et l'intensité ; la voix sera alors perçue par le locuteur lui-même, son entourage et le praticien lors de la consultation comme altérée (voix éraillée, rauque, soufflée, monocorde, etc.). Dans certains cas, la dysphonie se manifestera, plutôt, comme une gêne ressentie par le locuteur (fatigue vocale, manque d'efficacité dans le geste vocal, etc.) sans pour autant présenter de manifestation sur le plan acoustique. D'autre part, nous aborderons une forme particulière des troubles de la parole, dénommée dysarthrie, définie comme un trouble de la réalisation motrice de la parole dû à une atteinte du système nerveux central ou périphérique. La dysarthrie est l'un des symptômes de différentes maladies neurodégénératives comme par exemple la maladie de Parkinson, la sclérose en plaque, la Sclérose Latérale Amyotrophique (SLA), les ataxies cérébelleuses, ou encore les Accidents Vasculaires Cérébraux (AVC). Elle peut avoir une incidence sur le fonctionnement d'un ou plusieurs composants impliqués dans la production de la parole comme la respiration, la phonation (induisant une dysphonie), la résonance, l'articulation ou encore la prosodie. En lien direct avec la localisation de la ou des lésions neurologiques, la dysarthrie pourra conduire à une faiblesse musculaire ou, au contraire, à une augmentation exagérée du tonus des muscles au repos, à des mouvements involontaires ou encore à une mauvaise coordination des mouvements volontaires.

La dysarthrie (incluant les troubles dysphoniques) a été étudiée suivant différents axes disciplinaires dans la littérature. Comme le rapporte (Grewel, 1957), les premières études, au début et milieu du siècle dernier, avaient pour objectif de définir la dysarthrie et ses différentes formes en fonction des troubles neurologiques associés (exemple de classes : *"Dysarthria with severe epilepsy"*). Bien que précieuse, (Grewel, 1957) souligne que cette classification, purement "neurologique", n'est pas suffisante du point de vue de la prise en charge médicale des troubles de la parole chez les patients dysarthriques et de leur diagnostic par un praticien Otho-Rhino-Laryngologiste (ORL) ou un orthophoniste. Les auteurs proposent, par conséquent, une approche multi-disciplinaire visant à étudier la dysarthrie conjointement dans les domaines de la neurologie, de la phoniatrie et de l'orthophonie et de définir une classification des formes de dysarthrie selon ces 3 domaines (exemple de classes : *Dysarthria with flaccid paralysis : depending on the aspect of paralysis disorders affecting articulation of the lips, cheeks, tongue and/or velum muscles are noticed, sometimes in combinations with laryngeal paralysis and therefore dysphonia*). Une dizaine d'années plus tard, cette nécessité de mieux comprendre la dysarthrie et ses conséquences sur la production de la parole pour une meilleure prise en charge clinique du patient conduit Darley et ses collègues de la Mayo Clinic de Cleveland à proposer une nouvelle classification des formes de dysarthrie (Darley et al., 1969b,a, 1975). Cette classification, obtenue à partir d'un regroupement des anomalies perçues dans la parole, repose sur une association explicite entre troubles moteurs et atteinte anatomique. Elle a été complétée plus récemment par les travaux de (Duffy, 2005). Toujours considérée comme une référence pour l'évaluation de la dysarthrie dans la pratique clinique, elle est également jugée toujours insuffisante dans la littérature d'un point de vue étiologique et pour une prise en charge optimale du patient (Pinto et Ghio, 2009).

Pour répondre à ce besoin, une partie des travaux sur la dysarthrie (et la dysphonie) ont alors été orientés vers une meilleure compréhension des dysfonctionnements physiologiques intervenant sur chaque composant impliqué dans la production de la parole. De ce fait, les études ont porté sur l'observation, chez des populations de patients dysarthriques, des mouvements des différents organes de l'appareil de production de la parole par l'analyse de mesures aérodynamiques, cinématiques, cinéradiographiques, acoustiques, articulatoires, etc. Dans le même objectif, un grand nombre de chercheurs, notamment issus du domaine de la phonétique, se sont concentrés sur l'analyse acoustique et/ou articulatoire du signal de parole afin d'acquérir une connaissance plus approfondie des altérations observées dans la parole dysarthrique sur la base de mesures instrumentales ou perceptives.

En parallèle, le besoin d'outils d'évaluation fiable, objective et reproductible de la dysarthrie (et de la dysphonie) exprimé par le corps médical a suscité de la part de la communauté scientifique de nombreuses investigations. En effet, le seul outil d'évaluation utilisé en pratique clinique pour détecter la présence d'une dysarthrie ou pour mesurer son degré de sévérité global reste l'"oreille" du praticien - évaluation perceptive - malgré le caractère subjectif avéré de ce type d'évaluation (Revis, 2004). Pourtant, ces évaluations sont cruciales, toujours pour la prise en charge thérapeutique du patient, dans le suivi de sa dysarthrie et de son évolution dans le cadre d'un traitement médicamenteux ou d'une rééducation.

La difficulté majeure dès lors qu'on étudie la parole dysarthrique est de faire face à la très large variabilité qui peut être observée dans des populations présentant des pathologies différentes (cf. classification des dysarthries), mais également chez des patients d'une même population (Tomik et Guiloff, 2010). En effet, outre la maladie et la localisation de la lésion neurologique, le niveau de dégradation de la parole sera conditionné par la nature et le temps d'apparition des symptômes au fur et à mesure que la maladie évolue. Par exemple, un patient atteint de la maladie de Parkinson pourra ne pas souffrir de dysarthrie alors que d'autres symptômes de la maladie sont apparus. De même, une étude sur une population de patients atteints de SLA montre que l'apparition de la dysarthrie peut intervenir de 33 mois avant le diagnostic officiel (la dysarthrie seule n'étant pas suffisante pour établir le diagnostic) jusqu'à 66 mois après (Yorkston et al., 1993). Ces différences peuvent être attribuées : (1) au patient lui-même en raison de son âge, de sa résistance physique, de ses antécédents de santé, etc. ; (2) à la progression de la maladie et/ou au degré d'altération des organes de l'appareil de production de la parole qui peuvent différer d'un patient à l'autre ; (3) ou encore aux stratégies d'adaptation et de compensation que les patients peuvent développer pour surmonter leurs troubles de la parole.

Pour tenir compte de cette variabilité, les travaux de recherche devraient s'appuyer sur de larges populations de patients, présentant une variété suffisante de maladies et de types de dysarthrie dans le cas d'une étude comparative entre populations. Nous verrons dans le document qu'il existe un petit nombre de corpus de parole dysarthrique et dysphonique dédiés à la recherche. Néanmoins, ils sont loin de répondre à tous les critères que nous venons d'énoncer, le contexte particulier du domaine clinique ne fa-

vorisant pas la collecte de données.

Finalement, les avancées de ces vingt dernières années dans le domaine du traitement automatique de la parole ont conduit la communauté scientifique à s'intéresser de très près aux technologies disponibles pour l'étude de la dysarthrie. Dès 1992, les approches automatiques ont été considérées comme des solutions potentielles pour répondre, par exemple, aux besoins d'outils d'évaluation objective de la dysarthrie (Ferrier et al., 1992; Doyle et al., 1997) - *la machine se substituant à l'"oreille" de l'expert*. Depuis, de nombreuses études ont porté sur leur application comme outils de détection, outils d'évaluation, outils thérapeutiques, ou encore comme outils de communication alternative pour aider les patients atteints d'handicap lourd (en plus des symptômes de dysarthrie) à intéragir avec leur environnement. Néanmoins, comme nous le soulignerons dans le document, les approches automatiques se heurtent aux difficultés que nous avons mentionnées ci-dessus, à savoir la variabilité de la parole dysarthrique, son caractère dégradée comparée à la parole normale et la faible quantité de données disponibles pour la représenter.

Les travaux que j'ai menés au LIA sur l'étude de la dysphonie et de la dysarthrie s'inscrivent dans ce cadre et, plus particulièrement, sur les thèmes de la détection et de l'évaluation automatique de la dysphonie et de la dysarthrie. Au delà des aspects applicatifs, l'objectif à long terme de ces travaux sont d'étudier en quoi et comment les systèmes de traitement automatique de la parole peuvent apporter de la connaissance à des experts humains et faciliter leur analyse du signal de parole et des conséquences des troubles de la parole et de la voix sur ce dernier.

Le document a été organisé comme suit :

- la première partie du document est consacrée à la description détaillée des mécanismes impliqués dans la production de la parole (chapitre 1). En effet, il nous est apparu important de revenir sur ces mécanismes afin de montrer la complexité des processus mis en jeu et de mieux appréhender les conséquences que des perturbations peuvent avoir sur le signal de parole résultant. Dans cette même partie, nous abordons ensuite les troubles de la parole et de la voix, et plus précisément, la dysphonie et la dysarthrie, qui sont au coeur de ce document (chapitre 2). Nous introduisons les différentes approches d'évaluation de ces troubles dans le domaine clinique et nous en exposons les limites. Nous concluons sur la place que peut tenir le traitement automatique de la parole dans l'évaluation de ces troubles.
- La deuxième partie du document est dédiée au traitement automatique de la parole et à son application à la parole pathologique (incluant la voix). Dans un premier temps, nous proposons un inventaire des corpus de parole pathologique disponibles pour la recherche et montrons leurs limites actuelles. Un état de l'art des approches de traitement automatique de la parole appliquées à la parole pathologique est ensuite proposé (chapitre 1). Au travers de ce dernier, nous avons tenté de mettre en avant les difficultés rencontrées, et notamment celles que nous avons évoquées précédemment. Nous présentons dans un dernier chapitre les activités majeures du LIA autour de la dysphonie et de la dysarthrie (chapitre 2). Les différents systèmes de traitement de la parole que nous avons développés dans

ce contexte sont décrits, évalués et critiqués. Nous exposons également quelques unes des connaissances que ces systèmes nous ont permis de mettre en lumière et ce, dans notre objectif, d'apporter de la connaissance à nos collègues phonéticiens et cliniciens avec qui nous collaborons depuis de nombreux années maintenant.

– la dernière partie de ce document a pour vocation de faire un bilan de l'application du traitement automatique de la parole à la parole pathologique, que ce soit au travers de la revue de la littérature que nous avons réalisée ou de nos propres travaux et de proposer des perspectives dans ce domaine en lien avec les activités du LIA.

Première partie

Les troubles de la parole et de la voix : de la production à l'évaluation

Chapitre 1

Les mécanismes de production de la parole

Sommaire

[...] ***Speech is a complex behaviour*** *that requires the coordinated contraction of a large number of muscles for its production, including the muscles of the lips, jaw, tongue, soft palate, pharynx and larynx as well as the muscles of respiration. Contraction of the muscles of the speech mechanism is controlled by nerve impulses, which originate in the motor areas of the cerebral cortex and then pass to the muscles by way of the motor pathways.* [...] (Murdoch, 1998).
Les troubles de la parole et de la voix pouvant être la conséquence d'une lésion localisée sur l'un ou plusieurs des éléments intervenant dans la production de la parole, il nous a semblé intéressant de revenir sur tous les mécanismes de production de la parole : du neurone au son produit. Ce chapitre est, par conséquent, dédié à la description des mécanismes neurologiques, phonatoires et articulatoires impliqués dans la production de la parole.

1.1 Mécanismes neurologiques

La parole est considérée comme un mouvement volontaire en opposition à un mouvement automatique (assurer, par exemple, la tonicité des muscles dans l'exécution d'un mouvement volontaire), un mouvement réflexe (réponse musculaire involontaire à un stimulus) ou encore un mouvement involontaire (régulation des fonctions vitales du corps humain : respiration, circulation sanguine). Tout mouvement volontaire, qu'il s'agisse d'un mouvement simple comme lever un bras ou d'un mouvement plus complexe comme produire de la parole va engendrer une activité cérébrale complexe mais hiérarchisée, articulée autour de trois étapes : (1) la planification du mouvement (définition, sélection et élaboration du mouvement en fonction du contexte, de l'environnement, mais également de notre propre expérience/apprentissage), (2) la préparation du mouvement (sélection de l'ensemble des muscles à contracter pour la réalisation du mouvement), (3) l'exécution du mouvement (envoi et transfert de l'information vers les muscles pour la réalisation proprement dite du mouvement). Les sous-sections suivantes vont être dédiées à la description des différents éléments du système nerveux intervenant dans ces différentes phases et leur rôle en vue de produire *in fine* un mouvement de parole.

Le système nerveux se décline en deux sous-systèmes :
- le système nerveux cérébro-spinal, au coeur des relations de l'organisme avec le milieu extérieur et par conséquent à l'origine des mouvements volontaires, automatiques ou réflexes ;
- le système nerveux végétatif ou autonome dont les fonctions sont de réguler les fonctions vitales de l'organisme, de manière involontaire et inconsciente pour l'être humain. Ce système n'intervenant en rien dans la production de la parole, il ne sera pas développé ici.

Le système nerveux cérébro-spinal est lui-même divisé en deux sous-systèmes : le système nerveux central et le système nerveux périphérique. L'ensemble des éléments qui

composent chacun de ces systèmes sont représentés sur la figure 1.1.

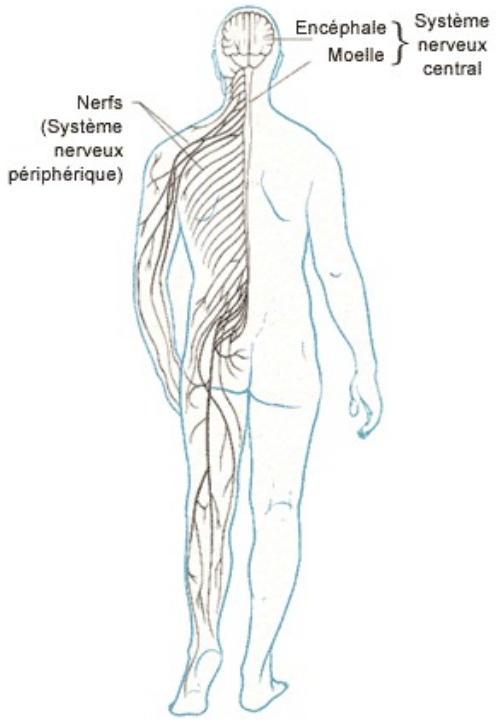


FIGURE 1.1 – *Illustration des deux sous-systèmes du système nerveux cérébro-spinal : le système nerveux central, comprenant l'encéphale (cerveau, tronc cérébral et cervelet) et la moelle épinière, et le système nerveux périphérique, composé de l'ensemble des nerfs crâniens (rattachés à l'encéphale) et rachidiens (rattachés à la moelle épinière).*

1.1.1 Le système nerveux central

Le système nerveux central comprend (1) l'encéphale localisé dans la cavité de la boîte crânienne et composé du cerveau, du tronc cérébral et du cervelet et (2) la moelle épinière située dans le canal rachidien.

Le cerveau

Le cerveau est situé dans la boite crânienne, protégé par trois membranes (les méninges). Il est composé de deux hémisphères - hémisphère droit et gauche - , quasi symétriques sur le plan anatomique, séparés par un profond sillon, mais reliés entre eux par les commissures inter-hémisphériques (voir encadré sur les commissures inter-hémisphériques). Sur le plan fonctionnel, les activités des hémisphères droit et gauche - symétriques ou asymétriques, concurrentielles ou coopératives, unilatérales ou bilatérales - ainsi que leurs interactions ont été largement débattues et le sont encore au-

Les commissures inter-hémisphériques.

Elles contiennent des fibres nerveuses (voir encadré sur les fibres nerveuses) qui établissent des relations entre les deux hémisphères cérébraux. Ces fibres sont appelées fibres d'association inter-hémisphériques. On compte quatre commissures principales : (1) le corps calleux, la plus volumineuse, reliant les territoires corticaux des deux lobes frontaux, des deux lobes pariétaux et des deux lobes occipitaux, (2) le fornix, formé de deux cordons de fibres nerveuses appelés piliers, chacun reliant l'hippocampe au tubercule mamillaire, tous deux jouant un rôle notamment dans la mémoire et l'apprentissage, (3) la commissure blanche antérieure, reliant les deux noyaux amygdaliens (placés dans le lobe temporal) qui appartiennent aux systèmes olfactif et limbique antérieur (4) et la commissure blanche postérieure, de structure très complexe, contenant des fibres d'association entre les noyaux des nerfs crâniens, et entre les deux moitiés du mésencéphale et du diencéphale.
Il existe des aires cérébrales symétriques mais non reliées fonctionnellement par des fibres du corps calleux. Elles sont appelées aires primaires. Toutes les autres régions sont connectées par des fibres calleuses. Elles sont appelées aires associatives.

Sources : http ://www.anatomie-humaine.com/Le-Cerveau-1.html

Substances grise et blanche.

Les organes du système nerveux central sont formés de deux substances :
- La matière grise : périphérique et superficielle au niveau du cerveau et du cervelet, elle est centrale dans la moelle épinière. Elle est composée des corps cellulaires des neurones, des dendrites et de la partie initiale des axones (voir encadré sur les neurones). Toutes les fibres qui pénètrent dans la substance grise perdent leur gaine de myéline.
- La matière blanche : profonde dans le cerveau et le cervelet, elle est superficielle au niveau de la moelle épinière. Elle est formée de fibres nerveuses myélinisées. Elle est de ce fait plus ferme et plus résistante que la substance grise.

Sources : http ://www.anatomie-humaine.com/Presentation-Generalites.html

Les fibres nerveuses, voies afférentes et efférentes.

On distingue deux types de fibres nerveuses :

- les **fibres sensitives** qui assurent l'acheminement des informations en provenance des récepteurs thermiques, tactiles, musculaires, articulaires et algiques vers le système nerveux central. Un ensemble de fibres nerveuses sensitives ayant une même origine, terminaison et fonction sont organisées en faisceau pour former une **voie sensitive dite ascendante ou afférente**. Parmi ces faisceaux, nous pouvons citer les *faisceaux spino-cérébelleux* transportant les informations musculaires, articulaires et tactiles, les *faisceaux spino-thalamiques* transportant les informations liées aux sensations thermiques, algiques, tactiles (tact grossier), et de pression, et les *faisceaux gracile et cunéiforme* transportant les impulsions proprioceptives et les informations liées au tactile (tact fin) ;
- les **fibres motrices** qui assurent l'acheminement des informations en provenance du système nerveux central vers les organes moteurs (e.g. les muscles). Un ensemble de fibres nerveuses motrices ayant une même origine, terminaison et fonction sont organisées pour former une **voie motrice dite descendante ou efférente** et dont l'objectif est d'assurer le contrôle des mouvements volontaires et involontaires, du tonus musculaire, mais également des fonctions autonomes et des mouvements réflexes. Nous pouvons citer les *faisceaux pyramidaux*, composés du *faisceau cortico-spinal* et *cortico-bulbaire*, assurant le contrôle volontaire fin, le *faisceau rubro-spinal* contrôlant le tonus des muscles fléchisseurs, le *faisceau tecto-spinal* contrôlant les mouvements réflexes en réponse à des stimuli visuels, les *faisceaux vestibulo-spinaux* assurant l'excitation des muscles extenseurs et finalement les *faisceaux réticulo-spinaux* contrôlant les fonctions vitales (respiration, circulation sanguine), les activités réflexes et le tonus musculaire.

Les nerfs pouvant être composés de plusieurs faisceaux de fibres nerveuses sont dits sensitifs s'ils ne contiennent que des fibres sensitives, moteurs, s'ils ne contiennent que des fibres motrices, ou mixtes s'ils contiennent les deux types de fibres.

Sources : http ://www.ufrsdv.u-bordeaux2.fr/siteIML/Maste1biosante/Master1biosantecours/NAF%20topographie%20Fenelon.pdf

jourd'hui dans la littérature. Il est établi, sur la base des connaissances anatomiques du système nerveux central (notion de décussation évoquée dans les sections suivantes), que chaque hémisphère commande les organes moteurs de la partie du corps opposé, de même pour les informations sensitives et sensorielles. Néanmoins qu'en est-il des autres fonctions du cerveau ?
Paul Broca, en 1861, est le premier à invoquer la notion de latéralisation des fonctions cérébrales dans un bulletin de la société française d'anthropologie (Broca, 1862), en démontrant la présence d'une zone de l'hémisphère gauche - l'aire de Broca - dédiée spécifiquement au langage. En effet, il observe chez certains de ses patients, ayant perdu la faculté de parler tout en gardant leur lucidité et faculté de compréhension, des lésions localisées au niveau de cette aire dans l'hémisphère gauche uniquement. Cette découverte d'une latéralisation est accréditée peu d'années après par Karl Wernicke, neurologue allemand, qui soutient la présence d'une nouvelle zone de l'hémisphère gauche - l'aire de Wernicke - dédiée spécifiquement à la compréhension du langage. En effet, les patients atteints de lésions localisées au niveau de cette aire dans l'hémisphère gauche uniquement peuvent parler, mais présentent de graves incohérences dans leur communication verbale. De très nombreuses études ont ensuite été rapportées au fil des époques se basant sur l'observation de patients atteints de lésions cérébrales de l'hémisphère droit ou gauche, de patients ayant subi une ablation d'une partie d'un hémisphère pour répondre à un besoin clinique, de patients ayant subi une section des commissures inter-hémisphériques, de patients sur lesquels une anesthésie d'un des deux hémisphères a pu être pratiquée au cours d'une opération du cerveau (test de Wada), ou plus récemment lors d'expérimentations sur des sujets sains grâce aux avancées technologiques liées à l'imagerie cérébrale fonctionnelle.
Si la théorie d'une dominance majeure voire exclusive de l'hémisphère gauche (l'hémisphère droit étant jugé mineur, voire inutile) pour toutes les activités cognitives a perduré une bonne partie du XX^e siècle, les travaux de Roger Sperry dans les années 60 ont permis la "réhabilitation" de l'hémisphère droit. En effet, ces travaux et d'autres réalisés par la suite ont permis de montrer des spécialisations fonctionnelles différentes du traitement de l'information de la part de chaque hémisphère. En d'autres termes, chaque hémisphère traiterait un type d'informations donné de manière différente, seul ou de manière dominante vis à vis de l'autre hémisphère. Ainsi, l'hémisphère droit interviendrait plutôt dans le traitement des mouvements spatiaux, dans la cognition spatiale, dans le traitement de l'image et de la musique et plus généralement dans la communication non verbale tandis que l'hémisphère gauche jouerait un rôle dans le contrôle des séquences des mouvements volontaires et dans la communication verbale. Une autre dichotomie attribuerait à l'hémisphère gauche une efficacité accrue dans le traitement local et séquentiel de l'information contrairement à la nature plus globale du traitement dans l'hémisphère droit. Cette dernière est donnée sous la terminologie "modèle des fréquences" par Joseph Dien dans (Dien, 2008) qui propose une revue des cinq théories les plus connues avant de proposer la sienne.

Finalement, malgré les avancées technologiques de ces dernières années, la complexité du cerveau tant au niveau de ses activités fonctionnelles, de son fonctionnement (rôle de chaque hémisphère et des commissures inter-hémisphériques, leur interaction)

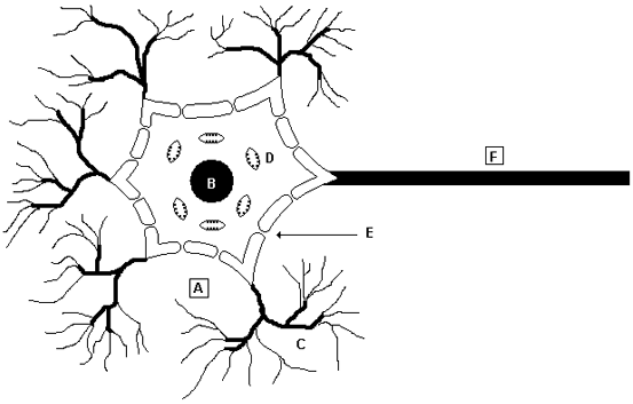
que de son évolution (notion de plasticité du cerveau) est loin d'être maîtrisée par les scientifiques encore aujourd'hui et reste un domaine de recherche très ouvert.

D'un point de vue purement anatomique, la surface visible des deux hémisphères est recouverte d'une couche de substance grise (voir encadré sur les substances grise et blanche) constituant le cortex cérébral. Cette surface présente de nombreux plis, délimitant, pour les plis les plus profonds, différentes parties du cerveau. Trois sillons principaux situés sur la face latérale de chaque hémisphère délimitent ainsi les lobes frontal, pariétal et temporal étoccipital. D'autres sillons moins prononcés situés sur les différents lobes vont délimiter des circonvolutions cérébrales appelées gyrus. On soulignera également la présence, dans ces différentes parties du cerveau, de l'hippocampe (impliqué dans le processus de mémorisation et de perception dans l'espace par exemple) et d'un 5e lobe cérébral, le lobe de l'insula.
Outre le cortex cérébral qu'on retrouve en surface du cerveau, ce dernier se compose principalement de substance blanche au sein de laquelle sont répartis des amas de substance grise, appelés les noyaux gris centraux ou encore ganglions de la base (voir encadré sur les noyaux gris centraux) et un ensemble de cavités centrales appelées ventricules cérébraux, contenant du liquide cérébro-spinal. Par ailleurs, la substance blanche est délimitée en plusieurs zones : les capsules interne, externe et extrême et le centre ovale dont certaines sont formées par le passage de faisceaux de fibres nerveuses intervenant dans la motricité (voir section 1.1.3 sur les voies pyramidales).
En 1905, puis en 1909, Korbinian Brodmann publia des cartographies du cerveau humain (et de primates), permettant d'identifier différentes zones du cortex cérébral en fonction de la forme des neurones (voir encadré sur les neurones) et des réseaux qui la constituent (analyse cyto-architectonique). Ces zones, appelées aires de Brodmann et illustrées par la figure 1.2, ont fait l'objet de nombreux débats au cours des années. Elles restent néanmoins d'actualité, notamment en raison de corrélations établies entre certaines de ces aires (prises individuellement, partiellement ou groupées avec d'autres) et des fonctions propres du cerveau comme les fonctions sensitives et sensorielles ou encore la motricité, qui nous intéresse tout particulièrement ici.
Il est établi que la fonction de motricité est étroitement liée à un sous-ensemble de ces aires, comme nous le verrons au cours des sections suivantes, dont l'aire 4 (également appelée aire somato-motrice ou cortex moteur primaire et notée M1), typiquement dédiée à la motricité en raison de sa composition corticale très particulière. En effet, l'aire 4 est singulièrement composée de cellules motrices pyramidales uniquement (grandes et petites cellules), se traduisant localement par l'absence des couches I, II, IV, VI porteuses des cellules sensitives et sensorielles et des cellules d'association au sein du cortex (voir encadré sur la composition du cortex). L'étude approfondie de cette aire par W. Penfield, neurochirurgien canadien, entre les années 30 et 50, par le biais de la stimulation électrique du cortex cérébral permet à ce dernier de co-publier une cartographie, appelé Homoncule moteur [1], illustrant la répartition des commandes motrices des muscles du corps humain sur l'aire 4 (voir figure 1.3). De manière pragmatique, la stimulation électrique d'une zone particulière de l'aire 4 entraîne systématiquement

1. une cartographie similaire est proposée dans la même publication pour les aires 1, 2 et 3 liées aux fonctions sensitives.

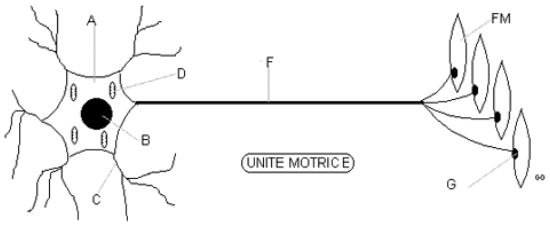
Le neurone.

Le **neurone** ou cellule nerveuse est l'unité fondamentale du système nerveux, dont le rôle est de transmettre un **influx nerveux** d'un neurone à l'autre ou d'un neurone vers l'organe innervé. Ce flux nerveux, porteur de messages tels que des stimulations sensitives ou sensorielles, des ordres de réponse, des réponses réflexes, etc., est transmis d'un neurone à un autre par l'intermédiaire de substances chimiques : les **neurotransmetteurs**.
Un neurone est composé d'un (1) corps cellulaire contenant le noyau de la cellule et présentant une ou plusieurs ramifications ou **dendrites**, reliées par des connexions ou **synapses** à d'autres neurones, (2) d'un **axone**, prolongement du corps cellulaire, possédant des branches collatérales et se terminant par une arborisation de fibres dont chacune des branches aboutit à la **plaque motrice** d'une **fibre musculaire**, dans le cas d'un axone moteur. Recouvert de gaines isolantes (gaines de **myéline** et/ou gaines de **Schwann**), l'axone prend le nom de **cylindraxe** ou **fibre nerveuse**. Un ensemble de fibres nerveuses parallèles ayant toutes même origine et même destination forme un **faisceau**.
L'ensemble du corps cellulaire d'un neurone moteur, appelé **moto-neurone**, de l'axone et des fibres musculaires qui en dépendent, constitue l'**unité motrice**.



- La cellule nerveuse - Morphologie

A : Corps cellulaire. B : Noyau. C : Dendrite. D : Mitochondries. E : canaux ou pores membranaires
F : Axone.

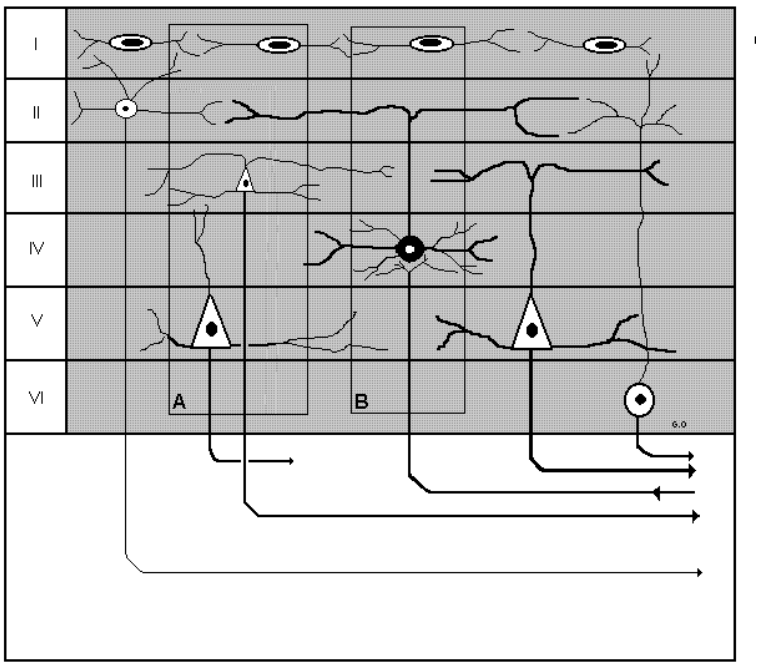


La cellule nerveuse motrice (motoneurone), son axone et les fibres musculaires qui en dépendent constituent l' UNITE MOTRICE.
A: Corps cellulaire B: Noyau C: Dendrites F: Axone FM: Fibres Musculaires
G: Plaque motrice des fibres musculaires

Composition d'un neurone et d'une unité motrice. Illustration des différents composants d'un neurone ou cellule nerveuse, unité fondamentale du système nerveux, et d'une unité motrice impliquant un neurone particulier dit moteur : le motoneurone. (Source : http ://www.anatomie-humaine.com/Presentation-Generalites.html)

Le cortex cérébral.

Le cortex cérébral est composé de substance grise, ie de cellules nerveuses disposées dans sa grande majorité en 6 couches différentes comprenant des cellules d'association (couches I, II et VI), des cellules réceptrices des informations sensitivo-sensorielles (couche IV) et des cellules motrices pyramidales petites et grandes, les grandes étant appelées cellules pyramidales de Betz. L'illustration suivante retrace la disposition de ces différentes couches. Une organisation en colonnes verticales est également visible sur le schéma (colonnes A et B), mettant en jeu des types de cellules différentes mais liées au niveau fonctionnel ie intervenant dans la même activité cérébrale.



I : cellules d'association superficielles. II : cellules d'association intra-hémisphériques.
III : petites cellules pyramidales. IV : Cellules de projection sensitives et sensorielles.
V : grandes cellules pyramidales de Betz (origine du faisceau pyramidal).
VI : cellules d'association inter- hémisphériques (fibres calleuses).
A et B: Structure fonctionnelle en colonnes

Source : http ://www.anatomie-humaine.com/Le-Cerveau-1.html

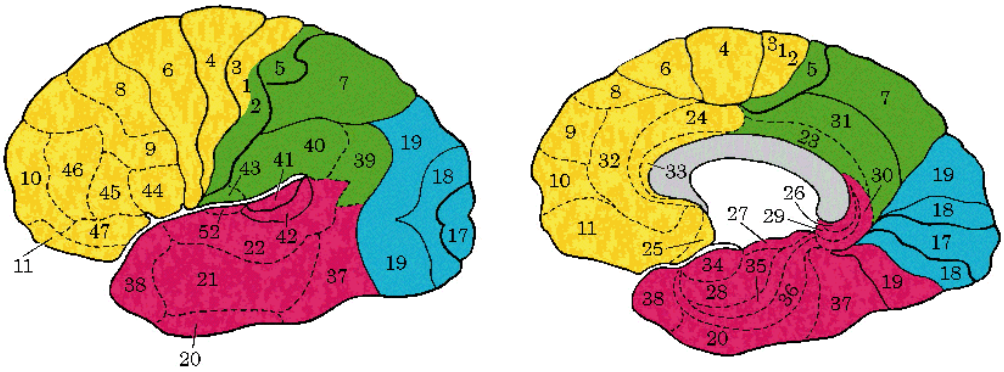


FIGURE 1.2 – *Cartographie du cortex cérébral en zones numérotées en fonction de la forme et des réseaux qui la constituent. Coupe latérale du cerveau (à gauche), coupe médiale du cerveau (à droite) (source : http ://geeam.free.fr).*

une contraction musculaire localisée dans une zone précise du corps humain. Outre l'apport considérable de cette étude sur la connaissance du cerveau humain, Penfield démontre également que la complexité des mouvements mis en oeuvre par les zones anatomiques désignées sur sa cartographie, en terme de précision d'exécution notamment, est proportionnelle à la surface de l'aire 4 correspondante. Ainsi, la surface de l'aire correspondant à la main est largement plus étendue que celle de la cuisse par exemple.

Le tronc cérébral

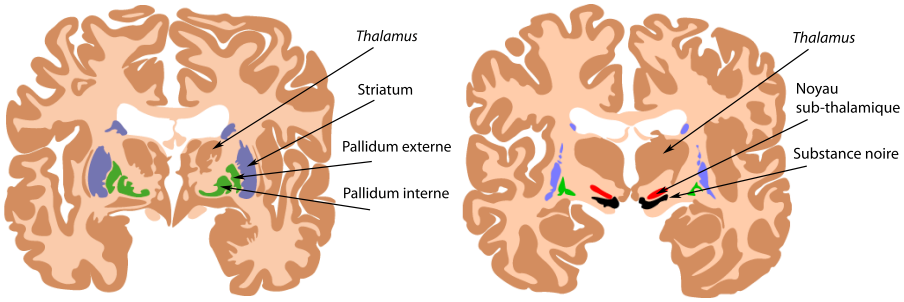
Le tronc cérébral est situé au dessus de la moelle épinière, sous le cerveau et en avant du cervelet. Il comprend trois structures différentes : la moelle allongée, également appelée bulbe rachidien, la protubérance annulaire, également appelée le pont, et le mésencéphale.
Il est composé en interne de substance blanche et grise. Dans cette dernière, se dessinent différents noyaux, divisés en deux sous-groupes : (1) les noyaux ou centres segmentaires, noyaux d'origine des nerfs crâniens (voir section 1.1.2) et les noyaux ou centres supra-segmentaires constituant des relais moteurs, sensitifs ou sensoriels, illustrés sur la figure 1.4.

Le cervelet

Le cervelet est localisé dans la boite crânienne dans une fosse appelée : fosse crânienne ou cérébelleuse et recouvert, similairement au cerveau, par les méninges. Sous-entité de l'encéphale, il est situé à la base postérieure du cerveau, accolé en partie à la face postérieure du tronc cérébral. Il est connecté à ce dernier au niveau du bulbe rachidien, de la protubérance annulaire et du mésencéphale par l'intermédiaire de trois cordons de substance blanche, les pédoncules cérébelleux inférieur, moyen et supérieur respectivement.

Les noyaux gris centraux.

Ils désignent un ensemble fermé de noyaux plus ou moins volumineux composés de substance grise sous-corticale, entourés de substance blanche et localisés au centre du cerveau. On les retrouve également sous la terminologie de ganglions de la base (basal ganglia/nuclei en anglais). La composition de cet ensemble diverge suivant les sources, incluant ou non d'autres noyaux du cerveau dont le thalamus. Sur une base plutôt commune, nous pouvons citer : (1) le striatum (également appelé noyau strié), 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 également en substance noire pars compacta (dont la particularité tient à ses neurones qui libèrent de la dopamine comme neuro-transmetteur) et pars reticula et finalement (4) le noyau subthalamique (aussi appelé corps de Luys).
Tous ces noyaux gris sont interconnectés par un réseau dense de voies afférentes et efférentes, excitatrices (le neurotransmetteur étant le glutamate) ou inhibitrices (le neurotransmetteur étant l'acide gamma-amino-butyrique : GABA) et impliqués dans différentes activités cérébrales dont le contrôle moteur et le processus d'apprentissage. Cette implication est néanmoins indirecte. En effet, dans le cas de la motricité, ces noyaux vont intervenir en amont du processus d'exécution du mouvement en proposant une boucle de régulation (appelée ici "boucle motrice") dont l'objectif est de définir et ajuster ce dernier. Le lecteur pourra se référer à (Pinto, 2007) pour une description détaillée de la boucle motrice.



Sources : http ://www.ibags.info/, (Pinto, 2007)

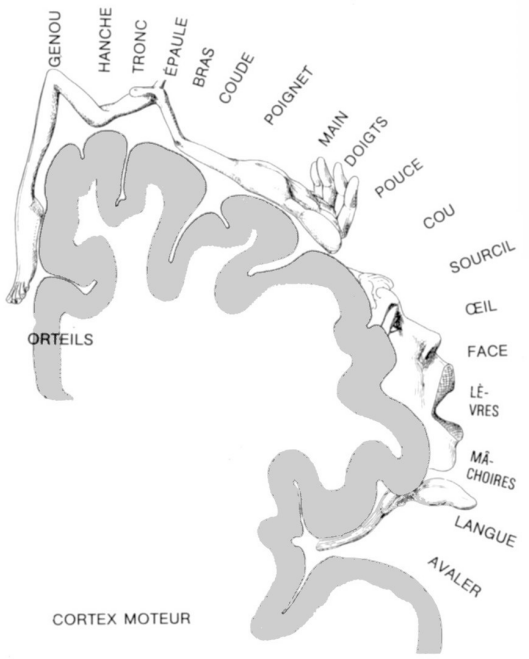


FIGURE 1.3 – *Illustration de W. Penfield de la répartition des commandes motrices des muscles du corps humain au niveau de l'aire somato-motrice (aire 4 de Brodmann) en montrant la relation proportionnelle entre étendue de la zone sur l'aire et complexité des mouvements mis en oeuvre par la zone du corps humain concernée.*

Le cervelet est composé d'un lobe médian appelé le vermis, auquel sont accolés de part et d'autre deux hémisphères cérébelleux (ou lobes cérébelleux). Un quatrième lobe, le lobe flocculo-nodulaire termine cette composition. Le cervelet est généralement divisé en sous-régions, chacune intégrant un composant structurel différent de ce dernier, et dédiée à une activité cérébrale bien définie :

- l'archéocervelet qui comprend le lobe flocculo-nodulaire et intervient dans le processus d'équilibre ;
- le paléocervelet qui comprend principalement le vermis et intervient dans la régulation de l'activité musculaire par action sur le tonus musculaire ;
- le néocervelet qui comprend les deux hémisphères cérébelleux et contribue à la coordination et précision des mouvements volontaires par action sur les muscles antagonistes.

En interne, le cervelet se compose, de manière quasi similaire au cerveau, de trois structures cellulaires. La première, localisée en périphérie, est composée de substance grise, sous une forme très plissée constituant l'écorce cérébelleuse, appelée cortex cérébelleux. Ce dernier comporte trois couches différentes - couche moléculaire en périphérie, couche des cellules de Purkinje (cellules les plus volumineuses du cerveau), et la couche granulaire - qui sont le lieu d'une convergence importante d'informations en

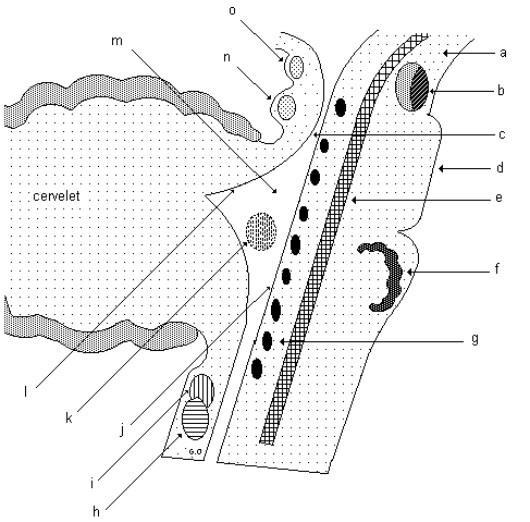


FIGURE *1.4* – *Illustration des différents noyaux présents dans la substance grise du tronc cérébral (source : http ://www.anatomie-humaine.com/IMG/gif/_S.77_T._C._Morphologie_interne.gif).*

provenance du cortex cérébral, du tronc cérébral ou de la moelle épinière formant des voies afférentes et efférentes à l'origine des activités citées ci-dessus.
La seconde structure est composée de substance blanche contenant d'une part des fibres nerveuses myélinisées efférentes (axones des cellules de Purkinje) connectées aux noyaux gris, dernière structure du cervelet et d'autre part des fibres afférentes, provenant du cortex, de la moelle épinière et du tronc cérébral et s'articulant avec les dendrites des cellules Purkinje. Quatre paires de noyaux gris (troisième structure) sont localisées en profondeur à l'intérieur du cervelet (chaque paire répartie entre les deux hémisphères) : noyau du toit ou fastigial, noyau emboliforme, noyau globuleux et le noyau dentelé (ou denté), qui servent de relais entre le cortex cérébelleux et les autres composantes du système nerveux central.

La moelle épinière

La moelle épinière est le prolongement de l'encéphale et plus particulièrement de la moelle allongée du tronc cérébral. D'une longueur variant entre 40 et 50cm, elle est située dans le canal rachidien. Elle est protégée par les méninges et vertèbres et s'étend jusque dans le bas du dos, au niveau de la deuxième vertèbre lombaire. Elle présente deux renflements indiquant chacun le départ d'un réseau formé par l'entrelacement de nerfs appelé plexus : le renflement cervical lié au plexus brachial (innervant princi-

palement les membres supérieurs) et le renflement lombaire lié au plexus lombo-sacré (innervant principalement les membres inférieurs).

Une coupe transversale de la moelle épinière met en évidence la présence de substance grise et blanche disposée à l'inverse comparé aux différents composants de l'encéphale (voir figure 1.5). Ici, la substance grise est positionnée au centre de la moelle, entourée par une couche de substance blanche. Elle est généralement assimilée à la forme d'un papillon ou de la lettre H, dont les deux parties latérales, reliées par la commissure grise, constituent en amont les cornes antérieures et en aval les cornes postérieures. Leurs centres sont appelés cornes latérales. A chaque type de cornes correspond un type de neurones différents : (a) les cornes antérieures (ou ventrales) contiennent les neurones moteurs dont les axones forment les fibres motrices des nerfs spinaux (ou rachidiens - voir section suivante pour plus de détails sur les nerfs), (b) les cornes postérieures (ou dorsales) contiennent, d'une part, les neurones sensitifs des fibres sensitives de ces mêmes nerfs spinaux et, d'autre part, des neurones participant aux mouvements réflexes, et finalement (c) les cornes latérales contiennent les moto-neurones viscéraux relatifs aux mouvements involontaires.

De son côte, la substance blanche est formée de fibres nerveuses recouvertes de leurs gaines et groupées en faisceaux. Ces faisceaux sont divisés en trois zones de substance blanche appelées cordons ventraux, latéraux et dorsaux. Finalement, ces faisceaux émergent de chaque côté de la moelle épinière soit par la racine antérieure (ou dorsale, prolongement de la corne antérieure) prolongée par le ganglion spinal pour les fibres sensitives, soit par la racine postérieure (ou ventrale, prolongement de la corne postérieure) pour les fibres motrices afin de former un nerf rachidien (nerf mixte). Ce schéma se répète, tout au long de la colonne vertébrale pour chaque nerf rachidien émergeant de la moelle épinière, formant à chaque niveau un segment constitué de la paire de nerfs spinaux et des organes innervés.

1.1.2 Le système nerveux périphérique

Le système nerveux périphérique comprend l'ensemble des nerfs rattachés à l'encéphale, dénommés nerfs crâniens et ceux rattachés à la moelle épinière dénommés nerfs spinaux ou rachidiens. L'ensemble de ces nerfs sont situés à l'extérieur de la boîte crânienne ou du canal rachidien.

Les nerfs crâniens

On dénombre 12 nerfs crâniens émergeant de la boite crânienne, nommés de n.I à n.XII[2]. Leur noyau d'origine est localisé au niveau du tronc cérébral excepté pour les deux premiers (n.I et n.II). Les nerfs crâniens peuvent être moteurs (n.III, n.IV, n.VI,

2. n pour nervus

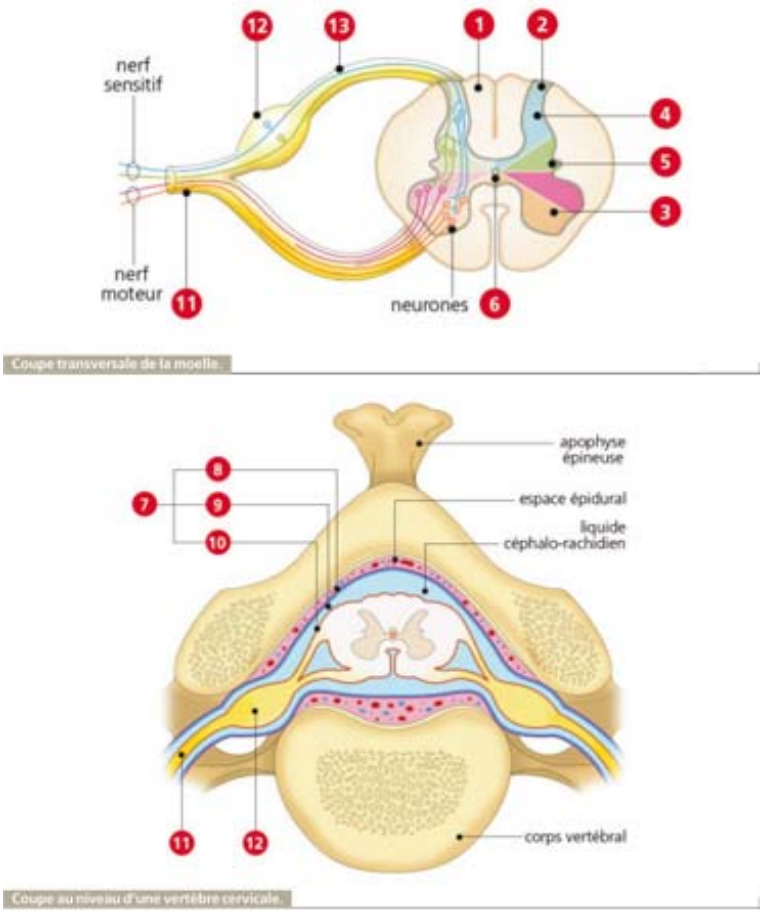


FIGURE 1.5 – *Illustration des différents composants de la moelle épinière : (1) substance blanche, (2) substance grise, (3) cornes antérieures, (4) cornes postérieures, (5) cornes latérales, (6) canal de l'épendyme, (7) trois méninges (8) dure-mère, (9) arachnoïde, (10) pie-mère, (11) nerf rachidien, (12) ganglion spinal, (13) racine antérieure (source : Encyclopédie le corps humain, édition Atlas).*

n.XI, n.XII), sensitivo-sensoriels (n.I, n.II, n.VIII), mixtes ie sensori-moteur (n.V, n.VII, n.IX, n.X) voire pour certains d'entres eux végétatifs (n.III, n.VII, n.IX, n.X). Le tableau 1.1 fournit la dénomination, le type, les fonctions, la ou les zone(s) innervée(s) par chaque nerf crânien ainsi que la localisation de leur noyau d'origine dans l'encéphale. Il est à noter que les informations relatives aux fibres végétatives ne sont pas développées ici.

Dénomination	Type	Fonctions	Zone(s) innervée(s)	Localisation noyau d'origine
n.I : Olfactif	Sensoriel	Olfaction	Fosses nasales	Bulbe olfactif
n.II : Optique	Sensoriel	Vision	Rétine	Chiasma optique
n.III : Oculo-moteur	Moteur	Motricité des muscles de l'oeil	Région oculaire	Mésencéphale, tubercules quadrijumeaux antérieurs
n.IV : Trochléaire (ou pathétique)	Moteur	Motricité des muscles de l'oeil	Région oculaire	Mésencéphale, tubercules quadrijumeaux postérieurs
n.V : Trijumeau	Sensitif, moteur	Sensibilité d'une grande majorité de la face, motricité des muscles de la mastication et tenseur du tympan	Face	Protubérance
n.VI : Abducens (ou oculo-moteur externe)	Moteur	Muscles de l'oeil	Région oculaire	Bulbe, protubérance
n.VII : Facial	Sensitivo-sensoriel, moteur	Gustation, sensibilité au niveau de la peau du pavillon de l'oreille, motricité des muscles peauciers de la face et du cou	Face	Bulbe, protubérance
n.VIII : Auditif (ou vestibulo-cochléaire)	Sensoriel	Audition, équilibre	Canal auditif	Bulbe, protubérance
n.IX : Glosso-pharyngien	Sensitivo-sensoriel, moteur	Gustation, sensibilité du pharynx, motricité bucco-pharyngée	Base de la langue	Bulbe
n.X : Vague (ou pneumo-gastrique)	Sensitivo-sensoriel, moteur	Gustation, sensibilité des muqueuses inférieures du larynx/pharynx et de la peau, du conduit auditif, motricité partielle du pharynx, larynx et voile du palais	Abdomen et Thorax	Bulbe
n.XI : Spinal (ou accessoire)	Moteur	Motricité du muscle sterno-cléido-mastoïdien (muscle du cou) et du muscle trapèze	Cou	Bulbe
n.XII : Hypoglosse	Moteur	Motricité des muscles de la langue (mastication, déglutition, parole)	Base de la langue	Bulbe

TABLE *1.1* – *Descriptif des douze nerfs crâniens : dénomination, type, fonctions, zone(s) innervée(s) et localisation du noyau d'origine dans l'encéphale.*

Les nerfs spinaux

Les nerfs spinaux sont au nombre de 31 paires, émergeant de la moelle épinière (à gauche et à droite) entre deux vertèbres le long de la colonne vertébrale. Comme précisé précédemment, chacune de ces paires de nerfs sont mixtes, véhiculant à la fois les informations sensitives (des zones du corps humain vers le système nerveux central) et motrices (du système nerveux central vers les zones du corps humain). Ils sont par conséquent responsables de la motricité et de la sensibilité des membres (peau et muscles), sphincters [3] et du périnée. Parmi ces 31 paires, on distingue :

- 8 nerfs cervicaux (nommés de C1 à C8) ;
- 12 nerfs thoraciques (nommés de T1 à T12) ;
- 5 nerfs lombaires (nommés L1 à L5) ;
- 5 nerfs sacrés (nommés de S1 à S5) ;
- 1 nerf coccygien (nommé Co).

1.1.3 Les voies motrices

On distingue trois grandes voies motrices au sein du système nerveux :

- le système des voies pyramidales : à l'origine des mouvements volontaires, il permet la conduction de l'influx nerveux du cortex cérébral vers les muscles innervés ;
- le système des voies extra-pyramidales : à l'origine des mouvements automatiques, il permet la régulation des mouvements volontaires (tonus musculaire, posture, équilibre, ...) ;
- le système des voies réflexes : à l'origine des mouvements réflexes (inconscients) d'un ou plusieurs muscles suite à un stimulus particulier. Ce système fait intervenir soit la moelle épinière (réflexes spinaux), soit le tronc cérébral (réflexes crâniens) en tant que centre d'intégration/réflexe recevant un stimulus et produisant une réponse appropriée sous la forme d'une contraction musculaire. Ce système particulier ne sera pas développé ici.

Le système des voies pyramidales

Les voies pyramidales prennent naissance dans le cortex cérébral au niveau des neurones corticaux. Elles prennent ensuite la forme de deux faisceaux de fibres nerveuses (axones des neurones) - faisceau cortico-bulbaire (ou cortico-nucléaire) et faisceau cortico-spinal. Elles traversent la substance blanche du cerveau au niveau de la capsule interne, puis descendent dans le tronc cérébral et la moelle épinière pour faire jonction avec les nerfs crâniens et les nerfs spinaux respectivement. Ces voies sont dites directes ou mono-synaptiques dans le sens où l'influx nerveux est véhiculé du cortex vers le muscle innervé via deux neurones seulement et une seule synapse, comme

3. Muscle en forme d'anneau qui entoure un orifice du corps humain et qui en assure l'occlusion ou l'ouverture.

illustré en figure 1.6. Ainsi, dans le cadre du faisceau cortico-bulbaire, un seul neurone, nommé moto-neurone central ou supérieur, relie le cortex au noyau d'un nerf crânien, localisé dans le bulbe rachidien du tronc cérébral ; un second neurone, part de ce noyau pour innerver le muscle correspondant. Un schéma similaire se retrouve dans le cadre du faisceau cortico-spinal pour lequel le second neurone est localisé au niveau de la moelle épinière.

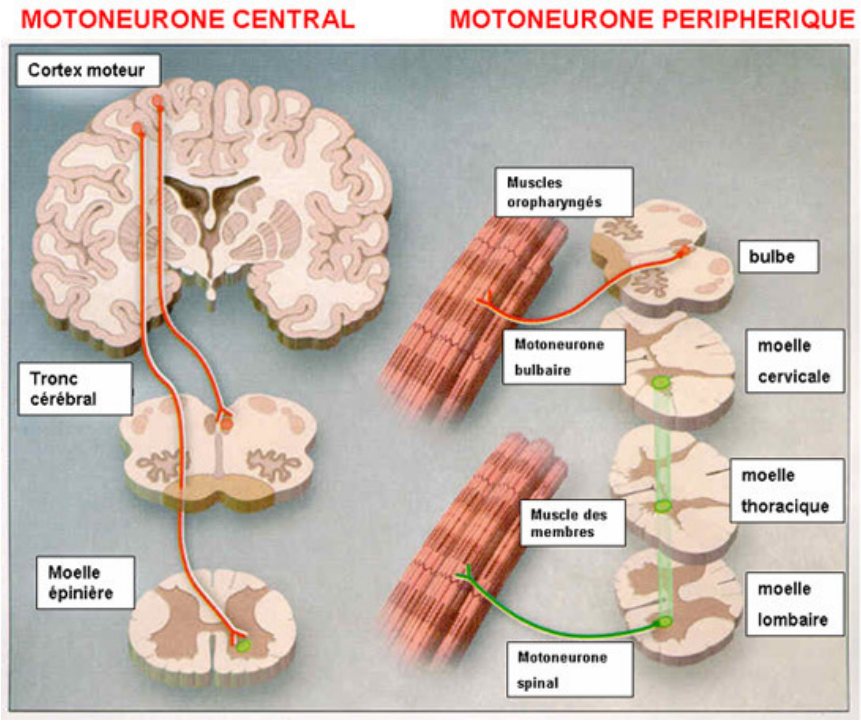


FIGURE *1.6* – *Illustration (1) du faisceau cortico-bulbaire reliant le cortex aux noyaux (localisés au niveau du bulbe rachidien du tronc cérébral) des nerfs crâniens par le biais des moto-neurones centraux, puis reliant ces noyaux aux muscles par le biais des moto-neurones périphériques et (2) le faisceau cortico-spinal reliant le cortex aux noyaux (localisés au niveau de la moelle épinière) des nerfs spinaux, toujours par le biais des moto-neurones centraux, puis reliant ces noyaux aux muscles par le biais des moto-neurones périphériques.*

En relation avec les nerfs crâniens (voir section 1.1.2), le faisceau cortico-bulbaire est par conséquent impliqué dans le contrôle des mouvements de la tête et du cou. Il est constitué des axones des neurones pyramidaux de la couche V du cortex moteur primaire M1 (aire 4 de Brodmann).

Le faisceau cortico-spinal est, quant à lui, le support des mouvements volontaires du corps humain liés aux nerfs spinaux : muscles du cou, tronc, membres supérieurs et inférieurs, ... Il est composé d'axones dont la provenance est davantage diversifiée. En effet, il comprend des axones du cortex moteur primaire (M1), du cortex pré-moteur et aire motrice supplémentaire (AMS) (tous deux formant l'aire 6 de Brodmann) et du

gyrus cingulaire antérieur mais également des aires 1, 2, 3 (cortex somato-sensoriel primaire) et des aires 5 et 7 (cortex somato-sensoriel associatif). Une fois constitué, le faisceau descend dans le tronc cérébral pour atteindre le bulbe rachidien. A ce niveau, 10% des fibres (formant le faisceau pyramidal direct) vont poursuivre leur trajet latéralement dans le cordon ventral de la moelle épinière. A chaque segment de la moelle, une partie d'entre elles se détachent, traversent la ligne médiane de la moelle pour faire finalement jonction avec des interneurones (eux-mêmes reliés à des motoneurones) de la corne antérieure opposée. Les 90% restant (faisceau pyramidal croisé) vont se décusser au niveau du bulbe et poursuivre leur descente dans le cordon latéral, côté opposé, de la moelle épinière pour atteindre également les motoneurones de la corne antérieure, localisés sur différents segments de la moelle. Ainsi, le faisceau cortico-spinal croisé contrôle les muscles distaux des membres, responsables des mouvements fins et précis tandis que le faisceau direct contrôle les muscles axiaux impliqués dans les mouvements grossiers et les ajustements de postures.

Le système des voies extra-pyramidales

Les voies extra-pyramidales naissent pour leur majorité dans les noyaux du tronc cérébral. Contrairement aux voies pyramidales dites directes en raison de leurs liaisons mono-synaptiques entre le cortex cérébral et les muscles innervés, les voies extra-pyramidales sont dites indirectes et poly-synaptiques. En effet, elles empruntent différents trajets, et surtout font intervenir différentes structures de l'encéphale avant d'atteindre les cordons de la moelle épinière, pour faire jonction finalement avec les interneurones et moto-neurones, associés aux différents muscles innervés. Par ailleurs, même si elles interviennent dans le cadre des mouvements volontaires, elles sont à l'origine de mouvements automatiques permettant de moduler l'activité de ces derniers (sous le contrôle des voies pyramidales), de contrôler le tonus musculaire et d'intervenir sur l'équilibre et la posture du corps humain.
Les voies extra-pyramidales sont pour la grande majorité originaires des noyaux gris du tronc cérébral (d'où elles tirent leurs noms respectifs) et comprennent les faisceaux suivants :

- le faisceau rubro-spinal : prend naissance dans le noyau rouge du tronc cérébral. Sous l'influence directe du cortex cérébral (via la voie cortico-rubrique), et indirecte du cervelet et du pallidum (noyau gris central), il vient seconder le faisceau cortico-spinal (voies pyramidales) en agissant sur la flexion des muscles des membres contro-latéraux. Après décussation, le faisceau descend dans le cordon latéral de la moelle pour rejoindre le faisceau cortico-spinal ;
- le faisceau tecto-spinal : prend naissance dans le colliculus supérieur du mésencéphale. Sous l'influence d'informations afférentes de type visuel (en provenance de la rétine), somato-sensorielles et auditives, ce faisceau intervient dans les mouvements automatiques de la tête et du cou pour moduler l'orientation du regard en réponse à l'environnement extérieur. Après décussation, le faisceau descend dans le cordon ventral de la moelle pour atteindre les segments relatifs aux nerfs cervicaux C1 à C4 ;

- les faisceaux vestibulo-spinaux médian et latéral : prennent naissance dans le noyau vestibulaire. Intégrant des informations en provenance de l'oreille interne, les faisceaux descendent dans la moelle épinière, sans décussation, le faisceau médian, de manière bilatérale dans les cordons latéraux pour contrôler la position de la tête lors des déplacements, le faisceau latéral, dans le cordon ventral (sans décussation), pour contrôler le tonus des muscles extenseurs pour le maintien de la posture et de l'équilibre.
- les faisceaux réticulo-spinaux médian et latéral : prennent naissance au niveau du tronc cérébral dans la substance réticulée [4] après intégration d'informations en provenance pour une grande part du cortex cérébral (via la voie cortico-réticulaire) et, indirectement, du cervelet. Les faisceaux médian et latéral descendent dans le cordon ventral et latéral de la moelle respectivement sans décussation. Ils facilitent ou inhibent l'activité des moto-neurones des muscles innervés, assurant ainsi le maintien de la posture et l'équilibre.

Boucles motrices de régulation

En complément des voies motrices pyramidales et extra-pyramidales d'origines diverses, mais partageant une même destination, la moelle épinière, il existe d'autres éléments de contrôle, régulation et correction des mouvements volontaires : les boucles motrices de régulation. Contrairement aux voies motrices citées précédemment, les boucles motrices ne font pas intervenir la moelle épinière et n'agissent donc pas directement sur les noyaux moteurs des nerfs. Elles interviennent en amont de ces dernières, dans la préparation du mouvement volontaire, en régulant l'activité des neurones moteurs du cortex ou du tronc cérébral, ou en parallèle pour corriger le mouvement. Cette régulation intervient dans le choix et le contrôle des mouvements à exécuter mais également dans le choix des paramètres des mouvements en fonction de l'environnement extérieur : paramètres de durée, de direction et de force (liée notamment au tonus musculaire approprié). Ces boucles font intervenir différentes structures de l'encéphale, dont notamment le cervelet et les noyaux gris centraux.

La boucle cortico-ponto-cérébello-thalamo-corticale permet au cervelet de recevoir des afférences corticales (informations sensitivo-sensorielles) en provenance des aires associées à la motricité. Celles-ci, une fois analysées, lui permettent de participer à la programmation du mouvement et de ses paramètres (séquencement et durée) et aux ajustements de posture à des fins de maintien de l'équilibre par renvoi d'efférences motrices vers le cortex via le thalamus. Le cervelet participe donc par cette boucle à la programmation de mouvements fins, précis et harmonieux.

De son côté, la boucle cortico-striato-pallido-thalamo-corticale permet aux noyaux gris centraux de recevoir des afférences corticales provenant de différentes aires du

4. partie du système réticulaire qui s'étend de la moelle épinière, en passant par le tronc cérébral, pour atteindre le thalamus, et projeter ses voies dans le cortex cérébral. Le système réticulaire est prégnant dans les différentes activités cérébrales inhérentes aux systèmes végétatif, moteur et sensitif.

cortex cérébral, notamment le cortex moteur et les aires somato-sensorielles. Ces afférences lui permettent de définir des mouvements adaptés et précis (en supprimant des mouvements parasites par exemple) par ajustement de leurs paramètres de force (par action sur le tonus musculaire), de direction et d'amplitude avant de renvoyer des efférences via le thalamus vers les différentes aires motrices du cortex cérébral. Il est à noter que cette boucle fait intervenir deux voies différentes, une voie directe (striato-pallidale) facilitant les mouvements choisis (via l'activation d'une partie de l'aire 6 - l'aire motrice supplémentaire - par libération de l'activité thalamique), une voie indirecte (striato-subthalamo-pallidale) inhibant les mouvements parasites (par freinage de l'activité thalamique).

Suivant le modèle "classique" de fonctionnement des noyaux gris centraux proposé par (Albin et al., 1989), ces deux voies ont pour entrée le striatum (réception des afférences). Celui-ci a pour rôle d'influencer l'activité (inhibitrice ou excitatrice) qu'exercent le pallidum interne et la substance noire pars reticula, considérés comme les points de sortie, sur leur cible, en l'occurrence ici le thalamus. Cette influence est soit directe (liaison entrée-sortie : voie striato-pallidale), soit indirecte ie faisant intervenir successivement le pallidum externe et le noyau sub-thalamique (voie striato-subthalamo-pallidale). Cette influence et l'équilibre fonctionnel des deux voies seraient dus à la dopamine, neurotransmetteur libérée par la substance noire pars compacta vers le striatum (voie nigro-striatale), qui suivant le type de récepteurs dopaminergiques touchés (D1 ou D2), présenterait un comportement opposé.

Il est également établi que cette boucle intervient dans l'automatisation des mouvements par implication de la mémorisation dite procédurale, l'attention et le choix des stratégies motrices. Par ailleurs, il est à noter que d'autres boucles non motrices impliquent les noyaux gris centraux et d'autres structures cérébrales. Ces boucles interviennent dans le traitement d'informations cognitives et comportementales (émotion, motivation, attention, vigilance, ...).

1.1.4 La parole : acte moteur volontaire

La parole est considérée comme un des supports du langage (langage parlé) au même titre que l'écriture (langage écrit). La notion de support sous-entend que la parole, en tant qu'acte moteur, ne peut intervenir qu'après l'élaboration de l'acte langagier proprement dit. Comme nous l'avions précisé en introduction, la parole, en tant qu'acte moteur volontaire, suit un schéma similaire à celui d'un mouvement volontaire quelconque. Ainsi, une fois l'acte langagier établi, s'en suit une phase d'élaboration et de planification du mouvement de parole inhérent à l'acte, une phase de programmation de ce mouvement, pour finir par une phase d'exécution.

Dans le cadre plus général d'un mouvement volontaire, la phase d'élaboration du mouvement impliquerait l'activation des aires corticales pré-frontales (aires 8, 9 et 10)

pour la planification du mouvement, en association avec le cortex somato-sensoriel associatif (aires 5 et 7) et les noyaux gris centraux, ces derniers intervenant davantage sur des processus comportementaux tels que la vigilance, l'attention ou encore la motivation. La phase de programmation du mouvement qui consiste, plus précisément, à établir la séquence de contractions musculaires nécessaires au mouvement, impliquerait l'activation de l'aire motrice supplémentaire et du cortex pré-moteur (aire 6) en association avec le cervelet et les noyaux gris centraux au travers des boucles motrices de régulation. Finalement, la phase d'exécution du mouvement, qui consiste à produire et transmettre aux muscles les influx nerveux nécessaires à l'établissement du mouvement, impliquerait l'activation du cortex primaire moteur (aire 4), les voies pyramidales et extra-pyramidales, le tronc cérébral et les nerfs crâniens associés et finalement, la moelle épinière et les nerfs spinaux associés.
Nous allons voir par la suite comment ces différents processus opèrent dans l'acte complexe de production de la parole. Le lecteur pourra également se référer à (Pinto, 2007; Riecker et Ackermann, 2007) pour des informations plus détaillées.

Elaboration du langage et de la parole

Le langage parlé est un processus complexe faisant intervenir un ensemble de composantes, relatives au langage uniquement, plus ou moins imbriquées telles que la sémantique, syntaxe, morphologie, phonologie, ... auxquelles viennent se greffer des composantes liées davantage à la parole telle que la prosodie et finalement les mécanismes moteur impliqués dans la production. Les modèles psycholinguistiques (Ferrand, 2001) s'accordent pour définir trois étapes dans le langage parlé, impliquant ces composantes (excepté la prosodie) : (1) étape de conceptualisation du message - définition des concepts et idées qui doivent être exprimés verbalement -, (2) étape de lexicalisation du message - sélection des mots appropriés et définition de la structure sémantique, syntaxique, morphologique et phonologique adaptée - et (3) étape d'articulation du message. Au delà de ces étapes, de nombreuses questions restent en suspend sur la nature exacte, la chronologie et les interactions des processus cognitifs mis en jeu lors d'un acte de parole, induisant différents modèles proposés et débattus dans la littérature.
Au niveau de l'organisation cérébrale, ce degré élevé de complexité et les questionnements qui en découlent se traduisent par un ensemble d'hypothèses sur le rôle des aires cérébrales suivantes, activées lors d'un acte de parole :

- L'aire de Broca (aires 44 et 45) : au delà des hypothèses initialement émises par Paul Broca sur l'implication de cette aire dans des fonctions purement motrices liées au langage, de récentes études basées sur l'imagerie cérébrale fonctionnelle ont permis de montrer d'autres fonctions dont une participation à la planification du mouvement de parole et à la mémoire verbale. Ces fonctionnalités supplémentaires feraient, par ailleurs, intervenir les deux hémisphères, contrairement aux hypothèses de Paul Broca sur la présence d'un hémisphère dominant. Ainsi, l'aire 44 serait impliquée dans le traitement phonologique et la production de la parole tandis que l'aire 45 participerait au traitement sémantique du langage. Néanmoins, il semblerait que l'activation de l'aire de Broca n'intervienne que lors

de tâches complexes de production de la parole impliquant une construction sémantique ou grammaticale.
- Le cortex auditif (aires 41, 42, 21, 22) : s'il est établi que l'activation bilatérale du cortex auditif primaire (aire 41) lors d'un acte de parole est due au phénomène de feedback auditif ("nous nous entendons parler"), des hypothèses sur l'activation des aires 21 et 22 (aires temporales associatives) ont également été émises, notamment au niveau de leur implication dans le traitement phonologique et sémantique de l'information linguistique du message perçu.
- L'aire de Wernicke (aire 22 majoritairement, et 37, 39, 42 partiellement selon les auteurs) : tout comme l'aire de Broca, les hypothèses émises au XIXe siècle sur cette zone corticale ont évolué récemment. Les études conduites sembleraient montrer, d'une part, une activation bilatérale de l'aire 22, même si une dominance de l'hémisphère gauche semble bien présente pour le langage. La partie supérieure gauche participerait à la construction phonémique lors de production de parole (mais également de perception) alors que la partie droite prendrait part au processus de compréhension.
- L'insula : différentes études ont montré l'implication d'une petite zone de l'insula (hémisphère gauche) dans le langage. Néanmoins, deux hypothèses sont émises : l'une que cette zone serait impliquée dans la planification motrice des mouvements articulatoires de la production de la parole, et plus précisément agirait sur l'ajustement des muscles laryngés et respiratoires avant la production de parole (Riecker et Ackermann, 2007), l'autre qu'elle pourrait plutôt influencer la motivation et la spontanéité verbale au regard d'autres fonctionnalités imputées à l'insula autour des émotions et du désir conscient.
- Les aires 5 et 7 : même si leur implication n'a pas été démontrée dans le cadre de la production de la parole, leurs relations avec le cortex somato-sensoriel (aires 1, 2, 3) d'une part pour l'aire 5, et les aires visuelles d'autre part pour l'aire 7 font d'elles des acteurs potentiels.
- Les aires 39 (gyrus angulaire) et 40 (gyrus supra-marginal) : des hypothèses sont émises sur leurs implications respectives dans le traitement sémantique et dans le traitement phonologique et articulatoire des mots, globalement au niveau de l'hémisphère gauche, mais également de l'hémisphère droit pour l'aire 39.

Préparation du mouvement

La planification et l'organisation du mouvement de la parole semblent suivre le même schéma qu'un mouvement volontaire quelconque. En effet, nous retrouvons ici l'implication de l'aire 6 de Brodmann, composée du cortex pré-moteur latéral et de l'aire motrice supplémentaire (AMS), côté hémisphère gauche. Cette dernière, associée au cervelet, détermine la séquence précise des contractions musculaires à réaliser pour exécuter le mouvement. A cela, il faut ajouter l'implication d'une région supplémentaire, le gyrus cingulaire antérieur, qui semblerait intervenir dans la préparation des mouvements de parole, notamment au niveau de la supervision de l'attention qui leur est accordée.

Exécution du mouvement

Nous avons vu précédemment que le cortex moteur primaire (M1) correspondant à l'aire 4 de Brodmann était au coeur de la motricité, comme a pu le démontrer Penfield au travers de son Homoncule moteur, encore récemment confirmé dans (Grabski et al., 2009) par imagerie cérébrale. Comme le montre la figure 1.3, les zones attribuées à la parole sont très nombreuses, puisqu'elles comprennent les zones oro-faciales ainsi que celle liée au tronc pour les fonctions respiratoires. Nous pouvons également remarquer l'étendue de ces zones, comparées à celles attribuées au reste du corps humain, qui dénote la précision nécessaire pour l'exécution des gestes de parole. De plus, l'activation bilatérale des zones liées aux mouvements de parole a pu être démontrée par imagerie cérébrale. Ainsi, après réception des différentes informations en provenance de l'aire 6, du cortex somato-sensoriel et grâce aux différentes boucles motrices de régulation (abordées dans la section 1.1.3), le cortex moteur primaire peut transmettre aux différents moto-neurones les informations relatives à la contraction des différents muscles impliqués dans le mouvement 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.

1.1.5 Discussion

Nous venons de passer en revue les différents composants du cerveau humain et leur implication dans la production du langage parlé. La complexité du système nerveux est à la hauteur de celle du langage parlé et les scientifiques sont loin d'avoir percé les nombreux processus et mécanismes inhérents. En effet, plusieurs verrous doivent être levés pour y parvenir. Le premier verrou concerne la nature même du langage parlé. Comme mentionné précédemment, des investigations dont l'objectif serait de décrire précisément les processus cérébraux relatifs à la production de la parole uniquement doivent parvenir en premier lieu à dissocier le langage de la parole (en tant que support). En second lieu, ils doivent contrôler les activités sensorielles comme la vision, l'audition, mais également les activités cognitives comme la perception. Ces deux éléments réunis montrent déjà les difficultés liées au choix et à la mise en oeuvre du protocole expérimental à utiliser. Cette difficulté est d'autant plus accrue que suivant la complexité du langage parlé utilisé lors du protocole expérimental, des zones différentes du système nerveux peuvent s'activer. En effet, suivant la tâche à réaliser, les mécanismes langagiers liés à la syntaxe, à la sémantique, à la pragmatique, à la prosodie, voire à l'automatisme vont s'enclencher ou non.

Le deuxième verrou provient des technologies récentes utilisées. Basées sur l'imagerie cérébrale fonctionnelle, on distingue deux grandes techniques : la tomographie par émission de positions (TEP) ou l'imagerie par résonance magnétique fonctionnelle (IRMf). Ces deux techniques sont apparues courant des années 70-80, et sont utilisées pour la recherche biomédicale, uniquement, depuis la fin des années 80. Si elles ont permis de grandes avancées sur la connaissance des mécanismes cérébraux impliqués

dans les processus cognitifs, elles n'en restent pas moins limitées dans leur usage en raison de leur localisation quasiment exclusive en milieu clinique, du caractère invasif de la TEP ou encore la sensibilité aux mouvements (pouvant induire des artefacts) et la présence d'un bruit acoustique non négligeable pour l'IRMf (Ghio et Teston, 2004). Par ailleurs, le sujet est obligatoirement en position couchée, limitant les scénarios d'investigation (interaction avec l'autre dans le langage parlé par exemple).

Le dernier verrou que nous voudrions souligner réside dans l'unicité de l'être humain. Nous sommes tous différents sur le plan physique, et l'anatomie du système nerveux n'échappe pas à cette règle (taille et volume différents, localisation des sillons, profondeur du cortex cérébral, ...). Si des différences anatomiques sont observées d'une manière générale entre les deux hémisphères cérébraux, des différences sont également observables entre les individus, voire entre des classes d'individus. Ces différences sont dues initialement à nos gènes, mais peuvent s'accentuer en fonction de notre environnement, expériences vécues. On parle alors de plasticité cérébrale, ie la capacité du cerveau à se réorganiser (au niveau neuronal par formation ou disparition de synapses) pour intégrer une nouvelle information en phase d'apprentissage et de mémorisation ou pour pallier un déficit fonctionnel en cas de lésion. Les études sur l'organisation fonctionnelle du système nerveux nécessite par conséquent une certaine attention sur le choix et le nombre de sujets étudiés.

Finalement, nous avons pu voir, au travers de ces différentes sections, que de nombreux composants et mécanismes interviennent dans la production de la parole, à divers niveaux et à des fins bien précises. Il devient par conséquent évident que la multitude de lésions neurologiques possibles en termes de localisation dans le système nerveux va induire des troubles divers et variés dans les mécanismes de production de la parole [5], à savoir des troubles potentiellement respiratoires, phonatoires, articulatoires, prosodiques, ... Nous verrons dans le chapitre suivant les liens entre pathologies et lésions du système nerveux.

1.2 Mécanismes phonatoires

La phonation se définit par la production de sons. Elle est présente chez les êtres vivants disposant d'un appareil phonatoire comme l'homme, les mammifères ou encore les oiseaux. Suivant leur forme, ces sons peuvent être des cris, des rires, des pleurs, des vocalisations, du chant, de la parole, etc. Chez l'être humain, la voix est le résultat de la production de ces sons dès lors que les cordes vocales sont impliquées dans le processus de phonation. Tout comme la parole est considérée comme le support du langage, la voix est le support de la parole - sans voix, pas de parole. Par ailleurs, par sa modulation, la voix apporte du relief à la parole et du sens au mot. La voix est également le vecteur de l'identité du locuteur, de son genre, de son état psychique, de son

5. Des troubles du langage peuvent également être observés suivant la localisation de la lésion.

état émotionnel, de son état de santé, de son intention de communiquer, etc. La voix est un facteur essentiel de la communication parlée. Nous allons par conséquent nous intéresser dans les sections suivantes à l'appareil phonatoire de l'être humain et aux mécanismes mis en oeuvre pour produire des sons.

1.2.1 Organes de la phonation

L'appareil respiratoire

Chez l'être humain, l'appareil respiratoire doit répondre à deux fonctions : la respiration et la phonation. Il s'acquitte de ces fonctions grâce à deux composants principaux : le système pulmonaire et le système thoraco-abdominal. Le système pulmonaire comprend, d'une part, les poumons, lieu d'échanges gazeux avec les cellules sanguines et l'extérieur, attachés à la cage thoracique par le biais de la plèvre, et, d'autre part, les voies aériennes pulmonaires ayant pour fonction de conduire l'air de la trachée vers les poumons tout en le conditionnant (mécanismes de défense des poumons par filtrage de particules étrangères). Le système thoraco-abdominal est composé de la cage thoracique (incluant côtes, vertèbres et sternum) - armature entourant les poumons et le coeur -, du diaphragme - muscle séparant la cage thoracique de la cage abdominale - et de l'abdomen. Ce dernier comprend un nombre important de muscles impliqués dans les phases d'inspiration et d'expiration de l'air. Localisés à différents niveaux du corps humain - abdomen, cage thoracique, dos, épaule, cou -, l'innervation de ces muscles fait intervenir 1 nerf crânien (nerf XI au niveau du cou) et une part importante des nerfs spinaux (part des nerfs cervicaux, thoraciques et lombaires). Le tableau 1.2 fournit la liste de ces muscles, leur localisation ainsi que leur fonction dans la respiration. La liste des nerfs impliqués dans leur innervation est également fournie afin de faire le lien avec le système nerveux périphérique.

Un cycle respiratoire normal (automatique) comprend une phase d'inspiration dite active, ie nécessitant un travail musculaire, et une phase d'expiration dite passive. Lors de l'initiation d'un cycle, en phase d'inspiration, la contraction en parallèle des muscles intercostaux externes et du diaphragme avec l'aide d'autres muscles secondaires entraîne respectivement le soulèvement des côtes et du sternum, et l'abaissement du diaphragme, suivi consécutivement de l'abaissement du contenu de la cavité abdominale. Cette configuration entraîne une augmentation de la taille de la cage thoracique et, par conséquent, une augmentation du volume pulmonaire (facilitée par la grande élasticité de chacun des poumons). A cet instant, la pression de l'air à l'intérieur des poumons, pression intra-pulmonaire, diminue et devient alors inférieure à celle de l'air ambiant. Cette différence de pression est immédiatement compensée par une entrée d'air ambiant dans les poumons.
Lorsque les muscles contractés se relâchent, la cage thoracique et les poumons reprennent leur place et forme initiales, initiant la phase (passive) d'expiration. En effet, la pression intra-pulmonaire augmente suite à la diminution du volume pulmonaire et entraîne cette fois-ci l'expulsion de l'air contenu dans les poumons vers l'extérieur afin de réta-

Dénomination	Fonctions respiratoires	Innervation
Muscles du cou		
Sterno-cléido-mastoïdien	inspiration	nerf crânien XI, C2, C3
Scalènes antérieurs, moyens et postérieurs	inspiration	C3 à C8
Muscles de l'épaule		
Grand pectoral	inspiration	C5 à C8, T1
Petit pectoral	inspiration	C5 à C8
Sous-clavier	inspiration	C5, C6
Grand dentelé	inspiration	C4 à C6, T2, T3
Muscles du dos		
Grand dorsal	inspiration ou expiration	T1 à T11
Petit dentelé inférieur	expiration	C6 à C8
Petit dentelé supérieur	inspiration	T9 à T12
Carré des lombes	expiration	T12, L1 à L3
Iliocostal cervical	inspiration	C4 à T3
Iliocostal thoracique	stabilisation	T1 à T11
Iliocostal lombaire	expiration	T7 à L2
Muscles de la cage thoracique		
Intercostaux externes	inspiration	T1 à T11
Intercostaux internes	inspiration ou expiration	T1 à T11
Thoracique transverse	expiration	T2 à T6
Elévateurs des côtes	inspiration	C8 à T11
Muscle du diaphragme	inspiration	C3 à C5
Muscles de la paroi abdominale		
Droit	expiration	T7 à T11
Transverse	expiration	T7 à T12, L1
Oblique externe	expiration	T7 à T12
Oblique interne	expiration	T8 à T10, L1

TABLE 1.2 – *Muscles impliqués dans les mécanismes respiratoires, classés suivant leur localisation, et accompagnés de leur fonction respective et de la liste des nerfs qui les innervent (pour rappel - C : nerfs cervicaux, T : nerfs thoraciques et L : nerfs lombaires).*

blir l'équilibre entre pression de l'air ambiant et pression intra-pulmonaire.
Il est à noter que si l'expiration est considérée comme passive dans un cycle respiratoire normal, elle ne l'est plus en cas d'expiration forcée ie due à un acte volontaire. Dans ce cas, elle peut faire intervenir différents muscles (tableau 1.2) permettant notamment d'agir sur la taille de la cage thoracique et, par conséquent, sur le volume pulmonaire. De manière similaire, l'inspiration forcée va avoir une incidence sur le type de muscles impliqués et leur force de contraction.

Le cycle respiratoire impliqué dans la phonation diffère sensiblement du cycle respiratoire normal. Ici, l'expiration devient une phase active puisqu'elle est à l'origine du flux d'air sortant (des poumons vers l'extérieur) nécessaire à toute production sonore. Par ailleurs, cette expiration doit être adaptée à la production sonore ciblée et, par conséquent, contrôlée pour fournir le débit d'air expiré attendu. Ainsi, le cycle respiratoire se résume en une inspiration plus courte (de l'ordre de 0.5s), suivi d'une expiration plus lente (de l'ordre de 3 à 4s) (Solomon, 2007) pour une production de parole normale. Dans ce contexte particulier, on observe également une configuration atypique du système thoraco-abdominal pendant le cycle de respiration, montrant une contraction de

l'abdomen et un soulèvement de la cage thoracique en début de production, suivi d'un rétrécissement de chacun au cours de cette dernière, sous l'action de différents muscles liés à l'expiration. Pour finir, la phonation nécessite un volume pulmonaire plus important de l'ordre de 25% de la capacité totale contre 10 à 15% dans le cadre d'une respiration normale (Auzou et al., 2001). Le lecteur pourra se reporter à l'ouvrage suivant (Hixon et Hoit, 2005) pour une description plus détaillée du fonctionnement du système respiratoire dans la production de la parole.

Le larynx

Le larynx est situé au carrefour des voies aériennes et digestives, et plus précisément, en amont de la trachée, en aval du pharynx et en avant de l'oesophage. C'est un conduit ostéo-cartilagineux composé (de haut en bas) de l'os hyoïde et de différents cartilages (comme illustré sur la figure 1.7) : le cartilage thyroïde (formant la pomme d'Adam visible chez l'homme), le cartilage épiglottique, les deux cartilages aryténoïdiens et le cartilage cricoïde. De plus, le larynx abrite les cordes vocales, également appelées plis vocaux. La rigidité du larynx dans le cou est maintenue grâce aux cartilages thyroïde et cricoïde, les cartilages épiglottique (appelés plus communément épiglotte) et aryténoïdiens étant, pour leur part, mobiles. Ces différents éléments sont reliés par des muscles et des ligaments et recouverts de muqueuse, le tout contribuant à assurer le rôle sphinctérien (ouverture/fermeture) du larynx dans la respiration (ouverture du larynx pour laisser passer l'air entrant ou sortant), la déglutition (fermeture du larynx par basculement du cartilage épiglottique et fermeture des cordes vocales afin d'éviter les fausses routes ie l'introduction d'aliments dans les voies aériennes) et la phonation que nous décrirons plus bas.
Les muscles du larynx sont généralement dissociés en deux groupes :

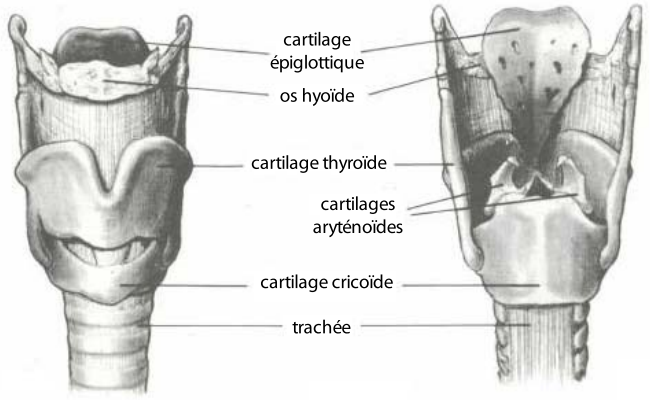


FIGURE 1.7 – *Illustration des différents composants du larynx, face antérieure (à gauche), face postérieure (à droite).*

– les muscles intrinsèques qui contrôlent le mouvement des cordes vocales : abduction (écartement/ouverture), adduction (rapprochement/fermeture) mais également le réglage de leur tension. Suivant leur fonction respective, on distinguera l'unique muscle abducteur - le crico-aryténoïde postérieur (PCA), 3 muscles adducteurs - (1) thyro-aryténoïdien (TA), (2) le crico-aryténoïdien latéral (LCA), (3) l'inter-aryténoïdien (IA) et un muscle "régulateur" - le crico-thyroïdien (CT)-, qui allonge et peut mettre en tension les cordes vocales en association avec le muscle TA. Si l'innervation de tous ces muscles a pour origine le nerf crânien X (nerf vague), on distinguera deux de ses branches, le nerf laryngé supérieur qui innerve le muscle crico-thyroïdien et le nerf laryngé inférieur (ou récurrent) pour les autres.

– les muscles extrinsèques qui contrôlent la position du larynx dans le cou, favorisant ainsi l'action des muscles intrinsèques. Il est à noter que l'action de ces muscles peut également avoir une influence sur la longueur des cordes vocales (Simonyan et Ludlow, 2007). Ces muscles se déclinent en deux sous-groupes : les muscles sous-hyoïdiens, innervés par le nerf crânien XII (nerf hypoglosse) et les muscles sus-hyoïdiens, dont l'innervation dépend de branches issues soit du nerf V (nerf trijumeau), soit du nerf VII (nerf facial), soit du nerf XII.

L'intérieur du larynx est tapissé d'une muqueuse formant deux paires de replis. La première paire, dite supérieure, constitue les bandes ventriculaires, également appelées "fausses cordes vocales", à différencier des "vraies" cordes vocales, relatives à la deuxième paire de replis, dite inférieure. S'il est indéniable que les cordes vocales sont partie intégrante de l'appareil de phonation, le rôle des bandes ventriculaires dans la production de certaines formes phonatoires autres que la parole usuelle (chuchotement, coup de glotte, etc.) reste à l'étude étant donnée leur implication observée dans certains types de chants diphoniques ou dans des cas de forçage vocal. Nous nous intéresserons ici uniquement à la structure et rôle des cordes vocales et invitons le lecteur à lire (Bailly, 2009) pour un état de l'art des travaux réalisés sur les bandes ventriculaires.
Les cordes vocales sont donc de fines bandes musculo-membraneuses, positionnées de part et d'autre du larynx. Elles sont attachées, conjointement, à l'avant au cartilage thyroïde et séparément, à l'arrière à la base des deux cartilages arythénoïdes. Elles ont une structure hétérogène en couches, comme le montre la figure 1.8 comprenant :

– dans la zone la plus profonde, le muscle thyro-aryténoïdien inférieur, également appelé muscle vocal ou vocalis ;
– dans la zone transitoire, la lamina propria, tissu composé de fibres élastiques et de collagène, également organisée en couches, formant en superficie l'espace de Reinke et dans sa couche intermédiaire et profonde le ligament thyro-aryténoïdien inférieur, également appelé le ligament vocal ;
– dans la zone superficielle, l'épithélium, formant avec la couche superficielle de la lamina propria, une muqueuse recouvrant le bord libre des cordes vocales.

La taille des cordes vocales est variable suivant les individus, et notamment leur genre. On estime à 22mm la taille des cordes vocales chez les hommes et entre 18 et 20mm chez les femmes. Elles délimitent une cavité appelée glotte, donnant son nom à la "fente

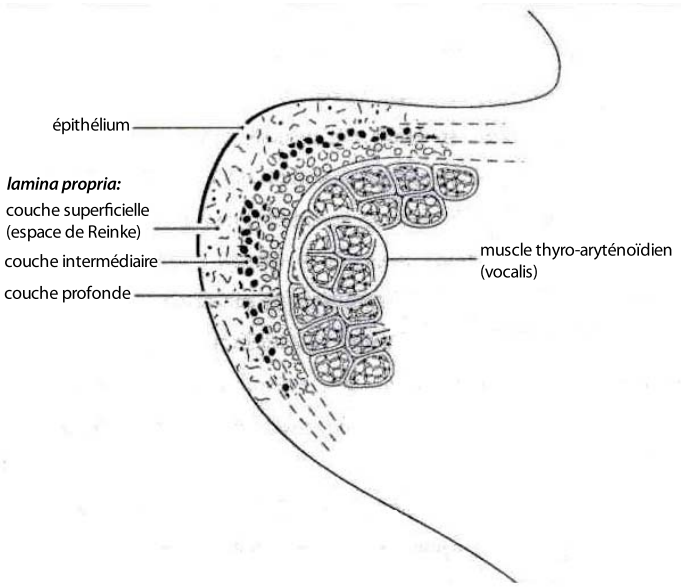


FIGURE 1.8 – *Illustration des différentes couches composant une corde vocale.*

glottique" - espace entre les cordes vocales. La glotte est considérée d'un point de vue anatomique comme un plan médian (plan glottique), séparant les deux autres parties du larynx : l'étage sus-glottique au dessus des cordes vocales (partie supérieure) qui s'étend jusqu'à l'épiglotte et l'étage sous-glottique au dessous des cordes vocales (partie inférieure) qui s'étend jusqu'au cartilage cricoïde.
Les cordes vocales peuvent prendre trois positions principales :

- en position écartée, la glotte est ouverte, l'air circule librement permettant la mise en place des phénomènes de respiration ;
- en position accolée, la glotte est fermée, l'air ne passe plus. On parle alors de phénomène d'apnée ;
- en position rapprochée, l'ouverture de la glotte est variable, l'air circule de manière saccadée permettant la mise en place des phénomènes de phonation.

Les cavités supraglottiques

Au nombre de trois, les cavités supraglottiques principales - cavités pharyngale, orale et nasale - forment ce qui est appelé le tractus ou conduit vocal (cf. figure 1.9). Ce dernier s'étend de la glotte (larynx) aux lèvres, et atteint une longueur moyenne d'environ 18cm chez l'homme.

Cavité pharyngale

Supportée par le pharynx, carrefour des voies respiratoires et digestives, la **cavité pharyngale** est un conduit musculo-membraneux qui s'étend des cavités orale et nasale au larynx et à l'oesophage. Assimilable à une demi-gouttière, le pharynx, élargi au niveau supérieur (de 4 à 5 cm au niveau des cavités orale et nasale), se rétrécit finalement au

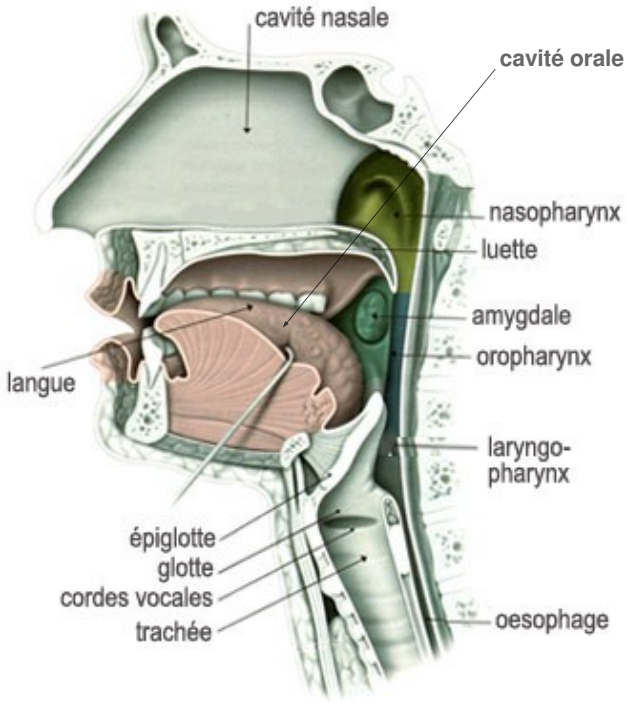


FIGURE *1.9* – *Illustration des cavités pharyngale (comprenant les 3 étages du pharynx), orale et nasale (source : http ://lecerveau.mcgill.ca).*

niveau de l'oesophage (environ 2cm). Il peut être divisé, de haut en bas, en 3 sous-composants : le naso-pharynx localisé entre le voile du palais et les fosses nasales (cavité nasale), l'oro-pharynx, localisé entre le palais mou (cavité orale) et l'os hyoïde et le laryngo-pharynx localisé entre l'arrière de l'épiglotte et l'oesophage (cf. figure 1.9). Le pharynx est muni de différents muscles jouant un rôle dans la déglutition, la respiration et la phonation. Parmi eux, nous pouvons relever :

- les muscles constricteurs supérieur, moyen et inférieur qui agissent sur l'avancement de la paroi postérieure du pharynx et le rapprochement des parois latérales, modifiant ainsi le diamètre du pharynx ;
- les muscles élévateurs - le stylo-pharyngien, le palato-pharyngien et le salpingo-pharyngien - permettent le déplacement vertical du pharynx (et également du larynx), modifiant ainsi la longueur du conduit vocal ;
- les muscles du naso-pharynx agissant sur le voile du palais : les muscles péristaphylin interne et palato-staphylin, élévateurs du voile, les muscles pharyngo-staphylin et palatoglosse, abaisseurs du voile et le muscle péristaphylin externe, tenseur du voile.

L'innervation sensitivo-motrice de ces muscles est assurée par les nerfs crâniens n.V, n.IX, n.X et n.XI.

Cavité orale
Elle est délimitée à l'avant par les lèvres, en latéral par les joues, sur la partie inférieure par le plancher oral incluant la langue, sur la partie supérieure par le palais et à l'arrière par la cavité pharyngale. Premier élément de l'appareil digestive, elle y joue différents rôles : insalivation, mastication, gustation et déglutition des aliments.

Les **lèvres** sont des replis musculo-membraneux délimitant la fente orale et réunies à leurs extrémités par les commissures. Elles sont composées de couches, comprenant de l'extérieur vers l'intérieur la peau, des muscles, des glandes (notamment salivaires) et une muqueuse. Sur le plan musculaire, on distingue deux grands types de muscles :

- les muscles constricteurs incluant le muscle orbiculaire des lèvres responsable de la fermeture de la bouche, le pincement des lèvres et la projection des lèvres en avant (protrusion des lèvres) et le muscle compresseur des lèvres, responsable de la compression des lèvres d'avant en arrière et principalement actif dans le phénomène de succion (en particulier, chez le nourrisson) ;
- les muscles dilatateurs, responsables du mouvement des commissures des lèvres (muscles canin, buccinateur, grand zygomatique, triangulaire des lèvres, peaucier du cou, risorius), du mouvement de la lèvre inférieure (muscles houppe du menton, carré du menton, triangulaire des lèvres, peaucier du cou) et du mouvement de la lèvre supérieure (releveur de l'aile du nez et de la lèvre supérieure, petit zygomatique).

L'innervation sensitivo-motrice de ces muscles est assurée par les nerfs crâniens n.V (innervation sensitive) et n.VII (innervation motrice).

Les **joues** sont composées, à l'extérieur, de la peau richement vascularisée (participation à l'expression du visage et des sentiments par le changement de sa coloration) et à l'intérieur d'un plan musculaire formé de muscles superficiels (muscles peauciers) et d'un muscle profond, le muscle buccinateur. Ce dernier est à l'origine de la tonicité des joues. Il intervient également dans le mouvement des commissures des lèvres (voir ci-dessus) et joue un rôle important dans la mastication des aliments, la réalisation du sifflement et du soufflement. Tout comme pour ceux des lèvres, l'innervation sensitivo-motrice des muscles des joues est assurée par les nerfs crâniens n.V (innervation sensitive) et VII (innervation motrice). Finalement, il est à noter que le mouvement de protrusion des lèvres permet de former une sous-cavité - cavité labiale - constituant ainsi l'extrémité antérieure du conduit vocal.

Le **plancher oral** est délimité à l'avant par la mandibule (mâchoire inférieure), à l'arrière par l'os hyoïde, tous deux en forme de fers à cheval. Il accueille en son centre la langue à laquelle il est rattaché par le frein lingual. Il comprend un muscle principal, le mylo-hyoïdien, dont la forme plate et triangulaire constitue la majeure partie du plancher et deux muscles plus secondaires, le genio-hyoïdien et le ventre antérieur du muscle gastrique. Ce plan musculaire est chargé des mouvements ascendants de l'os hyoïde et de la langue dès lors que la mandibule est fixe et des mouvements descen-

dants de la mandibule dès lors que l'os hyoïde est fixe. Le muscle mylo-hyoïdien et le ventre antérieur du muscle gastrique sont innervés par le nerf crânien n.V, le muscle genio-hyoïdien par le nerf n.XII.

La **langue** est un organe composé d'un squelette ostéo-fibreux et d'un plan musculaire complexe, le tout enveloppé par une fine muqueuse, porteuse de papilles gustatives. Le squelette ostéo-fibreux comprend l'os hyoïde auquel la langue est rattachée à la base et deux membranes fibreuses que sont le septum lingual (séparant la langue en deux parties sur sa longueur) et la membrane hyo-glossienne. Un premier sillon dénommé V lingual (en raison de sa forme en V), présent sur toute la largeur de la langue, sépare cette dernière en deux sections distinctes : une partie fixe, la racine, rattachée à l'oro-pharynx, et une partie mobile, le corps, rattachée à la cavité orale. Le corps s'étend du sillon V lingual jusqu'à la pointe de la langue, nommée apex, et est parcouru par un sillon médian.
Le plan musculaire rattaché à la langue est très complexe. Il participe activement à la grande mobilité de la langue au sein de la cavité orale et à sa capacité à faire varier la configuration de cette dernière. Ces caractéristiques lui confèrent un rôle important dans la mastication, la déglutition, la production de la parole mais également la gustation grâce à la présence des papilles gustatives mentionnées plus haut. Le plan musculaire comprend 17 muscles (8 muscles pairs et 1 muscle impair) qui se caractérisent soit par leur caractère intrinsèque - localisés uniquement dans la langue - soit par leur caractère extrinsèque - leur origine s'effectue hors de la langue, sur laquelle ils se rattachent. Ils peuvent être catégorisés suivant leur action sur la mobilité de la langue. On distingue :

- les muscles permettant d'abaisser et de rétracter la langue (mouvement de recul) : génio-glosse (pair, extrinsèque - langue plaquée contre le plancher oral, apex rétractée vers le bas), hyo-glosse (pair, extrinsèque), longitudinal inférieur (pair, intrinsèque - apex rétractée vers le bas), longitudinal supérieur (impair, intrinsèque - raccourcissement de la langue) ;
- les muscles permettant d'élargir la langue et de la "tracter" en arrière vers le haut : stylo-glosse (pair, extrinsèque), palato-glosse (pair, extrinsèque - élévation du corps de la langue), amygdalo-glosse (pair, extrinsèque - élévation de la base de la langue), pharyngo-glosse (pair, extrinsèque - traction arrière de la base de la langue) ;
- les muscles permettant de jouer sur la forme de la langue : muscles transverse (pair, intrinsèque - réduction de la largeur de la langue) et vertical (pair, intrinsèque - aplatissement de la langue).

L'innervation de la langue est à la fois motrice, sensitive et sensorielle (gustation). Ainsi, le nerf crânien n.XII intervient dans l'innervation motrice d'une grande majorité des muscles de la langue. Seuls deux muscles peuvent être traités séparément, le muscle stylo-glosse, qui serait innervé par le n.VII et le muscle palato-glosse innervé par le n.X. Côté sensitif, le nerf n.V interviendrait sur l'avant du V lingual et le n. IX sur l'arrière. Côté sensoriel, le nerf n.VII interviendrait sur l'avant du V lingual, le nerf n.IX sur l'arrière.

Le **palais** est la cloison qui sépare la cavité orale (bas) de la cavité nasale (haut). Il est composé sur les 2/3 postérieurs du palais dur (également appelé voûte palatine [6] ou palais osseux) et sur le 1/3 antérieur du palais mou ou voile du palais. De 4 à 5 cm de large et 7 à 8 cm de long, le palais dur est une voûte osseuse, formée par la réunion des lames horizontales des deux os maxillaires supérieurs d'une part, et la réunion des lames horizontales des deux os palatins d'autre part. Cette voûte est recouverte d'une muqueuse comportant des glandes salivaires. Dans le prolongement postérieur du palais dur, est situé le palais mou, partie flexible du palais qui se termine au centre par la luette (ou uvule). Il a un rôle essentiel dans la déglutition grâce à sa position relevée permettant d'isoler la cavité orale de la cavité nasale et ainsi d'empêcher les aliments de refluer vers cette dernière. Le palais mou est constitué d'une charpente fibreuse (aponévrose palatine) sur laquelle s'insèrent les muscles pairs chargés de sa mobilité et d'une muqueuse recouvrant les deux faces du voile pourvue de glandes palatines et de récepteurs du goût. Les muscles, au nombre de cinq, peuvent être classés en fonction de leur rôle dans le mouvement du voile. On distingue ainsi :

- les muscles élévateurs : péristaphylin interne et palato-staphylin (appelé également azygos de la luette ou muscle uvulaire) ;
- les muscles abaisseurs : palato-pharyngien et palato-glosse ;
- le muscle tenseur : péristaphylin externe.

Hormis le muscle tenseur dont l'innervation motrice est assurée par le nerf mandibulaire V3 (branche du nerf crânien n.V), les muscles du palais mou sont pris en charge par le nerf crânien n.X. Côté sensitif, l'innervation est assurée par le nerf maxillaire V2 (branche du nerf crânien n.V).

Cavité nasale

Le rôle premier de la cavité nasale est de réchauffer et d'assainir l'air inspiré et de fournir les éléments organiques impliqués dans l'olfaction. Elle est constituée d'une charpente ostéo-cartilagineuse formée de nombreux os du crâne et de la face (os maxillaire supérieur, éthmoïde, sphénoïde, palatin, unguis, cornet inférieur, propre du nez et vomer), d'une muqueuse tapissant cette charpente et d'organes en lien avec l'olfaction. Elle regroupe les fosses nasales, deux cavités situées symétriquement de part et d'autre de la cloison nasale (septum nasal), en avant du rhino-pharynx, au dessus de la cavité orale et en dedans des os maxillaires supérieur et inférieur. La partie antérieure des fosses nasales communiquent avec l'extérieur via les narines, la partie postérieure avec le rhino-pharynx via les narines internes appelées choanes. Les fosses nasales communiquent également avec les sinus paranasaux, cavités creusées dans les os avoisinants (sinus maxillaires, frontaux, éthmoïdaux, sphénoïdaux). L'innervation de la cavité nasale est pour l'essentiel assurée par le nerf crânien n.VII.

6. suivant certains auteurs, la voûte palatine est assimilée au palais dans sa globalité.

1.2.2 Production sonore

Lorsqu'un son est émis, les trois grandes classes d'organes que nous venons de décrire, à savoir l'appareil respiratoire, le larynx et les cavités supra-glottiques vont intervenir successivement pour produire une onde sonore. Les qualificatifs de soufflerie pulmonaire, vibrateur et résonateurs sont donnés respectivement à chacun d'eux dans la littérature.

L'énergie aérodynamique nécessaire à l'établissement de cette onde est fournie par l'appareil respiratoire, par expiration d'un flux d'air stocké dans les poumons. Différemment de celle impliquée dans la respiration (expiration passive), cette expiration doit être contrôlée de manière très précise par l'être humain afin d'assurer une pression pulmonaire suffisante à la production sonore. A l'origine de la phonation, les cordes vocales sont accolées (phase de fermeture), empêchant l'écoulement du flux d'air de l'expiration. La pression de l'air sous-glottique augmente sous les cordes vocales jusqu'à atteindre une force suffisamment importante (appelée pression de seuil de phonation) pour pousser les cordes vocales et les éloigner l'une de l'autre (phase d'ouverture). Avec l'ouverture des cordes vocales, une bouffée d'air est libérée, provoquant une baisse de la pression sous-glottique. Sous cet effet, les cordes vocales peuvent reprendre leur position initiale (fermeture), relançant le processus d'augmentation de la pression sous-glottique et leur ré-ouverture. La répétition de ces cycles d'ouverture et de fermeture rentre dans le processus de vibration des cordes vocales, permettant de transformer le flux d'air expulsé en onde acoustique. A cette étape, l'onde acoustique est perçue comme un simple bourdonnement. Seule la propagation de l'onde dans les cavités supra-glottiques (résonateurs ou caisses de résonance) va permettre de modifier ses propriétés acoustiques, conduisant à l'élaboration et la distinction des différents sons voisés [7] présents dans la parole.

La vibration vocale

Le fonctionnement vibratoire des cordes vocales a suscité de nombreux travaux de recherche, dédiés à la compréhension des phénomènes d'ouverture/fermeture des cordes vocales, de maintien de la vibration ou de l'auto-entretien de l'oscillation, de synchronisation entre cordes, ... Il en a résulté, depuis le XVIIIe siècle, différentes théories de la phonation (voir (Crevier-Buchman, 2007), (Le Huche et Allali, 2010a) pour une description comparative et détaillée de ces théories), dont les principales sont les suivantes :

– la théorie myo-élastique proposée par Ewald (1898) - dans laquelle seul le flux d'air expiré couplé à l'élasticité des cordes vocales contribue à leur mise en vibration. L'air écarte les cordes vocales tandis que l'élasticité les ramène en position fermée. Cette théorie a été complétée en 1958 par Van Den Berg (théorie aérodynamique-myoélastique complétée) ajoutant la contribution de l'effet de

7. par opposition aux sons non voisés pour lesquels le processus de vibration des cordes vocales n'est pas activé.

Bernoulli [8] pour expliquer la fermeture des cordes vocales (par rétro-aspiration ou de succion) ; En effet, l'accélération d'un flux d'air traversant la glotte (accélération due au rétrécissement du canal à ce niveau) entraîne une baisse de la pression de l'air (effet Bernoulli). Cette baisse de pression provoque l'accolement des cordes vocales (phénomène de rétro-aspiration ou de succion), et par conséquent la fermeture de la glotte et l'arrêt du flux d'air. La pression sous-glottique augmentant, les cordes vocales s'écartent pour laisser passer un nouveau flux d'air, permettant au cycle fermeture/ouverture de se reproduire ;

- la théorie neuro-chronaxique de Husson (1962), largement révoquée, dans laquelle la contraction des cordes vocales est due uniquement à des influx nerveux (nerf récurrent) réfutant même la présence de phénomènes vibratoires au niveau des cordes vocales ;
- la théorie impulsionnelle de Cornut et Lafon (1962), qui en appuyant la théorie de Van Den Berg, apporte la notion d'impulsion laryngée, assimilant le larynx à un oscillateur à relaxation produisant des impulsions rythmées ;
- la théorie neuro-oscillatoire de Mac-Leod et Sylvestre (1968) qui s'inspirent de l'observation physiologique des ailes des insectes et notamment du muscle associé dit "asynchrone". La particularité de ce muscle est d'entrer en vibration, après une simple impulsion nerveuse, par le biais de contractions répétées. Ce mode vibratoire permet ainsi d'augmenter considérablement la fréquence de battements des ailes de l'insecte (indépendamment du rythme des impulsions nerveuses) et dépend, au niveau de sa fréquence, de la masse et de l'élasticité des structures en mouvement. Cette théorie repose par conséquent sur l'assimilation des muscles du larynx à un muscle asynchrone permettant le maintien en vibration des cordes vocales ;
- la théorie oscillo-impédancielle de Dejonckère (1981), qui, s'appuyant sur la théorie de Van Dan Berg et celle de Cornut et Lafon, assimile plutôt le larynx à un oscillateur à faible amortissement.

Nous savons aujourd'hui que ces théories, bien qu'ayant inspiré grand nombre de chercheurs dans le domaine, ne permettent pas de prendre en compte toute la complexité du fonctionnement vibratoire du larynx. En effet, les travaux de recherche menés depuis les années 70 ont permis de mettre en évidence l'importance de différents facteurs dans la modélisation de ce fonctionnement, comme par exemple :

- la structure en couches hétérogènes des cordes vocales (décrite en sous-section 1.2.1), mélangeant fibres musculaires et muqueuse, joue un rôle important dans le phénomène vibratoire. Dans la théorie du "body-cover" d'Hirano (Hirano, 1974), il est proposé de différencier le corps musculo-ligamentaire des cordes vocales ("body"), comprenant le muscle vocal et le ligament vocal, de la couverture muqueuse ("cover"), composée de l'épithélium et de la couche superficielle de la lamina propria. En effet, la composition différente de ces deux couches couplée au fait qu'elles soient au final peu connectées, notamment par la présence de l'espace de Reinke, leur permettent de se mouvoir de manière indépendante (on parle de

8. Le principe de Bernoulli, du nom de son auteur, met en relation la vitesse d'écoulement d'un fluide dans un canal et sa pression : plus la vitesse augmente, plus la pression diminue.

glissement de la couverture vis à vis du corps) et leur confèrent, par conséquent, un rôle distinct à jouer dans le mécanisme vibratoire ;
- le facteur myo-élastique, mis en avant par la théorie de Ewald, est un élément important dans les mécanismes vibratoires des cordes vocales. D'après (Titze, 1994), il tient compte de la masse des cordes vocales, de leur élasticité (pour le retour à la position d'équilibre), de leur longueur, de leur tension déterminant la rapidité de retour à l'état antérieur, de leur raideur déterminant leur résistance au déplacement ainsi que de leur inertie pour le maintien dans un état donné. Par ailleurs, les muscles thyro-aryténoïdien, crico-thyroïdien et crico-aryténoïdien postérieur jouent également un rôle important dans ces mécanismes (Hirano, 1988; Titze, 1994) ;
- le mécanisme vibratoire ne se réduit pas à un simple mouvement horizontal d'ouverture et de fermeture des cordes vocales. Il serait la combinaison de mouvements indépendants produits par le corps et la couverture des cordes vocales. L'observation des cordes vocales par stroboscopie ou cinématographie ultra-rapide a, en effet, permis de montrer une ondulation de bas en haut de la couverture (mucosal wave), qui est due aux mouvements asynchrones entre les parties inférieure et supérieure des cordes vocales. Comme le montre la figure 1.10, suite à l'accolement des cordes vocales (illustration a), on observe un mouvement latéral d'ouverture, démarrant de la partie inférieure de la couverture (illustration b) qui se propage vers sa partie supérieure (illustration c) jusqu'à écartement complet (illustration d). Dès lors, la partie inférieure amorce un mouvement latéral (inverse) de fermeture (illustration e) qui se propage à nouveau vers le haut de la couverture (illustration f) jusqu'à l'accolement des cordes vocales (illustration g). Un nouveau cycle peut alors démarrer (illustration h). Ce déphasage entre partie inférieure et supérieure conduit à deux formes glottiques distinctes dites divergente (partie inférieure fermée, supérieure ouverte) et convergente (partie inférieure ouverte, supérieure fermée), caractérisées par des pressions intra-glottiques différentes. Cette différence de pression, sous l'effet de Bernoulli, va engendrer une alternance de formes divergentes et convergentes et pourra ainsi fournir une oscillation auto-entretenue. Plusieurs modèles ont été proposés dans la littérature pour représenter cette superposition de mouvements entre corps et couverture des cordes vocales dont le plus abouti reste encore le modèle à trois masses de (Story et Titze, 1995) (également appelé "body-cover model") issu de la théorie oscillo-impédancielle.

Malgré ces avancées, les cordes vocales et leur mode vibratoire sont loin d'avoir livré tous leurs secrets. Comme le souligne Lise Crevier Buchman, "*[...] la vibration des cordes vocales nécessite une combinaison complexe de facteurs aérodynamiques, biomécaniques, musculaires et élastiques à l'origine d'ajustements non parfaitement connus [...]*. (Crevier-Buchman, 2007)". Les notions de seuil phonatoire (seuil de pression à partir duquel les phénomènes cycliques de la vibration se produisent), de différences physiologiques entre cordes vocales d'un même individu, de synchronisation des cordes vocales, de force de contact lors de l'accolement des cordes restent encore peu abordées.

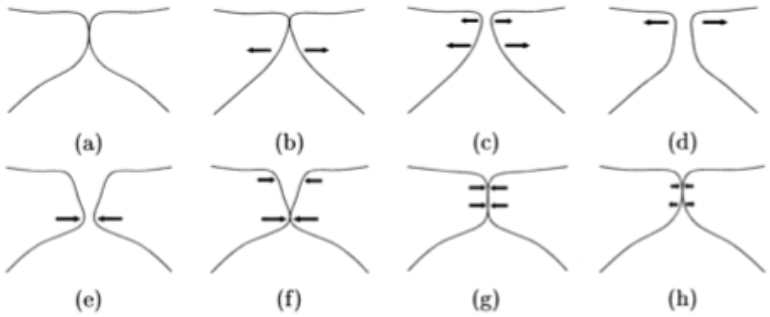


FIGURE 1.10 – *Illustration du phénomène d'ondulation de la couverture due aux mouvements asynchrones des parties inférieure et supérieure des cordes vocales lors d'un cycle d'ouverture et de fermeture. (source : (Story, 2002))*

Quelques caractéristiques de l'onde glottique

L'onde glottique, émise par la vibration des cordes vocales, est, comme tout son, un signal complexe. Ce dernier, suivant la théorie de Fourrier peut être décomposé en un ensemble de signaux sinusoïdaux élémentaires, ayant des amplitudes et des fréquences déterminées. Par application de la transformée de Fourier, l'onde glottique peut donc être représentée sous la forme d'un spectre, fournissant les caractéristiques en amplitude et en fréquence des signaux élémentaires qui la composent. Oscillant en principe de manière cyclique, l'onde glottique est assimilée à un son périodique. La sinusoïde de fréquence la plus basse correspond, par conséquent, à la composante principale et détermine la fréquence fondamentale F0 du signal complexe, les sinusoïdes de fréquences multiples de F0 déterminent ses harmoniques (2 F0, 4 F0, ...), comme illustré schématiquement sur la figure 1.11.

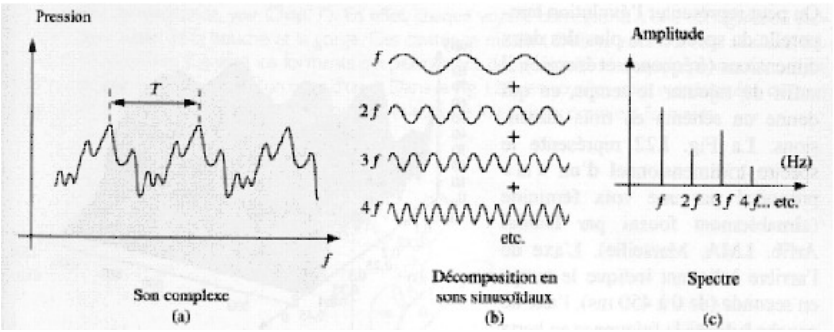


FIGURE 1.11 – *Illustration de la décomposition d'un signal complexe en signaux sinusoïdaux élémentaires et représentation sous la forme d'un spectre avec présence de la fréquence fondamentale f et de ses harmoniques (paires et impaires).*

La fréquence de vibration des cordes vocales ie le nombre de vibrations par seconde détermine la **fréquence fondamentale** de l'onde glottique. En cours d'élocution, cette fréquence fondamentale varie, apportant le caractère mélodique de la parole. Néanmoins, la courbe de variation de la F0 est fréquemment interrompue durant l'acte de

parole en raison des pauses silencieuses ou au cours de la production de sons non voisés. L'ensemble des fréquences produites par un individu est appelé **étendue de la voix**. Il est à noter que la différence de mesures de F0 entre deux cycles successifs est appelée le **jitter** et pourra être utilisée pour caractériser une instabilité dans le cycle d'oscillations.

De nombreux travaux ont porté sur la modélisation de la source glottique et du signal qui en résulte dans le domaine temporel, et dans une moindre mesure dans le domaine spectrale. Le lecteur pourra se référer à (Henrich, 2001) pour une revue de ces travaux.

Mécanismes laryngés
D'après (Henrich, 2001), les mécanismes laryngés, caractéristiques principales de la production vocale humaine, représentent des configurations glottiques particulières identifiables par la forme des cordes vocales (longueur, épaisseur) et la tension musculaire mise en jeu. On distingue quatre mécanismes laryngés :

- Le mécanisme 0 ou "vocal fry", caractérisé par des cordes vocales courtes, très épaisses et peu tendues et une durée d'ouverture glottique très brève. Ce mécanisme se traduit par des fréquences les plus basses ;
- Le mécanisme I, caractérisé par des cordes vocales épaisses vibrant sur toute leur longueur avec une grande amplitude comme illustré sur la figure 1.12. Ce mécanisme se retrouve couramment dans les voix parlées masculines, et plus rarement dans les voix féminines lors de production de parole très grave ;
- le mécanisme II, caractérisé par des cordes fines, vibrant uniquement sur les 2/3 de leur longueur comme illustré sur la figure 1.12. A l'inverse du mécanisme I, celui-ci se retrouve couramment dans les voix parlées féminines, et plus rarement dans les voix masculines lors de production de parole très aiguë ;
- Le mécanisme III ou voix de sifflet, caractérisé par des cordes vocales fines et très tendues, vibrant avec une amplitude réduite. La fermeture des cordes vocales est quasi inexistante, permettant de réaliser des fréquences très aiguës.

Les mécanismes I et II couvrent 90% de la production vocale humaine.

D'un bourdonnement vers un son...

Comme souligné précédemment, l'"écoute" de l'onde acoustique produite par la vibration des cordes vocales se réduit à un bourdonnement. Seul le passage de cette onde dans le conduit vocal, composé des trois cavités supra-glottiques principales [9] - cavités pharyngale, orale et nasale décrites dans la section 1.2.1 - lui confère le terme de "son". Lors de ce passage, le conduit vocal va se comporter comme une caisse de résonance, en renforçant ou diminuant l'amplitude de certaines oscillations de l'onde glottique dont

9. Dans la littérature, la cavité orale peut être sous-divisée en cavité orale, entre la langue et le palais, et la cavité buccale, entre les joues et les lèvres. On peut également considérer une dernière cavité, la cavité labiale, lors de la protrusion des lèvres.

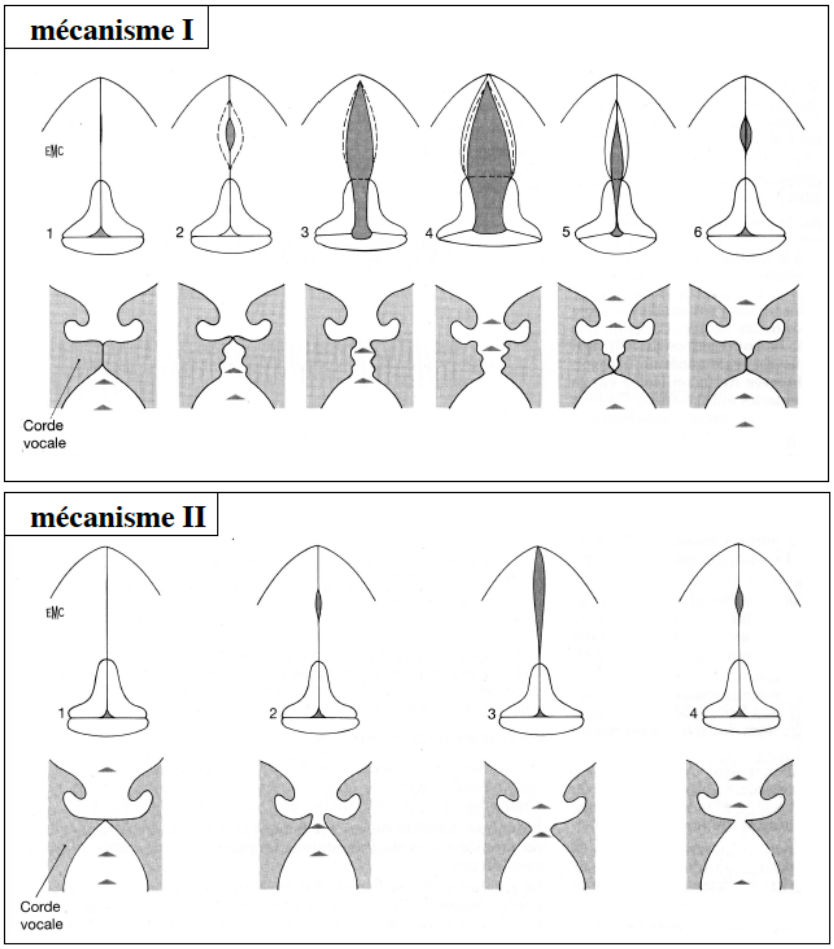


FIGURE 1.12 – *Illustration de la forme (premier jeu d'images) et des phénomènes vibratoires des cordes vocales (deuxième jeu d'images) dans la réalisation des mécanismes laryngés I et II (source : (Henrich, 2001).*

la fréquence est proche de la fréquence de résonance du conduit. Ce phénomène va se traduire par une amplification de certaines harmoniques, apportant des modifications notables dans la composition spectrale de l'onde glottique. La modélisation de ce phénomène par (Fant, 1960) sous la forme d'un modèle linéaire source-filtre dans lequel l'onde glottique est la source, le conduit vocal est le filtre est toujours d'actualité.
Les zones fréquentielles de l'onde glottique accentuées par le conduit vocal s'appelle les **formants**. Le premier harmonique présentant la plus forte amplification (maxima d'énergie) sera assimilé au premier formant et noté F1, le second harmonique au deuxième formant et noté F2, etc. Généralement, les formants F1, F2 et F3 sont bien marqués sur un spectrogramme [10] ; les formants F4 et F5 sont plus difficilement visibles. Un exemple de spectrogramme est fourni sur la figure 1.13. Caractéristiques de certaines classes

10. Représentation tridimensionnelle d'une analyse temps-fréquence d'un signal. Le temps est représenté sur l'axe des x, la fréquence sur l'axe des y et le niveau d'énergie de chaque fréquence sur l'axe des z, symbolisé par un niveau de gris. Plus le point est foncé, plus il représente un fort niveau d'énergie.

de phonèmes, les formants permettent principalement de distinguer les voyelles entre elles. Ainsi, la forme, la dimension ainsi que la constitution physiologique des résonateurs seront autant d'éléments qui détermineront l'ensemble des harmoniques mises en valeur. Il est d'ailleurs important de noter que plus la taille du résonateur est grande, plus sa fréquence de résonance est basse, et par conséquence, plus la valeur du formant sera basse (voir la section 1.3 pour plus de précisions).

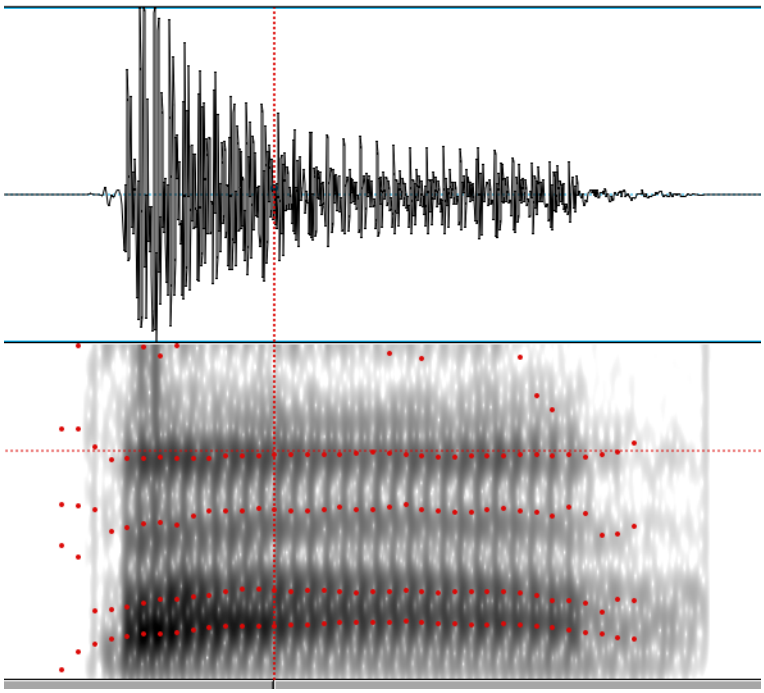


FIGURE *1.13* – *Exemple de signal temporel accompagné de son spectrogramme pour la production sonore /ka/. Les quatre premiers formants F1 (bas) à F4 (haut), de la voyelle /a/ détectés automatiquement par le logiciel Praat (Boersma et Weenink), sont marqués en pointillé.*

Paramètres acoustiques fondamentaux de la voix et de sa qualité

La voix se résume à l'ensemble des sons produits par l'être humain lors de la vibration des cordes vocales. La qualité d'une voix est communément caractérisée sur le plan acoustique par trois paramètres que sont la hauteur, l'intensité et le timbre.

La **hauteur** de voix des locuteurs est fixée par la **fréquence fondamentale** (F0) de l'onde glottique et se situe entre 100 à 150Hz pour les hommes, de 200 à 300Hz pour les femmes et de 300 à 450Hz pour les enfants. Les fréquences les plus basses (vibration lente) correspondent à des voix graves, les plus hautes (vibration rapide) à des voix aiguës. La hauteur de la voix dépend de la longueur des cordes vocales (de 5 à 12mm chez l'enfant, de 14 à 18mm chez la femme et de 18 à 25mm chez l'homme (Le Huche et Allali, 2010a)), de leur masse et de leur tension.

L'**intensité** de la voix traduit, quant à elle, l'amplitude des vibrations des cordes

vocales et est principalement contrôlée par la variation de la pression sous-glottique. En principe, plus la pression sous-glottique sera grande, plus l'intensité vocale le sera également. Du point de vue perceptif, plus l'intensité est importante, plus la voix sera considérée comme forte. Mesurée en décibel (dB), on considère à 10dB l'intensité d'une voix chuchotée, de 35 à 40dB celle d'une voix calme, de 50 à 70dB celle d'une conversation normale, de 80 à 90dB pour un cri. Il est à noter que la différence de mesures d'amplitude entre deux cycles successifs (à fréquence constante) est appelée le **shimmer**.

Finalement, le **timbre** de la voix est peut-être l'élément essentiel dans le principe d'individualité chez l'être humain, puisqu'il nous permet d'identifier une personne à l'écoute de sa voix. Il joue également un grand rôle dans les notions d'esthétisme que l'on peut attribuer à une voix. Le timbre est directement lié au nombre et à l'intensité des harmoniques présents dans l'onde acoustique, ce qui en fait sa richesse. Il dépend par conséquent à la fois des propriétés physiologiques (masse et longueur) et biomécaniques des cordes vocales, notamment au niveau de leur accolement, et des caractéristiques anatomiques des cavités de résonance et dynamiques (mobilité du voile du palais et de la langue, tonicité des lèvres et des joues). Tout comme la fréquence fondamentale, le timbre va varier dans le temps en fonction des changements opérés au niveau des cordes vocales et/ou du volume des cavités de résonance en fonction de leurs configurations (voir section 1.3). On parle communément de la richesse du timbre et de sa couleur.

1.3 Mécanismes articulatoires

Nous venons de voir que l'onde glottique subit des modifications majeures lors de son passage dans le conduit vocal dues aux phénomènes de résonance et que ces phénomènes dépendent de la conformation des cavités supra-glottiques composant ce conduit. Néanmoins, ce fonctionnement ne permet pas d'expliquer toutes les réalisations sonores présentes dans le langage parlé, comme par exemple, les sons non voisés ou sourds (à opposé aux sons voisés ou sonores), réalisés sans excitation de la source glottique (les cordes vocales restent ouvertes tout au long de la production du son), simplement par le passage d'un flux d'air expulsé par l'appareil respiratoire et modulé dans le conduit vocal. Nous allons décrire dans cette section le rôle du conduit vocal et de certains organes dans la production des sons du langage parlé, en se focalisant sur l'ensemble des phonèmes [11] du français.

Tout d'abord, comme nous l'avons déjà souligné le conduit vocal est modulable au sens où sa configuration peut évoluer dans le temps. Ces modifications sont dues principalement à des mécanismes dits articulatoires qui font intervenir tout ou partie

11. Le phonème est la plus petite unité sonore permettant de distinguer, dans un langage parlé, un mot d'un autre : bateau/gâteau. Le français compte 37 phonèmes. Un phonème peut être associé, dans une langue donnée, à plusieurs sons ou allophones. Pour exemple, en français, le phonème /o/ pourra se prononcer dans sa forme ouverte dans le mot *corps* ou fermée dans le mot *seau*.

d'organes du conduit vocal comme le pharynx, le voile du palais, la langue, la mâchoire inférieure et les lèvres. Ces derniers par rapprochement ou contact avec des parties fixes du conduit vont permettre :

- d'augmenter ou de réduire le volume d'une cavité ;
- d'augmenter ou de réduire le diamètre du conduit vocal. On parle alors de constriction du conduit et de son degré ;
- de fermer complètement le conduit vocal. On parle alors d'occlusion du conduit.

Ces multiples configurations du conduit vocal vont conduire à la réalisation d'environ 150 productions sonores potentielles, à la base de l'ensemble des langues du monde (Teston, 2007). Les termes de **mode** et de **lieu d'articulation** sont fréquemment utilisés dans la littérature pour décrire et classer ces productions sonores. Ils permettent de caractériser la conformation du conduit vocal lors de la production d'un son ou d'une classe de sons. D'après (Munot et al., 2002), le mode d'articulation représente la manière selon laquelle le flux d'air provenant des poumons se dirige vers l'extérieur. On distinguera ainsi cinq principaux modes :

- le mode expiré ou inspiré, relatif au sens de circulation du flux d'air utilisé dans la production sonore ; le mode inspiré étant très minoritairement représenté (cas des clics) ;
- le mode **sonore** (également noté **voisé**) et **sourd** (également noté **non voisé**) faisant intervenir ou non la vibration des cordes vocales ;
- le mode **oral** et **nasal** suivant que le flux d'air passe, soit uniquement dans la cavité buccale (la cavité nasale étant fermée par appui du voile du palais contre la paroi pharyngale), soit dans les deux cavités buccale et nasale suite à l'abaissement du voile du palais provoquant l'ouverture de la cavité nasale ;
- le mode **libre**, **constrictif** (également noté **fricatif**) et **occlusif** relatif à la manière dont le conduit vocal laisse passer le flux d'air, soit librement (le conduit vocal ne présente aucune constriction), soit avec un écoulement limité suite à un rétrécissement du conduit vocal provoquant un bruit de turbulence (ou friction), soit momentanément stoppé suite à une occlusion du conduit vocal ;
- **labialisé** (également noté **arrondi**) ou **délabialisé** suite à une protrusion et un arrondissement des lèvres (mise en place d'une cavité labiale) dans le premier cas ou à une position neutre ou d'étirement dans le second.

L'Alphabet Phonétique International (API [12] illustré en figure 1.14) fait état de modes d'articulation supplémentaires, relatifs à une petite minorité de sons. On pourra ainsi parler également du mode *roulé* (ou **vibrant**) qui fait mention à un son pour lequel l'organe articulateur va rentrer en vibration durant sa production, du mode **approximant** qui fait mention à un rétrécissement du conduit, dans une moindre mesure pour s'apparenter au mode constrictif et notamment pour provoquer un bruit de turbulence et finalement le mode **battu** qui fait mention d'une fermeture si courte du conduit vocal qu'elle ne peut s'apparenter au mode occlusif.

12. publié en 1888 et révisé en 2005, l'API a pour objectif de transcrire les sons des langues du monde d'un point de vue phonétique.

Toujours d'après (Munot et al., 2002), le lieu ou point d'articulation est le point de rapprochement ou de contact de l'articulateur avec les parties fixes du conduit vocal. L'API fait état de 9 lieux d'articulation : bilabial (rapprochement ou contact des deux lèvres), labiodental (de la lèvre inférieure et des incisives supérieures), dental (de la langue et des incisives supérieures), alvéolaire (de la langue et les alvéoles), post-alvéolaire (de la langue et la jonction entre les alvéoles et le palais dur), rétroflexe (langue retournée vers le palais dur), palatal (de la langue et du palais dur), vélaire (de la langue et du palais mou), uvulaire (de la langue et de la luette), pharyngal (de la langue et du pharynx) et glottal (resserrement des cordes vocales avec ou sans vibration).

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 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	B			r					R		
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			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

FIGURE *1.14* – *Classification des sons des langues du monde suivant leur mode (en ligne) et leur lieu d'articulation (en colonne).*

Pour finir, le conduit vocal, du fait de sa complexité, a fait l'objet de nombreux travaux portant sur sa modélisation en vue, soit de mieux comprendre les mécanismes d'articulation par la simulation, soit pour affiner les systèmes de synthèse de la parole. Dans (Ghio, 2007), l'auteur répertorie les différentes approches de modélisation du conduit vocal présentes dans la littérature – modélisation acoustique, géométrique, articulatoire, biomécanique – et propose de les organiser hiérarchiquement en cascade, comme illustré en figure 1.15, afin de montrer leur possible combinaison.

1.3.1 Distinction voyelles/consonnes

Deux éléments caractérisent les voyelles sur le plan acoustique. Le premier est le libre passage du flux d'air dans le conduit vocal lors de leur production, contrairement aux consonnes où un rétrécissement plus ou moins important va s'opérer. Le deuxième repose sur le fait que la production des voyelles (non chuchotées) s'accompagne toujours d'une mise en vibration des cordes vocales, contrairement aux consonnes qui

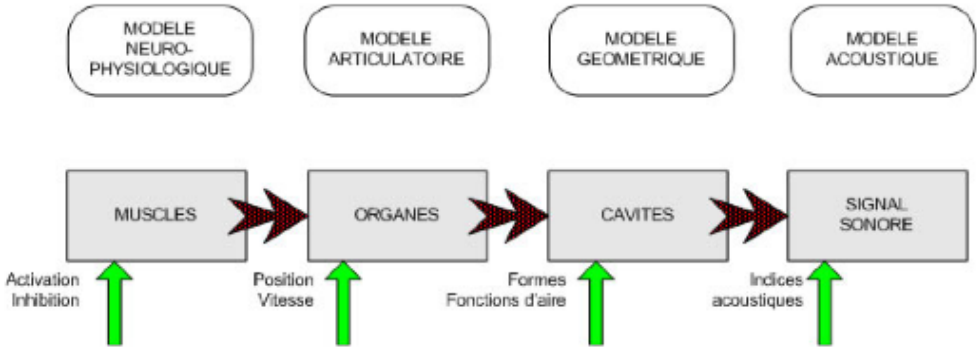


FIGURE *1.15* – *Hiérarchisation en cascade des différentes approches de modélisation du conduit vocal proposées dans la littérature, montrant leur possible combinaison : les sorties d'un modèle deviennent les entrées du modèle suivant (flèches horizontales) (source : (Ghio, 2007)).*

peuvent être voisées ou non voisées.

1.3.2 Réalisation des voyelles

Les 6 voyelles écrites du français se déclinent sous la forme de 16 voyelles dans le domaine phonétique. Chacune de ces voyelles étant associées à des configurations différentes du conduit vocal, elles seront caractérisées par des valeurs de formants différentes permettant leur distinction. On distingue quatre critères rentrant en jeu dans la configuration variable du conduit vocal lors de la production de voyelles :

- l'aperture de la mandibule correspondant au degré d'ouverture de la bouche. Les voyelles seront qualifiées de fermée, mi-fermée, mi-ouverte, ou ouverte. Le premier formant F1 étant déterminé par ce degré d'aperture et la hauteur de la langue, les voyelles ouvertes seront caractérisées par un F1 élevé (ex : voyelle /a/) dû à une cavité pharyngale très petite suite à l'ouverture maximale de l'aperture et un abaissement de la langue contrairement aux voyelles fermées qui présenteront un F1 bas (ex : voyelle /i/) dû à une cavité pharyngale très grande ;
- la position de la langue en avant ou en arrière de la cavité buccale. Les voyelles seront qualifiées d'antérieures ou de postérieures. Le changement de position ayant partiellement une incidence sur le deuxième formant F2, les voyelles antérieures seront caractérisées par un F2 élevé (ex : voyelle /i/) contrairement aux voyelles postérieures qui présenteront un F2 bas (ex : voyelle /u/) ;
- la position des lèvres, soit neutre ou étirée, soit formant la cavité labiale par protrusion et arrondissement. Les voyelles seront qualifiées de labiales/arrondies ou de non labiales/non arrondies/étirées. La formation de la cavité labiale ayant pour conséquence un abaissement des valeurs de formants F2 et F3, les valeurs de F2 et F3 des voyelles arrondies (ex : voyelle /y/) sont plus basses que les voyelles étirées (ex : /i/) ;
- le caractère de nasalité présent suite à l'abaissement du voile du palais et à l'ouverture de la cavité nasale. Les voyelles seront qualifiées d'orales (cavité buccale

uniquement) ou de nasales (cavités buccale et nasale). Du fait du couplage entre les deux cavités, l'analyse acoustique des voyelles nasales est largement plus complexe que celle des voyelles orales. Les observations rapportées dans la littérature font état de modifications spectrales des formants associés à la seule cavité orale interprétées par l'ajout d'anti-formants induits par la cavité nasale. Ces modifications peuvent se traduire par un affaiblissement de certains formants (notamment F1) et un décalage des valeurs des premiers formants (le lecteur pourra se référer à (Amelot, 2004) pour une revue des travaux réalisés autour des analyses acoustiques des voyelles nasales).

L'utilisation de ces caractéristiques dans la distinction des voyelles orales de l'ensemble des langues du monde est classiquement représentée sous la forme d'un trapèze vocalique comme illustré sur la figure 1.16 (les voyelles du français sont entourées). Les valeurs moyennes des formants de ces mêmes voyelles orales du français issues de (Tubach, 1989) sont fournies pour indication dans la table 1.3. Rappelons que F1 est déterminé par le degré d'aperture de la mandibule et la hauteur de la langue, F2 par la position de la langue (avant ou arrière) et des lèvres (étirée ou arrondie), F3 par la position des lèvres.

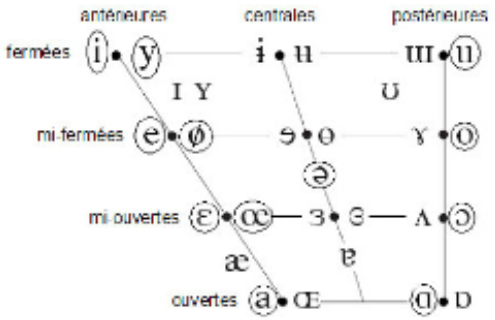


FIGURE 1.16 – *. Représentation des voyelles orales des langues du monde (les voyelles du français sont entourées), sous la forme d'un trapèze vocalique, suivant les critères articulatoires : aperture de la mandibule, position de la langue et position des lèvres (dans les paires de voyelles séparées par un point, la première voyelle représente une réalisation étirée, la seconde une réalisation arrondie).*

1.3.3 Réalisation des consonnes

Les consonnes sont considérées comme une classe de phonèmes plus hétérogènes que les voyelles car elles présentent des distinctions articulatoires plus larges. Le mode et le lieu d'articulation permettent de différencier chacune des 19 consonnes [13] du système consonantique français. Néanmoins, dès lors qu'il s'agit de proposer une classification de ces consonnes en macro-classes, les avis peuvent diverger, notamment si l'on

13. Les consonnes /ɲ/ et /ŋ/, bien que présentes dans l'API, ne seront pas décrites ici, car considérées comme non anatomiques mais comme la réalisation de deux consonnes atomiques.

Voyelle	Voyelle en contexte mot	F1	F2	F3
i	am**i**	308	2064	2976
y	ém**u**	300	1750	2120
u	c**ou**p	315	764	2027
e	**été**	365	1961	2644
ə	**je**u	381	1417	2235
o	s**eau**	383	793	2283
ɛ	m**e**r	530	1718	2558
œ	fl**eu**r	517	1391	2379
ɔ	p**o**rte	531	998	2399
a	p**a**pa	684	1256	2503

TABLE 1.3 – *Valeurs moyennes des formants F1, F2 et F3 pour les différentes voyelles orales du français, d'après (Tubach, 1989).*

considère des critères purement articulatoires ou acoustico-articulatoires.

Si l'on se réfère à l'API (figure 1.14), les consonnes du français peuvent se décliner en cinq macro-classes (par ordre d'apparition dans le tableau) : les consonnes occlusives ou plosives (/p/, /b/, /t/, /d/, /k/, /g/), les consonnes nasales (/m/, /n/), les consonnes vibrantes (/r/), les consonnes fricatives (/f/, /v/, /s/, /z/, /ʃ/, /ʒ/), les approximantes (ou spirantes) (/j/, /w/, /ɥ/) incluant les approximantes latérales (/l/). D'autres classifications, comme celle proposée dans (Munot et al., 2002), met en avant deux macro-classes, les plosives englobant les nasales d'un côté, les constructives de l'autre incluant les fricatives évidemment mais également les approximantes et les vibrantes en raison de la présence du rétrécissement (même plus faible) du conduit vocal (voir ci-dessus les différents modes d'articulation). D'autres préfèreront mettre en avant la présence ou non de vibrations des cordes vocales lors de la production des consonnes en proposant les deux macro-classes : consonnes sonores (/b/, /d/, /g/, /v/, /z/, /ʒ/, /m/, /n/, /r/, /j/, /w/, /ɥ/, /l/) et consonnes sourdes (/p/, /t/, /k/, /f/, /s/, /ʃ/) ; une dichotomie similaire pouvant être faite sur l'ouverture ou non de la cavité nasale avec, d'un côté, les consonnes orales (/p/, /b/, /t/, /d/, /k/, /g/, /r/, /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /j/, /w/, /ɥ/, /l/) et, de l'autre, les nasales (/m/, /n/). Finalement, une classification quelque peu différente est proposée dans (Meunier, 2007), reposant davantage sur des caractéristiques acoustico-articulatoires. Si les macro-classes consonnes plosives et fricatives restent présentes, cette classification suggère de regrouper toutes les autres consonnes (nasales, vibrantes, approximantes) sous la dénomination de consonnes *vocaliques* (ou sonantes). Ce regroupement est motivé par le fait que toutes ces consonnes, bien que très différentes, présentent un rétrécissement du conduit vocal, caractéristique des consonnes, mais également une structure formantique, (très) caractéristique des voyelles. Cette structure formantique, due au faible rétrécissement du conduit vocal, est bien plus instable (en termes de fréquences et d'intensité) que celle des voyelles et est très dépendante du contexte phonémique de réalisation de la consonne ciblée. La description (ci-dessous) un peu plus détaillée de

chacune de ces classes se fera sur ce schéma de classification.

Les occlusives

Les occlusives peuvent être sonores (/b/, /d/, /g/), ou sourdes (/p/, /t/, /k/). Compte tenu de leurs caractéristiques articulatoires (notamment le mode), elles produisent un signal complexe composé de deux séquences temporelles : (1) la tenue établie tout au long de l'occlusion (le flux d'air est stoppé), qui se traduit par un silence dans le cas des occlusives sourdes ou par un signal périodique (appelé barre de voisement et visible sur un spectrogramme) d'une fréquence égale à F0, pour les occlusives sonores, (2) un bruit impulsionnel d'explosion (ou "burst") dû au relâchement du ou des organes articulateurs (fin de l'occlusion du conduit vocal, laissant échapper le flux d'air préalablement stoppé).

Seule cette deuxième séquence permet de différencier les plosives entre elles sur le plan acoustique (différence en termes de fréquence, durée et énergie du burst), ces différences étant liées au lieu de leur articulation. Toujours d'après (Meunier, 2007), les plosives bilabiales /p/ et /b/ produisent un bruit d'explosion de très faible énergie, court et de fréquence basse. Les plosives alvéolaires /t/ et /d/ produisent un bruit intense, court et de haute fréquence. Les plosives vélaires /k/ et /g/ produisent un bruit intense et long, dépendant de la voyelle suivante en termes de fréquences. L'analyse des formants des voyelles réalisées consécutivement aux plosives permet également une très bonne identification de ces dernières et de leur lieu d'articulation.

Pour finir, les plosives sonores se distinguent des sourdes par une tenue moins longue, un bruit d'explosion moins intense (Munot et al., 2002) et une mesure de Délai d'Etablissement du Voisement [14], (noté VOT pour Voice Onset Time) en français, négative (VOT positif pour les consonnes sourdes), ces particularités étant directement liées à la mise en vibration des cordes vocales.

Les fricatives

Les fricatives peuvent être sonores (/v/, /z/, /ʒ/) ou sourdes (/f/, /s/, /ʃ/). Le bruit de friction issu du resserrement du conduit vocal forme à lui seul le signal apériodique caractéristique des fricatives sourdes qui vient se superposer à l'onde périodique issue de la source glottique dans le cas des fricatives sonores. Toujours en raison de la mise en vibration des cordes vocales, les fricatives sonores sont de plus courte durée que les sourdes et présentent également un affaiblissement d'intensité au niveau du bruit. De manière similaire aux plosives, la fréquence et l'intensité du bruit de friction apportent une information discriminatoire sur le lieu d'articulation des fricatives. Les fricatives labiodentales /f/ et /v/ produisent un bruit de faible intensité avec des pics diffus vers 3.5 et 8kHz. Les fricatives alvéolaires /s/ et /z/ produisent un bruit plus

14. intervalle de temps entre le burst et la mise en vibration des cordes vocales.

intense entre 4 et 8kHz avec des pics autour de 5 et 8kHz. Les fricatives palatales /ʃ/ et /ʒ/ produisent un bruit entre 2 et 7kHz avec un pic diffus autour de 4kHz.

Les vocaliques

Comme souligné plus haut, cette classe de consonnes regroupe des consonnes ayant deux traits communs - faible rétrécissement du conduit et structure formantique - mais qui sont, au final, très différentes :

- les consonnes nasales /m/ et /n/ présentent les mêmes modes et lieux d'articulation que les plosives bilabiale /b/ et alvéolaire /d/ respectivement. La seule différence réside dans la complexification du signal en raison de l'ouverture de la cavité nasale apportant au spectre, compris entre 200 et 2000Hz, des formants et anti-formants. Il est à noter que les formants du /m/ sont plus graves que ceux du /n/.
- la consonne /r/ est la consonne française la plus polymorphe, au sens où elle peut changer de mode et de lieu d'articulation en fonction du contexte dans lequel elle est réalisée. Apparentée au /r/ (vibrante alvéolaire) en présence d'accent régional, sa réalisation la plus fréquente repose sur un lieu d'articulation uvulaire. Au sein d'un groupe consonantique, elle peut être, soit vibrante, /ʀ/, si précédée d'une occlusive ou d'une plosive sonore, soit fricative /ʁ/ couplée à un phénomène de dévoisement si précédée d'une occlusive ou plosive sourde. Finalement, en contexte intervocalique, la réalisation de la consonne se rapproche d'une approximante ;
- les consonnes approximantes /j/, /w/, /ɥ/ sont des réalisations très atypiques, entre consonnes et voyelles, leur conférant très souvent l'appellation de semi-voyelle ou semi-consonne. Un parallèle, au niveau de l'articulation, a été établi entre ces consonnes et trois voyelles fermées, apparaillant le /j/ avec la voyelle /i/, le /w/ avec le /y/ et le /ɥ/ avec le /u/, bien que les consonnes présentent une structure formantique plus faible, sans partie stable contrairement à l'ensemble des voyelles ;
- la consonne approximante latérale /l/ tient ce qualificatif particulier du positionnement de l'apex de la langue, qui en s'appuyant sur les alvéoles, rétrécit le conduit vocal tout en laissant passer le flux d'air sur l'un des deux côtés ou les deux (latéraux) de la langue. Le signal qui en résulte est périodique (vibration des cordes vocales), sans bruit de friction et présente un F1 autour de 300Hz et un F2 dépendant des voyelles adjacentes.

1.4 Des sons à la parole continue...

Les informations descriptives fournies dans la section précédente sont issues d'études prototypiques des différentes réalisations sonores ciblées, faites, pour la grande majorité, sur des données très contrôlées. Hors, nous savons que la parole n'est pas la

simple juxtaposition d'éléments prototypiques et qu'un certain nombre de facteurs interviennent dans sa grande complexité et la variabilité de réalisation des phonèmes qui la composent. Ainsi, dans (Vaissière, 2006), l'auteur rapporte pas moins de six axes pouvant influer sur cette variabilité :

- les caractéristiques anatomiques de chaque individu, qui permettent d'identifier les locuteurs mais également leur état pathologique et leurs émotions ;
- l'adaptation par le locuteur de son élocution à une situation de communication particulière en terme de registres (du plus formel au plus informel), de styles de parole (narration, conversation, ...), d'attitudes exprimées, etc.
- le contexte phonétique dans lequel est prononcé un phonème et qui influence les gestes des organes articulateurs associés, faisant référence aux mécanismes de coarticulation (le lecteur pourra se référer aux premiers chapitres des thèses suivantes (Robert, 2008; Ma, 2008) pour une revue détaillée de ces phénomènes) ;
- la position du phonème dans la syllabe, le mot et la phrase influençant le geste articulatoire. En position dominante, le phonème pourra imposer ses caractéristiques aux phonèmes environnants ; en position dominée, il subira les caractéristiques de ses voisins, au point dans certains cas de disparaître ;
- la vitesse d'élocution (liée aux mécanismes prosodiques) susceptibles d'engendrer des phénomènes de sous-articulation (réduisant les différences acoustiques des voyelles et des consonnes), voire de suppression de certains phonèmes dans la prononciation de mots ou d'expressions ;
- les contrastes phonétiques dus aux accents régionaux.

Tous ces facteurs induisent des phénomènes singuliers ou combinés, qui viennent s'ajouter à ceux, déjà très complexes, des mécanismes de phonation et d'articulation.

Avant de clore ce premier chapitre, nous voudrions revenir sur quelques bases des mécanismes prosodiques, partie intégrante de la parole dans son rôle communicatif. D'après (Vaissière, 1999), la prosodie concerne, sur le plan de la production, l'ensemble des variations acoustiques dans la réalisation des phonèmes qui ne dépendent pas de leur identité ni de celle de leur entourage phonétique immédiat. En d'autres termes, elle désigne tous les aspects de la parole qui ne sont pas liés à la production et à l'identification des sons (Duez, 2007). Elle englobe les phénomènes :

1. d'intonation, souvent définie comme la courbe mélodique de la parole, qui permet de structurer un énoncé à la fois sur le plan syntaxique (fonction démarcative à différents niveaux syntaxiques : syllabes, mots, syntagmes, phrases, ...) et pragmatique (fonction informationnelle : focalisation, contraste, ...), mais également expressif (nuances de sens, attitudes et émotions) ;

2. d'accentuation lexicale, caractéristique intrinsèque des mots et généralement localisée au niveau de la syllabe, qui permet soit d'enlever une ambiguïté homomorphique dans les langues à tons lexicaux (ex : le mandarin) ou à accent lexical (ex : l'anglais), soit d'apporter une fonction démarcative (ex : accentuation de la dernière syllabe d'un groupe ou d'une phrase pour marquer une fin) ou pragmatique (ex : insistance par l'accentuation d'une syllabe donnée) ;

3. et des facteurs de performance dont le rythme, paramètre perceptif reposant sur l'organisation temporelle des syllabes accentuées et non accentuées basée sur la régularité. Le rythme permet, tout comme l'intonation, de structurer l'énoncé tant sur le plan syntaxique, pragmatique (information sur le style de parole par exemple (Duez, 1987)), voire expressif (attitudes principalement).

D'un point de vue acoustique, la prosodie se résume principalement aux variations ou mesures des paramètres suivants :

- fluctuation de la fréquence fondamentale (F0) (définie en section 1.2.2) ;
- variation de l'intensité (définie en section 1.2.2) ;
- variation dans le timbre (défini en section 1.2.2) ;
- mesure de la durée des phonèmes, des syllabes ou autres unités prosodiques ;
- analyse de la position et mesure de la durée des pauses.

Comme le souligne (Cristo, 2004), la prosodie a fait l'objet d'un *engouement hors du commun* ces vingt dernières années, suscitant un grand nombre de travaux de la part de chercheurs des domaines couverts par les sciences du langage (phonologie, syntaxe, sémantique, pragmatique) mais également des disciplines apparentées comme la psycholinguistique, la neurolinguistique, la phonétique clinique et le traitement automatique des langues.

1.5 Bilan

Comme nous avons pu le voir au travers de ce chapitre, la production de la parole est un processus très complexe faisant intervenir bon nombre de mécanismes différents. Le signal de parole qui en résulte présente de manière intrinsèque une très grande variabilité. Par ailleurs, comme souligné dans (Fougeron, 2013), [...] *Le signal de parole est façonné non seulement par des contraintes intrinsèques au système de production (c'est-à-dire des contraintes physiologiques, articulatoires, biomécaniques, aérodynamiques, acoustiques, perceptives) mais aussi par des contraintes linguistiques propres à la grammaire de la langue et par des contraintes liées à la communication. Toutes ces contraintes entraînent des variations dans la réalisation des productions, et l'on peut donc interpréter les variations phonétiques de surface comme étant (i) soit le résultat automatique, non-contrôlé, des manoeuvres mises en place pour la production de la parole, (ii) soit le reflet de manoeuvres contrôlées par le locuteur et exécutées pour satisfaire aux spécificités de sa langue et à l'efficacité de la transmission du message.* [...]. De très nombreuses études ont porté ces dernières décennies sur la variation de la parole, dans des conditions de production de moins en moins contrôlées - de la parole lue vers la parole spontanée (le lecteur pourra se référer à (Meunier, 2014) pour une revue de la littérature). La mise à disposition de grands corpus de parole ainsi que l'enrichissement que peuvent apporter les systèmes de traitement automatique de la parole (alignement automatique en phonèmes, syllabification, transcription de la parole, etc.) ont contribué au développement de ces travaux, notamment pour le français (Durand et al., 2003; Galliano et al., 2005; Adda-Decker, 2006; Bertrand et al., 2008; Bürki et al., 2008; Torreira et al., 2010; Meunier, 2014). Un autre moyen d'analyser la parole et ses variations est de se positionner dans un contexte d'observation où la parole est connue

pour être dégradée - parce qu'elle est perçue en tant que telle -, tel que la parole pathologique, la parole produite par un locuteur en apprentissage d'une seconde langue ou, encore, en présence d'accents régionaux bien prononcés. En effet, l'observation conjointe de parole considérée comme normale et de parole attendue comme dégradée peut permettre de déterminer des formes de variations communes, considérées alors comme "typiques" (même si celles-ci ne sont pas forcément présentes dans le même type de parole entre les deux populations) et des formes de variations uniquement présentes dans la parole dégradée directement attribuables au contexte d'observation (parole pathologique par exemple). Le projet ANR TyPaLoc, dans lequel nous sommes partenaire et que nous présenterons plus tard dans le document, était orienté dans cette optique.

L'étude des variations du français reste encore un vaste chantier pour répondre aux questions que nous pouvons nous poser :
"la variation de la parole a-t-elle des bornes fixées par des contraintes intrinsèques au système de production, notamment les contraintes linguistiques et communicationnelles citées plus haut ?"
"Si ces bornes existent, permettent-elles de définir une parole considérée comme "typique" à opposé à une parole "atypique" ?"

Nous ne sommes évidemment pas en mesure de répondre à ces questions. Néanmoins, elles sont au coeur de nos activités de recherche sur la parole pathologique que nous détaillerons plus tard dans ce document et le traitement de ce type de parole par des systèmes automatiques.

Chapitre 2

Dysphonie et Dysarthrie : deux cas particuliers de troubles de la communication

Sommaire

2.1 Généralités sur les troubles de la communication

Un trouble de la communication est défini par l'American Speech-Language Hearing Association (ASHA, 1993) comme une déficience dans la capacité à recevoir, envoyer, traiter et comprendre les informations de type verbal, non-verbal ou graphique. Cette définition renvoie à quatre grandes sous-classes de déficiences que sont les troubles de la voix, les troubles de la parole, les troubles du langage et les troubles de l'audition.

Troubles de la voix

La **dysphonie** (du grec : dys - idée de difficulté, phonie - voix) est le terme qui désigne couramment les troubles de la voix. A l'extrême, le terme **aphonie** peut venir le suppléer lorsque l'individu n'est plus en mesure de produire un seul son. La dysphonie est généralement définie comme une altération de la voix d'origine laryngée, et plus particulièrement, une altération des paramètres acoustiques caractéristiques de la qualité vocale, qui, comme nous l'avons vu dans le chapitre précédent (section 1.2.2), sont la hauteur, l'intensité et le timbre. (Le Huche et Allali, 2010c) propose une définition quelque peu différente : *[...] la dysphonie est un trouble momentané ou durable de la fonction vocale ressenti comme tel par le sujet lui-même et son entourage. Elle se traduit le plus souvent, mais non obligatoirement, par une altération d'un ou plusieurs paramètres acoustiques de la voix et par ordre de fréquence, du timbre, de l'intensité et de la hauteur tonale [...]*. Cette définition, qui peut, dans sa première phrase, semblait plus vague, est au contraire très ouverte au sens où elle fait intervenir deux notions intéressantes et importantes dans la dysphonie : la notion temporelle - momentané ou durable - et la notion de ressenti vocal par le patient et les personnes qui le côtoient. En effet, les remarques suivantes : (1) *Ma voix me fatigue, je n'ai plus de voix en fin de journée, j'ai mal en parlant*, (2) *On ne m'entend pas* ou encore (3) *Je n'aime pas ma voix* sont autant de plaintes qu'un phoniatre pourra entendre lors de la première consultation avec un patient atteint de dysphonie (Crevier-Buchman, 2001). Comme nous le verrons plus amplement dans la section suivante, la dysphonie peut être soit d'origine (dys)fonctionnelle ie entretenue par une perturbation du comportement vocal, soit d'origine organique, les deux étant parfois difficilement séparables, le premier pouvant engendrer le second (Le Huche et Allali, 2010c) ou inversement.

Troubles de la parole

Les troubles de la parole regroupent l'ensemble des déficiences affectant la production de la parole, dont voici quelques exemples :

– le **bégaiement**, défini par l'Organisation Mondiale de la Santé comme *un trouble du rythme de la parole dans lequel l'individu sait exactement ce qu'il veut dire, mais est incapable de le dire en raison d'une répétition involontaire, d'une prolongation ou de l'arrêt d'un son* (Hirsch, 2007), et touchant 1% de la population mondiale ;

- la présence de malformations dans la cavité buccale, soit au niveau du palais (**fente palatine**), soit au niveau de la lèvre supérieure (**fente labiale**), soit au niveau des deux (**fente labio-palatine**), pouvant engendrer des difficultés d'ordre articulatoire et de nasalité et touchant un enfant sur mille au niveau européen (Bechet, 2011), soit, également, au niveau du frein de la langue, trop court, pour réaliser les mouvements attendus dans la réalisation de certains phonèmes, voire de la dentition ;
- l'**apraxie de la parole**, définie comme une altération d'origine neurologique affectant la programmation des mouvements propres à la parole ;
- la **dysarthrie**, sur laquelle nous reviendrons plus en détails dans la suite du document, qui rassemble l'ensemble des troubles de la réalisation motrice de la parole dus à des lésions du système nerveux central ou périphérique ;
- la **dysprosodie**, définie comme l'atténuation ou la disparition complète des gestes prosodiques (rythme, intonation, timbre de la voix), se traduisant par une parole sans mélodie, monotone et un débit très lent ;
- la **dyslalie infantile d'ordre fonctionnel** (à opposé à la dyslalie d'ordre organique due à une malformation des articulateurs comme cité ci-dessus, à la perte de l'audition, à une dysarthrie ou une apraxie), définie comme un trouble de l'articulation, touchant les enfants lors de l'acquisition du langage. Elle se traduit par une mauvaise réalisation d'un ou plusieurs phonèmes (voire l'omission complète ou le remplacement du phonème par un autre plus facile à prononcer) due à des erreurs motrices systématiques et permanentes dans le placement des articulateurs en jeu. L'exemple le plus "parlant" est le phénomène de *zézaiement* pour lequel les phonèmes /s/ et /j/ seront systématiquement remplacés par le phonème /z/. Ces troubles sont généralement considérés comme pathologiques chez l'enfant après l'âge de 4 ou 5 ans ;
- **les troubles de l'articulation** dus à une ablation partielle ou totale d'un ou plusieurs articulateurs, suite à une exérèse chirurgicale.

Troubles du langage oral

Les troubles du langage oral regroupent l'ensemble des déficiences affectant les mécanismes cérébraux complexes mis en oeuvre pour établir un acte langagier, préliminaire aux phases d'élaboration, planification, programmation et exécution de l'acte moteur de parole. Généralement, on distingue deux grandes classes de troubles : les dysphasies - troubles du langage oral relatifs à son apprentissage, affectant typiquement les enfants, mais pouvant perdurer chez l'adulte -, les aphasies - troubles du langage oral acquis, suite à une ou des lésions cérébrales, affectant typiquement les adultes. Selon la lésion et sa localisation dans le cerveau, ces troubles pourront se manifester sous différentes formes : troubles phonologiques (difficulté dans la discrimination et l'assemblage des sons), troubles lexicaux (perte de la signification attachée aux mots - un mot est utilisé hors de son contexte ou à la place d'un autre - ou tout simplement l'accès lexical au mot - perte du mot), troubles syntaxiques (non respect de la grammaire, phrases mal construites), troubles pragmatiques (manque d'adaptation des choix linguistiques au contexte du discours), voire sous des formes combinées, mêlant différents troubles.

Les aphasies les plus "célèbres" sont les aphasies de Broca et de Wernicke. La première fait référence à une aphasie dite d'expression (également appelée aphasie motrice) pour laquelle les patients, conscients de leurs troubles, sont incapables de formuler oralement leurs idées, bien que tout à fait intactes. De forme télégraphique, le langage est alors réduit à sa plus simple expression, voire à une simple succession de mots sans structure grammaticale. Des troubles phonémiques (mots déformés par omission, ajout, déplacement ou substitution d'un ou plusieurs phonèmes) et de l'articulation peuvent également apparaître. La seconde fait référence à une aphasie dite de réception (également appelée aphasie sensorielle) qui se manifeste par de grandes difficultés pour les patients, inconscients de leurs troubles, à comprendre le langage oral et écrit. La parole est tout à fait fluente (contrairement à une aphasie de Broca), mais comporte des incohérences sémantiques (utilisation de mots inappropriés, voire des non-mots) pour les formes peu sévères jusqu'à être complètement incompréhensible sémantiquement pour les formes très sévères. Aucun trouble articulatoire n'est alors observé.

Troubles de l'audition

Les troubles de l'audition rassemblent l'ensemble des déficiences altérant la réception et/ou le traitement des sons nécessaires à tout acte de communication et indispensable pour un développement harmonieux du langage oral. L'apparition de déficits auditifs chez le nouveau né, l'enfant, l'adulte ou les personnes âgées n'aura donc pas la même incidence sur la production de la parole.

On distingue généralement 3 types de déficits auditifs, (1) la surdité de transmission due à une atteinte de l'oreille externe ou moyenne, qui empêche le libre passage des ondes sonores vers la cochlée ; l'intensité de l'onde sonore s'en trouve diminuée, mais sans déformation lorsqu'elle parvient aux centres auditifs ; la personne s'entend généralement parler ; sa voix a donc toutes les caractéristiques de timbre, d'intonation et de rythme d'un entendant ; (2) la surdité de réception due à une atteinte de l'oreille interne et/ou des voies et centres nerveux auditifs (au niveau du cerveau), qui entrave la transformation de l'onde sonore en un message nerveux, entraînant un manque d'amplification et une distorsion plus ou moins importante du son lors de son traitement ultérieur au niveau cérébral ; Dans le cas de surdité grave, la personne ne s'entend pas parler ; elle a, par conséquent, des difficultés à contrôler l'intensité et le timbre de sa voix ; et (3) la surdité mixte mélangeant les deux précédentes.

Les troubles de communication sont directement dépendants du degré de sévérité de la surdité. Dans la surdité dite légère, la parole normale est correctement perçue excepté pour des niveaux sonores faibles. Des troubles articulatoires et des confusions phonétiques (sur les consonnes par exemple si les fréquences aiguës sont perturbées) peuvent être observés dans le langage oral, ainsi qu'un léger retard dans son apprentissage. Dans la surdité dite moyenne, la parole est perçue uniquement si elle s'accompagne d'un fort niveau sonore. L'articulation de la parole ainsi que la mélodie de la voix sont défectueuses. La gêne est accentuée dans des ambiances bruitées, au téléphone ou lorsque la personne ne peut s'aider de la gestuelle labiale dans son mécanisme de compréhension de la parole. Dans la surdité sévère, la parole est perçue que sommairement.

Dans certains cas, la voix est entendue mais la parole incompréhensible. L'acquisition du langage est très difficile. La parole produite s'avère déformée avec un rythme haché. Dans la surdité profonde, la parole, même accompagnée d'une forte intensité sonore, n'est plus perçue.

La prévalence de la surdité permanente néonatale, pouvant entraîner des déficits auditifs de perception ou de transmission, à des degrés de sévérité variables est estimée à 1 pour 1000 naissances en France.

Résumé

Nous venons de voir que les troubles de la communication sont d'origines très diverses, qu'ils touchent un nombre non négligeable d'individus dans le monde et ont surtout des conséquences très variables sur leurs facultés de communication et leur vie de tous les jours : que ce soit à l'école ou au travail, dans leurs relations sociales avec autrui,...

Dans la suite de ce chapitre nous allons nous intéresser plus particulièrement à la dysphonie et à la dysarthrie, ces deux types de pathologie étant au coeur de nos activités de recherche depuis 2004.

2.2 Dysphonie

D'après un rapport produit par l'ASHA [1], la prévalence de la dysphonie dans le monde variait entre 3 et 7% de la population suivant une revue de la littérature menée en 1981 (Healy et al., 1981). Par ailleurs, ces valeurs augmentent de manière très significative suivant la profession exercée et l'usage de la voix qui en résulte (enseignants, chanteurs, acteurs, commerciaux, ...). Ainsi, pour les Etats-Unis, (Titze et al., 1997) avance que pour 5 à 10% de la population active américaine, la voix est un outil majeur de la profession tandis que (Roy et al., 2004) met en relation une prévalence de 11% pour des enseignants contre 6,2% pour des non-enseignants suite à une étude menée auprès de plus de 2000 individus. Dans cette même étude, il est également souligné que les troubles vocaux augmentent avec l'âge avec un pic maximal dans le groupe d'âge 50-59 ans. Certains enseignants sont contraints d'abandonner définitivement leur profession de manière prématurée en raison de leurs difficultés vocales (Sapir et al., 1993). Une étude similaire réalisée en Finlande a également montré que les troubles vocaux ont augmenté de manière significative en 12 ans (1988 à 2000) (Simberg et al., 2005). Finalement, le coût des troubles vocaux chez les enseignants uniquement, en termes de soins de santé et de journées de travail perdues, est estimé aux Etats-Unis à 2,5 milliards de dollars par an suivant l'analyse réalisée dans (Verdolini et Ramig, 2001). Ces différents chiffres et d'autres, présentés dans un rapport français consacré à la voix et ses troubles chez les enseignants (INSERM, 2007) (rédigé par un groupe d'experts réunis

1. disponible à cette URL : http ://www.asha.org/policy/TR2005-00158.htm#r41.

par l'Institut National de la Santé et de la Recherche Médicale (INSERM) pour répondre à la demande de la Mutuelle Générale de l'Education Nationale (MGEN)) révèlent toute l'importance à donner aux troubles de la voix. Comme nous le verrons ci-après, même si certains de ces troubles peuvent avoir que peu de conséquences sur les paramètres vitaux d'un patient, les retombées sur sa vie professionnelle et sociale, sur son mental, son physique, son état émotionnel et sa communication sont loin d'être négligeables. Néanmoins, contrairement à certains pays européens comme l'Angleterre, la Finlande ou la Pologne, les troubles de la voix ne sont toujours pas reconnus comme pathologies professionnelles en France.

2.2.1 Terminologie et voix pathologiques

Dans la pratique clinique, ainsi que dans la littérature, les termes utilisés pour décrire une altération de la qualité vocale, du point de vue perceptif, sont très nombreux. Ces termes peuvent porter d'une manière très globale directement sur la voix - voix claire, forte, faible, blanche, rocailleuse, rauque, soufflée, éraillée, criarde, grésillante... - ou, de manière plus précise, sur les trois critères - hauteur, intensité et timbre - associés à la qualité vocale (cf. section 1.2.2 du chapitre précédent). Dans (Dupessey et Coulombeau, 2003), le degré d'altération vocale est retranscrit selon les critères de hauteur (tonalité) - très basse, basse, normale, élevée, très élevée -, d'intensité - faible, normale, forte - et de timbre - voilé avec comme dénomination équivalente soufflé/désonorisé, éraillé équivalent à grésillant/rauque/bitonal, forcé équivalent à serré/rauque. Dans (Hammarberg et al., 1980), un ensemble de 28 critères de jugement de la voix sont proposés, sélectionnés parmi les 50 termes les plus fréquemment utilisés par des cliniciens suédois de l'Hôpital Universitaire de Huddinge pour décrire un ensemble de 115 voix. Parmi ces 28 critères (23 décrivant la voix, 2 la hauteur et 3 le registre), on retrouve de manière individualisée les termes de voix soufflée et de désonorisation (synonyme de dévoisement), d'éraillement, de voix rauque et de bitonalité, de voix forcée et de voix rauque. Dans (Anders et al., 1988), la voix éraillée *est une voix contenant clairement des composantes de bruit, et qui peut être considérée comme rauque et soufflée (c'est-à-dire éléments de bruit dans la source associés à des bruits de friction)*. D'après ces trois exemples, il semble difficile de parvenir à un consensus quand à la définition propre de ces différents termes, leur nuance étant parfois difficile à discerner. Néanmoins, sachant que certains de ces termes seront utilisés dans la section suivante dédiée à la classification des dysphonies, notamment pour caractériser les altérations du timbre, nous nous permettons de reporter ici quelques citations les décrivant :

- *[...] timbre voilé : il s'agit de la sensation de fuite d'air pendant l'émission du son laryngé. On retrouve également les caractérisations suivantes dans la littérature : soufflé, désonorisé [...] (Dupessey et Coulombeau, 2003).*

- *[...] L'accolement des plis vocaux peut être plus ou moins ferme. Lorsque cette fermeté d'accolement augmente, le timbre vocal s'enrichit et l'on dit que la voix acquiert du mordant. Sur le plan physique, les ouvertures glottiques sont plus brusques et plus brèves. Cela se traduit sur le plan acoustique par un "enrichissement en aigus" du spectre sonore. Lorsqu'au contraire, l'accolement des plis vocaux est relâché, ou incomplet, on a une*

voix de timbre pauvre. Parfois, ce timbre vocal s'accompagne d'un bruit de souffle et l'on dit que ce timbre est voilé [...] (Le Huche et Allali, 2010a).

- *[...] Le timbre voilé : on observe une perte des harmoniques aigus. Les cordes vocales présentent un défaut de fermeture modéré. La voix est sans portée et de faible intensité. Elle manque de netteté, de brillant. Le timbre soufflé : il y a adjonction d'un bruit de souffle par défaut important de fermeture glottique et la présence en excès d'harmoniques aigus [...] (Coudière, 2003).*

- *[...] voix soufflée : fuite audible d'air à travers la glotte à cause d'une fermeture glottique insuffisante [...]. (Revis, 2004) (traduction française de (Hammarberg, 1986)).*

- *[...] timbre éraillé : il correspond aux irrégularités vibratoires perçues dans le son émis. Autres appellations : grésillant, rauque, bitonal [...] (Dupessey et Coulombeau, 2003)..*

- *[...] Le timbre éraillé : il s'agit d'une superposition d'une vibration parasite irrégulière sur le son fondamental laryngé, lorsqu'il y a lésion du bord libre de la corde vocale qui vibre à sa fréquence propre. La vibration perd de sa souplesse et de sa régularité. La glotte n'est pas totalement fermée. Il peut aussi s'agir d'une voix bitonale non reconnue comme telle par l'oreille, qui n'arrive pas à différencier les deux fréquences. Le timbre rauque : la voix est grave, soufflée, son signal apériodique. L'étendue vocale est diminuée. Les cordes vocales s'accolent mal, semblent rigides du fait d'une altération de leur capacité vibratoire ou du tonus glottique. La voix est émise dans le bas pharynx. On a une sensation d'effort avec un son rugueux, dur. [...] (Coudière, 2003).*

- *[...] voix éraillée : bruit non-périodique avec haute fréquence, probablement en rapport avec un certain type de vibrations irrégulières des cordes vocales. Voix rauque : bruit non-périodique avec basse fréquence, probablement en rapport avec un certain type de vibrations irrégulières des cordes vocales. Diplophonique/Bitonalité : deux fréquences différentes perçues simultanément [...] (Revis, 2004) (traduction française de (Hammarberg, 1986)).*

- *[...] timbre forcé : il désigne l'impression auditive d'effort pendant la phonation. Autres dénominations : serré, rauque [...] (Dupessey et Coulombeau, 2003)..*

- *[...] Le timbre serré : le patient force sur son larynx qui a une position trop haute. Cela entraîne une contraction et une diminution des résonateurs. La pression expiratoire est très importante et provoque un accolement brutal des cordes vocales. La voix a beaucoup trop d'harmoniques aigus [...] (Coudière, 2003).*

Nous reviendrons, plus tard dans ce chapitre, sur la problématique de l'évaluation, notamment perceptive, des troubles de la voix et de la parole.

2.2.2 Classification des dysphonies

Comme vu précédemment, la dysphonie est définie comme un trouble de la fonction laryngée pouvant aller jusqu'à l'aphonie, perte totale de la voix, dont l'origine est soit fonctionnelle, soit organique, mais qui, au fur et à mesure de son évolution, pourra mêler les deux. Apparaissant de manière progressive ou brutale, se manifestant de manière intermittente ou constante, elle constituera une gêne plus ou moins importante pour la personne dysphonique et son entourage, sous la forme d'une diminution du confort vocal et/ou par une altération des qualités acoustiques de la voix. Ainsi, l'altération vocale peut être audible (altération des paramètres acoustiques) mais, également, apparaître sous la forme de difficultés vocales ressenties (diminution du confort vocal) par le patient sans présence obligatoire d'altérations acoustiques.

(Le Huche et Allali, 2010b) propose une classification des dysphonies à deux niveaux : les dysphonies d'origine (dys)fonctionnelle et celles d'origine organique. Nous adopterons cette classification par la suite bien que d'autres classifications aient également été proposées.

Dysphonie d'origine (dys)fonctionnelle

La dysphonie d'origine fonctionnelle est définie par (Le Huche et Allali, 2010c) comme une altération de la fonction vocale **entretenue** essentiellement par une **perturbation du geste vocal**. L'usage de la terminologie "dysphonie dysfonctionnelle" est d'ailleurs suggérée par Le Huche à la place de "dysphonie fonctionnelle" afin d'éviter toute confusion et de rendre compte de la présence possible de lésions organiques concomitantes, voire de lésions temporaires du larynx à l'origine de la perturbation du geste vocal en cause dans l'apparition de la dysphonie. Pour revenir aux gestes vocaux, on parlera de malmenage lors d'une mauvaise utilisation du geste vocal, de surmenage lors d'une utilisation normale (voire excessive) de la voix sans repos compensatoire suffisant, pouvant tous deux entraîner un forçage vocal. Comme précisé dans (Garnier et al., 2007), le forçage vocal est défini, dans la majorité des études *comme une utilisation excessive de la voix dans le cadre de situations de communication particulières ou de modes de production s'écartant de la parole "confortable"*. Néanmoins, comme le souligne également cette étude, il est très difficile d'apporter une seule définition à cette terminologie tant elle peut être étudiée sous différents points de vue (le lecteur pourra se référer à la thèse de Maëva Garnier (Garnier, 2007) pour une revue complète de la littérature francophone et anglophone sur le forçage vocal).

Dysphonie dysfonctionnelle sans complication laryngée
Si plusieurs théories ont tenté de trouver des causes à l'origine des dysphonies dysfonctionnelles sans présence de lésions du larynx : causes physiologique, auditive, hormonale, psychologique, ...sans faire l'unanimité, il semblerait que la notion de cercle vicieux du forçage vocal, de facteurs favorisants et déclenchants soient les paramètres les mieux acceptés pour comprendre le phénomène d'installation, d'entretien voire de

complications des dysphonies dysfonctionnelles (Le Huche et Allali, 2010c). Le cercle vicieux du forçage vocal s'explique tout simplement par le fait qu'un individu va essayer de compenser une diminution de son efficacité vocale de manière inadéquate. Plus un individu sentira une gêne en parlant, plus il va forcer sur sa voix ; plus il va forcer sur sa voix, plus la gêne va devenir importante, il forcera alors sur sa voix encore davantage, ... Lorsqu'il est récurrent, ce cercle vicieux peut conduire alors à une baisse progressive et régulière du rendement ; le geste vocal est différent et de plus en plus inefficace, la dysphonie s'est installée.

Les conséquences de ce cercle vicieux sont multiples, principalement dues à une évolution des gestes vocaux vers la voix d'insistance (à opposé à la voix d'expression simple et voix implicatrice dite projetée, classiquement utilisées en mode communicationnel). La voix d'insistance présente un changement de posture caractéristique (affaissement thoracique, arrondissement du dos et déplacement de la face vers l'avant), pouvant engendrer un étirement des muscles du larynx entravant ses mouvements et par compensation une crispation de tous les autres muscles impliqués dans les mécanismes articulatoires, voire d'autres muscles du visage et membres supérieurs. Les conséquences ressenties par l'individu peuvent être une fatigabilité vocale, une impression de manque d'efficacité (*"on ne m'entend pas"*), la présence d'une pathologie au niveau de la gorge (picotement, gêne, ...), etc. Dans certains cas, des altérations laryngées peuvent effectivement apparaître comme une irritation de la muqueuse laryngée par exemple.

Les facteurs déclenchants sont soit d'ordre organique (affaiblissement des résistances physiques, affectations Otho-Rhino-Laryngologiques (ORL), toux, période prémenstruelle, grossesses et interventions abdominales, etc.), soit d'ordre psychologique (contrariétés d'ordre émotionnel ou choc psychologique dus à des événements professionnels, familiaux ou sentimentaux). Outre la situation professionnelle obligeant un usage intensif de la voix (surmenage) et l'usage défectueux de la voix (malmenage), aspects que nous avons déjà évoqués ci-dessus, les facteurs favorisants le forçage vocal sont multiples et très divers. Voici les principaux (1) caractéristiques psychologiques telles que le stress, l'émotivité, l'anxiété, qui entraînent (2) d'autres facteurs, d'ordre physiologique tout aussi défavorables, tels que la sécheresse de la muqueuse, la rigidité posturale, l'hypertonicité des muscles laryngés, la perturbation du geste respiratoire, (3) les affections de type ORL ou allergies, (4) le tabagisme et l'alcoolisme (provoquant des irritations de la muqueuse laryngée), (5) le facteur environnemental tel que le bruit, les conditions climatiques (vent, sécheresse de l'air, froid, ...), la présence de climatisation, l'exposition à la poussière ou à des vapeurs irritantes.

Au niveau clinique et phonatoire, on distinguera :

– les dysphonies à larynx hypotonique[2] présentant un défaut d'accolement des cordes vocales sur tout ou partie des bords libres. Au niveau acoustique, le défaut d'accolement pourra se traduire par une fuite d'air résultant en un timbre voilé ; le manque de tonus par une F0 abaissée ; les deux réunis par une intensité faible ;

2. Diminution de la tonicité musculaire (tension), responsable ici du relâchement du larynx.

– les dysphonies à larynx hypertonique[3], présentant un serrage important des plis vocaux, pouvant aller jusqu'au rapprochement des bandes ventriculaires. Ces phénomènes peuvent se traduire au niveau acoustique par une fréquence fondamentale et une intensité plus élevées et un timbre forcé.

Dysphonie dysfonctionnelle avec complication laryngée

Les laryngopathies dysfonctionnelles regroupent l'ensemble des altérations du larynx apparaissant suite à un comportement de forçage vocal. On y distingue :

– le **nodule**, épaississement de la muqueuse laryngée, soit unilatéral, soit bilatéral, localisé généralement au niveau des bords libres entre le tiers moyen et antérieur de la corde vocale - le point nodulaire. Il est à noter que l'apparition de nodules est plus fréquente chez la femme que chez l'homme. Au niveau acoustique, la présence d'un nodule pourra se traduire par un abaissement de F0 (augmentation de la masse cordela), une difficulté dans les aigus avec de possibles désonorisations (perturbations du mode vibratoire), un timbre éraillé et/ou voilé (gêne lors de l'accolement des cordes vocales entraînant une fuite d'air), parfois des sons bitonaux (vibrations du nodule avec la muqueuse) ;

– le **pseudo-kyste séreux**, gonflement de la muqueuse, situé au point nodulaire, le plus unilatéral, caractérisée par son aspect translucide le différenciant du nodule ou d'un polype. Ses conséquences sur le plan acoustique sont similaires à celles du nodule, excepté qu'il peut induire un timbre plus éraillé ;

– l'**épaississement muqueux fusiforme (ou oedème en fuseau)**, lésion de la muqueuse, très étalée sur le bord libre de la corde vocale, uni ou bi-latérale, caractérisé par un timbre plus rauque comparé au pseudo-kyste séreux et un abaissement de F0 ;

– le **polype**, pseudo-tumeur bénigne d'origine inflammatoire, qui pourra se traduire au niveau acoustique par un abaissement de F0 (augmentation de la masse cordela), un timbre éraillé (vibration du polype, de manière synchrone ou asynchrone vis à vis de la vibration laryngée) et des désonorisations et des irrégularités (arrêt localisé de la vibration des cordes vocales) ;

– **kyste muqueux par rétention**, tuméfaction sous le bord libre d'une corde vocale liée à l'obstruction du canal excréteur d'une glande muqueuse. Il se traduit par une intensité plus faible, un timbre assourdi, parfois éraillé ou soufflé ;

– l'**hémorragie sous-muqueuse**, rupture vasculaire au niveau de la corde vocale, pouvant se doubler d'une rupture du muscle thyro-aryténoïdien (coup de fouet laryngien), se traduisant par un hématome sur la face supérieure de la corde vocale. Au niveau acoustique, on observe une baisse subite de l'intensité s'accompagnant d'une altération plus ou moins marquée du timbre, pouvant aller dans certains cas vers une quasi aphonie ;

3. Excès de tonicité musculaire (tension).

– l'**ulcère de l'aryténoïde**, altération de la muqueuse due à des accolements brutaux des cordes vocales (coups de glotte) et favorisée par un reflux gastro-oesophagien, siégeant le plus souvent sur la partie postérieure des cordes vocales. Il se traduit généralement par un timbre pauvre et des coups de glotte fréquents.

Dysphonie d'origine organique

En opposition aux dysphonies (dys)fonctionnelles, la dysphonie d'origine organique *implique l'existence d'une lésion de l'appareil phonatoire dont la responsabilité dans le déficit de la fonction vocale apparaît comme prépondérante (Le Huche et Allali, 2010c).* Néanmoins, comme nous l'avons déjà évoqué, il est rare que ce type de dysphonie ne soit pas associé à une composante dysfonctionnelle suite à des mécanismes réactionnels et compensatoires inadaptés. Différentes causes organiques à l'origine de ces dysphonies sont répertoriées :

– les **laryngites**, (1) dites aigües (suite à une classique affection rhino-pharyngée que tout individu peut attraper par une exposition au froid) pouvant donner lieu à des complications fonctionnelles, (2) dites chroniques, définies par une altération inflammatoire de la muqueuse laryngée et favorisées principalement par l'alcool, le tabac et le malmenage vocal et se traduisant par une altération progressive du timbre (voilé puis rauque et éraillé) ou (3) plus spécifiques telles que, par exemple, l'oedème chronique des cordes vocales, dû principalement au tabagisme, qui, par un oedème localisé dans l'espace de Reinke déforme la face supérieure et le bord libre et se traduit par une diminution de l'intensité et un assourdissement du timbre. On trouve également comme autre exemple de laryngite spécifique la papillomatose laryngée, très fréquente chez l'enfant, qui apparaît sous la forme de tumeurs bénignes au niveau des cordes et se traduit par une diminution de l'intensité, un abaissement de la F0 et par un timbre assourdi et "mouillé" ;

– la **paralysie laryngée**, uni- ou bi-latérale (plus rare), totale, partielle voire intermittente, peut avoir pour origine une atteinte du système nerveux central ou périphérique (paralysie bi-latérale), d'une atteinte du nerf récurrent (d'où la terminologie paralysie récurrentielle uni-latérale) innervant le larynx ou tout simplement une lésion du larynx lui-même, suite à une arthrite crico-aryténoïdienne bloquant l'articulation par exemple. Les conséquences sur le plan acoustique peuvent être les suivantes : une augmentation de la F0 (perte de masse suite à une atrophie du muscle vocal), une intensité faible, un timbre voilé avec des désonorisations irrégulières (fuite glottique), présence de sons bitonaux (fréquences de vibration des cordes différentes) ;

– le **cancer du larynx**, entraînant une laryngectomie (ablation) partielle ou totale du larynx. Dans le cas d'une laryngectomie partielle, les altérations sur le plan acoustique seront très dépendantes du type d'ablation réalisé ainsi que des méthodes de reconstruction chirurgicale (si possible) mises en oeuvre. Dans le cas

d'une laryngectomie totale, le patient se voit démuni de ses mécanismes phonatoires, voire articulatoires (dérivation du courant d'air pulmonaire qui ne passe plus par le nez, ni la bouche). Il peut très rapidement recouvrir des mécanismes communicationnels proches de la parole chuchotée. Néanmoins, la réhabilitation vocale devra passer par l'apprentissage de nouveaux mécanismes de production sonore basés par exemple sur la voix oesophagienne ;

- les **anomalies laryngées congénitales**, qui peuvent survenir au niveau de la structure laryngée suite à des anomalies neurologiques (entraînant des paralysies uni- ou bi-latérales), à des syndromes liées à des aberrations chromosomiques (trisomie 21 par exemple), ou pour finir à des formations pseudo-tumorales telles que des kystes ou vergetures au niveau des cordes vocales ;

- les **traumatismes laryngés**, soit externes provoqués, par exemple, lors d'accidents de la route ou d'accidents sportifs et entraînant différentes lésions du larynx : hématomes, contusions, étirements ou désinsertion des cordes vocales, ..., soit internes suite à une intervention chirurgicale concernant la filière respiratoire, à une intubation, à de la radiothérapie cervicale, etc.

- un **dérèglement hormonal**, provoquant par exemple une masculinisation de la voix et se traduisant initialement par des irrégularités dues à une perte de contrôle de la voix, puis par un abaissement de la F0 ;

- les **atteintes du système nerveux central ou périphérique** ayant pour incidence des troubles de la réalisation motrice de la parole (dysarthrie), et notamment des mécanismes phonatoires. Ces atteintes peuvent conduire à une hypotonie de la musculature laryngée et respiratoire donnant lieu à une faible intensité et un abaissement de la F0, ou au contraire, à une hypertonie donnant lieu à des hésitations au démarrage du voisement, des discontinuités dans la production vocale, une augmentation de la F0, un timbre sourd et voilé (mauvais accolement des cordes vocales). Elles peuvent également induire des spasmes ou des paralysies, provoquant des difficultés dans le mécanisme vibratoire (Teston, 2001). (Remacle, 1999) propose dans ce contexte le terme de dysphonie dysarthrique.

Cas particulier de la dysphonie spasmodique

Trouble rare, la dysphonie spasmodique est affiliée au groupe des dystonies focales, définies comme un trouble moteur qui se caractérise par des contractions musculaires inopinées affectant les muscles impliqués dans une même fonction telle que par exemple le regard, la respiration, la déglutition, et ce qui nous intéresse ici la fonction vocale, d'où la terminologie de dystonie laryngée. L'origine des dysphonies focales reste encore très floue, soit attribuée à une lésion des noyaux gris de part leur rôle dans la boucle motrice de régulation (cf. section 1.1.3), soit d'ordre psychologique, les troubles intervenant très souvent dans un contexte de stress et affectant des individus au profil psychologique particulier (Le Huche et Allali, 2010c). Caractérisée par des

spasmes laryngés et/ou respiratoires, la dysphonie spasmodique se traduit par des blocages ou des décrochages intermittents de la voix, résultats de la crispation des muscles laryngés lors de la fermeture des cordes vocales, ou plus rarement lors de leur accolement. La voix peut également présenter un tremblement (voix chevrotante). Sur le plan acoustique, on observe des baisses d'intensité suivies de remontées excessives, des brusques augmentations de la F0 pouvant s'accompagner d'un timbre voilé ou éraillé.

Compléments

Toujours sur des bases étiologiques, d'autres classifications ont été proposées dans la littérature. Dans (Remacle, 1999), les dysphonies d'origine neurologique sont considérées comme une classe à part entière, comprenant les dysphonies dysarthriques et les paralysies du larynx dues à un dysfonctionnement de son innervation. D'autres font état de dysphonies d'origine psychogène, comprenant notamment la dysphonie spasmodique.

2.3 Dysarthrie

La définition de la dysarthrie, la plus communément admise dans la littérature comme, par exemple, dans (Murdoch, 1998) ou (Lechevalier et al., 2008) est celle donnée par (Darley et al., 1969b) : *[...] a collective name for a group of speech disorders resulting from disturbances in muscular control over the speech mechanism due to damage of the central or peripheral nervous system. It designates problems in oral communication due to paralysis, weakness or incoordination of the speech musculature [...]*. Cette définition sera complétée, par la suite, par (Duffy, 2005), aboutissant à l'énoncé suivant : *[...] Dysarthria is a collective name for a group of neurologic speech disorders resulting from abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for control of the respiratory, phonatory, resonatory, articulatory, and prosodic aspects of speech production. The responsible pathophysiologic disturbances are due to central or peripheral nervous system abnormalities and most often reflect weakness ; spasticity ; incoordination ; involuntary movements ; or excessive, reduced, or variable muscle tone [...]*. Les éléments importants à retenir dans ces deux définitions sont (1) que la dysarthrie est un trouble de l'exécution ou réalisation motrice de la parole, la différenciant par conséquent, de l'apraxie, trouble de la programmation des commandes motrices, (2) d'ordre neurologique due à une lésion du système nerveux central ou périphérique (cf. section 1.1 du chapitre précédent pour une description du système nerveux et des mécanismes neurologiques), la différenciant des troubles de la parole en liant avec une malformation ou ablation des articulateurs (suite à un cancer par exemple) et (3) pouvant affecter les différents mécanismes et sous-mécanismes intervenant dans la production de la parole au niveau laryngé et supra-laryngé : respiration, phonation, résonance, articulation et/ou prosodie, ne la réduisant pas, de manière souvent erronée, à de simples troubles de l'articulation (voir la discussion autour de la définition de la dysarthrie dans (Pinto et Ghio, 2009)). Ce troisième point met l'accent

sur le fait qu'une dysphonie d'origine neurologique (comme définie plus haut) peut accompagner une dysarthrie. Dans ce sens, (Pinto et Ghio, 2009) propose la terminologie de neurodysarthrophonie pour définir le schéma complexe et souvent mal défini de la dysarthrie.

2.3.1 Déficiences motrices

Avant d'enchaîner sur la classification des dysarthries, il est important de rappeler ici très brièvement les grande lignes de la motricité volontaire (le lecteur pourra se référer au chapitre précédent pour une description détaillée) et d'apporter quelques informations en cas de déficience de cette dernière, que nous jugeons importante pour la suite de ce chapitre. La motricité volontaire met en jeu : (1) une commande et une régulation réalisée par les centres du système nerveux central (cerveau, cervelet, tronc cérébral et centres médullaires), (2) la transmission de cette commande par le biais des voies pyramidale et extra-pyramidale puis des nerfs périphériques (ces derniers formant le système nerveux périphérique), et (3) l'exécution de cette commande par les muscles, os et articulations.
En cas de lésion du système nerveux central et/ou périphérique, les déficiences motrices observées vont dépendre directement de la localisation de la ou des lésions. Dans le cadre de la dysarthrie, les cliniciens pourront faire référence à (liste non exhaustive) :

- un **syndrome pseudo-bulbaire**, conséquence d'une lésion bilatérale de la voie pyramidale au niveau cortical ie des motoneurones centraux, caractérisé, par une paralysie musculaire totale ou partielle (hypotonie) associée, de manière synchrone ou asynchrone, à des signes d'hypertonie musculaire de type spastique. La spasticité est un trouble moteur caractérisé par des contractures musculaires involontaires touchant les muscles volontaires des membres et du tronc, se traduisant par une rigidité (spastique) des mouvements ;

- un **syndrome bulbaire**, conséquence d'une lésion des motoneurones périphériques localisés au niveau du bulbe rachidien, directement en relation avec les nerfs crâniens. Il se traduit par une atrophie musculaire, une paralysie totale ou partielle (paralysie bulbaire) et une flaccidité musculaire ie une dégradation du tonus musculaire ;

- un **syndrome cérébelleux**, conséquence d'une lésion au niveau du cervelet, caractérisé notamment par des troubles de la coordination des gestes moteurs. Connus sous le terme d'ataxie, ces troubles induisent des erreurs dans l'organisation temporelle. Ils se traduisent par un retard d'initialisation du mouvement et un mouvement lent, des erreurs dans l'amplitude, la force et la direction des gestes qui perturbent évidemment les fonctions motrices visées ;

- un **syndrome extra-pyramidal**, conséquence d'une lésion des voies extra-pyramidales et des noyaux gris centraux (intervenant dans la boucle motrice de régulation cf. chapitre précédent, section 1.1.3), caractérisé par une rigidité musculaire (dite plastique) et des mouvements involontaires "anormaux" tels que tremblements,

chorée (mouvements brusques, imprévisibles, saccadés, rapides et amples), athétose (mouvements lents et ondulatoires), dystonie (contractions musculaires involontaires qui peuvent être brèves - spasmes -, prolongées - mouvements dystoniques - , ou soutenues - posture dystonique), ...

– un **syndrome parkinsonien** est un cas particulier de syndrome extra-pyramidal, touchant dans la majorité des cas le système dopaminergique au niveau des noyaux gris centraux. Il est caractérisé par la présence d'une akinésie (lenteur dans l'initialisation des mouvements) ou d'une bradykinésie (lenteur dans l'exécution des mouvements) dues au déficit en dopamine, souvent associées à la rigidité et tremblement extra-pyramidaux. La maladie de Parkinson est la principale cause du syndrome parkinsonien, mais ce n'est pas la seule.

2.3.2 Classification des dysarthries

Les études menées par Darley et ses collègues de la clinique Mayo (Etats-Unis) sur la classification des dysarthries (Darley et al., 1969a,b, 1975) font toujours référence dans la pratique clinique et dans la littérature. Cette classification a été établie à partir d'une étude réalisée auprès de 212 patients dysarthriques, répartis en 7 groupes homogènes représentant chacun une atteinte neurologique différente (d'après un diagnostic clinique complet). L'hypothèse inhérente à cette étude était la présence d'une association univoque possible entre une atteinte neurologique donnée et un ensemble d'altérations "observables auditivement" dans la production de la parole.

Cette étude comportait, en premier lieu, une évaluation perceptive réalisée par un jury de 3 experts cliniciens, d'un échantillon de parole lue de 30s produit par chacun de ces patients. Cette évaluation perceptive reposait sur 38 critères, fournis dans le tableau 2.1 (Auzou, 2007b) et répartis selon 7 axes majeurs : la respiration, l'articulation, la prosodie, la qualité de la voix [4], la hauteur de la voix, l'intensité de la voix, autres caractéristiques (comprenant l'évaluation du niveau d'intelligibilité et du caractère *bizarre* de la production). Chacun de ces critères devait être côté par le jury d'experts suivant une échelle à 7 points : 1 représentant la normalité, 7 le degré le plus sévère d'altération. Une première analyse des résultats de cette évaluation perceptive a permis de mettre en évidence les caractéristiques les plus déviantes (associées à des scores moyens supérieurs ou égaux à 1,5) pour chacune des atteintes neurologiques ciblées. Néanmoins, cette analyse a également démontré la présence de certaines caractéristiques dans plusieurs atteintes, ne permettant donc pas leur discrimination.

Pour pallier ce problème, une seconde analyse des résultats, toujours menée individuellement pour chacune des atteintes neurologiques, a porté sur la recherche de **clusters** basés sur la concomitance des caractéristiques les plus déviantes. L'hypothèse sous-jacente à la construction de ces clusters est qu'un ensemble de paramètres pertur-

4. La hauteur et l'intensité vocale sont considérées comme des axes à part entière. La qualité de la voix sera donc basée, ici, principalement sur la perception du timbre vocal et sur des caractéristiques de nasalité.

bés simultanément, mis en relation avec l'atteinte neurologique, reflèterait un processus physiopathologique particulier. Ainsi, 8 clusters physiopathologiques ont pu être définis : imprécision articulatoire, excès prosodique, insuffisance prosodique, incompétence résonance et articulatoire, sténose phonatoire [5], incompétence phonatoire et incompétence résonance. Chaque atteinte s'est vue assigner un sous-ensemble de ces 8 clusters, chaque sous-ensemble étant unique malgré le fait que certains clusters pouvaient être communs à plusieurs atteintes. Finalement, ces études ont conduit à l'élaboration d'une organisation des dysarthries en 6 classes (Darley et al., 1975), dénommée par certains auteurs *the Mayo clinic system*, élargie à 8 classes en 2005 par Duffy (également membre de la clinique Mayo) (Duffy, 1995). Nous allons passer chacune de ces classes en revue en spécifiant les atteintes neurologiques affiliées ainsi que les phénomènes d'altération de la parole qui les caractérisent.

Articulation	**Respiration**
imprécision des consonnes	inspiration-expiration forcées
dégradations articulatoires	inspiration audible
allongement des phonèmes	bruit en fin d'expiration
répétition de phonèmes	
distorsion des voyelles	
Prosodie	**Qualité de la voix**
débit	voix rauque
débit variable	voix humide
augmentation du débit (global)	voix soufflée (continu)
augmentation du débit (segment)	voix soufflée (intermittent)
diminution de l'accentuation	voix forcée
accentuation excessive	arrêts vocaux
phrases courtes	hypernasalité
allongement des pauses	hyponasalité
accélérations paroxystiques	émission nasale
silences inappropriés	
Hauteur vocale	**Intensité vocale**
niveau global	niveau global
rupture de la hauteur	mono-intensité
monotonie	variation excessive d'intensité
tremblement vocal	décroissance d'intensité
	instabilité d'intensité
Autres caractéristiques	
intelligibilité	
bizarrerie	

TABLE 2.1 – *Liste des 38 critères perceptifs utilisés pour l'élaboration de la classification des dysarthries proposée dans (Darley et al., 1969a,b, 1975) suivant un modèle physiopathologique.*

5. rétrécissement de l'ouverture laryngée.

Dysarthrie spastique

La dysarthrie spastique décrit les altérations de la parole dues à une lésion bilatérale des moto-neurones centraux à la base d'un syndrome pseudo-bulbaire. Ces altérations, d'après (Darley et al., 1975), se traduisent par une spasticité, une diminution de la force et de l'amplitude des mouvements ainsi qu'une lenteur. Elles peuvent toucher tous les mécanismes de production de la parole, avec des degrés variables de sévérité d'un individu à l'autre. La parole est lente et laborieuse, démontrant des efforts importants de la part du locuteur dysarthrique dans sa phase de production.

Lors de l'analyse de l'évaluation perceptive (Darley et al., 1969a,b), les 14 paramètres les plus déviants de la dysarthrie spastique mis en avant par Darley et ses collègues sont (par ordre décroissant des scores attribués par le jury) : imprécision des consonnes, monotonie, diminution de l'accentuation, voix rauque, mono-intensité, niveau global de la hauteur (bas), débit lent, hyper-nasalité, voix forcée, phrases courtes, distorsion des voyelles, rupture de la hauteur, voix soufflée, accentuation excessive. Finalement, la dysarthrie spastique est associée à 3 clusters principaux représentant l'incompétence résonance et articulatoire, la sténose phonatoire et l'insuffisance prosodique, auxquels se rajoute une version simplifiée du cluster d'excès prosodique réduite aux seuls paramètres d'accentuation excessive et de débit lent.

Plus tard, Enderby complète cette description par une observation des mouvements oro-moteurs de patients souffrant de dysarthrie spastique et par une évaluation de leur altération (faiblesse observée dans les muscles de la langue et des lèvres) en identifiant les caractéristiques dominantes suivantes (Enderby, 1986) : mouvement réduit de la langue et des lèvres lors de production de parole, débit lent, faiblesse dans la phonation et l'intonation, intelligibilité faible lors de tâches conversationnelles et de description, limitation dans les mouvements alternatifs de la langue et dans le maintien du palais en position haute, hypernasalité, manque de contrôle du volume.

Dysarthrie ataxique

La dysarthrie ataxique (ou dysarthrie cérébelleuse) décrit les altérations de la parole dues à un syndrome cérébelleux. Comme évoqué plus haut, ces altérations, localisées au niveau du cervelet, touchent principalement la coordination des gestes. En production de parole, cela se traduit par des gestes respiratoires, phonatoires et articulatoires imprécis et non coordonnés dus à des erreurs de temporalité (délai d'initialisation du mouvement), de force, d'amplitude, de direction et de lenteur de l'ensemble des mouvements simples qui les composent. Certains patients peuvent présenter une altération de la prosodie, une mauvaise gestion de l'intensité vocale (parole explosive), un tremblement et un trouble de la résonance nasale (Auzou et Rolland-Monnoury, 2006).

Les 10 paramètres les plus déviants mis en avant par Darley et ses collègues sont ici : imprécision des consonnes, accentuation excessive, dégradation articulatoire, dis-

torsion des voyelles, voix rauque, allongement des phonèmes, allongement des pauses, monotonie, mono-intensité, débit lent, représentés sous la forme de 3 clusters que sont l'imprécision articulatoire, l'excès prosodique et l'insuffisance phonatoire et prosodique.

Dysarthrie flasque

La dysarthrie flasque décrit les altérations de la parole dues (1) soit à des atteintes des moto-neurones périphériques situés au niveau de la moelle épinière ou du bulbe rachidien (syndrome bulbaire), (2) soit de la jonction neuro-musculaire ou (3) soit des muscles impliqués dans les mécanismes de parole. Ces atteintes, en lien direct avec les nerfs crâniens V, VII, IX, X, XI et XII et les nerf spinaux, innervant les muscles intervenant dans les différents mécanismes de production de la parole, peuvent avoir une incidence sur chacun d'eux. Le type d'altérations dépendra de la localisation de la ou des lésions, et par conséquent des nerfs et muscles affectés. La faiblesse musculaire ainsi que la réduction du tonus musculaire, impactant la vitesse, l'amplitude et la précision des mouvements en seront la principale cause.

9 paramètres les plus déviants ont été mis en avant par Darley et ses collègues : hypernasalité, imprécision des consonnes, voix soufflée en continu, monotonie, émission nasale, inspiration audible, voix rauque, phrases courtes et mono-intensité, représentés sous la forme de 3 clusters que sont l'incompétence phonatoire, l'incompétence de la résonance et l'insuffisance phonatoire et prosodique. Similairement à l'observation des dysarthries spastiques, Enderby met en évidence ici (Enderby, 1986) une faiblesse dans la fermeture des lèvres, une posture anormale des lèvres au repos et en position ouverte, une limitation dans les mouvements alternatifs de la langue, une réduction du temps de phonation, une intelligibilité faible lors de tâches de répétition et de description.

Dysarthrie hypokinétique

La dysarthrie hypokinétique décrit les altérations de la parole dues à une atteinte localisée des noyaux gris centraux. Elle se réfère typiquement aux patients dysarthriques souffrant de la maladie de Parkinson.

La maladie de Parkinson est la forme la plus pure des syndromes parkinsoniens. Il s'agit d'une maladie neurologique, dégénérative progressive, qui après la maladie d'Alzheimer, est considérée comme la plus fréquente, justifiant le nombre volumineux de travaux de recherche la concernant. Elle est associée à une perte sélective des neurones dopaminergiques dans la substance noire (pars compacta), un des noyaux gris centraux, intervenant dans la boucle motrice de régulation. Les autres maladies assimilées à ce syndrome présentent des lésions plus diffuses. Ainsi, elles peuvent présenter des signes dysarthriques très similaires à ceux de la maladie de Parkinson, comme par exemple, la paralysie supra-nucléaire progressive, mais également présenter d'autres signes (empruntés notamment à la dysarthrie ataxique ou spastique), leur conférant une dysarthrie de type mixte (voir les détails sur la dysarthrie mixte ci-dessous).

Les signes moteurs de la maladie de Parkinson sont l'akinésie (lenteur dans l'initialisation des mouvements) et la bradykinésie (lenteur dans l'exécution des mouvements), la rigidité et la présence de tremblement. Leur incidence peut s'observer sur l'ensemble des mécanismes de production de parole. Ainsi, *[...] les troubles de la production de parole au cours de la maladie de Parkinson constituent un des marqueurs cliniques représentatifs de l'évolution du handicap moteur et cognitif [...]* (Viallet et Teston, 2007). Ces troubles se caractérisent principalement par une réduction des mouvements articulatoires et une diminution de la modulation prosodique de la parole, qualifiée de monotone, d'où le terme de dysarthrie hypokinétique. Néanmoins, différentes études (Logeman et al., 1978; Ho et al., 1998), rapportées dans (Viallet et Teston, 2007), ont montré que la dysphonie est le premier signe "observable" de la maladie de Parkinson, rattrapée au cours de la progression de la maladie par les troubles articulatoires et de débit, principales causes de la perte d'intelligibilité chez les patients parkinsoniens.

Les 10 paramètres les plus déviants mis en avant par Darley et ses collègues sont ici : monotonie, diminution de l'accentuation, mono-intensité, imprécision des consonnes, silences inappropriés, accélérations paroxystiques, voix rauque, voix soufflée continue, niveau global, débit variable. Typiquement, l'analyse en clusters ne révèle qu'un seul cluster - insuffisance prosodique. Rassemblant uniquement, pour la majorité des lésions neurologiques concernées, les paramètres de monotonie, mono-intensité, diminution de l'accentuation et phrases courtes, ce cluster a été exceptionnellement étendu pour caractériser la maladie de Parkinson et tenir compte d'autres paramètres tout aussi importants et typiques de celle-ci et en lien avec l'insuffisance prosodique tels que le débit variable, les accélérations paroxystiques et l'imprécision des consonnes.

Dysarthrie hyperkinétique

La dysarthrie hyperkinétique décrit les altérations dues à un syndrome extra-pyramidal, caractérisées principalement par des mouvements involontaires anormaux. Comme évoqué plus haut, ces mouvements involontaires peuvent comporter des formes (chorée, athétose, dystonie, ...) et des localisations très diverses selon l'atteinte des noyaux gris. Cela explique la grande hétérogénéité observée dans les altérations de parole produite par les patients atteints de ce syndrome. (Darley et al., 1969a) distingue deux types de dysarthrie hyperkinétique :

- la dysarthrie hyperkinétique rapide, incluant notamment les mouvements choréïques, pour laquelle les paramètres les plus déviants sont au nombre de 16 : imprécision des consonnes, allongement des pauses, débit variable, monotonie, voix rauque, silences inappropriés, distorsion des voyelles, variation excessive d'intensité, allongement des phonèmes, mono-intensité, phrases courtes, dégradation articulatoire, accentuation excessive, hyper-nasalité, diminution de l'accentuation, voix forcée. Face à la diversité de ces paramètres, 6 clusters ont été identifiés pour représenter ce premier type : l'imprécision articulatoire, l'excès prosodique, l'insuffisance prosodique, l'incompétence résonance et articulatoire, la sténose phonatoire et finalement l'incompétence résonance ;

– la dysarthrie hyperkinétique lente, incluant notamment les mouvements dystoniques et athétosiques, pour laquelle les paramètres les plus déviants sont au nombre de 15 : imprécision des consonnes, distorsion des voyelles, voix rauque, dégradation articulatoire, monotonie, mono-intensité, silences inappropriés, allongement des pauses, allongement des phonèmes, variation excessive d'intensité, diminution de l'accentuation, arrêts vocaux, débit lent. Ici, les clusters identifiés sont communs avec 4 de ceux de la dysarthrie cinétique rapide, à savoir l'imprécision articulatoire, l'excès prosodique, l'insuffisance prosodique et la sténose phonatoire.

Dysarthrie mixte

La dysarthrie mixte n'est pas une forme pure de dysarthrie, comme celles décrites jusqu'à présent. Comme son nom l'indique, elle combine plusieurs formes de dysarthrie, généralement de deux à trois. Dans une étude menée par Duffy (Duffy, 1995) sur 300 cas de dysarthrie mixte, les combinaisons les plus fréquemment trouvées sont : dysarthrie flasque/spastique (42%), dysarthrie ataxique/spastique (23%), hypokinétique/spastique (7%), dysarthrie ataxique/flasque/spastique (6%) et dysarthrie hypokinétique/hyperkinétique (3%).

L'évaluation perceptive de Darley et ses collègues (Darley et al., 1969a) a porté, dans le cas des dysarthries mixtes, exclusivement sur des patients atteints de Sclérose Latérale Amyotrophique (SLA), également appelée maladie de Charcot. La SLA est due à une atteinte, à la fois des moto-neurones centraux et périphériques. Elle se caractérise, par conséquent, dans la grande majorité des cas, par une dysarthrie flasque/spastique. Dans ce cas précis, 18 paramètres les plus déviants ont été identifiés : imprécision des consonnes$^{(f,s)}$, hypernasalité$^{(f,s)}$, voix rauque$^{(f,s)}$, débit lent$^{(s)}$, monotonie$^{(s)}$, phrases courtes$^{(f,s)}$, distorsion des voyelles$^{(s)}$, niveau global de la hauteur (basse)$^{(s)}$, mono-intensité$^{(f,s)}$, accentuation excessive$^{(s)}$, allongement des pauses$^{(m)}$, diminution de l'accentuation$^{(s)}$, allongement des phonèmes$^{(m)}$, voix forcée$^{(s)}$, voix soufflée (continue)$^{(f,s)}$, inspiration audible$^{(f)}$, silences inappropriés$^{(m)}$ et émission nasale$^{(f)}$. Parmi ces paramètres, 6 sont présents dans la dysarthrie flasque et spastique (identifiables par les symboles (f,s) en indice), 2 sont présents uniquement dans la dysarthrie flasque (symbole (f)), 7 sont présents uniquement dans la dysarthrie spastique (symbole (s)) et 3 n'apparaissent ni dans l'une, ni dans l'autre ((m)). L'analyse en clusters a révélé la nécessité de 6 clusters sur les 8 disponibles pour représenter la variabilité des altérations de la dysarthrie mixte. 4 clusters sont précisément ceux de la dysarthrie spastique (incompétence résonance et articulatoire, sténose phonatoire, insuffisance prosodique et excès prosodique), les 2 autres emprunts à la dysarthrie flasque (incompétence phonatoire et incompétence résonance).

La dysarthrie mixte pourra se manifester dans d'autres maladies telles que la sclérose en plaque (SEP) à prédominance néanmoins ataxique pour laquelle des évaluations perceptives ont montré des troubles vocaux (hauteur et intensité), une voix rauque et des troubles articulatoires ou des traumatismes crâniens mêlant soit une dysarthrie

ataxique/spastique, soit une dysarthrie spastique/ataxique/flasque (d'après différentes études rapportées dans (Auzou et Rolland-Monnoury, 2006)).

Dysarthrie par atteinte unilatérale des neurones centraux

Cette dysarthrie, introduite par (Duffy, 1995) pour compléter la classification de Darley et ses collègues, décrit les altérations dues à une lésion uni-latérale des motoneurones centraux (à différencier de la lésion bilatérale de la dysarthrie spastique). Cette dysarthrie, contrairement aux autres types de dysarthrie, est généralement peu sévère et temporaire. Elle est caractérisée principalement par des troubles articulatoires (imprécision des consonnes), qui peuvent s'accompagner d'une voix rauque, et encore plus rarement d'altérations prosodiques touchant principalement le débit.

2.4 L'évaluation, au coeur des bilans cliniques de la dysphonie et de la dysarthrie

Dans le contexte des troubles de la voix et/ou de la parole, l'examen clinique comportera généralement un examen physique du larynx (troubles de la phonation) et/ou du conduit vocal (troubles de la parole). Il s'accompagnera d'un interrogatoire du patient afin de reconstituer l'histoire de la maladie (anamnèse) : antécédents médicaux, prises de médicaments, description et périodicité des troubles, ... et d'une évaluation qualitative, voire quantitative de la voix et/ou de la parole et de leurs altérations.

Dans (Auzou et Rolland-Monnoury, 2006), il est souligné que le bilan clinique de la dysarthrie doit permettre aux cliniciens de répondre aux questions suivantes :
(1) Quelle est la sévérité de la dysarthrie ?
(2) Quelles sont les principales anomalies perceptives qui permettent de la décrire (et donc de communiquer entre les différents thérapeutes) ?
(3) Quels sont les dysfonctionnements des effecteurs qui rendent compte des perturbations de la parole ?
(4) Quelle perception le patient a-t-il de sa parole et quelle est sa plainte ?

Il est également précisé que ce bilan doit ainsi conduire à la description des anomalies, à l'obtention de pistes physiopathologiques pour guider la rééducation et fournir des données quantitatives pour le suivi thérapeutique du malade.
Même si nous n'avons pas trouvé de questionnement similaire concernant la dysphonie en tant que telle dans la littérature (hors dysarthrie), il nous semble que les questions concernant la sévérité de la dysphonie, les anomalies perceptives qui permettent de la décrire, les dysfonctionnements des effecteurs qui rendent compte des altérations de la voix, dès lors que ces derniers sont réduits à l'observation du seul larynx voire du système pneumo-phonatoire, et la perception du patient de sa voix et l'énoncé de sa plainte s'appliquent tout à fait dans le cadre d'un bilan clinique de la dysphonie.

Comme nous le détaillerons dans cette section, l'évaluation qualitative et quantitative de la voix et/ou de la parole pathologique d'un patient peut avoir plusieurs objectifs :

- identifier le caractère pathologique ou non de la voix et/ou de la parole ;
- aider à poser un diagnostic, notamment neurologique, lorsqu'une dysarthrie ou une dysphonie d'origine neurologique est identifiée et qu'elle constitue le seul signe précoce de la pathologie sous-jacente. Le patient pourra, dès lors, être orienté vers un neurologue pour la prise en charge thérapeutique de l'origine de la dysphonie et/ou dysarthrie (sans que celle-ci ne soit mise de côté évidemment). Dans ce cas précis, la caractérisation des altérations de la voix et/ou de la parole, suite à une évaluation perceptive inspirée de la classification proposée par (Darley et al., 1969a), pourra donner au neurologue de premières pistes intéressantes d'investigation sur la localisation de la lésion.
- aider à poser un diagnostic dans le cas d'une dysphonie d'origine autre que neurologique, afin d'apprécier la présence d'un comportement d'effort, de fatigue vocale, de mécanismes de serrage (voix rauque, forcée, ...), d'atténuation de la voix (voix sourde, soufflée), ... et dans le cas d'une dysarthrie à caractériser les altérations de la voix et/ou de la parole ;
- aider à définir les objectifs de la prise en charge thérapeutique suivant l'évaluation de la sévérité de la dysphonie et de la dysarthrie et l'évaluation des altérations observées ;
- aider à mesurer les résultats d'une thérapie en comparant des évaluations initiale, intermédiaires et finale ;
- aider le patient à prendre conscience de tous les troubles qui l'affectent (et non pas uniquement ceux qui l'ont amené à consulter) et de leur atténuation/disparition en cours et après thérapie ;
- faire progresser la connaissance sur les pathologies et permettre leur comparaison d'un point de vue fondamental.

L'évaluation qualitative et/ou quantitative de la voix et/ou de la parole fait référence à un ensemble de méthodologies, toutes ou partiellement utilisées en pratique clinique, mais également dans les travaux de recherche. Chacune d'elles sera détaillée ci-dessous.

2.4.1 Evaluation perceptive

La méthode d'évaluation perceptive consiste à juger la qualité de la voix et/ou de la parole en se basant sur le système de perception auditive (évaluation "à l'oreille"). La voix et la parole ont pour premier sens d'être perçues. *Aussi, quoi de plus naturel que l'écoute pour percevoir le caractère pathologique d'une voix ou de la parole et leurs anomalies ?* Le patient (ou son entourage) est la première personne à utiliser ce système pour déceler un changement dans la voix/parole, ce qui, en présence d'une gêne, l'amènera à consulter. Il est communément admis que les spécialistes de la voix et de la parole utilisent leurs oreilles comme outil d'évaluation principal des dysphonies et/ou des dysarthries et que cette méthode d'évaluation est la plus répandue dans la pratique clinique. Outre son caractère naturel qui la rend simple à mettre en place, cette méthode

offre également l'avantage d'être très peu coûteuse et robuste. En effet, (Kent, 1996) souligne que le système auditif est remarquable si l'on considère sa capacité à comprendre la parole dégradée par des interruptions temporelles, des substitutions, des omissions ou sa capacité à évaluer les troubles de la voix et/ou de la parole sur un large spectre de degré de sévérité. Nous verrons par la suite que ce terme de "robuste" ne fait pas l'unanimité.

L'évaluation perceptive peut intervenir à différents niveaux comme nous le verrons dans les sous-sections suivantes.

Evaluation perceptive analytique

Comme nous l'avons vu avec les études menées par Darley et ses collègues (Darley et al., 1969a,b), une évaluation perceptive analytique peut permettre de qualifier les altérations de la voix et de la parole chez les patients dysarthriques et de les quantifier. Dans ce cadre, elle repose généralement sur un ensemble des paramètres à "observer" auditivement (qualification des altérations) accompagné d'une échelle d'évaluation (quantification des altérations). L'ensemble de paramètres est généralement choisi, dans le cadre de la dysarthrie, de façon à caractériser les différents mécanismes liés à la production de parole (respiration, phonation, articulation, résonance et prosodie). Concernant l'échelle d'évaluation, (Revis, 2004) distingue les échelles bipolaires - utilisées pour spécifier la présence / absence d'un paramètre ou pour opposer deux dimensions contraires (stable/instable) - les échelles de classes (également connues sous la dénomination d'échelles à points d'intervalles équivalents) - la présence d'un paramètre est quantifiée par plusieurs niveaux pré-définis -, les échelles ouvertes (également connues sous la dénomination d'estimation directe de magnitude) - la quantification, sous format numérique, est laissée au libre arbitre de l'expert sur une échelle allant de 0 à l'infini - et les échelles visuelles numériques - similaires aux échelles ouvertes au niveau de la quantification libre, excepté que l'échelle est ici bornée et non plus infinie.

Pour rappel, les études de (Darley et al., 1969a,b) reposaient sur 38 paramètres évalués sur une échelle linéaire à 7 points. De nombreuses études ont été rapportées dans la littérature sur l'évaluation perceptive des dysarthries en vue de mieux les caractériser et permettre de les catégoriser en vue d'une aide au diagnostic différentiel en pratique clinique. Ces études montrent une grande diversité dans le nombre et le choix des paramètres utilisés, très souvent orientés par la population de patients étudiée, dans le choix de l'échelle d'évaluation (celle-ci peut même varier au sein d'une même étude suivant le paramètre ciblé) et dans le choix des populations de patients étudiés. Concernant la dysphonie, la même constatation peut être faite. Dans (Revis, 2004), 4 tentatives de standardisation des méthodes d'évaluation perceptive de la dysphonie sont comparées (d'après une étude de revue menée dans (De Bodt et al., 1996)) : le Buffalo Voice Profile System (BVP) (Wilson, 1987), le Hammarberg Scheme (Hammarberg et al., 1980), le Vocal Profile Analysis Scheme (VPAS) (Laver, 1980) et l'échelle GRBAS (Hirano, 1981). Ces 4 méthodes d'évaluation présentent effectivement un nombre et un choix de paramètres différents ; certains "observant" uniquement les altérations au niveau laryngé

(Hammarberg Scheme et GRBAS), d'autres ajoutant le niveau supra-laryngé (BVP et VPAS) mais également le comportement vocal (BVP). Les divergences portent également sur l'échelle d'évaluation, toutes de type "échelle de classe" mais comportant des niveaux différents : 4 niveaux pour l'échelle GRBAS, 5 niveaux pour le BVP et le Hammarberg Scheme et 6 niveaux pour le VPAS. Finalement, ces méthodes d'évaluation se distinguent sur le type de tâches utilisé lors des passations : lecture seule (Hammarberg Scheme), lecture et parole spontanée seule (GRBAS) ou associée à d'autres tâches (comptage et voyelle tenue pour BVP) et finalement, lecture et parole spontanée pour VPAS.

Evaluation de la sévérité

Si l'identification des altérations dans la production de la parole est nécessaire pour poser un diagnostic en pratique clinique, mesurer la sévérité d'une dysphonie ou d'une dysarthrie est une étape cruciale pour définir les objectifs d'une prise en charge thérapeutique ou pour évaluer ses bienfaits en cours ou après thérapie. Dans ce contexte, la sévérité est généralement définie suivant différents paramètres que sont l'**intelligibilité**, la **compréhensibilité**, la **compréhension de l'auditeur**, l'**efficacité** et la **sévérité perceptive** (Auzou, 2007c; Hustad, 2008; Lowit et Kent, 2010). Les quatre premiers facteurs sont directement liés à la faculté d'un locuteur à transmettre un message oral à un auditoire. Dans le cas de troubles de la parole, cette faculté peut être profondément altérée, conséquence d'une pathologie avérée, pouvant conduire à un handicap social sévère entraînant difficultés professionnelles et perte de relations avec autrui pouvant aller jusqu'à l'isolement (McAuliffe et al., 2010).

Inspirées des domaines de l'acoustique et des télécommunications (cf. (Fontan, 2012) pour un état de l'art des tests d'intelligibilité de la parole dans ces domaines), l'**intelligibilité** se définit par la précision avec laquelle un message émis par un locuteur est "décodé" par un auditeur. La notion de "décodage" à laquelle nous faisons mention vient du fait que les mesures d'intelligibilité reposent pour une grande majorité sur le comptage d'unités de parole émises par le locuteur et correctement reconnues par l'auditeur. Ces mesures diffèrent suivant les études par (le lecteur pourra se référer aux revues de la littérature suivantes : (Barreto et Ortiz, 2008; Hustad, 2008; Fontan, 2012)) :

- la granularité des unités de parole considérées : phonèmes, syllabes, mots, phrases, conversations ; chaque type d'unités présentant un cadre d'observation et d'analyse bien spécifique (analyse quantitative des déviances au niveau du phonème vs analyse prosodique au niveau de la phrase ou des conversations) et, par conséquent, des objectifs et limites différents comme détaillé dans (Lowit et Kent, 2010) ;
- la tâche à réaliser par l'auditeur : transcription orthographique des unités de parole reçues par l'auditeur, sélection d'une transcription ou d'une image en relation avec le message reconnu parmi une liste d'items fermés, évaluation perceptive sur des échelles représentant le taux d'intelligibilité perçu ;
- le système de cotation : binaire, échelle numérique discrète, échelle analogique, taux entre les unités de parole correctement "décodées" par l'auditeur sur le nombre d'unités produites.

Pour exemple, (Auzou, 2007d) décrit trois tests d'intelligibilité disponibles pour le français, partie intégrante de la Batterie d'Evaluation Clinique de la Dysarthrie (BECD), développés ou adaptés de travaux existants par le même auteur (Auzou et Rolland-Monnoury, 2006). Ces tests, spécifiquement adaptés pour le français reprennent quelques uns des items cités ci-dessus : (1) l'**item perceptif** basé sur une estimation globale de l'intelligibilité par l'auditeur lors de conversations, selon une cotation sur une échelle à 5 points accompagnés d'une interprétation (0 - pas d'anomalie à 4 - anomalie sévère ou quasi permanente), (2) **le score d'intelligibilité**, adapté pour la langue française d'épreuves issues de la batterie de tests "Frenchay Dysarthria Assessment" (Enderby, 1983) et reposant sur la somme de scores d'intelligibilité (entre 0 et 8) obtenus à partir de la lecture de mots, phrases et parole spontanée (24 - parole normale, entre 18 et 23 - dysarthrie légère, entre 12 et 17 - dysarthrie modérée, entre 7 et 11 - dysarthrie sévère, en dessous de 6 - dysarthrie très sévère atteignant l'inintelligibilité), (3) le **test phonétique d'intelligibilité**, adapté des travaux de (Kent et al., 1989) pour l'anglais et de (Gentil, 1992) pour le français, portant sur un score d'intelligibilité couplé à une identification des troubles articulatoires qui s'y rapportent ; ce test repose sur la lecture de 13 séries de 4 mots bisyllabiques, présentant un ou deux contrastes phonétiques au sein de chaque quadruplet (ex. "ses oui", "si oui", "série", "scierie"), et la sélection par l'auditeur du mot perçu parmi l'ensemble fermé de mots constituant le quadruplet.

La **compréhensibilité** est le plus souvent définie dans la littérature comme une forme particulière d'intelligibilité pour laquelle des informations contextuelles sont prises en compte au moment de son évaluation. (Yorkston et al., 1996) utilise le terme "contextual intelligibility". Ces informations contextuelles peuvent être liées à la connaissance du patient, de sa pathologie et des troubles de parole qui en découlent, à des indices sémantiques, syntaxiques et orthographiques accompagnant le message énoncé par le locuteur ou encore, à des indices visuels (mimiques, gestes) dès lors que le locuteur est en face de son auditeur. Cette définition explique le fait que la compréhensibilité puisse ne pas être évoquée dans certaines publications, car implicitement confondue à l'intelligibilité, et qu'elle soit par conséquent évaluée en utilisant les mêmes mesures citées précédemment. Néanmoins, les études menées, différenciant intelligibilité et compréhensibilité dans le protocole de tests, montrent une réelle sous-estimation du degré de sévérité des patients dans le deuxième cas et que, par conséquent, la présence d'informations contextuelles influence le jugement des auditeurs. Le cas le plus simple est la connaissance du message énoncé par le locuteur, pouvant influencer l'évaluation de l'auditeur. Cette situation se présente par exemple dans le cas de pratiques cliniques répétées où le patient doit lire une liste de mots parmi des listes utilisées régulièrement par le clinicien.

L'évaluation de la **compréhension de l'auditeur** a pour vocation de se rapprocher d'une situation réelle de communication entre un patient et autrui, ie une situation où le message émis par le locuteur atteint son objectif d'interaction. L'évaluation de l'intelligibilité/compréhensibilité, telle que décrite précédemment, ne tient pas compte de cette dimension. En effet, un mot mal prononcé et, par conséquent, mal transcrit au cours d'un test d'évaluation ne signifie pas qu'il sera mal compris par l'auditeur en

situation de communication dans laquelle le contexte peut fortement aider à la compréhension. Il est bien connu que l'être humain a des capacités d'adaptation et de compensation dans des conditions adverses (test du "cocktail party" par exemple). A l'inverse, certaines altérations de la parole peuvent venir perturber la compréhension de l'auditeur, comme par exemple une parole monotone et à faible intensité, engendrant des difficultés dans la segmentation de la parole nécessaire au processus de compréhension. Néanmoins, ces difficultés pourraient être masquées lors de la prononciation de simples mots isolés lors d'un test d'intelligibilité. L'étude menée par (Hustad, 2008) sur la relation entre les scores d'intelligibilité et de compréhension de l'auditeur montre une corrélation faible chez des groupes de patients dysarthriques homogènes en terme de sévérité. Les conclusions de celle-ci, malgré les limites soulevées par ses auteurs (même maladie pour tous les patients, simplicité des tests de compréhension comparés à une situation de communication réelle) ainsi que par (Fontan, 2012) qui remet en question la validité statistique des résultats présentés, s'avèrent tout de même intéressantes. Elles mettent en avant le fait que l'évaluation de l'intelligibilité et de la compréhension de l'auditeur sont nécessaires et complémentaires à un bilan clinique pré- et/ou post-traitements tout en ne ciblant pas les mêmes objectifs : la première doit être utilisée pour évaluer la capacité du patient à produire un signal acoustique, la seconde sa capacité à faire passer un message. Néanmoins, cette distinction est loin d'être clairement établie dans la littérature comme souligné dans (Fontan, 2012) ; le terme de compréhension venant très souvent appuyer la définition d'intelligibilité (*"une intelligibilité faible associée à une parole difficilement compréhensible"*).

Cela étant entendu, la mise en place de tests permettant l'évaluation de la compréhension de l'auditeur est une tâche très complexe, due notamment aux différents processus perceptifs et cognitifs mis en jeu dans la compréhension d'une séquence de parole. Les recherches réalisées dans ce cadre sont encore très exploratoires. La littérature fait état de quelques travaux portant sur la perception de la parole dysarthrique par des auditeurs, travaux qui tentent de répondre aux question suivantes : "comment et pourquoi un auditeur échoue (ou réussit) dans sa compréhension d'une séquence de parole émise par un locuteur[6]". La réponse à ces questions pourrait permettre la prise en compte du facteur "auditeur" dans la pratique clinique, son intégration dans le traitement thérapeutique et la rééducation des patients en vue d'améliorer leur capacité à communiquer avec autrui (cf. (McAuliffe et al., 2010)). Inspirés pour une grande part du champ de recherche dédié à la perception de la parole dégradée (bruit, accent étranger, ...), ces travaux portent notamment sur l'analyse des stratégies mises en place par les auditeurs lorsqu'ils sont confrontés à de la parole dysarthrique, les erreurs de transcription qu'ils commettent (Hustad, 2006), les effets de la connaissance du contexte linguistique en lien avec la séquence de parole émise (première lettre du mot prononcé, informations syntaxiques ou sémantiques des phrases émises, ...)(Hanson et al., 2004), les effets de l'exposition préalable des auditeurs à la parole dysarthrique (degré de familiarité de l'auditeur) (Liss et al., 2002; J. D'Innocenzo, 2006), ou encore les effets d'un apprentissage sur les auditeurs (Borrie et al., 2012a,b).

6. "[...] how and why a communication partner fails to comprehend what is said by the speaker [...](McAuliffe et al., 2010)."

Les rares tests d'évaluation de la compréhension de l'auditeur proposés dans la littérature reposent sur la capacité de ce dernier à répondre à des questions ayant trait au message produit par le patient dysarthrique (Beukelman et Yorkston, 1979; Hustad, 2008) ou sur sa capacité à produire un résumé du discours narratif du patient. Ces tests sont très loin de représenter une situation réelle de communication puisqu'ils n'impliquent, en premier lieu, aucune interaction entre le locuteur et l'auditeur, ni d'implication réelle ou attente de leur part. Seuls les travaux de (Fontan, 2012) semblent, à l'heure actuelle, aller dans cette direction, même si la solution proposée reste encore très "basique" (toujours au regard de la complexité de la tâche). En effet, l'évaluation de la compréhension de l'auditeur est mesurée en fonction d'actions qu'il doit réaliser et qui répondent à des ordres verbaux fournis par le locuteur. Lors du test, le patient et l'auditeur partagent le "même environnement" comportant des objets sur lesquels va porter la communication. Seule la parole est utilisée comme media de communication (pas de geste, ni mimique). L'environnement commun est dématérialisé par le logiciel EloKanz, utilisé par le locuteur et l'auditeur. Le logiciel comporte un ensemble d'images disséminées sur l'écran. Il permet au locuteur de dicter une série d'actions à l'auditeur, comme par exemple : "Mettez le chat sous la voiture", et à ce dernier de les exécuter.

L'**efficacité** se mesure par la quantité de messages intelligibles transmis par unité de temps. Plus elle sera faible, plus elle dénotera une altération du débit de parole ou de l'intelligibilité.

Dans certains cas, l'étude de tous ces facteurs ne révèle pas de pathologie particulière ; le patient est intelligible, compréhensible et sa parole est efficace. Et pourtant, sa parole a perdu de son caractère naturel ou sa voix ne correspond plus à l'image sociale attendue (le lecteur pourra se référer à (Barkat-Defradas et al., 2012) pour une étude sur la dimension esthétique des voix normales et dysphoniques). En d'autres termes, la parole ou la voix apparaissent comme déviantes. La **sévérité perceptive** est le facteur permettant de mesurer le degré de déviance de la voix et/ou de la parole. Elle est généralement associée à des critères globaux évaluant, par exemple, le caractère bizarre de la parole (Darley et al., 1969b), le caractère naturel de la parole (Auzou et Rolland-Monnoury, 2006) ou, tout simplement, le degré de sévérité global des altérations présentes dans la voix ou la parole d'un patient. Ce degré de sévérité se traduira par exemple par le critère G de l'échelle d'évaluation perceptive GRBAS (Hirano, 1981) dans le cas des voix dysphoniques ou encore par l'item "parole" dans l'échelle UPDRS (Unified Parkinson's Disease Rating Scale) pour l'évaluation des patients atteints de la maladie de Parkinson.
L'évaluation de la sévérité perceptive est couramment utilisée en pratique clinique en raison de sa simplicité, de sa rapidité d'exécution et de son faible coût (en temps et en ressource). Néanmoins, malgré le caractère robuste mentionné plus haut de ce type d'évaluation par un expert humain, l'évaluation perceptive de la sévérité reste très controversée en raison de son caractère subjectif. En effet, la variabilité intra-auditeur (un auditeur peut émettre un jugement variable dans le temps pour un même échantillon de voix/parole) et inter-auditeur (deux auditeurs peuvent émettre des jugements différents pour un même échantillon de voix/parole) sont les deux facteurs principaux

de cette subjectivité (Revis, 2004; Ghio et al., 2007).

Evaluation phonétique

L'analyse phonétique a pour objectif d'étudier l'impact des troubles de la parole et/ou de la voix sur la production des phonèmes. Elle repose généralement sur une transcription manuelle en phonèmes, plus ou moins enrichie, réalisée à partir du signal sonore. (Auzou et Rolland-Monnoury, 2006) distingue la transcription *phonémique* - annotation de la séquence de phonèmes prononcés uniquement - de la transcription *phonétique* - annotation de la séquence de phonèmes prononcés et de leurs caractéristiques articulatoires. Les deux approches permettent [7] de catégoriser les erreurs de production commises en termes d'omissions de phonèmes, de substitutions, de délétions, de répétitions et de les quantifier. La seconde permet d'analyser les phonèmes considérés comme distordus en s'appuyant sur les informations complémentaires récoltées et d'affiner la cause des erreurs ; Par exemple, l'observation de distorsions sur certaines consonnes pourra fournir des indications précises sur la présence de troubles articulatoires et leur localisation comme dans (Antolík et Fougeron, 2013).

L'analyse phonétique est, par conséquent, une approche très intéressante. Néanmoins, elle s'avère très coûteuse en temps d'annotation et n'est donc pas compatible avec une pratique clinique. Le test phonétique d'intelligibilité, décrit plus haut, peut-être considéré comme une version dégradée de l'analyse phonétique, très rapide mais limité en terme d'analyse.

Evaluation motrice des organes

Contrairement aux approches évoquées jusqu'à présent qui mettaient la parole au centre de l'évaluation, l'attention se porte ici sur les organes impliqués dans la production de la parole. En effet, l'observation de ces derniers et de leur potentiel dysfonctionnement peut permettre, d'une part, de définir une topographie de la lésion plus ou moins précise, suivant le nombre d'organes déficients (plus leur nombre est important, plus la localisation de la lésion devient complexe), et d'autre part, de comprendre dans quelle mesure la déficience du ou des organes est corrélée aux perturbations observées dans la parole. Comme précisé dans (Auzou, 2007a), *"[...] Idéalement, il faudrait pouvoir relier l'organe touché (étage respiratoire, larynx...), le type de perturbation en cause (déficit de force musculaire, de coordination...) et les conséquences sur la parole". [...]*

Dans ce cadre, différentes grilles d'évaluation ont été proposées afin d'évaluer la motricité des différents organes dans un contexte verbal et non verbal. Le Frenchay Dysarthria Assessment (Enderby, 1983) ou la BECD (Auzou et Rolland-Monnoury, 2006) en font partie. Elles tiennent compte, par exemple, de la respiration, de la phonation, de la motricité oro-faciale, ... à partir d'un jeu d'épreuves à réaliser par le patient et cotées suivant des niveaux qui leur sont propres. Elles présentent par ailleurs un intérêt

7. suivant le degré d'intelligibilité de la parole.

dans le suivi longitudinal des patients car elles sont reconnues comme étant sensibles au changement.

2.4.2 Evaluation instrumentale

Face à la complexité des mécanismes mis en oeuvre dans la production de la parole et à leur interdépendance, il peut s'avérer utile, dans le cadre d'un bilan clinique, d'évaluer le fonctionnement de chaque composant impliqué (respiratoire, phonatoire, articulatoire, ...) indépendamment des autres. Si l'évaluation perceptive peut donner un certain nombre d'indications, comme évoqué avec l'évaluation motrice des organes, l'instrumentation peut venir compléter cette évaluation en apportant des informations sur l'intégrité et l'état fonctionnel de ces composants et valider les observations perceptives. En effet, comme souligné dans (Murdoch, 1998), *"[...] instrumentation has given the clinician the ability to determine the contributions of malfunctions in the various components of the speech production mechanism to the production of disordered speech [...]"*. Ainsi, l'instrumentation est considérée comme un moyen (1) d'accroître la précision des diagnostics en apportant une observation objective et, par conséquent, une validation du dysfonctionnement d'un ou plusieurs composants, (2) de vérifier par des mesures objectives l'efficacité d'un traitement thérapeutique à court terme ou lors d'un suivi longitudinal, (3) d'auto-contrôle biologique (biofeedback) dans le cadre thérapeutique (Murdoch, 2011). Une revue des différents types d'instruments utilisés pour évaluer les mécanismes respiratoire, phonatoire et articulatoire est fournie dans (Lowit et Kent, 2010). Seuls quelques un d'entre eux sont utilisés en pratique clinique, les autres sont, soit en usage limité en raison de leur caractère trop invasif pour le patient, soit destinés uniquement au domaine de la recherche. Si l'utilisation de l'évaluation instrumentale continue à progresser pour une meilleure compréhension des dysfonctionnements et des troubles inhérents dans la production de la parole, les instruments eux-mêmes progressent également pour mieux répondre aux besoins des chercheurs et des cliniciens (Murdoch, 2011).

2.4.3 Evaluation psycho-sociale

Depuis une quinzaine d'années, la littérature fait état d'une nouvelle dimension à prendre en compte dans le bilan clinique des troubles de la voix et de la parole. Cette nouvelle dimension concerne l'impact psycho-social de ces troubles chez le patient et dans sa vie au quotidien. Plusieurs éléments sont à l'origine de cette évolution. En premier lieu, chaque individu peut réagir différemment à une situation donnée en fonction du contexte. Deux personnes atteintes d'une même pathologie et présentant un même degré de sévérité peuvent être plus ou moins affectées par leur état de santé dans leur vie au quotidien suivant leur situation professionnelle, leur rôle dans la société, leur niveau de sociabilité, leur comportement face au regard des autres, leur état émotionnel, etc. De même, il n'est pas rare pour un clinicien d'établir chez un patient un profil sévère de dysfonctionnement de certains mécanismes de production de la parole sans que ce dernier ne se considère comme sévèrement atteint, sa vie au quotidien n'ayant

pas été influencée outre mesure. Par ailleurs, l'adoption en 2001 d'un nouveau cadre de travail par l'Organisation Mondiale de la Santé (OMS) pour classifier et normaliser sur le plan international les conséquences des maladies (en complément de la classification des maladies - CIM) a fortement contribué aux investigations autour de l'évaluation psycho-sociale. En effet, le précédent cadre de travail, connu sous le nom de "Classification Internationale des Handicaps et santé mentale ou CIH", définissait les déficiences, incapacités et désavantages associés à chaque maladie sur la base d'un modèle pleinement orienté santé. Son successeur, rebaptisé "Classification internationale du Fonctionnement, du handicap et de la santé ou CIF", remplace le modèle orienté santé par un modèle orienté société. Ainsi, les conséquences d'une maladie sont étudiées à la fois en fonction d'un contexte médical mais également d'un contexte sociétal. Le CIF fait ainsi intervenir les notions de changements des fonctions organiques et des structures anatomiques, de capacités et performances à réaliser une tâche dans un environnement standard et réel/sociétal, de facteurs environnementaux (réalité physique ou sociale) et personnels affectant le fonctionnement ou le handicap. La prise en charge clinique du patient doit, dans ce contexte, tenir compte de sa maladie et des troubles qui en résultent mais également être adaptée à son environnement sociétal afin de répondre au mieux à ses besoins et maintenir (si possible) voire améliorer sa qualité de vie.

Différentes tentatives ont été menées pour fournir un questionnaire propre aux troubles de la parole et de la voix permettant de mesurer leurs impacts psychologiques et sociaux chez les patients. Nous pouvons citer le Voice Handicap Index (VHI) (Jacobson et al., 1997), spécifiquement élaboré pour les troubles de la voix. Sur la base de 30 items, il explore le domaine physique au travers de l'inconfort ressenti par le patient au niveau du larynx et de la qualité de la voix, le domaine fonctionnel en considérant l'impact du trouble sur les activités quotidiennes du patient et le domaine émotionnel de ce dernier. Il a été reconnu en 2002 par l'agence américaine pour la recherche et la qualité dans les soins de santé comme le seul outil de mesure du handicap vocal répondant à leur critère de qualité (normes, fiabilité et validité) (Auzou et Rolland-Monnoury, 2006). Si le VHI est reconnu et communément utilisé en pratique clinique, il n'existe pas à l'heure actuel de questionnaire équivalent en terme de reconnaissance pour les troubles de la parole malgré les différents travaux menés dans ce sens (Speech Handicap Index (Rinkel et al., 2008), Dysarthria Impact Profile (Walshe et al., 2009), French Dysarthria Impact Profile (Letanneux et al., 2013), Parole Handicap Index (Fichaux-Bourin et al., 2009). Comme souligné par Margaret Walshe dans (Lowit et Kent, 2010), *"We need to learn more about the psychosocial impact of motor speech disorders in order to assist the development of appropriate interventions and outcome measure and to assist in policy planning and development"*.

2.5 Une demande de la part des cliniciens

Face à la complexité des mécanismes mis en oeuvre dans la production de la parole, l'évaluation perceptive, quelle que soit sa forme, ne permet pas d'appréhender tous les dysfonctionnements et de comprendre leur répercussion sur la parole. Pourtant, elle reste actuellement l'évaluation la plus répandue en milieu clinique malgré son caractère subjectif bien reconnu, et ce, en l'absence de méthodologie bien établie. Dans le même ordre d'idée, la classification de (Darley et al., 1975), bien que controversée, reste également une des références dans le domaine.

Si les travaux de recherche se poursuivent dans l'analyse phonético-acoustique de la parole pathologique et dans la mise en place d'évaluations psycho-sociales mieux adaptées aux troubles de la voix et de la parole, les avancées de ces dix dernières années en traitement automatique de la parole ont ouvert de nouvelles perspectives dans le développement de l'évaluation instrumentale pour la pratique clinique. Ainsi comme mentionné dans (Baghai-Ravary et Beet, 2012) : *[...] Numerous techniques for automatic evaluation of speech disorders have been proposed over the last 20 years, but the relative merits of different approaches and the relative usefulness of some aspects of the speech signal are only now becoming apparent [...] Such techniques offer the promise of a simple and cost-effective, yet objective, assessment of a range of medical conditions, which are potentially of great value to clinicians.*

La prochaine partie de ce document sera dédiée à une revue des différentes applications possibles du traitement automatique de la parole dans le cadre de l'évaluation clinique, de l'aide aux patients ou encore à une meilleure compréhension des troubles de la parole et de la voix. Nous présenterons dans un deuxième chapitre les travaux menés au LIA dans ce même cadre.

Deuxième partie

La "machine" au service des praticiens

Chapitre 1

La place du Traitement Automatique de la Parole

Sommaire

1.1 Corpus disponibles

Les avancées dans le traitement automatique de la parole, observées ces vingt dernières années, ont été rendues, en grande partie, possibles grâce à un important investissement de structures spécialisées (citons par exemple LDC - Linguistic Data Consortium - ou ELRA - European Language Resources Association) ou de projets de recherche dans la collecte et l'enrichissement de données en vue de fournir des corpus structurés et annotés, proches des conditions réelles d'application des systèmes.

Le domaine clinique, de part sa spécificité, est loin de pouvoir bénéficier de telles structures. La collecte des données est le plus souvent régie par les établissements hos-

pitaliers sous une réglementation et des procédures propres à chaque pays. Aussi, cette collecte est très dépendante de la disponibilité des patients (critères d'inclusion), de leur nombre, de leur consentement à participer au programme d'enregistrements, de l'évolution de leur maladie, de leur degré de fatigabilité, etc. Répondre à des exigences quantitatives (large nombre de patients et de sujets contrôles appareillés en genre et en âge, équilibre des populations de patients) et qualitatives (pour répondre aux critères d'inclusion, à des niveaux de variabilité différents dans le stade de la maladie, dans les degrés de sévérité, dans l'âge des patients, etc.) est un objectif difficile à atteindre. Par ailleurs, l'enrichissement des données audio est une phase très importante puisqu'il pourra aider dans l'analyse des résultats et la comparaison entre patients. Ainsi, les informations relatives au patient, à sa pathologie, au stade d'évolution de sa maladie, aux traitements qu'il suit, aux évaluations perceptives réalisées par des cliniciens lors de consultation peuvent être aussi importantes que le signal lui-même (Ghio et al., 2012).
Le domaine du traitement automatique de la parole appliqué aux troubles de la parole et de la voix se retrouve, par conséquent, confronté aux mêmes problèmes posés par le manque flagrant de données qu'il a connus, il y a quelques années, pour la parole *normale* ; problèmes qu'on peut assimiler à ceux des langues peu dotées, de la problématique du traitement des voix d'enfants ou encore des personnes âgées qui restent encore non résolus.

Nous répertorions ci-dessous les quelques bases de données dédiées aux troubles de la parole et de la voix, en se focalisant sur la dysphonie et la dysarthrie, accessibles pour des travaux de recherche.

1.1.1 MEEI

Le corpus MEEI est une base de données commerciale dédiée aux troubles de la voix. Elle a été conçue conjointement par le laboratoire voix et parole du Massachusetts Eye and Ear Infirmary (MEEI) à Boston et la société Kay Elemetrics Corporation dans les années 90. Elle comprend des enregistrements de plus de 700 patients atteints de différentes pathologies et présentant différentes formes de troubles de la voix : organique, neurologique, traumatique, psychogène, etc. Le protocole d'enregistrement, mené à l'hôpital de Boston sur plusieurs années, comporte pour chaque patient la production sur au moins 3s d'un /a/ tenu sur 3 sessions successives, la meilleure production étant retenue, ainsi que la lecture d'un texte ("The Rainbow") pour lequel les 12 premières secondes sont conservées. Le même protocole a été appliqué sur 53 sujets sains par la société Kay Elemetrics Corporation dans des conditions similaires d'enregistrement. Chaque enregistrement des patients s'accompagne d'informations personnelles et médicales, comprenant notamment le diagnostic posé par les cliniciens ainsi que des paramètres acoustiques issus du logiciel *Multi Dimensional Voice Program (MDVP)* développé par la société Kay Elemetrics.

De part, le nombre important de patients enregistrés et la diversité des troubles de

voix répertoriés, cette base de données est largement utilisée dans la littérature, notamment pour la mise au point de systèmes automatiques de classification des voix normales et dysphoniques (voir (Saenz-Lechon et al., 2006) pour une revue de la littérature). Néanmoins, un certain nombre de précautions sont à prendre lors de son utilisation comme reporté dans (Saenz-Lechon et al., 2006). Nous en citons ici quelques unes parmi les plus pertinentes :

- la base de données n'est pas conforme à la description donnée par ses auteurs : sur les 657 enregistrements disponibles, correspondant à 566 patients (plusieurs enregistrements peuvent appartenir au même patient), seuls 345 sont fournis avec un diagnostic mentionnant une dysphonie (306 sans diagnostic, 6 assimilés à des voix normales). Par ailleurs, l'hétérogénéité des diagnostics donnés par les cliniciens est très importante (environ 200 énoncés de diagnostics différents). Le manque de filtrage et d'homogénéisation rend très difficile l'établissement d'une catégorisation des troubles de la voix. Pour finir, aucune évaluation perceptive de la sévérité de la dysphonie n'est fournie en complément du diagnostic. Aussi, il est préconisé dans (Parsa et Jamieson, 2000) de s'en tenir à une classification exclusivement binaire : normal versus pathologique.

- les enregistrements des voix pathologiques et normales ont été réalisés de manière asynchrone dans deux lieux bien distincts (laboratoire voix et parole du MEEI et la société Kay Elemetrics respectivement). Les sujets contrôles ont été sélectionnés par Kay Elemetrics sans évaluation clinique et l'appariement en âge n'est pas respecté. En effet, un appariement en âge entre les locuteurs normaux et dysphoniques ramènent à environ 175 le nombre de locuteurs dysphoniques pertinents. Pour finir, le nombre de sujets normaux (53) est très disproportionné comparé à celui des patients dysphoniques. Bien qu'il soit signalé que des conditions similaires d'enregistrement ont été respectées, aucune garantie n'est donnée sur une éventuelle influence de ce contexte sur des systèmes automatiques.

- Concernant les /a/ tenus, les auteurs de la base de données ont fait le choix de diffuser uniquement la partie stable de la voyelle. Néanmoins, des études ont montré que les parties de démarrage et de fin de la phonation peuvent également contenir des informations acoustiques pertinentes. Par ailleurs, ce découpage ne permet pas de calculer le rapport signal sur bruit des enregistrements.

1.1.2 Whitaker

La base de données Whitaker est l'une des premières bases à avoir été distribuée librement à des fins de recherche sur la parole dysarthrique (Deller et al., 1993). Elle comporte des enregistrements de 6 patients hommes atteints de paralysie cérébrale et présentant des dysarthries de degré de sévérité léger à très sévère et d'un sujet contrôle. Le protocole d'enregistrement consistait dans la lecture de 46 mots isolés comprenant 10 chiffres, les 26 lettres de l'alphabet et 10 mots contrôles ainsi que la lecture de 35 mots issus du texte *"The Grandfather"*. Les patients participaient à 30 sessions de lecture, le sujet contrôle à 15.

1.1.3 Nemours

Les laboratoires "Applied Science and Engineering Laboratories" (ASEL), réunissant des chercheurs de l'hôpital pour les enfants Alfred I. DuPont et de l'université du Delaware, sont à l'origine de la base de données Nemours, enregistrée au milieu des années 90 (Menéndez-Pidal et al., 1996). Cette base de données comprend des enregistrements de 7 jeunes patients américains atteints de paralysie cérébrale, de 4 patients atteints de traumatisme crânien et d'1 sujet contrôle. D'après la grille d'évaluation perceptive "Frenchay Dysarthria assessment" (Enderby, 1983), les 11 patients présentent une dysarthrie dont le degré de sévérité est très variable et peut former trois groupes distincts : 4 patients avec une dysarthrie légère, 3 patients avec une dysarthrie modérée et 4 patients avec une dysarthrie sévère, voire très sévère (un patient est inintelligible) (Dahmani et al., 2013).

Le protocole d'enregistrement comprend deux tâches principales. La première consiste dans la lecture de deux paragraphes de textes classiques ("My grandfather" and "The Rainbow"). La seconde est dérivée du test d'intelligibilité proposé par (Kent et al., 1989). Ici les mots sont remplacés par des phrases, incluant, toutes, deux noms monosyllabiques et un verbe bisyllabique, présentés sous la forme suivante : THE nom1 IS verbe-ING THE nom2. Les noms sont choisis aléatoirement parmi une liste de 74 mots présentant par sous-groupe des contrastes phonétiques proches (exemple de sous-groupe : *fin, thin, sin, bin, pin, inn*). De même, les verbes sont issus d'une liste de 37 mots (exemple de sous-groupe : *sipping, sinning, sitting*). Lors d'une session d'enregistrement, chaque patient entend de la part du sujet contrôle 37 phrases à prononcer, construites suivant le schéma précédent, et les répète dans la foulée. Ces 37 mêmes phrases sont à nouveau prononcées par le sujet contrôle en inversant la position des noms 1 et 2 dans la phrase et répétées par le patient. Ainsi, à la fin de l'enregistrement des 74 phrases, les patients ont prononcé deux fois chaque nom et verbe, les noms pouvant être positionnés en début ou fin de phrase.

Les enregistrements de 10 des 11 [1] patients sont accompagnés des résultats de l'évaluation perceptive réalisée par un praticien suivant la grille "Frenchay Dysarthria assessment", d'une transcription orthographique manuelle en mots et d'une transcription automatique en phonèmes corrigée manuellement (voir (Menéndez-Pidal et al., 1996) pour les détails du système automatique de transcription en phonèmes). De plus, les résultats de tests d'intelligibilité réalisés par 5 auditeurs naïfs (étudiants de l'université du Delaware) sur la base des enregistrements des 10 patients dysarthriques sont également fournis avec la base de données. Ces tests, similaires à ceux décrits dans (Kent et al., 1989), consistent à choisir, parmi des listes fermées de mots aux contrastes phonétiques proches, ceux (nom1, nom2 et verbe) prononcés par les patients dans chacune des 37 phrases (Polikoff et Bunnell, 1999).

Conçue dans le cadre de recherche sur l'intelligibilité (Polikoff et Bunnell, 1999)(Kain et al., 2007), la base de données Nemours est majoritairement utilisée pour la mise au

1. Le patient considéré comme inintelligible n'a pas été inclus ici.

point de systèmes automatiques comme les systèmes de transcription de la parole dysarthrique (Rudzicz, 2007)(Morales et Cox, 2009).

1.1.4 UA-speech

Le Universal Access speech corpus, nommé UA-speech database, a été développé par des chercheurs du "Statistical Speech Technology Group" de l'université de l'Illinois. Il est diffusé librement à des fins de recherche. Il est composé d'enregistrements audio-visuels de patients américains dysarthriques dédiés au développement de systèmes vocaux d'assistance pour des personnes ayant de profonds troubles moteurs (Kim et al., 2008).

Cette base comprend actuellement 19 patients - 1 quart de femmes, 3 quarts d'hommes - tous atteints de paralysie cérébrale et présentant majoritairement une dysarthrie spastique (quelques patients présentent d'autres formes de dysarthrie) d'après le diagnostic posé par un clinicien lors de l'écoute des enregistrements. L'âge des patients varie de 18 à 58 ans avec une moyenne de 31,14 ans et un écart type de 14,16 [2].

Le protocole d'enregistrement comporte la lecture de 765 mots, en trois passations successives mais distinctes. Chaque passation comporte la lecture de 255 mots, dont 155 sont communs entre les trois passations et 100 différents. Les 155 mots se répartissent entre une série de chiffres de 0 à 9, les 26 lettres de l'alphabet radio international (*"Alpha", "Bravo", "Charlie", etc.*), 19 commandes informatiques *"delete", "backspace", "enter", etc.*) et 100 mots communs les plus utilisés dans la langue anglaise écrite *"the", "it", "is", etc.*). Les 100 mots qui diffèrent entre les 3 passations sont issus de contes pour enfants, mais sélectionnés pour contenir un maximum de bi-phones peu fréquents et augmenter ainsi la diversité des séquences de phones. Parmi les 19 patients, seuls 15 d'entre eux ont suivi la totalité de ce protocole. Les 4 patients restant, ayant fait l'objet d'une première étude préliminaire, ont lu seulement 571 mots chacun, comprenant la série de chiffres prononcée 3 fois et la lecture unique des lettres, commandes et d'une liste de mots issue du texte "The Grandfather" et du protocole TIMIT. Les patients ont été soumis à un test d'intelligibilité réalisé par des auditeurs naïfs (5 par patient). Leur tâche était de retranscrire, pour chaque patient, 200 mots issus des enregistrements de la 2e passation. A partir des taux de mots correctement transcrits, moyennés sur l'ensemble des 5 auditeurs, chaque patient a été classé dans une des 4 catégories suivantes (de la sévérité la plus grande à la plus faible) : *very low* (taux entre 0 et 25%), *low* (entre 26 et 50%), *mid* (entre 51 à 75%) et *high* (entre 76 à 100%) (Kim et al., 2008).

1.1.5 TORGO

La base de données TORGO, initiée en 2008 et toujours en cours de collection, est à l'origine d'une collaboration entre les départements d'informatique et de pathologie

2. La moyenne et l'écart type sont calculés sur 14 patients seulement, 5 patients étant annotés uniquement comme ">18".

du langage et de la parole de l'Université de Toronto, de l'hôpital Holland-Bloorview pour la rééducation des enfants de Toronto et la fondation Ontario pour les paralysies cérébrales. Cette base de données a pour vocation de fournir des enregistrements de parole dysarthrique dédiés au développement de systèmes automatiques de reconnaissance de la parole mais également à l'étude de la production de parole atypique par des chercheurs ou des cliniciens du domaine. Pour répondre à cette ambition, la collection d'enregistrements se distingue par le fait de fournir des données alignées sur le plan acoustique et articulatoire, enregistrées tout au long des productions de parole de patients dysarthriques et de sujets contrôles (Rudzicz et al., 2012).

D'après la littérature, la base de données comporte des enregistrements de 7 patients dysarthriques (4 hommes et 3 femmes) dont l'âge varie entre 16 et 50 ans. 6 patients sont atteints de paralysie cérébrale présentant des dysarthries de différentes formes et 1 patient est atteint de sclérose latérale amyotrophique. Ces patients sont appariés en genre et en âge à des sujets contrôles du même nombre. Les fonctions motrices, et notamment celles des articulateurs, ainsi que le niveau d'intelligibilité de tous les patients ont été évalués suivant le Frenchay Dysarthria Assessment (Enderby, 1983) par un praticien.

Le protocole d'enregistrement de la parole comprend la lecture d'un nombre conséquent d'éléments très variés permettant d'analyser le comportement des patients dysarthriques sur différents types de production de parole. Il comporte ainsi la lecture de non-mots formés, par exemple, autour de plosives (/a p i/, /i p a/, /p a t a k a/, etc.) pour observer les comportements articulatoires, la lecture de mots courts (série de chiffres, alphabet radio international, commandes informatiques, les 50 mots du test d'intelligibilité du Frenchay Dysarthria Assessment, les 360 mots et les 162 phrases du test d'intelligibilité du Yorston-Beukelman Assessment of Intelligibility of Dysarthric Speech (Yorkston et Beukelman, 1981), toutes les paires de mots présentant des contrastes phonétiques issues du test d'intelligibilité de (Kent et al., 1989), les 10 mots anglais les plus communs, la lecture de phrases (3 phrases sélectionnées pour leur richesse phonémique), un passage du texte "The Grandfather" et les 460 phrases issues de la base de données TIMIT. Pour finir, chaque patient doit décrire 30 images afin de collecter de la parole "plus" naturelle.

Concernant les données articulatoires, elles ont été relevées grâce au système 3D d'articulographie électro-magnétique EMA [3] AG500 et à un système d'enregistrements vidéo de marqueurs faciaux phosphorescents. Ces systèmes permettent d'enregistrer les mouvements articulatoires à l'intérieur et à l'extérieur du conduit vocal en synchronie avec la parole.

1.1.6 PC-GITA - Maladie de Parkinson

Ce corpus, de langue espagnole, regroupe des enregistrements de 100 locuteurs colombiens. Dédié exclusivement à l'étude de la maladie de Parkinson, il comprend 50 patients et 50 sujets contrôles, équilibrés en genre (Orozco-Arroyave et al., 2014). Le

3. Electro-Magnetic Articulograph.

protocole d'enregistrement comprend différentes tâches ciblées pour l'évaluation de la phonation (production de voyelles tenues et de voyelles avec variation de la hauteur de voix), de la capacité articulatoire (séries diadococinésiques, séries de mots balancés sur le plan phonologique, séries de mots en lien avec des verbes et objets courants de la langue) et de la prosodie (répétition de phrases, lecture d'un dialogue balancé phonétiquement, lecture de phrases mettant le focus sur certains mots, monologue durant lequel le locuteur doit raconter une journée type). Tous les enregistrements ont été réalisés dans une chambre sourde avec le même matériel. Le diagnostic de la maladie de Parkinson a été posé par une équipe de neurologistes de la Clinique Noel de Médellin en Colombie. La sévérité de la maladie a été évaluée suivant les échelles UPDRS et Hoehn et Yahr (H&Y) par un membre de l'équipe. Les patients ont été enregistrés pas plus de 3 heures après la prise de leur traitement thérapeutique (Etat On).

L'ensemble des enregistrements des 50 patients et ceux de 11 patients supplémentaires réalisés dans des conditions de bruits non contrôlées sont diffusés, comme corpus, dans le challenge COMPARE - COMputational PAralinguistics ChallengE - d'Interspeech 2015 dans une sous-tâche nommée *Parkinson's Condition (PC)* et dédiée à la prédiction de l'état neurologique des patients atteints de la maladie de Parkinson sur la base de l'échelle UPDRS (Schuller et al., 2015). Dans le cadre de ce challenge (et de manière récurrente pour chacune des sous-tâches proposées), un ensemble de 6373 paramètres acoustiques incluant des descripteurs bas niveau d'énergie, spectraux, cepstraux, de qualité vocale, de rapport signal sur bruit, ... extraits grâce au toolkit openSMILE (Eyben et al., 2010) sont fournis aux participants.

1.1.7 Les autres bases de données

D'autres bases, distribuées librement à des fins de recherche (ou en voie de l'être), existent. Elles concernent d'autres langues que la langue anglaise. Nous pouvons citer :

- la base *Alborada-I3A* qui comporte 11 enfants ou très jeunes adultes espagnols (de 11 à 21 ans) présentant une dysarthrie (mais également des troubles du langage) et ayant prononcé, sur plusieurs sessions, des mots isolés, des phrases courtes et longues (Saz et al., 2008) ;

- une base coréenne de parole dysarthrique, initiée en 2010 et dont l'objectif réside dans l'enregistrement de 600 patients atteints de paralysie cérébrale d'ici 2014 en vue de développer des systèmes automatiques d'assistance. Le premier jeu de données, enregistré entre 2010 et 2011 et décrit dans (Choi et al., 2011), fait état de 120 patients dysarthriques et 40 sujets contrôles, ayant prononcé entre 200 et 360 mots isolés (chiffres, lettres de l'alphabet coréen, commandes, mots phonétiquement équilibrés) ;

- la base de données de Saarbrücken, comportant plus de 2000 voix de patients dysphoniques et sujets contrôles allemands, dédiée à l'analyse des troubles de la voix (Barry et Pützer). Chaque session comporte l'enregistrement de voyelles

tenues (/a/, /i/, /u/) à différentes hauteurs de voix (normal, basse, haute) et la lecture d'une phrase, déclinée sous sa forme audio et sous la forme d'un signal issu d'un Electro-GlottoGramme (EGG), renseignant sur la vibration des cordes vocales.

1.1.8 Et un corpus français ?

En France, aucune base n'est encore distribuée librement à des fins de recherche, bien que de nombreux corpus soient disponibles dans les laboratoires de recherche ou les centres hospitaliers. Un des freins majeurs à la diffusion de la plupart de ces corpus est que la collecte de données n'a pas été réalisée à des fins de diffusion, mais à des fins de recherche immédiate. Aussi, aucune attention particulière a été apportée sur le "formatage" au sens large des données et sur leur homogénéisation, que ces données soit d'ordre acoustique, personnelles ou cliniques. Reprendre ces corpus en vue d'une diffusion est un travail conséquent et très coûteux. Une partie du travail réalisé dans le projet DesPhoAPaDy [4] (Fougeron et al., 2010) financé par l'Agence Nationale de la Recherche (ANR) entre 2009 et 2011 avait pour objectif de réfléchir à cette problématique en s'appuyant sur plusieurs corpus de patients dysphoniques et dysarthriques, disponibles dans les laboratoires Parole et Langage d'Aix-en-Provence (LPL) et de Phonétique et de Phonologie de Paris (LPP) et les établissements hospitaliers partenaires. Cette réflexion a donné lieu à la conception d'un modèle de données en vue d'utiliser un système de gestion de bases de données relationnelles et de mettre en place une interface de requêtage permettant la consultation des informations et leur récupération sous forme de collections de données (Ghio et al., 2012). A l'heure actuelle, la base est alimentée par différentes sources d'enregistrements mais n'est pas encore diffusée.

1.2 Méthodologies et approches

Nous présentons dans les sections qui suivent quelques uns des champs d'application du traitement automatique de la parole dans le cadre des troubles de la parole et de la voix ainsi que les approches et méthodologies proposées dans la littérature. Bien qu'un certain nombre de ces travaux portent sur la synthèse automatique, nous avons fait le choix de ne pas les aborder ici au vu de l'expertise très limitée du LIA dans ce domaine.

1.2.1 La reconnaissance automatique de la parole

Dans le contexte des troubles de la parole, les attendus applicatifs de la reconnaissance automatique de la parole (RAP) ont évolué au cours du temps. On peut relever dans la littérature de ces 30 dernières années 3 grandes classes d'applications : (1) les

4. http ://despho-apady.univ-avignon.fr/

outils thérapeutiques, (2) les outils d'évaluation, et (3) les outils de communication alternative (Griffin et al., 2000).

La RAP comme outil thérapeutique

Dans les années 80-90, différentes études ont porté sur l'utilisation des systèmes de RAP dans la thérapie des troubles de la parole. Le dispositif mis en oeuvre dans ces études se décline généralement sous la forme d'une interface dédiée associée à un système de RAP. L'interface affiche une liste de mots, un par un, que le patient doit prononcer. Suivant la réponse du système de RAP - mot reconnu ou pas - le système passe au mot suivant ou demande au patient de répéter le mot mal reconnu. Le système de RAP est considéré comme un système de feedback dont l'objectif est d'aider les patients à mieux réaliser et comprendre les difficultés qu'ils rencontrent dans leur production de parole et à les corriger (dans la mesure du possible) (Ferrier et al., 1992; Parsons, 1997; Shriberg et al., 1990). Dans (Ferrier et al., 1992), il est montré qu'un tel dispositif peut permettre à des sujets contrôles et à des patients dysarthriques d'améliorer la précision de leurs articulateurs suite à la mauvaise reconnaissance des mots par le système de RAP DragonDictate. Par ailleurs, plusieurs de ces études soulignent le fait qu'un système de RAP peut avoir comme potentiel d'accroître la motivation des patients pour participer aux séances de thérapie (Parsons, 1997; Shriberg et al., 1990). Si ces études semblaient prometteuses, peu de systèmes opérationnels basés sur un système de RAP ont vu le jour ; Nous pouvons citer deux exemples : les systèmes OLP (Oster et al., 2002) et CATSEAR (Turk et Arslan, 2005). Nous verrons un peu plus loin les raisons de cette situation. En parallèle, la notion de "système de feedback" a été maintenue et a conduit au développement de systèmes connus sous différentes terminologies : CBST (*Computer-based speech therapy system*), CA(M)ST (*Computer-assisted (method for) speech therapy system*) ou encore CASLT (*Computer-Aided Speech and Language Therapy*) (Saz et al., 2009; Schipor et al., 2012). A notre connaissance, aucun de ces systèmes n'est utilisé en France.

Dans (Hamidi et Baljko, 2013), la raison majeure invoquée dans le très faible développement de systèmes de feedback à base de RAP est *[...] A recognized obstacle for the use of ASR in speech intervention systems has been that this technology oftentimes does not perform well for non-standard pronunciations and can lead to inconsistent feedback [...]*. En effet, dès lors que le système est utilisé en vue de corriger les mauvaises productions du patient, ce dernier doit garantir un taux très faible de faux positifs. En d'autres termes, le système de RAP doit être fiable et précis dans les décisions qu'il prend. Hors, malgré les différentes avancées technologiques relatées dans la littérature décrites plus loin dans la section 1.2.1, ce niveau de fiabilité n'est pas encore suffisant et maîtrisé en fonction des pathologies et de leur degré de sévérité (Young et Mihailidis, 2010). Par ailleurs, (Griffin et al., 2000) soulève une question intéressante sur l'utilisation des systèmes de RAP dans un tel contexte : quel type de système serait le mieux adapté ?

– un système indépendant du locuteur dont le taux d'erreurs mots, s'il est trop important, risque d'influer négativement sur la thérapie du patient (il est bien connu qu'un système de RAP indépendant réagit mal en situations adverses comme le

bruit ou des variantes de prononciation non attendues comme dans le cas de locuteurs non natifs) ;

– un système dépendant du locuteur, qui prend en compte les spécificités de prononciations mais qui nécessite de la part du patient de produire des séries de mots de la manière la plus correcte possible lors des phases d'apprentissage (uniquement possible si la dysarthrie est très légère et dans le cas de maladies dégénératives pour lesquelles la maladie est vouée à évoluer) ;

– un système adapté au locuteur qui va apprendre les productions incorrectes des patients et qui ne saura plus rejeter les prononciations erronées. Cette question semble être toujours en suspend à l'heure actuelle.

Une étude des besoins en pratique clinique menée auprès de cinq spécialistes de la parole et du langage dans (Hamidi et Baljko, 2013) souligne qu'un système de feedback doit indiquer si la production d'un segment de parole a été réalisée de manière correcte ou pas mais doit également analyser et évaluer le degré de déviation entre le segment de parole attendu et celui produit de manière erronée. Les systèmes de RAP actuellement disponibles répondent pour l'instant uniquement au premier objectif (certains travaux de recherche pourraient y remédier comme, dans (Morales et Cox, 2009), qui propose une modélisation des erreurs commises par des patients dysarthriques, ou de manière plus générale, tous les travaux sur les mesures de confiance des systèmes de transcription). La conclusion donnée dans (Hamidi et Baljko, 2013) reste pourtant positive sur l'utilisation des systèmes de RAP mais leur confère un rôle différent dans le processus thérapeutique : *[...] We recognize the limitation of ASR to analyse non-standard speech and instead use it to facilitate and motivate the use of speech as input mode. The task of providing detailed feedback can be left to the SLP* [5] *(the "human agent") and the use of ASR, and the computational media more generally, can be recruited for user engagement, motivation and the elicitation of speech production [...]*, rejoignant les réflexions antérieures de (Parsons, 1997; Shriberg et al., 1990) citées plus haut.

Pour finir, (Hawley et al., 2005) donne une vision différente sur l'utilisation de la RAP en tant qu'outil thérapeutique. Les auteurs prônent tout d'abord le fait qu'un système de RAP peut avoir un comportement différent de celui attendu par un praticien et qu'il n'est pas possible de focaliser son "attention" sur un trouble particulier (production des occlusives) dès lors que le dysfonctionnement au niveau des articulateurs est varié (allongement des voyelles). Par ailleurs, il met en avant le fait que des patients atteints d'une dysarthrie sévère présentent généralement une inconsistence dans la production d'une même combinaison de sons, ce qui rend leur parole difficilement prédictible (notamment par des systèmes automatiques dépendants du locuteur). Ils préconisent, par conséquent, d'utiliser les systèmes de RAP dans une optique différente : interagir avec un système de RAP, avec ou sans un praticien, peut conduire ces patients à mieux maîtriser leurs productions de parole et à diminuer la variabilité de ces dernières.

5. Speech and Language Therapist.

La RAP comme outil d'évaluation

Dans le chapitre précédent, nous avons vu que les tests d'intelligibilité sont couramment utilisés dans l'évaluation (perceptive) de la sévérité des troubles de la parole. Ces tests reposent sur l'écoute par des auditeurs experts ou naïfs de productions de parole pour lesquelles ils doivent évaluer par un score, retranscrire manuellement ou bien désigner dans une liste fermée d'items ce qu'ils perçoivent. Nous avons vu que les défauts majeurs de ces tests sont leur subjectivité (évaluation par un score) et le coût important en temps et ressources humaines pour les réaliser. Nous pouvons ajouter à ces défauts le phénomène d'habituation dans la transcription manuelle ou le choix des items dès lors que les listes de mots à prononcer sont en nombre limité et que les tests sont réalisés par les mêmes auditeurs ; ces derniers "devinent" ou "reconstruisent" le mot, même, dans une version sonore très dégradée. Les systèmes de RAP ont été considérés comme des candidats idéaux pour pallier ces limites par la fourniture d'une transcription rapide, objective et à moindre coût. Dans (Carmichael et Green, 2004; Carmichael, 2007), un système de RAP est utilisé dans une version informatisée du FDA (Enderby, 1983) pour fournir des scores d'intelligibilité. Néanmoins, nous sommes confrontés ici aux mêmes exigences que précédemment, à savoir que le test d'intelligibilité doit faire preuve de fiabilité et de précision pour être pertinent et validé pour une pratique clinique. Par ailleurs, la grande majorité des systèmes de RAP sont conçus pour émettre une transcription en sortie quel que soit le signal d'entrée et les confusions qu'il peut engendrer. Prenons le cas simplifié d'un système de transcription ayant dans son vocabulaire d'entrée le mot cible et trois autres possibilités (cas d'une liste fermée d'items). Si la prononciation du mot cible par le patient conduit à une production erronée, mais néanmoins plus proche acoustiquement du mot cible que des trois autres possibilités, la sortie de la RAP pourra être de manière également erronée le mot cible lui-même. Ce mode de fonctionnement n'est pas le mieux adapté à un test d'intelligibilité pour lequel on s'attendrait à une non réponse de la part du système.
Malgré les réserves émises dans (Griffin et al., 2000) : *"[...] it is unclear to what extent ASR will be a useful assessment tool [...]"*, notamment sur le dernier point évoqué, la littérature fait état de deux grandes écoles sur ce sujet (Martinez et al., 2013). Dans la première, les systèmes de RAP sont utilisés directement pour fournir une transcription et un taux d'erreurs mots à partir duquel un score ou degré d'intelligibilité est dérivé (Doyle et al., 1997; Sharma et al., 2009; Maier et al., 2010; Christensen et al., 2012). Dans la deuxième école, les systèmes de RAP (ou technologies assimilées comme l'alignement en phonèmes) sont utilisés pour extraire des caractéristiques du signal de parole. Celles-ci sont, dans un second temps, injectées, dans un modèle de prédiction du degré d'intelligibilité (Middag et al., 2009; Nuffelen et al., 2009).

La RAP comme outil de communication alternative

Lorsque la dysarthrie s'accompagne d'handicaps physiques graves, les patients éprouvent d'énormes difficultés pour intéragir avec leur environnement, limitant sévèrement leur indépendance et leur inclusion dans la société. Ces difficultés touchent de manière non

exhaustive (1) le contrôle des éléments majeurs d'une maison comme les équipements audio-visuels, le téléphone, les portes et volets, l'appel à une assistance, (2) l'accès à un ordinateur et au réseau Internet, qui sont devenus des incontournables dans la vie quotidienne, tout particulièrement lorsque l'accès vers l'extérieur est limité et (3) finalement la communication avec autrui, notamment lorsque la parole est proche de l'inintelligibilité ou tout simplement absente. De nombreux dispositifs électroniques sont maintenant disponibles pour permettre à ces personnes d'accomplir des tâches simples de la vie quotidienne (Fager et al., 2012). L'accès à ces dispositifs s'effectue, selon les capacités de la personne, soit directement par des claviers adaptés ou des écrans tactiles, soit indirectement par des systèmes électroniques de balayage d'un ensemble d'items différents que la personne peut sélectionner par pression d'un commutateur, une vocalisation ou un geste particulier dès que l'item désiré est balayé. De même, les systèmes électroniques de communication améliorée et alternative (Augmentative and Alternative Communication - AAC) ont connu un fort accroissement, permettant aux personnes handicapées souffrant de troubles de la parole de compléter ou remplacer la parole ou l'écriture et de ce fait d'améliorer leur communication avec le monde extérieur (le lecteur pourra se référer à (Antoine, 2008) pour une revue des différents systèmes). La synthèse vocale associée à des systèmes de génération de messages par sélection de lettres, de mots, de phrases voire de pictogrammes est un exemple de dispositifs AAC. Pourtant, certains de ces systèmes d'accès ou de communication, notamment ceux par sélection indirecte, sont très lents et peuvent conduire à certaines frustrations dans leur utilisation au quotidien (Hawley et al., 2005).

Par son côté naturel et immédiat, la parole reste le moyen d'accès le plus convoité, même pour des personnes atteintes de troubles profonds de la parole. Aussi, le bénéfice d'un système à base de RAP, capable de traiter efficacement de la parole produite par une personne atteinte d'une dysarthrie sévère, soit pour contrôler des dispositifs électroniques, soit pour générer de la parole plus intelligible (par synthèse vocale du message transcrit) est unanimement reconnu dans la littérature (Griffin et al., 2000; Rosen et Yampolsky, 2000; Hawley et al., 2005; Selouani et al., 2009; Fager et al., 2010) et dans les projets de recherche dédiés comme STARDUST [6] (Speech Training And Recognition for Dysarthric Users of aSsistive Technology, 2000-2003), SPECS [7] (Speech-driven Environmental Control System, 2006-2009), ENABL (ENabler for Access to computer-Based vocational tasks with Language and speech) ou VIVOCA [8] (Voice Input Voice Output Communication Aid, 2004-2007).

Récentes avancées des systèmes de RAP appliqués aux troubles de la parole

D'énormes progrès ont été réalisés dans le domaine de la RAP ces 10 dernières années. Les avancées technologiques, l'accroissement de la puissance des machines ainsi que le développement de grandes bases de données de parole ont participé à ces évolutions et à l'amélioration des performances des systèmes de RAP. Les systèmes actuels,

6. http ://spandh.dcs.shef.ac.uk/projects/stardust/
7. https ://www.shef.ac.uk/cast/projects/specs
8. http ://www.shef.ac.uk/cast/projects/vivoca

dits à grands vocabulaires et dédiés à la parole continue, sont appris sur des centaines d'heures de parole multi-locuteurs afin d'obtenir des modèles acoustiques de qualité, capable de capturer une part importante et pertinente de la variabilité de la parole. Ils reposent sur des millions de mots, issus de documents textuels pour affiner leurs modèles de langages. Ils comportent différents traitements, appliqués en parallèle ou successivement en vue d'améliorer la qualité des transcriptions. Un type important de traitements peut reposer, suivant la tâche visée, sur les approches d'adaptation dont l'objectif est d'affiner les modèles acoustiques ou de langage initialement appris, à la parole du locuteur qui parle et au thème qu'il aborde dans son discours respectivement. Si les applications basées sur des technologies vocales sont maintenant à la portée du grand public au travers des téléphones portables notamment, il reste encore une grande marge de progression pour que ces systèmes soient opérationnels dans n'importe quelles conditions d'utilisation : environnement bruité, canaux de transmission dégradés, superposition de parole, locuteurs non natifs, accents régionaux, parole émotionnelle, parole d'enfants, adaptation à de nouvelles langues peu dotées en corpus, etc. En effet, les technologies actuelles ne permettent pas, encore, de tenir compte de toute la variabilité qui peut être observée dans le signal de parole en fonction du locuteur (Benzeghiba et al., 2007) ou des dégradations que le signal peut subir lors de sa transmission et de son acquisition.

L'application de la RAP aux troubles de la parole pose différents verrous :

– les altérations de la parole observées chez les patients atteints de dysarthrie peuvent être une source plus ou moins grande de perturbations des systèmes de RAP, mis au point spécifiquement pour décoder de la parole normale. Parmi ces altérations, on notera par exemple, l'imprécision des consonnes (notamment les plosives et les fricatives (Rudzicz, 2010)) et des voyelles, conduisant à des erreurs de substitutions par les systèmes de RAP (Rosen et Yampolsky, 2000; Rudzicz, 2007; Morales et Cox, 2009), l'allongement des phonèmes dû à un débit de parole ralenti (étude comparative sur la longueur des voyelles et des consonnes sonores entre sujets contrôles et patients dysarthriques (Rudzicz, 2010)), conduisant à l'interprétation par le système de RAP de la production d'un mot à deux syllabes au lieu d'une seule (Morales et Cox, 2009), la présence de longues occlusions silencieuses à l'intérieur d'un mot, conduisant à la détection de deux mots au lieu d'un seul, la non production de certains phonèmes, notamment en début ou fin de mots, conduisant à des omissions et/ou substitutions de mots (Mumtaz et al., 2014). Par ailleurs, les patients dysarthriques, et notamment ceux présentant des dysarthries modérées à très sévères démontrent plus de difficultés à prononcer de la parole continue que des mots isolés (Hawley et al., 2005; Young et Mihailidis, 2010). Pour finir, d'autres facteurs peuvent venir perturber les systèmes de RAP : la fatigue que les patients peuvent ressentir très rapidement dès lors qu'on leur demande de fournir un effort prolongé ou l'inconsistance de la parole chez les patients atteints de dysarthrie sévère, tous deux se traduisant par une plus grande variabilité dans les signaux de parole (et leurs altérations) produits pour une même combinaison de phonèmes (Green et al., 2003; Young et Mihailidis, 2010). Ainsi, les études menées depuis ces 20 dernières années montrent que les systèmes de

RAP peuvent être tout à fait performants lors de leur utilisation par des patients atteints de dysarthrie légère (de manière similaire à des sujets contrôles), sur des tâches de reconnaissance de mots isolés ou de parole continue. En revanche, ces performances se dégradent de manière drastique en présence de dysarthrie modérée et sévère, même en reconnaissance de mots isolés (Ferrier et al., 1995; Rosen et Yampolsky, 2000; Green et al., 2003; Fager et al., 2010; Young et Mihailidis, 2010). Comme souligné dans (Green et al., 2003), *"[...] For severely dysarthric speech, recognizers trained on normal speech corpus cannot be expected to work well [...]"*.

– Malgré la volonté de développer des bases de données de parole dysarthrique comme les corpus Nemours ou UA speech (voir section 1.1), les données nécessaires pour tenir compte, au sein des modèles acoustiques des systèmes de RAP, de la variabilité de la parole en fonction de la maladie, de sa progression, du type de dysarthrie, de sa sévérité, de l'état de fatigabilité du patient sont très loin d'être en quantité suffisante. Les corpus disponibles comportent des enregistrements de lecture de mots isolés, de phrases ou de paragraphes de textes uniquement, la parole conversationnelle étant très peu ou pas du tout prise en compte. Autre aspect négligé, bien que secondaire, tous les enregistrements sont réalisés dans des environnements normalement contrôlés, sans bruit afin de ne pas perturber la qualité du signal de parole. Par conséquent, les dégradations du signal par l'environnement extérieur ne sont pas prises en compte dans les modèles acoustiques, ce qui réduit l'utilisation des systèmes dans des environnements moins "aseptisés" comme le domicile d'un patient par exemple, pouvant présenter des bruits extérieurs, comme le volume de la télévision ou de la radio, des écoulements d'eau, des bruits de ventilation, etc.

– En fonction des applications décrites précédemment, les objectifs fixés à la RAP ne sont pas les mêmes. Entre commander un appareil électronique et communiquer au travers d'un message électronique, la tâche de reconnaissance n'est pas la même - tâche de reconnaissance de mots isolés ou de mots connectés versus reconnaissance de parole continue -, la taille du vocabulaire à identifier n'est pas la même - quelques dizaines de mots de commandes vs quelques milliers au minimum. Entre commander un appareil électronique (hors assistance) et évaluer la sévérité de la dysarthrie d'un patient ou encore fournir un retour dans un cadre thérapeutique, les taux de performance attendus au niveau de la RAP ne sont pas les mêmes. Bien que le patient puisse se sentir frustré de devoir répéter plusieurs fois une commande pour obtenir un résultat probant, ce résultat est acceptable. En revanche, il n'est pas possible de l'induire en erreur dans le cadre d'une rééducation thérapeutique. Sur la même comparaison, les environnements d'utilisation peuvent être très différents. Eteindre la télévision à son domicile, alors que celle-ci est en fonctionnement (générant, par conséquent, un bruit de fond non négligeable), n'est pas équivalent à être seul avec un praticien dans son cabinet, ie dans une pièce sans bruit ambiant, face à un ordinateur. Pour finir, contrôler son environnement et accéder à des services publics, comme son compte bancaire par exemple, ne peuvent pas faire intervenir les mêmes moyens ou outils de person-

nalisation d'un système de reconnaissance vocale. Dans le premier cas, un système dédié, intégrant différentes phases d'apprentissage adaptées aux difficultés motrices de la personne, pourrait être envisagé. Dans le second cas, il s'agit d'un système "grand public", initialement conçu pour être indépendant du locuteur afin d'en permettre l'accès à tous. Aussi, les outils mis en place pour améliorer son utilisabilité ne pourront pas être aussi spécifiques que dans le cas précédent.

Les systèmes classiques de RAP utilisent de grands volumes de données de parole (des centaines d'heures) pour apprendre des modèles acoustiques indépendants du locuteur ; on parle alors de systèmes indépendants (SI) à opposer aux systèmes dépendants (SD) du locuteur qui requièrent des données de parole propres à chaque locuteur afin d'apprendre des modèles spécifiques. Au cours de la reconnaissance, des techniques d'adaptation, telles que MAP (Gauvain et Lee, 1994) ou MLLR (Leggetter et Woodland, 1995) par exemple, permettent d'adapter ces modèles génériques aux données du locuteur (ou à un environnement particulier) qui est en train de parler afin de les rendre plus fiables et améliorer les performances de reconnaissance du système (on parle alors de système adapté au locuteur - SA).

Ce schéma a été étudié de nombreuses fois dans la littérature dans le cas de la dysarthrie, notamment dans des études pionnières basées sur les premiers systèmes commerciaux de dictée vocale (Ferrier et al., 1995; Kotler et Thomas-Stonell, 1997; Doyle et al., 1997; Thomas-Stonell et al., A998). En effet, au vu de la faible quantité de parole dysarthrique disponible, de la difficulté à la collecter (milieu clinique, fatigabilité des patients, etc.), des déviances observées entre parole dysarthrique et parole normale, quel type de système (SI, SD, SA) est le mieux adapté au traitement de la parole dysarthrique ?

(Morales et Cox, 2009) montre de manière plutôt attendue qu'un système SI, appris sur de la parole normale, fournit des performances moindres comparé à un système SA. En effet, sur 10 patients dysarthriques de la base de données Nemours pour lesquels 34 phrases sont utilisées comme données d'adaptation et 40 comme données de tests, tous améliorent leurs taux de reconnaissance mots en mode SA. Les gains absolus en termes de taux d'erreurs mots (Word Error Rate - WER) s'échelonnent entre 15% et 40%.
Dans (Sanders et al., 2002), les modes SI et SD sont comparés dans le cadre de reconnaissance de parole néerlandaise produite par deux sujets contrôles et deux sujets souffrant de dysarthrie légère [9]. Les expériences menées montrent que les performances de reconnaissance sur différentes tâches (nombres, mots isolés et phrases) sont bien meilleures pour les patients dysarthriques en mode SD alors que peu de différence, voire une légère dégradation des performances est observée pour les sujets contrôles.

Dans (Rudzicz, 2007), un système SD est comparé à un pseudo-système [10] SA sur la base de données Némours (11 locuteurs dysarthriques, 1 sujet contrôle). Les mo-

9. Il est à noter que les auteurs précisent que la parole de ces deux patients peut être plutôt inintelligible pour des auditeurs non familiers, ce qui, à notre avis, rend l'expression "mild form of dysarthria" peut-être sous-estimée.

10. Le terme pseudo est utilisé ici car il ne s'agit pas d'une approche par adaptation classique.

dèles acoustiques du pseudo-système SA sont initialisés sur une base de données de parole normale lue, puis ré-estimés (et non adaptés) individuellement en utilisant une partie des données de parole dysarthrique de chaque patient. Les résultats montrent que les patients donnés comme légèrement et modérément dysarthriques obtiennent de meilleurs résultats de reconnaissance en mode SA comparé au SD (23,1% et 4,9% de réduction du WER respectivement, 30,7% pour le sujet contrôle) alors que les patients présentant une dysarthrie sévère obtiennent des résultats tout à fait similaires dans les deux modes.
Des observations comparables sont données dans (Raghavendra et al., 2001) sur 4 patients dysarthriques (léger, modéré, sévère, très sévère) et un sujet contrôle, excepté le fait que dans le cas de la dysarthrie sévère, le mode SD obtenait des résultats bien supérieurs que le mode SA.
De manière contrastive, l'étude comparative menée dans (Sharma et Hasegawa-Johnson, 2010) d'un système de RAP en mode SD et SA, basé ici sur l'approche MAP à partir de modèles appris sur de la parole normale, montre la supériorité du mode SA, pour 5 des 7 patients testés, issus de la base de données Universal Access. Parmi les 7 patients testés, les deux patients présentant le plus faible taux d'intelligibilité (évalué par le FDA entre 2 et 6%) améliorent de manière plus ou moins franche leurs taux de mots corrects (15% de gain en absolu pour l'un, 1,4% pour l'autre) ; les deux patients n'obtenant pas d'amélioration avec le mode SA sont associés quant à eux à des taux d'intelligibilité également faibles (28 et 29%).

Dans une étude récente (Mumtaz et al., 2014), les auteurs étudient dans le cadre d'un système SA quel type de données est le plus approprié pour initialiser les modèles acoustiques (normal ou dysarthrique) et comparent deux méthodes d'adaptation différentes. Les expériences sont conduites en utilisant la base TORGO pour initialiser les modèles acoustiques représentatifs de la parole dysarthrique et la base Nemours pour les données d'adaptation et de tests. Les conclusions montrent qu'il est préférable d'utiliser des données normales pour initialiser les modèles acoustiques des patients légèrement dysarthriques avant adaptation et des données dysarthriques pour les patients modérés et sévères. Pourtant, ces résultats vont à l'encontre de ceux obtenus dans (Christensen et al., 2012). Dans cette dernière étude, parmi les différentes expériences conduites à partir d'un système en mode SD et SA, en présence de parole normale et dysarthrique issue de la base de données Universal Access, les auteurs montrent que l'adaptation de modèles acoustiques estimés sur de la parole dysarthrique - versus parole normale - permet d'atteindre des taux de mots corrects meilleurs [11] sur les 15 patients dysarthriques de la base, dont les 5 présentant des taux élevés d'intelligibilité (évalués par le FDA entre 86 et 95%, pouvant être considérés comme légèrement dysarthriques) avec des gains allant de 0 à 10,9% en absolu. Par ailleurs, parmi les nombreuses configurations testées dans ces travaux, les auteurs soulignent l'hétérogénéité des systèmes en termes de comportement et de gain de performances observés chez les 15 patients : *[...] no 'one solution to fit all' [...]*. La raison principale invoquée est que le potentiel des techniques d'adaptation, telles que MAP ou MLLR, repose sur l'hypo-

11. 1 seul patient obtient strictement le même résultat.

thèse que les données d'entraînement des modèles et les données d'adaptation doivent être suffisamment *proches* acoustiquement parlant. Qu'en est-il de la parole normale et de la parole dysarthrique des patients testés dans ces travaux ? Des conclusions très similaires peuvent être relevées dans les travaux de (Sehgal et Cunningham, 2015), qui expérimentent également différentes techniques d'adaptation des modèles acoustiques et comparent les systèmes SI, SD et SA en fonction du degré d'intelligibilité des patients dysarthriques : *[...] Despite the results reported here, there is still no consensus on the best approach to model dysarthric speech with varying severity, aetiology or type [...]*. Finalement, dans (Christensen et al., 2014), les auteurs proposent de sélectionner les patients à utiliser dans les modèles acoustiques indépendants du locuteur appris sur la parole dysarthrique avant la phase d'adaptation au locuteur. L'idée sous-jacente est d'estimer un modèle indépendant sur des données acoustiques les plus proches de celles du locuteur visé afin de tenir compte de la variabilité, en terme de dysarthrie, entre patients. Les performances atteintes, comparées à un système SD et SI appris sur l'ensemble des patients, montrent tout l'intérêt de cette sélection.

Vers des approches différentes.

La différence de comportement des systèmes (notamment en modes SA et SD) suivant le degré de sévérité des patients (voire des patients eux-mêmes) a conduit les chercheurs dans différentes voies d'investigation, et notamment vers des systèmes de RAP divergeant des systèmes classiques à grand vocabulaire. En effet, excepté quelques travaux plus classiques (comme par exemple, architecture hybride basée sur des HMM et des réseaux de neurones dans (Polur et Miller, 2006), réseaux de neurones et sélection du meilleur ensemble de paramètres acoustiques dans (Shahamiri et Binti Salim, 2014), ou encore réseaux récurrents et paramètres glottiques dans (Nidhyananthan et al., 2016)), l'attention s'est portée sur l'étude de nouvelles approches permettant de mieux prendre en compte les spécificités de la parole dysarthrique, certains se focalisant sur les patients très dysarthriques et le développement de systèmes spécifiques pour le contrôle de leur environnement par exemple, d'autres conservant l'idée de systèmes génériques quel que soit le degré de sévérité des patients et cherchant à en améliorer les performances.

L'impulsion donnée par le projet STARDUST a amené les chercheurs impliqués à développer un système ASR dédié à la reconnaissance de mots de commandes pour aider les patients atteints de dysarthrie sévère. Dans (Green et al., 2003), le système proposé est un système dépendant du locuteur basé sur des modèles de mots entraînés sur des patients dysarthriques. Le vocabulaire est limité à 10 mots de commandes pour optimiser les taux de reconnaissance. Le système s'accompagne d'une interface permettant de choisir les mots les mieux adaptés aux difficultés de prononciation du patient. Des mesures fournies par le système permettent de guider ces choix. Ces mesures tiennent compte de la consistance d'un mot en comparant plusieurs répétitions de ce dernier à l'aide de son modèle acoustique estimé au préalable et de la confusion possible entre les mots par le système ASR. Si un mot n'est pas répété de manière constante

par un patient ou s'il peut être facilement confondu avec un autre mot du vocabulaire, les performances du système ne pourront pas être optimales. Une expérimentation menée sur 8 patients sévèrement dysarthriques, ayant suivi le protocole d'apprentissage individualisé du système précédent, montre une amélioration du taux de mots corrects moyen (de 88,5% à 95,4%) (Hawley et al., 2007). Cette même expérimentation, répétée au domicile de 5 patients en mode non contrôlé, conduit à des résultats très intéressants (moyenne de 87% de taux d'erreurs mots). Même si le système vocal s'avère moins performant qu'un système indirect de sélection par balayage, accomplissant la tâche de contrôle demandé avec 100% de succès après 3 tentatives dans la majorité des cas (entre 73 et 83% de réussite lors du premier essai suivant les patients), son temps d'exécution est deux fois plus rapide (tenant compte des différentes tentatives). La différence de performances entre les deux expérimentations (95% versus 87%) montre la difficulté (bien connue) des systèmes de RAP à maintenir leur efficacité dans des conditions de bruit non contrôlées. Des projets assez similaires sont actuellement en cours en Corée pour le développement d'une plateforme mobile de contrôle d'accès dédiée à des patients dysarthriques et basée sur un vocabulaire limité et personnalisable (Kim et al., 2013) et au Mexique pour le développement d'un robot assistant (Morales et al., 2013).

Pour pallier le manque flagrant de données relatives à la parole dysarthrique nécessaires pour estimer des modèles acoustiques de bonne qualité, (Christensen et al., 2013) et (Sharma et Hasegawa-Johnson, 2013) proposent deux approches alternatives à l'utilisation des techniques classiques d'adaptation. La première tire partie des travaux récents dans le domaine de la RAP sur l'apport de larges corpus de données dits "hors domaines" (Bell et al., 2012) vis-à-vis du champ d'application de la RAP pour lequel peu de données d'apprentissage sont disponibles. Ces corpus sont alors utilisés pour apprendre des réseaux de neurones profonds (Deep Neural Networks - DNN (Hinton et al., 2012)) ou des réseaux adaptatifs à plusieurs niveaux (Multi-Level Adaptive Network - MLAN (Bell et al., 2012)) en vue d'extraire des paramètres pertinents utilisés en entrée du système de RAP (au même titre que les paramètres acoustiques). Dans (Christensen et al., 2013), le large corpus de données correspond à la parole normale, le champ d'application de la RAP à la parole dysarthrique. Les paramètres acoustiques classiquement utilisés en RAP sont combinés aux paramètres extraits des réseaux de neurones pour former des vecteurs uniques impliqués dans l'estimation des modèles acoustiques. La seconde alternative proposée repose sur deux hypothèses fortes : (1) la variabilité de la parole entre patients d'une population dysarthrique est plus large que celle entre parole normale et parole dysarthrique, (2) directement issue des réflexions de (Weismer et al., 2010), la variabilité de la parole normale peut être bornée pour un mouvement donné d'un articulateur ou un paramètre acoustique, la distance entre ces bornes et celles déterminées pour un patient dysarthrique peut informer sur les capacités motrices (et leurs limites) de ce dernier. De ces hypothèses, (Sharma et Hasegawa-Johnson, 2013) met en avant le besoin de mieux prendre en compte la variabilité intra-locuteur dans les modèles acoustiques des systèmes de RAP. Une nouvelle technique d'adaptation, nommée BI-MAP est suggérée, reposant sur l'implication d'un modèle SD propre à chaque patient dysarthrique dont la forme simplifiée (un unique GMM inspiré du modèle UBM (Reynolds et al., 2000)) tire partie de la faible quantité

de données par patient dysarthrique disponible. Ce modèle GMM SD intervient, d'une part, pour contraindre l'estimation d'un modèle acoustique SI (appris sur de la parole normale) dans un espace acoustique proche de celui de la parole du patient et, d'autre part, dans l'estimation d'un modèle SA à partir du modèle SI et de données d'adaptation propres à la parole dysarthrique.

Les derniers travaux originaux que nous présentons envisagent une approche plus analytique dans leur prise en compte des altérations du signal de parole dysarthrique dans les systèmes de RAP. Dans (Saz et al., 2006), une normalisation des paramètres acoustiques est directement implantée dans l'algorithme de décodage du système de RAP. Cette normalisation a pour objectif de tenir compte de la grande variabilité de la parole dysarthrique, tant dans le domaine temporel que fréquentiel, et de la réduire.

Les erreurs de production de patients dysarthriques, en terme de délétions, substitutions et insertions, sont étudiées dans (Morales et Cox, 2009), modélisées et corrigées dans le processus de reconnaissance lors d'une seconde passe (la première passe étant basée sur un système de RAP classique). La modélisation est basée sur l'utilisation de transducteurs à états finis couplés à une matrice de confusion des phonèmes. Les transducteurs permettent de simuler la suppression et l'insertion de phonèmes tandis que la matrice de confusion gère les substitutions entre phonèmes. Comparé à un système SA basé sur MLLR, la modélisation des erreurs et leur correction permet d'obtenir des gains moyens de l'ordre de 2 à 5% en termes de taux de reconnaissance de mots corrects, suivant le degré d'intelligibilité des patients (5% étant le gain moyen obtenu chez les patients les moins intelligibles). Des travaux très similaires sont également reportés dans (Seong et al., 2012a,b). Finalement, dans la même lignée, les travaux menés dans (Christensen et al., 2013) se basent sur l'utilisation de réseaux neuronaux profonds (DNN) appris sur un large corpus de parole normale conversationnelle pour détecter des erreurs (ou variantes) de prononciation chez des patients dysarthriques. Une fois ces erreurs détectées automatiquement, des dictionnaires dépendants des patients, comportant de nouvelles prononciations pour les mots concernés sont générés et utilisés par un système de RAP. Les résultats de reconnaissance de mots isolés menée sur les 15 patients dysarthriques de la base UA démontrent encore un comportement très variable suivant les patients (gain ou perte en termes de taux corrects de mots comparé à un système SA), décorrélé du degré d'intelligibilité.

Finalement, l'intégration dans un processus de RAP de paramètres articulatoires issus de parole dysarthrique est expérimentée dans (Rudzicz, 2012). Ces travaux ont été rendus possibles grâce au développement par ces mêmes auteurs de la base de données acoustico-articulatoires, TORGO. Le système mis en place repose sur une première passe de reconnaissance de mots classique sur la base de modèles acoustiques. Une liste ordonnée des N meilleures hypothèses de mots est fournie à un premier composant, dédié à l'édition d'une séquence de gestes articulatoires canoniques associée à chacune d'elles. Chacune de ces séquences est ensuite transformée, en fonction de données réelles issues de la base TORGO et d'une base de données similaire (MOCHA) conçue uniquement à partir de sujets sains. Grâce à un modèle d'inversion de

l'acoustique vers l'articulatoire, une vraisemblance articulatoire pour chaque séquence transformée est calculée en fonction du signal d'entrée. Un nouveau score peut finalement être calculé pour chacune des hypothèses de mots basé sur une combinaison linéaire du score acoustique et du score articulatoire. Le réordonnancement de ces scores permet de désigner le (nouveau) mot reconnu. Les expériences conduites sur deux patients sévèrement dysarthriques montrent que l'utilisation des données articulatoires seules mènent à de meilleures performances que l'utilisation des données acoustiques (taux d'erreurs mots de 43,2% vs. 51,6% et 27,8% vs. 32,3% pour les deux patients en comparant données articulatoires vs. données acoustiques). En revanche, l'utilisation conjointe des deux types de données par le système décrit précédemment n'apporte que très peu de gain, même dans sa configuration optimale (taux d'erreurs mots moyen de 34,1% pour les deux patients dysarthriques contre 35,5% pour les données articulatoires seules). Néanmoins, cette étude a ouvert de nombreuses pistes d'investigations et peut être très prometteuse pour mieux comprendre les troubles articulatoires associés à la dysarthrie. Dans le même ordre d'idée, (Hahm et al., 2015) expérimente l'utilisation conjointe de données acoustiques et articulatoires après application de différentes approches de normalisation des données afin de réduire la variabilité inter-locuteur.

Bilan

Au travers de ce que nous venons d'évoquer, la RAP appliquée aux troubles de la parole apparaît comme un véritable challenge et les verrous que nous avons mentionnés précédemment sont encore loin d'être levés. Ce domaine d'application de la RAP semble nous ramener une quinzaine d'années en arrière avec, par exemple, les travaux sur la reconnaissance de digits ou de mots de commandes dans un environnement très bruité comme une voiture. Le parallèle peut également être fait avec un domaine plus récent concernant le développement de systèmes de RAP pour des langues peu dotées. Comme mentionné dans (Besacier et al., 2014), ce domaine doit également lever de nombreux verrous, dont certains assimilables ou transposables aux troubles de la parole : le manque de ressources disponibles (absence de larges corpus de parole, de dictionnaires phonétisés, de listes de mots dans certains cas, etc.), des langages comportant de nombreux dialectes présents dans différentes régions, des alternances de codes linguistiques possibles au cours du discours, la présence massive de locuteurs non natifs, l'observation de nouvelles formes de langages oraux qui peuvent amener à une remise en question des conventions adoptées pour des langages bien connus ("le phonème est il l'unité la mieux adaptée pour la phase de modélisation acoustique ?"), la nécessité d'un travail pluri-disciplinaire mêlant spécialistes du traitement de la parole, dialectologues, phonéticiens, voire des ethno-linguistes pour analyser une langue et concevoir les approches pour la traiter. Même si nous avons pu voir que les tâches de RAP dans le contexte des troubles de la parole sont pour l'instant très restreintes (reconnaissance de mots isolés ou de phrases) alors que ce domaine vise une reconnaissance plus élaborée, il doit, dans le même ordre d'idée, faire face à une parole pouvant dévier de celle couramment traitée, pouvant comporter une très grande variabilité (assimilable à la problématique des dialectes, des locuteurs non natifs) vis-à-vis d'une norme "fixée" par les langages déjà traités et pour lesquels les ressources nécessaires existent,

la norme de la langue elle-même pouvant être difficile à définir.

Pour conclure, il nous semble évident que les travaux sur une meilleure prise en compte de la variabilité intra- et inter-locuteur présente dans la parole dysarthrique en fonction des patients, de leur pathologie, du degré de sévérité de leur dysarthrie au sein des systèmes automatiques pourront apporter de l'information pertinente au domaine de la RAP en général et inversement.

1.2.2 Détection et Evaluation objective des troubles de la voix et de la parole (hors RAP)

De nombreux travaux sont dédiés dans la littérature à l'évaluation par des systèmes automatiques des troubles de la parole et de la voix, qu'il s'agisse d'une décision binaire - parole/voix pathologique vs normale - ou d'une évaluation sur une échelle de sévérité.

Une grande majorité de ces travaux porte notamment sur l'évaluation de la dysphonie. Nous présenterons également d'autres pistes récemment explorées dans la littérature sur l'évaluation de la dysarthrie.

Détection et Evaluation de la dysphonie

Ce thème est abordé sous deux angles principaux dans la littérature : le choix des paramètres à extraire du signal pour caractériser la présence ou non de dysphonie et le type d'approches pour fournir une décision.

Concernant le choix des paramètres, (Alonso et al., 2005) fait mention de quatre grandes classes de phénomènes susceptibles de refléter la présence de troubles de la voix :

- la périodicité et la stabilité de la voix. L'étude des cycles vibratoires permet d'observer la présence de perturbations de la fréquence fondamentale du signal sonore ou de son amplitude. Ces perturbations peuvent être le signe d'altérations du fonctionnement des cordes vocales. On retiendra parmi les mesures les plus utilisées pour mesurer ces perturbations : le jitter et le shimmer, généralement estimés sur des voyelles tenues, (Baken et Orlikoff, 2000; Teston, 2004; Vasilakis et Stylianou, 2009) et les mesures RAP (Relative Average Perturbation), PPQ (Pitch Perturbation Quotient), APQ (Amplitude Perturbation Quotient) considérées comme mieux adaptées à une extraction à partir de parole continue (Kiliç et al., 2004) ;

- la présence de bruit. Un dysfonctionnement des cordes vocales, conduisant à une fermeture glottique partielle, induit une fuite d'air et, par conséquent, la présence

de bruit dans la production sonore. Différentes mesures de bruit ont été proposées dans le cadre de la détection des voix pathologiques ou de l'évaluation de la qualité vocale : rapport signal sur bruit (SNR), rapport harmonique sur bruit (HNR), énergie du bruit normalisée (NNE), index de turbulence de la voix (VTI), rapport excitation glottique bruit (GNE) (Godino-Llorente et al., 2008; Daza-Santacoloma et al., 2009) ;

– la "richesse" spectrale. Le contenu spectral des voix dysphoniques est généralement atypique, présentant (1) des perturbations dans la structure des harmoniques dues en particulier au caractère non périodique du signal mentionné précédemment, (2) de l'énergie dans les hautes fréquences due à la fermeture glottique partielle (présence de bruit). L'extraction de paramètres classiquement utilisés en traitement de la parole a été proposée ici comme les Mel Frequency Spectral Coefficients (MFSC) ou Mel Frequency Cepstral Coefficients (MFCC)(Dibazar et al., 2002; Fredouille et al., 2005; Fraile et al., 2009c). D'autres auteurs ont proposé des analyses temps-fréquence basées sur la transformation en ondelettes discrètes (Fonseca et al., 2007) ;

– le comportement non linéaire. Différents travaux ont montré l'existence de caractéristiques non linéaires dans le signal de parole dues en particulier au fonctionnement intrinsèque des cordes vocales. L'analyse de ces caractéristiques non linéaires permettrait de discriminer les pathologies de la voix (Titze et Alipour, 2006). Deux principales mesures, utilisées en dynamique non-linéaire ont été proposées dans le cadre de la détection des voix pathologiques : le plus grand coefficient de Lyapunov (LLE) et la dimension de corrélation (CD) (Giovanni et al., 1999; Zhang et al., 2004; MacCallum et al., 2009). Néanmoins, ces mesures posent de nombreux problèmes algorithmiques et numériques pour leur estimation et rendent leur usage difficilement transposable dans des systèmes tout automatiques (Little et al., 2007).

Finalement, d'autres approches se sont intéressées aux paramètres bio-mécaniques des cordes vocales et à la paramétrisation de la source glottique indépendamment du conduit vocal (Gómez-Vilda et al., 2007, 2009).

Les systèmes de détection des voix pathologiques ou d'évaluation du niveau de sévérité s'inspirent largement du domaine du traitement de la parole et des approches de classification. On citera notamment l'utilisation de classifieurs simples comme l'approche des plus proches voisins (Shama et al., 2007) ou de classifieurs plus complexes à base de réseaux de neurones (Fraile et al., 2009c), de modèles à base de mixtures de gaussiennes adaptées à partir d'un modèle générique de parole (UBM/GMM) (Fredouille et al., 2005), de modèles de Markov cachés (HMM) (Dibazar et al., 2002), ou encore de séparateurs à vaste marge (Fonseca et al., 2007; Sáenz-Lechón et al., 2008).

Détection et Evaluation de la dysarthrie

Dans le cadre de la dysarthrie, peu de travaux impliquant des outils de traitement automatique autres que ceux déjà exposés sur la RAP dans le cadre de l'évaluation de l'intelligibilité, ont été finalement dédiés, à notre connaissance, à la détection et à l'évaluation de la dysarthrie. Il s'avère que très récemment, un certain nombre de chercheurs se sont intéressés à cette tâche particulière, en se focalisant principalement sur la maladie de Parkinson. Un des événements moteurs de cet intérêt est la proposition d'une sous-tâche du Challenge COMPARE - COMputational PAralinguistics ChallengE - d'Interspeech 2015, dédiée à l'étude des caractéristiques paralinguistiques véhiculées par le signal de parole. Cette sous-tâche, que nous avons mentionnée plus haut dans la section 1.1.6 dédiée aux corpus, a pour objectif la prédiction de l'état neurologique des patients atteints de la maladie de Parkinson sur la base de l'échelle UPDRS. Ce challenge a donné lieu à un certain nombre de travaux reposant sur un protocole expérimental commun fourni par les organisateurs pour l'évaluation des systèmes. Ce protocole comprend un ensemble de données d'apprentissage (35 patients), de développement (15 patients) et de test (11 patients) basé sur l'ensemble des enregistrements des patients du corpus PC-GITA (décrit en section 1.1.6). Pour comparaison, les résultats d'un système "baseline" (résultats donnés en termes de mesure de corrélation de Spearman - 0.390 sur l'ensemble de données de test) reposant sur l'extraction des 6373 paramètres fournis par le challenge aux participants et l'approche "Support Vector Regression (SVR)" pour la fourniture du score de prédiction (Schuller et al., 2015) sont également disponibles. Les axes de recherche de ces travaux sont très similaires à ceux dédiés à la détection et à l'évaluation de la dysphonie. Ils se focalisent principalement sur l'extraction des informations considérées comme les plus pertinentes dans le signal de parole pour caractériser les altérations dues à la maladie, estimer leur degré et corréler au mieux avec le degré de sévérité de la maladie donné par le score perceptif UPDRS (issu de l'évaluation réalisée par un clinicien pour chaque patient). Un axe secondaire concerne la méthode de régression utilisée pour fournir un score de sévérité automatique à corréler avec le score perceptif. Nous présentons ici une sélection de ces travaux jugés les plus intéressants sur le plan méthodologique. Dans (Hahm et al., 2015), les auteurs proposent de combiner les 6373 paramètres acoustiques fournis par le challenge à des paramètres articulatoires estimés à partir d'un modèle d'inversion acoustico-articulatoire. Deux approches de régression sont également proposées, l'approche SVR similaire au système "baseline" et une seconde basée sur des réseaux de neurones profonds (machines de Boltzmann). Les résultats obtenus sur les ensembles d'apprentissage (en cross-validation) et de développement montrent de légères améliorations des résultats, suivant les contextes expérimentaux, lors de l'utilisation conjointe des paramètres acoustiques et articulatoires (notamment les paramètres articulatoires en lien avec la mâchoire), quelle que soit la méthode de régression utilisée. Sur les données de tests, seule l'utilisation des DNN montre l'intérêt de combiner les deux types de paramètres, mais les performances obtenues sont largement plus faibles que le système "baseline". Dans (Orozco-Arroyave et al., 2016), l'apport d'informations articulatoires dans la tâche visée est étudiée de manière différente. En effet, il s'agit ici d'étudier les phases d'initiation et d'arrêt du mouvement vibratoire des cordes vocales, considérées

comme perturbées chez les patients atteints de la maladie de Parkinson. Ces phases sont observées à partir de mesures d'énergie (et statistiques associées) réalisées sur des phases de transitions entre phonèmes non voisés-voisés (initiation) et voisés-non voisés (arrêt). Une étude expérimentale, menée par ces mêmes auteurs, a effectivement montré que l'utilisation de tels paramètres permettait une bonne discrimination entre patients atteints de la maladie de Parkinson et sujets contrôles (Orozco-Arroyave et al., 2015). Dans le cadre du challenge, il s'agit de démontrer que ces mêmes paramètres peuvent également informer sur le score UPDRS des patients. Les auteurs proposent également d'utiliser une mesure d'intelligibilité basée sur le taux de reconnaissance de mots après application d'un système de RAP classique appris pour transcrire la parole normale et non adapté à la parole pathologique. L'approche de prédiction est basée sur la méthode SVR. Les résultats obtenus montrent un intérêt certain dans l'utilisation du taux de reconnaissance de mots et des mesures d'énergie sur l'initiation du mouvement vibratoire des cordes vocales. En particulier, les mesures d'énergie se montrent particulièrement pertinentes dans la prédiction des scores UPDRS élevés, signe d'un stade avancé de la maladie. Dans (An et al., 2015), les auteurs combinent l'utilisation de paramètres acoustiques bas niveaux (ceux fournis dans le cadre du challenge plus ceux de challenges précédents) avec des paramètres relatifs au debit de parole (durée des syllabes, durée des silences, nombre de syllabes par seconde, etc.), à des mesures de formants F1 et F2, lesquels, variant en fonction des mouvements des articulateurs, sont d'après (Sapir et al., 2010) caractéristiques des altérations dues à la maladie de Parkinson, des paramètres phonotactiques (durée des phonèmes et distribution des monophones, biphones et triphones), et de l'approche i-vecteur pour extraire et compresser de l'information acoustique utile véhiculée par le signal de parole sans information a priori (Martínez et al., 2015). Malgré le choix de paramètres ciblés sur les altérations attendues de la maladie de Parkinson, des gains relativement légers sont obtenus par comparaison avec le système "baseline" sur les données de tests, quel que soit le type de paramètres utilisés, excepté pour les paramètres phonotactiques qui n'apportent pas d'amélioration.

Finalement, d'autres pistes originales ont été explorées tout récemment dans le cadre de la dysarthrie. Dans (Oue et al., 2015), les auteurs s'intéressent dans une étude préliminaire aux phénomènes de répétitions de mots partiels, de mots ou encore d'expressions pouvant être accentués chez des patients dysarthriques en raison de leur difficulté de production de la parole ainsi qu'aux dysfluences (bruit de respiration, bruits produits par le locuteur dans la parole, pauses remplies, etc.) également prépondérantes chez ce type de locuteurs. La seconde étude que nous relevons ici s'intéresse non plus au signal de parole mais au contenu linguistique qu'il véhicule par une analyse du discours dans le cadre de parole spontanée (García et al., 2016). En effet, les auteurs s'intéressent aux informations sémantiques véhiculées par le discours, à la construction syntaxique des phrases et à la richesse lexicale du discours. Les résultats semblent montrer que ce type d'informations permet effectivement de dissocier les productions de parole de patients atteints de la maladie de Parkinson de sujets contrôles. Cette étude, bien que préliminaire au vu de la quantité limitée de données traitées, ouvre de nouvelles

perspectives puisqu'elle pourrait montrer que le langage et, non plus uniquement, les aspects moteurs de la production de la parole pourrait être altéré par la maladie de Parkinson.

Chapitre 2

Les travaux du LIA sur les troubles de la parole et de la voix

Sommaire

2.1 Contexte

Mes activités de recherche autour des troubles de la parole et de la voix ont été initiées en 2004 avec de premières collaborations impliquant le Laboratoire Parole et Langage (LPL) d'Aix-en-Provence et portant sur l'évaluation objective de la dysphonie. Ces collaborations ont notamment conduit à la participation et/ou aux co-encadrements d'étudiants dans le cadre de leur master recherche ou de leur mémoire d'orthophonie (école d'orthophonie de Marseille). Le tableau 2.1 fournit l'ensemble de ces collaborations. En parallèle, elles ont également donné lieu aux travaux de Gilles Pouchoulin sur la période 2005-2008, dans le cadre d'une thèse intitulée *Approche statistique pour l'analyse objective et la caractérisation de la voix dysphonique* co-encadrée par Jean-François

Bonastre et moi-même. Finalement, ces activités autour de la dysphonie ont conduit à une collaboration internationale avec l'Universidad Politecnica de Madrid et différentes publications (Fraile et al., 2009a)(Fraile et al., 2009c)(Fraile et al., 2009b) qui ne seront pas exposées ici.

Année	Etudiants	Intitulé du mémoire	Implication
2003-2004	Chloé Briffa	*Analyse de la dysphonie : application des méthodes statistiques de la reconnaissance automatique du locuteur*	Participation dans le cadre du mémoire d'orthophonie encadré par Joana Revis (LPL).
2005	Gilles Pouchoulin	*Caractérisation de la voix suivant des critères pathologiques*	Encadrement du mémoire de Master Recherche Informatique de l'université d'Avignon.
2005-2006	Marion Azzarello	*Analyse phonétique de la dysphonie : application des méthodes statistiques issues de la reconnaissance automatique du locuteur*	Participation dans le cadre du mémoire de Master Recherche Sciences du Langage de l'université AMU encadré par Antoine Giovanni (LPL).
2006-2007	Ondine Champavère	*Reconnaissance automatique de la dysphonie à partir de techniques issues de la reconnaissance automatique du locuteur*	Co-encadrement avec Joana Revis (LPL) du mémoire d'orthophonie.
2008-2009	Audrey Marques	*Qualité discriminante des consonnes sourdes dans la catégorisation des dysphonies par les systèmes de classification automatique*	Co-encadrement avec Joana Revis (LPL) du mémoire d'orthophonie - prix du "meilleur mémoire 2009" attribué par le jury des mémoires de l'école d'orthophonie de Marseille.
2011-2012	Camille Galant	*Cartographie automatique de la dysphonie et stratégies perceptives du jury d'écoute*	Co-encadrement avec Joana Revis (LPL) du mémoire d'orthophonie.

TABLE 2.1 – *Synthèse des encadrements d'étudiants en master Recherche ou en école d'orthophonie de Marseille sur le thème de l'évaluation de la dysphonie.*

A partir de 2009, le LIA s'est associé au Laboratoire de Phonétique et de Phonologie (LPP) de Paris pour travailler sur le thème des troubles de la parole, et notamment, de la parole dysarthrique. Les premiers travaux ont tout d'abord été initiés dans le cadre du projet *VML*, financé par l'association « Vaincre les maladies lysosomales » en collaboration avec le LPP et l'Hôpital La Pitié Salpêtrière (Paris). Ce projet, d'une durée d'un an, dont j'étais la responsable scientifique au sein du LIA, avait pour objectif d'étudier, d'un point de vue acoustique, l'évolution longitudinale de la dysarthrie chez des patients atteints de deux maladies lysosomales rares. Sous cette impulsion, deux autres

projets ANR ont été soumis et acceptés consécutivement pour lesquels j'ai conservé la responsabilité scientifique au LIA : le projet *DesPhoA-PaDy* en collaboration avec les laboratoires LPP (Porteur du projet - Paris) et LPL (Aix-en-Provence) portant sur la caractérisation acoustico-phonétique de la parole dysarthrique et le projet *TYPALOC*, toujours en collaboration avec les laboratoires LPL (Porteur du projet – Aix-en-Provence) et LPP (Paris) visant à mieux comprendre l'étendue des variations de la parole chez des populations tant saines qu'affectées d'une pathologie et dans des conditions de parole différentes – parole lue vs parole spontanée. Ces activités ont donné lieu à l'encadrement ou le co-encadrement de plusieurs étudiants dans le cadre de leur Master Recherche ou de leur mémoire d'orthophonie. Le tableau 2.2 fournit l'ensemble de ces collaborations. Elles ont également été menées dans le cadre d'une thèse de doctorat, réalisée par Imed Laaridh sur la période 2013-2016, co-encadrée par Christine Meunier du LPL et moi-même (sous la direction de Jean-François Bonastre) et financée par the Brain and Language Research Institute (BLRI) dont le LIA est membre et de deux post-doctorats financés par les projets *DesPhoA-PaDy* et *TYPALOC*.

Année	Etudiants	Intitulé du mémoire	Implication
2009	Pierre Clément	*Méthodes objectives issues du traitement automatique de la parole pour la recherche de zones "déviantes" dans la parole dysarthrique*	Encadrement du mémoire de Master Recherche Informatique de l'université d'Avignon.
2010-2011	Delphine Baculard et Sandy Guerineau-Condé	*Intelligibilité dans la sclérose latérale amyotrophique : comparaison des transcriptions de la parole dysarthrique par un jury d'écoute avec un système de reconnaissance automatique de la parole*	Co-encadrement réalisé avec Danièle Robert, docteur en médecine et chercheur au LPL du mémoire d'orthophonie.
2013	Imed Laaridh	*Détection de segments de parole anormaux chez des patients atteints de troubles de la parole*	Encadrement du mémoire de stage de fin d'étude d'ingénieur (5e année) de l'Ecole Nationale des Sciences de l'Informatique (ENSI) de Tunis (stage recherche réalisé au LIA).
2013-2014	Loriane Piro et Louisa Ziamni	*Analyse des sorties d'un système de détection automatique des zones de parole déviante et évaluation perceptive sur des patients dysarthriques et une population contrôle*	Co-encadrement avec Christine Meunier (LPL) du mémoire d'orthophonie.
2015-2016	Laura Pianelli et Laura Restivo	*Evaluation d'un système de détection de déviances dans la réalisation articulatoire de la dysarthrie*	Co-encadrement avec Christine Meunier (LPL) du mémoire d'orthophonie.
2016	Maxime Manos	*Etude de la confusion phonémique chez les patients dysarthriques*	Co-Encadrement avec Imed Laaridh du mémoire de Master Recherche Informatique de l'université d'Avignon.

TABLE 2.2 – *Synthèse des encadrements d'étudiants en master Recherche ou en école d'orthophonie de Marseille sur le thème du traitement de la dysarthrie.*

2.2 Evaluation objective des voix dysphoniques

Face aux limites de l'évaluation perceptive, pourtant incontournable dans le domaine clinique pour évaluer la qualité de la voix chez des patients atteints de dysphonie, l'objectif des travaux menés au LIA sur ce thème était d'étudier l'apport du traitement automatique de la parole en vue d'objectiver l'évaluation de la qualité vocale. Nous présentons ici les résultats les plus significatifs auxquels ces travaux ont mené ainsi que les limites et perspectives relevées. Ces travaux sont issus principalement de la thèse de Gilles Pouchoulin soutenue en 2008 (Pouchoulin, 2008) et des mémoires de master Recherche et d'orthophonie listés dans le tableau 2.2.

2.2.1 Méthodologie et contexte expérimental

Méthodologie

Sur la base de l'expertise du LIA sur les systèmes de Reconnaissance Automatique du Locuteur, un système de classification automatique des voix dysphoniques a été mis en place. L'idée originale était de s'appuyer sur les technologies état de l'art en RAL pour construire ce système de classification, à savoir les analyses spectrale ou cepstrale pour l'extraction et la formalisation de l'information pertinente véhiculée par le signal de parole, des modèles de mixtures de gaussiennes (GMM - Gaussian Mixture Model) pour représenter non pas un locuteur mais une classe de voix pathologiques ou contrôles, et la dérivation d'un modèle générique de parole (UBM - Universal Background Model), appris sur un grand corpus de données, par la technique d'adaptation MAP (Maximum A Posteriori) pour la construction des modèles GMM afin de pallier la faible quantité de données d'apprentissage disponible. Une fois la phase d'apprentissage des modèles terminée, la phase de test consiste à présenter un échantillon de voix inconnu au système de classification afin qu'il puisse la comparer aux différents modèles de classes disponibles et prendre une décision sur la base du maximum de vraisemblance entre la voix et les modèles. En d'autres termes, la classe dont le modèle présentera la plus grande mesure de similarité avec l'échantillon de voix inconnu sera désignée comme classe d'appartenance de ce dernier (le lecteur pourra se référer à (Fredouille et al., 2005; Pouchoulin, 2008) pour plus de détails).

Corpus

Le corpus utilisé dans ces travaux est constitué de 80 échantillons de voix de femmes, âgées de 17 à 50 ans, correspondant à 20 sujets témoins et 60 patientes dysphoniques présentant des pathologies diverses (nodules, polypes, oedèmes, kystes, etc.). Chaque sujet a été enregistré sur la lecture d'un paragraphe de "La chèvre de Monsieur Seguin" à l'hôpital de la Timone à Marseille. Les enregistrements ont été évalués perceptivement selon le grade global (critère G) de dysphonie de l'échelle GRBAS (Hirano, 1981) par un jury de 3 experts. Les décisions ont été prises par consensus afin de limiter la variabilité inter-auditeur et en une seule séance afin de limiter la variabilité intra-auditeur. L'ensemble du corpus est constitué de 20 voix normales de grade 0 (G0), 20 voix présentant

une dysphonie légère de grade 1 (G1), 20 voix une dysphonie moyenne de grade 2 (G2) et 20 voix de dysphonie sévère de grade 3 (G3).

Voix Contrôle	Voix Dysphoniques	Global
% TCC (nb/20)	% TCC (nb/60)	% TCC (nb/80)
95.0 (19)	91.7 (55)	92.5 (74)

TABLE 2.3 – *Résultats de la classification des voix dysphoniques sur la base de 2 classes : voix contrôles/voix dysphoniques, en termes de Taux de Classification Correcte (% TCC).*

Grade 0	Grade 1	Grade 2	Grade 3	Global
% TCC (nb/20)	% TCC (nb/20)	% TCC (nb/20)	% TCC (nb/20)	% TCC (nb/20)
95.0 (19)	65.0 (13)	70.0 (14)	85.0 (17)	78.75 (63)

TABLE 2.4 – *Résultats de la classification des voix dysphoniques sur la base de 4 classes représentant chacun un grade de l'échelle GRBAS, en termes de Taux de Classification Correcte (% TCC).*

Grade 0	Grade 1	Grade 2	Grade 3	Global
% TCC (nb/20)	% TCC (nb/20)	% TCC (nb/20)	% TCC (nb/20)	% TCC (nb/20)
95.0 (19)	70.0 (14)	80.0 (16)	100.0 (20)	86.25 (69)

TABLE 2.5 – *Résultats de la classification des voix dysphoniques sur la base de 7 classes représentant chacun un grade de l'échelle GRBAS auxquels se rajoutent des grades composites, en termes de Taux de Classification Correcte (% TCC).*

2.2.2 Résultats

Classification et pertinence des classes
Trois configurations différentes du système de classification automatique des voix dysphoniques ont été étudiées. Les différences de configuration portent sur le nombre de classes pris en compte par le système :

- 2 classes pour une distinction voix contrôles/voix dysphoniques ;
- 4 classes, de G0 à G3, correspondant aux différents grades de l'échelle GRBAS utilisée dans l'évaluation perceptive du corpus de voix par le jury d'experts ;
- 7 classes intégrant les 4 classes précédentes auxquelles se rajoutent 3 classes intermédiaires par composition de grades : G0, G01, G1, G12, G2, G23, G3. Cette classification affinée avait pour objectif d'assouplir l'approche catégorielle induite par l'échelle GRBAS dans l'évaluation perceptive. Le grade d'un échantillon de voix était considéré comme bien reconnu dès lors qu'il était rattaché à la classe correspondante ou à une classe composite.

Les résultats obtenus sont reportés dans les tableaux 2.3, 2.4 et 2.5 et sont donnés en termes de Taux de Classification Correcte (TCC). Comme souligné dans (Pouchou-

lin, 2008), la comparaison de ces résultats montre que la redéfinition des frontières de grades de l'échelle perceptive GRBAS, sous la forme d'une classification en 2 classes - voix contrôles/voix dysphoniques - ou d'une classification plus fine en 7 grades, permet une amélioration des performances de la classification en 4 grades avec des gains relatifs de +17.5% et +9.5% respectivement. Bien que la classification en 4 grades soit la plus conforme à l'évaluation perceptive pratiquée en milieu clinique et aux données d'apprentissage disponibles ici, cette observation tend à démontrer **premièrement la difficulté pour des experts humains d'évaluer de manière catégorielle la qualité globale des voix dysphoniques en tenant compte de frontières strictes, difficiles à définir de manière objective, et, en second, la difficulté de faire reposer un système supervisé sur des données annotées sur la base de cette évaluation**. En effet, contrairement à d'autres systèmes supervisés de traitement automatique de la parole, tels que les systèmes de transcription ou de reconnaissance automatique du locuteur par exemple, reposant sur des corpus annotés, le problème de l'objectivité et de la fiabilité des annotations se pose ici.

Caractérisation des voix dysphoniques et de leur degré de sévérité
Ce deuxième volet porte sur l'étude des informations pertinentes permettant de caractériser au mieux le grade de sévérité de la dysphonie, toujours dans le cadre d'un système automatique de classification. Malgré les limites citées dans la section précédente, cette étude a été conduite dans le cadre du système de classification en 4 grades des voix dysphoniques.

La pertinence des informations véhiculées par le signal de parole, dans l'optique d'un traitement automatique, a été examinée sous trois angles différents : (1) la représentation paramétrique la mieux adaptée, (2) la discrimination dans le domaine fréquentiel et (3) la pertinence des informations phonétiques. Nous donnerons ici une synthèse des résultats et observations obtenus à partir des nombreux contextes expérimentaux étudiés dans (Pouchoulin et al., 2007, 2008b; Pouchoulin, 2008; Pouchoulin et al., 2008a, 2007).

- **Représentation paramétrique la mieux adaptée.** L'analyse en bancs de filtres, qu'elle soit utilisée pour l'extraction de paramètres dans le domaine spectral ou cepstral semble être la plus pertinente (avec une préférence pour l'analyse spectrale notamment pour la classification des voix de grade 2) comparée à une analyse par prédiction linéaire. L'utilisation de l'échelle MEL, proche de la perception fréquentielle de l'oreille humaine, apporte un gain non négligeable dans le processus de classification. De même, la classification en grades tire partie de l'intégration d'informations dynamiques sous la forme des dérivées premières, secondes, voire troisièmes pour les grades 1 et 3 notamment. L'utilisation d'une fenêtre temporelle plus grande que celle classiquement utilisée en RAL (fenêtre de 5 trames) pour le calcul des coefficients dynamiques n'apporte pas d'information supplémentaire ;

- **Discrimination dans le domaine fréquentiel.** Une analyse en sous-bandes de

1kHz chacune montre que les bandes de fréquences entre 0 et 3000Hz sont les plus pertinentes pour la tâche de classification en 4 grades, la sous-bande [1000-2000]Hz atteignant des performances globales légèrement meilleures (TCC de 70%) que la bande fréquentielle totale (TCC de 65%). Au contraire, les bandes de fréquences entre 3000 et 5000Hz montrent une très nette dégradation des performances. De manière surprenante, la classification des voix de grade 3 tire partie des bandes de fréquences les plus hautes avec notamment le meilleur taux de classification (TCC de 80%) obtenu pour ce grade dans la bande [7000-8000]Hz alors que les bandes entre 5000 et 7000Hz atteignent le même score que sur la bande totale (TCC de 70%). Ce résultat peut trouver une explication dans la présence de bruit résiduel caractéristique des voix de grade 3.
Par ailleurs, considérant la bande de fréquences [0-3000]Hz, il s'avère qu'un gain de performances est observé par comparaison à l'utilisation de la bande fréquentielle totale, pour la classification des voix de grades 0, 1 et 2. Dans la même optique, la bande [0-300]Hz semble également porteuse d'informations utiles à la caractérisation du trouble vocal puisque son retrait conduit à une dégradation nette des performances. Ces observations corroborent tout à fait le fait que la dysphonie est souvent considérée comme atténuée au travers du canal téléphonique analogique en raison de sa bande fréquentielle restreinte ([300-3400]Hz). Dans cette optique, la réévaluation perceptive du corpus de voix par le même jury d'experts après filtrage des fréquences au delà de 3000Hz a été réalisée. Contrairement aux résultats observés sur le système automatique, elle a montré une surestimation de la dysphonie pour les voix de grades 0 et 1 comparée à la bande fréquentielle totale, probablement due à des tonalités de voix plus graves pouvant affecter le jugement des experts, l'évaluation de la sévérité des voix de grade 2 et 3 étant restée stable ;

– **Pertinence des informations phonétiques.** Une analyse des performances du système automatique de classification pour chacune des classes phonétiques suffisamment représentées dans le corpus a été réalisée. La distinction entre classes phonétiques n'intervenait qu'au niveau de la décision, tout le matériel phonétique disponible étant utilisé dans la phase d'apprentissage des modèles de classes. La segmentation en phonèmes était issue d'un alignement automatique contraint par le texte. Parmi les différents résultats relevés, il apparaît que la classe consonantique semble être la plus pertinente pour la classification des voix dysphoniques, comparée notamment à la classe vocalique. Cette observation, bien qu'inattendue en présence de troubles laryngés, est des plus intéressantes et montrerait **tout l'intérêt d'étudier des événements acoustiques plus longs, comme par exemple, les phénomènes vocaux transitoires dans des séquences de type CV**. Par ailleurs, la comparaison au sein de la classe consonantique des performances des consonnes sonores et sourdes montre, également de manière surprenante un très bon comportement des consonnes sourdes, n'impliquant pourtant pas de vibration laryngée. En effet, ces dernières peuvent obtenir des performances de classification égales voire supérieures aux consonnes sonores suivant les grades, leur TCC global atteignant 66.25% contre 62.5% pour les consonnes sonores.

Afin de vérifier la pertinence des consonnes sourdes dans l'évaluation de la qualité vocale, une étude spécifique a été conduite dans le cadre du mémoire d'orthophonie d'Audrey Marques (voir table 2.1). Cette étude reposait sur deux questionnements et hypothèses pouvant expliquer le comportement observé : (1) "les consonnes sourdes sont effectivement discriminantes dans l'évaluation de la dysphonie. Néanmoins, ne mettant en jeu aucun mécanisme de vibration laryngée, en quoi peuvent-elles être utiles dans sa caractérisation ?" ou (2) "des erreurs dans la segmentation automatique des consonnes sourdes conduisent à la présence de parties voisées dans les segments de parole qui leur sont associés, expliquant leur comportement". Les travaux réalisés dans ce mémoire ont permis de répondre aux deux hypothèses émises. En effet, il a été démontré que le caractère discriminant des consonnes sourdes dans la classification automatique des voix dysphoniques se maintenait après prise en compte des erreurs de segmentation automatique de ces dernières (hypothèse 2 - résultats publiés dans (Fredouille et al., 2007)). Ce caractère discriminant pourrait s'expliquer, pour les occlusives sourdes, par un allongement observé de la durée entre l'explosion et le début du voisement de la voyelle qui suit (durée connue sous le terme de Voice Onset Time - VOT). En effet, une augmentation de ce VOT a été observée en fonction du niveau de sévérité de la dysphonie montrant une difficulté croissante de mise en vibration des cordes vocales plus le grade augmente. Pour les fricatives sourdes, bien que jugées comme moins discriminantes que les occlusives, il a été observé un allongement de leurs durées en fonction du grade de dysphonie. **L'hypothèse émise ici serait que la difficulté de mise en vibration des cordes vocales pour la production de la voyelle qui suit, croissante avec le grade de sévérité, pourrait entraîner cet allongement de durées des fricatives sourdes et l'allongement du VOT des occlusives**.

Il est intéressant de noter qu'une étude très récente portant sur l'évaluation de la dysarthrie chez des patients atteints de la maladie de Parkinson pour laquelle la dysphonie est très souvent l'un des troubles observés en plus de la dysarthrie tire partie de ce type d'observations (Orozco-Arroyave et al., 2015) (cf. section 1.2.2).

Phonetic labelling

La dernière étude sur la dysphonie, présentée ici, a porté sur la technique du "phonetic labelling" utilisée par Joanas Revis dans sa thèse de doctorat (Revis, 2004). Cette technique a pour objectif d'évaluer perceptivement le degré du trouble vocal au niveau de chaque phonème produit au sein d'une phrase en vue de fournir une cartographie qualitative très précise de la dysphonie d'un patient. En pratique clinique, la méthode du "phonetic labelling" pourrait permettre de suivre l'évolution de la qualité vocale d'un patient.

Dans cette optique, le système de classification en 4 grades a été utilisé pour mettre en place une méthodologie similaire, nommée "automatic phonetic labelling" dans laquelle la décision du grade est prise au niveau de chaque phonème. Ce travail a été mené conjointement par Gilles Pouchoulin dans le cadre de sa thèse et de Marion Azzarello dans le cadre de son mémoire de master recherche en Sciences du Langage (voir

table 2.1). Le parallèle entre les deux techniques de "phonetic labelling" automatique et perceptive a montré que :

- tout comme relevé dans la méthode perceptive, le système détecte des phonèmes pathologiques dans les échantillons de voix normales (à hauteur de 15% contre 21% pour la méthode perceptive) ;

- 85% et 88% des phonèmes sont attribués au bon grade de sévérité pour les échantillons de voix normales (grade 0) et sévèrement dysphoniques (grade 3) respectivement ; Ces performances chutent de manière drastique pour les échantillons de voix de grade 1 et 2 pour lesquels seulement 62% et 37% des phonèmes sont attribués au bon grade respectivement.

- le nombre de phonèmes considérés comme "pathologiques" par la méthode automatique (ie attribués aux grades 1, 2 ou 3) est fortement corrélé avec le niveau de sévérité de la dysphonie. Egalement présente dans le cas de la méthode perceptive, cette observation suggère que d'une manière plus globale la perception du trouble vocal est basée sur la sommation d'événements dysphoniques intervenant au cours du signal de parole et, non, comme un événement continu. L'approche automatique trouve, par conséquent, toute sa légitimité.

Ces constatations sont très intéressantes et prometteuses au regard de la fiabilité potentiellement discutable de la méthode automatique, basée sur le système de classification en 4 grades, comme discuté précédemment. En effet, l'attribution du grade par l'expert est prise ici sur des durées très courtes, à l'échelle du phonème contrairement aux expériences reportées dans les sections précédentes (grade global). Par ailleurs, elle repose sur la segmentation en phonèmes issue de l'alignement automatique contraint par le texte. Des erreurs d'alignement pourraient induire que la décision soit prise sur une partie des phonèmes adjacents au phonème considéré, expliquant les erreurs de grades observées. Pour finir, ces résultats mettent à nouveau en exergue la difficulté de différencier les grades intermédiaires 1 et 2 comparés aux grades 0 et 3, que ce soit à un niveau segmental ou global.

2.2.3 Conclusions

Au travers de l'analyse des résultats du système de classification automatique en 4 grades, un certain nombre de connaissances acoustiques sur la dysphonie ont pu être mises en évidence, de manière surprenante pour certaines (intérêt de la bande fréquentielle [0-3000]Hz, intérêt des consonnes et en particulier des consonnes sourdes, intérêt de l'étude du VOT, etc.) (Fredouille et al., 2009) (Revis et al., 2012) ; connaissances qu'une analyse purement perceptive ou réalisée par un expert humain n'aurait pas forcément pu mettre en exergue. Une des perspectives majeures de ces travaux serait une phase de validation sur un corpus bien plus conséquent (limité ici à 80 échantillons de voix). Néanmoins, cette perspective est fortement contrainte par le besoin d'annotation en grades des échantillons de voix par un jury d'experts, difficilement réalisable sur un grand corpus au vu de la nécessité de limiter la variabilité intra- et inter-auditeurs. Par

ailleurs, ce constat soulève une des limites également de ces travaux qui réside dans le fait d'appuyer l'apprentissage des modèles statistiques internes au processus de classification sur des annotations de référence issues d'une analyse perceptive. Considérée comme subjective, cette analyse peut, en effet, comporter une part d'imprécision, voire d'erreurs, pouvant nuire, par conséquent, à la qualité des modèles et au processus de décision : cas des échantillons de voix de grade 1 et 2 pour lesquels la frontière est difficile à définir sur le plan perceptif et pour lesquels les performances du système de classification sont les plus dégradées.

2.3 Traitements automatiques dans la parole dysarthrique

Comme nous l'avons mentionné précédemment, l'imprécision des consonnes, la distorsion des voyelles, le débit faible ou encore l'hypernasalité sont des perturbations fréquemment observées dans la parole dysarthrique. Cependant, des analyses acoustiques et phonétiques plus précises restent nécessaires afin de considérer la large variabilité des phénomènes observables à l'intérieur d'un même groupe pathologique et entre groupes (Tomik et Guiloff, 2010) ou entre différents styles de parole (parole lue, parole spontanée, mots isolés, etc.). En effet, la maladie du patient, les symptômes de la maladie variables d'un patient à l'autre (ex. un patient atteint de la maladie de Parkinson peut ne pas souffrir ou de manière très légère de dysarthrie), son stade, sa vitesse de progression, le type et la sévérité de la dysarthrie, la prise en charge thérapeutique des patients (médicamenteuse, chirurgicale, rééducation, etc.), les stratégies de compensation mises en place par le patient pour contourner les difficultés de production, son degré de fatigabilité sont autant de facteurs pouvant jouer sur la variabilité des phénomènes observables dans la parole dysarthrique. Ce constat implique naturellement que les analyses acoustiques et phonétiques soient réalisées sur de grandes populations de patients, ce qui est difficilement réalisable, si ces données étaient disponibles, par des experts humains considérant l'aspect chronophage de la tâche. Dans ce contexte, une grande part des études manuelles présentes dans la littérature reposent soit sur un nombre très restreint de patients soit sur une pathologie bien ciblée.

Le premier objectif des travaux du LIA sur la parole dysarthrique est de répondre à cette problématique en étudiant comment les outils du traitement automatique de la parole pourraient permettre de traiter de larges populations de patients et de focaliser l'attention des experts humains sur des zones de déviance bien identifiées du signal en vue d'analyses plus fines. Le second objectif de ces travaux est d'analyser le comportement des systèmes de traitement automatique de la parole suivant différents contextes d'observation et de les comparer : comparaison entre parole normale et parole dysarthrique, comparaison entre maladies des patients, comparaison entre styles de parole, etc. Ces objectifs sont plutôt différents des applications du traitement automatique que nous avons évoquées dans notre revue de la littérature (cf. chapitre 1), ce qui leur confère une certaine originalité. Néanmoins, nous verrons dans les résultats présentés et dans les perspectives de nos travaux qu'ils peuvent conduire à la même finalité en termes d'évaluation de la sévérité notamment.

Nous présenterons dans cette section les résultats les plus significatifs auxquels ces travaux ont mené ainsi que les limites et perspectives relevées. Ces travaux sont issus de la thèse d'Imed Laaridh dont la soutenance est prévue d'ici janvier 2017 et de mémoires de master Recherche et d'orthophonie. Ils ont été réalisés principalement dans le cadre des trois projets de recherche cités en début de section : les projets *VML*, *DesPhoAPady* et *TYPALOC*.

2.3.1 Corpus de parole dysarthrique

Une des activités majeures des laboratoires impliqués dans les projets *DesPhoAPady* et *TYPALOC* a été consacrée à la constitution de différents corpus francophones en vue de répondre aux problématiques posées dans chacun des projets de recherche. Le défi méthodologique était de définir des populations importantes de patients dysarthriques pour limiter l'influence de la variabilité intra- et inter-population dans un contexte clinique où l'enregistrement des locuteurs est, rappelons le, loin d'être facilité (disponibilité requise des patients et des cliniciens, consentement des patients, difficulté de maîtriser les périodes d'enregistrements, fatigabilité des patients, etc.). Nous avons, par conséquent, pris le parti de tirer bénéfice des enregistrements de patients déjà existants dans les différents centres hospitaliers partenaires et de réduire autant que possible la mise en place de nouvelles sessions d'enregistrements.

L'implication du LIA dans ce travail de constitution a été à deux niveaux :

- participer aux discussions de projets concernant le choix des populations, la définition des critères d'inclusion des patients, le choix des enregistrements (en fonction de la qualité sonore, la durée et autres caractéristiques fixées dans le cadre des projets), la définition d'une liste de critères pour la mise en place d'une évaluation perceptive des échantillons de parole (Grille d'Evaluation Perceptive de la Dysarthrie (G.E.P.D.) (Lhoussaine, 2012)) et la définition de la base de données informatique permettant la structuration, le stockage et l'accès facilité aux enregistrements et à toutes les informations disponibles (civiles, médicales, sessions d'enregistrement, annotations, etc.) s'y reportant (Ghio et al., 2012) ;

- l'enrichissement des corpus en fournissant pour chaque enregistrement sonore une segmentation en phonèmes issue du système d'alignement automatique contraint par le texte développé au LIA et adapté au traitement de la parole dysarthrique et une cartographie des déviances détectées automatiquement (méthodologie détaillée plus tard dans le document).

La spécificité du traitement de la parole dysarthrique dans le cadre de l'alignement phonétique a demandé une réflexion toute particulière sur la transcription orthographique manuelle requise pour sa réalisation et les règles d'annotation sur lesquelles cette dernière devait reposer. Deux optiques ont été envisagées : (1) privilégier la qualité de l'alignement automatique en définissant des conventions d'annotation conduisant à une transcription la plus proche phonétiquement de la production de parole traitée et à

une meilleure prise en compte des altérations dues à la dysarthrie ou (2) privilégier la qualité de la détection automatique des déviances (reposant sur les sorties de l'alignement automatique) et, par conséquent, trouver un compromis entre prise en compte a minima de ces altérations [1] et perturbation de l'alignement automatique. Le facteur "coût" en temps et ressources humaines n'étant pas négligeable dans la première option, nous avons opté pour la seconde et proposé les règles d'annotation fournies en annexes A et B pour la lecture et la parole spontanée respectivement. La faisabilité de la mise en place de ces conventions de transcription manuelle, notamment pour la tâche de lecture, a été validée par deux annotateurs experts du domaine sur un échantillon d'enregistrements après différentes phases de révision.

Ce travail collaboratif, impliquant phonéticiens, cliniciens et informaticiens, a conduit à l'élaboration des corpus suivants :

– le corpus *VML* est issu du projet du même nom. Il a été le premier corpus de parole dysarthrique mis à notre disposition par un établissement hospitalier. Il comprend 8 patients atteints de maladies lysosomales - maladies génétiques rares touchant le système nerveux chez l'enfant et l'adulte - dont la dysarthrie, de type mixte, est l'un des symptômes. Le type de pathologie explique le faible nombre de patients concernés ; à cette époque, seules 16 personnes étaient diagnostiquées pour ces maladies en France et 8 de ces personnes faisaient partie d'un protocole thérapeutique dédié mis en place par le service neurologique de l'hôpital la Pitié Salpétrière (Paris). Ces patients ont été enregistrés à l'hôpital sur une période de deux années, à intervalle plus ou moins régulier (six mois environ entre deux sessions d'enregistrements), donnant lieu à un nombre variable d'enregistrements par patient (entre 3 et 6). Ces enregistrements étaient programmés lors d'une journée de consultation au cours de laquelle les patients subissaient toute une batterie d'examens cliniques et de tests psycho-moteurs.

  Le protocole d'enregistrement comprenait différentes tâches dont la lecture du texte *le cordonnier* [2]. Un groupe de 6 sujets contrôles appareillés en âge a également été enregistré suivant le même protocole, sur une période d'1 mois, avec un espacement d'une semaine entre deux sessions d'enregistrement pour reproduire un nombre comparable d'enregistrements par sujet. La durée moyenne des enregistrements est de 60s pour les sujets contrôles et de 85s pour les patients avec des durées pouvant varier de 48 à 196s suivant le locuteur. Il est à noter que les enregistrements d'un patient sont incomplets en raison de son haut niveau de fatigabilité l'empêchant d'aller au bout du texte à lire. Par ailleurs, le protocole d'enregistrement a dû être interrompu pour un autre patient atteint de démence dans l'évolution de la maladie. Ces deux exemples sont typiques des difficultés pouvant être rencontrées dans la collecte de données cliniques.

1. Une représentation fidèle des altérations dans la transcription par un codage phonétique, par exemple, rendra inopérante la détection automatique d'anomalies.

2. Le texte comprend environ 550 phonèmes répartis entre 290 consonnes et 260 voyelles de la langue française.

Le corpus audio a été enrichi par différents matériaux : (1) une transcription orthographique manuelle en suivant les conventions de transcription citées ci-dessus, (2) une segmentation phonétique issue du système d'alignement automatique contraint par le texte du LIA s'appuyant sur la transcription orthographique, (3) une correction manuelle des frontières des phonèmes issues de la segmentation automatique, (4) une annotation manuelle (sur la base d'une évaluation perceptive et des informations fournies par le logiciel PRAAT (Boersma et Weenink)) réalisée par un expert phonéticien du caractère déviant de chaque phonème par rapport à une production prototypique (décision binaire) en indiquant le cas échéant le type d'altération relevé (bruit, trouble du voisement, distorsion spectrale, etc.) et (5) d'une évaluation perceptive portant notamment sur le grade global de sévérité de la dysarthrie (voir détails dans le corpus *DesPhoAPady* décrit ci-après). Le tableau 2.6 reporte pour chaque patient le nombre d'enregistrements associés, le grade global de sévérité (évalué sur le premier enregistrement) et le pourcentage de phonèmes annotés comme déviants par l'expert phonéticien (Laaridh et al., 2015a). Nous pouvons remarquer que les patients sont atteints d'une dysarthrie modérée (1.5) à très sévère (3.0) et que le pourcentage de phonèmes annotés comme déviants croît de manière plutôt linéaire avec le degré de sévérité.

Patients (genre H/F)	# d'enregistrements	Grade global de sévérité	% de phonèmes annotés comme déviants
PSN-GALE00 (H)	4	1.5	9.5
PSN-NGHI00 (H)	3	2.5	37.5
PSN-NRON00 (H)	5	1.5	15.3
PSN-NVAL00 (H)	6	2.1	30.2
PSN-NCHR00 (F)	5	1.6	16.5
PSN-NDIA00 (F)	5	3.0	76.3
PSN-GHEL00 (F)	4	1.5	15.2
PSN-GSAN00 (F)	3	2.5	33.5

TABLE 2.6 – *Informations relatives au corpus de parole dysarthrique VML.*

– le corpus *DesPhoAPaDy* (Fougeron et al., 2010) regroupe des enregistrements de 129 locuteurs [3], réalisés dans différents établissements hospitaliers (Hôpital Européen Georges Pompidou, Hôpital la Pitié-Salpétrière, Hôpital la Timone, les Hôpitaux d'Aix-en-Provence). Cet ensemble de locuteurs se répartit entre 99 patients atteints de dysarthrie et 30 sujets contrôles dont la tranche d'âge se situe entre 32 et 89 ans. Parmi les patients, nous pouvons distinguer 4 populations différentes en fonction de leur maladie neuro-dégénérative et du type de dysarthrie associé : maladie de Parkinson associée à une dysarthrie hypokinétique, ataxie cérébelleuse (atteinte du cervelet) associée à une dysarthrie ataxique, Sclérose Latérale Amyotrophique (SLA) associée à une dysarthrie mixte et deux cas de maladies

3. Le nombre de locuteurs précisé ici diffère de celui reporté dans (Fougeron et al., 2010) en raison du rejet au final de certains enregistrements.

lysosomales [4] (maladies génétiques rares) également associées à une dysarthrie mixte. Le protocole d'enregistrement reposait sur la lecture du texte *Le cordonnier* (cf. note de bas de page 2), avec pour consigne de lire le texte le plus naturellement possible. Les enregistrements ont été réalisés pour une grande majorité des enregistrements en chambre sourde lors d'une consultation en hôpital.
L'enrichissement de ce corpus comprend (1) la transcription orthographique manuelle de tous les enregistrements en suivant les conventions d'annotation de la parole lue mentionnées ci-dessus, (2) une évaluation perceptive réalisée par un jury de 11 experts selon différents critères définis dans le cadre du projet *DesPhoAPady* (Grille d'Evaluation Perceptive de la Dysarthrie (G.E.P.D.) (Lhoussaine, 2012)) dont le grade global de sévérité quoté sur une échelle de 0 (pas de dysarthrie) à 3 (dysarthrie sévère), (3) une segmentation phonétique issue du système d'alignement automatique contraint par le texte du LIA s'appuyant sur la transcription orthographique, (4) une cartographie des déviances au niveau phonème pour chaque enregistrement issue du système de détection automatique du LIA.

Le tableau 2.7 fournit par population des détails sur ce corpus : effectifs, répartition hommes/femmes, statistiques sur le grade de sévérité issues de l'évaluation perceptive. D'après les analyses menées dans (Lhoussaine, 2012) à partir de l'évaluation perceptive réalisée par le jury de 11 experts, les patients atteints de la maladie de Parkinson présentent globalement une dysarthrie légère (moyenne du grade de sévérité de 0.9), néanmoins l'écart-type (0.7) et les valeurs extrêmes (0.0 ;2.7) montrent des degrés de sévérité très variables, indicateurs d'un groupe de patients hétérogène. Les patients atteints d'ataxie cérébelleuse montrent plutôt une dysarthrie légère à modérée qui, d'après l'écart-type, pourraient être considérés comme un groupe plus homogène que précédemment (moyenne : 1.5 avec un écart type à 0.5). On peut tout de même observer des écarts entre les valeurs extrêmes également importants qui montrent la palette de variation possible sur ce type de population. Concernant les patients atteints de SLA, le degré de sévérité global moyen fait référence à une dysarthrie très marquée (moyenne : 1.9) avec un écart-type (0.7) relativement faible. Malgré des écarts entre les valeurs extrêmes encore très importants (0.5 ;3.0), cette population est considérée comme moyennement homogène. Un comportement tout à fait similaire est observé chez les patients atteints de maladie lysosomale, associés comme pour la SLA à une dysarthrie mixte.

– le corpus *TyPaLoc* (Meunier et al., 2016) est un sous-ensemble d'enregistrements extraits du corpus *DesPhoAPaDy*. Dédié à l'étude de la variabilité de la parole sur la base d'une comparaison - parole normale vs parole pathologique - et - parole lue vs parole spontanée - le choix des enregistrements a été dicté par la disponibilité d'une quantité de parole spontanée suffisante chez les patients représentés dans le corpus *DesPhoAPaDy*. Ce critère de sélection a conduit à une réduction drastique du nombre de locuteurs par population comme le montre le tableau 2.8 avec un total de 28 patients couvrant la maladie de Parkinson, l'ataxie cérébel-

4. Ces enregistrements proviennent initialement du projet et du corpus **VML**.

Pathologie	Effectifs	Grade global de sévérité	
	Hommes/Femmes	Moyenne (Ecart-type)	Min - Max
Sujets contrôles	15H / 15F	0.1 (0.2)	0.0 - 0.6
Maladie de Parkinson	23H / 8F	0.9 (0.7)	0.0 - 2.7
Ataxie cérébelleuse	13H / 9F	1.5 (0.5)	0.6 - 2.5
SLA	14H / 24F	1.9 (0.7)	0.5 - 3.0
Maladies lysosomales	4H / 4F	2.0 (0.6)	1.5 - 3.0

TABLE 2.7 – *Informations relatives au corpus de parole dysarthrique DesPhoAPaDy.*

leuse et la SLA. La tranche d'âge de ce corpus s'étale à présent de 32 à 81 ans. Concernant les valeurs du grade de sévérité global de la dysarthrie, toujours évalué sur la parole lue, ce sous-ensemble conserve les comportements observés sur le corpus *DesPhoAPady* avec des valeurs statistiques légèrement plus faibles notamment pour la population de patients atteints de la maladie de Parkinson qui devient plus homogène.
La population de sujets contrôles a été entièrement revue pour satisfaire les objectifs du projet et notamment de l'étude de la parole spontanée. Elle comprend une première sous-population de 6 seniors âgés de 63 à 82 ans, enregistrés dans le nord de la France et une deuxième sous-population de 6 juniors âgés de 29 à 47 ans et originaires du sud-est de la France, dont les enregistrements ont été extraits d'un corpus plus large de parole conversationnelle (Corpus CID - (Bertrand et al., 2008)).
Tous les locuteurs ont été enregistrés sur la tâche de lecture du texte *Le cordonnier* et sur une tâche de production de parole spontanée, soit lors d'un entretien mené par un chercheur ou un clinicien dans lequel ils devaient s'exprimer via un monologue sur leur quotidien, leur travail, un événement de leur vie, etc. (cas des patients et de la sous-population de personnes la plus âgée), soit lors d'une situation d'interaction en mode décontracté avec une deuxième personne (cas de la sous-population de sujets contrôles plus jeunes) durant laquelle ils évoquaient un événement particulier (seules les portions de parole narrative les plus longues ont été considérées pour se rapprocher au mieux du mode entretien). Il est à noter des différences très conséquentes dans les durées des productions de parole spontanée entre les populations de patients et les sujets contrôles. Ces différences sont principalement dues à la difficulté de faire parler certains patients en raison de blocages inhérents à leur maladie ou tout simplement de leur niveau élevé de fatigabilité.
D'une manière similaire au corpus *DesPhoAPady*, ce corpus a été enrichi par (1) une transcription orthographique manuelle de tous les enregistrements en suivant soit les conventions de transcription de la lecture mentionnées plus haut, soit celles de la parole spontanée, (2) d'une évaluation perceptive (pour les patients uniquement) déjà disponible pour la lecture et reproduite à l'identique avec le même jury d'experts sur la parole spontanée, (3) d'une segmentation en pho-

nèmes [5] et (4) d'une cartographie des déviances au niveau phonème pour chaque enregistrement. A cette base, il a été ajouté une segmentation en syllabes issue du système automatique SPPAS développé au LPL (Bigi, 2015) et une correction manuelle par un expert humain de toutes les segmentations en phonèmes. Concernant l'évaluation perceptive et la comparaison entre les deux styles de parole lue et spontanée, des notations moyennes quasi similaires ont été apportées par le jury d'experts, quelle que soit la population observée, et ce, considérant les différentes dimensions perceptives évaluées par le jury (degré de sévérité global, intelligibilité, altérations articulatoires, etc.) (Meunier et al., 2016).

Pathologie	Effectifs	Grade global de sévérité (sur la parole lue)		Durée parole (en sec.)	
	Hommes/Femmes	Moyenne (Ecart-type)	Min - Max	Lecture	Spontané
Sujets contrôles juniors	3H / 3F	-	-	48	229
Sujets contrôles seniors	3H / 3F	-	-	53	514
Maladie de Parkinson	6H / 2F	0.8 (0.4)	0.4 - 1.4	50.4	31.1
Ataxie cérébelleuse	4H / 4F	1.3 (0.4)	0.8 - 2.1	68	48
SLA	5H / 7F	2.0 (0.6)	0.9 - 2.9	83.2	68.6

TABLE 2.8 – *Informations relatives au corpus de parole dysarthrique TyPaLoc.*

La constitution de ces différents corpus a demandé un travail conséquent de réflexion sur leur composition (critères d'inclusion des patients, choix des enregistrements, types de maladies, types d'enrichissement nécessaires, processus d'enrichissement, etc.), de vérification et validation des enregistrements disponibles, de collecte des informations se rapportant aux patients (informations civiles, médicales, sessions d'enregistrement, ...), d'enrichissement, notamment au niveau des différentes annotations manuelles qui le composent et de structuration pour la mise en place d'une base de données permettant le stockage et la consultation des données associées.

Malgré les réflexions menées pour leur constitution au regard des objectifs visés et les précautions prises, ces corpus présentent encore un certain nombre de limites dues notamment au contexte clinique et à la contrainte de temps :

– taille limitée du corpus *TyPaLoc* pour l'étude de la parole spontanée en termes de nombre de patients et de durée par patient. La levée de cette limite demanderait un accès en plus grand nombre à des patients pour les enregistrer, loin d'être trivial dans le contexte clinique ; par ailleurs, la durée d'enregistrement par patient étant fortement corrélée à leur degré de fatigabilité, ce point est difficilement contournable ;

– l'hétérogénéité intra-population, en termes de degré de sévérité de la dysarthrie ou de variabilité dans les troubles de parole observés (débit très lent ou très rapide au sein d'une même population), pouvant rendre difficile l'interprétation de

5. Ici, la segmentation en phonèmes s'est appuyée sur une annotation manuelle préalable des unités inter-pauses, ie des séquences de parole séparées par des pauses silencieuses de plus de 250ms.

certains résultats, tout particulièrement, lorsque la taille de la population est petite ;

- les corpus ne sont pas forcément bien équilibrés en genre, ce qui peut rendre difficile l'interprétation de certains résultats, notamment basés sur des analyses acoustiques, dès lors qu'un déséquilibre est présent ; par ailleurs, vu la taille de certains corpus, il n'est pas envisageable de sous-diviser les populations en fonction du genre.

- la qualité des annotations, qui, pour une grande majorité d'entre elles, ont été réalisées par un seul expert sans contre-validation et ce, en raison du coût en temps et ressources humaines requis.

2.3.2 Système de détection des déviances

La détection d'anomalies consiste à rechercher des échantillons dans un ensemble de données qui ne sont pas conformes à un comportement donné. Ses domaines d'applications sont nombreux comme par exemple la détection d'intrusions sur un réseau informatique, la détection de fraudes, la surveillance médicale, la surveillance environnementale (données climatiques par exemple), la recherche de dysfonctionnement ou de défauts dans le comportement d'un système électronique ou mécanique, le diagnostic de pannes, etc. Comme souligné dans (Chandola et al., 2007), la détection d'anomalies peut s'avérer dans certains contextes d'application une tâche très difficile en raison de différents facteurs, qui peuvent très facilement être transposables au domaine de l'analyse de la parole pathologique :

1. la frontière entre le comportement normal et anormal peut être dans certains contextes imprécis et/ou difficilement définissable, ce qui contribue à des erreurs de détection (faux positifs/négatifs) ;

2. dès lors qu'une anomalie est associée à une action malicieuse (cas des fraudes par exemple), des stratégies de compensation peuvent être mises en place pour masquer cette dernière. Ce facteur peut être facilement transposable aux stratégies de compensation que les patients peuvent mettre en jeu pour masquer leurs troubles de la parole ou contourner les difficultés rencontrées même si cela ne se rapporte évidemment pas à une action malicieuse ;

3. dans le cadre de systèmes de détection d'anomalies basés sur un apprentissage supervisé, les données annotées, incluant des anomalies, nécessaires pour l'apprentissage de modèles et pour l'évaluation des systèmes peuvent ne pas être disponibles ou en quantité très limitée ;

4. les données peuvent contenir du bruit au sens large du terme, assimilable à des anomalies et, par conséquent, sujet à erreurs de détection.

Sur la base de certains de ces facteurs, deux principales "écoles" se distinguent dans la manière de répondre à la tâche de détection des anomalies. La première "école" propose un ensemble de techniques permettant de modéliser le comportement normal uniquement. Ainsi, en présence d'une nouvelle donnée à traiter, le système doit déterminer si celle-ci répond au modèle de normalité. Les données en lien avec des anomalies sont nécessaires uniquement pour l'évaluation de la pertinence des systèmes, ce qui limite la quantité requise. La seconde "école" repose sur un ensemble de techniques tenant compte à la fois du comportement normal et anormal dans leurs modèles et nécessitant, de ce fait, des données annotées en quantité suffisante pour chacun des comportements à modéliser.

Sur la base de ces deux "écoles", le LIA a développé ces dernières années deux systèmes de détection d'anomalies dédiés à la parole dysarthrique (Fredouille et Pouchoulin, 2011)(Laaridh et al., 2015a). La détection des anomalies étant réalisée au niveau du phonème, ces deux systèmes reposent sur une première étape de segmentation automatique. Les sections suivantes décrivent brièvement cette étape ainsi que les approches mises en oeuvre dans chacun des systèmes.

Segmentation automatique en phonèmes
La segmentation en phonèmes des enregistrements de parole est réalisée grâce à l'outil d'alignement automatique contraint par le texte développé au LIA. Cet outil utilise, de manière classique en entrée la séquence de mots prononcée dans chaque enregistrement ainsi qu'un lexique phonétisé présentant une variété phonologique de chaque mot basé sur un ensemble de 37 phonèmes de la langue française. La séquence de mots est le résultat d'une transcription orthographique manuelle réalisée par un humain (voir section 2.3.1 pour une discussion sur la transcription manuelle de la parole dysarthrique). L'alignement automatique repose sur un décodage du signal de parole par l'algorithme Viterbi basé sur des modèles statistiques - modèles de Markov cachés, HMM - associés aux phonèmes. Ici, chaque phonème est représenté par un HMM à trois états, indépendant du contexte construit par estimation du maximum de vraisemblance à partir de 200 heures d'enregistrements radiophoniques français issus du corpus ESTER (Galliano et al., 2005). Une adaptation de type Maximum A Posteriori (MAP) à 3 itérations est appliquée afin de créer des modèles dépendants du locuteur. Ce processus d'alignement résulte en une segmentation temporelle des enregistrements avec, pour chaque phonème produit, ses frontières de début et de fin dans le signal.

Approche à 1 modèle : modélisation de la parole normale
La première approche de détection d'anomalies proposée par le LIA (Fredouille et Pouchoulin, 2011) repose sur l'idée d'attribuer un score de normalité à un segment de parole associé à un phonème, étant donné un *modèle* de parole normale. En effet, sur la base des sorties de l'alignement contraint par le texte, il s'agit de calculer un score normalisé et d'apprécier, sur la base d'une échelle de normalité, le degré de déviance de chaque phonème issu de la segmentation automatique.

Dans ce contexte, le score normalisé est défini comme suit :

$$LL_p^{norm}(y_p) = LL_p^{Constrained}(y_p) - LL^{Unconstrained}(y_p) \tag{2.1}$$

où $LL_p^{norm}(y_p)$ est le score acoustique normalisé du phonème p calculé sur le segment de parole associé y_p. $LL_p^{Constrained}(y_p)$ est le score acoustique attribué au phonème p par le processus d'alignement automatique contraint par le texte ie la log-vraisemblance calculée entre le segment de parole y_p et le modèle HMM associé au phonème p. Afin de pouvoir comparer les scores acoustiques entre phonèmes, une normalisation du score $LL_p^{Constrained}(y_p)$ est appliquée, reposant sur le second score $LL^{Unconstrained}(y_p)$. Ce dernier représente la meilleure séquence d'états issus des modèles HMM qu'un algorithme de type Viterbi peut produire, considérant aucune contrainte sur le texte prononcé, ni contrainte de temporalité entre états des HMM. Ce processus pourrait être assimilé à un alignement phonétique non contraint par le texte, appliqué à un segment de parole donné (ici y_p). Dans ce contexte, le score de normalisation $LL^{Unconstrained}(y_p)$ peut être défini de la manière suivante :

$$LL^{Unconstrained}(y_p) = \sum_{t=1}^{T_{y_p}} \arg\max_{k \in K} LL_k(y_p^t) \tag{2.2}$$

avec K l'ensemble des états des modèles HMM disponibles, T_{y_p} le nombre de trames du segment de parole y_p et $LL_k(y_p^t)$ la log-vraisemblance entre la t-ième trame de y_p et le k-ième état des modèles HMM.
Plus les scores $LL_p^{Constrained}(y_p)$ et $LL^{Unconstrained}(y_p)$ sont proches, plus le score normalisé ($LL_p^{norm}(y_p)$) est proche de la valeur 0. Dans ce cas, les processus d'alignement contraint et non contraint par le texte convergent vers le même phonème p. A l'inverse, plus les scores $LL_p^{Constrained}(y_p)$ et $LL^{Unconstrained}(y_p)$ sont éloignés, plus le score normalisé ($LL_p^{norm}(y_p)$) tend vers $-\infty$. Dans ce deuxième cas, cela aura pour signification que la séquence d'états optimale fournie par le processus d'alignement non contraint par le texte sera différente de celle produite par le processus d'alignement contraint par le texte. En d'autres termes, le signal acoustique du segment de parole associé au phonème p ne sera pas celui attendu.

La dernière étape du système de détection des anomalies consiste à exploiter les scores normalisés assignés à chaque phonème individuellement, en déterminant de manière automatique si le phonème doit être considéré comme normal ou anormal du point de vue acoustique. Cette décision est établie grâce à un indice de normalité attribué à chaque phonème par projection de leur score acoustique normalisé sur une échelle de référence. Cette échelle est construite à partir d'une population de sujets sains produisant de la parole considérée comme normale sur laquelle un ensemble de scores acoustiques normalisés sont calculés. A partir de ces scores, des valeurs de scores minimum, maximum et médian sont estimées et utilisées pour définir de manière très basique une échelle de référence.
Les scores acoustiques normalisés calculés à partir d'un signal de parole test à évaluer sont alors projetés sur cette échelle de référence, résultant en un indice de normalité compris dans un intervalle [-100 ;100].

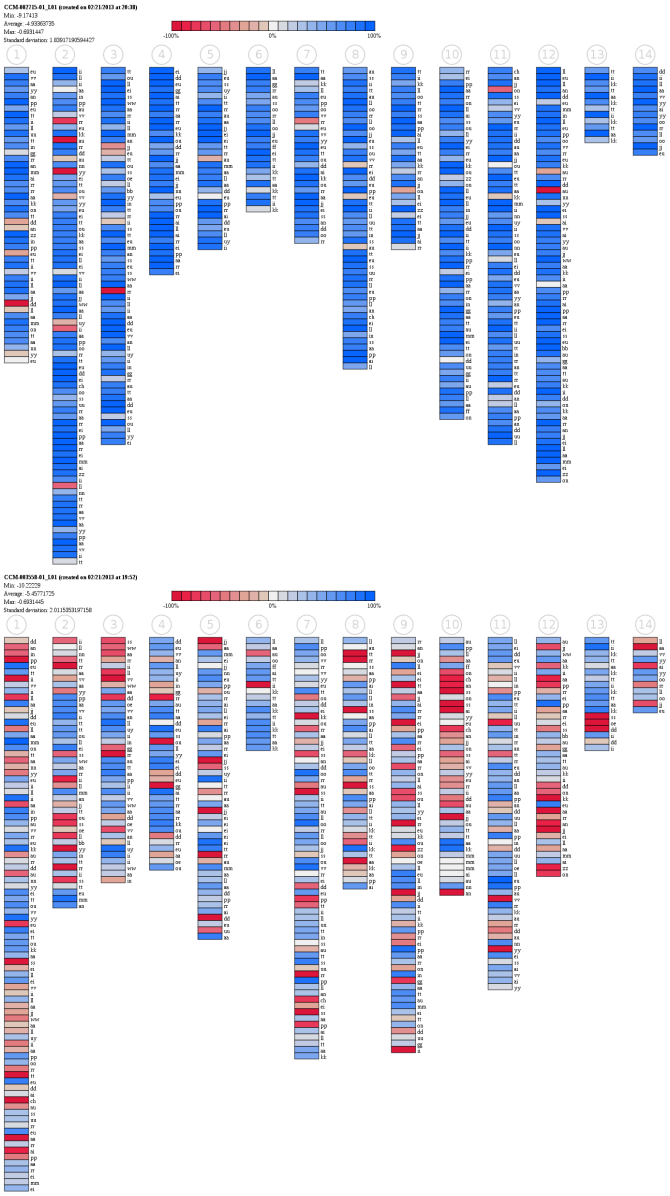


FIGURE 2.1 – *Exemples de cartographie représentant pour chaque phonème leur indice de normalité. (En haut) Cartographie d'un sujet témoin. (En bas) Cartographie d'un patient atteint de maladie de Parkinson dont la dysarthrie a été évaluée à 2.7 (sur 3) sur l'échelle de sévérité globale.*

En vue de faciliter la lecture des résultats issus de cette projection, une cartographie est associée à chaque enregistrement, permettant de représenter graphiquement grâce à une palette de couleurs l'ensemble des phonèmes et leurs indices de normalité. Cette représentation permet ainsi de visualiser très rapidement les zones de déviance (accumulation de plusieurs phonèmes présentant des indices de normalité faible) et de comparer, par exemple, des cartographies établies sur des productions de parole différentes issues d'un même patient. La figure 2.1 fournit deux exemples de cartographie

associés pour la première (cartographie du haut) à un sujet contrôle, la seconde (cartographie du bas) à un patient atteint de la maladie de Parkinson dont la dysarthrie a été évaluée à 2.7 sur l'échelle de sévérité globale par le jury d'experts.

Approche à deux modèles : modélisation de la parole normale et déviante
La deuxième approche de détection d'anomalies mise en place dans ces travaux a pour vocation de tenir compte simultanément des informations propres à la parole normale et celles relatives à la parole "anormale" contrairement à l'approche précédente, focalisée uniquement sur la parole normale (Laaridh et al., 2015a). Elle repose sur une classification supervisée des phonèmes sur la base de deux classes - phonème normal et phonème anormal. Dans chaque classe, les phonèmes sont caractérisés par un ensemble de paramètres jugés pertinents pour la tâche visée. Ces paramètres sont essentiellement tirés de l'alignement automatique de la parole et se déclinent ainsi, considérant un phonème p et son segment de parole associé y_p :

- la durée du segment y_p associé au phonème p, exprimée en terme de nombre de trames de 10ms ;
- le nombre de trames de y_p pour lesquelles la recherche du meilleur état au sein des modèles HMM conduit à un état du phonème p ;
- le score acoustique du phonème p' émanant de la recherche du meilleur modèle HMM sur la base du segment y_p ; si p est le meilleur phonème, le second dans l'ordre est considéré à sa place ;
- la catégorie phonétique de p' ;
- le score acoustique du second meilleur phonème p'' correspondant au segment y_p ; si p est l'un des deux meilleurs phonèmes, le troisième est considéré à sa place ;
- la catégorie phonétique de p'' ;
- le score acoustique de p et son rang par rapport aux scores de tous les modèles HMM.

La tâche de classification repose sur une méthode à base de SVM (Support Vector Machines) associées à des noyaux polynomiaux (Vapnik, 1995; Scholkopf et Smola, 2001), le tout disponible dans l'outil SVMlight (Joachims, 1999). Dans ce travail, la méthode de classification porte sur les deux classes qui nous concernent : les phonèmes normaux et anormaux (anomalies). Chaque phonème est caractérisé par l'ensemble de paramètres décrit ci-dessus. Différents modèles de classes ont été estimés, tenant compte du genre des locuteurs et d'un ensemble fermé de catégories phonétiques. En effet, nous avons fait le choix de distinguer les consonnes sourdes, consonnes sonores, voyelles orales, voyelles nasales produites par des hommes et par des femmes en vue d'apprendre des modèles de classes différenciés et d'affiner le processus de classification (Laaridh et al., 2015a).

2.3.3 Comparaison des deux approches de détection d'anomalies

Les deux approches de détection des anomalies décrites dans la section précédente requérant des données annotées en anomalies pour l'apprentissage du modèle de pa-

role anormale pour la détection à base de SVM et pour leur évaluation [6], cette comparaison repose sur le seul corpus annoté manuellement à notre disposition, le corpus *VML* (voir détails section 2.3.1).

Le tableau 2.9 fournit des résultats comparatifs, issus de mesures de rappel et de précision calculées sur l'ensemble des anomalies détectées automatiquement pour chacune des approches. Ces résultats sont issus d'une comparaison plus complète entre les deux approches conduite dans (Laaridh et al., 2015a). Les valeurs sont données par patient et calculées sur l'ensemble des enregistrements disponibles pour chacun d'entre eux (cf. tableau 2.6 pour le détail du nombre d'enregistrements par patient). Finalement, il est à noter que le comptage des anomalies détectées automatiquement tient compte d'une fenêtre intégrant le phonème observé et les deux phonèmes adjacents. En effet, si un phonème est considéré comme anormal par l'annotateur humain, une concordance sera accordée si le phonème en question ou ses deux voisins sont considérés comme une anomalie par le système automatique, ceci afin de tenir compte d'un décalage de quelques trames possibles dans la segmentation automatique en phonèmes (stratégie 2 d'évaluation dans (Laaridh et al., 2015a)).

Nous pouvons constater que l'approche à deux modèles obtient, dans une grande majorité, de meilleurs résultats, notamment en termes de rappel (moyenne globale de 0.81 vs 0.74). Par ailleurs, nous soulignons que les résultats des tests réalisés sur la population de sujets contrôles du corpus *VML*, exprimés en termes de taux d'erreurs de détection d'anomalies [7] a conduit à des taux de 8.4% et 6.9% respectivement pour les approches à un modèle et à deux modèles.

Cette supériorité de l'approche à deux modèles semble reposer sur la meilleure prise en compte des anomalies dans le processus de détection. Elle peut également être le résultat d'une modélisation plus fine de la parole normale entre les deux approches considérant l'ensemble des paramètres utilisés par l'approche à deux modèles pour caractériser chaque phonème.

En considérant à présent uniquement la meilleure approche (à deux modèles), l'analyse des résultats montre des mesures de rappel très satisfaisantes sur la population de femmes avec des valeurs au delà de 0.8. Concernant les hommes, les mesures de rappel sont plus variables avec deux patients pour lesquels les mesures sont excellentes (proches de 0.9), un présentant une valeur un peu moindre (0.72) et le dernier présentant une valeur très inférieure (0.37).

Les mesures de précision sont, quant à elles, bien inférieures (proches de 0.6) ; le système détecte davantage d'anomalies qu'il ne le devrait comparé à l'annotation manuelle. Il est, par conséquent, plus sévère (à tord ou à raison) que l'annotateur humain. Ce comportement se retrouve, d'ailleurs, dans les taux d'erreurs de détection relevés chez la population de sujets contrôles, à hauteur de 7%. Etant donné que les enregistrements des sujets contrôles n'ont pas été annotés, nous pourrions envisager que

6. au vu du peu de données annotées disponibles, la technique du *Leave-one-out* est utilisée pour l'apprentissage des modèles et l'évaluation du système à base de SVM.

7. Les enregistrements de la population de sujets contrôles n'ayant pas été annotés en termes de déviance, seul un taux d'erreurs de détection peut être applicable ici.

parmi ces 7%, une part d'anomalies, considérées actuellement comme des faux positifs, ne le soient pas en réalité. En effet, de la même manière que nous avons reporté dans la section dédiée à la dysphonie et au phonetic labelling (voir section 2.2.2) que des phonèmes pathologiques pouvaient être relevés chez les sujets contrôles, certains phonèmes pourraient effectivement être considérés comme déviants ici par un expert humain, et par conséquent, en concordance avec le système. Pour finir, il est intéressant de constater que les mesures de rappel ou de précision les plus importantes sont attribuables aux patients présentant des degrés de sévérité globaux les plus élevés (au delà de 2). Ce sont également ces patients pour lesquels le système détecte le plus grand nombre d'anomalies.

Patients : Corpus *VML*	Approche à 1 modèle : parole normale		Approche à 2 modèles : parole normale et anormale	
	Rappel	Précision	Rappel	Précision
PSN-GALE00 (H)	0.36	0.30	**0.37**	**0.52**
PSN-NGHI00 (H)	0.77	**0.77**	**0.90**	0.68
PSN-NRON00 (H)	**0.76**	0.53	0.72	**0.57**
PSN-NVAL00 (H)	0.73	0.64	**0.89**	**0.65**
Moyenne Hommes	0.65	0.56	**0.72**	**0.61**
PSN-NCHR00 (F)	0.79	**0.45**	**0.85**	0.42
PSN-NDIA00 (F)	0.87	0.98	**1.00**	0.98
PSN-GHEL00 (F)	0.79	0.55	**0.80**	**0.58**
PSN-GSAN00 (F)	0.88	**0.63**	**0.90**	0.60
Moyenne Femmes	0.83	0.65	**0.89**	0.65
Moyenne globale	0.74	0.61	**0.81**	**0.63**

TABLE 2.9 – *Résultats comparatifs des deux approches de détection des anomalies : approche à base d'un modèle unique de parole normale et approche à base de deux modèles de parole normale et anormale estimés sur les enregistrements annotés manuellement des patients du corpus VML. En gras, les valeurs optimales issues de la comparaison.*

2.3.4 Validation de l'approche à deux modèles sur corpus non annoté

Malgré le fait que le corpus *DesPhoAPady* ne soit pas annoté en déviances/anomalies au niveau du phonème, nous avons voulu tirer partie du grand nombre de patients qui le composent en vue de valider le comportement du système de détection à deux modèles.

Etude de la relation entre taux d'anomalies détectées et degré de sévérité global de la dysarthrie

La première étude que nous avons menée a pour objectif d'observer la relation entre les taux d'anomalies détectées par le système automatique et le degré de sévérité global moyen attribué par le jury d'experts sur chacun des enregistrements des patients du corpus *DesPhoAPady*, illustrée en figure 2.2. Sur cette figure, les populations de patients femmes (figure du haut) et hommes (figure du bas) ont été séparées pour une

meilleure visibilité. Ces figures montrent une relation très forte entre ces deux mesures automatique et perceptive. Ce constat est appuyé par le taux de corrélation de Pearson calculé sur chacune des populations, qui atteint 0.89 et 0.86 pour les populations de patients femmes et hommes respectivement. L'étude détaillée des taux de corrélation par maladie (disponible dans (Laaridh et al., 2015a)) montre que ces taux sont stables quels que soient la maladie et le genre observé (autour de 0.9 et plus) excepté pour la population de femmes atteintes d'ataxie cérébelleuse pour laquelle le taux de corrélation chute à 0.52 (le taux de corrélation pour les hommes étant de 0.81). Nous n'avons pas d'explication à l'heure d'aujourd'hui à ce comportement très particulier. Pour finir, le taux d'erreurs de détection d'anomalies [8] relevé sur la population de sujets contrôles du corpus *DesPhoAPady* s'élève à 6.6%. Ce taux est tout à fait comparable à celui de 6.9% relevé sur le corpus *VML*.

Qualité de l'alignement automatique et détection des anomalies
La deuxième étude que nous avons menée repose sur l'observation conjointe du comportement de l'alignement automatique et de la détection des anomalies. Sur la base des corpus *VML* et *TyPaLoc* pour lesquels des corrections de la segmentation automatique en phonèmes sont disponibles, l'idée ici est d'observer le comportement de la détection des anomalies en fonction des erreurs de segmentation en comparant le corpus *VML* annoté en anomalies et le corpus *TyPaLoc* non annoté sur ce type d'information (Laaridh et al., 2015b)(Laaridh et al., 2016a).

La qualité du système d'alignement automatique est mesurée sur la base des décalages temporels de deux points d'ancrage sur le phonème - début et point central - entre les segmentations automatique et manuelle. Emprunt de la littérature (Bürki et al., 2008), nous considérons que deux valeurs d'un point d'ancrage entre segmentation automatique et manuelle sont en concordance dès lors que le décalage ne dépasse pas un intervalle [-20ms ;+20ms].

Nous présentons ici uniquement les résultats sur le point d'ancrage de début des phonèmes obtenus sur les corpus *VML* et *TyPaLoc* publiés dans (Laaridh et al., 2016a). Nous pouvons relever que 85% des phonèmes produits par les sujets contrôles et alignés automatiquement sont en concordance avec la segmentation manuelle contre 79%, 71%, 71% et 64% pour les patients atteints de la maladie de Parkinson, d'ataxie cérébelleuse, de maladies lysosomales (*VML*) et de SLA respectivement. Le plus faible taux de concordance est attribué aux patients atteints de SLA, qui rappelons le, présentent également le degré de sévérité le plus élevé. D'ailleurs, nous pouvons remarquer sur la figure 2.3, illustrant la répartition des phonèmes suivant le décalage [9] en fonction des différentes populations, que les patients atteints de SLA atteignent le plus fort taux de phonèmes présentant des décalages au delà de 60ms comparé à la segmentation manuelle.

Le tableau 2.10 fournit, pour le corpus *VML*, (1) la distribution des phonèmes, (2)

8. Ce taux est estimé sur l'hypothèse qu'aucune anomalie ne devrait être présente sur la population contrôle.

9. le décalage est exprimé ici en valeur absolue ie la barre des 10 inclut à la fois les décalages de -10ms et de +10ms entre la segmentation automatique et manuelle.

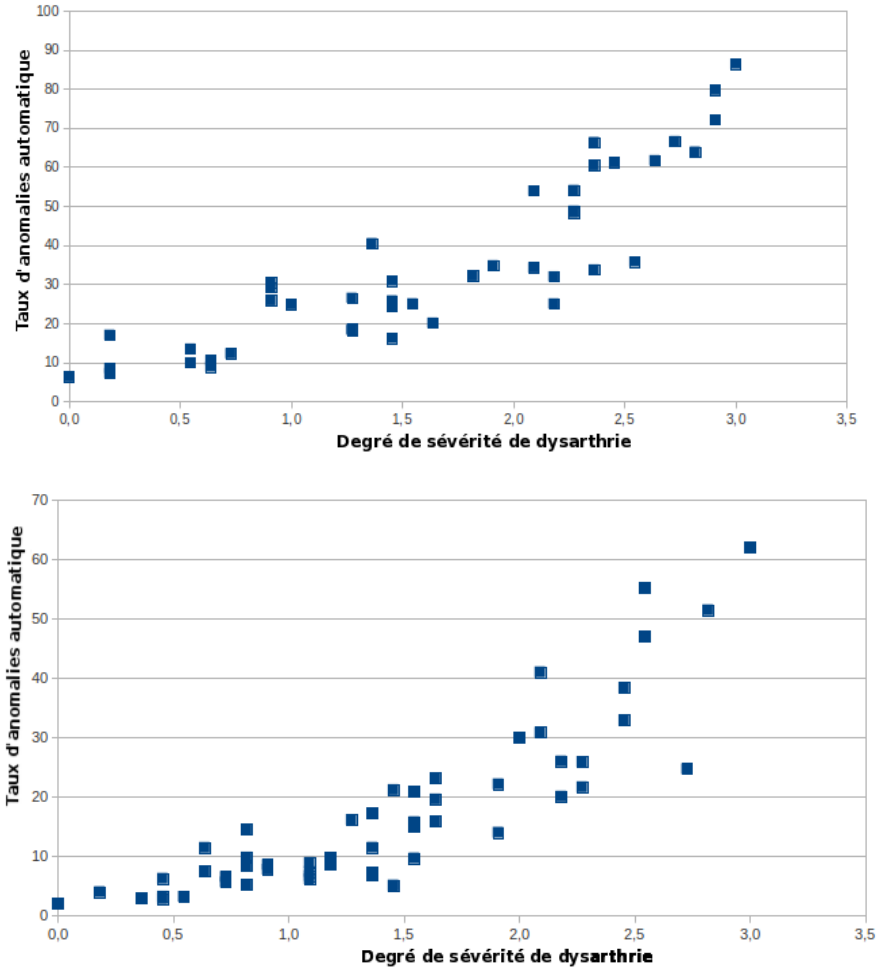


FIGURE 2.2 – *Relation entre taux d'anomalies détectées et le degré de sévérité global de la dysarthrie chez les patients du corpus non annoté DesPhoAPady. (En haut) Population de femmes. (En bas) Population d'hommes.*

celle des anomalies manuellement annotées, (3) celle des anomalies automatiquement détectées et (4) celle des anomalies automatiques concordantes avec l'annotation manuelle, en fonction des décalages en valeurs absolues des débuts des phonèmes entre segmentation manuelle et automatique. Les mesures de rappel et de précision sont indiquées par décalage [10]. Nous pouvons remarquer que 44% des anomalies détectées par le système automatique se situent hors de l'intervalle [-20ms ;+20ms] dont 29% au delà de 60ms. Le système détecte, par conséquent, un grand nombre d'anomalies sur des phonèmes présentant des erreurs importantes d'alignement. Au vu des paramètres utilisés pour caractériser les phonèmes dans l'approche de détection automatique, ce com-

10. Les valeurs de rappel et de précision sont ici plus faibles que celles fournies dans le tableau 2.9 du fait que les concordances sont données strictement au niveau du phonème sans tenir compte des phonèmes adjacents (stratégie 1 d'évaluation).

portement n'est pas surprenant. Nous pouvons noter que parmi ces 29% d'anomalies détectées, seules 35.8% sont concordantes avec l'annotation manuelle malgré une mesure de rappel plus élevée à 0.61. Nous noterons également que l'annotation manuelle ne considère d'ailleurs que 29% d'anomalies dans cette tranche des + de 60ms, considérant l'ensemble des anomalies étiquetées par l'expert humain. Ces résultats montrent que le système détecte à tort un très grand nombre d'anomalies (faux positifs), que nous pouvons probablement expliquer par les erreurs d'alignement. La question que nous pouvons nous poser ici est : **quels phénomènes dans le signal de parole peuvent perturber le fonctionnement de l'alignement automatique, amenant à de tels décalages (au delà de 60ms) sans pour autant conduire l'expert humain à annoter les phonèmes comme des anomalies ?** La première hypothèse que nous pouvons émettre et qui reste à vérifier serait qu'un phénomène d'altération global au signal de parole ne serait pas reporté par l'expert humain (ralentissement notable du débit par exemple) sur chacun des phonèmes mais pourrait en revanche perturber le système d'alignement, entraînant l'augmentation des faux positifs.

Nous nous intéressons maintenant à l'intervalle [-20ms ;+20ms] pour lequel les phonèmes sont considérés comme bien alignés. Ici, la détection automatique des anomalies repose non plus sur une erreur d'alignement mais bien sur des altérations sur la forme acoustique du phonème produit et traité par le système. Nous pouvons relever dans le tableau 2.10 que le système automatique détecte environ 18.5% d'anomalies sur la totalité des phonèmes bien alignés contre environ 10% pour l'expert humain. Parmi ces 18.5%, 31% sont concordantes avec l'expert humain avec un rappel de 0.56 (cette mesure atteignant 0.63 pour des décalages d'alignement quasi nuls). Le système semble donc capable de détecter une partie des anomalies annotées par l'expert humain sur la seule base des distorsions acoustiques. Malheureusement, il reste toujours très "sévère", détectant à tort un grand nombre d'anomalies. Ces résultats soulèvent plusieurs questions : **quels phénomènes dans la forme acoustique des phonèmes bien alignés peuvent amener le système de détection à surévaluer le nombre d'anomalies ? Parmi le type d'anomalies annotées par l'expert humain, le système est-il plus apte à détecter certaines altérations acoustiques que d'autres ?**

La dernière étude que nous avons menée consiste à comparer le comportement du système de détection sur le corpus annoté en anomalies *VML* et le corpus non annoté *TyPaLoc*. Nous pouvons indiquer que les taux de phonèmes détectés comme anomalies par le système automatique sur le corpus *TyPaLoc* sont de 5.4% pour les sujets contrôles, 10.6% pour les patients atteints de la maladie de Parkinson, 20.6% d'ataxie cérébelleuse, 35.8% de SLA contre 23.5% pour le corpus *VML*. Ces résultats restent tout à fait cohérents avec les degrés globaux de sévérité observés dans chacune des populations.
Toujours en référence aux erreurs d'alignement entre segmentation manuelle et automatique, la figure 2.4 fournit pour chaque décalage (en valeurs absolues) la fréquence des anomalies détectées par le système sur les différentes populations. Nous pouvons relever que ces fréquences sont de 77%, 64%, 60% et 55% pour les populations des sujets contrôles, des patients atteints de Parkinson, d'ataxie cérébelleuse et de SLA respectivement dans l'intervalle [-20ms ;+20ms], contre 56% mentionnés précédemment pour

le corpus *VML*. Ces distributions sont en adéquation avec les taux d'erreurs d'alignement en fonction des décalages reportés plus haut pour chacune des populations (cf. figure 2.3). En effet, plus le nombre d'erreurs d'alignement augmente, plus le système détectera des anomalies dans les zones d'erreurs, augmentant ainsi le taux d'anomalies globalement détectées.

La comparaison des différentes populations montre des comportements cohérents en adéquation avec les degrés de sévérité de la dysarthrie qui les caractérisent et les observations faites sur le corpus *VML*. Elle semble supporter l'hypothèse que le comportement du système automatique de détection des anomalies observé précédemment sur le corpus annoté soit transposable aux autres pathologies du corpus *TyPaLoc*.

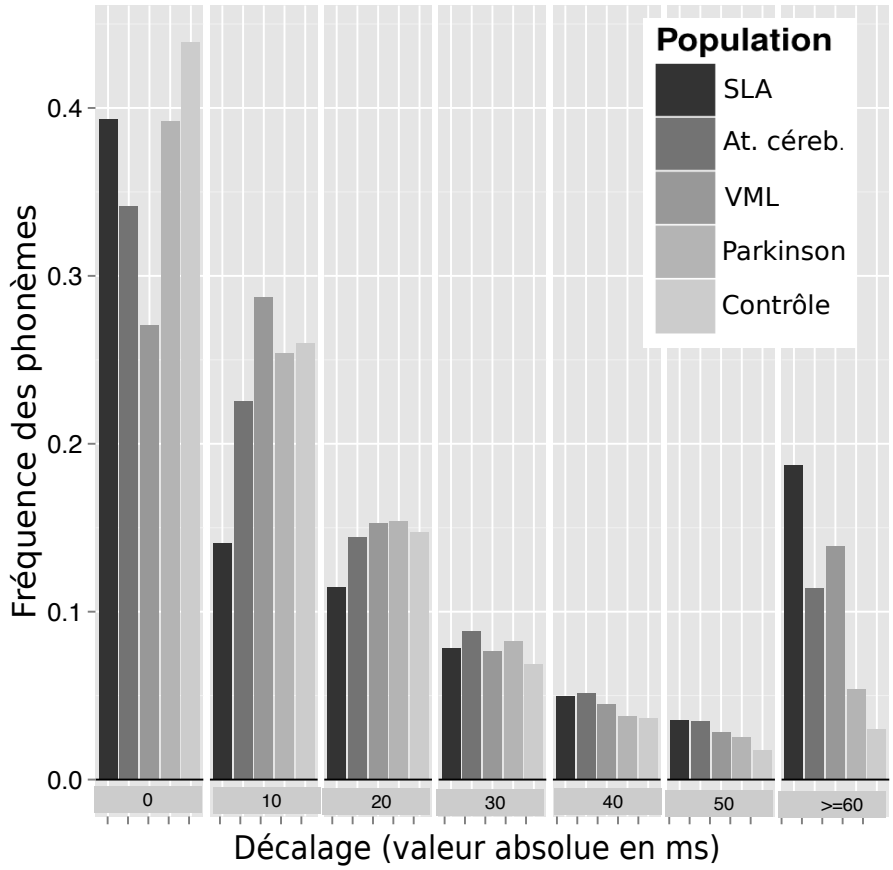


FIGURE 2.3 – *Distribution des phonèmes issus des corpus VML et TyPaLoc en fonction de la valeur de décalage du début des phonèmes entre segmentation automatique et manuelle. Décalage exprimé en valeur absolue - ex. décalages de -10ms et +10ms représentés par la barre des 10.*

|Décalage|	≥60ms	=50ms	=40ms	=30ms	=20ms	=10ms	=0ms	>20ms	[0 ;20ms]
Nb. Phonèmes	1263	259	410	696	1392	2614	2465	2628	6471
Nb. Anomalies Automatiques	620	92	100	127	199	400	595	939	1194
Nb. Anomalies Manuelles	363	61	70	81	132	228	310	575	670
Nb. Anomalies Concordantes	222	32	32	29	59	120	194	315	373
Rappel	0.61	0.53	0.46	0.36	0.48	0.53	0.63	0.55	0.56
Précision	0.36	0.35	0.32	0.23	0.30	0.30	0.33	0.34	0.31

TABLE 2.10 – *Distribution (1) des phonèmes, (2) des anomalies manuellement annotées, (3) des anomalies automatiquement détectées et (4) celles concordantes avec l'annotation manuelle en fonction des décalages en valeurs absolues des débuts des phonèmes entre segmentation manuelle et automatique pour le corpus annoté VML.*

2.3.5 Apport de la détection automatique des anomalies

Nous venons de voir dans les deux sections précédentes que le système de détection automatique des anomalies obtient des résultats plutôt prometteurs en terme de mesures de rappel sur le corpus *VML* tout en présentant des mesures de précision relativement faibles, indiquant une sur-détection des anomalies comparée à un expert humain. Néanmoins, malgré ce déséquilibre, il est intéressant d'observer le comportement du système et de le comparer suivant différents contextes d'application. Une première étude préliminaire menée sur l'approche à un seul modèle appliquée au corpus *VML* avait permis par exemple de mettre en évidence des comportements particuliers du système lors d'une analyse en bandes de fréquences en distinguant certaines catégories phonétiques (Fredouille et Pouchoulin, 2011) (Fredouille et Pouchoulin, 2012).

Nous nous intéresserons ici à une étude plus récente impliquant l'approche à deux modèles évaluée dans les sections précédentes (Laaridh et al., 2016b). Cette étude a pour vocation de comparer le comportement du système de détection d'anomalies sur les enregistrements de parole lue et parole spontanée. Les taux d'anomalies détectées par le système sont donnés dans le tableau 2.11 par population pour ces deux styles de parole. Ce dernier montre que les sujets contrôles présentent davantage d'anomalies détectées sur la parole spontanée (+8.3% en absolu). Cette augmentation est probablement liée au fait que les modèles de phonèmes normaux et anormaux, utilisés par le système de détection, ont été appris sur de la parole lue exclusivement. En effet, la parole spontanée peut présenter plus de variabilité acoustique qui est, à la fois, atypique - comparée ici à la parole lue - et non pathologique due à un débit de parole plus rapide, aux faux départs, hésitations ainsi qu'à des phénomènes de réduction plus fréquents dans le cadre de la parole spontanée.

Si l'on s'intéresse maintenant aux populations dysarthriques, nous pouvons tout d'abord constater que les différences en termes de taux d'anomalies entre la parole lue et spontanée ne sont pas aussi importantes qu'avec la population de sujets contrôles. Ces différences décroissent, d'ailleurs, avec l'augmentation du degré de sévérité de la

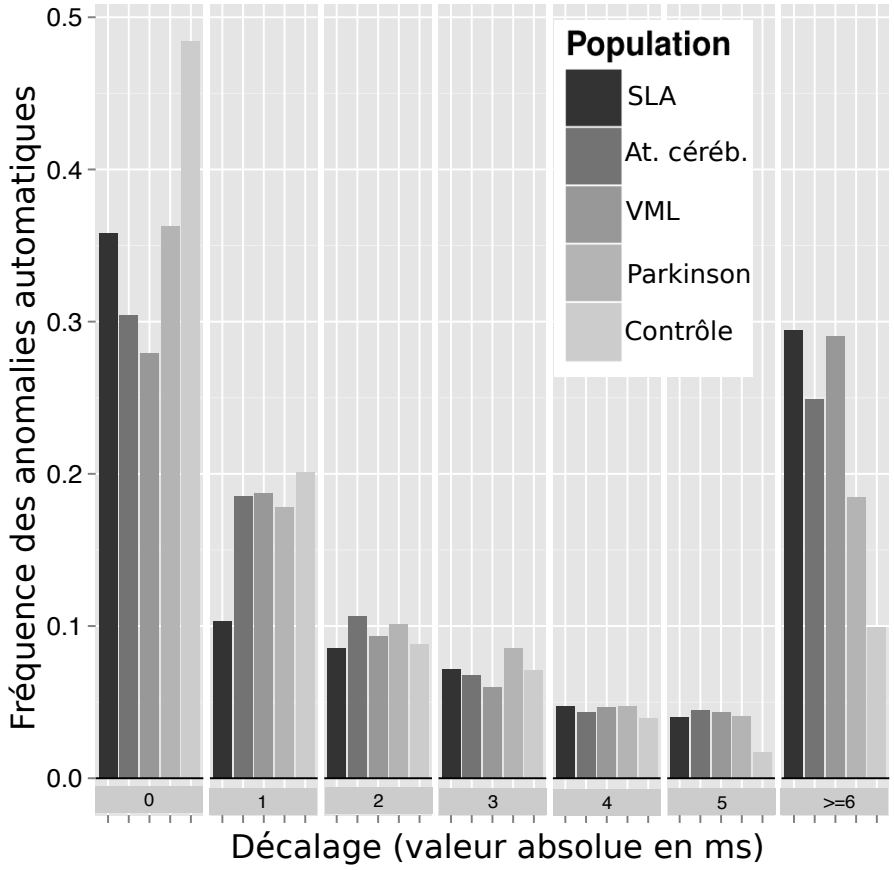


FIGURE 2.4 – *Distribution des anomalies détectées automatiquement sur les corpus VML et TyPaLoc en fonction de la valeur de décalage du début des phonèmes entre segmentation automatique et manuelle. Décalage exprimé en valeur absolue - ex. décalages de -10ms et +10ms représentés par la barre des 10.*

dysarthrie des patients jusqu'à s'inverser pour les patients atteints de SLA. En effet, nous pouvons constater une augmentation de +7.2% pour les patients atteints de la maladie de Parkinson, +3.8% pour ceux atteints d'ataxie cérébelleuse et une baisse de -3.9% pour ceux atteints de SLA. Si l'explication donnée pour les sujets contrôles reste valide ici - apprentissage des modèles sur de la parole lue -, il est intéressant de constater que l'impact est plus limité en présence de dysarthrie voire nulle pour les SLA. L'hypothèse que nous émettons à ce sujet est que **les sujets sains adaptent leur production selon le style de parole conduisant à une grande variabilité dans la production de parole. Les patients dysarthriques, quant à eux, peuvent perdre graduellement cette capacité d'adaptation aux différents styles de parole au fur et à mesure que le degré de sévérité de la dysarthrie augmente**. L'observation individuelle et par population des différences de taux entre la parole lue et spontanée par locuteur du corpus *TyPaLoc* par le biais de la figure 2.5 tend à montrer que le degré de sévérité n'est pas le seul facteur à prendre en compte et que la pathologie semble jouer également un rôle dans

cette distinction entre styles de parole (présence de clusters plutôt caractéristiques de chaque population sur la figure).

Population	Parole lue	Parole spontanée
Maladie de Parkinson	10.6	17.8
Ataxie Cérébelleuse	20.6	24.4
SLA	35.8	31.9
sujets contrôles	5.4	13.7

TABLE 2.11 – *Taux d'anomalies moyens (%) par population et style de parole calculés sur tous les phonèmes du corpus TyPaLoc.*

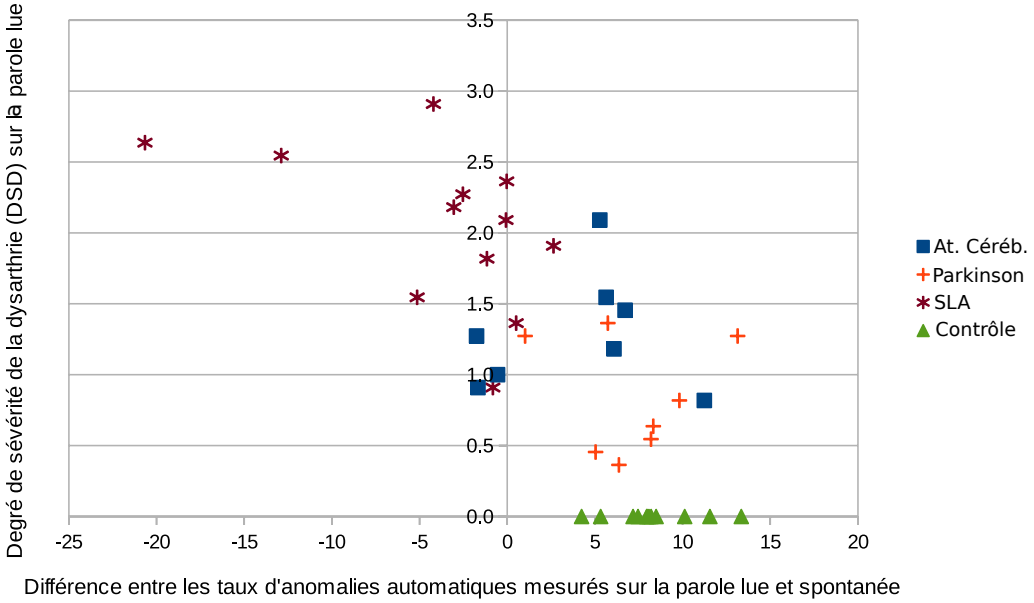


FIGURE 2.5 – *Différence des taux d'anomalies détectées sur la parole lue et spontanée en fonction du degré de sévérité des locuteurs du corpus TyPaLoc.*

2.3.6 Perception des déviances

Le dernier volet que nous voulions aborder ici repose sur le travail réalisé par Laura Pianelli et Laura Restivo dans le cadre de leur mémoire d'orthophonie encadré par Christine Meunier et moi-même sur la perception des déviances. L'objectif de ce mémoire est de confronter les sorties du système de détection automatique d'anomalies à un jury d'experts sur un nombre limité d'échantillons de parole issus des corpus *TyPaLoc* et *VML*, toutes les populations de patients dysarthriques étant représentées. Dans cette optique, un protocole d'évaluation a été proposé permettant à un jury d'experts d'annoter au niveau du mot la présence ou non d'une déviance, en terme de réalisation articulatoire, sur l'ensemble des échantillons de parole sélectionnés. Le choix du

mot a été guidé par un précédent mémoire d'orthophonie réalisé par Loriane Piro et Louisa Ziamni (toujours sous la direction de Christine Meunier et moi-même) dans lequel nous avions relevé la difficulté pour un auditeur d'évaluer la présence de déviances au niveau du phonème. La détection automatique des anomalies étant basée sur le phonème, il a été décidé de considérer comme déviant un mot monosyllabique comportant au moins une anomalie détectée par le système et un mot polysyllabique comportant au moins deux anomalies.

La sélection des échantillons s'est faite autour de trois catégories de mots présentant les caractéristiques suivantes en lien avec le comportement du système automatique :

1. les concordances (représentant 50% des échantillons) : cette catégorie regroupe les concordances entre le système automatique et l'évaluation perceptive a posteriori des deux élèves orthophonistes sur la présence ou l'absence d'une déviance dans le mot. Cette catégorie a été elle-même divisée en deux sous-catégories : les concordances évidentes (25%) et les concordances douteuses dès lors que les deux élèves orthophonistes n'étaient pas elles-mêmes en accord (75%) ;

2. les faux positifs (25%) : une déviance détectée par le système mais non perçue par les deux élèves orthophonistes ;

3. les faux négatifs (25%) : une déviance perçue par les deux élèves orthophonistes mais non détectée par le système.

La volonté de varier la nature des mots cibles (mots grammaticaux/lexicaux), la longueur des mots cibles (courts/long) et le positionnement des mots cibles dans la phrase a également guidé le choix des échantillons.

Ainsi, 98 passages, comportant de 1 à plusieurs mots cibles et produits par 41 locuteurs (34 dysarthriques et 7 sujets contrôles) ont été annotés perceptivement par le jury, soit un total de 289 mots jugés comme normaux par les deux élèves orthophonistes et 126 mots jugés comme déviants. Ce jury d'experts est composé de 18 étudiants de dernière année d'orthophonie, de 13 orthophonistes et d'1 ORL/phoniatre.

De manière synthétique (le lecteur pourra se reporter au mémoire (Pianelli et Restivo, 2016) pour le détail des analyses), l'analyse des résultats a montré que :

- le taux de concordance entre le jury et le système automatique sur la présence d'une déviance s'élève à 81% pour les concordances évidentes, à 53% pour les concordances douteuses, 35% pour les faux négatifs, 11% pour les faux positifs ; Si le dernier résultat est sans équivoque et demande une analyse acoustique fine des échantillons afin de mieux comprendre les erreurs du système automatique, il est intéressant de constater qu'environ 1 expert sur 3 est en accord avec le système sur les faux négatifs et qu'environ 1 expert sur 2 pour les concordances douteuses. **Ces résultats montrent bien la difficulté intrinsèque d'appréhender et de détecter une déviance, même pour un expert humain** ;

– le taux de concordance entre le jury et le système automatique sur l'absence d'une déviance (tous les mots non cibles) s'élève à 88% sur l'ensemble des échantillons ;

– l'analyse des résultats par population semble montrer que **les jurés, malgré leur expertise et les consignes reçues de tenir compte uniquement des déviances d'ordre articulatoire, peuvent être facilement influencés par l'ensemble des informations véhiculées par le signal de parole et, particulièrement, lorsque l'impression de parole dégradée est globale à l'échantillon**. En effet, le nombre de déviances relevées par les jurés a été significativement plus important chez les patients atteints de SLA pour lesquels la perception d'une altération globale du signal de parole peut effectivement être plus présente (voix rauque, nasonnement, degrés de sévérité les plus élevés, etc.) avec environ 20% de concordance avec le système sur les faux positifs (moyenne de 11% sur l'ensemble des populations) et 79% de concordance avec le système sur les concordances douteuses (contre 53% en moyenne sur l'ensemble des populations) ;

– L'analyse de la variabilité inter-jury montre une très grande disparité dans la quotation des experts, allant du juré le plus sévère annotant 40% des mots comme déviants au juré le plus clément annotant uniquement 7.5% (moyenne : 23, écart-type : 8.6). **Cette variabilité rend à nouveau compte de la difficulté pour un expert humain d'apprécier une déviance sur le plan perceptif** ;

– la sélection d'un sous-groupe de 7 jurés, dont les quotations ont été jugées les plus homogènes [11], a permis d'atteindre un taux de concordance avec le système automatique de 95% pour les concordances évidentes, de 64% pour les concordances douteuses, de 24.2% pour les faux négatifs et de 11.3% pour les faux positifs. Ce jury tend à donner raison au système dans les cas favorables (concordances) et être plus sévères dans les cas défavorables (en présence probables d'erreurs du système). La comparaison des réponses de ces 7 jurés avec l'ensemble du jury semble montrer que **le niveau d'expertise des jurés influence par conséquent leurs décisions et la concordance de leurs réponses avec celle du système**.

2.3.7 Conclusions

Nous venons de présenter quelques uns des travaux réalisés au LIA sur la parole dysarthrique et notamment autour de la détection automatique des anomalies. Nous avons montré que ce système obtient des résultats potentiellement intéressants mais que des progrès sont encore à fournir notamment afin de le rendre moins "sévère" en terme de nombre d'anomalies détectées (amélioration des mesures de précision). Ces améliorations pourraient être apportées par une augmentation significative des données d'apprentissage pour l'estimation du modèle de parole déviante et une meilleure

11. Il s'avère que ces 7 jurés ont une expérience plus régulière de la dysarthrie ou des troubles de la voix et de la parole en général. Néanmoins, tous les jurés ayant le même niveau d'expérience n'ont pas été sélectionnés, montrant l'hétérogénéité possible du jugement perceptif à degré d'expérience équivalent.

fiabilité des annotations qui, pour rappel, reposent ici sur l'expertise d'un seul annotateur humain. Dans cette perspective, nous nous heurtons aux mêmes problèmes soulevés dans le cadre des travaux sur l'évaluation objective du degré de sévérité de la dysphonie. En effet, l'augmentation des données d'apprentissage requiert (1) la disponibilité de corpus d'enregistrements présentant un grand nombre de patients et une variabilité suffisante pour une bonne couverture des phénomènes observables et (2) l'annotation en déviances de ces mêmes corpus. Outre la disponibilité des corpus, la phase d'enrichissement de ces derniers pose le problème de la fiabilité de l'annotation utilisée comme *Gold Standard* pour l'apprentissage des modèles statistiques - l'annotation doit être la plus objective et précise possible. Hors, dans le cas des déviances dues à des troubles de la parole, comme nous avons pu le constater avec le mémoire d'orthophonie présenté dans la section précédente, et de manière générale pour toute évaluation perceptive (Revis, 2004), objectivité et précision ne sont pas garanties. Premièrement, pour tout jury d'évaluation, l'obtention d'une variabilité intra- et inter-juge réduite demande la mise en place d'un protocole d'évaluation très couteux en temps et en ressources humaines, difficilement envisageable sur de grands corpus. En second, la précision d'une évaluation nécessite une définition bien établie des éléments à évaluer : ici, la définition d'une anomalie en lien avec des troubles de la parole et de la voix. En d'autres termes, qu'est-ce qu'une déviance ?
Si certaines anomalies peuvent être facilement "définissables" comme, par exemple le dévoisement [12], d'autres le sont plus difficilement - une distorsion dans le domaine spectral. Rappelons que, toujours dans le mémoire précédent, 1 expert sur 2 du jury complet n'était pas d'accord sur le fait de donner raison ou pas au système pour les concordances douteuses.

Un autre parallèle possible avec l'évaluation des dysphonies repose sur le fait que le système de détection des anomalies s'avère être très performant sur les patients les plus atteints par la dysarthrie (degrés de sévérité au delà de 2 sur 3) à comparer avec les forts taux de classification correcte obtenus par le système sur les patients de grade 3 de dysphonie. Dans les deux cas de figure, les erreurs sont accentuées sur les patients légèrement ou plus modérément atteints (grades intermédiaires). **La question que l'on peut se poser ici est de savoir si ces erreurs sont dues à une difficulté accrue des experts pour exécuter la tâche qui leur est demandée face à des frontières plus floues entre classes - frontières des grades 1 et 2 de dysphonie - frontières entre parole normale/anormale chez des patients peu ou modérément dysarthriques**.

12. Faut-il encore être capable d'apprécier un dévoisement accepté dans la parole normale d'un dévoisement "pathologique" ?

Troisième partie

Bilan et perspectives

Complexité et variabilité
Ce document aborde le thème des troubles de la parole et de la voix et l'application des outils de traitement automatique de la parole à ce contexte particulier. Au travers de la première partie, nous avons voulu souligner, bien que ce soit une caractéristique bien connue des experts du domaine, la complexité des nombreux processus mis en jeu dans la réalisation motrice de la parole pour produire ne serait-ce qu'un simple phonème, sans parler d'une syllabe ou d'un mot. Nous avons d'ailleurs pu voir que tous ces processus ne sont pas encore entièrement "décortiqués" et que des recherches sont encore nécessaires pour cerner certains d'entre eux, notamment sur le plan neurologique. Dès lors qu'un ou plusieurs de ces processus sont perturbés, cette complexité rend difficile la tâche d'analyse, de compréhension et d'évaluation des répercutions sur le signal de parole. Si la réalisation motrice est complexe, le signal de parole qui en découle l'est tout autant et présente une très grande variabilité lui permettant de véhiculer les informations propres au langage mais également les informations extra-linguistiques propres au locuteur. Par ailleurs, chaque locuteur étant constitué différemment sur le plan physiologique, la variabilité inter-individuelle est également à prendre en compte dans le signal de parole. De même, peut-on parler d'une variabilité intra-locuteur puisque ce dernier ne pourra produire deux fois le même son strictement de la même façon, ce phénomène étant amplifié dès lors que des informations extra-linguistiques telles que les émotions ou l'état pathologique viennent ajouter de la variation.
L'approche par corpus et la disponibilité de corpus de plus en plus larges d'enregistrements de parole a permis de faire avancer la connaissance sur la parole normale et sa variabilité dans le domaine de la phonétique. Néanmoins, des recherches sont encore nécessaires, notamment pour aborder cette notion de variabilité en fonction de différents contextes de parole, comme par exemple, la parole contrôlée vs parole non préparée.

Dans le chapitre dédié aux troubles de la parole et de la voix, nous avons pu voir quels mécanismes impliqués dans la production de la parole pouvaient être perturbés par les troubles dysphoniques ou dysarthriques, de manière individuelle ou conjointe, et leurs conséquences potentielles sur le signal de parole. Nous avons pu voir qu'il existait différentes formes de troubles (cf. classification des dysphonies ou des dysarthries) en fonction de la localisation des lésions et que les conséquences de ces troubles chez les patients pouvaient être très variables en fonction de différents facteurs propres à la maladie, à la manifestation de ces troubles, au patient lui-même ou encore à sa prise en charge thérapeutique. En plus de la variabilité intrinsèque de la parole que nous venons d'évoquer, se rajoutent, par conséquent, ici, d'autres formes de variabilité qui doivent être prises en compte et étudiées.

Un des problèmes majeurs auquel les praticiens doivent faire face actuellement est l'évaluation des troubles de la parole et de la voix, lors du diagnostic, de l'évolution de la maladie chez le patient, pour son suivi lors de prises en charge thérapeutique, etc. Malgré de nombreuses recherches sur le sujet, l'évaluation perceptive (ie "à l'oreille" du praticien) reste la méthode la plus utilisée en pratique clinique et ce, malgré, son carac-

tère subjectif très controversé renforcé par le facteur de variabilité que nous venons de mentionner. Au vu des progrès réalisés ces dernières années, le traitement automatique de la parole se veut être une solution idéale pour apporter une approche objective et peu coûteuse à cette problématique d'évaluation. Hors, la revue de la littérature que nous avons réalisée montre que les avancées sont encore loin de répondre à cette demande, et ce, pour deux raisons majeures :

1. à nouveau, la variabilité induite par la parole doublée ici par son caractère pathologique ;
2. le manque de corpus disponibles pour pouvoir modéliser correctement cette variabilité.

Même si la recherche bénéficie aujourd'hui des avancées technologiques récentes (traitement de la variabilité, apprentissage profond, etc.), nous revenons en quelques sorte une dizaine d'années en arrière lorsque les systèmes de traitement de la parole souffraient d'un manque cruel d'enregistrements de parole normale et variée.

Une des perspectives à envisager dans ce cadre pour permettre des avancées aussi significatives est la collecte de grands volumes de données tenant compte d'un maximum de variabilité intra-et inter-populations, à l'instar, des corpus de parole normale collectés dans le cadre de projets de recherche ou d'instituts spécialisés comme LDC ou ELDA. Néanmoins, si la collecte de ces corpus de parole normale a pu bénéficier du développement des médias numériques (collectes massives de journaux radiophoniques ou télévisés par exemple), la collecte d'enregistrements dans le domaine clinique reste, à notre avis, un cas très particulier, qui nécessite un plus grand contrôle des conditions d'enregistrement et des informations collectées autres que les simples données audio. En effet, au vu des travaux de recherche à mener dans ce domaine, les données portant sur le patient lui-même (informations personnelles et médicales) afin de pouvoir bien cerner, lors des analyses, ce qui relève de la production de parole normale de ce qui a trait à la pathologie du patient sont aussi importantes que les données audio. Pour exemple, il serait important de contrôler pour des patients atteints de la maladie de Parkinson si les sessions d'enregistrement ont eu lieu alors que ces derniers étaient sous traitement thérapeutique ou non (stimulation, traitement L-Dopa, etc.).
A notre connaissance, un seul projet de très grande envergure a été mis en place jusqu'à présent. Il s'agit de la plateforme de collecte d'enregistrements par téléphone développée par l'équipe Parkinson's voice Initiative (impulsée par le chercheur Max Little - http ://www.parkinsonsvoice.org). Cette plateforme a pour vocation de recueillir le plus grand nombre d'échantillons de parole sur une période donnée afin, d'une part, de pallier le manque de données évoqué précédemment, et, d'autre part, de constituer une cohorte de patients atteints de la maladie de Parkinson suffisante pour couvrir la grande variabilité des troubles à observer. L'objectif in fine du projet et de cette collecte est de développer un outil automatique de détection précoce de la maladie. Depuis son ouverture en 2012, la plateforme a permis de recueillir 10000 échantillons de parole, mêlant sujets sains, sujets potentiellement atteints de maladie de Parkinson mais non diagnostiqués et patients diagnostiqués, qui sont en cours d'analyse par l'équipe de Max Little. Néanmoins, contrairement à ce que nous avons évoqué précédemment, très peu de contrôle de l'appelant (voire aucun) ne semble être réalisé et peu de documen-

tation est disponible sur le type d'informations demandées à ces derniers, hormis la connaissance d'un diagnostic de la maladie de Parkinson. La base de données n'est actuellement pas diffusée à la communauté. Les seuls travaux pour l'instant disponibles restent ceux ayant initié cette collecte, ie basé sur une population restreinte d'une cinquantaine de patients.

L'autre voie a envisagé est celle initiée dans le projet *DesPhoAPady* qui consiste à récupérer et à structurer les données déjà existantes dans différents établissements hospitaliers. Le projet a permis d'aboutir à un modèle de données permettant de stocker les enregistrements audio mais également toutes les informations disponibles s'y reportant (informations du patient, de la session d'enregistrement, du contexte expérimental, des évaluations, etc.) (Ghio et al., 2012). La mise en place pratique de ce modèle permet à l'heure d'aujourd'hui de stocker, de manière structurée, une grande partie des enregistrements réalisés dans les hôpitaux d'Aix et de La Timone (Marseille) et leurs données associées en vue de leur utilisation à des fins de recherche. Les données également rassemblées dans le projet *DesPhoAPady* ont également été intégrées. L'étape suivante serait de généraliser cette pratique auprès d'autres établissements hospitaliers afin de mutualiser les ressources existantes.

Travaux du LIA sur les thèmes de la dysphonie et de la dysarthrie
Nous avons rapporté dans ce document les activités que nous avons menées au LIA depuis une dizaine d'années sur la dysphonie puis la dysarthrie, en étroite collaboration avec nos collègues des laboratoires LPL d'Aix-en-Provence et LPP de Paris.

Les systèmes automatiques que nous avons développés répondent à des tâches précises de classification comme méthode d'évaluation objective dans le cadre de la dysphonie ou comme méthode de détection d'anomalies dans le cadre de la dysarthrie. Néanmoins, l'objectif de ces travaux sont également de pouvoir apporter de la connaissance à nos partenaires phonéticiens et cliniciens sur les troubles étudiés en vue, par exemple, **d'orienter leur analyse** par la mise à jour d'une nouvelle caractéristique - les résultats inattendus obtenus par le système automatique sur les consonnes sourdes dans un contexte de dysphonie, soulignés par l'analyse du VOT -, **de faciliter ces analyses** - détection de déviances dans une large population de patients permettant de focaliser l'attention du phonéticien sur des zones particulières - ou **d'orienter ces analyses** - comparaison du comportement des patients et des sujets contrôles sur la parole spontanée. A l'opposé, nos travaux ont également besoin de l'expertise des cliniciens et des phonéticiens pour comprendre le comportement du système dans certains contextes, comme par exemple, l'analyse des erreurs de type "faux positifs" du système de détection sur les sujets contrôle - *est-ce bien une erreur du système ou une production atypique de la part du locuteur ?* ou l'analyse des différences de comportement du système entre sujets contrôles et patients sur la parole spontanée. Cette orientation montre l'importance des collaborations multi-disciplinaires que nous avons tissées au travers des différents projets de recherche auxquels nous avons participé et que nous devons maintenir et amplifier dans la suite de nos activités.

De manière similaire aux travaux de la littérature, l'interprétation et la généralisation de nos résultats souffrent du même manque de données pour faire face au facteur de variabilité maintes fois mentionnée dans ce document, même si nous disposions pour certaines études de 122 locuteurs. Nous avons également évoqué, au cours de ce document, la problématique de l'annotation par un expert humain de corpus audio dès lors que la frontière entre deux annotations possibles est difficile à définir et/ou revêt un caractère subjectif très dépendant de l'annotateur et de son niveau d'expertise (notion de déviance). Estimer des modèles statistiques sur des données "bruitées" induit forcément de la confusion entre modèles et des erreurs dans le comportement du système. Que ce soit avec les voix dysphoniques ou la parole dysarthrique, le problème a été persistant, limitant, à notre avis, grandement, les performances de nos systèmes. Les perspectives que nous proposons ci-dessous sont orientées dans une grande majorité pour répondre à cette problématique.

Perspectives
Le système automatique de détection des anomalies, basé sur l'approche à deux modèles (normal et anormal), que nous avons développé pour le traitement de la parole dysarthrique repose sur l'annotation manuelle des anomalies pour l'apprentissage du modèle correspondant. L'idée initiale était d'utiliser l'ensemble des anomalies relevées par l'expert humain sans tenir compte du type d'anomalies que ce dernier avait identifié et reporté. Au vu des difficultés pour définir ce qu'est une déviance et du coût pour les annoter, nous pourrions envisager d'aborder le problème différemment. En effet, nous pourrions dans un premier temps tirer bénéfice des connaissances déjà disponibles dans le domaine de la phonétique clinique pour identifier un ensemble fermé de déviances, considérées comme "facilement" mesurables, de manière automatique, dans le signal de parole (présence de dévoisement dans un contexte non attendu par exemple). Ces déviances seraient alors extraites automatiquement sur de plus larges corpus non annotés de parole normale et pathologique en vue de les intégrer dans un système de classification (phase d'apprentissage des modèles) permettant ensuite de les identifier.

Le LPL dispose d'un nombre considérable d'enregistrements de patients dysphoniques ou de patients atteints de dysarthrie (plus de 2500 patients enregistrés (Ghio et al., 2012)). Ces enregistrements de parole sont pour la plupart accompagnés d'une évaluation perceptive ("à l'oreille") du praticien ayant suivi le patient (grade global de sévérité de la dysphonie suivant le paramètre G de l'échelle GRBAS, évaluation globale de la dysarthrie suivant l'échelle UPDRS pour les patients atteints de la maladie de Parkinson, etc.). Une des perspectives de nos activités de recherche à venir serait d'exploiter au mieux ces données.
En premier lieu, il serait intéressant d'appliquer le système de détection des anomalies sur ces larges corpus en se focalisant, dans un premier temps, sur les patients les plus atteints (dysarthrie modérée à sévère). En effet, le système a démontré sa pertinence dès lors que le degré de sévérité de la dysarthrie est très élevé. En se focalisant sur les anomalies de patients modérément ou sévèrement atteints, et par conséquent, sur des anomalies plus "typiques et marquées", nous pouvons émettre l'hypothèse que le système sera moins enclin à détecter de fausses anomalies, ce qui est, rappelons le, l'un de

ces principaux défauts relevés. L'utilisation de ces anomalies pour l'apprentissage du ou des modèles de déviance du système de détection automatique pourrait permettre de pallier partiellement le besoin en annotation manuelle. Ce travail demandera de lever deux verrous au préalable. D'une part, il n'est pas envisageable de procéder à une transcription orthographique manuelle d'un tel volume de données. Aussi, un filtrage des alignements automatiques, requis par le système de détection, sera nécessaire afin de ne conserver que ceux associés à un indice de confiance élevé. Cette opération devra permettre, par exemple, de rejeter des insertions de mots faites par le locuteur lors de la lecture du texte et qui vont perturber le système d'alignement contraint par le texte. Il est à noter que ce filtrage, s'il est trop sévère, peut avoir pour conséquence d'orienter la détection des anomalies uniquement vers un ensemble de phonèmes bien alignés. La mise en place du filtrage et son réglage devra par conséquent tenir compte de ce facteur. Le deuxième verrou à lever sera de pouvoir associer également un indice de confiance aux anomalies détectées en vue de rejeter les faux positifs. En effet, nous avons mentionné que les erreurs d'annotation dues à un expert humain étaient une source de perturbation des systèmes automatiques dès lors que ces annotations étaient utilisées dans l'apprentissage des modèles. Ce filtrage des anomalies devraient, par conséquent, permettre de limiter l'utilisation de données bruitées au sein du système.
Ce volume de données devrait également permettre de répondre à la perspective précédente que nous avons mentionnée, portant sur la recherche de déviances ciblées au sein de grands corpus de parole pathologique.

Les travaux sur la détection des anomalies et nos collaborations avec nos partenaires phonéticiens nous ont montré toute la difficulté de définir une déviance dans la parole pathologique au vu de la complexité et de la variabilité, déjà avérées et non encore bien délimitées, de la parole normale. Aussi, dans une optique différente mais complémentaire, il nous semble intéressant dans la perspective de ces travaux, de revenir à des observations plus basiques. La confusion entre phonèmes est l'un des volets que nous voudrions étudier. Dans la lignée de notre objectif initial d'apport de connaissances par le biais du traitement automatique, cette confusion serait analysée au travers des sorties du système automatique d'alignement contraint par le texte. Il s'agirait de quantifier et de qualifier les confusions les plus fréquentes chez les différentes populations de patients à notre disposition et ce, en comparant, les deux styles de parole lue vs spontanée. Nous avons déjà initié ce travail au travers d'un mémoire de master recherche. Nous espérons que l'analyse et l'interprétation des résultats, avec l'aide de nos collègues phonéticiens et cliniciens, nous permettra d'apporter des connaissances supplémentaires au domaine.

Pour finir, nous sommes impliqués depuis 2014 dans le projet C2SI - Carcinologic Speech Severity Index - financé par l'Institut National sur le CAncer (INCA) en partenariat avec le CHU de Toulouse (porteur du projet), les laboratoires IRIT et Octogone (Toulouse) et LPL (Aix-en-Provence). Ce projet a pour vocation de montrer qu'un indice de sévérité fourni par un outil de traitement automatique peut être équivalent ou supérieur à un score d'intelligibilité obtenu par des auditeurs humains, en terme de

prédiction du handicap de parole, pour des patients traités pour un cancer des voies aéro-digestives supérieures. La cause des perturbations dans la réalisation motrice de la parole n'étant pas d'ordre neurologique, le terme de dysarthrie ne s'applique plus ici. Néanmoins, les conséquences sur la production de la parole n'en sont pas moins très similaires, la grande différence étant ici que nous connaissons plus précisément la localisation de la lésion et l'organe de l'appareil de production touché en premier lieu. Nous sommes actuellement en phase de finalisation de la collecte d'un corpus d'enregistrements audio et des données cliniques qui s'y rattachent, sur la base de 100 patients et 35 sujets contrôles. Outre la participation à la définition du protocole d'enregistrements (Ghio et al., 2016) et le suivi de la qualité des enregistrements, le travail de recherche du LIA sur le corpus débutera en novembre 2016. Notre objectif est de tirer partie de l'expertise acquise sur le traitement de la dysarthrie et de la connaissance plus précise que nous aurons des lésions du patient afin de l'utiliser dans le traitement automatique. Ce projet sera l'opportunité d'utiliser le large volume d'enregistrements disponibles au LPL, et notamment, de l'ensemble des 600 patients atteints de la maladie de Parkinson. En effet, nous avons montré dans la revue de nos résultats que le taux de détection des anomalies chez nos différentes populations de patients était fortement corrélé au degré de sévérité global de leur dysarthrie. Il s'agira d'observer si ce résultat est reproductible sur cette population élargie de patients atteints de la maladie de Parkinson pour laquelle nous disposons pour chacun des patients d'une évaluation perceptive sur la base de l'échelle UPDRS (travaux rejoignant ceux réalisés dans le cadre du Challenge Interspeech 2015, évoqué dans la revue de la littérature du chapitre 1), et, le cas échéant, de reporter cette étude sur la population de 100 patients atteints d'un cancer des voies aéro-digestives supérieures collectée dans le cadre du projet C2SI.

Au delà du projet C2SI et de ses objectifs, le rapprochement des observations faites sur la population de patients atteints d'un cancer et celles de patients dysarthriques nous semble être une démarche des plus pertinentes à mettre en oeuvre. En effet, tirer partie de la connaissance de la localisation des lésions des patients atteints d'un cancer pour lesquels nous pourrions observer des comportements similaires chez d'autres populations de patients dysarthriques pourrait permettre d'apprendre davantage sur les troubles de ces derniers (et vice-versa) et les conséquences des lésions neurologiques sur le fonctionnement de l'appareil de production de la parole. Cette démarche fera partie intégrante des activités futures du LIA sur les troubles de la voix et de la parole.

Annexe A

Convention de transcription des enregistrements pour la parole lue

Convention de transcription v6.0

Ce document réunit les règles et conventions de transcription à respecter en vue des traitements manuelles et automatiques.

Le texte original doit être modifié uniquement pour les différents cas énumérés ci-dessous. Pour le traitement automatique, un certain nombre de balises sera supprimé avant traitement. Néanmoins, elles pourront être réutilisées pour l'analyse.

1. Délétion (suppression)

Est considérée comme une délétion, l'absence (par rapport au texte original) d'un mot entier ou d'une ou plusieurs syllabes, quelle que soit sa position dans le mot.

Convention 1.1 : le mot ou les syllabes supprimés sont encadrés par des parenthèses (sans espace à l'intérieur des parenthèses). Quand seule une portion d'un mot est supprimée (se retrouvant alors entre parenthèses), le reste du mot, prononcé par le locuteur, doit être phonétisé en SAMPA.

Exemples :

a) suppression d'un ou plusieurs mots :
ex : suppression du mot "petit" => *dans un (petit) village*
ex : suppressions des mots « un » et « petit » => *dans (un) (petit) village*
(attention : chaque mot doit être balisé par les parenthèses)

b) suppression d'une ou plusieurs syllabes :
ex : suppression de la syllabe "pe" en début du mot "petit" => *dans un (pe)ti village*
ex : suppression de la syllabe "lla" au milieu du mot "village" => *dans un petit vi(lla)Z*
ex : suppression des syllabes "llageois" dans le mot "villageois" => *les vi(llageois)*

Attention : ne sera pas considérée ici une délétion inférieure à une syllabe (une consonne ou un groupe de consonnes). Le mot produit sera transcrit normalement.
ex : la non réalisation du groupe "rl" dans "horloge" ou de la consonne « r » dans « pauvre »

2. Insertion

Est considéré comme une insertion tout ajout (par rapport au texte original) d'un segment équivalent à au moins une syllabe. Trois types d'insertion peuvent être rencontrées :

- l'insertion d'un mot (ex : « un très petit village »)
- les faux départs et palilalie : répétition d'une syllabe ou d'un mot (ex : « un petit petit village », « le co co cordonnier)
- les hésitations ou pauses remplies (ex. « un euh petit village »)

Convention 2.1 : chaque segment inséré est transcrit sous une forme phonétisée en code SAMPA (voir annexe) et est délimité par un symbole "*" en début et en fin de segment (sans espaces). Dans le cas d'une forme inintelligible, celle-ci sera phonétisée le plus fidèlement possible et de manière à restituer au mieux la durée de la production.

ex : ajout du mot « très » => *dans un *tRE* petit village*
ex : ajout d'un segment inintelligible => *dans un petit *uwilaSe**

Convention 2.2 : lors d'une répétition (d'un mot ou de la première syllabe d'un mot), chaque

répétition sera balisée séparément par des "*", excepté la dernière qui ne sera pas balisée.

ex : répétition d'un mot => *un *povR* *povR* pauvre cordonnier*

ex : répétition de la première syllabe d'un mot => *pauvre *ko* *ko* cordonnier*

Attention : lors de répétitions de syllabe(s) en milieu de mot (ex : le cordodonnier), tout le mot sera phonétisé et balisé entre "*". ex => *koRdodonje*

Convention 2.3 : les hésitations et les pauses remplies ("euh", "hum", …) seront balisées par des "*" comme des insertions et phonétisées en code SAMPA le plus fidèlement possible ("euh" => "9", "hum"=>"9m", ...) de manière à restituer au mieux la durée de la production.

ex : hésitation => *dans un *9* *9* petit village*

3. Intervention d'un signal extérieur : autre locuteur, bruit, rire…

Toute intervention d'un autre locuteur, présence d'un bruit assez long ou d'un rire (ex. rire spasmodique), etc. doivent être balisées, notamment si cette intervention chevauche la voix du locuteur principal dont la parole doit être transcrite.

Convention 3.1 : la forme sonore devra être phonétisée en code SAMPA le plus fidèlement possible, de manière à restituer au mieux la durée de la production. La transcription de cette forme sonore sera délimitée par des "*" et encadrée par les balises "[ext]" en début et fin de segment.

Convention 3.2 : lors d'un chevauchement, la forme sonore principale sera retranscrite. La transcription de cette forme sonore sera délimitée par des "*" et encadrée par les balises "[ext]" en début et fin de segment.

ex : Intervention d'un autre locuteur => *[ext] *vwala* *R@liR* *l* *tEkst* [ext]*
(attention : chaque mot doit être balisé par des "")*

ex : bruit de feuilles=> *dans un [ext] *SSSS* [ext] petit*

4. Substitution

Est considérée comme une substitution : 1/ tout remplacement d'un mot entier par un autre segment (suite phonétique), 2/ tout remplacement d'au moins trois phonèmes **consécutifs** dans un mot par un autre segment.

Convention 3.1 : une substitution est encadrée par les balises [su=mot_cible] et [su], où «mot_cible» correspond au mot du texte. La forme produite sera transcrite entre les balises en code SAMPA.

ex 1 : substitution d'un mot entier ("les") par un autre groupe phonétique ("bien")=> *jamais je ne pourrai [su=les] bie~ [su] réparer*

ex 2 : substitution de trois phonèmes consécutifs (ex. "décou" dans décourager ou "bien" dans bientôt) => *[su=découragé] bipuRaZe [su] ; [su=bientôt] pato [su]*

Attention : ne sera pas considérée comme une substitution la modification de moins de 3 phonèmes consécutifs dans un mot. Le mot produit sera transcrit normalement.

ex : - un seul phonème : "le" prononcé "la"
- deux phonèmes consécutifs : "villageois" prononcé "vilaSa" ou "vilaZo"
- trois phonèmes non consécutifs dans un mots : "réparer" prononcé "REbavE"

5. Inversion

Est considérée comme une inversion toute permutation de deux mots consécutifs attendus dans la transcription originale : « Y X » à la place de « X Y ».

Convention 5.1 : une inversion de deux mots consécutifs sera notée de la manière suivante: en considérant « X Y » l'expression constituée des deux mots attendus dans la transcription et « Y' X' » l'expression prononcée par le locuteur (avec X=X' et Y=Y'), « Y' » sera considéré comme une insertion (voir §2), X' sera laissé tel quel, et « Y » sera considéré comme une délétion (voir §1).

Ex : permutation des deux mots "donc" et "a" dans la phrase "Qui donc a rangé la maison ?" => *"Qui *a* donc (a) rangé la maison ?"*

6. Application des conventions sur des exemples concrets de production sur le texte "le cordonnier"

Texte original :
"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 y 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. Ô joie ! Qui a préparé ce beau gâteau ? Qui donc a rangé la maison ? Tic tac tic tac dit la vieille horloge.".

Ex1: <s> dans un petit village de la montagne (il) y a *9~* un pauvre cordonnier tout vieux et tout cassé </s> <s> les villageois *lHi* lui apportent des chaussures à réparer mais il ne travaille pas vite </s> <s> tous les soirs il mange *9* tout seul bien tristement </s>

Ex2: <s> [ext] *vwala* *R@liR* *l* *tekst* [ext] dans un petit village de la montagne il y a un pauvre cordonnier tout vieux et tout cassé </s> <s> les villageois *lHi* lui apportent des chaussures à réparer mais il ne travaille pas vite </s> <s> tous les soirs il mange tout seul bien tristement […] <s> *oRa~z* rangeons les étagères réparons les souliers recousons le linge dit tic </s> <s> préparons *le* un gâteau mettons *l* *gi* du gui au plafond *sa~Zo~* changeons ces vieux rideaux ajoute tac </s> <s> minuit sonne </s> <s> les deux vaillants petits lutins rentrent dans la pendule </s> <s> le lendemain le pauvre cordonnier s' éveille </s> <s> *o* *Zwa* *ki* *a* *pRepaRe* *s* ô joie </s> <s> qui a préparé ce beau *gato* gâteau *se* *sa* </s> <s> qui donc a *di* rangé la maison </s> <s> *tik* *tik* *no~* tic tac *u* tic tac dit la vieille horloge </s>

Ex3: <s> dans un petit village de la montagne il y a un pauvre cordonnier tout vieux et tout cassé </s> <s> [ext] *Z* *a~lEv* *l* *talib* *Op* *alE* *zi* [ext] les villageois lui apportent des chaussures à réparer mais il ne travaille pas vite </s> <s> tous les soirs il mange tout seul bien tristement </s>

Ex4: <s> *da~* *da~* *9~* *p@ti* *vizE* dans un petit village de la montagne il y a *9~* un pauvre cordonnier tout vieux et tout cassé </s> <s> des villageois lui apportent des chaussures à réparer mais il ne travaille pas vite </s> <s> tous les soirs il mange *tu* *s9l* tout seul bien tristement </s> <s> ce soir il y a devant lui un gros tas de souliers et de guêtres à recoudre </s> <s> jamais je ne pourrai [su=les] bie~ [su] ré(parer) </s> <s> je suis trop âgé et trop malade </s> <s>

près de lui la grosse horloge fait tic tac tic tac </s> <s> *l* *povR* *l* *povR* *povR* pauvre vieux tout découragé s' endort </s> <s> aussitôt 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> *Ra~Z* rangeons les étagères réparons les souliers *R* *k* *k* recousons le linge dit tic </s> <s> préparons un gâteau mettons du gui au plafond changeons ces vieux rideaux ajoute tac </s> <s> minuit sonne </s> <s> *dE* *dE* *d2* vaj* les deux vaillants petits lutins rentrent *d* dans **le** pendule </s> <s> le lendemain le pauvre *ko* cordonnier s' éveille </s> <s> ô joie </s> <s> qui a préparé ce beau gâteau </s> <s> qui donc a rangé la maison </s> <s> tic tac tic tac dit la vieille horloge </s>

7. Codes Sampa

SAMPA : consonnes françaises		**SAMPA : voyelles françaises**	
SAMPA	**Exemples**	**SAMPA**	**Exemples**
p	*pont*	**i**	*si*
b	*bon*	**e**	*ses*
t	*temps*	**E**	*seize*
d	*dans*	**a**	*patte*
k	*coût*, *quand*, *koala*	**A**	*pâte*
g	*gant*	**O**	*comme*
f	*femme*	**o**	*gros*
v	*vent*	**u**	*doux*
s	*sans*, *dessus*, *cerise*	**y**	*du*
z	*zone*, *rose*	**2**	*deux*
S	*champ*	**9**	*neuf*
Z	*gens*, *jambon*	**@**	*justement*
j	*ion* [jo~]	**e~**	*vin*
m	*mont*	**a~**	*vent*
n	*nom*	**o~**	*bon*
J	*oignon*	**9~**	*brun*
N	*camping*		
l	*long*		
R	*rond*		
w	*quoi* [kwa]		
H	*juin* [Zhe~]		

Annexe B

Convention de transcription des enregistrements pour la parole spontanée

Conventions de transcription – TYPALOC

23 avril 2014

Actualisation des conventions:

- Pour tous les **corpus lus**, les conventions de transcription de Desphoapady sont appliquées.
- Pour les corpus de **parole spontanée**, les conventions TYPALOC (ci-dessous) sont appliquées. Une version standard (le minimum nécessaire pour l'alignement) est proposée (pour ceux qui veulent transcrire vite). Un deuxième niveau d'annotation (complémentaire) est proposé et pourra être supprimé automatiquement lorsque l'homogénéisation des transcriptions sera nécessaire (alignement, etc.). Pas de ponctuation!

	Transcription standard	***Code***	***Exemples***
1	Noms propres	**$**	$pornik$ (pour Pornic)
2	réalisations non attendues	**[**attendu, réalisé en SAMPA**]**	[aéroport, aReopOR] [pneu, p2n2]
3	Liaisons non attendues	**=**liaison**=**	donne moi =z= en
4	pauses remplies, hésitations	liste d'items spécifiques	euh, mmh, ben, hein, , hum
5	onomatopées	liste d'items spécifiques	ah, oh, eh, ouh, ouahou, hop, aïe, paf, boum
6	insertions non traitées	*****insertion*****	locuteur: *rires*, *respiration*, *chuchotements*; non locuteur: *externe* * (sans contenu si pas besoin de préciser)
7	Séquences non intelligibles	**?**	? (dans un chunck à part)
8	Pauses	**#**	# (dans un chunck à part)

	Annotations complémentaires	***Code***	***Exemples***
9	élision de phonèmes	**(**élision**)**	mE(r)kR@di, kE(l)k(e)fwa, tu (v)wa élision entre parenthèses et le reste du mot **en SAMPA**
10	Palilalies, répétitions, faux départs	rep**-** rep**-** tiret (du 6) suivi d'un blanc	le- le- le- téléphone

Commentaires spécifiques:

- 4 et 5: les listes seront données aux transcripteurs (ainsi que le code SAMPA: http://www.phon.ucl.ac.uk/home/sampa/french.htm)
- 6: les étoiles servent à baliser des séquences qui ne seront pas traitées dans le projet; il s'agit soit d'émissions du locuteur, soit de bruits extérieurs.
- 7: différents de * car ces séquences (inintelligibles) seront traitées à part
- 8: sont considérées comme pause: blancs + respirations + bruits de bouche doivent être > 250ms

Liste des illustrations

Liste des tableaux

Bibliographie

(Adda-Decker, 2006) M. Adda-Decker, 2006. De la reconnaissance automatique de la parole à l'analyse linguistique de corpus oraux. *Journées d'Etude sur la Parole, JEP'06*, Dinard, France, 389–400.

(Albin et al., 1989) R. Albin, A. Young, et J. Penney, 1989. The functional anatomy of basal ganglia disorders. *Trends in neurosciences 12(10)*, 366–375.

(Alonso et al., 2005) J. B. Alonso, F. Díaz-de María, C. M. Travieso, et M. A. Ferrer, 2005. Optimal size of time window in nonlinear features for voice quality measurement. *International Conference on Nonlinear Analyses and Algorithms for Speech Processing*, 206–218. Springer.

(Amelot, 2004) A. Amelot, 2004. *Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français*. Thèse de Doctorat, Université de Paris III, la Sorbonne Nouvelle.

(An et al., 2015) G. An, D. G. Brizan, M. Ma, M. Morales, A. R. Syed, et 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, Germany.

(Anders et al., 1988) L. Anders, H. Hollien, P. Hurme, A. Sonninnen, et J. Wendler, 1988. Perceptual evaluation of hoarseness by several classes of listeners. *Clinical Linguistics and Phonetics 40*, 91–100.

(Antoine, 2008) M. Antoine, 2008. *Adaptive word prediction and its application in an assistive communication system*. Thèse de Doctorat, Eberhard-Karls Universität Tübingen.

(Antolík et Fougeron, 2013) T. K. Antolík et C. Fougeron, 2013. Consonant distortions in dysarthria due to parkinson's disease, amyotrophic lateral sclerosis and cerebellar ataxia. *Proceedings of Interspeech'13*, Lyon, France, 2152–2155.

(Auzou, 2007a) P. Auzou, 2007a. Analyse sensori-motrice des effecteurs. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part II(20)*, 216–221.

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

(Auzou, 2007c) P. Auzou, 2007c. Les objectifs du bilan clinique de la dysarthrie. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part II(16)*, 189–195.

(Auzou, 2007d) P. Auzou, 2007d. L'intelligibilité. *Les dysarthries, édition Solal Evaluation, Part II(18)*, 204–2009.

(Auzou et al., 2001) P. Auzou, C. Ozsancak, et V. Brun, 2001. *Les dysarthries*. Edition Masson.

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

(Baghai-Ravary et Beet, 2012) L. Baghai-Ravary et S. W. Beet, 2012. *Automatic speech signal analysis for clinical diagnosis and assessment of speech disorders*. Springer Science & Business Media.

(Bailly, 2009) L. Bailly, 2009. Interaction entre cordes vocales et bandes ventriculaires en phonation : exploration in-vivo, modélisation physique, validation in-vitro. *Thèse de doctorat, Université du Maine*.

(Baken et Orlikoff, 2000) R. J. Baken et R. F. Orlikoff, 2000. *Clinical measurement of speech and voice*. Cengage Learning.

(Barkat-Defradas et al., 2012) M. Barkat-Defradas, C. Busseuil, O. Chauvy, F. Hirsch, C. Fauth, J. Révis, et B. A. de la Bretèque, 2012. Dimension esthétique des voix normales et dysphoniques : Approches perceptive et acoustique. *Travaux Interdisciplinaires sur la parole et le langage (TIPA) 28*.

(Barreto et Ortiz, 2008) S. D. S. Barreto et K. Z. Ortiz, 2008. Intelligibility measurements in speech disorders : a critical review of the literature. *Pró-Fono Revista de Atualizacao Cientifica 20(3)*.

(Barry et Pützer,) W. J. Barry et M. Pützer. Saarbrücken voice database.

(Bechet, 2011) M. Bechet, 2011. *Perturbation de la production des occlusives chez des locuteurs présentant une division palatine ou labio-palatine*. Thèse de Doctorat, Université de Strasbourg.

(Bell et al., 2012) P. Bell, M. Gales, P. Lanchantin, X. Liu, Y. Long, S. Renals, P. Swietojanski, et P. C. Woodland, 2012. Transcription of multi-genre media archives using out-of-domain data. *IEEE Spoken Language Technology workshop (SLT)*, 324–329.

(Benzeghiba et al., 2007) M. Benzeghiba, R. D. Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, et C. Wellekens, 2007. Automatic speech recognition and speech variability : A review. *Speech Communication 2007(49)*, 763–786.

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

(Besacier et al., 2014) L. Besacier, E. Barnard, A. Karpov, et T. Schultz, 2014. Automatic speech recognition for under-resourced languages : A survey. *Speech Communication 56*, 85–100.

(Beukelman et Yorkston, 1979) D. R. Beukelman et K. Yorkston, 1979. The relationship between information transfer and speech intelligibility of dysarthric speakers. *Journal of Communication Disorders 12*, 189–196.

(Bigi, 2015) B. Bigi, 2015. Sppas - multi-lingual approaches to the automatic annotation of speech. *The Phonetician - International Society of Phonetic Sciences, ISSN 0741-6164 111-112/2015-I-II*, 54–69.

(Boersma et Weenink,) P. Boersma et D. Weenink. Praat : doing phonetics by computer. http ://www.praat.org/.

(Borrie et al., 2012a) S. A. Borrie, M. J. McAuliffe, et J. Liss, 2012a. Perceptual learning of dysarthric speech : a review of experimental studies. *Journal of Speech, Language, Hearing Research 55(1)*, 290–305.

(Borrie et al., 2012b) S. A. Borrie, M. J. McAuliffe, J. Liss, C. Kirk, G. A. O'Beirne, et T. Anderson, 2012b. Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes 27(7–8)*, 1039–1055.

(Broca, 1862) P. Broca, 1862. Bulletin de la société francaise d'anthropologie. *Texte numéro 1 : séance du 18 avril 1861 tome 2*, 235–238.

(Bürki et al., 2008) A. Bürki, C. Gendrot, G. Gravier, G. Linarès, et C. Fougeron, 2008. Alignement automatique et analyse phonétique : comparaison de différents systèmes pour l'analyse du schwa. *Traitement Automatique des Langues 49*(3), 165–197.

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

(Carmichael et Green, 2004) J. Carmichael et P. Green, 2004. Revisiting dysarthria assessment intelligibility metrics. *Proceedings of the 8th International Conference on Spoken Language Processing, ICSLP'04*, 742–745.

(Chandola et al., 2007) V. Chandola, A. Banerjee, et V. Kumar, 2007. *Anomaly detection : a survey*. University of Minnesota (USA).

(Choi et al., 2011) D. Choi, B. Kim, Y. L. Y. Um, et M. Chung, 2011. Design and creation of dysarthric speech database for development of QoLt software technology. *Speech Database and Assessment (Oriental COCOSDA)*, Hsinchu City, Taiwan, 47–50.

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

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

(Christensen et al., 2013) H. Christensen, P. Green, et T. Hain, 2013. Learning speaker-specific pronunciations of disordered speech. *Proceedings of Interspeech'13*, Lyon, France.

(Coudière, 2003) C. Coudière, 2003. De l'utilité des logiciels pour la voix en rééducation en dysphonies dysfonctionnelles. Mémoire de Master, Université Paul Sabatier, Faculté de Médecine Toulouse-Rangueil.

(Crevier-Buchman, 2001) L. Crevier-Buchman, 2001. Les dysphonies chronicles. *La lettre d'Oto-rhino-laryngologie et de chirurgie cervico-faciale 263*, 25–28.

(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.

(Cristo, 2004) A. D. Cristo, 2004. La prosody au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions. *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence 23*, 67–211.

(Dahmani et al., 2013) H. Dahmani, S.-A. Selouani, D. O'shaughnessy, M. Chetouani, et N. Doghmane, 2013. Assessment of dysarthric speech through rhythm metrics. *Journal of King Saud University - Computer and Information Sciences 25(1)*, 43–49.

(Darley et al., 1969a) F. L. Darley, A. E. Aronson, et 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, et 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, et J. R. Brown, 1975. *Motor speech disorders*. Philadelphia : W. B. Saunders and Co.

(Daza-Santacoloma et al., 2009) G. Daza-Santacoloma, J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruíz, et G. Castellanos-Domínguez, 2009. Dynamic feature extraction : an application to voice pathology detection. *Intelligent Automation & Soft Computing 15*(4), 667–682.

(De Bodt et al., 1996) M. S. De Bodt, P. H. Van de Heyning, F. L. Wuyts, et L. Lambrechts, 1996. The perceptual evaluation of voice disorders. *Acta Oto-rhino-laryngolica Belg. 50(4)*, 283–291.

(Deller et al., 1993) J. R. Deller, M. S. Liu, L. J. Ferrier, et P. Robichaud, 1993. The whitaker database of dysarthric (cerebral palsy) speech. *Journal of the Acoustical Society of America 93*.

(Dibazar et al., 2002) A. A. Dibazar, S. Narayanan, et T. W. Berger, 2002. Feature analysis for automatic detection of pathological speech. *Engineering Medicine and Biology Symposium'02*, Volume 1, 182–183.

(Dien, 2008) J. Dien, 2008. Looking both ways through time : The janus model of lateralized cognition. *Brain and Cognition 67*(3), 292–323.

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

(Duez, 1987) D. Duez, 1987. *Contribution à l'étude de la structuration temporelle en français*. Thèse de Doctorat, Université de Provence.

(Duez, 2007) D. Duez, 2007. Prosodie et rythme. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(15)*, 181–188.

(Duffy, 1995) J. R. Duffy, 1995. *Motor speech disorders : substrates, differential diagnosis and management*. Motsby- Yearbook, St Louis, 1st edition.

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

(Dupessey et Coulombeau, 2003) M. Dupessey et B. Coulombeau, 2003. *A lécoute des voix pathologiques*. Symétrie.

(Durand et al., 2003) J. Durand, B. Laks, et C. Lyche, 2003. Le projet "phonologie du français contemporain" (pfc). *La tribune internationale des langues vivantes*, Numéro 33, 3–10.

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

(Enderby, 1986) P. Enderby, 1986. Relationships between dysarthric groups. *British Journal of disorders of communication 21*.

(Eyben et al., 2010) F. Eyben, M. Wöllmer, et 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.

(Fager et al., 2012) S. Fager, D. R. Beukelman, M. Fried-Oken, T. Jakobs, et J. Baker, 2012. Access interface strategies. *Assistive Technology 24(1)*, 25–33.

(Fager et al., 2010) S. K. Fager, D. R. Beukelman, T. Jakobs, et J.-P. Hosom, December 2010. Evaluation of a speech recognition prototype for speakers with moderate and severe dysarthria : A preliminary report. *Augmentative and Alternative Communication (AAC) 26(4)*, 267–277.

(Fant, 1960) G. Fant, 1960. *The Acoustic Theory of Speech Production*.

(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, et H. C. Shane, 1992. A case study of a dysarthric speaker using the dragondictate voice recognition system. *Journal for Computer Users in Speech and Hearing 8(1)*, 33–52.

(Ferrier et al., 1995) L. J. Ferrier, H. Shane, H. Ballard, T. Carpenter, et A. Benoit, 1995. Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. *Augmentative and Alternative Communication (AAC) 11*, 165–175.

(Fichaux-Bourin et al., 2009) P. Fichaux-Bourin, V. Woisard, S. Grand, M. Puech, et S. Bodind, 2009. Validation d'un questionnaire d'auto-évaluation de la parole (parole handicap index). *Revue de Laryngologie Otologie Rhinologie, 130*, 45–51.

(Fonseca et al., 2007) E. S. Fonseca, R. C. Guido, P. R. Scalassara, C. D. Maciel, et J. C. Pereira, 2007. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. *Computers in Biology and Medicine 37*(4), 571–578.

(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, 2013) C. Fougeron, 2013. *Renforcements et réductions segmentales : influence de la structure prosodique de la phonologie, et des contraintes motrices en parole pathologique*. Thèse de Doctorat, Mémoire d'Habilitation à Diriger des Recherches (HDR), Université Lyon 2.

(Fougeron et al., 2010) C. Fougeron, L. Crevier-Buchman, C. Fredouille, A. Ghio, C. Meunier, C. Chevrie-Muller, J.-F. Bonastre, A. Colazo-Simon, C. Delooze, D. Duez, C. Gendrot, T. Legou, N. Lévêque, C. Pillot-Loiseau, S. Pinto, G. Pouchoulin, D. Robert, J. Vaissière, F. Viallet, et C. Vincent, 2010. The DesPho-APaDy project : Developing an acoustic-phonetic characterization of dysarthric speech in french. *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC'10*, Valletta, Malta, 2831–2838.

(Fraile et al., 2009a) R. Fraile, J. I. Godino-Llorente, N. Saenz-Lechon, V. Osma-Ruiz, et C. Fredouille, 2009a. Mfcc-based remote pathology detection on speech transmitted through the telephone channel. *International Conference on Bio-inspired Systems and Signal Processing, BIOSIGNALS'2009*, Porto, Portugal.

(Fraile et al., 2009b) R. Fraile, N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, et C. Fredouille, 2009b. Effect of a simulated analogue telephone channel on the performance of a remote automatic system for the detection of pathologies in voice : Impact of linear distortions on cepstrum-based assessment-band limitation, frequency response and additive noise. *International Joint Conference on Biomedical Engineering Systems and Technologies*, 173–186. Springer.

(Fraile et al., 2009c) R. Fraile, N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiza, et C. Fredouille, 2009c. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mfcc parameters and differentiation of patients by sex. *Folia phoniatrica et logopaedica, International Journal of Phoniatrics, Speech Therapy and Communication Pathology, Special issue : COST Action 2103 - A Joint European Project for Advanced Voice Assessment*, *61*(3), 146–52.

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

(Fredouille et Pouchoulin, 2012) C. Fredouille et G. Pouchoulin, 2012. Détection automatique de zones de déviance dans la parole dysarthrique : étude des bandes de fréquences.

(Fredouille et al., 2005) C. Fredouille, G. Pouchoulin, J.-F. Bonastre., M. Azzarello, A. Giovanni, et A. Ghio, 2005. Application of automatic speaker recognition techniques to pathological voice assessment (dysphonia). *Proceedings of Interspeech'05*, Lisboa, Portugal.

(Fredouille et al., 2007) C. Fredouille, G. Pouchoulin, J.-F. Bonastre., A. Ghio, A. Marques, et A. Giovanni, 2007. Are the unvoiced consonants relevant for dysphonia phenomenon observation ? *3rd Advanced Voice Function Assessment International workshop, AVFA'09*, Madrid, Spain.

(Fredouille et al., 2009) C. Fredouille, G. Pouchoulin, A. Ghio, J. Revis, J. F. Bonastre, et A. Giovanni, 2009. Back-and-forth methodology for objective voice quality assessment : from/to expert knowledge to/from automatic classification of dysphonia. *EURASIP Journal on Advances in Signal Processing 2009*, 8.

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

(García et al., 2016) A. M. García, F. Carrillo, J. R. Orozco-Arroyave, N. Trujillo, J. F. V. Bonilla, S. Fittipaldi, F. Adolfi, E. Nöth, M. Sigman, D. F. Slezak, et al., 2016. How language flows when movements don't : An automated analysis of spontaneous discourse in parkinson's disease. *Brain and Language 162*, 19–28.

(Garnier, 2007) M. Garnier, 2007. *Communiquer en environnement bruyant : de l'adaptation jusqu'au forçage vocal*. Thèse de Doctorat, Université de Paris 6.

(Garnier et al., 2007) M. Garnier, D. Dubois, et N. Henrich, 2007. *Bruit et voix : de l'adaptation au forçage vocal. Conception et implication méthodologiques*. B. Vaxelaire, R. Sock, G. Kleiber et F. Marsac.

(Gauvain et Lee, 1994) J. L. Gauvain et 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.

(Gentil, 1992) M. Gentil, 1992. Phonetic intelligibility testing in dysarthria for the use of french language clinicians. *Clinical linguistics and phonetics 6*, 179–189.

(Ghio, 2007) A. Ghio, 2007. Modélisation du conduit vocal. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(11)*, 140–156.

(Ghio et al., 2016) A. Ghio, L. Giusti, E. Blanc, S. Pinto, M. Lalain, D. Robert, C. Fredouille, et V. Woisard, 2016. Quels tests d'intelligibilité pour évaluer les troubles de production de la parole ? *Journées d'Etude sur la Parole, JEP'16*, Paris, France.

(Ghio et al., 2007) A. Ghio, G. Pouchoulin, A. Giovanni, C. Fredouille, B. Teston, J. Révis, J.-F. Bonastre, D. Robert-Rochet, P. Yu, M. Ouaknine, et al., 2007. Approches complémentaires pour l'évaluation des dysphonies : bilan méthodologique et perspectives. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA) 26*, 33–74.

(Ghio et al., 2012) A. Ghio, G. Pouchoulin, B. Teston, S. Pinto, C. Fredouille, C. D. Looze, D. Robert, F. Viallet, et A. Giovanni, 2012. How to manage sound physiological and clinical data of 2500 dysphonic and dysarthric speakers ? *Speech Communication 54(5)*, 664–679.

(Ghio et Teston, 2004) A. Ghio et B. Teston, 2004. Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. *International Conference on Voice Physiology and Biomechanics*, 55–58.

(Giovanni et al., 1999) A. Giovanni, M. Ouaknine, et J. M. Triglia, 1999. Determination of largest lyapounov esponents of vocal signal : Application to unilateral laryngeal paralysis. *Journal of Voice 13 (3)*, 341–354.

(Godino-Llorente et al., 2008) J. I. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, I. Cobeta-Marco, R. González-Herranz, et C. Ramírez-Calvo, 2008. Acoustic analysis of voice using wpcvox : a comparative study with multi dimensional voice program. *European Archives of Oto-Rhino-Laryngology 265*(4), 465–476.

(Gómez-Vilda et al., 2007) P. Gómez-Vilda, R. Fernández-Baillo, A. Nieto, F. Díaz, F. Fernández-Camacho, V. Rodellar, A. Álvarez, et R. Martínez, 2007. Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *Journal of Voice 21*(4), 450–476.

(Gómez-Vilda et al., 2009) P. Gómez-Vilda, R. Fernández-Baillo, V. Rodellar-Biarge, V. N. Lluis, A. Álvarez-Marquina, L. M. Mazaira-Fernández, R. Martínez-Olalla, et J. I. Godino-Llorente, 2009. Glottal source biometrical signature for voice pathology detection. *Speech Communication 51*(9), 759–781.

(Grabski et al., 2009) K. Grabski, J.-L. Schwartz, L. Lamalle, C. Vilain, N. Vallée, I. Tropres, M. Baciu, J.-F. L. Bas, et M. Sato, 2009. Représentations cérébrales des articulateurs de la parole. *Actes du 8ème colloque Rencontres Jeunes-Chercheurs en Parole, Avignon, France*.

(Green et al., 2003) P. Green, J. Carmichael, A. Hatzus, P. Enderby, M. Hawley, et M. Parker, 2003. Automatic speech recognition with sparse training data for dysarthric speakers. *Proceedings of Interspeech'03*, Geneva, Switzerland, 1189–1192.

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

(Griffin et al., 2000) S. Griffin, L. Wilson, et 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, et 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*, Dresden, Germany, 47–54.

(Hamidi et Baljko, 2013) F. Hamidi et M. Baljko, 2013. Automatic speech recognition : a shifted role in early speech intervention. *4th workshop on Speech and Language processing for assistive technologies, SLPAT'13*, Grenoble, France.

(Hammarberg, 1986) B. Hammarberg, 1986. Perceptual and acoustic analysis of dysphonia. *Dept. of Logopedics and Phoniatrics, Karolinska Institutet*.

(Hammarberg et al., 1980) B. Hammarberg, B. Tritzell, J. Gauffin, J. Sundberg, et L. Wedin, 1980. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol 90*, 441–451.

(Hanson et al., 2004) E. Hanson, K. Yorkston, et D. Beukelman, 2004. Speech supplementation techniques for dysarthria : A systematic review. *Journal of Medical Speech Language Pathology 19*, 2–30.

(Hawley et al., 2007) M. S. Hawley, P. Enderby, P. Green, S. Cunningham, S. Brownsell, J. Carmichael, M. Parker, A. Hatzis, P. O'Neill, et R. Palmer, 2007. A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering & Physics 29(5)*, 586–593.

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

(Healy et al., 1981) W. C. Healy, B. L. Ackerman, C. R. Chappell, K. L. Perrin, et J. Stormer, 1981. *The prevalence of communicative disorders : A review of the literature*.

(Henrich, 2001) N. Henrich, 2001. *Etude de la source glottique en voix parlée et chantée. Modélisation et estimation, mesures acoustiques et électroglottographiques, perception*. Thèse de Doctorat, Université de Paris 6.

(Hinton et al., 2012) G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al., 2012. Deep neural networks for acoustic modeling in speech recognition : The shared views of four research groups. *Signal Processing Magazine, IEEE 29(6)*, 82–97.

(Hirano, 1974) M. Hirano, 1974. Morphological structure of the vocal cord as a vibrator and its variations. *Clinical Linguistics and Phonetics 26*, 89–94.

(Hirano, 1981) M. Hirano, 1981. Psycho-acoustic evaluation of voice : GRBAS scale for evaluating the hoarse voice. *Clinical Examination of voice, Springer Verlag*.

(Hirano, 1988) M. Hirano, 1988. *Laryngeal muscles and glottic shaping*. Vocal Physiology. Voice production, mechanisms and functions, Raven Press, New-York. Fujimura O. edition.

(Hirsch, 2007) F. Hirsch, 2007. *Le bégaiement : Perturbation de l'organisation temporelle de la parole et conséquences spectrales*. Thèse de Doctorat, Université de Strasbourg.

(Hixon et Hoit, 2005) T. Hixon et J. Hoit, 2005. *Evaluation and Management of Speech Breathing Disorders Principles and Methods*. San Diego : Plural publishing inc.

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

(Hustad, 2006) K. C. Hustad, 2006. A closer look at transcription intelligibility for speakers with dysarthria : Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech Language Pathology 15*, 268–277.

(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.

(INSERM, 2007) INSERM, 2007. La voix. ses troubles chez les enseignants. *Edition INSERM 2006. ISBN 2-85598-854-3*.

(J. D'Innocenzo, 2006) G. G. J. D'Innocenzo, K. Tjaden, 2006. Intelligibility in dysarthria : Effects of listener familiarity and speaking condition. *Clinical Linguistics and Phonetics 20*, 659–675.

(Jacobson et al., 1997) B. H. Jacobson, A. Johson, C. Grywalski, G. Jacobson, M. S. Benninger, et C. Newman, 1997. The Voice Handicap Index (VHI) : Development and validation. *American Journal of Speech-Language Pathology 6*, 66–70.

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

(Kain et al., 2007) A. B. Kain, X. N. J.-P. Hosom, J. P. H. V. Santen, M. Fried-Oken, et J. Staehely, 2007. Improving the intelligibility of dysarthric speech. *Speech Communication 49(9)*, 743–759.

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

(Kent et al., 1989) R. D. Kent, G. Weismer, J.-F. Kent, et J. Rosenbek, 1989. Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders 54*, 482–499.

(Kiliç et al., 2004) M. A. Kiliç, F. Öğüt, G. Dursun, E. Okur, I. Yildirim, et R. Midilli, 2004. The effects of vowels on voice perturbation measures. *Journal of Voice 18*(3), 318–324.

(Kim et al., 2008) H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin, et S. Frame, 2008. Dysarthric speech database for universal access research. *Proceedings of Interspeech'08*, Brisbane, Australia.

(Kim et al., 2013) S. Kim, Y. Hwang, D. Shin, C.-Y. Yang, S.-Y. Lee, J. Kim, B. Kong, J. Chung, N. Cho, J.-H. Kim, et al., 2013. Vui development for korean people with dysarthria. *Journal of Assistive Technologies 7(3)*, 188–200.

(Kotler et Thomas-Stonell, 1997) A.-L. Kotler et N. Thomas-Stonell, 1997. Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment. *Augmentative and Alternative Communication (AAC) 13(2)*, 71–80.

(Laaridh et al., 2015a) I. Laaridh, C. Fredouille, et C. Meunier, 2015a. Automatic detection of phone-based anomalies in dysarthric speech. *ACM Transactions on Accessible Computing (TACCESS) 6*(3), 9.

(Laaridh et al., 2015b) I. Laaridh, C. Fredouille, et C. Meunier, 2015b. Automatic speech processing for dysarthria : A study of inter-pathology variability. *International Congress of Phonetic Sciences, ICPhS'15, Glasgow, UK*.

(Laaridh et al., 2016a) I. Laaridh, C. Fredouille, et C. Meunier, 2016a. Automatic speech processing for dysarthric speech across different speech styles. *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC'16*, Portoroz, Slovenia.

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

(Laver, 1980) J. Laver, 1980. *The Phonetic Description of Voice Quality*. Cambridge.

(Le Huche et Allali, 2010a) F. Le Huche et A. Allali, 2010a. *la Voix : Anatomie et physiologie des organes de la voix et de la parole*, Volume 1. Masson, 4e édition.

(Le Huche et Allali, 2010b) F. Le Huche et A. Allali, 2010b. *la Voix : Pathologie vocale d'origine fonctionnelle*, Volume 2. Masson, 3e édition.

(Le Huche et Allali, 2010c) F. Le Huche et A. Allali, 2010c. *la Voix : Pathologie vocale d'origine organique*, Volume 3. Masson, 2e édition.

(Lechevalier et al., 2008) B. Lechevalier, F.Eustache, et F. Viader, 2008. *Traité de neuropsychologie clinique*.

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

(Letanneux et al., 2013) A. Letanneux, M. Walshe, F. Viallet, et S. Pinto, 2013. The dysarthria impact profile : A preliminary french experience with parkinson's disease. *Parkinson's Disease 2013*.

(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).

(Liss et al., 2002) J. M. Liss, S. M. Spitzer, J. N. Caviness, et C. Adler, 2002. The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America 112*, 3032–3030.

(Little et al., 2007) M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, et I. M. Moroz, 2007. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine 6*(1), 1.

(Logeman et al., 1978) J. A. Logeman, H. B. Fisher, B. Boshes, et 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 et R. D. Kent, 2010. *Assessment of motor speech disorders*, Volume 1. Plural publishing.

(Ma, 2008) L. Ma, 2008. *La coarticulation en français et en chinois : étude expérimentale et modélisation*. Thèse de Doctorat, Université d'Aix-Marseille I, Université de Provence.

(MacCallum et al., 2009) J. K. MacCallum, L. Cai, L. Zhou, Y. Zhang, et J. J. Jiang, 2009. Acoustic analysis of aperiodic voice : perturbation and nonlinear dynamic properties in esophageal phonation. *Journal of Voice 23*(3), 283–290.

(Maier et al., 2010) A. Maier, T. Haderlein, F. Stella, E. North, E. Nkenke, F. Rosanowski, A. Schutzenberger, et M. Schuster, 2010. Automatic speech recognition system for the evaluation of voice and speech disorders in head and neck cancer. *EURASIP Journal on Audio, Speech and Music Processing 2010*.

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

(Martínez et al., 2015) D. Martínez, E. Lleida, P. Green, H. Christensen, A. Ortega, et A. Miguel, 2015. Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace. *ACM Transactions on Accessible Computing (TACCESS) 6*(3), 10.

(McAuliffe et al., 2010) M. J. McAuliffe, S. Borrie, P. V. Good, et L. E. Hughes, 2010. Consideration of the listener in the assessment and treatment of dysarthria. *Speech, Language, and Hearing, 12(1)*, 16–19.

(Menéndez-Pidal et al., 1996) X. Menéndez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, et H. T. Bunnell, 1996. The nemours database of dysarthric speech. *Proceedings of the Fourth International Conference on Spoken Language Processing*, Philadelphia, USA.

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

(Meunier, 2014) C. Meunier, 2014. *Variation de la parole : contraintes linguistiques et mécanismes d'adaptation*. Thèse de Doctorat, Mémoire d'Habilitation à Diriger des Recherches (HDR), Université Lyon 2.

(Meunier et al., 2016) C. Meunier, C. Fougeron, C. Fredouille, B. Biggi, L. Crevier-Buchman, E. Delais-Roussarie, L. Georgeton, A. Ghio, I. Laaridh, T. Legou, C. Pillot-Loiseau, et 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*, Portoroz, Slovenia.

(Middag et al., 2009) C. Middag, J.-P. Martens, G. V. Nuffelen, et M. D. Bodt, 2009. Automated intelligibility assessment of pathological speech using phonological features. *EURASIP Journal on Applied Signal Processing 2009(1)*.

(Morales et Cox, 2009) A. O. C. Morales et S. J. Cox, 2009. Modelling errors in automatic speech recognition for dysarthric speakers. *EURASIP journal on advances in signal processing*.

(Morales et al., 2013) S. O. C. Morales, G. B. Enríquez, et F. T. Romero, 2013. Speech-based human and service robot interaction : An application for mexican dysarthric people. *International Journal of Advanced Robotic Systems 10(11)*.

(Morris et al., 2016) M. A. Morris, S. K. Meier, J. M. Griffin, M. E. Branda, et S. M. Phelan, 2016. Prevalence and etiologies of adult communication disabilities in the united states : Results from the 2012 national health interview survey. *Disability and health journal 9*(1), 140–144.

(Mumtaz et al., 2014) B. M. Mumtaz, S. S. Siti, M. Noraini, A.-A. Bassam, et E. S. Chng, 2014. Severity-based adaptation with limited data for asr to aid dysarthric speakers. *PloS one 9*(1).

(Munot et al., 2002) P. Munot, F. X. Nève, et P. Toussaint, 2002. *Une introduction à la phonétique* (Liège : Edition du Céfal ed.).

(Murdoch, 1998) B. E. Murdoch, 1998. *Dysarthria : a physiological approach to assessment and treatment*.

(Murdoch, 2011) B. E. Murdoch, 2011. Physiological investigation of dysarthria : Recent advances. *International journal of speech language pathology 13 (1)*, 28–35.

(Nidhyananthan et al., 2016) S. S. Nidhyananthan, V. Shenbagalakshmi, et al., 2016. Assessment of dysarthric speech using elman back propagation network (recurrent network) for speech recognition. *International Journal of Speech Technology 19*(3), 577–583.

(Nuffelen et al., 2009) G. V. Nuffelen, C. Middag, M. D. Bodt, et 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., 2015) J. Orozco-Arroyave, F. Hönig, J. Arias-Londono, J. Vargas-Bonilla, S. Skodda, J. Rusz, et E. Nöth, 2015. Voiced/unvoiced transitions in speech as a potential bio-marker to detect parkinson's disease. *Proceedings of Interspeech'15*, Dresden, Germany.

(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. *Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP'16)*, Shanghai, China, 6490–6494.

(Orozco-Arroyave et al., 2014) J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. V. Bonilla, M. C. Gonzalez-Rátiva, et E. Nöth, 2014. New spanish speech corpus database for the analysis of people suffering from parkinson's disease. *Proceedings of the International Conference on Language Resources and Evaluation, LREC'14*, Reykjavik, Iceland, 342–347.

(Oster et al., 2002) A. M. Oster, D. House, A. Protopapas, et A. Hatzis, 2002. Presentation of a new EU project for speech therapy : OLP (Ortho-Logo-Paedia. *Proceedings of Fonetik, TMH-QPSR*, Volume 44 (1), 45–48.

(Oue et al., 2015) S. Oue, R. Marxer, et F. Rudzicz, 2015. Automatic dysfluency detection in dysarthric speech using deep belief networks. *6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, 60.

(Parsa et Jamieson, 2000) V. Parsa et D. G. Jamieson, 2000. Identification of pathological voices using glottal noise measures. *Journal of Speech, Language, and Hearing Research 43(2)*, 469–485.

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

(Pianelli et Restivo, 2016) L. Pianelli et 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.

(Pinto et Ghio, 2009) S. Pinto et A. Ghio, 2009. Troubles du contrôle moteur de la parole : contribution de l'étude des dysarthries et dysphonies à la compréhension de la parole normale. *Revue française de linguistique appliquée 13*(2), 45–57.

(Polikoff et Bunnell, 1999) J. B. Polikoff et H. T. Bunnell, 1999. The nemours database of dysarthric speech : a perceptual analysis. *XIVth international Congress of Phonetic Sciences*, San Francisco, USA, 783–786.

(Polur et Miller, 2006) P. D. Polur et G. E. Miller, 2006. Investigation of an hmm/ann hybrid structure in pattern recognition application using cepstral analysis of dysarthric (distorted) speech signals. *Medical engineering & physics 28(8)*, 741–748.

(Pouchoulin, 2008) G. Pouchoulin, 2008. *Approche statistique pour l'analyse objective et la caractérisation de la voix dysphonique*. Thèse de Doctorat, Université d'Avignon.

(Pouchoulin et al., 2007) G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, et A. Giovanni, 2007. Frequency study for the characterization of the dysphonic voices. *Proceedings of Interspeech'07*, Antwerp, Belgium.

(Pouchoulin et al., 2008a) G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, et A. Giovanni, 2008a. Analyse phonétique dans le domaine fréquentiel pour la classification des voix dysphoniques. *Journées d'Etude sur la Parole, JEP'08*, Avignon, France.

(Pouchoulin et al., 2008b) G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, et A. Giovanni, 2008b. Dysphonic voices and the 0-3000Hz frequency band. *Proceedings of Interspeech'08*, Brisbane, Australia.

(Pouchoulin et al., 2007) G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, et J. Revis, 2007. Characterization of pathological voices (dysphonia) in the frequency space. *International Congress of Phonetic Sciences, ICPhS'07*, Saarbrücken, Germany.

(Raghavendra et al., 2001) P. Raghavendra, E. Rosengreen, et S. Hunnicutt, 2001. An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition system. *Augmentative and Alternative Communication (AAC) 17(4)*, 265–275.

(Remacle, 1999) M. Remacle, 1999. *Troubles de la voix*. In Rondal, J.A., Seron, X. (eds), Troubles du langage. Bases théoriques, diagnostic et rééducation, Sprimont (Belgique), Mardaga.

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

(Revis et al., 2012) J. Revis, C. Galant, C. Fredouille, A. Ghio, et A. Giovanni, 2012. Influence du contexte phonétique dans la manifestation dysphonique : apport des nouvelles méthodologies pour l'analyse de la voix pathologique. *Revue de laryngologie, d'otologie et de rhinologie 133*(1), 33–40.

(Reynolds et al., 2000) D. A. Reynolds, T. F. Quatieri, et R. B. Dunn, 2000. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing (DSP), a review journal - Special issue on NIST 1999 speaker recognition workshop 10 (1-3)*, 19–41.

(Riecker et Ackermann, 2007) A. Riecker et H. Ackermann, 2007. Circuits cérébraux du contrôle moteur de la parole ; apport de l'imagerie cérébrale fonctionnelle. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(2)*, 13–23.

(Rinkel et al., 2008) R. N. Rinkel, I. M. V. de Leeuw, E. J. van Reij, N. K. Aaronson, et R. Leemans, 2008. Speech handicap index in patients with oral and pharyngeal cancer : better understanding of patients' complaints. *Head and Neck 30 (7)*, 868–874.

(Robert, 2008) V. Robert, 2008. *Modélisation de la co-articulation labiale : Mise en oeuvre sur une tête parlante*. Thèse de Doctorat, Université d'Henri Poincaré, Nancy I.

(Rosen et Yampolsky, 2000) K. M. Rosen et S. Yampolsky, 2000. Automatic speech recognition and a review of its functioning with dysarthric speech. *Augmentative and Alternative Communication (AAC) 16(1)*, 48–60.

(Roy et al., 2004) N. Roy, R. M. Merril, R. Parsa, S. S. Gray, et E. M. Smith, 2004. Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research 47(2)*, 281–293.

(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*, 80–88. Association for Computational Linguistics.

(Rudzicz, 2012) F. Rudzicz, 2012. Using articulatory likelihoods in the recognition of dysarthric speech. *Speech Communication 54(3)*, 430–444.

(Rudzicz et al., 2012) F. Rudzicz, A. K. Namasivayam, et 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*, 523–541.

(Saenz-Lechon et al., 2006) N. Saenz-Lechon, J. I. Godino-Llorentea, V. Osma-Ruiza, et P. Gomez-Vilda, 2006. Methodological issues in the development of automatic systems for voice pathology detection. *Journal of Biomedical Signal Processing and Control, Elsevier*.

(Sáenz-Lechón et al., 2008) N. Sáenz-Lechón, V. Osma-Ruiz, J. I. Godino-Llorente, M. Blanco-Velasco, F. Cruz-Roldán, et J. D. Arias-Londono, 2008. Effects of audio compression in automatic detection of voice pathologies. *IEEE Transactions on Biomedical Engineering 55*(12), 2831–2835.

(Sanders et al., 2002) E. Sanders, M. Ruiters, L. Beijer, et H. Strik, 2002. Automatic recognition of dutch dysarthric speech : a pilot study. *Proceedings of International Conference on Spoken Language Processing (ICSLP'02)*, 661–664.

(Sapir et al., 1993) S. Sapir, A. Keidar, et B. Mathers-Schmidt, 1993. Vocal attrition in teachers : survey findings. *European Journal of disorders of communication 28*, 177–185.

(Sapir et al., 2010) S. Sapir, L. O. Ramig, J. L. Spielman, et C. Fox, 2010. Formant centralization ratio : a proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research 53*(1), 114–125.

(Saz et al., 2006) O. Saz, A. Miguel, E. Lleida, A. Ortega, et L. Buera, 2006. Study of time and frequency variability in pathological speech and error reduction methods for automatic speech recognition. *Proceedings of Interspeech'06*, Pittsburgh, US.

(Saz et al., 2008) O. Saz, W.-R. Rodriguez, E. Lleida, et C. Vaquero, 2008. A novel corpus of children disordered speech. *First Workshop on Child, Computer and Interaction*, Chania, Greece.

(Saz et al., 2009) O. Saz, S. C. Yin, E. Lleida, R. Rose, C. Vaquero, et W. R. Rodriguez, 2009. Tools and technologies for computer-aided speech and language therapy. *Speech communication 51*, 948–967.

(Schipor et al., 2012) O. A. Schipor, S. G. Pentiuc, et M. D. Schipor, 2012. Improving computer based speech therapy using a fuzzy expert system. *Computing and Informatics 29*(2), 303–318.

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

(Schuller et al., 2015) B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hönig, J. R. Orozco-Arroyave, E. Nöth, Y. Zhang, et F. Weninger, 2015. The interspeech 2015 computational paralinguistics challenge : Nativeness, Parkinson's & eating condition. *Proceedings of Interspeech'15*, Dresden, Germany.

(Sehgal et Cunningham, 2015) S. Sehgal et S. Cunningham, 2015. Model adaptation and adaptive training for the recognition of dysarthric speech. *6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, Dresden, Germany, 65.

(Selouani et al., 2009) S.-A. Selouani, M. S. Yakoub, et D. O'Shaughnessy, 2009. Alternative speech communication system for persons with severe speech disorders. *EURASIP Journal on advances in signal processing 2009*.

(Seong et al., 2012a) W. K. Seong, J. H. Park, et H. K. Kim, 2012a. Dysarthric speech recognition error correction using weighted finite state transducers based on context–dependent pronunciation variation. *Computer helping people with special needs, Lecture notes in computer science 7383*, 475–482.

(Seong et al., 2012b) W. K. Seong, J. H. Park, et H. K. Kim, 2012b. Performance improvement of dysarthric speech recognition using context-dependent pronunciation variation modeling based on kullback-leibler distance. *Advanced Science and Technology Letters 14(1)*, 53–56.

(Shahamiri et Binti Salim, 2014) S. R. Shahamiri et S. S. Binti Salim, 2014. Artificial neural networks as speech recognisers for dysarthric speech : Identifying the best-performing set of mfcc parameters and studying a speaker-independent approach. *Advanced Engineering Informatics 28(1)*, 102–110.

(Shama et al., 2007) K. Shama, N. U. Cholayya, et al., 2007. Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP Journal on Applied Signal Processing 2007*(1), 50–50.

(Sharma et Hasegawa-Johnson, 2010) H. V. Sharma et M. Hasegawa-Johnson, 2010. State-transition interpolation and map adaptation for hmm-based dysarthric speech recognition. *Workshop on Speech and Language processing for assistive technologies, SLPAT'10*.

(Sharma et Hasegawa-Johnson, 2013) H. V. Sharma et M. Hasegawa-Johnson, 2013. Acoustic model adaptation using in-domain background models for dysarthric speech recognition. *Computer Speech & Language 27(6)*, 1147–1162.

(Sharma et al., 2009) H. V. Sharma, M. Hasegawa-Johnson, J. Gunderson, et 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, et T. Synder, 1990. Tabletop versus microcomputer-assister speech management : response evocation phase. *Journal of speech and hearing disorders 55*, 635–655.

(Simberg et al., 2005) S. Simberg, E. Sala, et K. Vehmas, 2005. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *Journal of Voice 19*, 95–102.

(Simonyan et Ludlow, 2007) K. Simonyan et C. L. Ludlow, 2007. La production vocale. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(5)*, 67–73.

(Solomon, 2007) N. P. Solomon, 2007. La fonction respiratoire dans la production de parole. *Les dysarthries, édition Solal Neurophysiologie et production de la parole, Part I(4)*, 44–55.

(Story, 2002) B. H. Story, 2002. An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology 23(4)*, 195–206.

(Story et Titze, 1995) B. H. Story et I. R. Titze, 1995. Voice simulation with a body-cover model of the vocal folds. *Journal of the Acoustical Society of America 97(2)*, 1249–1260.

(Teston, 2001) B. Teston, 2001. L'évaluation objective des dysfonctionnement de la voix et de la parole : 2e partie : les dysphonies. *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence 20*, 169–232.

(Teston, 2004) B. Teston, 2004. L'évaluation instrumentale des dysphonies. etat actuel et perspectives. *Le bilan d'une dysphonie*, 105–169.

(Teston, 2007) B. Teston, 2007. L'étude instrumentale des gestes dans la production de la parole ; importance de l'aérophonométrie. *Les dysarthries, édition Solal Evaluation, Part II(24)*, 248–258.

(Thomas-Stonell et al., A998) T. Thomas-Stonell, A.-L. Kotler, H. A. Leeper, et P. C. Doyle, A998. Computerized speech recognition : influence of intelligibility and perceptual consistency on recognition accuracy. *Augmentative and Alternative Communication (AAC) 14*, 51–56.

(Titze, 1994) I. R. Titze, 1994. *Principles of voice production*. Prentice-Hall. Englewood Cliffs.

(Titze et Alipour, 2006) I. R. Titze et F. Alipour, 2006. *The myoelastic aerodynamic theory of phonation*. National Center for Voice and Speech.

(Titze et al., 1997) I. R. Titze, J. Lemke, et D. Montequin, 1997. Populations in the u.s. workforce who rely on voice as a primary tool of trade : A preliminary report. *Journal of Voice 11*, 254–259.

(Tomik et Guiloff, 2010) B. Tomik et J. Guiloff, 2010. Dysarthria in amyotrophic lateral sclerosis : a review. *Amyotrophic Lateral Sclerosis 11 (1–2)*, 4–15.

(Torreira et al., 2010) F. Torreira, M. Adda-Decker, et M. Ernestus, 2010. The nijmegen corpus of casual french. *Speech Communication 52*(3), 201–212.

(Tubach, 1989) J. Tubach, 1989. *La parole et son traitement automatique*. Masson, collection technique et scientifique des télécommunications.

(Turk et Arslan, 2005) O. Turk et M. Arslan, 2005. Software tools for speech therapy and voice quality monitoring. *13th European Signal Processing Conference - EUSIPCO*, Antalya, Turkey.

(Vaissière, 1999) J. Vaissière, 1999. Utilisation de la prosodie dans les systèmes automatiques : un problème d'intégration des différentes composantes. *Faits de Langues, Oral-écrit : Formes et théories, Ophys (9)16*.

(Vaissière, 2006) J. Vaissière, 2006. *La phonétique*. Paris : Que sais-je ? Presses Universitaires de France (PUF).

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

(Vasilakis et Stylianou, 2009) M. Vasilakis et Y. Stylianou, 2009. Voice pathology detection based eon short-term jitter estimations in running speech. *Folia Phoniatrica et Logopaedica 61*(3), 153–170.

(Verdolini et Ramig, 2001) K. Verdolini et L. O. Ramig, 2001. Review : Occupational risks for voice problems. *Logopedics, phoniatrics, vocology journal 26*, 37–46.

(Viallet et Teston, 2007) F. Viallet et 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.

(Walshe et al., 2009) M. Walshe, R. K. Peach, et N. Miller, 2009. Dysarthria impact profile development of a scale to measure psychosocial effects. *International Journal of Language and Communication Disorders 44(5)*, 693–715.

(Weismer et al., 2010) G. Weismer, Y. Kim, B. Maassen, et P. van Lieshout, 2010. Classification and taxonomy of motor speech disorders : What are the issues. *Speech motor control : New developments in basic and applied research*, 229–241.

(Wilson, 1987) D. K. Wilson, 1987. *Voice problem of children* (3rd ed. Baltimore MD : Williams & Wilkins ed.).

(Yorkston et Beukelman, 1981) K. M. Yorkston et D. R. Beukelman, 1981. Assessment of intelligibility of dysarthric speech. *Tigard, OR : C.C. Publications*.

(Yorkston et al., 1996) K. M. Yorkston, E. Strand, et M. Kennedy, 1996. Comprehensibility of dysarthric speech : implications for assessment and treatment planning. *American Journal of Speech Language Pathology 55*, 55–66.

(Yorkston et al., 1993) K. M. Yorkston, E. Strand, R. Miller, A. Hillel, et K. Smith, 1993. Speech deterioration in amyotrophic lateral sclerosis : implications for the timing of intervention. *Journal Med Speech Language Pathology 46*, 35–46.

(Young et Mihailidis, 2010) V. Young et 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.

(Zhang et al., 2004) Y. Zhang, C. McGilligan, L. Zhou, M. Vig, et J. J. Jiang, 2004. Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. *The Journal of the Acoustical Society of America 115*(5), 2270–2277.