



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 13043

To cite this version : Maclair, Julie and Pellegrini, Thomas and Le Coz, Maxime and Robert, Marina and Gatignol, Peggy *[Caractérisation acoustico-phonétique de parole provenant de patients atteints de paralysies faciales](#)*. (2014) In: Journées d'Etudes sur la Parole - JEP 2014, 23 June 2014 - 27 June 2014 (Le Mans, France).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Caractérisation acoustico-phonétique de parole provenant de patients atteints de paralysies faciales

Julie Mauclair^{1, 2} Thomas Pellegrini¹ Maxime Le Coz¹

Marina Robert³ Peggy Gatignol³

(1) IRIT-Université Paul Sabatier, Toulouse, France

(2) Université Paris Descartes, Paris, France

(3) Service ORL et de chirurgie de la face et du cou, Hôpital de la Pitié Salpêtrière, Paris, France

{mauclair,pellegrini,lecoz}@irit.fr

{marina.robert,peggy.gatignol}@psl.aphp.fr

RÉSUMÉ

La caractérisation de la parole pathologique est un problème important pour la détection et le diagnostic précoces. La parole pathologique étudiée dans cet article, provient de patients paralysés faciaux, souffrant notamment de problèmes de prononciation des plosives. Plusieurs paramètres sont extraits d'un détecteur automatique de bursts en vue de caractériser cette altération. L'une des autres hypothèses étudiées ici est que les cibles vocaliques souffrent de la mauvaise prononciation des consonnes, et ont une réalisation approximative. Une analyse des voyelles et des bursts est donc proposée en fonction des différents grades de sévérité de la paralysie des patients. Une diminution significative du nombre de bursts, des écart-type des durées de ceux-ci et du F2 des voyelles /i, e, ε, a/ ont été observés pour les patients du corpus. Le F3 des voyelles /i, y/ est affecté, montrant que les patients ont du mal à produire ces voyelles, faisant intervenir les lèvres.

ABSTRACT

Phonetic-acoustic characterization of speech pronounced by patients impaired with facial paralysis

Characterization of pathological speech is important for early detection and diagnosis. The pathological speech that is studied in this paper is from patients who have a peripheral facial paralysis which, among other pronunciation impairments, affects their ability to pronounce bilabial sounds. The idea is then to analyse the bursts that should characterize the bilabial sounds and the vowels that suffer from the bad pronunciation of the consonants. Indeed, this article analyses the vowels and the bursts that are uttered by patients, in correlation to the various levels of the paralysis impairment. No significant differences were observed concerning the vowel durations. A decrease in the number and in the standard deviation of the duration of the bursts, as well as a decrease in the second formant values were observed for the following vowels : /i, e, ε, a/. Differences in the third formant values appeared for the posterior vowels /i, y/, showing that the patients have difficulties with moving their lips.

MOTS-CLÉS : Paramètres acoustico-phonétiques, traitement automatique de la parole, parole pathologique.

KEYWORDS: Acoustical and phonetical features, automatic speech processing, pathological speech.

1 Introduction

La paralysie faciale périphérique est une pathologie fréquente qui a de nombreuses conséquences dans le quotidien des patients. Parmi les conséquences notables de cette paralysie, l'articulation des patients est affectée (Gatignol et Lamas, 2004). De nos jours, il y existe toute une batterie d'outils disponible, permettant de mesurer le degré de l'atteinte mais la plupart ne tiennent compte que de l'aspect physique du visage et ne mesure pas la capacité du patient à prononcer correctement des mots ou des phrases.

L'échelle d'évaluation clinique la plus utilisée au niveau internationale est celle de House et Brackmann (House et Brackmann, 1985) qui permet une cotation du degré de dysfonction faciale rapide et intégrant l'évaluation des séquelles, allant du grade I (fonctionnement normal) au grade VI (paralysie flasque totale). C'est une échelle à 6 niveaux qui permet uniquement de noter les aspects physiques de la paralysie (principalement le front, les yeux et la bouche) grâce à une observation directe effectuée par un praticien. Le tableau 1 met l'accent sur les difficultés au niveau de la bouche au travers des différents grades de l'échelle de House et Brackmann. Pour les patients atteints de paralysie faciale périphérique à partir du grade III, la littérature fait état d'un retentissement psychologique et de troubles fonctionnels tels que les troubles oculaires, les syncinésies, l'hémi-spasme, les troubles de la mastication, de la déglutition et de l'articulation (Gatignol et Lamas, 2004).

Grade	Caractéristiques
I Normal	Fonction faciale normale
II Atteinte discrète	Légère parésie. Au repos, symétrie et tonus normaux. Bouche : légère asymétrie
III Atteinte modérée	Asymétrie évidente mais non choquante. Au repos, symétrie et tonus normaux Bouche : faiblesse des muscles quand l'effort est maximum
IV Atteinte modérément sévère	Faiblesse ou asymétrie défigurante Au repos, tonus et symétrie normaux Bouche : asymétrique quand l'effort est maximum
V Atteinte sévère	Mouvement à peine perceptible Asymétrie au repos Bouche : mouvement très faible
VI Paralysie totale	Aucun mouvement

TABLE 1 – Les mouvements de la bouche référencés dans l'échelle de House et Brackman

Dans le domaine du traitement automatique de la parole, l'analyse de la voix et notamment de la parole pathologique est utilisée pour aider au diagnostic médical. L'analyse automatique des voix est une méthode non-invasive et est largement expérimentée en ce qui concerne les pathologies laryngées et la voix dysphonique (Drugman *et al.*, 2011; Henriquez *et al.*, 2009) afin de fournir une évaluation objective aux praticiens. L'analyse acoustique des voix pathologiques peut être pratiquée sur des voyelles tenues ou des phrases et utilise des méthodes de traitement du signal linéaires ou non linéaires (Henriquez *et al.*, 2009; Arias-Londoño *et al.*, 2011). Souvent le but de ces expériences est de discriminer la voix pathologique de la voix non-pathologique (Drugman *et al.*, 2011; Henriquez *et al.*, 2009; Arias-Londoño *et al.*, 2011). Les troubles de la parole, notamment la caractérisation de la parole dysarthrique ont par exemple été étudiés dans le projet DesPho APady (Fougeron *et al.*, 2010). Plusieurs articles font de la classification de parole dysarthrique pour caractériser plusieurs types de dysarthrie. La classification porte notamment sur des paramètres caractérisant les spécificités acoustico-articulatoires de ce type de parole (Auzou *et al.*, 2007) et plusieurs études utilisent des systèmes de traitement automatique de la parole (Gu *et al.*, 2005) pour objectiver la sévérité d'un trouble de parole. Pour les atteintes faciales, une analyse automatique de la parole pourrait aider le patient à mesurer objectivement ses progrès lors d'une thérapie. Celle-ci est un très long processus et il est souvent difficile pour les patients d'être objectifs sur l'amélioration de leur condition, même si celle-ci est faite par leurs proches ou même par les médecins.

Ici, le but de cet article est de vérifier s'il y a une corrélation entre l'atteinte physique faciale prise en compte dans l'échelle de House et Brackmann avec les difficultés articulatoires dont souffrent les patients. Afin d'appréhender les difficultés articulatoires, plusieurs paramètres caractérisant les réalisations acoustiques sont calculés pour représenter le degré d'atteinte des patients. A l'écoute de la parole des patients, l'impact le plus clair de la paralysie porte sur la réalisation approximative des consonnes et en particulier des plosives (Gatignol et Lamas, 2004). Les voyelles proches sont alors aussi altérées et leur réalisation est réduite. Pour vérifier ces hypothèses dégagées grâce à une analyse et un alignement manuel dans (Robert *et al.*, 2011), un système de détection automatique de bursts est utilisé (paragraphe 3.1). Ce système, déjà décrit dans une précédente étude pour une classification préliminaire (Mauclair *et al.*, 2013), permet ici une description plus précise du phénomène d'altération au niveau des plosives. De plus, une étude des altérations des voyelles au niveau des différents grade de paralysie faciale est également proposée (paragraphe 3.2).

Le plan de l'article est construit comme suit. La base de données dont un corpus de phrases isolées est extrait pour cet article, est décrite dans le paragraphe 2. Ensuite, nous décrirons les paramètres étudiés pour les expériences dans le paragraphe 3. Enfin, une discussion des résultats et une conclusion à ce travail seront abordées.

2 Corpus utilisé

Les expériences ont été réalisées sur une base de données enregistrée à l'hôpital de la Pitié Salpêtrière de Paris. Celle-ci contient les enregistrements de 70 personnes, des patients et des témoins sains. Les patients souffrent de paralysie faciale périphérique d'origine virale, dûe à des neurinomes (du nerf acoustique ou du nerf facial) ou encore d'origine traumatique. De plus amples détails sur les patients de ce corpus sont disponibles dans (Robert, 2011). Quelques

enregistrements proviennent du même patient à plusieurs grades de paralysie. Les patients ont été enregistrés dans une cabine insonorisée avec un microphone hypercardoïde, grâce à un enregistreur numérique au format 16 bits/44,1kHz PCM WAV.

Plusieurs types d'enregistrements ont été réalisés. Les patients ont eu plusieurs textes à lire, comme des mots isolés, des phrases isolées, des passages de journaux. Ils ont aussi répondu de façon spontanée à la question : "Comment faites vous une omelette ?"

Dans cet article, seule la partie où les patients et témoins ont lu les phrases isolées a été exploitée. 32 enregistrements de 32 patients ont été utilisés, équitablement répartis dans quatre classes que nous détaillerons par la suite. Chaque personne prononce 17 phrases (telles que : "Pierre et Paul préparent la purée de petits pois"), ce qui correspond approximativement à 1 minute par patient. L'épreuve de lecture de phrases comporte 14 phrases testant le lieu d'articulation des phonèmes : bilabiaux, alvéolaires, vélaux, labiodentaux et palataux. Trois autres phrases opposent les modes d'articulation : occlusives et constrictives constrictives sonores et sourdes et nasales.

Les quatre classes étudiées dans ce papier correspondent à 5 grades de paralysie faciale (voir Tableau 1), les grades I, III, IV et les V&VI. Ce choix d'une classification un peu différente des grades de la paralysie a été fait pour deux raisons. La première est que, les patients de grade II ne ressentent plus le besoin de venir consulter, ils n'ont donc pas pu être enregistrés. Ensuite, les différences entre les grades V et VI ne correspondent pas aux mouvements de la bouche et nous avons choisi de les regrouper dans la même classe. Le grade I est constitué de patients témoins n'ayant jamais présenté de paralysie faciale, car il est quasi-impossible pour les paralysés de revenir à une fonction de la face normale.

3 Expériences

Pour étudier ces différentes classes, plusieurs paramètres acoustico-phonétiques sont calculés. Dans un premier temps, nous avons choisi des paramètres caractérisant les plosives. En effet, ces phonèmes sont, à l'écoute des enregistrements des patients, assez mal prononcés (Robert *et al.*, 2011). Dans une précédente étude (Mauclair *et al.*, 2013), cette paramétrisation était faite à l'aide d'un détecteur de bursts. Nous reprenons ici cette paramétrisation pour en extraire une étude plus fine dans le paragraphe 3.1. Une autre hypothèse étudiée ici est que nous pensons que si plusieurs consonnes sont altérées par la paralysie faciale, les voyelles alentour peuvent l'être aussi, d'une part par le phénomène de coarticulation provenant de la consonne altérée précédente ou suivante et d'autre part par la nécessité pour la production de quelques-unes de faire intervenir les lèvres. Nous avons donc étudié l'espace vocalique de celles-ci (voir paragraphe 3.2). Pour les analyses statistiques, nous nous sommes basés sur un T-test avec un seuil de significativité de 10%.

3.1 Analyse des bursts

La détection des bursts se fait automatiquement grâce à la méthode développée dans (Hosom et Cole, 2000). Cette méthode permet d'obtenir les valeurs en amplitude des pics d'énergie en fonction du temps. Ici, le but n'est pas de connaître le nombre de bursts correctement détectés mais plutôt de voir s'il y a une différence dans le nombre de bursts détectés en fonction des

différents grades et si la représentation des bursts diffère. La figure 1 montre qu'il existe une diminution du nombre de bursts détectés en fonction du grade de la paralysie.

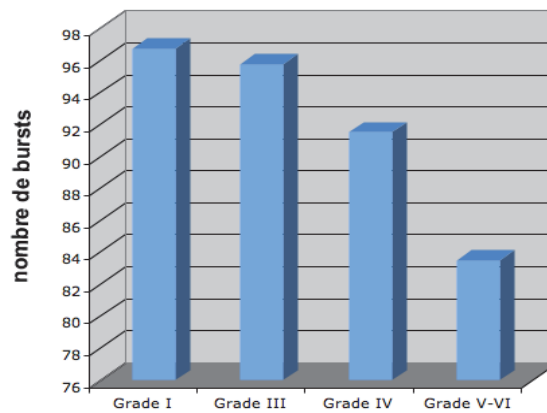


FIGURE 1 – Représentation du nombre de bursts par grade de paralysie faciale

Ces bursts forment des gaussiennes (voir figure 2)dont l'amplitude et la dispersion sont intéressantes à étudier pour caractériser le degré d'altération dans la prononciation des plosives. En effet, l'écart type des durées des bursts (et donc la dispersion de ces bursts) semble être plus élevé avec des grades de paralysies élevés, car la paralysie faciale entraîne un ralentissement dans la production de cette explosion.

Plusieurs paramètres sont extraits du résultat du détecteur de bursts pour caractériser ces plosives. Nous avons choisi d'étudier le nombre de bursts détectés, le nombre de bursts détectés dont l'amplitude est supérieure à la moyenne de nombre de bursts, la moyenne des écart type des durées des bursts et le nombre de bursts dont l'écart type des durées est inférieur à la moyenne des écart type des durées.

La différence en terme de nombre de bursts détectés n'est pas significative entre les locuteurs sains (grade I) et ceux des grades III et IV. Par contre elle est significative entre les patients des grades III et V&VI et entre les locuteurs sains et ceux du grade V&VI (p-value respectivement de 0.08 et 0.06). Si on tient compte maintenant uniquement du nombre de bursts dont l'amplitude est supérieure à l'amplitude moyenne, l'écart devient significatif entre les locuteurs sains et les grades IV ainsi qu'entre les locuteurs sains et ceux du grade V (p-value de 0.09 et p-value de 0.08). Ce qui est très intéressant est la mesure des écart-types de chacun des bursts pour les différents grades. La différence est significative entre la moyenne des écart-types des durées des bursts des locuteurs du grade III et du grade I (p-value de 0.004), ainsi qu'entre le grade IV et le grade I (p-value de 0.009) et entre le grade V&VI et le grade I (p-value de 0.02). Ceci est également vérifié entre les grades pathologiques, la différence est significative. Quand on comptabilise seulement les écart-types qui sont inférieurs à la moyenne des écart-types sur la durée des bursts, seul le grade V&VI a une différence significative avec les locuteurs sains (p-value de 0.07).

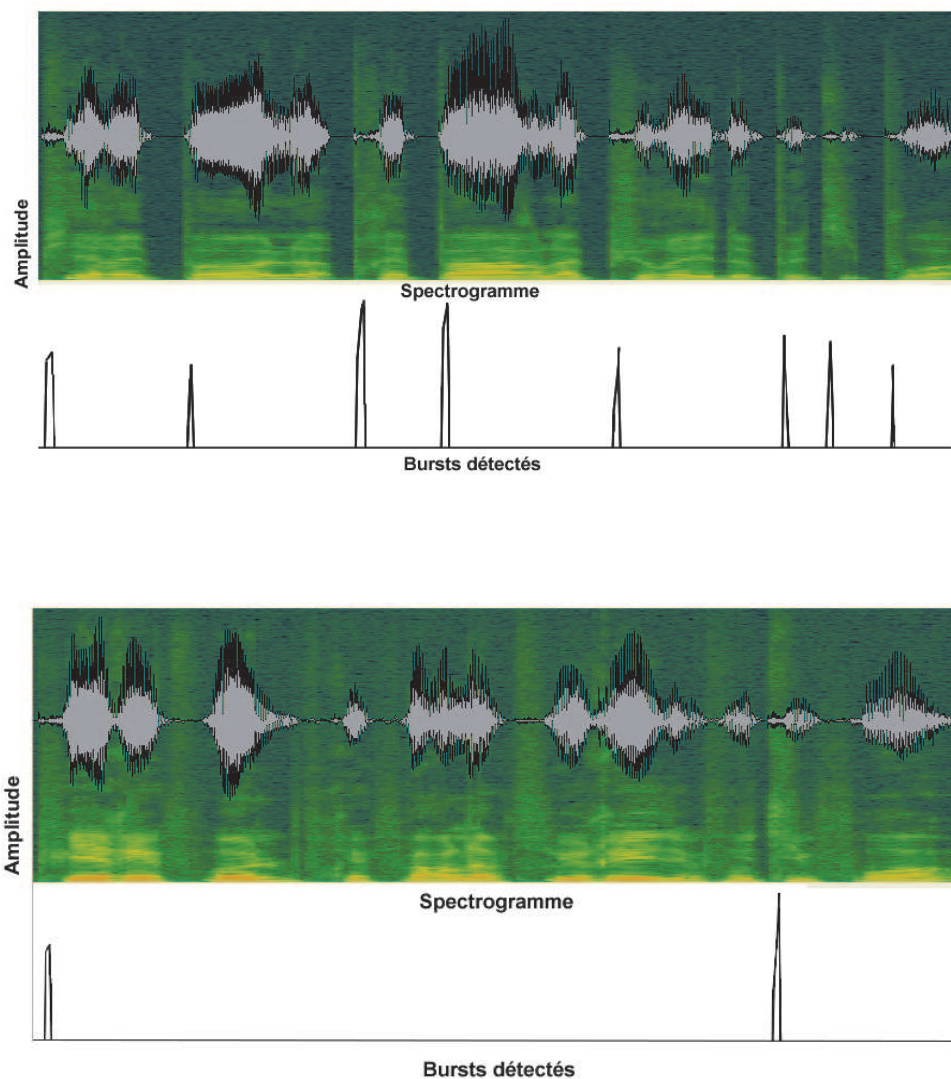


FIGURE 2 – Représentation des bursts détectés pour un sujet grade I (en haut) et un patient grade VI (en bas) prononçant : " Pierre et Paul préparent la purée de petits pois"

3.2 Espace vocalique

Pour calculer les valeurs des formants analysés, un alignement forcé est réalisé à l'aide du plugin Easyalign de Praat (Goldman, 2011; Boersma, 2001). Nous avons considéré les onze voyelles orales suivantes : a, e, i, o, u, y, ε, ə, ɔ, ø, œ. Le nombre total de voyelles segmentées est d'environ 5000. Le nombre d'occurrences de chaque voyelle correspond approximativement au pourcentage observé en français : les voyelles /a/ et /œ/ sont la plus fréquente (27%) et la moins fréquente (1%) respectivement. À l'aide du logiciel Praat, nous avons extrait F0, les trois premiers formants, l'énergie et les durées de ces voyelles. Un filtrage sur le deuxième formant a été réalisé pour écarter les voyelles donnant des valeurs aberrantes, en prenant une tolérance de plus ou moins 300Hz autour de valeurs canoniques tirées de (Gendrot et Adda-Decker, 2005).

La figure 3 représente les triangles vocaliques des locuteurs masculins (à gauche) et féminins (à droite) du corpus. Les triangles vocaliques des groupes III et IV sont très similaires au triangle du

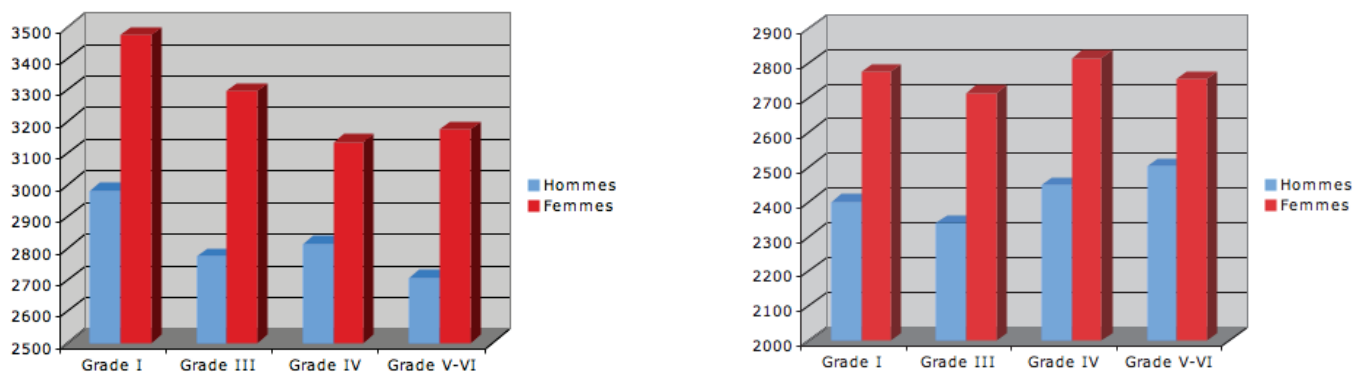


FIGURE 4 – Valeurs des troisièmes formants pour les voyelles /y/ (à gauche) et /i/ (à droite) pour les hommes et les femmes

2502Hz pour le groupe V&VI contre 2400Hz pour le groupe I (p -value=0.02). Chez les femmes, une augmentation est observée si l'on compare le groupe I (2775Hz) et le groupe IV (2814Hz). En revanche, une faible diminution de 20Hz est observée pour les groupes III et V&VI.

En ce qui concerne les durées des voyelles, aucune différence significative n'a été mise en évidence entre les différents groupes de locuteurs (durée moyenne autour de 100ms).

4 Discussion et conclusion

Nous avons étudiés ici deux groupes de phonèmes qui nous semblent altérés par la paralysie faciale. Les différences sont en générales très significatives entre les paramètres des locuteurs sains et ceux des locuteurs de grade V&VI. Cependant, elles s'atténuent pour les grades qui se trouvent entre ces extrêmes. Les seuls paramètres qui permettent de distinguer les classes médianes sont les moyennes des écart-types des bursts, les valeurs du F2 pour les voyelles /e, ε/ et les valeurs du F3 pour la voyelle /i/. Il est également vrai qu'à l'écoute des enregistrements, il est parfois difficile d'entendre la difficulté d'articulation, notamment de certains patients de grade III. Ces patients ont en effet parfois obtenu cette gradation plutôt à cause de la mauvais mobilité de leur oeil que de celle de leur bouche et ont, à l'écoute des enregistrement, plus d'aisance en lecture que d'autres patients de leur groupe.

Dans cet article, nous avons proposé une caractérisation de plusieurs phénomènes acoustico-phonétiques en vue d'étudier des enregistrements de patients atteints de paralysie faciale. La classification en terme de sévérité de l'atteinte proposée par les orthophonistes est une classification sur une échelle tenant compte non seulement des mouvements de la bouche mais également des mouvements de la joue et de l'oeil. Cette classification a donc des limites en terme de ce que l'on peut espérer corrélérer avec la prononciation des patients. Cependant, entre les grades extrêmes, la caractérisation au niveau des voyelles et en terme de détection des bursts donne des résultats significativement différente pour espérer une classification plus efficace que celle auparavant réalisée avec la détection de bursts seule dans (Mauclair *et al.*, 2013). Une paramétrisation agrégeant cette caractérisation en terme de bursts à une caractérisation de certaines voyelles pourraient améliorer les résultats.

Références

- ARIAS-LONDOÑO, J., GODINO-LLORENTE, J., SÁENZ-LECHÓN, N., OSMA-RUIZ, V. et CASTELLANOS-DOMÍNGUEZ, G. (2011). Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *IEEE Transactions on Biomedical Engineering*, 58(2):370–379.
- AUDIBERT, N. et FOUGERON, C. (2012). Distorsions de l'espace vocalique : quelles mesures ? application à la dysarthrie. *Journées d'Étude sur la Parole (JEP 2012)*.
- AUZOU, P., OZSANCAK, C., PINTO, S. et ROLLAND, V. (2007). *Les dysarthries*. Solal.
- BOERSMA, P. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5, 9:341–345.
- DRUGMAN, T., DUBUISSON, T. et DUTOIT, T. (2011). Phase-based information for voice pathology detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4612–4615, Prague, Czech Republic.
- FOUGERON, C., CREVIER-BUCHMAN, L., FREDOUILLE, C., GHIO, A., MEUNIER, C., CHEVRIE-MULLER, C., AUDIBERT, N., BONASTRE, J.-F., SIMON, A., DELOOZE, C., DUEZ, D., GENDROT, C., LEGOU, T., LÉVÊQUE, N., PILLOT, C., PINTO, S., POUCHOULIN, G., ROBERT, D., VAISSIERE, J., VIALLET, F. et VINCENT, C. (2010). Developing an acoustiphonetic characterization of dysarthric speech in french. In *7th international conference on Language Resources and Evaluation*, pages 2831–2838.
- GATIGNOL, P. et LAMAS, G. (2004). *Paralysies faciales*. Solal Editions.
- GENDROT, C. et ADDA-DECKER, M. (2005). Impact of duration on F1/F2 formant values of oral vowels : an automatic analysis of large broadcast news corpora in french and german. *Interspeech 2005*, pages 2453–2456.
- GOLDMAN, J. (2011). Easyalign : an automatic phonetic alignment tool under praat. *Interspeech 2011*, pages 3233–3236.
- GU, L., HARRIS, J.-G., RAHUL, S. et SAPIENZA, C. (2005). Disordered speech evaluation using objective quality measures. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 321–324.
- HENRIQUEZ, P., ALONSO, J., FERRER, M., TRAVIESO, C., GODINO-LLORENTE, J. et Diaz-de MARIA, F. (2009). Characterization of healthy and pathological voice through measures based on nonlinear dynamics. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6):1186–1195.
- HOSOM, J.-P. et COLE, R. A. (2000). Burst detection based on measurements of intensity discrimination. In *International Conference on Spoken Language Processing (ICSLP)*, pages 564–567, Beijing, China.
- HOUSE, J. et BRACKMANN, D. (1985). Facial nerve grading system. *Otolaryngology-Head and Neck Surgery*, 93:146–147.
- MAUCLAIR, J., KOENIG, L., ROBERT, M. et GATIGNOL, P. (2013). Burst-based features for the classification of pathological voices. *Interspeech 2013*, pages 2167–2171.
- ROBERT, M. (2011). Analyse acoustique des troubles articulatoires chez les patients atteints de paralysie faciale périphérique. Mémoire de Master, Ecole d'Orthophonie de Paris VI.
- ROBERT, M., MAUCLAIR, J., LANNADERE, E., TANKÉRE, F., LAMAS, G. et GATIGNOL, P. (2011). Analyse des troubles articulatoires au cours des paralysies faciales périphériques. *LXVII Congrès de la Société Française de Phoniatrie*.