



## LOGICIEL D'EVALUATION ACOUSTIQUE DES PATHOLOGIES PHONATOIRES CHEZ L'ENFANT

Fabrice PLANTE, Jocelyne BOREL, Christian BERGER-VACHON

Laboratoire "Audition et Voix" URA CNRS 1447  
Pavillon U, Hôpital E. Herriot, 69437 LYON Cédex

### RÉSUMÉ

Dans ce travail, les auteurs présentent un système d'évaluation acoustique des pathologies phonatoires. Contrairement aux multiples études réalisées, ce travail a porté spécifiquement sur l'enfant. Nous nous sommes intéressés à des pathologies des cordes vocales (dysphonies) et à l'insuffisance vélaire). De multiples paramètres ont été utilisés. Ils sont basés sur des considérations de traitement du signal et sur les connaissances des modifications spectrales dues aux pathologies. Les résultats présentés dans ce travail ont été obtenus sur plus de 200 sujets âgés de 5 à 12 ans.

### 1. INTRODUCTION

Les pathologies phonatoires touchent 30% de la population infantile. Une détection précoce permet d'une part, de dépister les pathologies malignes et ainsi d'éviter leur aggravation, et d'autre part, d'éviter que l'enfant n'acquière de mauvais gestes phonatoires en l'orientant tout de suite vers une rééducation adaptée. Cependant, il est difficile d'obtenir un diagnostic précis chez l'enfant car l'exploration clinique de l'appareil phonatoire n'est pas toujours aisée à réaliser. De plus, l'efficacité de la rééducation orthophonique dépend de la motivation de l'enfant. Il faut donc lui montrer des signes objectifs de ses progrès. L'analyse du signal de parole semble donc être une bonne méthode, complémentaire des examens classiques.

De nombreuses études ont déjà été réalisées sur la détection de la pathologie [1, 2, 3, 4]. La majorité d'entre elles se sont focalisées sur les pathologies laryngées de l'adulte. Dans ce travail, nous nous sommes intéressés à toutes les pathologies phonatoires de l'enfant, aussi bien au niveau des cordes vocales (dysfonction, kissing nodules, ...) que des cavités de résonances (insuffisance vélaire). Dans le cas de cette pathologie, l'analyse acoustique que nous avons réalisée est novatrice [5].

Afin de prendre en compte toutes les perturbations engendrées par les pathologies, plusieurs paramètres ont été utilisés. Nous avons considéré des paramètres issus de méthodes de traitement du signal (distances cepstrales). Ces paramètres n'ont d'ailleurs jamais été utilisés dans le cas des

### ABSTRACT

In this work, the authors present a computer system of acoustic evaluation of phonatory pathologies. Contrary to previous works, this study was achieved on the child. We interested in laryngeal pathologies and velar impairment. Many parameters were studied. They are stem from signal processing and take into account the spectral modifications produce by the pathologies. Results exposed in this work were obtained with more than 200 subjects aged 5 to 12.

pathologies laryngées. D'autres part nous avons développé ou repris des paramètres qui prenaient en considération les connaissances que nous avons sur les modifications acoustiques dues aux pathologies. Ces différentes méthodes, ainsi que les interfaces nécessaires à leur utilisation ont été programmées au sein d'un logiciel [5].

### 2. METHODES

L'insuffisance vélaire et les dysphonies étant des pathologies différentes, les supports phonétiques d'analyses et les paramètres utilisés sont différents.

#### A. Populations

La population étudiée est composée de 209 enfants témoins, 52 insuffisants vélaire et 88 sujets dysphoniques. Le tableau 1 montre la répartition des sujets dans les différents groupes suivant le sexe. Une orthophoniste a vérifié qu'aucun des sujets étudiés n'avait mué.

Groupes	Filles	Garçons	Total
Témoins références	60	60	120
Témoins tests	42	47	89
Total témoin	102	107	209
Insuffisants vélaire	19	33	52
Dysphoniques	36	52	88

Tableau 1 : Répartition des sujets suivant le sexe.



Le groupe des sujets témoins a été divisé en deux groupes. Le groupe des témoins références va permettre d'établir les références et les intervalles de normalité utilisés dans le système de décision. Les sujets tests vont permettre de définir le taux de fausses alarmes du système.

### B. Textes

L'insuffisance vélaire est caractérisée par une nasalisation des phonèmes et une difficulté à augmenter la pression au niveau du conduit vocal (difficulté à former les occlusives et les fricatives). Les voyelles isolées /a/, /i/ et /u/ qui représentent les extrema de l'espace vocalique et le /t/ du mot "gâteau" ont été analysés.

Les pathologies laryngées entraînent un mauvais accolement des cordes vocales. Il s'ensuit des perturbations du fondamental (fréquence et amplitude) [6], ainsi qu'une apparition de bruit [7]. Pour ces pathologies, les voyelles isolées /a/, /i/ et /u/ et la voyelle tenue /a/ ont été analysées.

L'acquisition s'est faite à une fréquence d'échantillonnage de 20 kHz avec un filtre antirepliement ayant une fréquence de coupure de 8 kHz. La segmentation s'est faite sur des critères visuels. Pour les voyelles isolées, la zone centrale a été conservée. Elle dure entre 75 et 130 ms. Pour le /a/ tenu, la portion finale de la voyelle a été conservée car elle est la plus pertinente pour la discrimination [8]. La durée varie entre 0.5 et 1 secondes. Pour le /t/, la segmentation a débuté un peu avant l'explosion et se finissait avant que ne commence la voyelle /o/. Sa durée varie entre 20 et 100 ms.

### C. Paramètres acoustiques

Les pathologies engendrent des modifications spectrales. Nous avons voulu, sur les voyelles isolées, quantifier ces modifications sans tenir compte de leur localisation. Pour cela, on sait que les distances cepstrales sont les meilleures [9]. Nous avons choisi trois jeux de coefficients cepstraux, correspondant à trois représentations spectrales différentes. Les coefficients linéaires (LFCC) qui correspondent à la définition classique du cepstre. Les coefficients Mel (MFCC) qui font l'analogie avec le mode de fonctionnement de l'oreille. Enfin les coefficients LPCC calculés à partir d'une prédiction linéaire [10] qui tiennent compte d'une modélisation du conduit vocal. A ces domaines de représentation nous avons associé des distances. Nous avons choisi la distance euclidienne, qui est la plus intuitive, et la distance pondérée par la variance [11]. Comme référence, nous avons pris la moyenne des coefficients cepstraux calculés sur les voyelles des témoins références.

Nous avons aussi étudié le pourcentage d'erreur de la LPC, c'est à dire la puissance quadratique moyenne de l'erreur sur la puissance du signal. En effet ce pourcentage devrait augmenter dans le cas où le modèle du conduit vocal ne correspond pas à un filtre tous pôles (ce qui est le cas pour l'insuffisance vélaire

ou la mise en parallèle de la cavité nasale avec le conduit vocal ajoute des zéros) ou que l'excitation glottique est perturbée.

Les nombreuses études réalisées sur la nasalité ont permis d'établir la zone d'apparition des formants nasaux. Pour le /a/, le premier formant nasal apparaît aux alentours de 200-300 Hz. Pour le /i/ et le /u/, la plus grosse modification est l'apparition du deuxième formant nasal entre les deux premiers formants oraux. Nous avons donc calculé des rapports d'énergie entre des bandes spectrales pour les mettre en évidence. L'apparition des ces formants entraîne une convergence des caractéristiques spectrales des voyelles vers la caractéristique de la cible nasale [12]. Nous avons donc calculé la distance cepstrale entre les trois voyelles. Celle-ci doit être plus petite dans le cas des voix nasalisées.

L'étude de l'occlusive /t/ est plus délicate d'une part à cause de son caractère transitoire et d'autre part à cause du peu d'information dont nous disposons sur ces modifications dans le cas d'une phonation nasalisée. La première contrainte nous a orienté vers l'utilisation de méthodes temps-fréquence et plus particulièrement vers le spectrogramme. La deuxième nous a obligé à choisir un paramètre descriptif, plutôt qu'un paramètre basé sur la production des occlusives. Au vu de plusieurs spectrogrammes, la différence qui nous a semblé la plus importante était une libération d'énergie plus importante en haute fréquence pour les témoins. Après un seuillage de l'image, nous avons calculé l'énergie dans les bandes fréquentielles de 1-3 kHz et 4-7 kHz et nous en avons fait le rapport.

Pour les pathologies laryngées, de nombreux paramètres ont été développés. On peut les classer en deux grandes catégories : les paramètres de perturbations du fondamental (fréquence et amplitude) et les paramètres d'évaluation du bruit. Nous les avons calculés sur les voyelles isolées et le /a/ tenu.

Les paramètres de perturbations reviennent à calculer la différence entre une valeur et son voisinage. La différence entre les divers paramètres décrits dans la littérature vient de la définition de la différence et du voisinage [13]. Dans ce travail, le fondamental est détecté sur une fenêtre de 25.6 ms, c'est à dire que la valeur correspond à une moyenne des valeurs cycle à cycle du fondamental (3 à 10).

Pour l'évaluation du bruit, plusieurs techniques sont utilisées. La plus utilisée consiste à calculer le rapport d'énergie entre les zones harmoniques et les zones non harmoniques [14, 15]. Nous avons aussi étudié le paramètre défini par Kitajima [16] utilisant un lissage du spectre. Nous avons comparé les résultats obtenus suivant trois plages de fréquence : 0-2.5 kHz, 0-5 kHz et 0-8 kHz.

Au niveau de ces deux derniers types de paramètres, notre rôle a surtout consisté à réaliser une synthèse [5].

D. Système de détection

Pour chaque paramètre, nous avons défini un intervalle de normalité sur l'ensemble des sujets témoins références. Ces intervalles ont été calculés séparément pour les filles et les garçons.

Dans un premier temps les paramètres pertinents ont été sélectionnés. Pour cela les pourcentages de fausses alarmes et de détection ont été calculés sur les témoins tests et les sujets pathologiques. Les paramètres pour lesquels le pourcentage de fausses alarmes étaient significativement inférieur au pourcentage de détection et à 5% (limite théorique du calcul de l'intervalle de normalité) ont été sélectionnés. On a noté que la variabilité était plus importante pour les enfants âgés de 10 à 12 ans. Il semblerait donc que la mue se fasse ressentir avant qu'elle ne soit perçue.

Un sujet est détecté comme pathologique si pour au moins un des paramètres sélectionnés sa valeur sort de l'intervalle de normalité.

3. RESULTATS

A. Insuffisance vélaire

La matrice de confusion obtenue pour la détection de l'insuffisance vélaire est indiquée dans le tableau 2. En ligne figure le classement réalisé par l'orthophoniste et en colonne celui obtenu par l'analyse acoustique.

	Classement acoustique	
	Témoins	Insuffisants vélaire
Témoins	66	7
Insuffisants vélaire	10	42

Tableau 2 : Matrice de confusion pour la détection de l'insuffisance vélaire.

Globalement le score de détection est de 81% avec un taux de fausses alarmes de 10%. Cependant ce score varie suivant le sexe. Les pourcentages de détection avec leur intervalle de confiance sont de :

Filles : 68% [46%, 85%]

Garçons 87% [72%, 95%]

Pour les filles, les meilleurs coefficients sont la distance nasale calculée avec les coefficients LPCC, l'indice spectral du /i/ et la distance cepstrale calculée sur les MFCC du /a/ (coefficients MFCC et LPCC). Pour les garçons, ce sont le pourcentage d'erreur LPC du /u/, l'indice spectral du /i/, la distance euclidienne calculée sur les LPCC du /u/ et la distance nasale calculée avec les LPCC. On voit apparaître quelques différences entre les filles et les garçons dans le choix des coefficients et des voyelles.

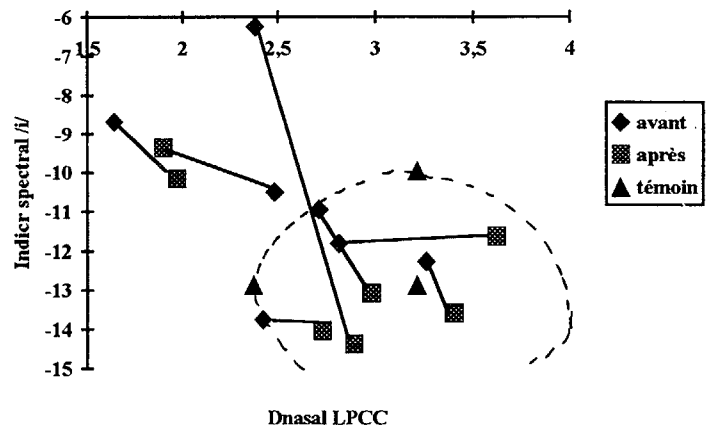


Figure 1 : Evolution des paramètres de nasalité pour 7 sujets avant et après rééducation.

Sur la figure 1, nous avons représenté dans l'espace formé par l'indice spectral du /i/ et la distance nasale, la position de 7 sujets avant et après la rééducation. L'ellipse de normalité est représentée en pointillé.

Les résultats sont assez encourageant. En effet on note un rapprochement vers le centre de normalité pour six sujets. Il est donc envisageable d'évaluer les progrès des patients. Il faut aussi noter que les paramètres acoustiques permettent d'évaluer la réussite de la vélopharyngoplastie (opération visant à réduire l'espace vélopharyngé) [5].

B. Pathologies laryngées

La matrice de confusion obtenue pour la détection des pathologies laryngées est indiquée dans le tableau 3.

	Classement acoustique	
	Témoins	Pathologiques
Témoins	85	14
Pathologiques	4	84

Tableau 3 : Matrice de confusion pour la détection des pathologies laryngées.

Globalement le score de détection est de 95% [88%, 98%] avec un taux de fausses alarmes de 16%. Il n'y a pas de différence pour les pourcentages de détection entre les filles et les garçons. Au niveau de la sélection des paramètres, par contre, des divergences sont apparues.

Comme pour la détection de l'insuffisance vélaire, il semblerait que les coefficients MFCC soient plus efficaces pour les filles, alors que pour les garçons se sont les LPCC. Ce résultat est sans doute à rapprocher de celui de la plage fréquentielle pour le calcul du bruit glottique. En effet, pour les filles la plage optimale est celle de 0 à 5 kHz, alors que pour les garçons elle est de 0 à 8 kHz. On a aussi remarqué que les paramètres de perturbation du fondamental étaient inadaptés



dans le cas des garçons et assez performants pour les filles. En ce qui concerne les distances cepstrales et le pourcentage d'erreur de la LPC, c'est le contraire.

Comme pour la détection de l'insuffisance vélaire, nous avons regardé la corrélation entre la qualité de la voix (établie perceptuellement par un jury d'orthophonistes) et les paramètres acoustiques. Il n'a pas été trouvé de corrélation significative. Sur la figure 2, nous avons représenté pour 12 sujets les différences acoustique et perceptuelle avant et après rééducation. Dans ce cas la corrélation est de 0.8. Il semble donc possible, par des paramètres acoustiques, de juger de l'évolution de la qualité vocale.

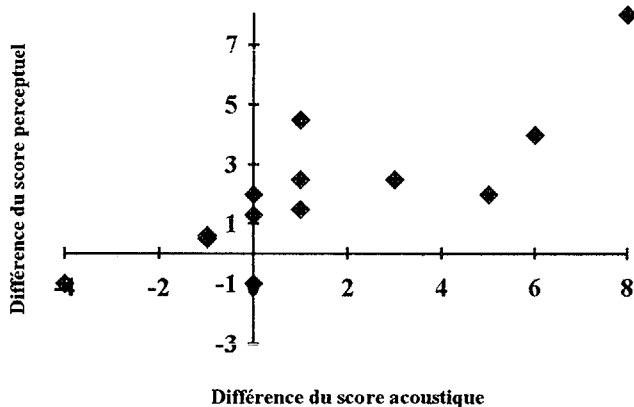


Figure 2 : Relation entre les différences des scores acoustique et perceptuel pour 12 sujets rééduqués.

#### 4. CONCLUSION

Cette étude montre la faisabilité de l'analyse acoustique pour la détection des pathologies phonatoires chez l'enfant. Ces bons scores ont pu être obtenus en différenciant les types de paramètres et donc les méthodes utilisées. Il apparaît aussi important de pouvoir prendre en compte, dans les paramètres acoustiques, des connaissances sur le système et ses modifications.

Une différence nette est apparue entre les filles et les garçons. Actuellement, aucune hypothèse n'est satisfaisante pour expliquer cette différence. De nombreuses études sont donc nécessaires pour comprendre cette divergence des résultats entre les filles et les garçons.

Il semble utopique de vouloir juger de la qualité de la voix par l'intermédiaire de paramètres acoustiques. Par contre, nous avons montré qu'il était possible d'objectiver son évolution que ce soit au cours d'une rééducation ou à la suite d'une opération. Là aussi de nombreuses études restent à faire pour définir les potentialités de la méthode. Il serait par ailleurs nécessaire de pouvoir coupler les résultats avec d'autres méthodes d'analyses (électroglottographique, aérodynamique, physiologique...).

#### REMERCIEMENTS

Cette étude a été réalisée avec le soutien de la Région Rhône-Alpes, de la société API (Audiologie Prothèse Innovation) et des Hospices Civils de Lyon. Les auteurs tiennent aussi à remercier Mme N. Gache du LTS de l'ICPI Lyon, Mr G. Chollet de l'IDIAP à Martigny et de Mr G. Feng de l'ICP Grenoble pour leurs conseils et leur aide dans la réalisation de cette étude.

#### REFERENCES

- [1] Banci G., Monini S., Falaschi A., De Sario N. "Vocal fold disorder by digital speech analysis". *J. Phonetics*, Vol.14, 495-499, 1986.
- [2] Childers D.G., Bae K.S. "Detection of laryngeal function using speech and electroglottographic data". *IEEE Trans. Biomed. Eng.* Vol.39, 19-25, 1992.
- [3] Kasuya H., Masubuchi K., Ebihara S., Yoshida H., "Preliminary experiments on voice screening". *J. Phonetics*. Vol.14, 463-468, 1986.
- [4] Laver J., Hiller S., Mackenzie J., Rooney E. "An acoustic screening system for the detection of laryngeal pathology". *J. Phonetics*, Vol.14, 514-524, 1986.
- [5] Plante F. "Détection acoustique des pathologies phonatoires chez l'enfant". Thèse INP Grenoble, spécialité Signal Image Parole, Mai 1993.
- [6] Liebermann P., "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges". *J. Acoust. Soc. Am.* Vol.35, 344-353, 1963.
- [7] Yanaghiara N. "Significance of harmonic changes and noise components in hoarseness". *J. Speech. Hear. Res.* Vol.10, 531-541, 1967.
- [8] Plante F., Berger-Vachon C., Kauffmann I., Collet L., Borel J. "Perturbations fréquentielles du fondamental laryngé chez des enfants porteurs de Kissing-nodules". *ITBM*. Vol.14, 239-252, 1993.
- [9] Gray A.H., Markel J.D "Distance measures for speech processing". *IEEE Trans. Acoust. Speech Signal Process.* Vol.24, 367-376, 1976.
- [10] Makhoul J. "Linear prediction : a tutorial review". *Proc. IEEE* Vol.63, 561-580, 1975.
- [11] Tohkura Y. "A weighted cepstral distance measure for speech recognition". *IEEE Trans. Acoust. Speech Signal Process.* Vol.35, 1414-1422, 1987.
- [12] Feng G. "Modélisation acoustique et traitement de la parole. Le cas des voyelles nasales". Thèse INP Grenoble. 1986.
- [13] Pinto N.B., Titze I.R. "Unification of perturbation measures in speech signals". *J. Acoust. Soc. Am.* Vol.87, 1278-1289, 1990.
- [14] Childers D.G., Lee C.K. "Vocal quality factors : analysis, synthesis and perception". *J. Acoust. Soc. Am.* Vol.90, 2394-2410, 1991.
- [15] Kasuya H., Ogawa S., Mashima K., Ebihara S. "Normalized noise energy as an acoustic measure to evaluate pathologic voice". *J. Acoust. Soc. Am.* Vol.80, 1329-1334, 1986.
- [16] Kitajima K. "Quantitative evaluation of the noise level in the pathologic voice". *Folia Phoniatrica*. Vol.33, 115-124, 1981.