



**HAL**  
open science

## Modélisation statistique et informations pertinentes pour la caractérisation des voix pathologiques (dysphonies)

Gilles Pouchoulin, Corinne Fredouille, Jean-François Bonastre, Alain Ghio, Marion Azzarello, Antoine Giovanni

### ► To cite this version:

Gilles Pouchoulin, Corinne Fredouille, Jean-François Bonastre, Alain Ghio, Marion Azzarello, et al.. Modélisation statistique et informations pertinentes pour la caractérisation des voix pathologiques (dysphonies). Journées d'Etude sur la Parole (JEP), Jun 2006, Dinard, France. pp.93-96. hal-00136742

**HAL Id: hal-00136742**

**<https://hal.science/hal-00136742>**

Submitted on 15 Mar 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Modélisation Statistique et Informations Pertinentes pour la Caractérisation des Voix Pathologiques (Dysphonies)

G. Pouchoulin<sup>1</sup>, C. Fredouille<sup>1</sup>, J.-F. Bonastre<sup>1</sup>, A. Ghio<sup>2</sup>, M. Azzarello<sup>3</sup>, A. Giovanni<sup>3</sup>

<sup>1</sup>LIA, Avignon (France)

<sup>2</sup>LPL-CNRS, Aix en Provence (France)

<sup>3</sup>LAPEC, Marseille (France)

### ABSTRACT

This paper investigates the class of information relevant for the task of automatic classification of pathological voices. By using a GMM-based classification system (derived from the Automatic Speaker Recognition domain), the focus was made on three main classes of information : energetic, voiced, and phonetic information. Experiments made on a pathological corpus (dysphonia) have shown that phonetic information is particularly interesting in this context since it permits to refine the selection of the relevant information by looking at phonem- or phonem class-level (e.g. nasal vowels).

### 1. INTRODUCTION

Dans le domaine de la phoniatry, l'évaluation de la qualité vocale est un sujet sensible, au centre de nombreuses études dans des domaines multi-disciplinaires [11][7]. Dans le cas de dysphonies (altération du son laryngé), sur lesquelles se concentre cet article, le dysfonctionnement vocal peut être évalué suivant deux approches, que sont l'analyse perceptive et l'analyse objective.

L'analyse perceptive ou auditive est la plus utilisée en pratique clinique. Elle consiste à caractériser la qualité vocale par une simple écoute attentive ; de par sa subjectivité, le recours à des jury d'experts peut être nécessaire afin d'augmenter la fiabilité de l'analyse. Les inconvénients majeurs de cette approche sont le manque de fiabilité dû à différents facteurs de variabilité, qu'elle soit intra ou inter-individuelle, ainsi que la difficulté de mise en place dans le cas d'un jury d'écoute (réunions périodiques de plusieurs experts, durée des séances d'écoute, ...).

L'analyse objective ou instrumentale (comme le système EVA<sup>TM</sup> [10], Evaluation Vocale Assistée - SQLab) s'appuie sur l'acquisition de nombreuses données quantitatives (mesures acoustiques, aérodynamiques et physiologiques) par le biais d'appareils de mesure. Elle offre une approche complémentaire à l'examen laryngoscopique et à l'interrogatoire du patient effectués par les praticiens. Dans [12], 86% de concordance entre l'analyse perceptive et objective sont atteints en utilisant 9 paramètres acoustiques et aérodynamiques (F0, intensité, jitter, coefficient de Lyapunov, rapport signal sur bruit, débit d'air buccal, pression sous-glottique estimée, étendue vocale, temps maximum de phonation). L'approche instrumentale offre donc des résultats très acceptables mais encore insuffisants pour son usage dans une pratique clinique quotidienne. Par ailleurs, l'acquisition des mesures sur le /a/ tenu (généralement utilisée dans ce cas) reste controversée dans la littérature [8] car elle tend à sous-estimer la dys-

phonie. Des mesures effectuées sur de la parole continue permettraient de prendre en compte par exemple, les phénomènes vocaux de l'attaque, reconnus comme pertinents dans l'évaluation des dysphonies.

En résumé, à ce jour, aucune approche d'évaluation de la qualité de la voix ne semble répondre correctement à l'attente des cliniciens, même si l'analyse perceptive reste incontournable dans la pratique quotidienne. En outre, cette dernière reste la seule référence à laquelle sont confrontées les méthodes objectives. Dans la continuité d'une étude préliminaire [4] menée par le LIA en vue d'adapter des techniques de reconnaissance automatique du locuteur (RAL) à l'évaluation de voix pathologiques (dysphonies), notre objectif est de proposer une méthode instrumentale mieux adaptée au suivi de la pathologie des patients : facilité et rapidité d'utilisation, non contraignante pour le patient et accessible pour les cliniciens. Comparée à des méthodes instrumentales classiques, les originalités de cette approche basée sur une modélisation statistique, reposent sur :

- sa capacité à analyser de la parole continue (et non des voyelles tenues) proche de l'élocution naturelle ;
- sa capacité à traiter de grands corpus, permettant de mener des études à grandes échelles et d'obtenir des informations statistiques significatives ;
- une analyse acoustique, simple et automatique, permettant une simplicité d'instrumentation et un faible coût humain.

Le système conçu pour cette tâche particulière s'appuie sur l'approche à base de GMM, état de l'art pour la RAL. Il est issu des outils de RAL, disponibles en « version libre » (LIA\_SpkDet et ALIZE) et développés au LIA.

Dans la continuité de ce travail, nous proposons ici une première étude sur l'extraction des informations utiles à la caractérisation des voix pathologiques. Nous nous intéresserons plus particulièrement à trois sources d'informations : (1) signal complet nettoyé des parties silencieuses, (2) segments phonétiques, (3) segments voisés uniquement.

Le corpus dysphonique utilisé dans cette étude ainsi que le système de classification automatique sont décrits en sections 2 et 3. La sélection des différentes classes d'information est détaillée en section 4, puis évaluée dans un contexte expérimental en section 5. Finalement, la section 6 fournit une conclusion à ce travail.

### 2. EVALUATION DES DYSPHONIES

Dans cette étude, nous nous intéressons à des dysphonies d'origine multiple (nodules, polypes, oedèmes, kystes...) classées perceptivement selon le paramètre G de l'échelle

d'évaluation GRBAS de Hirano [5]. Le corpus mis à disposition par le LAPEC (Hôpital de la Timone Marseille) est constitué de 80 échantillons de voix de femmes correspondant à 20 sujets témoins et 60 patientes dysphoniques, âgées de 17 à 50 ans (moyenne de 32.2 ans). Chaque sujet a été enregistré sur la lecture d'un paragraphe de « La chèvre de Monsieur Seguin » d'Alphonse Daudet. Les enregistrements ont été évalués selon le grade global G de dysphonie de l'échelle GRBAS par un jury d'experts ; les décisions ont été prises par consensus afin d'en limiter la variabilité inter-auditeur et en une seule séance afin d'en limiter la variabilité intra-auditeur. L'ensemble du corpus étiqueté se présente donc de la manière suivante : 20 voix normales G0, 20 voix présentant une dysphonie légère G1, 20 voix une dysphonie moyenne G2 et 20 voix une dysphonie sévère G3.

### 3. SYSTÈME DE CLASSIFICATION DES VOIX PATHOLOGIQUES

Le principe retenu consiste en l'adaptation d'un système état de l'art de RAL à la classification de voix pathologiques suivant leur degré de dysphonie. Trois phases sont nécessaires et sont décrites dans les sections suivantes.

#### 3.1. Extraction de l'information acoustique

L'extraction de l'information acoustique est issue des méthodes courantes employées en RAL : pour chaque trame de signal analysé (fenêtre de Hamming de 20ms avec un pas de 10ms), sont extraits 16 coefficients cepstraux (MFCC) obtenus à partir de 24 coefficients de banc de filtres répartis sur une échelle de MEL. Les dérivées premières des coefficients MFCC ( $\Delta$ ) sont ajoutées aux vecteurs de paramètres qui sont finalement normalisés pour obtenir une distribution de moyenne 0 et variance 1 (les moyennes et les variances sont estimées sur les portions jugées utiles du signal dont la sélection est décrite en section 4).

#### 3.2. Modélisation

En RAL, l'état de l'art repose sur une modélisation statistique (GMM : Gaussian Mixture Model)[1]. Un GMM  $X$  est une somme pondérée de  $M$  distributions gaussiennes multidimensionnelles, chacune caractérisée par un vecteur moyen  $\bar{x}$  de dimension  $d$ , une matrice de covariance  $\Sigma$  de dimension  $d \times d$  et un poids  $p$  de la gaussienne au sein de la mixture. Durant la phase d'apprentissage, les paramètres  $(\bar{x}, \Sigma, p)$  sont estimés par l'algorithme EM/ML<sup>1</sup>. Classiquement, deux phases d'apprentissage sont nécessaires en RAL pour pallier le manque de données d'apprentissage disponibles pour chaque locuteur [1] :

- apprentissage d'un modèle générique (aussi appelé « modèle du monde ») estimé par l'algorithme EM/ML sur une grande quantité de données (population de locuteurs) ;
- apprentissage du modèle locuteur dérivé du modèle du monde par application des techniques d'adaptation (MAP, Maximum a Posteriori) [9].

Dans le contexte pathologique, un modèle ne correspond plus à un locuteur donné mais à un niveau de sévérité de dysphonie. Il sera appelé **modèle de grade**  $G_g$  avec  $g \in \{0, 1, 2, 3\}$ . Le modèle de grade est appris en utilisant

l'ensemble des locuteurs de même grade. On s'assurera que les voix utilisées pour l'apprentissage des modèles de grade, sont exclues des jeux de tests afin de différencier la détection de la pathologie de la reconnaissance du locuteur (mise en oeuvre de la technique leave\_x\_out).

Les modèles GMM représentant les grades pathologiques sont établis comme suit :

- un modèle GMM générique est d'abord estimé par l'algorithme EM/ML sur un corpus français composé de 76 enregistrements de 2 mn chacun de voix de femmes exclusivement ; Cette population est extraite du corpus BREF [6] entièrement disjoint du corpus dysphonique
- les modèles de grade sont ensuite dérivés du modèle GMM générique suivant la technique d'adaptation MAP [9]. Seules les moyennes sont adaptées.

Tous les modèles GMM se composent de 128 composantes gaussiennes à matrices de covariance diagonales.

#### 3.3. Classification

Lors de la phase de test, une mesure de similarité entre des vecteurs acoustiques  $y_t$  issus d'un signal et un modèle  $X$  est calculée suivant :  $L(y_t|X) = \sum_{i=1}^M p_i L_i(y_t)$  où  $L_i(y_t)$  est la vraisemblance du signal  $y_t$  par rapport à la gaussienne  $i$ ,  $M$  le nombre de gaussiennes et  $p_i$  le poids de la gaussienne.

La **décision** correspondra au grade  $g$  du modèle  $G_g$  sur lequel la plus grande vraisemblance est obtenue. Cette définition de la décision est proche de celle de la tâche d'identification automatique du locuteur. On dira que le système a classé la voix du locuteur  $Y$  dans le grade  $g$ .

### 4. SÉLECTION DES SEGMENTS PERTINENTS

Dans ce papier, nous nous intéressons aux informations pertinentes pour la caractérisation des dysphonies. Trois niveaux d'extraction de l'information utile sont étudiés :

- « segments énergétiques » : le signal de parole est nettoyé des trames de silence (système de détection « parole/non parole » du LIA basé sur une modélisation statistique de l'énergie) ;
- « segments phonétiques » : extraite d'un alignement phonétique automatique contraint par le texte (système d'alignement du LIA basé sur un décodage Viterbi, à partir d'un lexique de mots et leurs variantes phonologiques - 38 phonèmes du français) ;
- « segments voisés » : extraction des sons de parole voisés par analyse de la fréquence fondamentale F0 (logiciel PRAAT [2] sur l'intervalle de fréquence [75,600] Hz).

Selon le niveau d'analyse choisi, est opérée une sélection de trames issues du signal qui sera utilisée lors de la normalisation des paramètres acoustiques, l'apprentissage des modèles et la décision.

### 5. EXPÉRIENCES

Les expériences ont été réalisées après « adaptation » du système de RAL du LIA. Ce système (appelé LIA\_SpkDet) repose entièrement sur la plateforme libre ALIZE [3] conçue et réalisée dans le cadre du programme Technolangue. LIA\_SpkDet est également distribué en logiciel libre.

<sup>1</sup>Expectation-Maximization/Maximum Likelihood

	Grade 0	Grade 1	Grade 2	Grade 3	Total
<b>Information</b>	% succès (nb/20)	% succès (nb/20)	% succès (nb/20)	% succès (nb/20)	% succès (nb/80)
Energétique	95,0 (19)	70,0 (14)	50,0 (10)	60,0 (12)	68,75 (55)
Voisement	95,0 (19)	65,0 (13)	50,0 (10)	75,0 (15)	71,25 (57)
Phonétique	95,0 (19)	60,0 (12)	55,0 (11)	75,0 (15)	71,25 (57)

**TAB. 1:** Résultats de la classification 4-Grades suivant différentes classes d'informations extraites

Classification	Grade 0	Grade 1	Grade 2	Grade 3
Locuteurs de Grade 0	19	1	0	0
Locuteurs de Grade 1	2	12	4	2
Locuteurs de Grade 2	2	5	11	2
Locuteurs de Grade 3	0	1	4	15

**TAB. 2:** Matrice de confusion - Phonétique

### 5.1. Protocole expérimental

Il s'agit de classer une voix suivant les 4 niveaux du grade global de l'échelle GRBAS. Par conséquent, 4 modèles de grade  $G_g$  sont à estimer avec  $g \in \{0, 1, 2, 3\}$ .

Lors de la phase de test, la mise en oeuvre de la technique leave\_x\_out (en vue de séparer les données de test et d'apprentissage) permet de comparer chaque voix  $y_t$  de grade  $g$  avec :

- 1 modèle  $G_g$  appris à partir des 19 voix restantes de grade  $g$  ( $y_t \notin$  aux 19 voix) ;
- 3 x 20 modèles  $G_{\bar{g}}$  appris chacun à partir de 19 voix de grade  $\bar{g}$  avec  $\bar{g} \in \{0, 1, 2, 3\} - \{g\}$ .

A l'issue de ces comparaisons, les moyennes des vraisemblances des tests sur chaque grade (1 modèle  $G_g$  et 3 x 20 modèles  $G_{\bar{g}}$ ) sont calculées et comparées pour fournir une unique décision pour la voix ( $y_t$ ) de grade  $g$ .

*Note : Le même nombre de voix (19) est utilisé pour l'ensemble des modèles de grade.*

### 5.2. Résultats

#### Classes d'informations utiles

Le tableau 1 donne les résultats des différentes classes d'informations sélectionnées en vue de la caractérisation des voix pathologiques. Les expériences relatives aux segments « voisés » et « phonétiques » obtiennent le meilleur résultat (71,25% de réussite). Quelle que soit la classe d'information, on peut remarquer que le grade 0 est le mieux reconnu (95,0%) et que la confusion provient principalement des grades 1 et 2 (voir matrices de confusion Tab. 2 et 3). Concernant les segments « énergétiques », à durée équivalente avec les segments « phonétiques », la classification en grade 3 est dégradée, pouvant montrer une faiblesse du détecteur parole/non parole sur des dysphonies très sévères. Les informations « voisées » obtiennent un taux de réussite équivalent aux informations « phonétiques » alors que leurs durées sont plus courtes de 10 à 20% suivant les grades. Cette observation tend à démontrer que les informations non voisées présentes dans la classe des informations « phonétiques » sont moins pertinentes, ce qui semble logique dans le contexte d'une analyse des dysphonies qui concernent une altération du voisement.

*Note : Tous les résultats fournis dans ce papier sont issus du classifieur GMM et doivent être interprétés d'un point de vue statistique.*

Classification	Grade 0	Grade 1	Grade 2	Grade 3
Locuteurs de Grade 0	19	1	0	0
Locuteurs de Grade 1	3	13	2	2
Locuteurs de Grade 2	3	5	10	2
Locuteurs de Grade 3	0	1	4	15

**TAB. 3:** Matrice de confusion - Voisement

### Analyse phonétique

Dans [7], une étude descriptive et perceptive des caractéristiques pathologiques de chaque phonème constitutif d'un échantillon de parole chez des patients dysphoniques est proposée sous le nom de "phonetic labelling". Nous proposons ici de suivre une démarche similaire en observant le comportement du système de classification automatique suivant différentes classes de phonèmes.

A partir des résultats obtenus sur les segments « phonétiques », une analyse a été réalisée sur la pertinence des différentes classes de phonèmes dans la classification des voix pathologiques. Il est à noter que cette analyse porte uniquement sur le pouvoir décisionnel de ces différentes classes (cette catégorisation n'est pas utilisée lors des phases de normalisation des paramètres ni d'apprentissage pour lesquelles l'ensemble des informations « phonétiques » a été utilisé). Le tableau 4 présente une première analyse des décisions suivant une catégorisation détaillée « Consonnes/Voyelles ». Il est intéressant de remarquer :

- la pertinence des différentes classes pour le grade 0 (85% en moyenne de bonne classification) ;
- des différences marquées entre les classes pour le grade 3 (de 75 à 85% pour les voyelles orales, les semi-consonnes et les occlusives contre 45 à 60% pour les voyelles nasales, les consonnes nasales, liquides et fricatives) ;
- pour les grades 1 et 2, les consonnes occlusives et les voyelles nasales apportent peu d'information (de 35% à 45% uniquement de bonne classification). En revanche, les semi-consonnes, liquides et fricatives réagissent plus favorablement au grade 2 (de 60% à 65% de bonne classification) ; les voyelles orales et consonnes nasales plus favorablement au grade 1 (60%) ;
- les semi-consonnes, malgré leurs courtes durées (de 0.39s à 0.44s), obtiennent un taux global de réussite de 67.5% contre 71.25% sur la totalité des informations « phonétiques ».

Un premier approfondissement de cette étude montre qu'au sein d'une même classe de phonèmes, les comportements peuvent être très différents suivant les grades. Par exemple, le tableau 5 présente les résultats de classification des voyelles nasales.

*Note : Dans ce cas, la décision est prise au niveau du phonème uniquement.*

On peut observer, par exemple, pour le grade 3 des taux de réussite variant de 25% pour le phonème [ɛ̃] à 70% pour le phonème [œ̃]. Ce même comportement a été observé sur d'autres classes de phonèmes.

Par ailleurs, des travaux en cours devraient permettre d'établir des corrélations entre ces résultats et ceux obtenus par la méthode de « phonetic labelling » (analyse perceptive).

## 6. DISCUSSION

Dans ce papier, nous proposons une analyse de l'information pertinente, véhiculée par le signal de parole, pour la caractérisation des voix pathologiques. Trois niveaux

	Grade 0	Grade 1	Grade 2	Grade 3	Total
<b>Classe phonét.</b>	<b>% succès (nb/20)</b> dur. moy.	<b>% succès (nb/20)</b> dur. moy.	<b>% succès (nb/20)</b> dur. moy.	<b>% succès (nb/20)</b> dur. moy.	<b>% succès (nb/80)</b>
<b>Voyelle</b>	<b>95,0 (19)</b>	<b>60,0 (12)</b>	<b>35,0 (7)</b>	<b>70,0 (14)</b>	<b>65,00 (52)</b>
Voyelle orale	95,0 (19) 4,14 (s)	60,0 (12) 4,09 (s)	45,0 (9) 4,15 (s)	75,0 (15) 4,12 (s)	68,75 (55)
Voyelle nasale	95,0 (19) 0,89 (s)	40,0 (8) 0,85 (s)	35,0 (7) 0,85 (s)	45,0 (9) 0,66 (s)	53,75 (43)
Consonne	90,0 (18)	50,0 (10)	60,0 (12)	65,0 (13)	66,25 (53)
Semi-consonne	90,0 (18) 0,39 (s)	35,0 (7) 0,43 (s)	60,0 (12) 0,43 (s)	85,0 (17) 0,44 (s)	67,50 (54)
Consonne liquide	80,0 (16) 1,68 (s)	45,0 (9) 1,66 (s)	60,0 (12) 1,72 (s)	60,0 (12) 1,86 (s)	61,25 (49)
Consonne nasale	75,0 (15) 1,40 (s)	60,0 (12) 1,41 (s)	50,0 (10) 1,57 (s)	50,0 (10) 1,46 (s)	58,75 (47)
Consonne fricative	90,0 (18) 1,53 (s)	40,0 (8) 1,56 (s)	65,0 (13) 1,64 (s)	45,0 (9) 1,77 (s)	60,00 (48)
Consonne occlusive	85,0 (17) 2,01 (s)	45,0 (9) 2,06 (s)	45,0 (9) 2,18 (s)	85,0 (17) 2,25 (s)	65,00 (52)

**TAB. 4:** Analyse par classe phonétique : % de réussite et durée moyenne par classe et par grade

Phonèmes	Grade 0	Grade 1	Grade 2	Grade 3
[ā]	<b>68,3 (41/60)</b>	<b>28,3 (17/60)</b>	<b>41,7 (25/60)</b>	<b>36,7 (22/60)</b>
[ē]	<b>47,5 (19/40)</b>	<b>35,0 (14/40)</b>	<b>35,0 (14/40)</b>	<b>25,0 (10/40)</b>
[ō]	<b>85,0 (51/60)</b>	<b>28,3 (17/60)</b>	<b>28,3 (17/60)</b>	<b>51,7 (31/60)</b>
[ā̃]	<b>75,0 (15/20)</b>	<b>30,0 (6/20)</b>	<b>25,0 (5/20)</b>	<b>70,0 (14/20)</b>

**TAB. 5:** Résultats de la classification 4-Grades suivant les voyelles nasales (décision au niveau du phonème)

d'informations ont été testés : segments « énergétiques », « phonétiques » et « voisés ». Cette étude a été menée en utilisant un système de classification de voix pathologiques dérivé du domaine de la RAL et basé sur une approche statistique (GMM).

D'un point de vue expérimental, les informations « phonétiques » semblent les plus intéressantes dans cette étude, au sens où elles permettent d'affiner la sélection de l'information utile. En effet, l'étude des différentes classes phonétiques a montré l'influence de certains phonèmes ou classes de phonèmes dans la tâche de classification des voix pathologiques (68.75% de réussite pour les voyelles orales contre 53.75% pour les voyelles nasales). Néanmoins, nous avons montré également que des comportements différents peuvent intervenir au sein d'une même classe. Par ailleurs, il est intéressant de constater que certaines classes de phonèmes sont plus discriminantes que d'autres pour les grades 1 et 2, qui restent problématiques dans le cadre d'une décision globale (totalité de l'information présente). Il est à noter cependant que la petite taille du corpus (80 voix dont 60 dysphoniques) est à prendre en compte dans la validité de ces résultats ainsi que les caractéristiques intrinsèques du système de classification utilisé.

Il serait à présent intéressant de définir un paradigme de décision basé sur les informations phonétiques, permettant d'améliorer les performances du système automatique. La définition d'un arbre de décision phonétique constitue une voie intéressante pour améliorer la fiabilité de la classification.

## 7. CONCLUSION

Si d'autres types d'analyses instrumentales permettent de meilleurs résultats pour le moment [12], le taux de classification correcte de 70% avec 80 locuteurs et 4 classes est encourageant. En effet, les performances du système peuvent être améliorées dans plusieurs directions : (1) l'augmentation du corpus d'apprentissage (élément très important dans les systèmes de RAL), (2) une extraction d'information acoustique mieux adaptée à l'analyse des dysphonies, (3) une adaptation sélective des segments pertinents. Enfin, nous avons conscience que l'intérêt majeur de ce type d'outil de classification automatique est un certain déterminisme qui fait actuellement défaut à l'analyse perceptive. Cet outil restera un instrument d'évaluation et non un outil de décision qui reste clairement entre les mains du clinicien.

## REMERCIEMENTS

Les auteurs tiennent à remercier le Laboratoire Audio-Phonologie Expérimentale et Clinique (LAPEC - Hôpital La Timone-Marseille) d'avoir mis à leur disposition le corpus dysphonique utilisé dans cette étude.

## RÉFÉRENCES

- [1] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska, and D. A. Reynolds. A tutorial on text-independent speaker verification. In *EURASIP Journal on Applied Signal Processing*, volume 39, pages 430–451, 2004.
- [2] P. Boersma and D. Weenink. Praat : doing phonetics by computer. <http://www.praat.org/>.
- [3] J.-F. Bonastre, F. Wils, and S. Meignier. Alize, a free toolkit for speaker recognition. In *ICASSP-05, Philadelphia, USA*, volume 39, pages 430–451, 2005.
- [4] C. Fredouille, G. Pouchoulin, J.-F. Bonastre, M. Azzarello, A. Giovanni, and A. Ghio. Application of automatic speaker recognition techniques to pathological voice assessment (dysphonia). In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech 05)*, 2005.
- [5] M. Hirano. Psycho-acoustic evaluation of voice : Grbas scale for evaluating the hoarse voice. In *Clinical Examination of voice*, Springer Verlag, 1981.
- [6] L. Lamel, J. Gauvain, and L. Eskénazi. Bref, a large vocabulary spoken corpus for french. In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech 99)*, 1991.
- [7] J. Revis. L'analyse perceptive des dysphonies : approche phonétique de l'évaluation vocale. In *Phd thesis, Université de la Méditerranée*, 2004.
- [8] J. Revis, A. Giovanni, FL. Wuyts, and J.M. Triglia. Comparaison of different voice samples for perceptual analysis. In *Folia Phoniatr Logop.*, pages 108–116, 1999.
- [9] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models, digital signal processing (dsp). In *a review journal - Special issue on NIST 1999 speaker recognition workshop 10 (1-3)*, pages 19–41, 2000.
- [10] B. Teston and B. Galindo. A diagnosis of rehabilitation aid workstation for speech and voice pathologies. In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech 95)*, pages 1883–1886, 1995.
- [11] F. L. Wuyts, M. S. De Bodt, G. Molenberghs, M. Remacle, L. Heylen, B. Millet, K. Van Lierde, J. Raes, and P. H. Van de Heyning. The dysphonia severity index : an objective measure of vocal quality based on a multiparameter approach. In *Journal of Speech, Language, and Hearing Research* 43, pages 796–809, 2000.
- [12] P. Yu, M. Ouakine, J. Revis, and A. Giovanni. Objective voice analysis for dysphonic patients : a multiparametric protocol including acoustic and aerodynamic measurements. In *Journal Voice* 15, pages 529–542, 2001.