Analyse Phonétique dans le Domaine Fréquentiel pour la Classification des Voix Dysphoniques

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

▶ **To cite this version:**

Gilles Pouchoulin, Corinne Fredouille, Jean-François Bonastre, Alain Ghio, Antoine Giovanni. Analyse Phonétique dans le Domaine Fréquentiel pour la Classification des Voix Dysphoniques. Journées d'Etude sur la Parole (JEP), Jun 2008, Avignon, France. pp.221-224. hal-00292400

HAL Id: hal-00292400
https://hal.science/hal-00292400v1

Submitted on 1 Jul 2008

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.

Analyse Phonétique dans le Domaine Fréquentiel pour la Classification des Voix Dysphoniques

G. Pouchoulin[1], C. Fredouille[1], J.-F. Bonastre[1], A. Ghio[2], A. Giovanni[2]

[1]Université d'Avignon, Laboratoire Informatique d'Avignon (EA331), F-84018 Avignon (France)
[2]CNRS-LPL, Aix en Provence (France)
{gilles.pouchoulin,corinne.fredouille,jean-francois.bonastre}@univ-avignon.fr, alain.ghio@lpl-aix.fr

ABSTRACT

Concerned with pathological voice assessment, this paper aims at characterizing dysphonia in the frequency domain for a better understanding of related phenomena while most of the studies have focused only on improving classification systems for diagnosis help purposes. Based on a first study which demonstrates that the low frequencies ([0-3000]Hz) are more relevant for dysphonia discrimination compared with higher frequencies, the authors propose in this paper to pursue by analyzing the impact of the restricted frequency subband ([0-3000]Hz) on the dysphonic voice discrimination from a phonetical point of view. In this sense, performance of the GMM-based automatic dysphonic voice classification system is measured according to different phoneme classes and frequency bands ([0-3000] and [0-8000]Hz).
Keywords : speech analysis, voice pathology, dysphonia, speech disorder, frequency domain analysis

1. INTRODUCTION

L'évaluation qualitative de la voix dysphonique est un sujet sensible, au centre de nombreuses études dans des domaines multi-disciplinaires. Deux principales approches peuvent être considérées. La première méthodologie, l'évaluation perceptive [3, 9] consiste à qualifier et à mesurer le dysfonctionnement vocal par une simple écoute attentive de la voix du patient. Cependant même si elle reste la plus utilisée par les cliniciens, elle est largement discutée dans la littérature à cause de sa subjectivité intrinsèque, le manque d'une échelle universelle, sa grande variabilité intra et inter-individuelle dans les jugements et, finalement, son coût non négligeable en temps et en ressources humaines lorsqu'un jury expert est constitué afin d'en réduire son manque de fiabilité. La deuxième méthodologie est l'analyse objective qui a été présentée comme une alternative à l'évaluation perceptive afin de faire face à ses inconvénients. Dans ce contexte, des mesures acoustiques, aéro-dynamiques et/ou physiologiques sont associées au système de classification automatique pour fournir une décision. La plupart des études sont proposées dans la littérature dans le but d'améliorer la performance de ces systèmes qui, d'un point de vue clinique, ne sont pas suffisamment efficaces [5, 7].
Comme la dysphonie concerne essentiellement la source vocale, la plupart des études se sont concentrées sur des paramètres directement liés à ce vibrateur (stabilité FO, jitter, shimmer, HNR, ..., [10, 12]). D'autres études ont porté sur le timbre global de la voix, en supposant que les caractéristiques acoustiques de la dysphonie sont distribuées uniformément sur l'ensemble du spectre. Finalement, l'information issue d'une analyse spectrale à long terme a aussi été étudiée, aboutissant à différentes classifications de voix pathologiques [11, 4].
Ce papier poursuit le travail rapporté dans [8], dans lequel une étude a porté sur les caractéristiques de la dysphonie dans le domaine fréquentiel à travers une analyse par sous-bandes de fréquences. Dans ce contexte, les auteurs ont montré que la plage de fréquences [0-3000]Hz semble plus pertinente en terme de discrimination des voix dysphoniques que les bandes de plus hautes fréquences, voire de la bande totale [0-8000]Hz. Dans ce papier, nous proposons d'étudier les manifestations de la dysphonie en fonction de la nature des segments phonétiques, au moyen d'une analyse en sous-bande de fréquences [0-3000]Hz. Dans ce sens, les décisions fournies par la classification automatique des voix dysphoniques seront analysées selon différentes classes de phonèmes.

2. LE CORPUS DE VOIX DYSPHONIQUES

Le corpus utilisé dans cette étude est composé de parole lue, prononcée par des sujets dysphoniques (affectés par nodules, polypes, oedèmes, kystes,...) et un groupe témoin. Les voix sont classées suivant le paramètre G de l'échelle GRBAS de Hirano [6] classiquement utilisée dans le cadre d'évaluations perceptives où une voix normale correspond au grade 0, une dysphonie légère à 1, une dysphonie modérée à 2 et une dysphonie sévère à 3.
Le corpus, fourni par le service ORL du CHU Timone à Marseille, est composé de 80 voix de femmes, âgées entre 17 et 50 ans, ayant reçu la consigne de lire un texte standardisé, la durée de lecture variant de 13.5 à 77.7 secondes. Les 80 voix sont réparties équitablement entre les 4 grades (20 voix dans chacun). Ces grades perceptifs ont été attribués par consensus, par un jury composé de 3 auditeurs experts, au cours d'une même session.
Ce corpus est utilisé pour l'ensemble des expériences présentées dans ce papier.

3. LE SYSTÈME DE CLASSIFICATION

Le système est dérivé d'un système classique de Reconnaissance Automatique du Locuteur (RAL) adapté à la classification des voix dysphoniques. Il est basé sur une modélisation statistique reposant sur un mélange de gaussiennes (GMM), « état de l'art » en RAL. Il repose sur la boite à outils (LIA_SpkDet et ALIZE [2]) développée au laboratoire LIA disponible en « open source ». Trois phases sont nécessaires (pour plus de détails voir en [8]).
Paramétrisation : le signal de parole pré-accentué (0.95) est décomposé en trames de 20ms, extraites toutes les 10ms, sur lesquelles une fenêtre de Hamming est appliquée. Pour chaque trame, une transformée de Fourier est calculée, suivie d'une analyse en banc de filtres (24 filtres triangulaires dont les fréquences centrales sont réparties sur une échelle linéaire) résultant en un vecteur de 24 coefficients LFSC (Linear Frequency Spectrum Coefficient). Les vecteurs de paramètres sont ensuite normalisés pour obtenir une distribution de moyenne-0 et de variance-1.
Modélisation : les techniques basées sur les Modèles de

Mélange de Gaussiennes (GMM) [1] sont utilisées pour construire un modèle statistique pour chaque niveau de sévérité de dysphonie, nommé **modèle de grade** G_g avec $g \in \{0, 1, 2, 3\}$. Le modèle de grade G_g est appris sur l'ensemble des voix évaluées perceptivement dans le grade g. 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. Tous les modèles GMM sont composés de 128 composantes gaussiennes avec des matrices de covariance diagonales.

Décision : Dans le contexte pathologique, la décision correspond au grade g du modèle G_g sur lequel la plus grande mesure de ressemblance est calculée pour une voix de test donnée. Cette mesure correspond à la vraisemblance définie par : $L(y_t|G) = \sum_{i=1}^{M} p_i \, L_i(y_t)$ où $L_i(y_t)$ est la vraisemblance du signal y_t de la gaussienne i, M le nombre de gaussiennes et p_i le poids de la gaussienne i.

4. SOUS-BANDES DE FRÉQUENCES ET ANALYSE PHONÉTIQUE

Dans [8], les auteurs ont étudié comment les caractéristiques acoustiques de la dysphonie sont dispersées sur l'ensemble de l'espace fréquentiel en analysant les performances du système de classification (décrit en section 3) sur différentes sous-bandes de fréquences obtenues par filtrage des signaux de parole sur les plages suivantes : [0-3000]Hz, [3000-5400]Hz et [5400-8000]Hz. Les tests de classification effectués sur le corpus filtré de voix dysphoniques montrent que la sous-bande [0-3000]Hz tend à être la zone la plus intéressante (comparée aux autres sous-bandes ou à la bande totale), permettant une meilleure discrimination entre les différents grades de dysphonie.

Dans ce papier, les auteurs proposent de poursuivre l'étude précédente en observant le comportement du système de classification automatique des voix dysphoniques selon différentes classes de phonèmes. Ces comportements seront analysés sur les 2 bandes de fréquences, [0-3000]Hz et [0-8000]Hz, par phonème ou classe de phonèmes, afin d'évaluer les manifestations de la dysphonie selon les grades. Cette analyse phonétique est très proche de la méthode « phonetic labelling » proposée par [9] dans laquelle une étude descriptive et perceptive des caractéristiques pathologiques de différents phonèmes est présentée.

Pour effectuer les tests de classification des voix dysphoniques selon différentes classes de phonèmes, une segmentation phonétique est nécessaire pour chaque signal de parole du corpus. Cette segmentation a été extraite automatiquement par un alignement phonétique contraint sur le texte, effectué en utilisant le système d'alignement du LIA, basé sur un algorithme de décodage Viterbi, un lexique de mots avec leurs variantes phonologiques et un ensemble de 38 phonèmes du français.

Il est à noter que cette catégorisation phonétique n'est utilisée que durant la phase de décision. En effet, les phases de paramétrisation et de modélisation utiliseront l'ensemble du matériau phonémique disponible dans chaque signal de parole du corpus. Par contre, lors d'un test de classification, la décision sera prise (voir à la section 3) sur l'ensemble des segments associés à une classe phonétique.

Le tableau 1 fournit les différentes classes de phonèmes disponibles dans le corpus de voix dysphoniques avec leur durée (en secondes) pour chaque grade. Il doit être noté que les classes de phonèmes avec une durée inférieure à 20s (notées en italique) *i. e.* moins de 1 seconde par locuteur, ne sont données qu'à titre indicatif et ne seront pas prises en compte pour l'analyse phonétique, ces résultats étant jugés peu fiables par les auteurs.

Classes	**Grades**				**Effectifs Totaux**		
Phonétiques	*G0*	*G1*	*G2*	*G3*	*nb*	μ	σ
Consonne	**135.13**	**139.21**	**149.83**	**167.28**	**6395**	**0.092**	**0.045**
. Sonore	88.80	90.56	95.36	106.57	4719	0.081	0.039
. Sourde	46.33	48.65	54.47	60.71	1676	0.125	0.046
Liquide	34.56	34.01	36.04	43.03	2181	0.068	0.033
Nasale	29.72	30.17	31.85	33.42	1279	0.098	0.039
Fricative	31.77	32.32	35.07	40.70	1144	0.122	0.057
. Sonore	*10.14*	*10.32*	*10.45*	*11.76*	*436*	*0.098*	*0.056*
. Sourde	21.63	22.00	24.62	28.94	708	0.137	0.052
Occlusive	39.08	42.71	46.87	50.13	1791	0.100	0.039
. Sonore	*14.38*	*16.06*	*17.02*	*18.36*	*823*	*0.080*	*0.030*
. Sourde	24.70	26.65	29.85	31.77	968	0.117	0.038
Voyelle	**103.58**	**98.77**	**103.46**	**109.79**	**5586**	**0.074**	**0.046**
Orale	84.37	80.45	85.22	93.66	4862	0.071	0.044
Nasale	*19.21*	*18.32*	*18.24*	*16.13*	*724*	*0.099*	*0.046*
Semi-voyelle	***2.80***	***2.98***	***3.37***	***3.45***	***159***	***0.079***	***0.040***
Tous phonèmes	**241.51**	**240.96**	**256.66**	**280.52**	**12140**	**0.084**	**0.046**
. Sonore	195.18	192.31	202.19	219.81	10464	0.077	0.043
. Sourde	46.33	48.65	54.47	60.71	1676	0.125	0.046

TAB. 1: Durée en secondes par classe phonétique et par grade - Informations quantitatives sur les phonèmes d'une classe phonétique : nombre (nb) avec durée moyenne (μ) et écart-type (σ) associé

5. EXPÉRIENCES

Cette section présente les performances du système de classification automatique des voix dysphoniques selon les différentes classes de phonèmes et les bandes de fréquences : totale [0-8000]Hz et restreinte [0-3000]Hz.

Les résultats fournis dans cette section sont exprimés en terme de Taux Correct de Classification (nommé TCC dans le reste du papier) ; le nombre de voix correctement classées est aussi fourni entre parenthèses.

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

5.1. Resultats

Les tableaux 2 et 3 fournissent les performances du système de classification automatique des voix dysphoniques en terme de % TCC par classe phonétique.

Performance globale Considérant l'ensemble des phonèmes[1], le TCC total de 65% est obtenu pour la bande totale ([0-8000]Hz) contre 72.5% TCC pour la sous-bande [0-3000]Hz comme observé dans [8][2]. Il est intéressant de souligner que l'amélioration concerne principalement le grade 1 (de 55% à 65%) et le grade 2 (de 50% à 65%), grades sur lesquels la plus grande confusion de classification est généralement observée.

Analyse phonétique en [0-8000]Hz Sur la bande de fréquences [0-8000]Hz, on peut observer que :

• le grade 0 obtient 80% TCC sur la classe des consonnes (fricatives). 70% TCC est obtenu sur la classe des voyelles (en dépit du 55% des voyelles orales). Comparé avec les autres grades, le grade 0 fournit le meilleur TCC (85%) sur l'ensemble des phonèmes ;

• le grade 1 affiche des TCC semblables pour les classes des voyelles et des consonnes, avec un TCC de 55% sur l'ensemble des phonèmes. De meilleurs TCC sont obtenus par les occlusives sourdes (70%) et les voyelles

[1] nommé « tous phonèmes » dans les tableaux

[2] une différence très légère peut être observée entre ce total % TCC et ceux publiés dans [8] (71.25% pour [0-3000]Hz), due à quelques modifications apportées à la segmentation « parole/non parole »

[0-8000]Hz	Grade 0	Grade 1	Grade 2	Grade 3	Total
Classes phonétiques	**% TCC** (nb/20)	**% TCC** (nb/20)	**% TCC** (nb/20)	**% TCC** (nb/20)	**% TCC** (nb/80)
Consonne	**80.0 (16)**	**50.0 (10)**	**50.0 (10)**	**85.0 (17)**	**66.25 (53)**
. sonore	70.0 (14)	45.0 (9)	60.0 (12)	75.0 (15)	62.50 (50)
. sourde	80.0 (16)	60.0 (12)	50.0 (10)	75.0 (15)	66.25 (53)
Liquide	50.0 (10)	35.0 (7)	40.0 (8)	70.0 (14)	48.75 (39)
Nasale	50.0 (10)	35.0 (7)	50.0 (10)	45.0 (9)	45.00 (36)
Fricative	80.0 (16)	40.0 (8)	50.0 (10)	85.0 (17)	63.75 (51)
. sourde	80.0 (16)	45.0 (9)	35.0 (7)	90.0 (18)	62.50 (50)
Occlusive	65.0 (13)	65.0 (13)	50.0 (10)	70.0 (14)	62.50 (50)
. sourde	75.0 (15)	70.0 (14)	50.0 (10)	70.0 (14)	66.25 (53)
Voyelle	**70.0 (14)**	**55.0 (11)**	**40.0 (8)**	**60.0 (12)**	**56.25 (45)**
Orale	55.0 (11)	60.0 (12)	40.0 (8)	60.0 (12)	53.75 (43)
Tous phonèmes	**85.0 (17)**	**55.0 (11)**	**50.0 (10)**	**70.0 (14)**	**65.00 (52)**

TAB. 2: Résultats de classification 4-G par classe phonétique en terme de % TCC selon la bande de fréquences totale [0-8000]Hz (24LFSC)

[0-3000]Hz	Grade 0	Grade 1	Grade 2	Grade 3	Total
Classes phonétiques	**% TCC** (nb/20)	**% TCC** (nb/20)	**% TCC** (nb/20)	**% TCC** (nb/20)	**% TCC** (nb/80)
Consonne	**90.0 (18)**	**55.0 (11)**	**70.0 (14)**	**75.0 (15)**	**72.50 (58)**
. sonore	75.0 (15)	70.0 (14)	30.0 (6)	65.0 (13)	60.00 (48)
. sourde	90.0 (18)	35.0 (7)	75.0 (15)	60.0 (12)	65.00 (52)
Liquide	80.0 (16)	50.0 (10)	20.0 (4)	45.0 (9)	48.75 (39)
Nasale	30.0 (6)	35.0 (7)	55.0 (11)	35.0 (7)	38.75 (31)
Fricative	85.0 (17)	40.0 (8)	35.0 (7)	85.0 (17)	61.25 (49)
. sourde	80.0 (16)	35.0 (7)	30.0 (6)	80.0 (16)	56.25 (45)
Occlusive	90.0 (18)	60.0 (12)	50.0 (10)	70.0 (14)	67.50 (54)
. sourde	95.0 (19)	50.0 (10)	65.0 (13)	50.0 (10)	65.00 (52)
Voyelle	**85.0 (17)**	**30.0 (6)**	**70.0 (14)**	**55.0 (11)**	**60.00 (48)**
Orale	75.0 (15)	35.0 (7)	50.0 (10)	50.0 (10)	52.50 (42)
Tous phonèmes	**90.0 (18)**	**65.0 (13)**	**65.0 (13)**	**70.0 (14)**	**72.50 (58)**

TAB. 3: Résultats de classification 4-G par classe phonétique en terme de % TCC selon la sous-bande de fréquences [0-3000]Hz (24LFSC)

orales (60%). Par contre, les consonnes liquides et nasales obtiennent les plus faibles TCC (35% pour les deux) ;
• concernant le grade 2, la plupart des classes obtiennent des TCC plutôt faibles (en dessous de 50%). Seules les consonnes sonores dépassent ce seuil avec un TCC de 60%. Ces faibles TCC sont très proches du TCC de 50% obtenu sur l'ensemble des phonèmes ;
• pour le grade 3, la différence en terme de TCC entre la classe des consonnes (85%) et celle des voyelles (60%) est la plus grande en comparaison avec les autres grades. De plus, les consonnes obtiennent le meilleur TCC (90%) pour les fricatives sourdes et le plus faible (45%) avec les consonnes nasales. Un TCC intermédiaire de 70% est obtenu sur l'ensemble des phonèmes.

Concernant la colonne « Total », la classe des consonnes (notamment les consonnes sourdes) et la classe « Tous phonèmes » obtiennent des résultats très similaires (66.25% contre 65%). L'écart entre la classe vocalique (56.25%) et la classe consonantique (66.25%) est assez important. Finalement, les consonnes liquides et nasales affichent les plus faibles TCC (48.75% et 45% resp.).

Analyse comparative [0-8000]Hz vs [0-3000]Hz Comparant les performances du système de classification des voix dysphoniques entre [0-8000]Hz et [0-3000]Hz, il peut être observé que :
• pour le grade 0, les valeurs de TCC sont améliorées sur l'ensemble des classes phonétiques en [0-3000]Hz, à l'exception des fricatives sourdes qui conservent leur TCC de 80% et des consonnes nasales pour lesquelles, le TCC baisse de 50% ([0-8000]Hz) à 30% ([0-3000]Hz). Les améliorations peuvent varier de 7% à 60% (consonnes liquides) en relatif, le meilleur TCC étant obtenu pour les occlusives sourdes (95 %) ;
• pour le grade 1, les valeurs TCC sont plus faibles en [0-3000]Hz, notamment pour la classe vocalique avec seulement 30% TCC. Les TCC des fricatives sourdes, occlusives sourdes et des voyelles orales, ont baissé par rapport à la bande [0-8000]Hz. Par contre, les consonnes liquides et sonores obtiennent les meilleurs TCC (50% et 70% respectivement contre 35% et 45%), améliorant le TCC de la classe « tous phonèmes » (65% en [0-3000]Hz contre 55% en [0-8000]Hz) ;
• pour le grade 2, les classes des consonnes et des voyelles obtiennent un TCC satisfaisant de 70% en [0-3000]Hz, légèrement au-dessus du TCC de l'ensemble des phonèmes (65%) mais largement au-dessus des TCC obtenus sur la bande [0-8000]Hz (50% et 40% TCC resp.). Par contre, des valeurs très faibles de TCC sont observées pour certaines classes de consonnes (sonores 30%, liquides 20%, fricatives sourdes 30%). Néanmoins, le TCC sur l'ensemble des phonèmes est bien meilleur en [0-3000]Hz avec 65% TCC contre 50% en [0-8000]Hz ;
• pour le grade 3, la plupart des classes présentent une baisse de TCC en [0-3000]Hz, à l'exception des fricatives et occlusives qui conservent leur TCC (85% et 70% resp.). Malgré cela, 70% de TCC est atteint sur l'ensemble des phonèmes pour les deux bandes de fréquences.

L'analyse des résultats de la colonne « Total » nous amène aux mêmes observations qu'en [0-8000]Hz. La classe des consonnes atteint la même performance que celle obtenue sur l'ensemble des phonèmes en [0-3000]Hz, avec un TCC de 72.5%. Les résultats de la classe des voyelles restent plus faibles comparés à ceux de la classe des consonnes. Finalement, les consonnes liquides (48.75%) et nasales (38.75%) affichent les plus faibles TCC en [0-3000]Hz.

5.2. Discussion

Les résultats présentés ci-dessus laissent apparaître que la classe consonantique semble être la plus pertinente pour la classification des voix dysphoniques quelque soit la bande de fréquences considérée dans ce contexte expérimental.
Cette observation est-elle conflictuelle au regard des évaluations perceptives et objectives qui s'appuient sur des voyelles tenues comme le [a] ?
Le choix de ce support phonétique permet des conditions expérimentales indispensables pour évaluer la stabilité et le bruit du vibrateur en régime permanent (fluctuations à court terme telles que jitter, shimmer, HNR, ...), sachant que l'instabilité vibratoire de la glotte est une manifestation/caractéristique essentielle des dysphonies. Néanmoins, ce support phonétique reste controversé dans la littérature car il tend à sous-estimer la dysphonie. Par ailleurs, certains phénomènes vocaux issus de la parole spontanée comme l'attaque sont reconnus comme pertinents dans l'évaluation des dysphonies. Au regard de ces éléments et de la pertinence des consonnes mise en évidence dans cette étude, il semblerait par conséquent intéressant d'élargir le cadre de l'analyse phonétique menée ici, à l'étude de certains phénomènes vocaux transitoires comme le passage entre phonème « voisé $\curvearrowright$ non voisé » ou la séquence CV induisant la présence simultanée d'informations de la consonne et de la voyelle. Cette approche pourrait être, en fait, complémentaire aux méthodes d'évaluation basées sur les voyelles tenues.
Deuxièmement, la plage [0-3000]Hz a tendance à améliorer les TCC des grades 0 et 2 pour la classe des voyelles et celle des consonnes alors qu'elle pénalise les grades 1

et 3 lorsque les classes phonétiques sont considérées individuellement. Le comportement en [0-3000]Hz du grade 1 (65% TCC sur l'ensemble des phonèmes) est particulièrement inattendu au regard des faibles TCC de la classe vocalique. Un phénomène compensatoire lors de la prise de décision sur l'ensemble des phonèmes semble ici être à l'origine de ce comportement. Il est intéressant de constater que l'observation des matrices de confusion (tableau 4) sur l'ensemble des phonèmes en [0-3000]Hz montre une tendance à la surévaluation pour le grade 1 (6 voix en grade 2) et à la sous-évaluation pour le grade 2 (6 voix en grade 1). Finalement, le comportement du grade 3 sur [0-3000]Hz peut probablement être dû au filtrage de « parole bruitée » présente dans les hautes fréquences et caractéristique de certaines voix sévèrement dysphoniques. Malgré cela, la constance du TCC de l'ensemble des phonèmes sur les deux bandes de fréquences, montre que l'information pertinente pour la discrimination du grade 3 vis à vis des autres grades est toujours présente en [0-3000]Hz.

TAB. 4: Matrices de confusion de la classification 4-G dans les plages de fréquences [0-8000]Hz et [0-3000]Hz (24LFSC)

	RG0	**RG1**	**RG2**	**RG3**
TG0	**17**	2	1	0
TG1	2	**11**	5	2
TG2	2	6	**10**	2
TG3	0	1	5	**14**

[0-8000]Hz (Bande Totale)

	RG0	**RG1**	**RG2**	**RG3**
TG0	**18**	1	1	0
TG1	1	**13**	6	0
TG2	0	6	**13**	1
TG3	0	2	4	**14**

[0-3000]Hz (Bande Restreinte)

6. CONCLUSION

L'objectif de ce papier est d'étudier la caractérisation de la voix dysphonique. Il poursuit une première étude dans laquelle les auteurs ont montré que la sous-bande [0-3000]Hz a tendance à être la zone la plus pertinente pour la discrimination automatique des voix dysphoniques selon le paramètre G de l'échelle GRBAS. L'étude proposée ici porte sur une analyse par classe de phonèmes, dans laquelle la discrimination des voix dysphoniques est mesurée sur différentes classes de phonèmes et sur les bandes de fréquences, totale [0-8000]Hz ou restreinte [0-3000]Hz. Deux principaux aspects ont été soulignés par cette analyse phonétique.

Premièrement, les consonnes paraissent être la classe la plus pertinente dans ce contexte expérimental pour la discrimination automatique des voix dysphoniques quelque soit la bande de fréquences observée. Cette observation permettrait d'étudier le comportement transitoire de séquences de phonèmes contigus afin de développer une méthode complémentaire aux méthodes d'évaluation utilisant la voyelle tenue pour analyser la voix pathologique.

Deuxièmement, les grades réagissent différemment en [0-3000]Hz. En effet, tandis que les grades 0 et 2 présentent quelques améliorations considérables en terme de Taux Correct de Classification (TCC) comparés avec la bande [0-8000]Hz, les grades 1 et 3 sont pénalisés sur la plupart des classes des phonèmes. Il serait 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.

Dans un futur travail, ces différents résultats seront comparés avec une analyse perceptive effectuée par un jury expert, sur les signaux de parole filtrés en [0-3000]Hz. Le but de cette analyse sera de comparer l'évaluation des auditeurs experts et le comportement du système automatique dans la configuration fréquentielle [0-3000]Hz.

REMERCIEMENTS

Les auteurs remercient le service ORL du CHU Timone de Marseille pour avoir placé à leur disposition le corpus utilisé pour 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. *EURASIP Journal on Applied Signal Processing*, 39 :430–451, 2004.

[2] J.-F. Bonastre, F. Wils, and S. Meignier. Alize, a free toolkit for speaker recognition. *ICASSP-05, Philadelphia, USA*, 2005.

[3] P. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. Van De Heyning, M. Remacle, and V. Woisard. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Guidelines elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)*, 258 :77–82, 2001.

[4] P. Dejonckere and D. Villarosa. Analyse spectrale moyennée de la voix. Comparaison de voix normales et de voix altérées par différentes catégories de pathologies laryngées. *Acta Oto-rhino-laryngolica Belg.*, 40 :426–435, 1986.

[5] C. Hernandez-Espinosa, M. Fernandez-Redondo, P. Gomez-Vilda, J. I. Godino-Llorente, and S. Aguilera-Navarro. Diagnosis of vocal and voice disorders by speech signal. *Neural Networks, IEEE-INNS-ENNS International Joint Conference*, 4 :253–258, 2000.

[6] M. Hirano. Psycho-acoustic evaluation of voice : Grbas scale for evaluating the hoarse voice. *Clinical Examination of voice, Springer Verlag*, 1981.

[7] Ji-Yeoun Lee, SangBae Jeong, and Minsoo Hahn. Classification of pathological and normal voice based on linear discriminant analysis. *Computer Science*, 4432 :382–390, 2007.

[8] G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, and A. Giovanni. Frequency study for the characterization of the dysphonic voices. *Interspeech'07, Antwerp, Belgium*, Aug 2007.

[9] J. Revis, A. Ghio, and A. Giovanni. Phonetic labeling of dysphonia : a new perspective in perceptual voice analysis. *7th International Conference Advances in Quantitative Laryngology, Voice and Speech Research*, Oct 2006.

[10] J. Schoentgen and F. Bucella. Acoustic analysis of dysphonic voices : descriptors and methods. In *LARYNX'97*, pages 37–46, 1997.

[11] N. Yanagihara. Significance of harmonic changes and noise components in hoarseness. *Journal Speech, Hear, Res*, 10 :531–541, 1967.

[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.