HAL
open science

Caractérisation et classification automatique des modes phonatoires en voix chantée

Léonidas Ioannidis, Jean-Luc Rouas, Myriam Desainte-Catherine

To cite this version:

Léonidas Ioannidis, Jean-Luc Rouas, Myriam Desainte-Catherine. Caractérisation et classification automatique des modes phonatoires en voix chantée. XXXèmes Journées d'études sur la parole, Jun 2014, Le Mans, France. hal-00992084

HAL Id: hal-00992084

https://hal.science/hal-00992084

Submitted on 4 Nov 2014

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.

Caractérisation et classification automatique des modes phonatoires en voix chantée

Léonidas Ioannidis Jean-Luc Rouas Myriam Desainte-Catherine
LaBRI - CNRS UMR 5800, Univ. Bordeaux, France
leonidas.ioannidis@labri.fr, jean-luc.rouas@labri.fr, myriam@labri.fr

RÉSUMÉ

Dans cette étude, nous présentons une analyse de paramètres caractérisant la voix pouvant permettre de caractériser et détecter automatiquement les modes phonatoires de la voix chantée. L'objectif est d'étudier les différences dans le contenu spectral entre les types de phonation et de valider leur capacité à les classer automatiquement avec des méthodes d'apprentissage. Cela revêt un intérêt lorsque l'on cherche à caractériser la qualité vocale du chant car le mode phonatoire en est un des principaux éléments. De nombreuses applications sont envisageables au niveau de l'indexation et de la recherche d'information dans les signaux audio. Nous présentons ici une méthode basée sur des méthodes d'apprentissage automatique en utilisant un ensemble de paramètres acoustiques et une base de données adéquate. Cette méthode nous permet d'obtenir des résultats de classification des quatre modes phonatoires principaux avec une F-mesure supérieure à 0.8.

ABSTRACT

Caractérisation and automatique classification of phonation modes in singing voice

In this study we present an analysis of voice related features that can help in a classification task to automatically detect and characterize phonation modes found in singing voice. The goal here is to examine the differences in spectral content between different phonation modes, and validate their ability to classify the phonation using machine learning techniques. This can be of interest when characterising the singing voice quality, since phonation is a essential part of the overall quality. It can have numerous applications for indexing and retrieving audio using content based search. We present here a method based on machine learning training, using an acoustic feature set on an audio library. This method is able to classify four widely used phonation modes with a F-measure above 0.8.

MOTS-CLÉS : Modes phonatoires, Voix chantée, Qualité vocale.

KEYWORDS: Phonation Modes, Singing Voice, Voice quality.

1 Introduction

La voix chantée est un des instruments de musique les plus utilisés. C'est pour cela que ses attributs sont étudiés d'une manière différente de ceux de la voix parlée (Sundberg, 1987). Bien que partageant le même appareil phonatoire, la voix parlée et la voix chantée l'utilisent de

manière différente et elles exigent différentes démarches que ce soit au niveau de l'analyse ou de la synthèse. La voix humaine est un instrument capable de produire une gamme très grande de sons. Dans le cadre de la voix chantée, cette variété est souvent exploitée comme un moyen d'expressivité, comme un violoniste peut pincer ou tirer avec l'archet afin de produire une couleur sonore différente. Cette variabilité est rendue possible par un ensemble de muscles qui pouvant se contracter ou se dilater (Sundberg, 1995).

La phonation est le sujet d'étude de cet article. Elle est définie comme étant le son produit par les vibrations des cordes vocales. Ce son peut varier selon la dynamique de vibration de la glotte et le contrôle de la musculature qui l'entoure.

La phonation est aussi perceptivement distinguée par des écouteurs entraînés (Sofranko et Prosek, 2012). Dans des études précédentes, des techniques permettant d'identifier les modes phonatoires ont été proposées, mais elles utilisent toutes des méthodes intrusives, telles que les mesures aérodynamiques, la résistance laryngée et l'électroglottographie (Grillo et Verdolini, 2008). Chacune de ces méthodes offre des performances variables. Cependant le problème est loin d'être considéré résolu. En dehors des considérations sur l'efficacité de ces méthodes, la classification de données déjà enregistrées est une nécessité qui ne peut pas être traitée par les méthodes intrusives. Des travaux ont été menés dans cette direction (Proutskova *et al.*, 2012b). Les auteurs de cette étude proposent une méthode basée sur la décomposition source-filtre, afin d'estimer la contribution de la source, et une classification dépendant de la voyelle prononcée, basée sur des descripteurs extraits du signal source. Cependant, les inconvénients d'une déconvolution inverse source-filtre appliquée à la voix chantée sont connus. Cette opération est sujette à des estimations erronées du signal source à cause de sa structure harmonique parcimonieuse, spécialement dans le cas des notes de haute fréquence (Alku, 2011). Pour pallier ces erreurs, il est possible d'obtenir une meilleure estimation de la source en utilisant des connaissances a priori sur le signal, mais la méthode devient alors inadaptée pour des sons inconnus. Les travaux précédents ont montré ces inconvénients et ont tenté de les contourner en proposant un nouveau descripteur basé sur la transformation en ondelettes pour mesurer l'amplitude sur différentes bandes de fréquences, afin de capturer la pente spectrale (Kane et Gobl, 2011).

Dans cet article, nous proposons un cadre de classification automatique des modes phonatoires avec l'extraction des descripteurs audio sans estimation préalable de la source glottique ni aucun a priori sur le signal. De par la nécessité croissante d'indexer des sons dans de grandes bases de données, ce travail est utile l'analyse sémantique de la voix chantée.

Dans la section 2, une explication détaillée des modes phonatoires est présentée. Dans la section 3, la base de données utilisée et les descripteurs acoustiques que nous utilisons sont expliqués et présentés. Dans la section 4, les expériences de classification automatique sont présentées et leurs résultats sont commentés. Finalement, des conclusions et des perspectives sont données dans la section 5.

2 Les modes de phonation en la voix chantée

Dans cette section, nous décrivons les modes phonatoires tels que définis par Sundberg. Selon lui, 4 modes phonatoires existent. D'un point de vue physiologique, ils résultent d'une différence dans les pressions sous-glottique et de la force appliquée pendant la phonation en contraction et

dilations de la glotte. Pendant la phonation soufflée, les plis vocaux restent partiellement ouverts pendant la vibration.Les muscles qui contrôlent l'adduction des plis vocaux sont détendus, ce qui permet à l'air de s'écouler à travers de glotte. Ceci donne une voix semblable au chuchotement (Sundberg, 1987). La voix modale, qui est la plus fréquente, correspond à une vibration régulière des plis vocaux tout au long de leur surface avec une feremeture complète à chaque période. Le debit d'air est notament moindre que dans le cas de la voix soufflée. Une voix serrée est produite quand une tension forte est appliquée à la musculature qui entoure le larynx. L'ensemble du larynx est alors soumis à une tension et le haut du larynx va se contracter fortement tandis que les plis ventriculaires pressent aussi sur les surfaces hautes des plis vocaux, rendant leur vibration inefficace et le debit d'air devient plus petit. (Sundberg, 1995). La phonation "Flow", aussi connue comme voix résonante, est une phonation assez fréquente dans la voix lyrique occidentale. Elle est caractérisée par une pression sous-glottique plus faible que la voix serrée et une faible force d'adduction aux plis vocaux (Titze, 2001). Elle permet d'obtenir des niveaux sonores comparables à la voix serrée mais autorise le chanteur a mieux contrôler son débit d'air. Cela contribue à en faire une voix "économique" comme la voix modale (Proutskova *et al.*, 2012a). Elle est très utilisée par les chanteurs lyriques car elle leur permet d'être entendus malgré la présence d'un ensemble orchestral important, en augmentant la puissance spectrale autour de 3KHz. Ce phénomène est aussi appelé "le formant du chanteur" (Sundberg, 1987). Des termes synonymes ou équivalents peuvent être trouvés dans la littérature. Par exemple "whisper" (chuchotée), "neutre", "harsh" au lieu de serrée. Mêmes s'ils ne sont pas toujours réellement équivalents, nous les mentionnons car ils sont relativement usités. Pour éviter les confusions, nous allons cependant pour cet article uniquement utiliser les termes que nous avons définis.

Les mécanismes de la voix sont très flexibles et nous permettent de produire une gamme des sons très variée. Lorsque l'on parle de mode phonatoires, nous devons aussi penser que ces modes sont pas toujours bien distincts et que l'on pourrait les positionner dans un espace continu, où à un extrême se trouverait la voix soufflée et à l'autre la voix serrée. Ainsi, vers le milieu de cet espace, nous trouverions la voix modale et "flow" mais également tous les cas intermédiaires. Les frontières ne sont pas strictement definies (Proutskova *et al.*, 2012a).

Pour analyser le contenu spectral de la voix chantée, il est nécessaire d'utiliser des outils adaptés aux modes de production humaine. Le modèle source-filtre (Fant, 1981) est souvent employé. Dans ce cadre, la production est modélisée en deux parties. La source, représentant la production du signal sortant de la glotte, et le filtre, qui est l'implémentation du filtrage effectué par la cavité vocale et les résonateurs. La démarche inverse est appelée filtrage-inverse. Après une estimation de l'enveloppe spectrale, en général en utilisant des coefficients de prédiction linéaire (Linear Predictive Coding -LPC), permettant d'obtenir une approximation de la cavité vocale, le signal est filtré pour obtenir une estimation de la source (Alku, 2011). Un autre modèle pouvant être utilisé pour la modélisation de la voix est le modèle sinusoïdal présenté par (McAulay et Quatieri, 1986). Dans ce modèle, les caractéristiques spectrales du signal sont approximées par des composantes sinusoïdales.

3 Base de données et Descripteurs

3.1 La base de données

Nous utilisons une base de données – distribuée librement – construite spécifiquement pour étudier les modes phonatoires et l'estimation de la source glottique. Le manque de données de référence et des données d'apprentissage en nombre suffisant ont conduit les auteurs à l'élaboration de cette base, constituée d'enregistrements de voix chantée étiquetés selon le mode phonatoire. Le fait d'utiliser cette base nous permet d'évaluer les résultats de notre méthode et permettra d'effectuer des comparaisons avec d'autres approches. Cette base est constituée de 790 enregistrements courts de 9 voyelles différentes que l'on peut trouver dans la langue russe. Ces voyelles sont toutes chantées avec les quatre modes phonatoires. Pour chaque mode, la totalité de l'étendue vocale de la chanteuse a été enregistrée lorsque cela était possible. Les auteurs notent cependant que pour les modes "Flow" et serrée, il n'était pas possible de produire des sons au delà d'une certaine fréquence car la phonation devenait alors instable. Tous les enregistrements ont été effectués par une seule chanteuse de tessiture soprano. La durée moyenne de chaque enregistrement est de 1.34 secondes, variant de 0.9 à 1.6 secondes. La base de données est fournie avec des méta-données indiquant la fréquence, la voyelle et le mode phonatoire exprimé dans chaque fichier. Les détails de la procédure d'enregistrement et des informations techniques sur la base de données peuvent être trouvés dans (Proutskova *et al.*, 2012a). Les enregistrements sont effectués avec un taux d'échantillonage de 96kHz. Avant d'effectuer nos traitements, nous avons sous-échantillonné l'ensemble des sons à 16kHz. C'est un taux d'échantillonage adéquat pour le traitement de la voix puisque 8kHz suffisent pour capturer le contenu spectral de la voix chantée. Enfin, nous n'utilisons qu'un extrait de 500ms situé au milieu de chaque fichier afin d'éliminer automatiquement les silences et les parties transitoires au début et à la fin de la phonation.

3.2 Descripteurs

Une analyse spectrale est conduite avant de choisir un ensemble final de descripteurs acoustiques permettant de modéliser les modes phonatoires. La pente spectrale est un des paramètres les plus pertinents pour différencier les quatre modes phonatoires (Kane et Gobl, 2011). Il est également connu que la différence entre les deux premières harmoniques du signal source est un moyen fiable de distinguer les modes soufflés et serrés (Alku, 2011), (Sundberg, 1987). C'est une autre manière de déterminer la rapidité de la chute de l'énergie sur les hautes fréquences et la richesse harmonique du signal. En analysant les spectrogrammes de la Fig. 2, nous pouvons observer des différences entre les modes soufflé et serré. La prédiction linéaire LPC est utilisée pouir visualiser les différences de fréquence des formants et leurs amplitudes selon les modes phonatoires. Dans ce cas, la comparaison est effectuée pour la même voyelle, puisque les valeurs des fréquences formatiques dépendent fortement de l'articulation. Les fréquences, les amplitudes et les bandes passantes des quatres premiers formants sont calculées. Dans la Fig. 1, nous observons que ces amplitudes varient pour chaque mode (même voyelle et même fréquence).

Nous pouvons observer que les formants sont sujets à des déplacements pour les modes serrée et soufflé. Cela peut être dû à l'étirement du larynx. Une pression sous-glottique plus important peut également être responsable de changements dans les fréquences des formants (Sundberg, 1987).

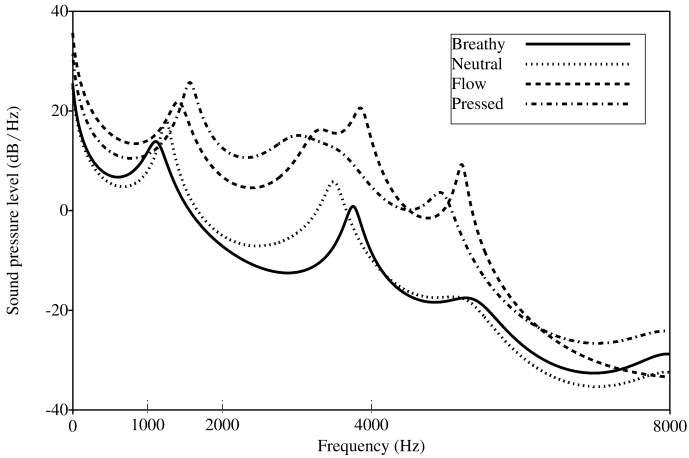

FIGURE 1 – Analyse LPC de la voyelle /a/ pour différents types de phonation.

Toutefois, en considérant que l'activité musculaire diffère pour chaque mode phonatoire, la forme du conduit vocal est alors également modifié, ce qui se reflète par des réponses fréquentielles différentes. Dans la Fig. 1, nous observons l'espace typique

F1-F2 où les voyelles */a/, /u/* et */i/* sont produites. Nous pouvons observer que la phonation "normale" occupe plus d'espace que les autres modes. Les modes serré et "flow" partagent des caractéristiques similaires, notamment un déplacement particulièrement important de la fréquence de F1 pour les voyelles */a/* et */i/*.

La richesse harmonique observée dans les modes serré et "flow" peut également être capturée en terme de Rapport Harmonique sur Bruit (HNR). Ce paramètre est communément utilisé dans des scénarios d'identification automatique d'instruments de musique, où certaines familles d'instruments peuvent avoir des excitations plus importantes dans les harmoniques élevées. Le HNR est mesuré par une technique basée sur le cepstre habituellement utilisé pour les signaux de voix parlée (de Krom, 1993). Dans cette méthode, la taille de la fenêtre d'analyse est fixée comme étant égale à 5 périodes du fondamental. La mesure de HNR est alors effectuée en prenant le composant fréquentiel du cepstre et en comparant le niveau d'énergie au seuil du bruit. Les mesures sont prises sur différentes bandes de fréquences : 0-0.5KHz, 0-1.5KHz, 0-2.5KHz et 0-3.5KHz. La fréquence fondamentale est calculée sur l'ensemble des données grâce à l'algorithme STRAIGHT (Kawahara *et al.*, 1998). En application une analyse FFT et étant donné le F0 calculé, les trois premières harmoniques H1 H2 et H3 sont extraites et leurs amplitudes normalisées sont calculées en dB ainsi que les différences H1-H2 et H2-H4. De plus, les différences entre l'amplitude du fondamental et l'amplitude des formants sont également calculées : H1-Amp1, H1-Amp2 et H1-Amp3. Enfin, nous ajoutons à l'ensemble de ces paramètres le pic de proéminence cepstral (Cepstral Peak Prominence - CPP), un paramètre qui a été élaboré pour la détection de voix soufflée et qui est basé sur une mesure de périodicité du ceptre (Hillenbrand *et al.*, 1994). L'ensemble de ces paramètres est calculé sur des fenêtres hanning de 25ms avec recouvrement toutes les 5 ms. Au final, la moyenne de chaque paramètre est calculée sur chaque fichier, résultant en un vecteur de 24 paramètres par fichier. Dans le tableau 1

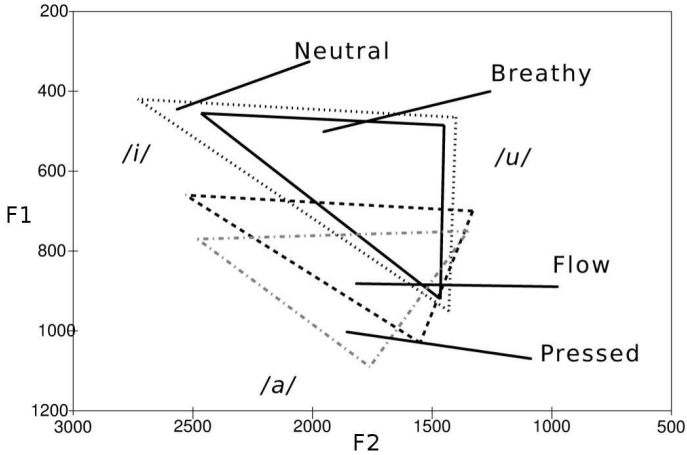

FIGURE 2 – Espace occupé dans le plan F1-F2 par chaque mode phonatoire pour les voyelles /a/, /i/, /u/.

ind.	nom	Desc.	ind.	nom	Desc.
1	H1	ampl. 1 harm.	13	HNR05	HNR (0-0,5KHz)
2	H2	ampl. 2 harm.	14	HNR15	HNR (0-1,5KHz)
3	H4	ampl. 3 harm.	15	HNR25	HNR (0-2,5KHz)
4	A1	ampl. 1 formant.	16	HNR35	HNR (0-3,5KHz)
5	A2	ampl. 2 formant.	17	F1	1 Formant
6	A3	ampl. 3 formant.	18	F2	2 Formant
7	H1H2	Difference H1 - H2	19	F3	3 Formant
8	H2H4	Difference H2 - H4	20	F4	4 Formant
9	H1A1	Difference H1 - A1	21	B1	1 Bandwidth
10	H1A2	Difference H1 - A2	22	B2	2 Bandwidth
11	H1A3	Difference H1 - A3	23	B3	3 Bandwidth
12	CP	Cepstral Peak Prominence	24	B4	4 Bandwidth

TABLE 1 – Recupulatif des descripteurs choisis

4 Classification

Après avoir collecté les données, nous apprenons un modèle pour chaque mode phonatoire en utilisant des techniques d'apprentissage automatique. Le logiciel Weka et la bibliothèque Libsvm ont été utilisés dans ces expériences.

4.1 Apprentissage automatique

Deux méthodes sont utilisées pour modéliser les modes phonatoires : les "Logistic Model Trees" (LMT) et les Machines à Vecteurs Support (SVM). Les SVM sont populaires pour la classification de motifs audio, notamment pour les tâches de classification supervisées comme dans le cas présent. Les SVM atteignent de bonnes performances pour la classification multiple et peuvent

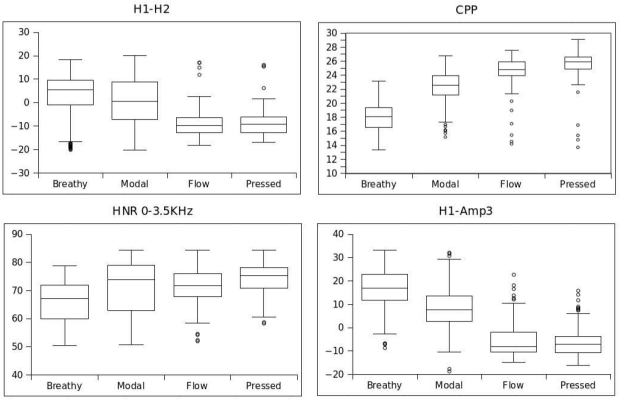

FIGURE 3 – Statistiques pour quatre paramètres. Toutes les valeurs sur l'axe Y sont en dB

traiter des vecteurs de paramètres de dimensions importantes. La méthode consiste à recherche un hyperplan qui permet de séparer linéairement les classes avec un critère de marge maximale. Afin d'éviter les problèmes de généralisations, nous effectuons une validation croisée en 10 itérations permettant d'obtenir des résultats moins biaisés que l'approche classique apprentissage-test. Les performances des deux méthodes de classification sont évaluées par le biais des mesures de précision et rappel, ainsi que la F-mesure qui est une combinaison de ces deux mesures. Parmi les deux méthodes employées, ce sont les LMT qui permettent d'obtenir les meilleurs résultats en termes de précision et de F-mesure. Les LMT sont des arbres de classification qui utilisent la régression logistique sur leur feuilles. Les SVM obtiennent des résultats à peu près similaires, mais avec un temps de calcul moindre par rapport aux LMT, ce qui est intéressant d'un point de vue performance vs. temps de calcul. Les résultats détaillées sont présentés dans la section suivante.

SVM			**LM trees**			
Pr.	Re.	F-m.	Pr.	Re.	F-m.	Class
0.87	0.921	0.894	0.918	0.933	0.925	BR
0.857	0.845	0.851	0.903	0.907	0.905	NE
0.721	0.676	0.698	0.727	0.717	0.722	FL
0.712	0.702	0.707	0.725	0.709	0.717	PR
0.807	0.81	**0.808**	0.84	0.842	**0.841**	Av.

TABLE 2 – Resultats de classification automatique avec SVM et LMT

4.2 Results

La précision maximale (84.15%) est atteinte en utilisant les LMT. Les résultats analytiques sont données dans le tableau 2. Nous pouvons observer que les models soufflé et modal (NE) sont les mieux classés. D'autre part, le modes "flow" (FL) et serrée (PR) ne sont pas très bien distingués. La matrice de confusion obtenue avec la classification par LMT est donnée dans le tableau 3.

Nous pouvons observer dans ce tableau que les vrais positifs pour les modes soufflé et modal sont nombreux et qu'un faible confusion est faite entre eux, alors qu'il sont bien distingués des autres modes. Le même phénomène peut être observé pour les modes "flow" et serrée, cependant cette fois la confusion entre ces deux classes est plus garnde.

BR	NE	FL	PR	←Classified as
223	16	0	0	BR
18	**205**	1	2	NE
0	5	**104**	36	FL
2	1	38	**100**	PR

TABLE 3 – Confusion matrix for the LMT classification experiment.

5 Discussion and Perspectives

La qualité vocale dans la voix chantée est un sujet de recherche actuel à cause de son importance dans la classification par le contenu de la musique. La phonation joue un rôle prépondérant dans la perception globale de la qualité vocale. Nous avons présenté ici une méthode pour la classification automatique des modes phonatoire soufflé, model, "flow" et serrée en employant des caractéristiques acoustiques extraites directement du signal. Les performances mesurées sont relativement élevées, particulièrement pour les deux premières classes qui semblent plus faciles à distinguer avec l'ensemble de paramètres que nous avons utilisé. Les confusions observées entre les modes "flow" et serrée démontrent cependant que l'utilisation de méthodes plus sophistiquées est nécessaire pour améliorer les résultats. C'est possible que cette confusion est à cause de l'énergie spectrale qui est plus élevé que dans les autres classes. Notre ensemble de descripteurs n'est pas capable de capter des différences que sont davantage détectables dans le signal de la source. Nous envisageons d'étendre ce travail en ajoutant à notre jeu de paramètres des mesures d'instants d'ouverture et de fermeture glottique, ainsi que leur quotient. Les signaux glottaux semblent en effet utiliser des motifs différents selon la phonation. De plus, nous devons examiner la pertinence de notre méthode, mise au point sur une base de voyelles prononcées isolément, pour classer et caractériser la voix chantée lors d'enregistrements continus. Il est aussi important de faire ce travail sur d'autres sujets et pas seulement chez les sopranos. Enfin, il est important d'évaluer dans quelle mesure des sujets humains sont capables de distinguer les différents modes phonatoires, en utilisant la même base de données que pour notre expérience de classification automatique. Il a en effet été démontré que selon le niveau d'expérience ou d'entraînement, nous ne possédons pas tous de manière égale la faculté de distinguer la qualité vocale (Henrich et Sotiropoulos, 2007), (Sofranko et Prosek, 2012). Il sera intéressant de comparer les résultats de classification automatique avec la perception humaine.

6 Remerciements

Ce travail est partiellement soutenu par le projet ANR DIADEMS ANR-12-CORD-0022.

Références

ALKU, P. (2011). Glottal inverse filtering analysis of human voice production — A review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana*, 36(October):623–650.

de KROM, G. (1993). A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals. *Journal of Speech, Language, and Hearing Research*, 36(2):254–266.

FANT, G. (1981). The source filter concept in voice production. *Speech, Music and Hearing Quarterly Progress and Status Report*, 22(1):021–037.

GRILLO, E. U. et VERDOLINI, K. (2008). Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects. *Journal of voice : official journal of the Voice Foundation*, 22(5):546–52.

HENRICH, N. et SOTIROPOULOS, D. (2007). Characterisation of Voice Quality in Western Lyrical Singing : from Teachers ' Judgements to Acoustic Descriptions. *Journal of Interdisciplinary Music Studies*, 1(2):62–91.

HILLENBRAND, J., CLEVELAND, R. A. et ERICKSON, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech, Language, and Hearing Research*, 37(4):769–778.

KANE, J. et GOBL, C. (2011). Identifying regions of non-modal phonation using features of the wavelet transform. *INTERSPEECH*, pages 177–180.

KAWAHARA, H., MASUDA-KATSUSE, I. et CHEVEIGN, A. D. (1998). Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction : Possible role of a repetitive structure in sounds.

MCAULAY, R. et QUATIERI, T. (1986). Speech analysis/synthesis based on a sinusoidal representation. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 34(4):744–754.

PROUTSKOVA, P., RHODES, C. et CRAWFORD, T. (2012a). Breathy or Resonant a Controlled and Curated Database for Phonation Mode Detecion in Singing. *Internasional Symposium on Music Information Retrieval*, (Ismir):589–594.

PROUTSKOVA, P., RHODES, C., CRAWFORD, T. et WIGGINS, G. (2012b). Breathy , Resonant , Pressed - Automatic Detection Of Phonation Mode From Audio Recordings of Singing. *Journal of New Music Research*, pages 1–28.

SOFRANKO, J. L. et PROSEK, R. a. (2012). The effect of experience on classification of voice quality. *Journal of voice : official journal of the Voice Foundation*, 26(3):299–303.

SUNDBERG, J. (1987). *The science of the singing voice*. Northern Illinois University Press.

SUNDBERG, J. (1995). Vocal fold vibration patterns and phonatory modes. *Folia Phoniatrica Logopedica*, 47:218–228.

TITZE, I. R. (2001). Acoustic interpretation of resonant voice. *Journal of voice : official journal of the Voice Foundation*, 15(4):519–28.