Approches complémentaires pour l'évaluation des dysphonies : bilan méthodologique et perspectives

Alain Ghio, Gilles Pouchoulin, Antoine Giovanni, Corinne Fredouille, Bernard Teston, Joana Révis, Jean-François Bonastre, Danièle Robert-Rochet, Ping Yu, Maurice Ouaknine, et al.

APPROCHES COMPLÉMENTAIRES POUR L'ÉVALUATION DES DYSPHONIES : BILAN MÉTHODOLOGIQUE ET PERSPECTIVES

A. Ghio, G. Pouchoulin**, A. Giovanni*, C. Fredouille**, B. Teston, J. Révis*, J.-F. Bonastre**, D. Robert*, P. Yu*, M. Ouaknine*, M.-D. Guarella*, C. Spezza*, T. Legou, A. Marchal

*LPL : Laboratoire Parole et Langage, CHU Timone, Marseille
**LIA : Laboratoire d'Informatique d'Avignon

Résumé

Nous proposons un bilan méthodologique fondé sur différentes expériences effectuées dans notre groupe de travail sur l'évaluation des troubles de la voix. Un premier axe d'étude a mis en parallèle un jugement perceptif de la qualité vocale de 449 participants avec des mesures instrumentales acoustique et aérodynamique effectuées sur le même groupe. Les résultats montrent que la combinaison de 7 paramètres instrumentaux permettent de classer 82 % des participants dans le même groupe que le jugement perceptif. Le deuxième axe d'étude, complémentaire, concerne l'adaptation de techniques de Reconnaissance Automatique du Locuteur à la catégorisation des dysphonies. Les expériences conduites sur 80 voix de femmes ont fourni des résultats plus que prometteurs et ont souligné l'intérêt d'une telle approche originale. Nous profiterons de la multiplicité de ces moyens expérimentaux pour faire un point méthodologique qui pointe des différences fondamentales entre ces approches complémentaires (montante vs *descendante, globale* vs *analytique). Nous discuterons aussi d'aspects théoriques notamment sur les relations entre mesures physiques et mécanismes de perception, considérations qui sont souvent mises de côté du fait de la course à la performance.*

Mots-clés : dysphonie, qualité vocale, évaluation perceptive, évaluation instrumentale, dispositif EVA d'évaluation vocale assistée, reconnaissance automatique du locuteur.

Abstract

This paper describes comparative studies of voice quality assessment based on complementary approaches. The first study was undertaken on 449 speakers whose voice quality was evaluated in parallel by a perceptual judgment and objective measurements on acoustic and aerodynamic data. Results showed that a non-linear combination of 7 parameters allowed the classification of 82 % voice samples in the same grade as the jury. The second study relates to the adaptation of Automatic Speaker Recognition (ASR) techniques to pathological voice assessment. Experiments conducted on 80 female voices provide promising results, underlining the interest of such an approach. We benefit from the multiplicity of theses techniques to evaluate the methodological situation which points fundamental differences between these complementary approaches (bottom-up vs *top-down, global* vs *analytic). We also discuss some theoretical aspects about relationship between acoustic measurement and perceptual mechanisms which are often forgotten in the performance race.*

Keywords: dysphonia, voice quality, perceptual assessment, GRBAS scale, instrumental assessment, EVA workstation, automatic speaker recognition, GMM.

GHIO, Alain *et al.,* Approches complémentaires pour l'évaluation des dysphonies : bilan méthodologique et perspectives, *Travaux Interdisciplinaires du Laboratoire Parole et Langage,* vol. 26, p. 33-74.

1. L'analyse de la qualité vocale

Dans le domaine de la phonétique, l'analyse de la qualité de la voix est généralement intégrée dans l'étude des phénomènes paralinguistiques de la communication parlée (Laver, 1981). Actuellement, la majorité des études sur cette thématique porte sur les relations entre l'état émotionnel du locuteur et les indices acoustiques portés par le signal vocal, ou encore sur la variabilité de la qualité de la voix en fonction de facteurs dialectaux et socioculturels (Campbell, 2000 ; Audibert *et al.,* 2004 ; Gobl *et al.,* 2003). Dans notre cas, depuis une quinzaine d'années, nous nous sommes penchés plus particulièrement sur les relations entre l'état physiologique du locuteur et sa qualité vocale, notamment dans un cadre clinique de dysfonctionnement du système pneumo-phonatoire. Dans cette prise en charge des dysphonies, l'étape de l'évaluation vocale apparaît nécessaire pour établir un bilan vocal au temps 't', pour contrôler l'évolution longitudinale de l'état vocal d'une même personne, pour dépister des situations de forçage, pour mesurer l'efficacité de différentes solutions thérapeutiques, pour permettre des comparaisons entre les différentes formes de pathologies et enfin, pour mieux comprendre les mécanismes mis en jeux d'un point de vue fondamental. Cette démarche s'inscrit dans une logique de bilan quantitatif, lequel se généralise dans le domaine de la santé.

Il existe une grande variété de méthodes pour établir un bilan vocal de personnes atteintes de troubles de la voix : interrogatoire avec le patient, examen endoscopique du larynx (Crevier *et al.,* 1993), appréciation du comportement postural du patient (Giovanni *et al.,* 2006), profil psychologique et étude comportementale (Roy *et al.,* 2000), questionnaire d'auto-évaluation (Woisard *et al.,* 2004), jugement perceptif de la qualité vocale (Révis, 2004), analyse instrumentale (Teston, 2004). La multiplication des angles d'observation s'avère nécessaire pour prendre en compte l'aspect multidimensionnel de la communication parlée, une méthode prise isolément se révélant souvent réductrice. Par la suite, nous ne présenterons que nos travaux effectués sur le jugement perceptif et les mesures instrumentales multiparamétriques.

2. Historique et genèse des différentes approches

La dimension perceptive de la voix reste un facteur essentiel pour l'évaluation de la qualité vocale, et ce pour plusieurs raisons. Tout d'abord, la parole et la voix sont produites pour être perçues. Les mécanismes perceptifs restent donc primordiaux dans le processus de communication parlée. D'ailleurs, la plupart des patients dysphoniques se décident à consulter au moment où la personne ou son entourage entend des changements dans le résultat vocal uniquement sur des sensations perceptives. À l'autre bout de la chaîne de prise en charge, les résultats thérapeutiques sont majoritairement appréciés par les cliniciens à l'écoute de la voix du patient : la perception auditive

est la modalité première, la plus accessible, pour évaluer la qualité vocale. Enfin, l'être humain et son système perceptif restent les plus performants pour décoder la parole même si on peut observer régulièrement des améliorations dans les performances des systèmes de reconnaissance automatique. Pour toutes ces raisons, le jugement perceptif de la voix demeure un procédé répandu en pratique clinique et fortement recommandé (De Jonckere *et al.,* 2001).

Pourtant, le jugement perceptif reste une méthode controversée car sujette à diverses imperfections. Les phénomènes de variabilité intra-auditeur[1] ou inter-auditeurs[2] en sont les éléments principaux. Pour obtenir une fiabilité raisonnable, l'évaluation doit être conduite par un jury d'experts. En effet, plusieurs auditeurs sont requis afin d'obtenir une appréciation moyenne ou consensuelle plus représentative de l'état vocal qu'un jugement isolé (Kreiman *et al.,* 1993). De même, il est préférable que soient sollicités des experts car diverses études comme celle d'Anders *et al.* (1998) ont montré que le niveau d'expertise contribue à améliorer la fiabilité des réponses, ce que l'on peut interpréter par le fait que des auditeurs non habitués à écouter des voix dégradées laissent une part plus importante à la subjectivité, génératrice de variabilité. Nous reviendrons sur cette question. De ce fait, une analyse perceptive fiable impliquant plusieurs auditeurs experts et plusieurs sessions d'écoute s'avèrent finalement consommatrices en temps et en ressources humaines, ne permettant pas une utilisation régulière en pratique clinique. Aussi, des approches instrumentales dites « objectives » fondées sur de la mesure physique ont été proposées pour pallier les faiblesses précédemment décrites de l'évaluation perceptive.

L'analyse instrumentale multiparamétrique est conçue pour qualifier et surtout quantifier les dysfonctionnements vocaux à partir de mesures acoustiques, aérodynamiques (débit d'air, pression intra-orale, pression sous-glottique) ou électrophysiologiques (électroglottographie[3], électro-myographie[4]…). Ces mesures sont réalisées sur le patient en cours de production vocale par le biais de capteurs conçus pour enregistrer et étudier de multiples paramètres de la production de parole. La majorité des études portant sur ces procédés font apparaître la nécessité de combiner différentes mesures complémentaires afin de tenir compte du caractère multidimensionnel de la production vocale (Wuyts *et al.,* 2000). En effet, une seule mesure isolée ne peut rendre compte à elle seule de dimensions différentes comme la raucité, le souffle ou le chevrotement. Nos

1. Un auditeur peut fournir des jugements variables dans le temps sur un même stimulus (inconstance).
2. Deux auditeurs fournissent des jugements différents sur un même stimulus.
3. Technique permettant d'observer l'accolement des cordes vocales par mesure de conductivité électrique à travers les plis vocaux.
4. Technique permettant d'observer l'activité musculaire à travers la mesure de différences de potentiels électriques au niveau d'électrodes placées *in vivo* ou en surface du muscle.

premières études sur cette thématique remontent à 1990 (Giovanni *et al.,* 1991) et nous n'avons cessé depuis d'étudier de nouvelles méthodologies et équipements spécifiques pour permettre de disposer de méthodes instrumentales d'évaluation vocale pour une utilisation clinique (Teston *et al.,* 1995 ; Giovanni *et al.,* 1996 ; Ghio *et al.,* 2004). Tout comme pour l'évaluation perceptive, les techniques instrumentales comportent un certain nombre de limites. Tout d'abord, la plupart des analyses sont fondées sur la production de voyelles tenues (/a/), contexte d'élocution très éloigné de la parole continue (Parsa *et al.,* 2001). Ensuite, l'analyse objective est souvent fondée sur des approches statistiques (analyse discriminante, corrélation…) appliquées sur des mesures qui peuvent être très dépendantes de la population de patients observés en termes de quantité et de qualité. Cela signifie qu'en changeant de cohorte clinique, la fiabilité des résultats peut s'effondrer notablement. Enfin, la nécessité d'utiliser des équipements particuliers de mesure peut s'avérer coûteuse, ce qui ne permet pas la diffusion à grande échelle de ces techniques pour une utilisation de routine clinique.

Ces restrictions nous ont conduits, récemment (Fredouille *et al.*, 2005), à tester l'adaptation, pour des locuteurs dysphoniques, de techniques issues de la Reconnaissance Automatique du Locuteur (RAL). Au départ, ces techniques sont conçues pour vérifier ou identifier automatiquement un locuteur, dans une certaine mesure, à partir d'une de ses productions vocales (Bimbot *et al.,* 2004). Des techniques similaires sont aussi capables, toujours dans une certaine mesure, de rattacher un locuteur à une catégorie comme dans le cadre de l'identification automatique de la langue parlée par un locuteur (Lamel *et al.,* 1994) ou d'accents régionaux (Ferragne *et al.,* 2006). Nous avons fondé notre travail sur l'hypothèse que la dysphonie pouvait être considérée de façon similaire à un accent régional, même si ce dysfonctionnement reste extralinguistique, et que les techniques utilisées en reconnaissance automatique étaient capables d'appréhender un tel phénomène. Comparée aux méthodes instrumentales classiques décrites dans le paragraphe précédent, l'originalité et l'intérêt d'une telle approche fondée sur de la modélisation statistique sont les suivants :

- une capacité à analyser de la parole continue proche de l'élocution naturelle ;
- une capacité à traiter de vastes corpora, autorisant des études à grande échelle et des résultats statistiques significatifs ;
- une analyse acoustique simple et automatique permettant une simplicité d'utilisation.

Les premières expériences fondées sur une simple analyse spectrale ou cepstrale associée à un système de classification automatique ont donné des résultats très encourageants pour l'évaluation vocale de patients dysphoniques (Fredouille *et al.,* 2005 ; Pouchoulin *et al.,* 2007).

Les parties suivantes seront consacrées aux résultats obtenus avec les trois approches que nous avons explorées et, enfin, un bilan sera présenté, ouvrant ainsi nos perspectives.

3. Évaluation vocale par analyse perceptive et mesures instrumentales multiparamétriques

Cette partie décrit un ensemble d'études ayant porté sur l'évaluation perceptive et instrumentale de patients dysphoniques.

3.1. Méthode

La méthode que nous avons choisie pour l'évaluation perceptive consiste à faire juger la qualité vocale de patients par des experts (phoniatres, orthophonistes) dont le rôle est de fournir un grade de dysphonie sur une échelle GRBAS proposée par Hirano (1981). Le principe est de faire lire au patient un texte normalisé dont l'énoncé enregistré est ensuite soumis en aveugle à divers juges expérimentés qui attribuent une note entre 0 (normal) et 3 (dysphonie sévère) par catégorisation directe ou à travers des échelles analogiques visuelles interprétées. Dans la plupart des cas, seul le grade G (global, général) de la dysphonie est exploité. Des détails sont disponibles dans Révis (2004).

Parallèlement, des mesures instrumentales sont effectuées sur les patients à l'aide du dispositif EVA® (Teston *et al.,* 1995 ; Ghio *et al.,* 2004) qui permet d'obtenir des mesures acoustiques primaires (F_0, intensité en dB SPL), des mesures de stabilité laryngée (jitter, shimmer, coefficient de Lyapounov[5]), des estimations de performance pneumo-phonatoire (étendue vocale, temps maximal de phonation) et des grandeurs aérodynamiques qui explorent de façon directe et sélective certains mécanismes comme la fuite glottique (par mesure de débit d'air oral) ou la tension de la source (par estimation de la pression sous-glottique). Suite au jugement perceptif, les patients sont regroupés par grade (de 0=normal à 3=dysphonie sévère) et les mesures instrumentales sont comparées à cette classification. En général, un test non paramétrique de Mann-Whitney permet de montrer s'il existe ou non des différences significatives entre les groupes pour chaque mesure instrumentale. De plus, l'utilisation d'analyse factorielle discriminante permet d'obtenir le pourcentage de concordance entre le jugement perceptif et les mesures instrumentales.

3.2. Historique

Plusieurs équipes ont abordé cette double évaluation perceptive et instrumentale. Dans Wolfe *et al.* (1995), les auteurs ont étudié la combinaison de 4 mesures acoustiques (F_0 moyenne, jitter, shimmer et rapport harmonique sur bruit) qu'ils ont confrontées à des jugements perceptifs du degré de dysphonie pour des patients porteurs de nodules, atteints de paralysie laryngée unilatérale ou de dysphonie dysfonctionnelle. À l'aide d'une analyse en régression, les auteurs ont montré que

5. Le calcul du coefficient de Lyapounov a été développé par M. Ouaknine et des détails sont disponibles dans Giovanni *et al.,* (1999).

la combinaison de ces paramètres permettait d'obtenir une corrélation de 0.56 entre les mesures perceptives et les mesures acoustiques. Dans Wuyts *et al.* (2000), les auteurs ont réalisé une étude multicentrique sur 319 sujets présentant une pathologie variée et 68 sujets témoins. Les patients ont été évalués par un jugement perceptif selon le grade G de la méthode GRBAS. A l'aide d'une analyse de régression multivariée, les auteurs ont sélectionné 4 mesures principales parmi une cohorte de 13 paramètres. À partir de ces 4 mesures sélectionnées (F_0 la plus élevée, intensité la plus basse, temps maximum de phonation, jitter), un index de sévérité de la dysphonie (Dysphonia Severity Index, DSI) est calculé par une simple équation linéaire[6]. En utilisant une analyse discriminante et la combinaison de ces 4 mêmes variables, les auteurs n'ont retrouvé que 56 % d'adéquation entre les deux méthodes d'évaluation instrumentale et perceptive. On peut signaler que dans les deux études précédentes, les mesures ne portaient que sur des paramètres acoustiques. Par contre, dans les études exposées par Piccirillo *et al.* (1998a, 1998b), portant sur un corpus de 97 sujets dysphoniques et 35 sujets témoins, ont été extraits 4 paramètres pertinents (pression sous-glottique, débit d'air oral, temps maximum de phonation et étendue vocale) parmi 14 mesures acoustiques, aérodynamiques et électroglottographiques, permettant d'obtenir une corrélation significative avec chaque paramètre de l'échelle GRBAS.

Cette même méthodologie a été conduite dans le service ORL du CHU Timone à Marseille depuis 1992. Le tableau 1 synthétise les principaux travaux relatifs à cette thématique. Plusieurs phénomènes sont observables :

a. la progression globale de la concordance entre le jury d'écoute et les mesures instrumentales que l'on peut expliquer par la prise en compte de mesures instrumentales de plus en plus pertinentes ;
b. un effet de saturation des performances dont nous discuterons plus loin ;
c. l'importance de la méthode d'évaluation perceptive, observable dans Yu *et al.* (2002) où, dans ce travail, ce sont les mesures instrumentales qui ont servi à confronter deux modes d'évaluation perceptive : une échelle ordinale classique et une échelle visuelle analogique discrétisée. L'écart de performance (64 % *vs* 88 %) est d'ailleurs éloquent et illustre les incertitudes liées à l'évaluation perceptive.

6. DSI = 0.13*TMP (s) + 0.0053*Fomax (Hz) – 0.26*Imin (dB) – 1.18*Jitter (%) + 12.4 (Wuyts *et al.,* 2000) *Remarque* : le DSI varie entre -5 pour les dysphonies sévères et +5 pour les voix normales. Il aurait été donc plus judicieux d'inverser la polarité de cet indice dit de « sévérité » (+5 pour les dysphonies sévères et -5 pour les voix normales) ou alors de le baptiser « Normality Index ».

Publication	Patients	Nb param.	Paramètres retenus		Concordance jury/mesures
Giovanni *et al.,* 1995, *Ann. Otolaryngol. Chir. Cervicofac.*	247 H/F/E divers	2	- jitter	- fuite glottique[7]	63 %
Giovanni *et al.,* 1996, *Folia Phoniatr Logop.*	245 H/F/E divers	4	- jitter - fuite glottique	- émergence harmonique - durée de l'attaque	66 %
Giovanni *et al.,* 1996, *Ann. Otolaryngol. Chir.*	23 normaux 34 Tucker[8] H	6 x 2	- jitter + Log(jitter) - shimmer + Log - coeff. variation F0 + Log	- coeff. variation dB + Log - débit oral + Log(dab) - fuite glottique + Log	84,7 %
Yu *et al.,* 2001, *Journal of Voice*	84 H divers	10	- F0 - jitter - intensité - rapport signal/bruit - rapport signal/bruit f>1kHz	- coef. de Lyapounov - débit d'air oral - pres. sous-glottique estim. - étendue vocale - temps max. phonation	86 %
Yu *et al.,* 2002, *Folia Phoniatr Logop.*	74 F divers	10	*idem*		64 % [9] 88 %
Yu *et al.,* 2007, *Folia Phoniatr Logop.*	449 H/F divers	6	- étendue vocale - coefficient de Lyapounov, - press. sous-glottique estim.	- temps max. phonation, - débit d'air oral - rapport signal/bruit	82 %

Tableau 1
Travaux effectués au CHU de la Timone sur la concordance entre jugement perceptif et mesures instrumentales

7. Le paramètre de fuite glottique est le rapport entre débit d'air oral et intensité. Elle correspond à la quantité d'air nécessaire pour émettre un son de 1 dB pendant une seconde.
8. 7 G1, 11 G2, 16 G3
9. 64% obtenus avec une échelle ordinale classique du grade G d'Hirano. 88% avec une échelle visuelle analogique discrétisée avec une segmentation non linéaire (voir détails dans Yu et al, 2002).

3.3. État actuel

Yu *et al.* (2007) détaillent les résultats les plus récents des expériences relatives à cette activité au sein de notre groupe de travail.

3.3.1. Patients

449 enregistrements sonores ont été sélectionnés dans les données du service ORL de l'hôpital de la Timone à Marseille, incluant 391 patients atteints de diverses pathologies vocales (308 femmes, 141 hommes) et 58 locuteurs de contrôle sans trouble vocal (38 femmes et 20 hommes). Les patients présentaient une variété de troubles vocaux typiquement rencontrés dans la pratique clinique (96 nodules, 91 polypes, 65 paralysies laryngées, 55 œdèmes de Reinke, 27 kystes, 24 dysphonies fonctionnelles, 19 dysplasies, 14 Sulcus).

3.3.2. Évaluation perceptive

Les instructions données aux sujets étaient de lire un texte standardisé (paragraphe de *La chèvre de M. Seguin* d'A. Daudet) à hauteur et intensité confortables, dans un local insonorisé. Les enregistrements ont été évalués par un jury composé de 4 auditeurs expérimentés. Trois sessions d'écoute ont été proposées. Ainsi, chaque production vocale a été évaluée 12 fois. La consigne donnée aux auditeurs était de fournir un grade pour les composantes G, R et B de l'échelle GRBAS d'Hirano (1981) mais seule la dimension G(lobale) a été analysée. Le jury d'écoute utilisait une échelle analogique visuelle et la position analogique était ensuite convertie en échelle discrète comme détaillée dans Yu *et al.* (2002). L'originalité était d'effectuer, non pas une discrétisation linéaire de l'échelle analogique, mais d'accorder des nuances réduites aux extrêmes (voix normales ou dysphonies très sévères) et plus importantes au milieu de l'échelle (dysphonie légère ou moyenne). Une telle démarche montre une amélioration des performances du juge en réduisant la variabilité inter juges et en renforçant la concordance avec les mesures instrumentales.

3.3.3. Analyse instrumentale multiparamétrique

Les mesures instrumentales ont été réalisées par l'intermédiaire du dispositif EVA®, SQLab-LPL, Aix-en-Provence, France (Teston *et al.,* 1995). Cet appareillage permet d'enregistrer simultanément des signaux acoustiques et aérodynamiques par l'utilisation d'une pièce à main (figure 1) contenant un microphone, un sonomètre, un pneumotachographe mesurant le débit d'air (Ghio *et al.,* 2004) et deux capteurs de pression. D'autres capteurs, comme l'électroglottographie, peuvent aussi être enregistrés en parallèle.

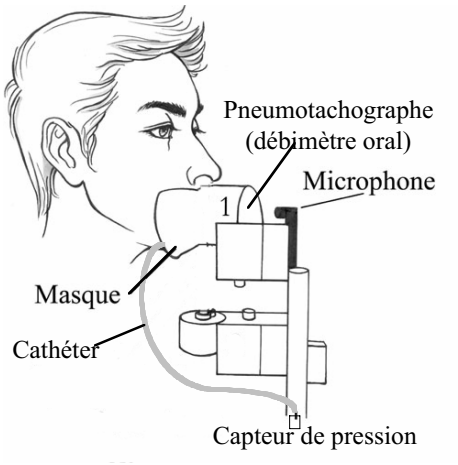


Figure 1
Pièce à main du dispositif EVA, permettant la mesure simultanée notamment de la voix, du débit d'air oral et de pressions (dessin réalisé par A. Michaud)

Les instructions données aux sujets sont de produire 3 voyelles tenues /a/ consécutives sur lesquelles est mesurée la F_0 (en Hz), l'intensité SPL[10] (en dB), le jitter factor (en %), le shimmer (en %), le rapport signal/bruit, le débit d'air oral (en dm3/s). Des détails sur ces différents paramètres sont disponibles dans Ghio, 2007. La pression sous-glottique est estimée de façon indirecte par la *« airway interrupted method »* de Smitheran & Hixon (1981) à l'aide d'un cathéter placé d'un côté dans la cavité orale du locuteur, de l'autre sur un capteur de l'appareillage EVA (figure 1). Le locuteur prononce alors une série de /papapa/. Sur la tenue de [p], les lèvres sont fermées. La glotte est ouverte. Un équilibre des pressions s'établit dans le conduit vocal : la pression sous-glottique peut être estimée en mesurant la pression intra-orale. Enfin, l'étendue vocale est mesurée en recueillant la F_0 la plus haute et la F_0 la plus basse que le locuteur peut produire après avoir reçu cette consigne de performance. De même, le temps maximal de phonation est mesuré après avoir demandé au locuteur de produire un /a/ le plus long possible suite à une longue inspiration.

10. Il s'agit d'une intensité calibrée analogue à celle fournie par un sonomètre. La comparaison d'intensité pour différents enregistrements est possible. Pour plus de détails, consulter le chapitre : *Le problème du calibrage de l'intensité*, dans Ghio (2007).

3.4. Résultats

La pertinence des mesures et les résultats détaillés des analyses discriminantes peuvent être consultés dans Yu *et al.* (2007). Chaque variable a été sélectionnée pour déterminer son effet sur le pouvoir discriminant entre grades. En utilisant une technique de « *stepwise backward* » dans laquelle toutes les variables sont introduites puis écartées une à une selon leur importance relative dans le modèle, nous avons pu identifier 7 paramètres pertinents pour les femmes et 6 pour les hommes pour prédire le grade global de dysphonie. Pour les femmes, il s'agit de l'étendue vocale, le coefficient de Lyapunov[11], la pression sous-glottique estimée, le temps maximal de phonation, le débit d'air oral, le signal *ratio* pour les fréquences au dessus de 1kHz[12] et la F_0. Pour les hommes, ont émergé l'étendue vocale, le coefficient de Lyapunov, le temps maximal de phonation, la pression sous-glottique estimée, la F_0 et le signal *ratio*. La figure 2 fournit les moyennes et écarts-type des mesures instrumentales décrites précédemment en fonction du grade de dysphonie (G0 à G3) et du genre (H/F).

3.4.1. Cohérence et interprétation des mesures instrumentales

Mise à part la mesure de la F_0 moyenne (figure 2e), nous observons une évolution cohérente des mesures instrumentales : plus la dysphonie est jugée sévère :

- plus le temps maximal de phonation (TMP) diminue (*cf.* figure 2a) ;
- plus l'étendue vocale (Voice Range ou VR) se réduit (*cf.* figure 2b) ;
- plus le débit d'air buccal (DAB) augmente (fuite glottique, *cf.* figure 2c) ;
- plus la pression sous-glottique (PSGE) augmente (forçage, *cf.* figure 2d) ;
- plus le jitter augmente (instabilité laryngée, *cf.* figure 2f) ;
- plus le taux d'énergie harmonique (Sr) diminue (pauvreté harmonique et présence de bruit, *cf.* figure 2g) ;
- plus le plus grand exposant de Lyapunov (PGEL) augmente (comportement chaotique du vibrateur laryngé, *cf.* figure 2h).

11. Une description simple des portraits de phase et des mesures de non-linéarité est disponible dans Ghio (2007).

12. Le Signal Ratio (Sr) se calcule à partir d'une méthode spectrale décrite dans (Hiraoka *et al.,* 1984). On considère le signal vocal comme la somme de deux composantes : une composante périodique (F_0 et harmoniques) et une composante de bruit (le reste). À partir d'un spectre effectué sur une portion de signal, l'énergie présente dans les raies est vue comme la composante périodique du signal. Le calcul de cette énergie dans les raies par rapport à l'énergie totale fournit un rapport appelé Sr. De plus, l'auteur préconise de faire le même type de calcul non pas sur tout le spectre mais uniquement sur des zones de moyennes et hautes fréquences comme, par exemple, au-dessus de 1 kHz. Dans ce cas-là, le Signal *Ratio* est particulièrement révélateur de dysphonie car il devient très faible chez ces patients dont le dysfonctionnement entraîne une composition spectrale appauvrie dans les aigus, caractéristique captée par le Sr calculé sans les graves.

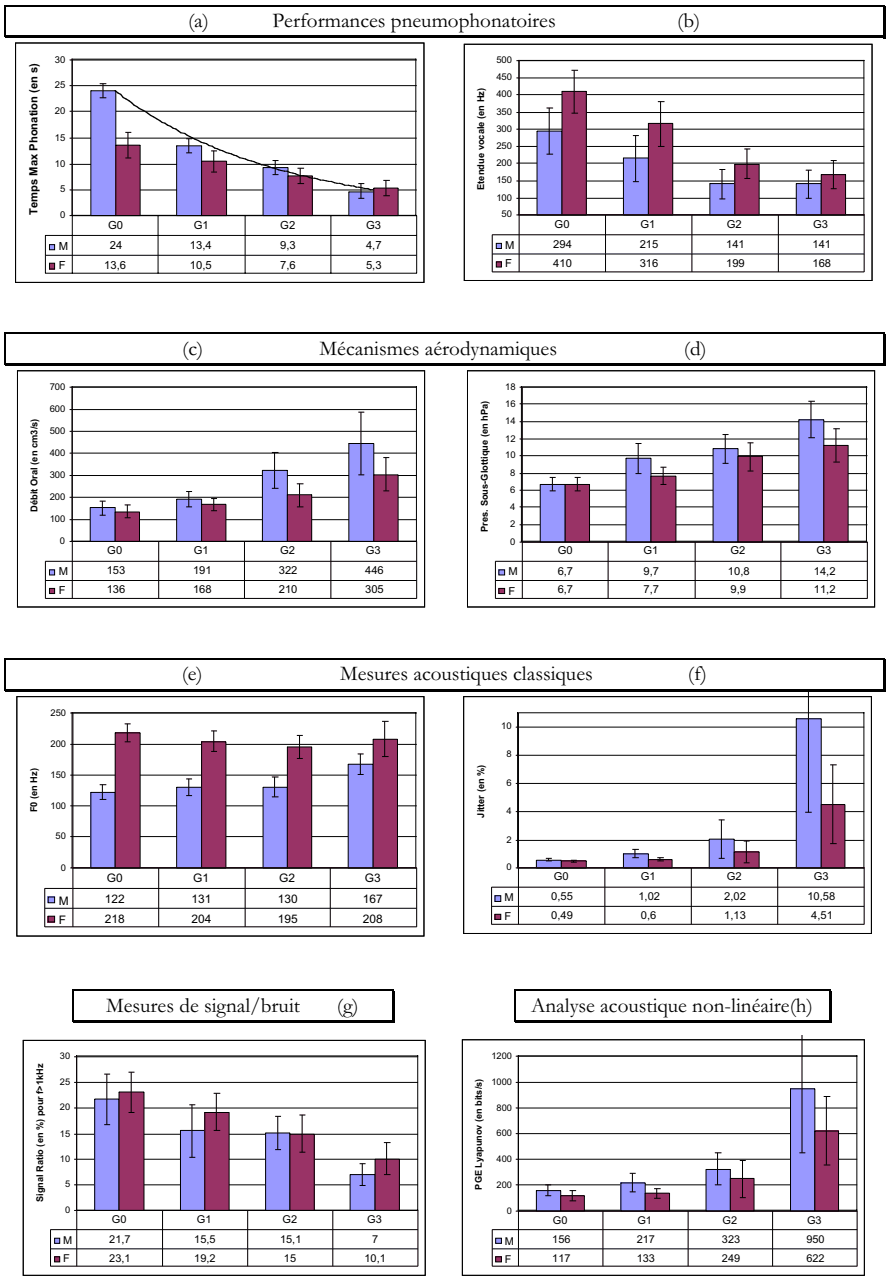
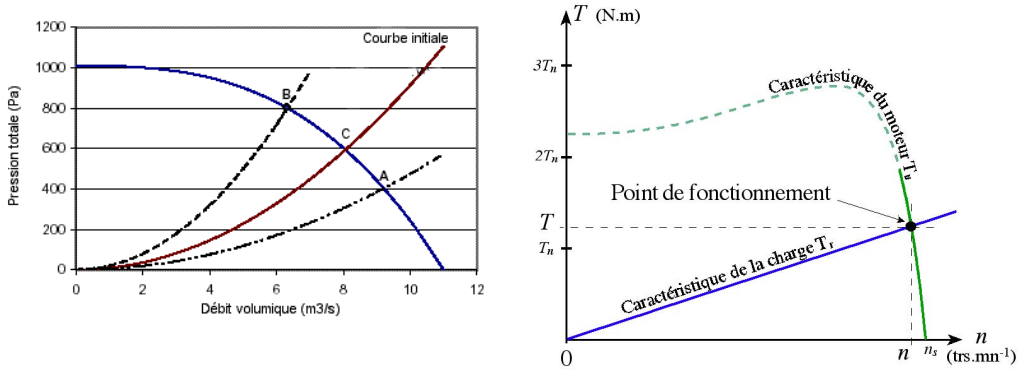


Figure 2 *Mesures instrumentales en fonction du grade de dysphonie (G0 à G3) et du genre*

Il est aussi important de noter la capacité d'interprétation de ces mesures avec un modèle fonctionnel de fonctionnement et dysfonctionnement de l'appareil phonatoire :

- les limites de l'espace de fonctionnement du système sont explorées par le TMP et le VR ;
- l'hypo ou hyperfonctionnement du système est exploré par la mesure de la PSGE ;
- le contrôle statique de la fréquence de vibration est exploré par la mesure du jitter et du PGEL ;
- le contrôle statique de l'amplitude de l'accolement est exploré par la mesure du DAB et du Sr.

Il est intéressant de faire des analogies avec d'autres systèmes mécaniques. Par exemple, la figure 3a représente un espace de fonctionnement d'un ventilateur. Le point C est le point de fonctionnement normal. Si, par exemple le filtre du ventilateur est encrassé, la perte de charge va croître dans le circuit de distribution aéraulique et le point de fonctionnement du ventilateur va passer du point C au point B. Le débit d'air va par conséquent décroître. La figure 3b représente un espace de fonctionnement d'un moteur électrique. Il permet de mettre en évidence les relations entre couple et vitesse de rotation du moteur, avec notamment les limites des performances et le point de fonctionnement optimal.



(a) Ventilateur

Source :
www.thermexcel.com/french/ressourc/dimensionnement_ventilateur_ventilateurs_pression.htm

(b) Moteur électrique

Source :
www.gmp.iut-tlse3.fr/IUT/cours/elec/Cours_moteurs_elec.ppt

Figure 3
Espaces de fonctionnement de systèmes mécaniques

La formalisation de tels espaces de fonctionnement pour le vibrateur laryngé, à partir des mesures instrumentales, pourrait permettre d'explorer les mécanismes de phonation, de comprendre les phénomènes de dysfonctionnement, de tirer parti de ces abaques personnalisés par locuteur pour adapter une stratégie thérapeutique.

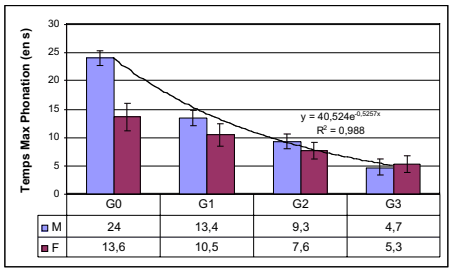
3.4.2. Vers un modèle prédictif de dysfonctionnement laryngé

La mesure de la F_0 moyenne (figure 2e) met en évidence un comportement opposé entre les locuteurs hommes et femmes, la F_0 des femmes diminuant globalement avec la sévérité de la dysphonie alors que celle des hommes augmente. Cela se traduit par une neutralisation du genre pour les dysphonies sévères, phénomène souvent observable chez un auditeur qui, écoutant en aveugle, rencontre des difficultés à définir le sexe d'un locuteur très dysphonique.

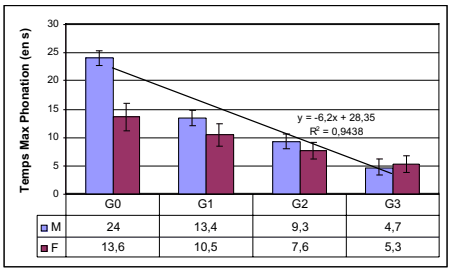
Excepté le point précédent, l'observation de la figure 2 laisse apparaître un « modèle de dysfonctionnement » identique pour les locuteurs hommes et femmes mais avec une dynamique plus réduite chez les femmes. Autrement dit, les mesures évoluent dans le même sens mais avec des écarts plus importants chez les hommes, d'où la nécessité de distinguer ces deux classes de locuteurs dans les modèles prédictifs de mesure du degré de dysphonie. Nous y reviendrons plus loin. De plus, il est important de remarquer que la métrique des mesures en fonction du grade n'est pas toujours linéaire. Par exemple, le temps maximal de phonation (figure 2a), le jitter (figure 2f) ou le plus grand exposant de Lyapunov (figure 2h) évoluent de façon exponentielle. Si l'on se place dans des modèles statistiques linéaires, il est nécessaire de linéariser ces mesures, démarche qui avait d'ailleurs été proposée dans Giovanni *et al.* (1996) où les auteurs avaient fait porter leur analyse statistique sur le Log(jitter).

Les deux remarques précédentes (modèles H/F, métrique non-linéaire) peuvent expliquer l'échec des modèles linéaires généraux comme le DSI de Wuyts *et al.,* 2000. En effet, cet index[13] est une simple combinaison linéaire qui d'une part, s'applique sans distinction pour les hommes et les femmes alors qu'au vu de nos résultats, cette distinction est nécessaire. D'autre part, il serait plus judicieux de linéariser les variables explicatives. Ainsi, par exemple, la courbe de tendance exponentielle du temps maximal de phonation (courbe noire superposée sur la figure 4a) est la suivante : $TMP = 40.524 \times e^{-0.5257(G+1)}$ avec un coefficient de détermination de R^2=0.988.

13. DSI = 0.13*TMP (s) + 0.0053*Fomax (Hz) – 0.26*Imin (dB) – 1.18*Jitter (%) + 12.4 (Wuyts *et al.,* 2000)



(a) Courbe de tendance exponentielle



(b) Courbe de tendance linéaire

Figure 4
Courbes de tendance de la variable Temps Maximal de Phonation
dans un modèle prédictif de la sévérité de la dysphonie

La transposition de l'équation exponentielle précédente permet de prédire :

$$G_{predictif} = \frac{3.7 - Ln(TMP)}{0.5257} - 1$$

alors qu'une prédiction linéaire, comme utilisée dans le DSI serait $G_{predictif} = \frac{28.35 - TMP}{6.2} - 1$

Les simulations exposées dans le tableau 2 montrent qu'une modélisation non linéaire permet de meilleures déterminations et prédictions.

Grade	0	1	2	3	
TMP mesuré (en s)	24	13,4	9,3	4,7	
TMP avec modèle linéaire (fig.4b)	22,2	16,0	9,8	3,6	$TMP = 28.35 - 6.2 \times (G+1)$
Écart modèle/mesure	-8%	19%	5%	-24%	
TMP avec modèle expon. (fig.4a)	24,0	14,2	8,4	4,9	$TMP = 40.524 \times e^{-0.5257(G+1)}$
Écart modèle/mesure	0%	6%	-10%	5%	
Grade prédit avec modèle linéaire	-0,3	1,4	2,1	2,8	$G_{predictif} = \frac{28.35 - TMP}{6.2} - 1$
Écart prédiction/grade réel	-0,30	0,41	0,07	-0,19	
Grade prédit avec modèle expon.	0,0	1,1	1,8	3,1	$G_{predictif} = \frac{3.7 - Ln(TMP)}{0.5257} - 1$
Écart prédiction/grade réel	-0,01	0,10	-0,20	0,09	

Tableau 2
Effet de la modélisation statistique de la variable Temps Maximal de Phonation *par rapport au grade G de dysphonie. La modélisation non linéaire permet de meilleures détermination et prédiction.*

3.5. Concordance des évaluations

Une analyse discriminante a été utilisée en mode prédictif pour construire une fonction de classement qui permet de prédire le groupe d'appartenance d'un locuteur à partir des mesures instrumentales. Cette technique s'apparente aux méthodes supervisées utilisées en apprentissage automatique ou à la régression logistique développée en statistique. L'intérêt pratique réside dans la fonction de classement qui s'exprime comme une combinaison linéaire des variables prédictives, c'est-à-dire les mesures instrumentales, et de fournir en sortie la probabilité d'appartenance à un groupe. Pour évaluer les performances de cette fonction de classement, l'idée est de confronter les prédictions, en l'occurrence le grade de dysphonie, issues des mesures instrumentales avec le grade proposé par le jugement perceptif considéré comme la vraie classe d'appartenance. Le tableau croisé qui en résulte est la matrice de confusion (tableau 3) avec en lignes les vraies classes d'appartenance, en colonnes les classes d'appartenance prédites. Le taux d'erreur est tout simplement le nombre de mauvais classement lorsque la prédiction ne coïncide pas avec la valeur attendue, rapporté à l'effectif des données. Cette opération a été réalisée séparément pour les hommes et les femmes, chaque groupe ayant une modélisation différente. Nous avons ensuite fait fusionner les matrices pour n'en obtenir qu'une seule présentée dans le tableau 3. Le résultat global montre que dans 82 % des cas, le locuteur est classé de façon concordante entre le jugement perceptif et les mesures instrumentales.

	G0 instrum.	G1 instrum.	G2 instrum.	G3 instrum.	Total	% correct
Grade 0 perceptif	**67**	5	0	0	72	93 %
Grade 1 perceptif	7	**94**	8	0	109	86 %
Grade 2 perceptif	2	29	**146**	21	198	74 %
Grade 3 perceptif	0	0	7	**61**	68	90 %
Total	76	128	161	82	447	82 %
% correct	88 %	73 %	91 %	74 %		

Tableau 3

Matrice de confusion entre prédiction du grade de dysphonie à partir des mesures instrumentales et jugement perceptif

Il faut noter que, dans ce cas-là, il s'agit d'un taux d'erreur en resubstitution, donc biaisé, car les données ont servi à la fois à la construction de la fonction de classement et à l'évaluation, autrement dit elles sont juges et parties dans ce schéma. C'est la raison pour laquelle une deuxième étude a été menée avec un deuxième jeu de données relatives à 46 nouveaux patients et locuteurs de contrôle. La concordance a alors atteint 80,5 % avec une majorité d'erreurs où l'analyse instrumentale sous-estime le grade de dysphonie (Yu *et al.,* 2007).

3.6. Bilan et perspectives de l'utilisation conjointe d'évaluation instrumentale et perceptive

3.6.1. Bilan des mesures instrumentales

La figure 5 retrace l'historique, à partir des études effectuées au sein de notre groupe de travail (tableau 1), du taux de concordance entre prédiction du grade de dysphonie issue des mesures instrumentales et jugement perceptif. Nous observons un effet de seuil autour de 80 %.

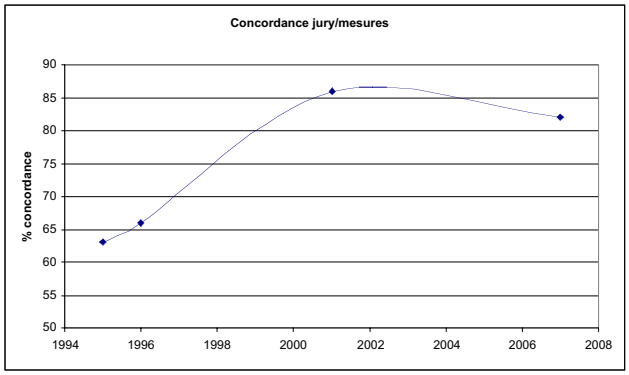


Figure 5
Évolution temporelle du taux de concordance entre prédiction du grade de dysphonie à partir des mesures instrumentales et jugement perceptif

3.6.2. Perspectives sur les mesures instrumentales analytiques

Les marges de progression autour de l'analyse instrumentale peuvent être les suivantes :

- améliorations des instruments de mesure ;
- nouvelles mesures complémentaires ;
- meilleure adaptation du contexte de production vocale.

3.6.2.1. Instruments et mesures

La marge de manœuvre liée aux mesures instrumentales nous semble plutôt restreinte, le matériel étant à présent intensément utilisé, validé, répandu dans divers centres hospitaliers européens et reconnu comme fiable. Le choix des mesures est à présent stabilisé autour des paramètres décrits précédemment. Seule l'utilisation de nouvelles mesures non redondantes avec les précédentes pourrait apporter de l'information complémentaire comme, par exemple, la mesure du quotient de fermeture à partir d'un dispositif d'électrogrottographie à l'instar de Fourcin *et al.* (2002) ou encore des mesures sur le trémor (Schoentgen, 2003).

3.6.2.2. Contexte de production vocale

On reproche souvent aux méthodes instrumentales la nécessité de productions phonatoires « artificielles » comme des /a/ tenus, des séquences /papapa/, des glissando. Il est souvent précisé que ce choix est lié à l'incapacité par ces méthodes à traiter de la parole continue. Cela n'est que partiellement exact. La raison principale est surtout liée à l'aspect sélectif et approprié de ce type d'élocution pour les besoins de l'évaluation vocale. En effet, l'objet de la mesure étant de quantifier le comportement du système pneumophonatoire, il n'est pas aberrant de placer ce système dans des situations certes peu naturelles, mais pertinentes pour l'observation. Cette situation se retrouve dans d'autres domaines de la santé comme en cardiologie, où le patient doit effectuer une série de flexions, attitude peu usuelle, pour évaluer la réponse cardiaque ou encore en audiologie où le principe même de l'audiogramme, qui consiste à écouter des sons purs, reste la référence malgré l'aspect très artificiel des stimuli sonores.

Dans le cadre de l'évaluation vocale, la production d'un /a/ tenu, où la consigne porte sur la stabilité, permet d'associer des variations de F_0 à un mauvais contrôle laryngé, raccourci bien plus difficile sur de la parole continue. En effet, les variations de F_0 peuvent rendre compte du double phénomène de contrôle et de régulation du vibrateur laryngé. Dans un fonctionnement correct, la capacité de pouvoir faire varier la fréquence de vibration des cordes vocales est révélatrice d'une bonne maîtrise vocale. C'est particulièrement le cas chez le chanteur qui possède une importante étendue tonale. On mesure alors une adéquation entre variation importante de F_0 et bon fonctionnement. Inversement, de fortes perturbations instantanées du cycle vibratoire se mesurent par des variations de F_0 révélatrices d'une dérégulation du système phonatoire. On retrouve alors une corrélation opposée à la précédente entre variations et dysfonctionnement. En fait, seules les variations contrôlées sont synonymes de bonne maîtrise. D'où l'importance cruciale du contexte de production vocale et des consignes.

Une consigne de production de voyelle tenue permet de focaliser la mesure sur la stabilité et régulation du vibrateur laryngé et d'écarter les variations contrôlées. C'est là l'intérêt majeur de l'utilisation de voyelles tenues. Dans ce cas-là, où la stabilité est demandée, comme une sorte de posture statique avec recherche d'immobilité, toute variation est synonyme de dysfonctionnement. Inversement, dans de la lecture de texte offrant une possibilité d'expressivité, tâche nécessitant un dynamisme s'apparentant à de la marche, les changements lents traduisent une bonne capacité vocale alors que les variations à court terme, voire les dévoisements inattendus, révèlent le dysfonctionnement.

De la même façon, une consigne de production de séquences /papapa/, de temps maximal de phonation, de glissando a pour vocation de placer le patient dans des situations d'élocution très

particulières, peu naturelles mais aussi très sélectives et analytiques. Contrairement au jugement perceptif qui se situe sur le plan de *l'utilisation de l'instrument* phonatoire, l'évaluation instrumentale est clairement tournée vers une détermination des caractéristiques « mécaniques » du *système* phonatoire. Il n'est donc pas illégitime que, dans ce cadre, l'évaluation porte sur des processus de production particuliers comme des voyelles tenues, des glissando… Par conséquent, du fait de la non-concordance des angles d'observation, il est légitime de constater une non-correspondance exacte des deux évaluations.

3.6.3. Bilan des méthodes conjointes instrumentales et perceptives

3.6.3.1. La non-relation directe entre réalité physique et sensation perceptive

Notre méthodologie actuelle consiste à considérer le jugement perceptif comme le *Gold Standard* et à confronter les mesures instrumentales à cette référence. La limite évidente de cette méthodologie réside dans l'utilisation d'appareillages de mesure comme « machines à écouter » avec lesquelles nous cherchons à obtenir des résultats identiques aux évaluations perceptives alors que ces deux procédés et approches sont fort différents. Il ne faut pas oublier que la perception, ou plus généralement la sensation, reste une réponse fortement non linéaire à une excitation (loi de Weber-Fechner citée dans Kitantou (1987), et qui s'applique d'ailleurs à d'autres sens que l'audition). Ainsi, Fletcher et Munson (1933) ont montré que non seulement la sensation auditive du niveau sonore n'est pas directement proportionnelle à la grandeur physique de l'amplitude (elle varie comme le logarithme de cette amplitude) mais que la fréquence du son influence aussi la perception du niveau sonore. Ils constataient, par exemple, qu'un son de 40 dB SPL à 1000 Hz devait être produit à 60 dB SPL à 100Hz pour être perçu au même niveau sonore. De même, les études menées par Mc Gurk & Mc Donald (1976) ont montré des effets surprenants et fortement non linéaires sur la perception audio-visuelle de la parole. Les auteurs ont monté artificiellement des stimuli audiovisuels où le son ne correspond pas systématiquement à l'image comme, par exemple, un visage articulant /ba/ avec un signal sonore /ga/. Or, confrontés à ce type de stimulus (ex : visuel /ba/, auditif /ga/), les participants à l'expérience perçoivent, de façon majoritaire, un mélange non linéaire des deux excitations (ex : visuel /ba/ + auditif /ga/ est perçu /da/). Imaginons que la réponse perçue soit considérée comme un *Gold Standard* (ex : /da/), une analyse acoustique fournissant /ga/ (qui correspond à la réalité) et un traitement de l'image proposant /ba/ (analyse exacte) seront alors considérés, de façon erronée, comme des résultats non concordants. Il n'est pas illégitime de penser que de tels phénomènes de fusion, de masquage, de relations non linéaires interviennent dans la perception de voix dysphoniques.

3.6.3.2. Des considérations méthodologiques orthogonales

L'évaluation instrumentale analytique a été conçue, à l'origine, pour fournir une réponse, sous la forme d'une ou plusieurs mesures, à une question claire au niveau physiologique. Prenons le cas des paralysies laryngées. L'immobilité d'une corde vocale se traduit par une importante fuite glottique et peut être traitée par médialisation (ex : goretex). Questions : la chirurgie a-t-elle réduit convenablement la fuite ? De combien ? Question subsidiaire : au niveau du résultat fonctionnel, cette technique réparatrice est-elle préférable à une autre (ex : injection de graisse, de collagène, thyroplastie) ? Pour mesurer une fuite d'air, le meilleur instrument reste le débitmètre qui peut fournir le débit d'air avant et après chirurgie, offrant directement une estimation chiffrée du taux de fermeture de la glotte en phonation et une mesure de l'impact de l'acte chirurgical. La démarche est clairement *analytique et descendante* : une hypothèse claire (fuite ?), une mesure adaptée à la question (le débit d'air oral en phonation), une réponse précise (le résultat de la mesure).

Les problèmes de subjectivité liés à l'évaluation perceptive de la voix ont conduit les cliniciens à adopter des mesures objectives. Ils ont donc utilisé les méthodes instrumentales analytiques, seules disponibles à l'époque, mais dans une approche non pas analytique descendante comme prévu, mais globale, montante et aveugle.

Aveugle ou tout du moins opaque dans la mesure où les cliniciens ont demandé à l'appareil de fournir des mesures permettant de classer le patient dans un grade 0, 1, 2 ou 3 de sévérité de la dysphonie, sans avoir jamais exprimé ou décrit clairement les caractéristiques de chaque grade.

Montante car la plupart des études portant sur l'évaluation des dysphonies sont fondées sur un recueil de nombreux paramètres avec pour objectif de faire émerger et mettre en évidence d'éventuels *clusters*. La démarche est dite montante car elle part des données (bas niveau d'observation) pour se diriger vers des catégories de plus haut niveau.

Globale car telle est la démarche utilisée dans le jugement perceptif, notamment pour la détermination du grade G de l'échelle GRBAS d'Hirano (1981). D'un point de vue géométrique, B, R et A+S peuvent être assimilés à 3 axes d'un espace métrique (fig. 5). Une voix peut être localisée par 3 coordonnées dans cet espace perceptif où R est la dimension raucité, B est relatif au souffle, A&S sont l'axe de l'hypo/hyper fonctionnalité. Dans cet espace, G apparaît comme une distance scalaire sans attribut de qualité. Ainsi, une voix évaluée comme R0 ; B2 ; A1 (fig. 5, G') aurait le même grade G=1 qu'une autre voix cotée R2 ; B0 ; S1 (fig. 5, G") alors que, d'un point de vue physiopathologique, ces deux patients sont très différents et pourtant référencés dans le même groupe G1, mettant ainsi en évidence la forte globalité de l'évaluation perceptive ainsi faite.

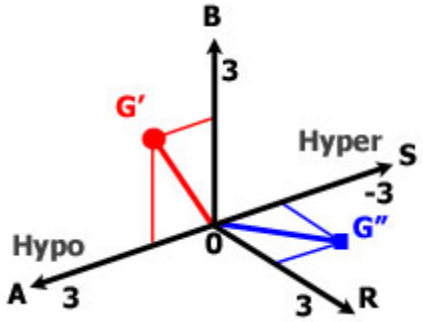


Figure 5
Espace perceptif GRBAS

Or, une technique analytique descendante utilisée dans une démarche globale, aveugle, montante ne peut qu'atteindre des limites, ce qui explique, actuellement, ce seuil de 80 % de concordances entre mesures instrumentales et perception.

3.6.3.3. Une alternative

Ce point méthodologique est fort comparable à la situation en reconnaissance automatique de la parole dans les années 90. Avant cette date, les scientifiques du domaine produisaient des systèmes analytiques descendants de type système expert, qui ont très vite été dépassés, par la suite, par les dispositifs stochastiques, qu'on peut qualifier, sans jugement de valeur, de globaux, aveugles, montants. En effet, si les phonéticiens sont parfaitement capables de mesurer, prédire et expliquer les phénomènes de coarticulation des /s/ dans le mot « soucis », ils sont démunis pour expliquer et décoder de façon exhaustive un flux continu de parole, notamment avec ses multiples formes de variabilité.

De façon identique, si les dispositifs instrumentaux classiques sont performants et gardent leur utilité, leur potentiel d'analyse et d'interprétation dans une approche analytique pour apporter une mesure précise à une problématique claire et restreinte relative aux dysfonctionnements de la parole, ils n'apportent pas de réelle réponse robuste, fiable et reproductible à la problématique très générale de l'évaluation globale de la dysphonie, aux variantes multiples. C'est pourquoi nous nous sommes tournés, de façon originale, dans l'évaluation des dysphonies, vers des approches stochastiques qui ont fait leurs preuves en reconnaissance automatique de la parole et du locuteur. Si ces techniques sont de loin les plus performantes dans des opérations d'identification, si elles ont l'avantage de pouvoir traiter de la parole continue, si elles sont capables d'intégrer par des

procédés statistiques une grande masse d'informations les rendant ainsi fiables, elles possèdent toutefois un inconvénient, en particulier pour les cliniciens, dans le sens où elles fonctionnent en boîte noire, ne permettant pas de comprendre précisément ou expliciter les mécanismes d'identification. Aussi, nous nous efforçons d'associer aux résultats issus du décodage automatique, une expertise d'ordre clinique ou phonétique et de ne pas nous satisfaire, uniquement, de la performance, sans justification d'ordre physiologique ou linguistique, faiblesse qui les empêcherait d'être adoptées par la communauté clinique.

4. Évaluation vocale par des techniques issues de la reconnaissance automatique du locuteur

4.1. La Reconnaissance Automatique du Locuteur (RAL)

Contrairement à la Reconnaissance Automatique de la Parole (RAP), la Reconnaissance Automatique du Locuteur (RAL) s'intéresse tout particulièrement aux informations extralinguistiques véhiculées par un signal de parole, informations porteuses de renseignements sur les spécificités d'un individu (identité, émotivité, caractéristiques physiques, particularités régionales…). Son objectif est d'identifier une personne à l'aide de sa voix grâce à la variabilité inter-locuteurs qui permet de reconnaître une voix parmi plusieurs possibles. L'état de l'art des systèmes actuels de reconnaissance automatique du locuteur utilise une approche statistique fondée sur les théories de la détection, de la décision bayesienne et de l'information.

4.1.1. Les tâches en RAL

Dans le domaine de la RAL, on distingue différentes tâches :

- *L'Identification Automatique du Locuteur :* qui consiste à déterminer la personne ayant prononcé un message donné, parmi un ensemble de locuteurs connus. On distingue deux modes :
  - en ensemble fermé : le locuteur à identifier est connu du système ;
  - en ensemble ouvert : le locuteur à identifier peut ne pas être connu du système.

  Ces applications sont peu nombreuses. En ensemble ouvert et dépendant du texte (par exemple, un même mot de passe pour les employés d'une même société), certaines applications d'IAL peuvent permettre le contrôle d'accès à un bâtiment, à un réseau ;
- *La Vérification Automatique du Locuteur :* consiste à déterminer la véracité de l'identité revendiquée par un individu, au moyen d'un message vocal. Ces applications sont multiples comme les serrures vocales pour le contrôle d'accès aux locaux, l'accès par le téléphone à des services distants sécurisés, la protection de matériel contre le vol (téléphones portables, voitures…) ;

- *Détection/Suivi de Locuteurs :* se rapproche de la VAL. Sa tâche consiste à déterminer si un locuteur donné intervient ou non dans un document audio (conférences, débats, conversations…). Ces applications sont principalement judiciaires et militaires. Cependant, en indexation de documents audio, elle peut faciliter la recherche d'un document audio particulier par la détection d'un locuteur connu (émission de télévision, de radio…) ;
- *L'Indexation de Locuteurs :* consiste à cibler les interventions de locuteurs dans un document audio (conférences, débats, conversations…). Ces applications sont principalement orientées sur le traitement de bases de données audio, comme par exemple la recherche de séquences d'émissions télévisées pour un locuteur particulier.

4.1.2. Description d'un système de RAL

Le processus de RAL comprend 3 phases : la paramétrisation, l'apprentissage et la phase de test.

4.1.2.1. La paramétrisation

La paramétrisation permet de réduire la redondance du signal de parole et d'en extraire les informations pertinentes en vue de la reconnaissance. Elle fournit ainsi une représentation simplifiée du signal nécessaire avant les phases d'apprentissage et de test. Cette représentation repose généralement sur des vecteurs de paramètres acoustiques correspondant à des trames de signal (longueur variant de 20 à 31,5 ms généralement) calculées périodiquement sur le signal de parole (par exemple, toutes les 10 ms). Suivant la nature des informations que l'on souhaite extraire du signal de parole, différentes représentations sont proposées. Ces représentations peuvent être classées en quatre grandes classes.

• Analyse spectrale

L'analyse spectrale met en évidence les caractéristiques physiques de l'appareil phonatoire (forme du conduit vocal et nasal) de chaque individu, à travers des vecteurs de paramètres qui en sont déduits. Les paramètres les plus pertinents en RAL sont les :

- LPC (Linear Predictive Coefficients) obtenus par prédiction linéaire ;
- LFC et MFC (Linear/Mel Frequency Coefficients) obtenus par analyse en banc de filtres.

Pour plus de détails, on se référera aux travaux de Charlet (1997), Homayounpour (1994) et Reynolds (1994).

• Analyse cepstrale

L'analyse cepstrale est une méthode qui vise à séparer la contribution de la source et du conduit vocal par déconvolution, en prenant comme hypothèse que le signal vocal est produit par un signal excitateur (source glottique) traversant le conduit vocal. Le spectre ainsi débarrassé de la

contribution de la source ne contient que des informations sur le conduit vocal. Les paramètres les plus pertinents en RAL sont les :

- LPCC (Linear Predictive Cepstral Coefficients) obtenus par prédiction linéaire ;
- LFCC et MFCC (Linear/Mel Frequency Cepstral Coefficients) obtenus par analyse en banc de filtres.

• Les paramètres prosodiques

Les paramètres prosodiques (mélodie, intensité et durée) mettent en évidence le style d'élocution (débit, durée et fréquence des pauses…), ainsi que les caractéristiques de la source glottale (fréquence fondamentale, énergie, taux de voisement…). Ces paramètres s'avèrent cependant fragiles en pratique et ne permettent pas, à eux seuls, de discriminer les locuteurs. En conséquence, ils sont souvent associés aux paramètres de l'analyse spectrale (par exemple l'énergie). De plus, ils restent difficiles à extraire de manière automatique.

• Les paramètres dynamiques

Souvent les paramètres dynamiques constituent un facteur d'amélioration des performances. Ils reflètent les phénomènes de co-articulation, les trajectoires formantiques ainsi que les informations temporelles à court terme (vitesse d'élocution, distribution des pauses). Un exemple d'exploitation des informations dynamiques a trait aux coefficients dérivés des spectres instantanés, appelés communément les coefficients Delta (ou Δ) pour la 1ère dérivée et Delta-Delta (ou $\Delta\Delta$) pour la 2ème dérivée.

4.1.2.2. Apprentissage statistique

En Reconnaissance Automatique de Locuteur, l'état de l'art repose sur une modélisation statistique s'appuyant sur le modèle de mélange de gaussiennes (GMM: Gaussian Mixture Model) (Reynolds, 1992). Cette densité de probabilité, répartition statistique des valeurs des vecteurs acoustiques, caractérise soit un locuteur, soit un ensemble de locuteurs. Un GMM X est une somme pondérée de M distributions gaussiennes multidimensionnelles, chacune caractérisée par un vecteur moyen $\overline{x}$ et une matrice de covariance Σ et un poids p.

Durant la phase d'apprentissage, les paramètres des GMM (le vecteur moyen $\overline{x}$ de dimension d, la matrice de covariance Σ de dimension d x d, la pondération p de chaque distribution gaussienne) sont estimés par l'algorithme EM (Expectation-Maximization)/ML (Maximum Likelihood) (Demster *et al.,* 1977). Suivant la loi de mélange de M gaussiennes, la densité de probabilité d'un vecteur y_t de dimension d s'écrit :

$$p(y_t|X) = \sum_{i=1}^{M} p_i N(y_t, \bar{x}_i, \Sigma_i)$$

Classiquement, deux phases d'apprentissage (figure 6) sont nécessaires en RAL pour pallier le manque de données d'apprentissage lié aux locuteurs (Bimbot *et al.,* 2004) :

1. apprentissage d'un modèle générique de parole (aussi appelé modèle du monde) estimé par l'algorithme EM/ML sur une grande quantité de données (population de locuteurs) ;
2. apprentissage du modèle locuteur dérivé du modèle du monde par application des techniques d'adaptation (MAP, Maximum a Posteriori) (Gauvain *et al.,* 1994).

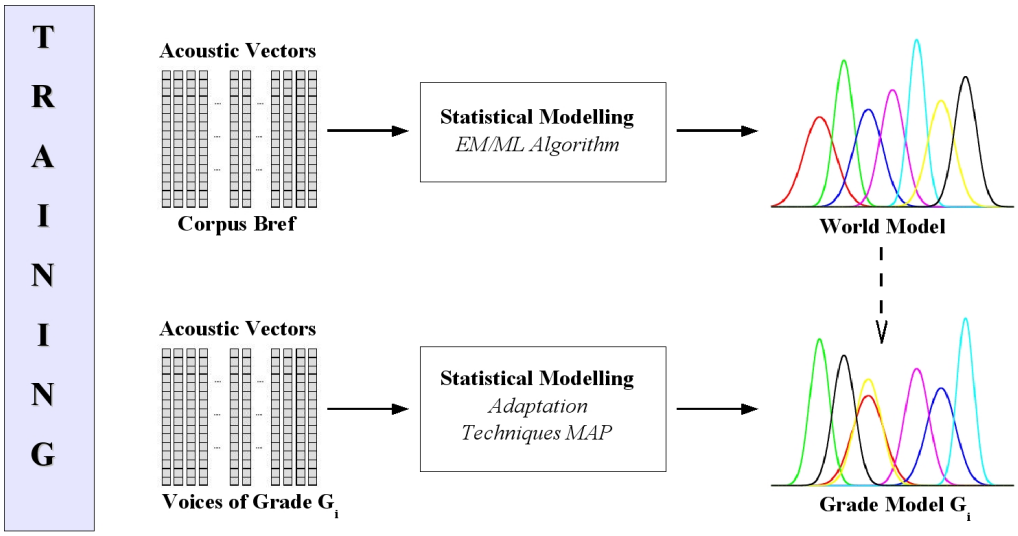


Figure 6
Phase d'apprentissage d'un système de RAL adapté aux voix pathologiques

4.1.2.3. Phase de Test : mesure de ressemblance et décision

Lors de la phase de test (figure 7), 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* qui s'exprime par :

$$L_i(y_t) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} e^{-\frac{1}{2}\left(y_t - \bar{x}_i\right)^T (\Sigma_i)^{-1}\left(y_t - \bar{x}_i\right)}$$

Le processus de décision est dépendant de la tâche visée. En VAL, la décision correspondra à une décision binaire *Acceptation* ou *Rejet* suivant que la mesure de ressemblance est supérieure à un seuil de décision. En revanche, en IAL (milieu fermé), elle correspondra à une décision 1 parmi N basée sur le *Maximum de Vraisemblance* désignant le locuteur le plus probable parmi les N connus du système.

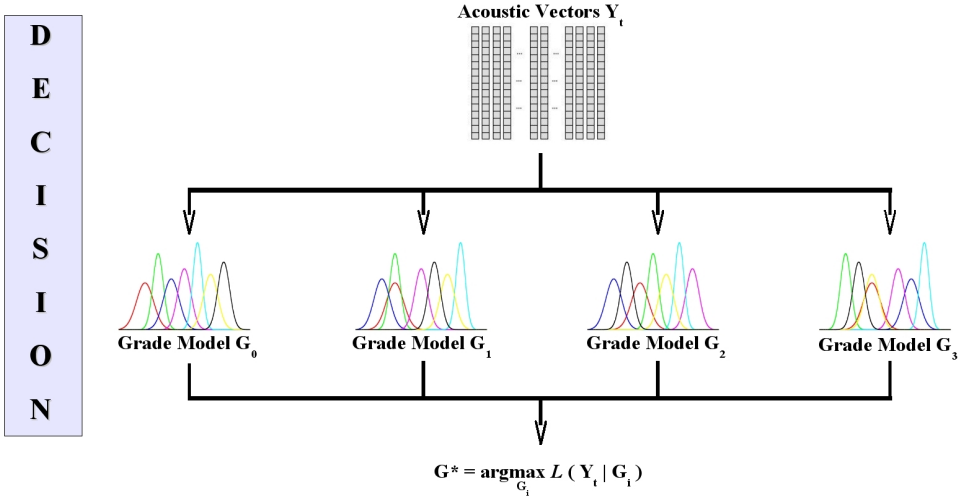


Figure 7
Phase de test d'un système de RAL adapté aux voix pathologiques

Le processus de décision est dépendant de la tâche visée. En VAL, la décision correspondra à une décision binaire *Acceptation* ou *Rejet* suivant que la mesure de ressemblance est supérieure à un seuil de décision. En revanche, en IAL (milieu fermé), elle correspondra à une décision 1 parmi N basée sur le *Maximum de Vraisemblance* désignant le locuteur le plus probable parmi les N connus du système.

4.2. Adaptation des techniques de RAL pour la reconnaissance des dysphonies

L'approche statistique à base de GMM a été mise à l'épreuve dans le cadre de la RAL lors de campagnes d'évaluation NIST (National Institute of Standards and Technologies) des systèmes. Elle sera adaptée à la tâche de reconnaissance des dysphonies à différents niveaux comme le montrent les sections suivantes.

4.2.1. Modèle de grade, World et Décisions

Dans le contexte pathologique, un modèle ne correspond plus ici 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.

Comme notre corpus est limité en taille pour chacun des grades, le schéma classique d'apprentissage des modèles locuteurs sera appliqué ici *(cf.* § 4.1.2.2). Les modèles de grade G_g seront dérivés du modèle générique par application d'une technique d'adaptation du type MAP. Dans la phase de classification, le signal de parole y_t du locuteur *Y* sera testé sur chacun des modèles de grade G_g avec $g \in \{0,1,2,3\}$. La **décision** correspondra au grade *g* du modèle G_g sur lequel la plus grande vraisemblance a été obtenue. Cette définition de la décision est proche de celle d'IAL. On dira que le système a classé la voix du locuteur *Y* dans le grade *g*.

4.2.2. Études des informations segmentales

Lors de l'expérimentation, une approche en mode segmental sera mise en œuvre. Pour chaque signal de parole du corpus, une segmentation sera extraite automatiquement par un processus d'alignement phonétique contraint sur le texte. Cette segmentation phonétique s'appuie 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. En revanche, lors d'un test de classification, la décision sera prise sur l'ensemble des segments associés à une classe phonétique. Dans ces conditions, il sera possible d'obtenir une décision *globale* sur un signal de parole ou une décision *locale* sur chacun de ses segments.

Cette démarche devrait permettre de mieux comprendre les phénomènes liés aux troubles de la voix. Une première étude pourra porter sur le caractère continu ou discontinu des phénomènes dysphoniques.

4.2.3. Les tâches de classification : Contrôle/Patho et 4-Grades

- la classification Contrôle/*Patho*
  La première action consistera à observer si un système de RAL, adapté à notre sujet, réagit favorablement à la classification binaire c'est-à-dire détecte si une voix donnée est reconnue en tant que *voix dysphonique* ou *voix normale* (= contrôle) ;
- la classification 4-Grades
  La seconde phase analysera le comportement du système à une classification par grades selon l'échelle GRBAS. L'ensemble des voix sera testé à travers le système et les résultats, comparés à ceux de l'analyse perceptive, permettront de mesurer les performances de différentes paramétrisations acoustiques et de l'apport des paramètres dynamiques.

4.2.4. Le corpus

Le corpus utilisé, issu du service ORL du CHU Timone à Marseille, est constitué de 80 échantillons de voix féminines correspondant à 20 sujets témoins et 60 patientes dysphoniques, âgés de 17 à 50 ans (moyenne de 32.2 ans). L'ensemble des patientes dysphoniques a eu un examen laryngoscopique faisant apparaître les pathologies énumérées dans le tableau 4.

Pathologies	Nombre	Pathologies	Nombre
nodules	22	cordes vocales normales	5
polype	10	Cordite	1
œdème	17	Sulcus	1
kyste	4	épaississement muqueux	1

Tableau 4
Répartitions des pathologies dans le corpus

En ce qui concerne le support vocal, 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 par consensus par un jury d'experts selon le grade global G de dysphonie de l'échelle GRBAS. L'ensemble du corpus étiqueté se présente de la manière suivante : 80 locuteurs équitablement répartis dans les 4 grades (20 voix normales, 20 voix avec dysphonie légère, 20 voix avec dysphonie moyenne, 20 voix avec dysphonie sévère). Il est à noter une grande variabilité dans les durées : de 13.5 à 77.7 secondes (18.7 secondes en moyenne).

4.3. Les expériences

Les expériences ont été réalisées après adaptation du système de RAL du LIA. Ce système de RAL (appelé LIA_SpkDet repose entièrement sur la plateforme libre Alize, Bonastre *et al.,* 2005) conçue et réalisée dans le cadre du programme Technolangue.

4.3.1. Les spécificités du système

4.3.1.1. Paramétrisation

Pour toutes les expériences, un retrait des silences est appliqué sur les signaux de parole. La durée moyenne conservée par grade est de :

[Grade 0 = 12.08 s] [Grade 1 = 12.05 s] [Grade 2 = 12.83 s] [Grade 3 = 14.03 s]

De plus, une normalisation à l'issue de la phase de paramétrisation de type centre-réduit (moyenne 0, variance 1) est appliquée sur les vecteurs de paramètres. En ce qui concerne la paramétrisation des signaux de parole, elle va dépendre du type de l'expérience :

- Tâche *Contrôle/Patho* : analyse en banc de filtres à échelle de Mel de type 16MFCC associé à 16 coefficients dynamiques **Δ** ;

- Tâche 4-Grades : analyse en banc de filtres à échelle de Mel de type 24MFC et 16MFCC, une analyse par prédiction linéaire de type 12LPC et 12LPCC. Les coefficients **Δ** et **ΔΔ** sont ajoutés suivant l'expérience.

4.3.1.2. Apprentissage des modèles

Le modèle du monde est appris par l'algorithme EM à partir de signaux issus du corpus Bref (Lamel *et al.*, 1991) c'est-à-dire de 76 enregistrements de 2 mn chacun de voix exclusivement féminines.

En raison de la taille réduite du corpus Patho, des protocoles particuliers à chacune des tâches visées sont nécessaires afin de :

- limiter l'influence de voix particulières au sein des modèles de grade G_g ;
- augmenter le nombre de tests et ainsi la pertinence (au sens statistique) des résultats obtenus.

Ces protocoles seront mis en place à l'aide de la technique leave_x_out, qui consiste à extraire la voix testée du corpus d'apprentissage et à recommencer le processus d'apprentissage avec une autre voix, elle-même exclue de cette phase.

4.3.2. Tâche 1 : protocole 2-Grades (Contrôle/Patho)

4.3.2.1. Protocole

Cette tâche consiste à déterminer si une voix est normale ou dysphonique. Par conséquent, deux modèles doivent être estimés, $G_{contrôle}$ et G_{patho} correspondant respectivement au modèle des voix normales (Grade 0) et au modèle des voix dysphoniques (Grades 1, 2 et 3).

La mise en œuvre de la technique leave_x_out permet de comparer, lors de la phase de test, chaque voix y_t avec :

- 6 modèles $G_{contrôle}$ appris chacun à partir de 18 voix normales ($y_t \notin$ aux 18 voix) ;
- 6 modèles G_{patho} appris chacun à partir de 18 voix dysphoniques également réparties sur les 3 grades ($y_t \notin$ aux 18 voix).

Le nombre identique de $G_{contrôle}$ et G_{patho} a été fixé pour homogénéiser le nombre de candidats (autant de candidats *patho* que de *contrôle*). De par la constitution du corpus, les 6 modèles $G_{contrôle}$ sont très proches les uns des autres. Par contre, les 6 modèles G_{patho} contiennent 18 locuteurs choisis aléatoirement (6 dans chacun des 3 grades). De même, chaque modèle est constitué du même nombre de locuteurs.

À l'issue de ces comparaisons, les moyennes des vraisemblances des tests *Contrôle* (6 modèles $G_{contrôle}$) et *Patho* (6 modèles G_{patho}) sont calculées et comparées pour fournir une unique décision pour la voix (y_t).

4.3.2.2. Résultats. Discussion

Le tableau 5 donne les résultats de la classification *Contrôle/Patho* qui obtient le score de 85,0 % indiquant un résultat prometteur et encourageant.

Contrôle	Patho	Total
% succès (nb sur 20)	% succès (nb sur 60)	% succès (nb sur 80)
95,0 (19)	81,7 (49)	85,0 (68)

Tableau 5

Résultat de la classification Contrôle/Patho - 16MFCC + Δ

La matrice de confusion (tableau 6) permet d'identifier les erreurs de classification suivant le grade de chaque voix.

1. la majorité des voix de grade 0 est bien classée (95,0 %) ;
2. les grades 1 et 2 sont reconnus à 70,0 % et 75,0 % ;
3. la totalité des voix de grade 3 est bien classée (100,0 %).

	Classification automatique	
Référence	Contrôle	Patho
Grade 0	**95 % (19)**	5 % (1)
Grade 1	30 % (6)	**70 % (14)**
Grade 2	25 % (5)	**75 % (15)**
Grade 3	0 % (0)	**100 % (20)**

Tableau 6

Résultat de la classification Contrôle/Patho *par grade - 16MFCC + Δ*

4.3.3. Tâche 2 : protocole 4-Grades

4.3.3.1. Protocole

Il s'agit de classer une voix suivant les 4 niveaux du grade général de l'échelle GRBAS. Par conséquence, 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 œuvre de la technique leave_x_out permet de comparer, chaque voix y_t de grade *g* avec :

- 1 modèle G_g appris à partir de 19 voix 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 $\bar{g}$ ($y_t \notin$ aux 19 voix) avec $\bar{g} \in \{0,1,2,3\} - \{g\}$

À 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*.

4.3.3.2. Résultats. Discussion. Les paramètres statiques

Le tableau 7 donne le résultat des différentes paramétrisations portant uniquement sur les coefficients statiques. Les paramétrisations MFCC et MFC ont les meilleurs résultats avec 70,00 % et 73,75 % respectivement. À l'exception du grade 2, leurs scores dans les différents grades sont similaires et montrent que le grade 0 est le mieux reconnu (95,0 % pour MFCC et 90,0 % pour MFC).

	Grade 0	Grade 1	Grade 2	Grade 3	Total
Expériences	% succès (nb sur 20)	% succès (nb sur 20)	% succès (nb sur 20)	% succès (nb sur 20)	% succès (nb sur 80)
12LPC	75,0 (15)	50,0 (10)	50,0 (10)	50,0 (10)	56,25 (45)
12LPCC	65,0 (13)	50,0 (10)	65,0 (13)	65,0 (13)	61,25 (49)
16MFCC	**95,0 (19)**	**55,0 (11)**	55,0 (11)	**75,0 (15)**	70,00 (56)
24MFC	90,0 (18)	**55,0 (11)**	**75,0 (15)**	**75,0 (15)**	**73,75 (59)**

Tableau 7

Résultats de la classification 4-Grades – Paramètres statiques

Les paramétrisations LPC et LPCC ont des scores inférieurs à ceux de l'analyse en banc de filtres. Les LPC obtiennent leur meilleur score pour le grade 0 (75,0 %), tandis que les LPCC l'obtiennent sur l'ensemble des grades (65,0 %) sauf pour le grade 1 (50,0 %).

1. le grade 0 obtient les meilleurs scores avec les MFCC et MFC (95,0 % et 90,0 %) ;
2. la discrimination est faible pour les grades dysphoniques (G1/G2) ;
3. les voix de grade 3 obtiennent un score de 75,0 % pour les MFCC et MFC.

Le comportement des coefficients cepstraux par rapport aux spectraux est plutôt inattendu, au regard de la suppression de toutes les informations relatives à la source glottale pour les coefficients cepstraux.

4.3.3.3. Résultats. Discussion. Ajout des paramètres dynamiques Δ et ΔΔ

Dans cette expérience, les coefficients dynamiques **Δ** et **ΔΔ** ont été successivement associés à la paramétrisation 24MFC. Le tableau 8 présente les résultats.

	Grade 0	Grade 1	Grade 2	Grade 3	Total
Expérience	% succès (nb sur 20)	% succès (nb sur 20)	% succès (nb sur 20)	% succès (nb sur 20)	% succès (nb sur 80)
24MFC + Δ	**95,0 (19)**	60,0 (12)	65,0 (13)	80,0 (16)	75,00 (60)
24MFC + Δ + ΔΔ	**95,0 (19)**	**65,0 (13)**	**70,0 (14)**	**85,0 (17)**	**78,75 (63)**

Tableau 8
*Résultat de la classification 4-Grades - 24MFC + **Δ** + **ΔΔ***

Le score global ne s'est amélioré que d'une seule voix (75,00 %) avec l'ajout des coefficients dynamiques **Δ** où seul le grade 2 a vu ses performances baisser de 75,0 % à 65,0 %. L'ajout des coefficients **ΔΔ** a un effet légèrement positif sur l'ensemble des voix dysphoniques (gain d'une voix pour G1/G2/G3), permettant au score global de progresser de 75,00 % à 78,75 %. On peut en déduire que l'ajout des coefficients dynamiques **Δ** et **ΔΔ** est pertinent pour cette tâche.

1. le grade 0 est le mieux reconnu (95,0 %) ;
2. les grades 1 et 2 ont respectivement 65,0 % et 70,0 % en score ;
3. le grade 3 obtient un score très intéressant de 85,0 %.

Classification	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**
Locuteurs G0	**19**		1	
Locuteurs G1	1	**13**	6	
Locuteurs G2		5	**14**	1
Locuteurs G3			3	**17**

Tableau 9
*Matrice de confusion de la classification 4-Grades - 24MFC + **Δ** + **ΔΔ***

La matrice de confusion (tableau 9) montre que les voix de grade G1/G2 mal classées ont tout de même été évaluées comme dysphoniques par le système, avec 6 voix G1 classées en G2 et 5 voix G2 classées en G1. Une analyse des résultats faite par une orthophoniste a montré que des facteurs étrangers à la dysphonie tels que les accents régionaux assez prononcés ou les rhinolalies, peuvent influencer le système. De plus, il semble que le système ne prenne pas suffisamment en compte les attaques/finales vocales qui sont des phénomènes pouvant être caractéristiques de dysphonie. Il en est de même pour les désonorisations, les chuchotements, les râlements légers… La visualisation de la classification segmentale sous l'outil Transcriber (Barras), a permis de constater que sur certaines voix, des portions de désonorisations ou de phénomènes de réductions importants comme l'absence de réalisation d'une voyelle se réduisant à un souffle, ont été faussement considérées comme du silence par le système.
Il est à noter que la classification locale des différents segments en grade permet de localiser des portions de parole très intéressantes pour les praticiens spécialistes de la voix.

4.4. Conclusion et perspectives sur l'évaluation vocale par des techniques issues de la reconnaissance automatique du locuteur

Nous avons présenté un système de RAL adapté à la classification de voix dysphoniques suivant le grade général de dysphonie de l'échelle GRBAS. Les résultats de l'approche statistique ont été comparés à ceux de l'analyse perceptive qui est la référence dans ce domaine.
Dans un premier temps, une classification basique *Contrôle/Patho* a permis de constater que le système répond plutôt favorablement à la caractérisation des voix normales et dysphoniques et plus particulièrement sur les grades 0 et 3.
La deuxième phase expérimentale a permis d'évaluer différentes paramétrisations sur une classification 4-Grades. La paramétrisation 24MFC + Δ + $\Delta\Delta$ a obtenu le meilleur score de 78,75 % dépassant le score du dispositif EVA à ses débuts.
Il est à noter que le manque manifeste de données (80 voix) peut influencer fortement la qualité des modèles des différents grades et par conséquent les résultats obtenus. Toutefois, grâce aux différents protocoles mis en œuvre (technique de leave_x_out), l'ensemble est plutôt prometteur (même si les résultats de l'analyse instrumentale affichent un score de 88 %) et nous permet d'envisager pour l'avenir différentes perspectives de recherche afin d'améliorer les résultats obtenus lors de cette étude.

4.4.1. Validation des résultats sur un plus grand nombre de voix dysphoniques

Nous travaillons actuellement sur la mise à disposition d'un corpus important de locuteurs dysphoniques enregistrés entre 1995 et 2003 dans le service ORL du CHU Timone à Marseille (Ghio *et al.,* 2007). Le résultat final de ce traitement de données permettra à terme de disposer d'un corpus d'environ 800 locuteurs, taille dix fois supérieure au corpus actuel. Il faut signaler que ce corpus pourrait devenir le corpus le plus important au niveau international, la référence actuelle dans le domaine étant celui du *Massachussetts Eye and Ear Infirmary Voice and Speech Lab* avec 710 sujets.

4.4.2. Les informations segmentales

L'exploitation des informations segmentales est aussi un axe important de recherche. En effet, une voix dysphonique n'est pas forcément un phénomène constant dans le discours, des caractéristiques spécifiques de dysphonie peuvent n'apparaître qu'à certains moments, portions de discours qui attirent l'oreille des spécialistes de la voix. Détecter des portions de parole caractéristiques de dysphonie faciliterait la décision à prendre. Fonder la décision sur une moyenne de vraisemblances semble ne pas être l'approche la plus pertinente dans le cas de voix dysphoniques. L'exploitation des informations segmentales pourrait permettre de redéfinir une décision plus appropriée à notre sujet.

De même, la nature des phonèmes peut influencer l'émergence du dysfonctionnement vocal et diverses études ont été menées sur l'importance du type d'information appropriée pour une tâche de classification automatique de voix produite par des patients atteints de dysfonctionnement vocal (Pouchoulin *et al.,* 2006).

4.4.3. Le travail en sous-bandes de fréquences

Comme la dysphonie concerne essentiellement la source vocale, la plupart des études se sont concentrées sur des paramètres directement liés à ce vibrateur. 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, ce qui a donné lieu à une analyse spectrale à long terme, aboutissant à différentes classifications de voix pathologiques. Nous avons mené diverses études sur les caractéristiques de la dysphonie dans le domaine fréquentiel à travers une analyse par sous-bandes de fréquences. Dans ce contexte, nous avons 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 (Pouchoulin *et al.,* 2007, a et b).

4.4.4. Apprentissage supervisé *vs* Automatique, évaluations complémentaires

Actuellement, notre méthodologie est fondée sur un apprentissage supervisé qui part de l'hypothèse que le jugement perceptif est la référence indiscutable permettant de constituer les groupes de locuteurs sur lesquels vont être appris les modèles de grade de dysphonie. Toute imprécision de la classification préalable, c'est-à-dire toute inclusion inappropriée de locuteur dans un groupe (ex : G2 dans G1…) entraînera une imperfection dans le système d'identification automatique en phase de test. Connaissant les imperfections de l'analyse perceptive *(cf.* §2), il est probable que les limites actuelles de performance de notre système sont en partie liées à cet apprentissage supervisé imparfait.

Une première solution pour sortir de cette impasse serait d'utiliser des techniques de classification automatique non supervisée, nous affranchissant ainsi d'une éventuelle précatégorisation floue. La seconde serait de confronter à la fois analyse perceptive, analyse instrumentale avec EVA et identification automatique et de mesurer la convergence ou divergence des résultats et ainsi, de n'être pas exclusivement dépendant d'une référence imparfaite.

5. Conclusions et perspectives générales

5.1. L'évaluation instrumentale analytique (dispositif EVA)

Même si sa définition peut être sujette à controverse (Le Huche *et al.,* 1997), la dysphonie est communément décrite comme une perturbation de l'émission du son laryngé. Dans cette étude, nous nous sommes placés volontairement dans le cadre de cette définition « acoustique » restreinte. Nous avons vu *(cf.* § 3.6.2.2) que l'évaluation instrumentale de la dysphonie était clairement tournée vers une détermination des caractéristiques « mécaniques » du système phonatoire. En revanche, le jugement perceptif ou les techniques d'évaluation automatique précédemment décrites se situent plus sur le plan de l'utilisation de l'instrument phonatoire, notamment en ayant pour support de la parole continue. Cette différence peut expliquer la non-concordance parfaite autour de 80 % entre ces modes d'évaluation *(cf.* § 3.6). Nous pensons notamment à des cas de « difficultés vocales sans traduction acoustique » (Le Huche *et al.,* 1997) où les instruments de mesure aérodynamique, comme celle de la pression sous-glottique estimée, peuvent mettre en évidence ces difficultés (par exemple liées à un forçage) qui ne se traduit pas nécessairement au niveau acoustique, ce phénomène passant ainsi inaperçu chez l'auditeur. Cet exemple montre aussi l'intérêt des mesures aérodynamiques qui, en plus de cette capacité à détecter certains dysfonctionnements non audibles, possèdent une sélectivité évidente (le débit d'air oral sur un /a/ tenu traduit directement une fuite glottique) et une robustesse reconnue,

contrairement à certains indices portant sur le calcul de la fréquence fondamentale, particulièrement difficile à détecter sur des voix très dégradées.

Par rapport à la concordance entre le jugement perceptif et les méthodes instrumentales, du fait des différences importantes entre ces approches *(cf.* § 3.6.3) et les échantillons vocaux utilisés *(cf.* § 3.6.2.2 *vs* parole continue), on peut même être surpris d'obtenir 80 % de concordances dans un ensemble à 4 grades. Si un tel seuil est atteint, cela provient du fait que les dysfonctionnements laryngés sont globalement captés par les mesures instrumentales que nous proposons *(cf.* § 3.4.2) : défaut de fréquence de vibration (jitter, Lyapounov), défaut d'amplitude de vibration, c'est-à-dire d'accolement (débit d'air oral, rapport signal/bruit), réduction de l'espace de fonctionnement en fréquence (étendue vocale), réduction de l'espace de fonctionnement au niveau temporel (temps maximal de phonation), tension inappropriée (pression sous-glottique). Un phénomène important n'est pour le moment pas pris en compte : les aspects transitionnels (phase d'initiation/extinction de la vibration). Or, ces dysfonctionnements, potentiellement captés par l'évaluation instrumentale analytique, vont apparaître en cours de production de parole continue et seront ainsi potentiellement détectés par le système perceptif de l'auditeur, ce qui conduit à une concordance, certes imparfaite, mais finalement assez fiable. Pour illustrer ce phénomène, nous pourrions faire une analogie avec un autre instrument sonore comme le piano. Imaginons un professionnel connaissant bien la mécanique de cet instrument et étant capable de mesurer et ainsi d'évaluer la plupart des caractéristiques mécaniques susceptibles d'entraîner un dysfonctionnement : désaccord le plus souvent des notes très graves et très aiguës, usure du feutre des marteaux des notes centrales, mauvais équilibre de touches, pédale « sourdine » mal ajustée positionnant de façon excessive le bandeau de feutre atténuant la percussion des marteaux sur les cordes… Une telle connaissance et des mesures bien adaptées pourraient permettre, dans une certaine mesure, de prédire le résultat musical d'un tel instrument, évaluation qui pourrait concorder avec le jugement d'un jury écoutant un pianiste jouant de cet instrument. Certes la dextérité du musicien, les techniques de compensation et tout simplement le style musical pourraient relativiser cette concordance, mais on peut penser que, même un bon pianiste jouant sur un instrument désaccordé, aux touches mal équilibrées, à la sourdine actionnée de façon permanente, ne pourra fournir qu'une prestation médiocre.

5.2. La reconnaissance automatique

En ce qui concerne les techniques de reconnaissance automatique adaptée aux voix dysphoniques (§ 4), nous manquons actuellement de recul du fait de la nouveauté de cette démarche (Fredouille *et al.,* 2005). Toutefois, nous pouvons penser que, par principe, du fait même de la phase

d'apprentissage *(cf.* § 4.1.2.2.) supervisée par la catégorisation obtenue par évaluation perceptive sur le même support langagier, il est probable qu'à terme cette méthode permette d'aboutir à un degré de concordance proche de 95 % avec l'évaluation perceptive d'un jury. L'intérêt résiderait alors en la non nécessité de réunir un jury d'experts et de multiplier les écoutes, configuration difficile à obtenir en pratique. L'autre avantage est l'aspect déterministe de la méthode, écartant ainsi toute inconstance que l'on peut observer chez un auditeur. Cependant, d'autres inconstances pourraient aussi apparaître comme la dépendance bien connue aux conditions d'enregistrement (changement de microphone, de chaîne d'acquisition, d'environnement sonore…). Toutes ces hypothèses restent à vérifier.

5.3. L'évaluation perceptive

S'il reste encore du chemin à parcourir pour perfectionner les méthodes instrumentales analytiques ou les techniques d'identification automatique, l'évaluation perceptive reste la clé de voûte de l'activité. Comme cela a été mentionné en introduction, ce processus est essentiel dans l'évaluation des dysphonies et restera probablement le standard de référence, qualité contestable à l'heure actuelle. Quelles sont les pistes possibles pour renforcer le pouvoir d'analyse de l'évaluation perceptive ?

La catégorisation des locuteurs dysphoniques en 4 grades de sévérité est, à l'origine, une simplification qui répond, partiellement, aux besoins cliniques d'évaluation globale et rapide. Pour une évaluation plus fine, ce classement à 4 niveaux peut s'avérer simpliste, voire problématique. En effet, au sein d'une même classe, par exemple les grades 2, vont être regroupées à la fois des voix proches de 2 (ex : 2.1 analogique) et proches de 3 (ex : 2.9). Or, une voix dont la valeur analogique serait de 2.1, classée grade 2, est finalement plus proche d'une voix de niveau analogique 1.9, classée grade 1 qu'une voix analogiquement évaluée à 2.9 et classée grade 2. Ce problème de frontière catégorielle a d'ailleurs donné lieu à l'utilisation de grades intermédiaires pour obtenir une meilleure granularité. Ainsi, à la place de 4 groupes, 7 catégories peuvent être définies : G0, G0.5, G1, G1.5, G2, G2.5, G3 (Révis, 2004). Finalement, la notion de voix proches ou éloignées fait référence à une notion de distance, qui peut être explorée par des échelles analogiques visuelles. Ce type d'évaluation comporte lui aussi ses inconvénients comme, par exemple, les différences de dynamique de jugement entre auditeurs, certains utilisant toute l'échelle, d'autres une partie restreinte de l'échelle de notation (Révis, 2004). Toutefois, il paraît évident que la notion de dysphonie, en plus de ses aspects multidimensionnels, apparaît avant tout comme une manifestation graduelle de type analogique, plutôt que comme un phénomène catégoriel comme le laisserait entrevoir la notion de Grade 0, 1, 2, 3. Il est très probable qu'une

part de la non-concordance mesurée entre l'évaluation perceptive et les évaluations instrumentales analytiques ou par identification automatique est liée à la non-concordance des frontières intercatégorielles, décalage probablement faible d'un point de vue analogique, mais dont la discrétisation excessive fait basculer brutalement des échantillons d'une catégorie à l'autre sans que la distance réelle justifie un tel changement d'état. Ainsi, il est probable qu'une part de la dispersion observée sur les mesures instrumentales (voir barres de dispersion, figure 2) est due, en fait, à des locuteurs situés aux frontières de grades et qui se retrouvent distribués sur deux grades adjacents, ce qui entraîne une dispersion des mesures et un recouvrement des distributions, altérant ainsi le pouvoir discriminant de la mesure.

Pour conclure, nous discuterons de la relation entre subjectivité et perception auditive. L'aspect perceptif de « l'évaluation perceptive des dysphonies » n'est pas réellement la source intrinsèque de la subjectivité. En effet, la perception n'est pas nécessairement subjective comme le montrent les tests d'intelligibilité, qui ont prouvé depuis longtemps leur efficacité dans des tâches de discrimination ou d'identification. Pour que la perception ne soit pas subjective, il faut des instructions clairement définies et des références partagées implicites ou explicites chez les auditeurs. Ce sont ces dernières qui sont absentes, actuellement, dans l'évaluation perceptive des dysphonies, ce qui conduit à des références mal définies, qui, par conséquent, rendent flous les résultats obtenus par des approches instrumentales ou encore dans des méthodes fondées sur de la modélisation statistique. Une des perspectives pour l'évaluation perceptive des dysphonies serait d'utiliser des méthodes telle que la *« Sentence Verification Task »* (Pisoni, 1986) qui est fondée sur le constat que lorsque les auditeurs doivent comprendre le contenu linguistique d'un message et exécuter une réponse appropriée, la qualité de l'information acoustico-phonétique du signal de parole, et donc la qualité vocale, joue un rôle important à la fois dans la vitesse et la justesse de la réponse fournie. La compréhension d'un message étant une capacité partagée par tous les auditeurs d'une même langue, cela permettrait de s'affranchir des aspects subjectifs de l'évaluation explicite de la qualité vocale.

5.4. Les trois approches conjuguées et complémentaires

Du fait de la lourdeur expérimentale et de la nécessité de manipuler de vastes corpus, nous n'avons pour le moment pas étudié de façon conjuguée les résultats fournis par les trois méthodes précédemment décrites, c'est-à-dire le jugement perceptif *(cf.* § 3.3.2, § 4.2.4), l'évaluation instrumentale avec le dispositif EVA *(cf.* § 3.3.3) et la reconnaissance automatique *(cf.* § 4.2.3). Une analyse conjuguée des sorties des trois approches permettrait, notamment, d'en savoir plus sur les

cas non concordants deux à deux (EVA *vs* GRBAS, système automatique *vs* GRBAS). On pourrait envisager, par exemple, une technique de vote qui fournirait comme « classification correcte » celle qui serait fournie par deux sorties sur trois. Cela permettrait, notamment, de « s'affranchir » du *Gold Standard* perceptif, régulièrement contesté en tant que tel.

Dans tous les cas, nous pensons que, bien plus que de se substituer l'une à l'autre, toutes ces méthodes restent avant tout complémentaires et apportent chacune un angle d'observation différent du même phénomène, au même titre que les autres types d'évaluation purement cliniques *(cf.* § 1).

6. Bibliographie

ANDERS, L.C. ; HOLLIEN, H. ; HURME, P. ; SONNINEN, A. & WENDLER, J. (1998) Perception of hoarseness by several classes of listeners, *Folia Phoniatrica,* 40, p. 91-100.

AUDIBERT, N. ; ROSSATO, S. & AUBERGÉ, V. (2004) Paramétrisation de la qualité de voix : EGG *vs* filtrage inverse, *Actes XXVèmes Journées d'Etude sur la Parole,* Fès, Maroc, p. 53-56.

BARRAS, C. Transcriber: a tool for segmenting, labeling and transcribing speech, http://www.etca.fr/CTA/gip/Projets/Transcriber/.

BIMBOT, F. ; BONASTRE, J.-F. ; FREDOUILLE, C. ; GRAVIER, G. ; MAGRIN-CHAGNOLLEAU, I. ; MEIGNIER, S. ; MERLIN, T. ; ORTEGA-GARCIA, J. & REYNOLDS, D.A. (2004) A tutorial on text-independent speaker verification, *EURASIP Journal on Applied Signal Processing,* 4, p. 430-451.

BONASTRE, J.-F. ; WILS, F. ; MEIGNIER, S. (2005) ALIZE, a free toolkit for speaker recognition, *Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2005)*, Philadelphia, USA, March.

BONASTRE, J.-F. ; FREDOUILLE, C. ; GHIO, A. ; GIOVANNI, A. ; POUCHOULIN, G. ; RÉVIS, J. ; TESTON, B. & YU, P. (2007) Complementary approaches for voice disorder assessment, *Proceedings Interspeech* (8 : août 27-31 : Antwerp, Belgium), Antwerp: ISCA, p. 1194-1197.

CAMPBELL, N. (2000) Databases of Emotional Speech, *Proceedings of ISCA 2000, Northern Ireland*, p. 34-38.

CHARLET, D. (1997) *Authentification vocale par téléphone en mode dépendant du texte,* Thèse de doctorat, ENST Paris.

CREVIER-BUCHMAN, L. ; MONFRAIS-PFAUWADEL, M.-C. ; LACCOURREYE, O. ; JOUFFRE, V. ; BRASNU, D. & LACCOURREYE, H. (1993) La Laryngostroboscopie, *Ann. Otolaryngol. Chir. Cervicofac.,* 110, p. 355-357.

DEJONCKERE, P. ; BRADLEY, P. ; CLEMENTE, P. ; CORNUT, G. ; CREVIER-BUCHMAN, L. ; FRIEDRICH, G. ; VAN DE HEYNING, P. ; REMACLE, M. & WOISARD, V. (2001) A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques, *Eur. Arch. Otorhinolar.,* 258, p. 77-82.

DEMPSTER, A.P. ; LAIRD, N. & RUBIN, D.B. (1977) Maximum-likelihood from incomplete data via the EM algorithm, *Journal of Acoustical Society of America*, 39, p. 1-38.

FERRAGNE, E. & PELLEGRINO, F. (2006) Les systèmes vocaliques des dialectes de l'anglais britannique, *Actes XXVIème Journées d'études sur la Parole,* Dinard, 12-16 juin, p. 411-414.

FLETCHER, H.F & MUNSON, W.A. (1933) Loudness, its definition, measurement and calculation, *Journal of Acoustical Society of America*, 5, p. 82-108.

FOURCIN, A. ; MCGLASHAN, J. & BLOWES, R. (2002) Measuring voice in the clinic - Laryngograph® Speech Studio analyses, *Proceedings 6th Voice Symposium of Australia,* Adelaide, Oct.

FREDOUILLE, C. ; POUCHOULIN, G. ; BONASTRE, J.-F. ; AZZARELLO, M. ; GIOVANNI, A. & GHIO, A. (2005) Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia), *Proc. Eurospeech*, Lisboa, ISCA, p. 149-152.

GAUVAIN, J.-L. & LEE, C.H. (1994) Maximum a Posteriori estimation for multivariate Gaussian mixture observations of Markov chains, April: 22, *Speech and Audio Processing, IEEE Transactions,* vol. 2, issue 2, p. 291-298.

GHIO, A. & TESTON, B. (2004) Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies, *Proceedings of International Conference on Voice Physiology and Biomechanics* (août 18-20 : Marseille, France), Marseille : Univ. Méditerranée, p. 55-58.

GHIO, A. (2007) Evaluation acoustique, in Auzou, P. ; Rolland, V. ; Pinto, S. & Ozsancak, C. (éds) *Les dysarthries,* ISBN 978-2-35327-021-7, Marseille : Solal, p. 236-247.

GHIO, A. ; TESTON, B. ; VIALLET, F. ; JANKOWSKI, L. ; PURSON, A. ; DUEZ, D. ; LOCCO, J. ; LEGOU, T. ; PINTO, S. ; MARCHAL, A. ; GIOVANNI, A. ; ROBERT, D. ; REVIS, J. ; FREDOUILLE, C. ; BONASTRE, J.-F. & POUCHOULIN, G. (2007) Corpus de parole pathologique, état d'avancement et enjeux méthodologiques au LPL, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA),* vol. 25, p. 109-126.

GIOVANNI, A. ; MOLINES, V. ; NGUYEN, N. & TESTON, B. (1991) L'évaluation objective de la dysphonie : une méthode multiparamétrique, *Proceedings of International Congress of Phonetic Sciences (ICPhS)* (12 : août 19-24 : Aix-en-Provence, France), p. 274-277.

GIOVANNI, A. ; MOLINES, V. ; N'GUYEN, N. ; TESTON, B. ; ROBERT, D. ; CANNONI, M. & PECH A. (1992) Une étude multiparamétrique d'évaluation vocale objective assistée par ordinateur, *Ann. Otolaryngol. Chir. Cervicofac.,* 109, p. 200-206.

GIOVANNI, A. ; TESTON, B. ; ROBERT, D. ; MOLINES, V. ; GALINDO, B. ; ZANARET, M. & CANNONI, M. (1993) Analyse multiparamétrique des dysphonies par l'appareillage Physiologia, *Revue de Laryngol.,* 114, p. 305-309.

GIOVANNI, A. ; ESTUBLIER, N. ; ROBERT, D. ; TESTON B. ; ZANARET M. & CANNONI, M. (1995) Evaluation vocale objective des dysphonies par la mesure simultanée des paramètres acoustiques et aérodynamiques à l'aide de l'appareillage EVA, *Ann. Otolaryngol. Chir. Cervicofac.,* 112, p. 85-90.

GIOVANNI, A. ; ROBERT, D. ; TESTON, B. ; GUARELLA, M.D. & ZANARET, M. (1996a) Etude préliminaire des paramètres acoustiques et aérodynamiques après laryngectomies frontales antérieures de Tucker, *Ann. Otolaryngol. Chir. Cervicofac.,* 113, p. 277-284.

GIOVANNI, A. ; ROBERT, D. ; ESTUBLIER, N. ; TESTON, B. ; ZANARET, M. & CANNONI, M. (1996b) Objective evaluation of dysphonia: preliminary results of a device allowing simultaneaous acoustic and aerodynamic measurements , *Folia Phoniatr. Logop.,* 48, p. 175-185.

GIOVANNI, A. ; OUAKNINE, M. & TRIGLIA, J.-M. (1999) Determination of the largest Lyapunov exponents of vocal signal. Applications to unilateral laryngeal paralysis, *Journal of Voice,* 13, p. 341-354.

GIOVANNI, A., ASSAIANTE, Ch. ; GALMICHE, A. ; VAUGOYEAU, M. ; OUAKNINE, M. & LE HUCHE, F. (2006) Forçage vocal et posture : études expérimentales chez le sujet sain, *Revue de laryngologie, d'otologie et de rhinologie,* vol. 127, nº 5, p. 285-291.

GOBL, C. & NI CHASAIDE, A. (2003) The role of voice quality in communicating emotion, mood and attitude, *Speech Communication,* 40, 1-2 (special issue), p. 189-212.

HAMMARBERG, B. ; FRITZELL, B. ; GAUFFIN, J. ; SUNDBERG, J. & WEDIN, L. (1980) Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngol.,* Nov-Dec, 90(5-6), p. 441-451.

HIRANO, M. (1981) *Clinical Examination of Voice,* Wien: Springer Verlag.

HIRAOKA, N. ; KITAZOE, Y. & UETA, H. (1984) Harmonic-intensity analysis of normal and hoarse voices, *Journal of Acoustical Society of America,* 76, p. 1648-1651.

HOMAYOUNPOUR, M.M. & CHOLLET, G. (1994) Performance comparison of some relevant spectral representations for speaker verification, *Workshop on Automatic Speaker Recognition, Identification, Verification,* Martigny, Suisse, p. 27-30.

KITANTOU, M. (1987) *La perception auditive. Le livre des techniques du son, vol. 1 : Notions fondamentales,* Paris : Eyrolles, éditions Fréquences, p. 155-181.

KREIMAN, J. ; GERRAT, B. ; KEMPSTER, G. ; ERMAN, A. & BERKE, G. (1993) Perceptual evaluation of voice quality: review, tutorial, and a framework for future research, *J. Speech Hear. Res.,* 36, p. 21-40.

LAMEL, L. ; GAUVAIN, J. & ESKÉNAZI, L. (1991) BREF, a large vocabulary spoken corpus for French, *Proceedings Eurospeech,* Genova, 24-26 September, vol. 2, p. 505-508.

LAMEL, L.F. & GAUVAIN, J.-L. (1994) Language Identification Using Phone-based Acoustic Likelihoods, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing,* Adelaide, Australia, April, p. 293-296.

LAVER, J. (1981) The Analysis of Vocal Quality: From the Classical Period to the Twentieth Century, part 3: History of Ideas in Phonetics, Voice Quality and Voice Dynamics, in Asher, R.E. & Henderson, E.J.A. (eds) *Towards A History Of Phonetics. Papers Contributed In Honour Of David Abercrombie,* p. 79-99.

LE HUCHE, F. & ALLALI, A. (1997) *La voix,* collection Phoniatrie, Paris : Masson.

MC GURK, H. & MC DONALD, J. (1976) Hearing lips and seeing voices, *Nature*, 264, n°5588, p. 746-748.

PARSA, V. & JAMIESON, D.G. (2001) Acoustic discrimination of pathological voice: sustained vowels versus continuous speech, *J. Speech Hear. Res.,* 44, p. 327-339.

PICCIRELLO, J. ; COLIN, P. ; DENNIS, F. & FREDERICKSON, J. (1998a) Multivariate analysis of objective vocal function, *Ann. Otol. Rhinol. Laryngol.,* 107, p. 107-112.

PICCIRILLO, J. ; COLIN, P. & DENNIS, F. (1998b) Assessment of two objective voice function indices, *Ann. Otol. Rhinol. Laryngol.*, 107, p. 396-400.

PISONI, D.B. & DEDINA, M.J. (1986) Comprehension of Digitally Encoded Natural Speech Using a Sentence Verification Task (SVT): A First Report, *Research on Speech Perception. Progress Report,* n° 12, Indiana University.

POUCHOULIN, G. ; FREDOUILLE, C. ; BONASTRE, J.-F. ; GHIO, A. & RÉVIS, J. (2007a) Characterization of the Pathological Voices (Dysphonia) in the frequency space, *Proceedings of International Congress of Phonetic Sciences (ICPhS)* (16 : 2007 août 6-10 : Saarbrücken, Germany), p. 1993-1996.

POUCHOULIN, G. ; FREDOUILLE, C. ; BONASTRE, J.-F. ; GHIO, A. & GIOVANNI, A. (2007b) Frequency Study for the Characterization of the Dysphonic Voices, *Proceedings of Interspeech* (8 : août 27-31 : Antwerp, Belgium), Antwerp: ISCA, p. 1198-1201.

POUCHOULIN, G. ; FREDOUILLE, C. ; BONASTRE, J.-F. ; GHIO, A. ; AZZARELLO, M. & GIOVANNI, A. (2006) Modélisation statistique et informations pertinentes pour la caractérisation des voix pathologiques (dysphonies), *Actes Journées d'Etude sur la Parole (JEP)* (26 : juin 12-16, Dinard, France), Rennes : Irisa, Afcp, Isca, p. 93-96.

REVIS, J. (2004) L'analyse perceptive des dysphonies, in A. Giovanni (éd.), *Le bilan d'une dysphonie : état actuel et perspectives,* Marseille : Solal, 244 p.

REYNOLDS, D.A. (1992) *A Gaussian mixture modeling approach to text-independent speaker identification,* PhD, Georgia Institute of Technology.

REYNOLDS, D.A. (1994) Experimental evaluation of features for robust speaker identification, *IEEE transactions Speech Audio Processing*, p. 639-643.

ROY, N ; BLESS, D.M. & HEISEY, D. (2000) Personality and voice disorders: a multitrait-multidisorder analysis, *Journal of Voice,* 14, p. 521-548.

SCHOENTGEN, J. (2003) Decomposition of Vocal Cycle Length Perturbations into Vocal Jitter and Vocal Microtremor, and Comparison of Their Size in Normophonic Speakers, *Journal of Voice,* 17, 2, p. 114-125.

SMITHERAN, J. & HIXON, T. (1981) A clinical method for estimating laryngeal airway resistance during vowel production, *J. Speech Hear. Dis.,* 46, p. 138-146.

TESTON, B. & GALINDO, B. (1995) A diagnosis of rehabilitation aid workstation for speech and voice pathologies, *Proc. European Conference on Speech Communication and Technology (Eurospeech)*, p. 1883-1886.

TESTON, B. (2004) L'évaluation instrumentale des dysphonies : État actuel et perspectives, in A. Giovanni (éd.), *Le bilan d'une dysphonie : état actuel et perspectives,* Marseille : Solal, 244 p.

WOISARD, V. ; BODIN, S. & PUECH, M. (2004) The Voice Handicap Index: impact of the translation in French on the validation, *Rev. Laryngol. Otol. Rhinol.,* 125(5), p. 307-312.

WOLFE, V. ; FITCH J. & CORNELL, R. (1995) Acoustic prediction of severity in commonly occurring voice problems, *J. Speech Hear. Res.,* 38, p. 273-279.

WUYTS, F.L. ; DE BODT, M.S. ; MOLENBERGHS, G. ; REMACLE, M. ; HEYLEN, L. , MILLET, B. ; LIERDE, K.V. ; RAES, J. & VAN DE HEYNING, P.H. (2000) The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approch, *J. Speech Hear. Res.,* 43, p. 796-809.

YU, P. (2001) *Méthodes instrumentales d'analyse de la dysphonie : corrélation avec l'analyse perceptive,* Thèse de l'Université de la Méditerranée, Marseille.

YU, P. ; OUAKNINE, M. ; REVIS, J. & GIOVANNI, A. (2001) Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements, *Journal of Voice,* 15 (4), p. 529-542.

YU, P. ; REVIS, J. ; WUYTS, F.L. , ZANARET, M. & GIOVANNI, A. (2002) Correlation of instrumental voice evaluation with perceptual analysis using a modified visual analogic scale, *Folia Phoniatr. Logop.,* 54, p. 271-281.

YU, P. ; GARREL, R. ; NICOLLAS, R. ; OUAKNINE, M. & GIOVANNI, A. (2007) Objective voice analysis in dysphonic patients. New data including non linear measurements, *Folia Phoniatr. Logop.,* 59, p. 20-30.