Bilan et perspectives de quinze ans d'évaluation vocale par méthodes instrumentales et perceptives

A. Ghio, A. Giovanni, B. Teston, J. Révis, P. Yu, M. Ouaknine, D. Robert, T. Legou

Laboratoire « Parole et Langage », Aix-Marseille Université, 29 Av. R. Schuman, F-13621 Aix-en-Provence, France
{ alain.ghio, antoine.giovanni, joana.revis...}@lpl-aix.fr

ABSTRACT

For fifteen years, we have developed and studied different techniques and methodologies to assess voice quality in a clinical context. This paper exposes recent results obtained by complementary approaches. 449 speakers (including 391 dysphonic patients) participated in the experiment where voice quality was evaluated by (1) perceptual voice assessment performed by a jury and (2) instrumental voice assessment using acoustic and aerodynamic data. Results showed that a combination of 7 instrumental measures allowed the classification of 82% voice samples in the same grade as the jury. We evaluate the methodological situation and we also discuss some theoretical aspects which are often forgotten in the performance race.

Keywords: voice quality, dysphonia, perceptual evaluation, instrumental voice assessment

1. INTRODUCTION

En phonétique, la qualité de la voix est généralement décrite comme un phénomène paralinguistique lié à l'état émotionnel du locuteur ou à des facteurs dialectaux et socioculturels. De notre côté, depuis une quinzaine d'années, nous nous sommes penchés 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 contrôler l'évolution temporelle de l'état vocal d'un patient, pour mesurer l'efficacité de différentes solutions thérapeutiques, pour du dépistage et enfin, pour mieux comprendre les mécanismes fondamentaux de la phonation. Il existe une grande variété de méthodes pour appréhender l'état vocal de personnes atteintes de troubles de la voix: interrogatoire avec le patient, examen endoscopique du larynx, questionnaire d'auto évaluation, appréciation du comportement postural, profil psychologique, jugement perceptif de la qualité vocale, analyse instrumentale... 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. Si certaines approches sont exclusivement cliniques, l'évaluation perceptive et les mesures instrumentales multiparamétriques relèvent de la phonétique. Nous ne présenterons que ces aspects.

2. INTERET ET LIMITES DES DIFFERENTES APPROCHES

Dans le cadre d'une évaluation de la qualité de la voix, l'appréciation perceptive reste essentielle. En effet, la plupart des patients dysphoniques se décide à consulter au moment où la personne ou son entourage entend des changements dans son résultat vocal. A 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 toutes ces raisons, le jugement perceptif demeure un procédé répandu en pratique clinique et fortement recommandé [1]. Pourtant, cette méthode reste controversée car sujette à diverses imperfections. Les phénomènes de variabilité intra-auditeur ou inter-auditeurs sont les éléments principaux. Pour obtenir une fiabilité raisonnable, l'évaluation doit être conduite par un jury d'experts. Ainsi, plusieurs auditeurs sont requis afin d'obtenir une appréciation moyenne ou consensuelle plus représentative de l'état vocal qu'un jugement isolé [2]. De même, il est préférable que soient sollicités des experts car le niveau d'expertise contribue à améliorer la fiabilité des réponses, ce que l'on peut interpréter par le fait que des auditeurs néophytes, non habitués à écouter des voix dégradées, laissent une part plus importante à la subjectivité, génératrice de variabilité. De ce fait, une analyse perceptive fiable impliquant plusieurs auditeurs experts et plusieurs sessions d'écoute s'avère finalement consommatrice en temps et en ressources humaines, ne permettant pas une utilisation régulière en pratique clinique. C'est pourquoi ont été proposées des approches instrumentales.

L'analyse instrumentale multiparamétrique est prévue pour quantifier les dysfonctionnements vocaux à partir de mesures acoustiques, aérodynamiques ou électrophysiologiques… Ces mesures sont réalisées sur le patient 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 [3]. En effet, une seule mesure isolée ne peut pas rendre compte à elle seule de dimensions différentes comme la raucité, le souffle ou le chevrotement. Nos premières études sur cette thématique remontent à [4] et depuis, nous n'avons cessé d'étudier de nouvelles méthodologies et équipements spécifiques pour permettre de disposer de méthodes instrumentales d'évaluation vocale pour une utilisation clinique ([5], [6], [7], [8]).

3. CORPUS, METHODES ET RESULTATS

3.1. Participants

449 locuteurs ont participé à l'étude dont 58 locuteurs de contrôle sans trouble vocal (38 femmes et 20 hommes) et 391 patients atteints de diverses pathologies vocales (308 femmes, 141 hommes). Ces patients du service ORL de l'hôpital de la Timone à Marseille 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).

3.2. Evaluation perceptive

Les instructions données aux locuteurs étaient de lire un texte standardisé (paragraphe de "la chèvre de M. Seguin", A. Daudet) à hauteur et intensité confortables. Tous les enregistrements étaient ensuite évalués en bloc 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 au jury d'auditeurs était de fournir un grade **G**lobal de dysphonie sur une échelle analogique visuelle. La position analogique était ensuite convertie en échelle discrète. L'originalité est 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 juge et en renforçant la concordance avec les mesures instrumentales.

3.3. Analyse instrumentale multiparamétrique

Les mesures instrumentales ont été réalisées par l'intermédiaire du dispositif EVA® (SQLab-LPL, France) [5]. Cet appareillage permet d'enregistrer simultanément des signaux acoustiques et aérodynamiques par l'utilisation d'une pièce à main contenant un microphone, un sonomètre, un pneumotachographe mesurant le débit d'air et deux capteurs de pressions [7]. Les instructions données aux sujets sont de produire 3 voyelles tenues /a/ sur lesquelles sont mesurés : la F0 (en Hz), l'intensité SPL (en dB), le jitter factor (en %), le shimmer (en %), le rapport signal/bruit, le débit d'air oral (en dm3/s) et le plus grand exposant de Lyapunov. La pression sous-glottique est estimée de façon indirecte par la "airway interrupted method" à l'aide d'un cathéter placé dans la cavité orale du locuteur qui produit alors des /papapa/. Enfin, l'étendue vocale est mesurée en recueillant la F0 la plus haute et la F0 la plus basse que le locuteur puisse 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 [8].

3.4. Résultats

Cohérence des mesures instrumentales

L'analyse perceptive a été considérée comme la référence pour constituer 4 groupes de locuteurs en fonction de leur niveau de dysphonie jugé auditivement : G0 = locuteurs normaux, G1 = dysphoniques légers, G2 = dysphoniques moyens, G3 = dysphoniques sévères. Une analyse discriminante a été utilisée pour évaluer la concordance entre les jugements perceptifs et les mesures instrumentales. 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, 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 et la F0. Pour les hommes, ont émergé l'étendue vocale, le coefficient de Lyapunov, le temps maximal de phonation, la pression sous-glottique estimée, la F0 et le signal ratio. La Figure1 fournit les moyennes et écart-type des mesures instrumentales décrites précédemment en fonction du grade de dysphonie (G0 à G3) et du genre (H/F).

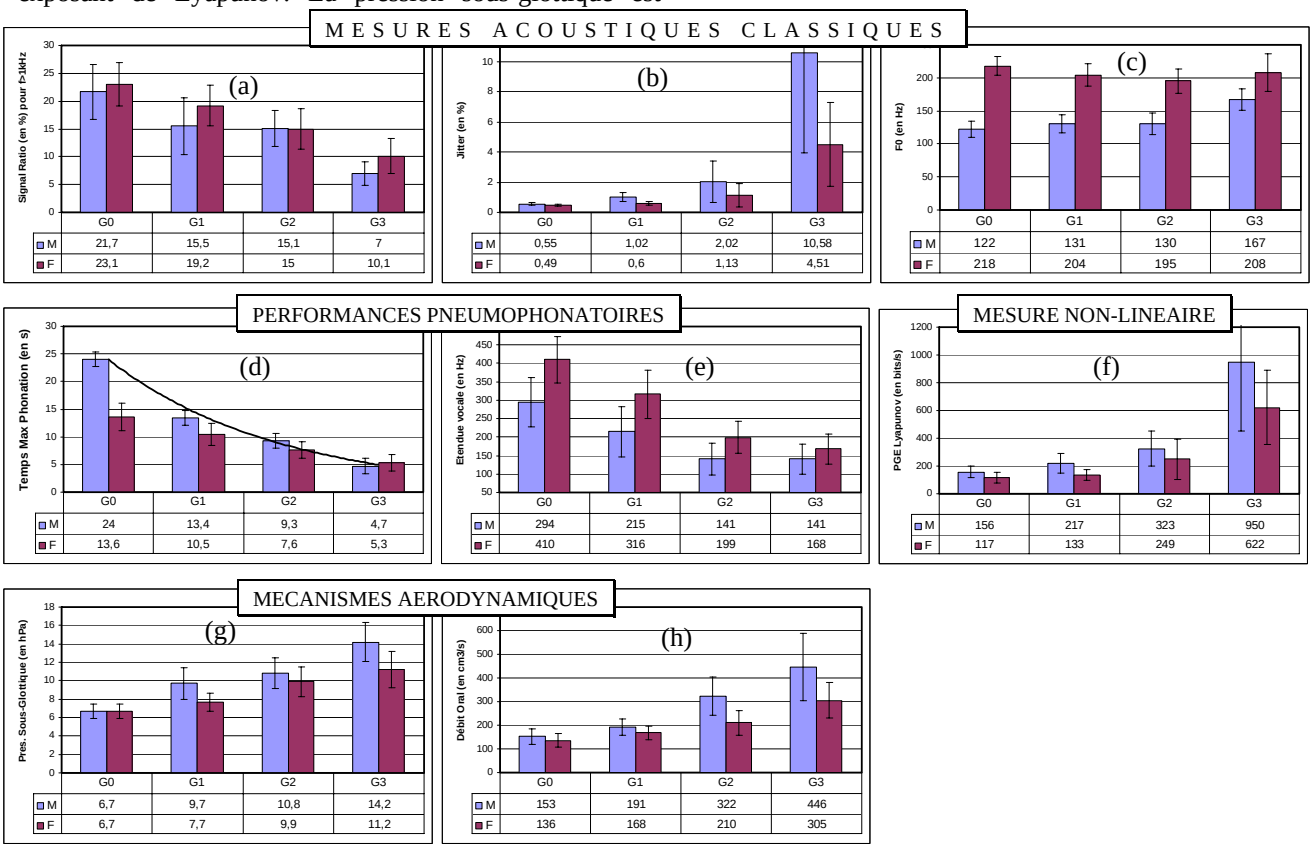

Figure 1 : Mesures instrumentales en fonction du grade de dysphonie (G0, G1, G2 , G3) et du genre (H,F)

Mise à part la mesure de la F0 moyenne (Fig1c), nous observons une évolution cohérente des mesures instrumentales. En effet, plus la dysphonie est jugée sévère

⊕ le temps maximal de phonation (TMP) diminue (Fig1d)
⊕ l'étendue vocale (VR) se réduit (Fig1e)
⊕ le débit d'air buccal (DAB) augmente, révélateur de fuite glottique (Fig1g)
⊕ la pression sous-glottique (PSGE) augmente, révélatrice de forçage (Fig1h)
⊕ le taux d'énergie harmonique (Sr) diminue, révélateur de pauvreté harmonique et présence de bruit (Fig1a)
⊕ le jitter augmente, révélateur d'instabilité laryngée (Fig1b)
⊕ le plus grand exposant de Lyapunov (PGEL) augmente, révélateur de comportement chaotique du vibrateur laryngé (Fig1f)

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 sous forme de performances pneumo-phonatoires
- 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

Vers un modèle prédictif de dysfonctionnement laryngé

Mise à part la mesure de la F0 moyenne (Fig1c), l'observation de la Figure 1 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. 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 (Fig1d), le jitter (Fig1b) ou le plus grand exposant de Lyapunov (Fig1f) 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 [6] où les auteurs avaient fait porté leur analyse statistique sur le Log(jitter). Ces deux remarques précédentes (distinction H/F, métrique non-linéaire) peuvent expliquer l'échec des modèles linéaires généraux comme le DSI de [3]. En effet, cet index est une simple combinaison linéaire $DSI = 0.13 \times TMP(s) + 0.0053 \times F0_{max}(Hz) - 0.26 \times I_{min}(dB) - 1.18 \times Jitter(\%) + 12.4$ qui d'une part, s'applique sans distinction pour les hommes et les femmes alors qu'à la vue 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 Fig1d) est $TMP = 40.524 \times e^{-0.5257(G+1)}$ avec un coefficient de détermination de R^2=0.988, ce qui fournit par transposition une nouvelle variable plus adaptée $G_{predictif} = \frac{3.7 - Ln(TMP)}{0.5257} - 1$

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. Pour évaluer les performances de cette fonction de classement, le principe 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 (Table 1) avec en ligne 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 fusionné les matrices pour n'en obtenir qu'une seule présentée en Table 1. Le résultat global montre que dans 82% des cas, le locuteur est classé de façon concordante entre le jugement perceptif et les mesures instrumentales.

Tab 1. : Matrice de confusion entre jugement perceptif et prédiction du grade de dysphonie à partir des mesures instrumentales

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

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 construire 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 [8]. La concordance a alors atteint 80.5% avec une majorité d'erreur où l'analyse instrumentale sous-estime le grade de dysphonie. Ce seuil récurrent autour de 80% de concordance entre jugement perceptif et mesures instrumentales apparaît comme une limite méthodologique qui nous amène à faire un bilan.

4. BILAN ET PERSPECTIVES

La limite évidente de notre méthodologie actuelle réside dans l'utilisation d'appareillages de mesure comme « machines à écouter ». Or, 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 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, une mesure adaptée à la question, une réponse précise.

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 mais globale, montante et aveugle.

Aveugle ou en tout cas opaque dans la mesure ou 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 fondés sur un recueil de nombreux paramètres avec pour objectif de faire émerger et mettre en évidence d'éventuels clusters.

Globale car telle est la démarche utilisée dans le jugement perceptif, notamment pour la détermination du grade G de l'échelle GRBAS d'Hirano [9]. D'un point de vue géométrique, B, R et A+S peuvent être assimilés à 3 axes d'un espace métrique (Fig. 2). 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.2, G') aura le même grade G=1 qu'une autre voix cotée R2;B0;S1 (Fig.2, 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.

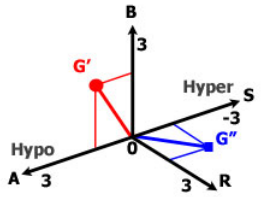

Figure 2: Espace perceptif GRBAS

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 concordance entre mesures instrumentales et perception.

Une autre différence fondamentale entre les deux approches réside dans le fait que l'évaluation instrumentale telle que décrite en § 3.3 est clairement tournée vers une détermination des caractéristiques "mécaniques" du **système** phonatoire alors que l'évaluation perceptive se situe plus sur le plan de **l'utilisation de l'instrument** phonatoire.

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 comme proposées par [10]. Une des perspectives pour l'évaluation perceptive des dysphonies serait d'utiliser des méthodes telle que la « Sentence Verification Task » [11] 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.

BIBLIOGRAPHIE

[1] P.Dejonckere, P.Bradley, P.Clemente, G.Cornut, L.Crevier-Buchman, G.Friedrich, P.Van de Heyning, M.Remacle, V Woisard. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques., *Eur Arch Otorhinola*, 258:77-82, 2001

[2] J.Kreiman, B.Gerrat, G.Kempster, A. Erman, G.Berke, Perceptual evaluation of voice quality : review, tutorial, and a framework for future research, *J. Speech Hear Res*, 36, 21-40, 1993

[3] F.Wuyts, M.De Bodt, G.Molenberghs, M. Remacle, L.Heylen, B.Millet, K.Lierde, J.Raes, P.VanDeHeyning, The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approch", *J Speech Hear Res*., 43:796-809, 2000

[4] A.Giovanni, V.Molines, N.Nguyen, B.Teston, L'évaluation objective de la dysphonie: une méthode multiparamétrique, *Proc. ICPhS*, 274-277, 1991

[5] B.Teston, B.Galindo, A diagnosis of rehabilitation aid workstation for speech and voice pathologies, *Proc. Eurospeech*, 1883–1886, 1995

[6] A.Giovanni, D.Robert., B.Teston., MD.Guarella, M.Zanaret, Etude préliminaire des paramètres acoustiques et aérodynamiques après laryngecto-mies frontales antérieures de Tucker", *Ann. Otolarl. Chir. Cerv*., 113, 277-284, 1996

[7] A.Ghio, B.Teston, Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. *Proc. Int. Conf. on Voice Physiol. & Biomechanics*, 55-58, 2004

[8] P.Yu, R.Garrel., R.Nicollas, M.Ouaknine, A. Giovanni, Objective voice analysis in dysphonic patients. New data including non linear measurements, *Folia Phon Log*, 59:20-30, 2007

[9] M.Hirano, *Clinical Examination of Voice*. Wien, Springer Verlag, 1981

[10]G.Pouchoulin, C.Fredouille, JF.Bonastre, A.Ghio, M.Azzarello, A.Giovanni. Modélisation statistique et infomations pertinentes pour la caractérisation des voix pathologiques (dysphonies), *JEP*, 93-96, 2006

[11]D.B.Pisoni, M.J. Dedina, Comprehension of Digitally Encoded Natural Speech Using a Sentence Verification Task (SVT): A First Report, In *Research on Speech Perception. Progress Report No. 12*, Indiana University, 1986