



HAL
open science

Bilan instrumental de la dysphonie

Alain Ghio

► **To cite this version:**

Alain Ghio. Bilan instrumental de la dysphonie. R. Garrel And B. Amy de la Bretèque And V. Brun. La voix parlée et la voix chantée, Sauramps Médical, pp.69-104, 2012, Echanges en réadaptation. hal-01482478

HAL Id: hal-01482478

<https://hal.science/hal-01482478v1>

Submitted on 12 Oct 2017

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

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

Bilan instrumental de la dysphonie

Alain Ghio

Laboratoire Parole et Langage, Université d'Aix-Marseille, Aix-en-Provence, France, alain.ghio@lpl-aix.fr

In R. Garrel; B. Amy de la Bretèque; V. Brun (eds.) La voix parlée et la voix chantée. (Echanges en réadaptation). Montpellier: Sauramps Médical. 2012, p. 69-104.

1. Quel bilan vocal ?

1.1. Imagerie, perception, instruments, questionnaires ?

Il existe une grande variété de méthodes pour établir un bilan vocal de personnes atteintes de troubles de la voix: examen endoscopique du larynx [5], jugement perceptif de la qualité vocale [46], analyse instrumentale [56], appréciation du comportement postural du patient [19], profil psychologique et étude comportementale [50], questionnaire d'auto évaluation [60]. 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. Cette approche multiple est d'ailleurs préconisée dans les recommandations de l'Union Européenne des Phoniatries [7].

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. 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 évaluer la qualité vocale. Pour toutes ces raisons (facilité de mise en œuvre, faible coût, directement accessible à tout clinicien), le jugement perceptif de la voix demeure le procédé le plus répandu en pratique clinique.

Pourtant, si l'analyse perceptive reste la référence en la matière, cette méthode soulève encore une question centrale: celle de la fiabilité. En effet, la littérature rapporte une importante variabilité dans les jugements perceptifs de la voix [3, 6, 34, 35, 47, 52, 63]. Cette variabilité se constate dans la possible inconstance d'un auditeur qui peut fournir des jugements différents entre plusieurs sessions d'écoute de la même voix (variabilité intra-auditeur). Elle est aussi observable dans le manque de cohérence des jugements à l'intérieur d'un groupe d'auditeurs (variabilité inter-auditeurs). Ces variabilités ont été largement étudiées dans le but de réduire ces phénomènes considérés comme indésirables en pratique clinique. Les auteurs ont notamment cherché différents points d'entrée pour en réduire son ampleur : recrutement d'auditeurs experts vs. naïfs [6], choix d'échelles analogiques vs. catégorielles [63], type d'énoncés (voyelle tenue vs. phrases dans Révis et al. 1999 [47]), dimensions perçues (qualité globale vs. dimension soufflée dans Shrivastav, 2006 [52])... Une explication partielle de ces phénomènes perceptifs indésirables en pratique clinique a été proposée dans Giovanni et al., 2009 [18]. Des compléments sont aussi apportés dans Ghio & Merienne, 2011 [15]. Dans la mesure où une fiabilité suffisante ne peut être obtenue que par le recrutement de plusieurs auditeurs experts, la nécessité de proposer plusieurs sessions d'écoute, la nécessité de mettre en place des séances d'apprentissage consensuel [16], l'évaluation perceptive s'avère finalement consommatrice en temps et en ressources humaines, ne permettant pas finalement une utilisation optimale en pratique clinique.

En complément de l'évaluation perceptive, l'analyse instrumentale 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 (électroglottographie¹, électromyographie² ...) 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

1 Technique permettant d'observer l'accolement des cordes vocales par mesure de conductivité électrique à travers les plis vocaux.

2 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.

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 [62].

La littérature abonde de comparaisons entre les méthodes perceptives et instrumentales. Baken (2000, p.146 [2]) affirme que les "*méthodes d'évaluation (perceptives) sont inadéquates... incapables de mettre en évidence de petites différences.*" De même, Morsomme et Estienne [40] confirment que si "*notre premier outil d'évaluation acoustique est l'oreille humaine...l'audition est insuffisante...(car) elle traite le signal acoustique dans sa globalité. L'oreille peut donc être dupée*". Enfin, Kent et al. [33] participent à ce tableau comparatif en signalant :

p1. "*Some would argue that voice quality can be described and assessed only by auditory-perceptual means... "*

p1. "*Still others might claim that auditory-perceptual methods are prone to various source of error... "*

p117. "*It is generally accepted in current research that the impressionistic nature of perceptual description should be augmented by more objective nature of instrumental analysis where possible.*"

p117. "*It is important however not to assume that instrumental approaches are somehow correct and infallible. Limits on accuracy exist with all techniques and the interpretation of results is not always straightforward*"

Pour résumer, bien que la perception soit considérée comme le meilleur moyen d'évaluer la qualité vocale, le moyen le plus répandue, le plus facile à mettre en œuvre, le plus économique, elle reste sujette à controverse. L'évaluation instrumentale s'avère finalement comme un complément plus objectif pour évaluer et mesurer les dysfonctionnements de la voix et de la parole. Cependant, ces techniques ne sont pas non plus infaillibles car, comme tout procédé de mesure, elles sont sujettes à des erreurs, à des problèmes de non reproductibilité, d'interprétation ou de généralisation.

1.2. Instruments ou perception : une vieille histoire

Ce débat entre mesure physique et perception de la parole, récurrent à la fois dans la pratique clinique ou dans des discussions académiques n'est pas sans rappeler un épisode fondateur de la phonétique expérimentale qu'il est intéressant de rappeler [56].

Etienne-Jules Marey est un physiologiste français du XIX siècle. Il a fourni une œuvre importante dans l'étude des mouvements des organismes vivants. On lui doit notamment la chronophotographie dont le principe est de photographier un être en mouvement sur de multiples poses, ce qui a permis pour la première fois de visualiser et décomposer les mécanismes de la locomotion naturelle. Pour capter le déroulement des phénomènes au cours du temps, il invente et perfectionne des méthodes graphiques d'enregistrement avec inscription au moyen de stylets, actionnés par des tambours manométriques, sur du papier noirici de fumée. Il applique ce principe à toutes sortes de capteurs pour « enregistrer » des phénomènes physiques ou physiologiques tels que le sphygmographe (battement de pouls), le pneumographe (rythme et l'amplitude des mouvements thoraciques au cours de la respiration), le myographe (contractions musculaires), le thermographe (température), l'odographe (distance parcourue et vitesse)...

En 1875, la Société de Linguistique de Paris, sous l'impulsion de son secrétaire Louis Havet, envoie une délégation pour consulter Marey dans le but « *d'appliquer la méthode graphique à l'étude des mouvements si complexes et si variés qui se produisent dans la parole [...] pour obtenir une trace objective des mouvements des organes articulateurs, cage thoracique, larynx, langue, lèvres, voile du palais, au cours de l'articulation des différentes unités phonétiques* ». Cette démarche est la conséquence d'un constat posé par certains linguistes de l'époque qui témoignent que « *l'oreille, la vue et la palpation ne sont pas suffisants pour appréhender les mouvements successifs et simultanés nécessaires à la réalisation d'un phonème* ». En effet, les linguistes cherchent à cette époque à vérifier leurs lois taxinomiques sur l'évolution des langues. Ils ont besoin de connaître avec précision les mécanismes de la formation des divers sons d'une langue pour en étudier les variations dialectales, sujets d'étude à la mode en cette fin du XIXe siècle. Comme ils se posent des questions d'ordre physiologique, ils se tournent naturellement vers Marey qui dispose d'un laboratoire et de tout un arsenal de capteurs nécessaires à l'étude de la nasalité, du voisement et de la labialité. Marey répond favorablement à cette démarche et écrit « *L'inscription des mouvements de la parole est une des applications les plus hardies de la méthode graphique. Quoi de plus complexe en effet que les mouvements des lèvres, de la langue, du voile du palais et du larynx d'une personne qui parle!* ».

De cette rencontre finissent par émerger divers instruments dont celui de Rosapelly équipé de trois tambours pneumatiques pour capter les vibrations du larynx, les mouvements des lèvres et ceux du voile du palais (Figure 1 gauche). Le dispositif le plus abouti reste celui mis au point par Rousselot (Figure 1 droite), qui va perdurer longtemps dans les études de phonétique articulatoire jusque dans les années 1970.

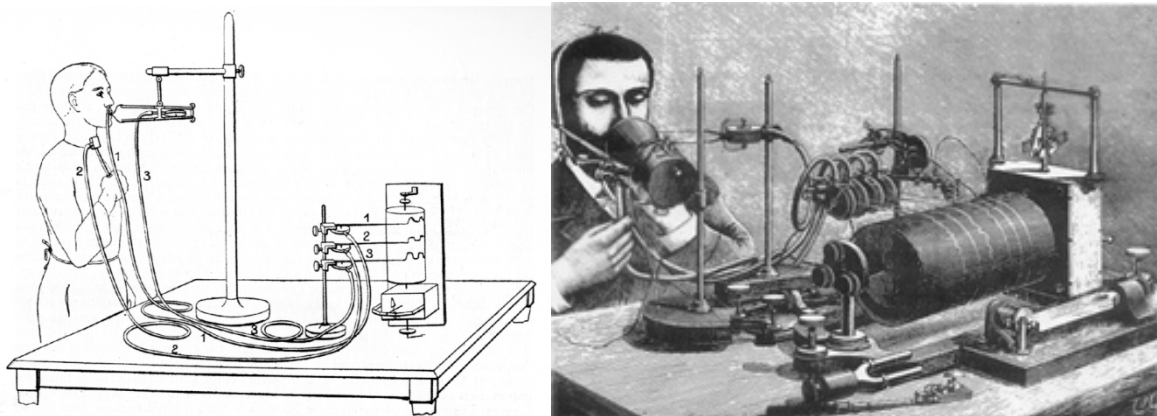


Figure 1 : Premiers dispositifs d'évaluation instrumentale de la parole.
 A gauche, montage expérimental de Rosapelly permettant l'enregistrement de la pression nasale, des vibrations laryngées et des mouvements des lèvres (1875).
 A droite, le kymographe de Rousselot (1892).

La phonétique expérimentale avec notamment la description des mécanismes de production de la parole est directement fondée sur les mesures obtenues avec ces dispositifs qui ont permis de s'extraire de spéculations théoriques. L'apparition, par la suite, de procédés permettant l'enregistrement et la restitution sonore puis de techniques d'analyse acoustique a définitivement ancré les trois composantes principales de la phonétique moderne : phonétique articulatoire, phonétique acoustique et phonétique perceptive. Par la suite, l'avancée des connaissances dans ces trois domaines s'est faite en parallèle, les progrès résidant dans la mise en relation des faits expérimentaux observés avec notamment la mise en relation entre phono-articulation et résultats acoustiques ainsi que la mise en relation entre acoustique et perception. Très vite, les phonéticiens ont donc intégré le fait que l'étude des mécanismes de production de la parole nécessite une approche instrumentale héritée de Marey, s'affranchissant de la composante perceptive appréhendée comme un autre phénomène. Utiliser la perception auditive pour observer la production de la parole d'un (inter)locuteur reste très contestable en phonétique sauf si l'objectif est d'étudier la perception en tant que telle ou l'intégration des deux mécanismes appréhendés dans un même phénomène de production/perception.

1.3. Les décalages entre mesures et perception : qui a raison ?

Dans le cadre de l'évaluation des troubles de la voix et de la parole, il est fréquent de recueillir le témoignage pointant le décalage entre ce qui est mesuré et ce qui est perçu : « je mesure des données qui ne correspondent pas à ce que j'entends » ou « j'entends des phénomènes que je ne mesure pas ». De nombreux résultats en phonétique ou en psycho-acoustique ont depuis longtemps pointé ce décalage. Divers exemples sont fournis dans Giovanni et al [18]. Pour n'en citer que certains, il est bien connu que la perception de la puissance sonore dépend à la fois de du logarithme de l'amplitude du stimulus sonore, de sa fréquence et de sa durée. Prenons le cas d'un observateur mal renseigné, convaincu que la perception du niveau ne dépend que de l'amplitude du signal acoustique. Imaginons que cet observateur écoute deux locuteurs, l'un étant perçu comme ayant un niveau sonore supérieur à l'autre. Considérons que pour le vérifier, il enregistre des échantillons sonores et qu'il découvre que les deux locuteurs produisent un signal acoustique de même amplitude. Fort de ses sensations, il en déduira que l'analyse acoustique n'est pas fiable. Et pourtant, il aura omis d'observer qu'un locuteur est un homme avec une fréquence fondamentale (F_0) de 80 Hz, perçue plus faiblement qu'une femme avec une F_0 de 300 Hz, les basses fréquences étant auditivement ressentie à un niveau moindre que des fréquences plus élevées.

De même, prenons un échantillon sonore réel produit par une locutrice française en parole spontanée. Dans cet extrait tiré de Meunier et al. [39], la majorité des auditeurs perçoivent sans problème le mot « spaghetti ». Pourtant, si une analyse spectrographique est effectuée (Figure 2 gauche), la mesure nous renseigne que le 1^{er} formant (F_1) de la voyelle /a/ de « spaghetti » est à 370 Hz et le 2^{ème} formant (F_2) est à 2000 Hz. Si l'on compare ces valeurs à des données normatives obtenues sur le français (Figure 2 droite), elles correspondent à la voyelle [e]. Une réaction sans précaution nous conduirait à affirmer que l'analyse spectrale est erronée et les valeurs de formants fausses car éloignés du /a/. Or, l'étude de Meunier et al. [39] confirme que l'écoute tronquée de la syllabe « spa » de « spaghetti » ou de la voyelle « a » du même mot est effectivement perçue comme un /e/, de façon encore plus nette en voyelle isolée. L'analyse instrumentale est donc correcte. Ce sont les processus cognitifs descendant avec notamment la connaissance du mot « spaghetti » dans le lexique qui masquent perceptivement les déformations acoustiques, elles-mêmes conséquentes d'ajustements articulatoires (le /a/ est produit en position plus fermé et plus antérieure). Ces

processus étant inhibés en écoute tronquée, l'auditeur perçoit la voyelle de façon différente, en accord avec les mesures observées. Pour information, cette prononciation du /a/ en [e] est le résultat d'un effet de coarticulation lié notamment à la présence du /e/ adjacent (spaghetti) qui entraîne une assimilation des lieux d'articulation. Ce phénomène peut s'observer aussi dans le mot « kaki » prononcé parfois [keki].

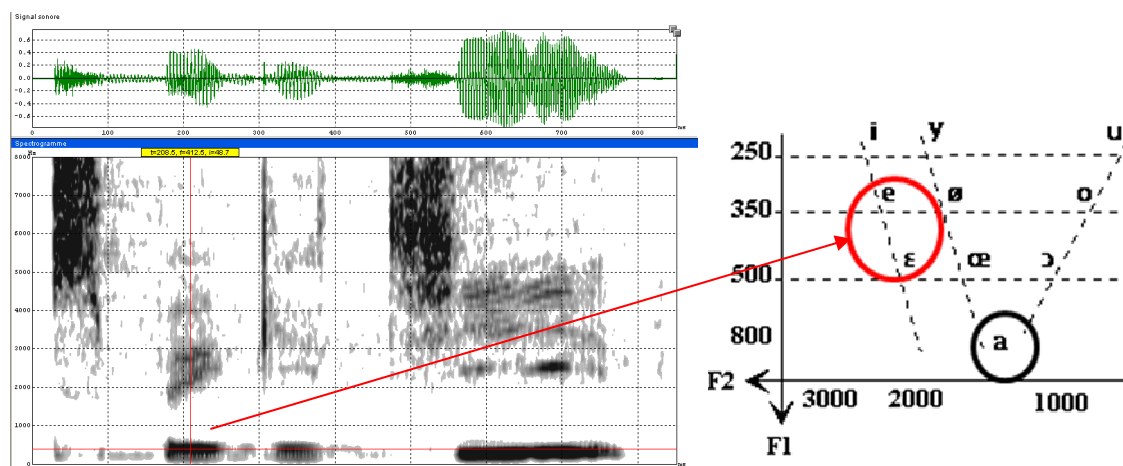


Figure 2 : A gauche, oscillogramme et spectrogramme du mot « spaghetti » prononcé en parole spontanée par une locutrice francophone. A droite, valeurs prototypiques des formants des voyelles du français. Les formants du /a/ de spaghetti sont ceux de [e].

Un autre exemple qui illustre la complexité de la perception de la parole sont les études menées par Mc Gurk & Mc Donald [38]. Ces études ont montré des effets surprenants et fortement non linéaires de 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.

Que conclure de ces exemples ? « La perception de la réalité sonore n'est pas un enregistrement direct de la réalité. C'est une construction mentale opérée à la suite d'un traitement de l'information disponible, contrainte par nos sens ainsi que nos habitudes sélectives » [12]. Dans le cadre de l'évaluation des troubles de la voix et de la parole, utiliser la perception comme méthode d'analyse reste délicat, sujette à un traitement de l'information disponible sensible au contexte [15], contrainte par des habitudes sélectives variables selon les individus et les situations.

Les exemples de la vie courante foisonnent de cas où le recours à des méthodes instrumentales s'est substitué à la simple perception humaine. Par exemple, l'arrivée d'un 100 m dans une compétition internationale n'est plus jugée à l'œil d'un chronométreur expert. Le recours à une « photo finish » est incontournable. De plus, l'intérêt des techniques instrumentales ne réside pas seulement dans leur capacité à mesurer quantitativement des phénomènes observés : elles permettent aussi de comprendre plus précisément la réalité des mécanismes observés. Pour illustrer cette affirmation, considérons le tableau de Géricault « Le Derby d'Epsom », 1821, Musée du Louvre (Figure 3). Dans cette œuvre, le peintre a représenté les chevaux aux jambes entièrement déployées volant dans les airs entre deux appuis. Cette représentation est la perception communément admise avant le XIX siècle. Les travaux de Muybridge adepte de Marey dans l'utilisation de la chronophotographie démontre que cette vision du mouvement est totalement erronée (Figure 3). En réalité, dans un galop, le cheval ramène alternativement ses jambes vers son corps à mesure qu'elles quittent le sol. Cette rectification de la représentation de la réalité s'est finalement diffusée à grande échelle au point de transformer même la vision artistique de chevaux de course au galop, en accord avec les observations issues de la photochronographie (Figure 3).

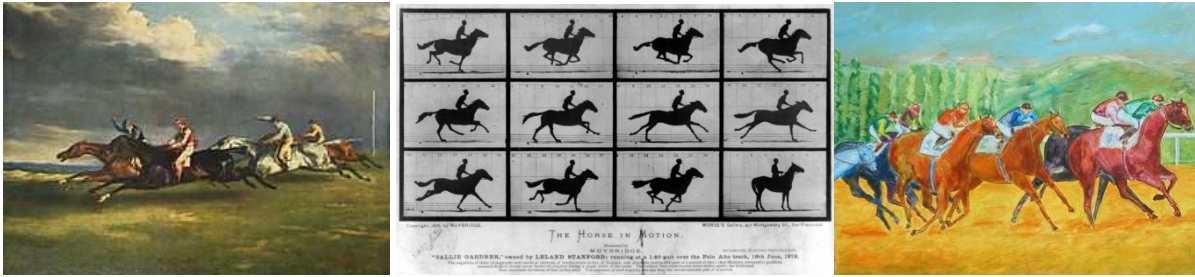


Figure 3 : A gauche, « Le Derby d'Epsom », tableau de Géricault (1821, Musée du Louvre). Au centre, chronophotographie décomposant le galop du cheval (1878, Muybridge), selon les principes de Marey. A droite, « chevaux de course au galop », 2007, François Vidal³

1.4. Comment concilier les approches perceptives, instrumentales et questionnaires dans un bilan vocal ?

Notre objectif n'est pas de démontrer la « supériorité » d'une méthode par rapport à l'autre. Nous considérons que les approches perceptives et instrumentales sont complémentaires à l'instar de Kent et. al, 2000 [33] qui écrivent « *Perceptual and instrumental approaches are complementary and the central task is to bring them into a unified picture.* » (p.1).

La justification de ce point de vue est à aller chercher dans la classification de l'Organisation Mondiale de la Santé, notamment dans la définition des concepts d'altération, invalidité et handicap⁴. Dans ce modèle, régulièrement adapté, l'altération désigne [49] « *une perte de substance ou l'altération d'une structure organique comme par exemple, la perte d'un membre ou la détérioration de certains tissus... L'invalidité désigne une restriction objective (du fait d'une altération corporelle) d'une capacité à réaliser certains actes... Le handicap désigne les difficultés qu'est susceptible de rencontrer un individu dans sa vie sociale ou personnelle du fait d'une altération corporelle ou d'une invalidité. L'importance de ces difficultés étant fonction de la situation particulière, matérielle, familiale et sociale, dans laquelle il se trouve.* » Adaptée aux troubles de la voix, l'apparition de nodules ou une cordectomie sont considérées comme des altérations. Ces altérations induisent des mécanismes de production vocale anormaux potentiellement mesurables par des techniques instrumentales, notamment fondées sur des capteurs physiologiques (débit d'air oral pour mesurer l'éventuelle fuite glottique, pression sous-glottique pour évaluer l'hypo/hyper fonctionnement pneumophonatoire, électroglottographie pour mesurer l'accolement des cordes vocales). A ce stade, l'objectif est de faire un contrôle technique de l'appareil phonatoire.

L'invalidité (ou le dysfonctionnement d'après Rossignol, 2007 [49]) pourrait résider dans la restriction des capacités langagières, notamment dans la perte de qualité vocale, d'étendue fréquentielle, d'intensité, d'endurance... Cette dimension, bien que mesurable partiellement (ex : temps maximal de phonation, phonéto gramme), apparaît comme le domaine où la perception s'avère adaptée. La parole étant produite pour être perçue, un dysfonctionnement en production entraîne la plupart du temps une restriction en perception, restriction qu'il s'agit de capter. A ce stade, l'objectif est de faire un contrôle de l'usage de l'appareil phonatoire.

Enfin, les questionnaires sont clairement dévolus à l'évaluation du handicap mais les situations particulière, matérielle, familiale et sociale s'avèrent si différentes selon les patients qu'il nous semble stérile de vouloir corrélérer résultats de questionnaire tels que le VHI [60] et mesures instrumentales sauf dans le cas d'études de grandes cohortes de patients où les situations particulières s'effacent devant la masse de données et le moyennage statistique.

1.5. Les concepts liés à l'évaluation

Comme le définit Condamines (1985, p.54 [4]), "*qui dit évaluation, dit chiffrage, donc passage de la comparaison qualitative à la comparaison quantitative.*" Indubitablement, l'évaluation est intimement liée à la mesure, opération qui consiste à donner une valeur à une observation. De façon sous-jacente, la mesure fait intervenir une notion de modèle qui est une représentation abstraite, simplifiée, d'un phénomène et qui peut se ramener à des paramètres et des grandeurs. Par exemple, dans le cas d'une voiture, le modèle

3 <http://didier-leveille.zevillage.org/news/une-page-pour-francois>

4 « International classification of Impairments, Disabilities and Handicaps, a manual of classification relating to the consequences of disease » Geneva : World Health Organisation, 1980, dont l'acronyme est ICIDH-80.

correspondant à cet objet peut être résumé par un ensemble de grandeurs définies : longueur, largeur, hauteur, masse, accélération, vitesse de pointe, couleur... La mesure ou, plus généralement, la métrologie couvre les méthodes et techniques qui permettent de paramétrer un modèle destiné à représenter la réalité. Une fois cette modélisation établie, l'objet et son modèle peut être étudié, manipulé, comparé.

Des modèles de production de la voix et de la parole existent depuis longtemps [14]. En revanche, des modèles rendant compte de la physiopathologie de la production de parole restent parcellaires, incomplets et en tout cas, insuffisamment annoncés pour permettre de clairement établir les mesures utiles à l'évaluation. Il n'est pas certain que l'orthophoniste, le phoniatre, l'ORL ou le neurologue attendent la même chose d'une évaluation car la voix et la parole ne sont pas, pour eux, exactement le même objet et donc, ils ne font pas nécessairement référence à un même modèle sous-jacent.

Pour l'orthophoniste, le diagnostic est généralement connu ; les objectifs de rééducation sont normalement définis. L'évaluation consiste alors à contrôler l'écart entre la situation réelle et les objectifs fixés. Cette vérification s'avère utile à la fois pour le thérapeute mais aussi pour le patient qui, à travers un bilan instrumental, peut prendre conscience de progrès non nécessairement perçus par lui-même. Dans ce cadre, chaque patient est son propre "modèle" et quelques mesures longitudinales bien choisies, liées au problème du patient, peuvent fournir une évaluation. Par exemple, une simple mesure d'intensité SPL effectuée à l'aide d'un sonomètre peut suffire à faire le bilan d'une rééducation d'hypophonie parkinsonienne pour laquelle le travail rééducatif porte sur l'amélioration du niveau vocal.

De son côté, l'ORL ou le phoniatre peut vouloir effectuer une évaluation vocale pour appréhender l'ampleur et le type de dysfonctionnement de façon à orienter la thérapie en conséquence ou encore pour évaluer a posteriori l'impact d'un acte chirurgical. Il s'agit plus alors d'une tâche de classification : la dysphonie est-elle légère, moyenne ou sévère ? Dans ce cadre, le bilan est nécessairement multiparamétrique de façon à rendre compte au mieux des multiples facettes de la production vocale. Bien souvent, dans ce cadre, le thérapeute est friand de mesures qui vont lui confirmer son impression auditive.

Enfin, le neurologue peut être amené à entreprendre une évaluation des troubles de la parole, non pas comme une finalité en soi, mais comme un révélateur ou un marqueur d'une dégénérescence neurologique ou encore pour mesurer l'effet d'une pharmacopée ou d'une stimulation électrophysiologique. L'évaluation instrumentale de la parole devient ainsi un moyen très indirect d'appréhender son modèle de dysfonctionnement neurologique général.

Il existe de multiples techniques pour faire de l'évaluation instrumentale de la voix et de la parole. La plus simple réside dans l'analyse acoustique, mais cette simplicité « apparente » ne doit pas masquer de multiples précautions d'usage.

2. L'évaluation vocale par analyse acoustique

2.1. Précautions techniques nécessaires à l'évaluation acoustique

2.1.1. Les conditions d'enregistrements nécessaires à l'évaluation acoustique

L'analyse acoustique de la voix et de la parole nécessite une bonne qualité du signal sonore, particulièrement dans l'étude des dysfonctionnements où il s'avère parfois nécessaire de mesurer de fines nuances qui ne doivent pas être entachées de bruit. Actuellement, la diffusion très large d'ordinateurs équipés en série d'entrées audio autorisant l'enregistrement sonore sans difficulté ne doit pas occulter le fait que certaines précautions doivent être prises. Typiquement, il est déconseillé d'utiliser un microphone bas de gamme connecté directement sur l'entrée micro de l'ordinateur : le résultat pourra être convenable à l'oreille mais les logiciels d'analyse auront des difficultés à fournir des mesures correctes sur des données altérées (Figure 4). Il est préférable, d'une part d'utiliser des microphones de qualité, le plus souvent électrostatiques. D'autre part, il est conseillé de connecter le microphone sur un préamplificateur externe et non directement à l'ordinateur. En effet, ces dispositifs périphériques (ex: Edirol USB Audio Capture UA25) génèrent beaucoup moins de bruit et, comme la plupart intègrent à présent les convertisseurs analogique-numérique, l'expérimentateur récupère alors directement le signal numérisé, évitant ainsi de multiples manipulations génératrices de problèmes.

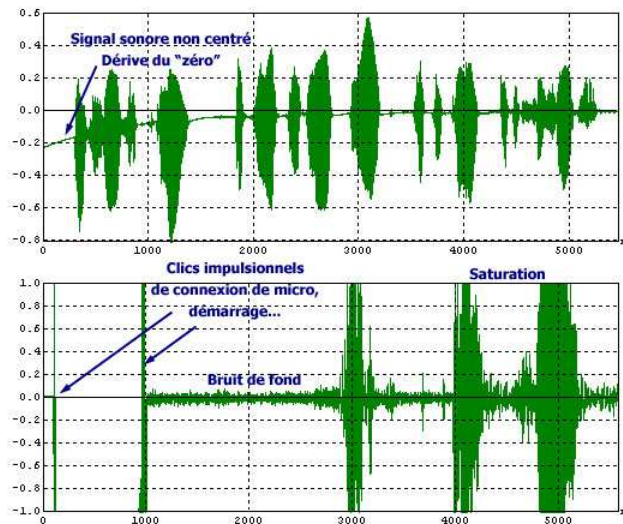


Figure 4 : Signaux de parole enregistrés sans précaution sur des dispositifs de mauvaise qualité. Une analyse acoustique fiable ne sera pas possible sur de telles données. En haut, on remarque un offset initial (le signal n'est pas centré sur zéro) et une dérive progressive. En bas, on observe des clics au démarrage, un bruit de fond résiduel important et la saturation du signal dans les parties énergétiques.

2.1.2. Le problème du calibrage de l'intensité

Le sonomètre reste l'instrument idéal pour mesurer l'intensité sonore car il fournit une mesure SPL calibrée. Il permet de faire des études comparatives fiables sur le niveau sonore. Ainsi, par exemple, il est possible de mesurer le fait que tel locuteur a augmenté de 6 dB sa puissance sonore en 3 mois de rééducation. A partir du moment où le signal sonore est enregistré sur un quelconque support (magnétique, informatique...) et/ou avec un microphone dont on ne connaît pas la sensibilité, ce type d'analyse comparative devient délicat. En effet, si les caractéristiques de la chaîne d'enregistrement constituée par le microphone, le préamplificateur et le convertisseur numérique ne sont pas parfaitement connues, il est impossible de connaître la valeur SPL finale de l'enregistrement, en particulier s'il est stocké sous forme numérique en fichier (ex: fichier wav). En effet, l'amplitude observée du signal dépend du microphone utilisé, du volume d'enregistrement utilisé et de divers autres réglages. Autrement dit, on peut obtenir pour un locuteur à voix forte un fichier audio avec de faibles valeurs d'amplitude tout simplement parce qu'on a utilisé un microphone à faible sensibilité, placé loin du locuteur, avec un faible volume d'enregistrement. Inversement, un locuteur hypophonique peut fournir un fichier audio avec de fortes valeurs d'amplitude si a été utilisé un microphone sensible placé à 10 cm du locuteur avec un enregistreur à fort gain.

Hormis le sonomètre, il existe deux moyens d'obtenir des données sonores permettant la mesure calibrée en intensité. La première consiste à maîtriser parfaitement toute la chaîne d'acquisition. C'est le cas dans le dispositif EVA2 fabriqué par SQLab-LPL (www.sqlab.fr) où un microphone précis est connecté à un préamplificateur réglé pour fournir 10 V crête à 110 dB SPL à 30 cm. Le signal est ensuite transmis via un canal d'acquisition indépendant des boutons de réglage de volume et la numérisation est effectuée par une carte fournissant un signal codé sur 16 bits pleine échelle à 10V. Dans ce cas là, il existe un lien direct entre l'amplitude du signal numérique en fin de chaîne et l'amplitude acoustique réelle du départ.

La deuxième solution consiste à utiliser une source sonore étalon, ceci avant chaque enregistrement. On peut utiliser un "buzzer" dont l'intensité est connue (ex: 83 dB à 30 cm). En émettant un bip en début d'enregistrement, on peut alors en déduire, par comparaison, l'intensité du signal de parole (ex: -10 dB par rapport au buzzer signifiera 73 dB). Si aucune précaution de calibrage n'a été prévue par l'expérimentateur, il est toutefois possible de faire des mesures d'intensité en valeurs relatives sur le même enregistrement. Autrement dit, il ne faut pas s'attacher à la valeur absolue d'intensité affichée mais en revanche, si l'on mesure une baisse de 10 dB, par exemple, lors d'une tâche de répétition, cette diminution de 10 dB est effective. Dans tous les cas, il faut alors être certain que l'enregistreur ne possède pas un circuit électronique de contrôle automatique de volume d'enregistrement dont la particularité est d'écraser la dynamique naturelle du signal de parole. Ainsi, les faibles amplitudes sont fortement rehaussées et, inversement, les fortes valeurs sont limitées. Cette technologie est quasi systématiquement utilisée dans les studios professionnels de radiophonie afin d'obtenir au final un signal sonore ayant approximativement une intensité sonore moyenne constante, évitant à l'auditeur de modifier sans cesse son volume d'écoute. Elle est aussi utilisée dans des enregistreurs portables pour lesquels il est alors nécessaire de désactiver ce réglage automatique. Bien

évidemment, sans information sur la méthode d'enregistrement, il convient de rester prudent sur les mesures d'intensité.

2.1.3. La délicate tâche de détection de la fréquence fondamentale

La majeure partie des indices exploités dans l'évaluation acoustique de la voix ou de la mélodie dérive de la mesure de fréquence fondamentale. Il est possible de faire des mesures manuelles de F_0 , comme proposées dans Baken et al. 2000 [2]. Toutefois, nous n'aborderons dans cette partie que le calcul automatique de cette grandeur très importante, des détails pouvant être consultés dans Hess, 1983 [28].

2.1.3.1. Les étapes et les différentes techniques de détection de la F_0

L'étape préliminaire à la mesure de la F_0 consiste à repérer le voisement dans le continuum acoustique. Cette détermination des segments voisés/non voisés n'est pas triviale, surtout dans le cas de dysfonctionnement où les cordes vocales peuvent "décrocher" ou se remettre en vibration très accidentellement. La plupart des algorithmes de détection explicite de voisé/non voisé sont fondés sur un ensemble de paramètres physiques tels que l'énergie, le taux de passage par zéro, le 1er coefficient LPC, le facteur de balance grave/aigu... dont les valeurs sont comparées à des seuils plus ou moins dynamiques. Par exemple, si une portion de signal a une faible énergie et un taux de passage par zéro important, le segment est déclaré non voisé. Des mécanismes de suivi temporel permettent ensuite de statuer sur le voisement ou pas des différents segments de parole.

Sur les parties détectées comme voisées, il est alors possible de mesurer la F_0 . Il existe différentes techniques pour cela [28]. On peut citer les techniques temporelles d'autocorrélation, d'AMDF (Average Magnitude Difference Function) ou encore les techniques spectrales d'analyse harmonique (technique du peigne) ou du cepstre. La plupart de ces méthodes sont fondées sur un calcul par trame, c'est-à-dire que le signal de parole est découpé en morceaux et une valeur de F_0 est fournie sur chaque section de durée fixe comme, par exemple, toutes les 10 ms (trame centiseconde). Ce principe est suffisant pour ensuite obtenir des courbes mélodiques rendant compte des montées ou descentes de la F_0 . En revanche, un tel procédé est complètement inadapté pour évaluer la stabilité à court terme du vibrateur laryngé. En effet, pour cela, il est nécessaire d'utiliser un détecteur de cycle fondé, par exemple, sur du passage par zéro ou de la détection de pic. L'objectif est alors de "marquer" très précisément le début et la fin de chaque cycle vibratoire de façon à en calculer les variations instantanées. L'utilisation de méthodes mixtes avec des algorithmes à plusieurs passes apparaît désormais comme un gage de robustesse. Dans ce type d'approche, une détection simultanée par AMDF, peigne et cepstre suivie d'un algorithme de vote permet d'obtenir une F_0 par trame fiable. On peut alors, dans un deuxième temps, déclencher une détection cycle à cycle rendue robuste par la détermination préliminaire de la F_0 sur la trame. La 2^{ème} passe consiste alors à affiner la détection et fournir très précisément le début et la fin de chaque cycle comme préconisé dans Titze, 1993 [54].

2.1.3.2. Les problèmes de détection de la F_0 : erreurs de voisé/non voisé et sauts d'octave

La plupart des problèmes de détection de F_0 porte sur 2 catégories : le voisement et les sauts d'octave. La décision du voisé/non voisé est un compromis délicat: soit le détecteur élimine trop de parties voisées en les considérant comme non voisées et par conséquent, la F_0 n'est pas calculée sur ces parties intéressantes; soit le détecteur favorise trop le voisement et des parties non voisées sont soumises à un calcul de F_0 qui se solde par des valeurs aberrantes car portant sur du bruit aperiodique. Il peut être intéressant d'introduire une contre-réaction sur la décision voisé/non voisé suite au calcul de F_0 en cas de valeurs aberrantes et déclarer non voisées des suites anarchiques de F_0 dues, par exemple, à un calcul sur du bruit. Toutefois, cette solution peut s'avérer malvenue notamment dans le cas de la parole pathologique. Dans ce cas précis, il est préférable d'introduire un réglage semi-manuel laissé à l'expertise de l'observateur.

Le deuxième problème fréquent dans la détection de F_0 est le saut d'octave. En effet, dans les techniques temporelles, si une périodicité est observée pour un temps Δt , on considérera qu'il s'agit alors de la période fondamentale T_0 , inverse de la F_0 . Mais $2T_0$, $3T_0$, $4T_0$ constituent aussi des candidats à la périodicité. Si l'algorithme, pour diverses raisons, sélectionne comme résultat $T=2T_0$, cela entraîne une erreur de calcul de fréquence fondamentale égale à $F_0/2$. Inversement, dans les techniques spectrales de type analyse harmonique (méthode du peigne), si un espacement Δf constant est observé entre deux raies spectrales, on considérera qu'il s'agit alors de la fréquence fondamentale F_0 , cette valeur étant l'écart systématique entre les harmoniques. Mais $2F_0$, constitue aussi un candidat à l'intervalle spectral. Si l'algorithme, pour diverses raisons, sélectionne comme résultat $F=2F_0$, cela entraîne une erreur de calcul de fréquence fondamentale.

2.1.3.3. Esprit critique et solutions

La détection de F0 n'est pas un processus sans faille comme peut l'être le calcul de l'intensité ou du spectrogramme. Il existe des seuils, des limites, des choix qui ont une incidence sur le résultat. Aussi, face à une courbe de F0, il faut toujours garder un œil critique. Tout saut brutal ou toute montée/descente tronquée doit être vue avec circonspection. En même temps, dans le cas de parole pathologique, ces "sauts" de F0 particulièrement visibles, par exemple, dans les diplophonies, font partie entière de la production vocale. L'œil expert reste donc important.

Une solution proposée et validée par De Looze, 2010 [8] consiste à effectuer une première détection de la fréquence fondamentale en laissant des paramètres larges de registre vocal (ex : [75 – 300 Hz] pour une voix d'homme, [100 – 500 Hz] pour une voix de femme). Dans un second temps, en fonction des mesures obtenues lors de la première détection, le registre vocal est adapté au locuteur en considérant comme valeur basse $F0_{min} = q15 * 0.83$ (où 'q' est un percentile) et comme valeur haute $F0_{max} = q65 * 1.92$. q15 représente la valeur de F0 du 15^{ème} percentile, c'est-à-dire la valeur seuil qui inclut 15% des mesures les plus faibles. Graphiquement, cela représente la partie gauche de l'histogramme de F0 (Figure 6). q65 représente la valeur de F0 du 65^{ème} percentile, c'est-à-dire la valeur seuil qui inclut 65% des mesures les plus faibles (ou qui exclut 35% des mesures les plus élevées). Graphiquement, cela représente la partie droite de l'histogramme de F0. Ces valeurs seuils étant fixées, une détection est à nouveau effectuée. Ces formules ont prouvé qu'elles apportaient une meilleure estimation de la F0 du locuteur, notamment en parole continue, en excluant plus efficacement les sauts d'octave inadéquats.

2.2. Evaluation acoustique de la voix

Les travaux portant sur l'évaluation acoustique des troubles vocaux sont innombrables. Nous ne présenterons que les principaux paramètres. Des détails peuvent être consultés dans Baken et al. 2000 [2].

2.2.1. Les mesures de vibration laryngée

2.2.1.1. La fréquence fondamentale moyenne

La fréquence fondamentale F0 reste le meilleur indicateur des caractéristiques biomécanique des cordes vocales. La F0 moyenne apporte une mesure globale de la hauteur de la voix du sujet (voix aiguë, grave...). Curieusement, dans la littérature, on trouve fréquemment des valeurs moyennes de F0 pour les hommes ou pour les femmes mais sans réelles mesures expérimentales. Nous avons donc repris les données de Peterson et Barney 1952 [42] qui portent sur 76 locuteurs américains (33 hommes, 28 femmes, 15 enfants) ayant prononcé 2 fois une liste de 10 mots contenant une voyelle sur laquelle a été notamment mesurée la F0 moyenne. Ces données sont légèrement supérieures (Table 1) à celles obtenues par Fitch & Holbrook, 1970, dans Baken et al. 2000 [2] qui ont porté sur 100 hommes et 100 femmes jeunes anglophones de 18 à 25 ans sur de la lecture de texte (Rainbow Passage). En revanche, les données de Fitch & Holbrook sont proches de celles obtenues sur le français par Yu et al. 2007 [64] où la population contrôle est composée de 38 femmes et 20 hommes de 19 à 59 ans (38 ans de moyenne) sur de la production de /a/ tenu.

Table 1: moyennes et écart type de F0 par groupe de locuteurs selon diverses sources

Source	Peterson et Barney, 1952			Fitch & Holbrook, 1970		Yu et al., 2007	
Age (ans)	Inconnu			18-25		19-59 (moy=38)	
Groupe	Hommes	Femmes	Enfants	Hommes	Femmes	Hommes	Femmes
Langue	Americ.	Americ.	Americ.	Americ.	Americ.	Français	Français
Effectif	33	28	15	100	100	38	20
F ₀ moyenne	131 Hz	223 Hz	264 Hz	116 Hz	217 Hz	122 Hz	218 Hz
Ecart-type F ₀	19 Hz	27 Hz	35 Hz	34 Hz	45 Hz	24 Hz	28 Hz

Sauf écart très important (ex : voix d'homme à 250 Hz) par rapport à ces données "normatives", il est difficile de corrélérer un dysfonctionnement et une simple valeur de F₀ moyenne. Ceci est d'autant plus vrai qu'il existe en fonction de l'âge un glissement différent de la F₀ moyenne pour les hommes et les femmes, la F₀ augmentant chez les hommes, diminuant chez les femmes (Figure 5). Yu et al. 2007 [64] ont retrouvé aussi cette tendance en fonction du degré de dysphonie (Figure 14).

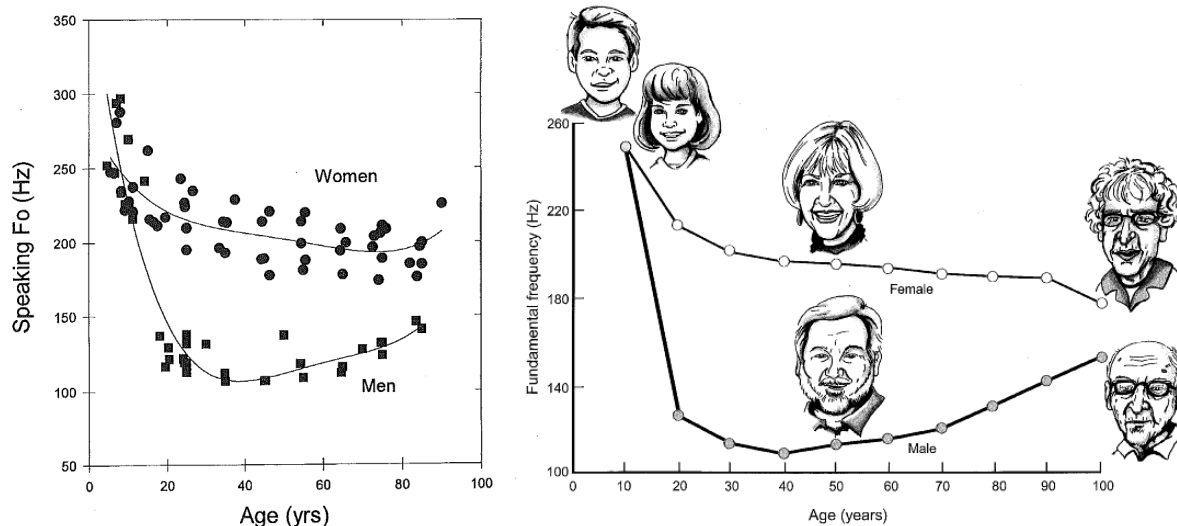


Figure 5 : Evolution de la F_0 moyenne chez les hommes et les femmes en fonction de l'âge. gauche, extrait de (Baken et al., 2000, p.174). A droite, extrait de (Hixon et al. 2008, p146)

2.2.1.2. L'ambiguïté contradictoire des variations de F_0

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 d'ailleurs l'un des intérêts 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. D'où la nécessité d'introduire plusieurs paramètres.

2.2.1.3. Instabilité de la vibration

La dynamique ou l'instabilité laryngée se traduit par des variations de F_0 au cours du temps. Elle peut se mesurer par l'**écart type de la F_0** , qui correspond à l'ampleur en Hz des variations de F_0 autour de la moyenne. Le **coefficient de variation** permet de relativiser l'écart type en le comparant à la F_0 moyenne. Il correspond donc à l'ampleur en % des variations de F_0 par rapport à la F_0 moyenne. Ainsi, à un écart type de 4.9 Hz pour une F_0 moyenne de 180 Hz correspond un coefficient de variation de 2.7%, valeur importante sur une voyelle tenue. Le même écart type pour une F_0 moyenne de 500 Hz fournit un coefficient de variation de 0.98%, valeur beaucoup plus normale sur ce type de production. Le coefficient de variation de la F_0 est donc le meilleur indice pour explorer globalement la dynamique ou l'instabilité du vibrateur laryngé (Figure 6). Sur des voyelles tenues, il devient important en présence de vibrato pathologique, chevrottement, instabilité d'origine neurologique.

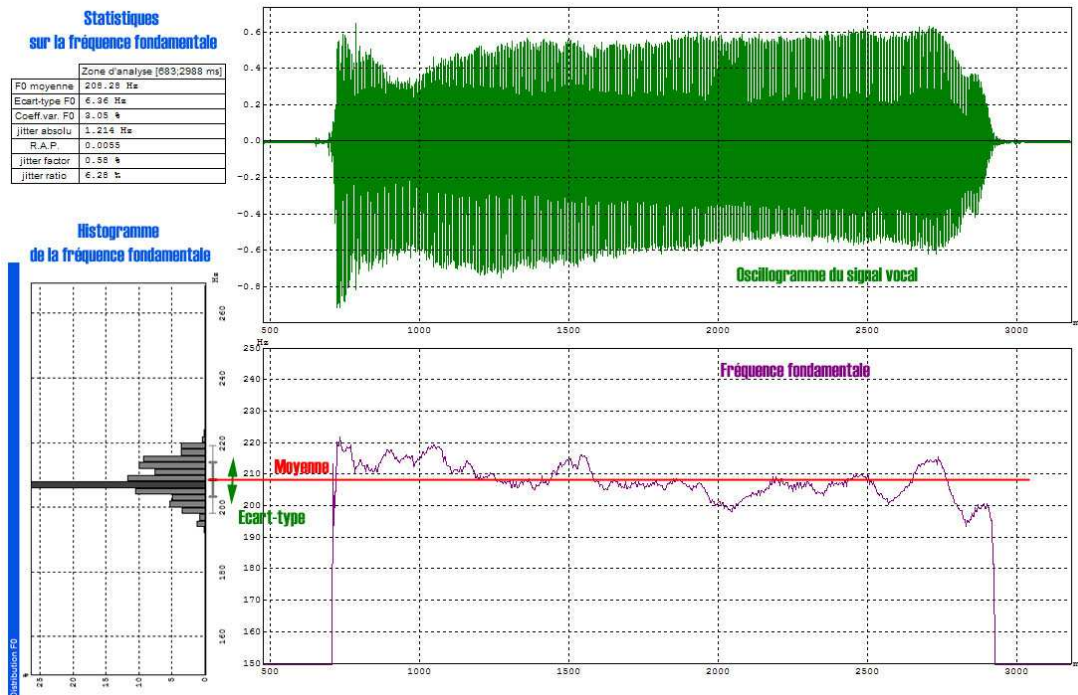


Figure 6 : signal vocal sur un /a/ tenu d'un adolescent en cours de mue "difficile". Décours temporel, histogramme et statistiques de la fréquence fondamentale. On observe les variations importantes de F₀ alors que la consigne est à la stabilité. Le coefficient de variation F₀ de 3% mesure cette instabilité globale. En revanche, le faible jitter de 0.58 % traduit une relativement bonne stabilité instantanée.

L'instabilité à court terme de la F₀ se traduit par des variations de fréquence entre chaque cycle d'oscillation (Figure 7). Elle se calcule par le **JITTER absolu moyen** qui est la moyenne de la différence de F₀ entre deux cycles de vibration consécutifs. Ces variations sont calculées cycle à cycle, de manière très précise. Le **JITTER factor** permet de relativiser le jitter moyen en le comparant à la F₀ moyenne. Ainsi, à un jitter moyen de 0.677 Hz pour une F₀ moyenne de 180 Hz correspond un jitter factor de 0.38%, valeur faible. Le même jitter moyen pour une F₀ moyenne de 80 Hz fournit un jitter factor de 0.84%, valeur plus problématique. Le jitter factor est donc un meilleur indice pour explorer la stabilité de la fréquence fondamentale à court terme. Il reste le meilleur corrélat acoustique de la raucité, bien que cette corrélation simple ne permette pas de rendre compte complètement de cette dimension vocale. Par consensus, il est admis qu'une valeur du jitter factor supérieure à 1% est une trace d'instabilité laryngée.

Le **RAP** (Relative Average Perturbation) mesure la moyenne des variations de trois périodes consécutives rapportée à la période moyenne du signal observé. Le **JITTER ratio** mesure la moyenne des variations de période, inverse de la fréquence, entre deux cycles de vibration consécutifs et relativise cette valeur par la période moyenne du signal observé.

Il faut signaler qu'un jitter élevé entraîne toujours un coefficient de variation important. La réciproque n'est pas vraie. En effet, de petites variations croissantes ou décroissantes de F₀ entre chaque cycle ne fournira pas un jitter important mais pourra donner lieu à des variations globales importantes de F₀ (exemple typique : un vibrato).

L'instabilité à court terme de l'amplitude de la vibration se traduit par des variations d'amplitude entre chaque cycle d'oscillation (Figure 7). Elle se calcule par le **SHIMMER moyen** qui est la moyenne des rapports d'amplitudes entre deux cycles de vibration consécutifs. Ce rapport est généralement exprimé en dB. Ces variations sont calculées cycle à cycle, de manière très précise. L'**APQ** (Amplitude Perturbation Quotient) mesure la moyenne des variations d'amplitude en tenant compte de 11 périodes consécutives, le tout rapporté à l'amplitude moyenne du signal observé.

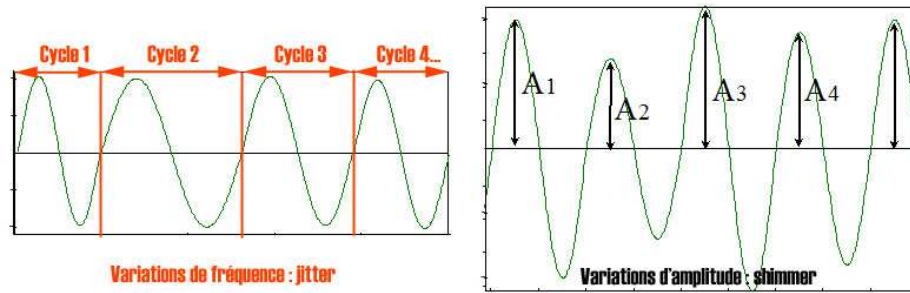


Figure 7 : Variations instantanées des caractéristiques de la vibration. A gauche, l'écart de fréquence entre deux cycles consécutifs entraînera un jitter important ; à droite, l'écart d'amplitude entre deux cycles consécutifs entraînera un shimmer important

Il existe de nombreux paramètres de ce type (Table 2) et il serait fastidieux de les présenter tous. Pour plus de détails, le lecteur pourra se référer à Baken et al. 2000 [2].

Table 2 : Principaux indices liés à la vibration laryngée et sa stabilité (Baken et al., 2000)

Moyenne de la F_0	$\text{moyenne (en Hz)} = \frac{1}{N} \sum_{i=1}^N F0_i$
Ecart type de la F_0	$\text{écart-type (en Hz)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (F0_i - F0_{\text{moy}})^2}$
Coefficient de variation	$\text{coefficient de variation (en \%)} = 100 \times \frac{\text{écart type}}{\text{moyenne}}$
Perturbation Cycle à Cycle	$\text{Jitter moyen (en Hz)} = \frac{1}{N-1} \sum_{i=1}^{N-1} F0_i - F0_{i+1} $
Perturbation Moyennée	$\text{Relative Average Perturbation (sans unité)} = \frac{1}{N-2} \sum_{i=2}^{N-1} \left \frac{T0_{i-1} + T0_i + T0_{i+1}}{3} - T0_i \right $ $T0_{\text{moy}}$
Pourcentage	$\text{Jitter factor (en \%)} = 100 \times \frac{\text{jitter moyen (Hz)}}{F0_{\text{moy}}}$
	$\text{Jitter ratio (en \%)} = 1000 \times \frac{\text{jitter moyen (ms)}}{T0_{\text{moy}}}$
Perturbation Cycle à Cycle	$\text{shimmer moyen (en dB)} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left 20 \cdot \log \left(\frac{A_i}{A_{i+1}} \right) \right $
Perturbation Moyennée	$\text{Amplitude Perturbation Quotient APQ} = \frac{1}{N-10} \sum_{i=6}^{N-5} \left \frac{A_{i-5} + \dots + A_i + \dots + A_{i+5}}{11} - A_i \right $ A_{moy}

2.2.2. Les mesures de signal/bruit

Les mesure de signal et de bruit explorent la présence (ou absence) de « bruits » au cours de la phonation. Ce bruit peut être dû à un écoulement turbulent causé par un mauvais accolement des cordes vocales. On parle alors de « bruit additif ». Toutefois, il faut signaler qu'une vibration très instable peut être elle aussi mise en évidence par cette analyse : l'écart de la F_0 et de ses harmoniques par rapport à la valeur moyenne est détectée comme un « bruit » de modulation.

Le rapport Harmonique/Bruit (Harmonic to Noise Ratio, **HNR**) se calcule à partir d'une méthode temporelle décrite dans Yumoto, 1982 [68]. 25 cycles consécutifs du signal voisé permettent d'établir une forme ondulatoire moyenne ; la soustraction entre le signal originel et la forme moyenne permet d'obtenir un résidu

considéré comme le bruit. Si H est l'énergie de l'onde moyennée périodique et N l'énergie du bruit résiduel, le rapport H/N calculé régulièrement fournit la valeur du HNR éventuellement converti en dB.

Le Signal Ratio (**Sr**) se calcule à partir d'une méthode spectrale décrite dans Hiraoka, 1984 [29]. On considère le signal vocal comme la somme de deux composantes: une composante périodique (F0 & harmoniques) et une composante de bruit (le reste). A 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 aussi de faire le même type de calcul non pas sur tout le spectre mais uniquement sur des parties 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 8).

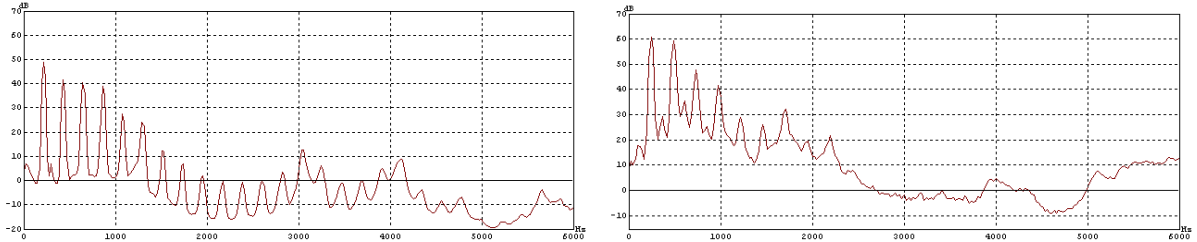


Figure 8 : Spectres d'une voix normale (à gauche) et d'une dysphonie (à droite).

Les composantes harmoniques disparaissent clairement au dessus de 2000 Hz sur l'échantillon dysphonique.

L'énergie Normalisée du Bruit (**NNE**) se calcule à partir d'une méthode spectrale décrite dans Kasuya, 1986 [32]. Cette méthode est assez proche de la précédente excepté dans la formulation finale qui indique non pas le pourcentage de périodicité mais plutôt la part de bruit exprimée en dB par rapport au signal complet.

Bien que calculés avec des techniques différentes (temporelles vs spectrales), ces indices restent corrélés. En effet, si on appelle S_H la composante harmonique et S_N la composante de bruit, le signal ratio Sr s'exprime sous la forme:

$$Sr = \frac{S_H}{S_H + S_N} \Leftrightarrow \frac{1}{Sr} = \frac{S_H + S_N}{S_H} \Leftrightarrow \frac{1}{Sr} = 1 + \frac{S_N}{S_H} \Leftrightarrow \frac{1 - Sr}{Sr} = \frac{S_N}{S_H} \Leftrightarrow \frac{Sr}{1 - Sr} = \frac{S_H}{S_N}$$

Ce qui conduit au résultat : $HNR = \frac{S_H}{S_N} = \frac{Sr}{1 - Sr}$

Ainsi, pour un Sr mesuré à 75% (0.75), le HNR équivalent est 3.

2.3. Les nouvelles tendances

Depuis une dizaine d'années, de nouvelles techniques ont été proposées pour établir un bilan acoustique de la voix, notamment en analysant de la parole continue.

2.3.1. Les portraits de phase et les mesures de non-linéarité

La représentation du signal sous la forme d'un portrait de phase est un outil classique de la physique non-linéaire. Dès la version 1 du dispositif EVA [58], un diagramme à deux dimensions permettait de distinguer les instabilités aléatoires vs. les instabilités « organisées ». Le principe de ce tracé (Figure) est de placer un point aux coordonnées $\{x=F_0(t) ; y=F_0(t+\tau)\}$, τ étant défini comme un délai temporel. Un ensemble de points peu espacés indique alors une bonne stabilité de la vibration laryngée car cela signifie que la valeur de F_0 au temps (t) est proche de celle au temps (t+ τ), synonyme de constance. En revanche, la présence de points éloignés de la droite bissectrice $F_0(t) = F_0(t+\tau)$ indique des variations instantanées de F_0 (Figure 9a). D'autre part, une distribution de points le long de la droite bissectrice témoigne d'une dérive à long terme (ex : vibrato) mais d'une bonne stabilité à court terme (Figure 9b). Enfin, un nuage dispersé met en évidence une vibration anarchique (Figure 9c). Il est possible de passer à une représentation en trois dimensions en rajoutant la coordonnée $z=F_0(t+2\tau)$ pour faire apparaître des trajectoires organisées en attracteur (Figure 9d) ou bien des chevelus (Figure 9e). L'analyse quantitative des attracteurs a fait l'objet de travaux dans le domaine. Le calcul des coefficients de Lyapunov a été proposé sur le plan théorique par Titze puis explicitement développé dans l'évaluation vocale par Giovanni et al., 1999 [20]. Le principe général est le calcul de la divergence de deux trajectoires voisines d'un portrait de phase. Toutefois, le Plus Grand Exposant de Lyapounov, bien que

paramètre d'analyse très prometteur, reste peu utilisé du fait de la difficulté de relier cette mesure à une réalité physiologique.

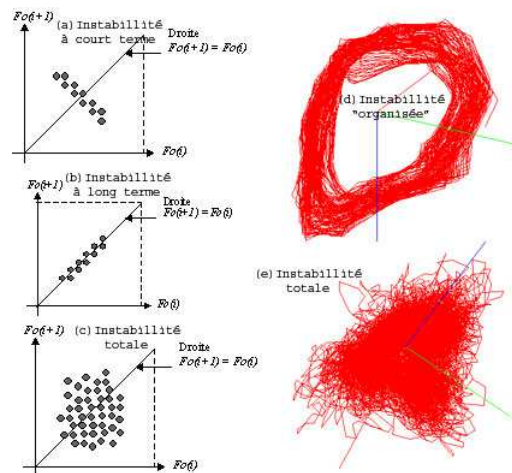


Figure 9 : Représentations en 2D de l'instabilité vocale (a, b, c) et portraits de phase en 3D (d, e)

2.3.2. Variogrammes et mesure de dyspériodicité

(Kacha et al., 2006) ont proposé une analyse de la dyspériodicité vocale fondée sur la méthode du variogramme généralisé qui permet l'estimation des dyspériodicités inter-cycles sur de la parole continue, quelles que soient leurs causes. Le principe est d'estimer les variations qui apparaissent entre des trames de signaux successifs en tenant compte de divers paramètres. Les auteurs ont montré sur divers corpus de voix dysphoniques que le rapport signal/dyspériodicité semble corrélé fortement avec le degré perçu de la sévérité d'une dysphonie [1].

2.3.3. Les techniques de classification automatique

La Reconnaissance Automatique du Locuteur (RAL) s'intéresse 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. Diverses tentatives ont été effectuées pour adapter ces techniques non pas à l'identification d'un locuteur mais à un groupe de locuteurs, en l'occurrence des dysphoniques [11]. Le principe est de fournir des données acoustiques de locuteurs atteints de troubles de la voix. Sur le signal sonore sont extraits des informations spectrales qui sont modélisées par une phase d'apprentissage supervisée. Durant cette étape, le système peut être confronté à des dysphonies légères, moyennes ou sévères. A la fin de l'apprentissage, il a établi de façon automatique des caractéristiques statistiques de chaque groupe. La phase d'utilisation proprement dite consiste par la suite à confronter le système à une nouvelle voix qui va être associée à l'un des groupes appris (ex : voix normale ou dysphonie légère ou dysphonie moyenne ou dysphonie sévère). D'autres techniques de classification ont été essayées : SVM (support vector machines) dans Godino-Llorente et al., 2005 [25], réseaux neuromimétiques dans Godino-Llorente et al., 2009 [26], techniques issues de l'extraction d'information musicale [9]. L'objectif principal de ces méthodes réside dans la classification d'une voix en normale vs. pathologique ou selon un nombre de catégories plus importantes comme dans Fredouille et al., 2009 [11].

Les limites actuelles de ces techniques sont diverses. La première faiblesse est d'ordre technique et réside dans leur forte dépendance au corpus d'apprentissage. Saenz-Lechon et al. 2006 [51] ont montré qu'il était possible de distinguer une voix normale d'une voix dysphonique si ces deux populations étaient enregistrées avec des microphones différents. En effet, les systèmes automatiques fonctionnent en aveugle et donc modélisent toute forme de variation (et donc des conditions d'enregistrement). Le deuxième obstacle est d'ordre métrologique. La plupart de ces techniques sont des classificateurs, c'est-à-dire qu'ils associent un échantillon vocal à une catégorie (ex: voix normale ou dysphonique), tâche intéressante mais qui peut être insuffisante si l'on souhaite obtenir une valeur analogique pour observer, par exemple, un résultat vocal en suivi longitudinal. Le dernier frein à l'usage de ces techniques est d'ordre épistémologique. Ces systèmes fonctionnent comme des « boîtes noires ». Il est très difficile de savoir sur quels éléments a porté la décision,

ce qui ne convient pas nécessairement au clinicien, au physiologiste ou au phonéticien qui cherche à comprendre et expliquer ses observations. Cependant, ces techniques innovantes ont la vocation à se développer, notamment du fait de leur simplicité d'utilisation et de leur capacité à appréhender de la parole continue.

3. L'évaluation vocale par électroglottographie

L'électroglottographe ou EGG est un dispositif qui permet d'obtenir une image de l'accolement et de l'ouverture des cordes vocales au moyen de la mesure de l'impédance électrique entre deux électrodes placées en surface de part et d'autre du larynx. Bien qu'il ne donne pas une information réelle de la surface d'ouverture de la glotte, il permet en revanche une très bonne représentation du cycle d'oscillation des cordes vocales et autorise une bonne mesure de sa fréquence car il est dénué de bruits aérodynamiques.

3.1. Composantes du signal E.G.G

Le signal EGG se présente sous la forme d'une onde caractérisée par une modulation rapide, synchronisée avec le cycle d'ouverture et de fermeture des cordes, ainsi qu'une modulation lente, révélatrice des mouvements du larynx. Sur la Figure 10, on peut observer deux composantes :

- une modulation rapide synchronisée avec le cycle d'ouverture et de fermeture des cordes vocales
- une composante lente (basse fréquence) qui traduit les mouvements lents du larynx (ex: déglutition) ou des mouvements de rotation du cou (artefact de mesure)

Pour séparer ces deux composantes, il est nécessaire d'effectuer du filtrage (passe-haut pour extraire le signal vibratoire).

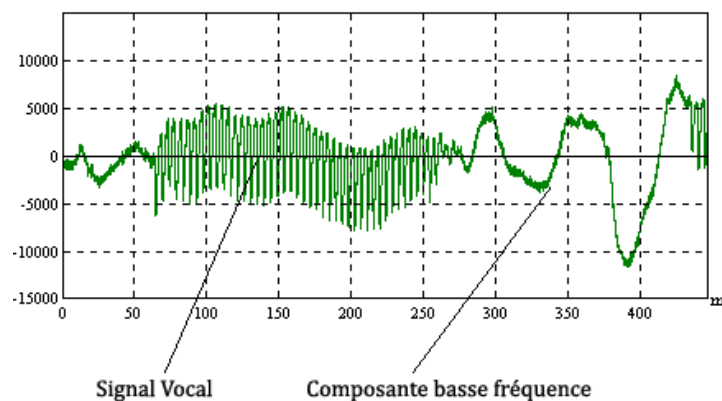


Figure 10 : Signal électroglottographique

3.2. Phase et amplitude du signal E.G.G

La phase, c'est-à-dire le signe du signal électroglottographique est extrêmement important. Les parties hautes de la courbe peuvent indiquer soit l'accolement, soit l'ouverture des cordes vocales (Figure 11). Cette phase dépend d'une part de la technologie du système d'électroglottographie (avec le Laryngograph de Fourcin, le signal indique l'accolement des cordes vocales). D'autre part, il dépend du système d'acquisition qui peut effectuer une inversion du signe lors de l'enregistrement des signaux.

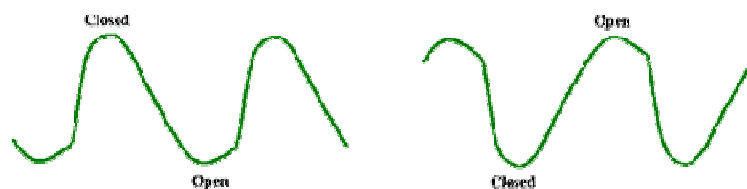


Figure 11 : Phase du signal électroglottographique (à gauche, signal d'accolement ; à droite, signal d'ouverture)

La qualité du signal EGG est fortement dépendante de la bonne application des électrodes. L'amplitude du signal EGG dépend de ce positionnement et de la conduction des tissus biologiques du locuteur. Il n'apporte

qu'une information relative. Autrement dit, une plus grande dynamique entre deux enregistrements n'implique pas nécessairement un meilleur fonctionnement vibratoire.

3.3. Le quotient de fermeture

Le quotient de fermeture (closed quotient, CQ) consiste à mesurer le rapport entre la phase d'accolement et la durée du cycle complet. Une méthode est proposée par Scherer et al., 1993. Cette technique nécessite un seuil H qui détermine les frontières entre phases fermées et ouvertes. Un seuil H de 25% et 35% est proposé par les auteurs, ces valeurs ayant été validées par des observations stroboscopiques. Le quotient de fermeture est alors le rapport entre phase fermée (T_c) et durée du cycle (T_c+T_o)

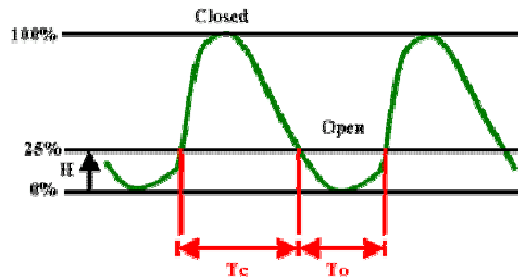


Figure 12 : Mesure du quotient de fermeture.

Une alternative a été proposée par Henrich et al., 2004 [27]. Elle consiste à identifier les phases fermées et ouvertes par le repérage des pics du signal dérivé du signal électroglottographique (dEGG). Les auteurs indiquent que « les pics positifs très marqués vont être reliés aux instants de fermeture glottique, c'est à dire les instants où le débit va commencer à diminuer jusqu'à s'annuler. Les pics négatifs moins marqués sont reliés aux instants d'ouverture glottique, c'est à dire les instants où le débit va commencer à s'accélérer et à passer à travers la glotte. » Cette technique a l'avantage de faire abstraction d'une notion de seuil, toujours délicate. L'inconvénient réside dans le fait que sur des signaux EGG de mauvaise qualité liée à un mauvais accolement, comme fréquemment dans la dysphonie, la détection des pics de vitesse devient très difficile du fait du faible rapport signal/bruit.

3.4. Utilisation clinique de l'EGG

Quelle que soit la technique de repérage des phases du cycle glottique, cette détection et ce calcul du CQ permettent de mettre en évidence les éventuels phénomènes d'hypo adduction ($CQ < 0.4$) ou d'hyper adduction ($CQ > 0.6$) des cordes vocales. Cette technique a été notamment utilisée pour étudier les liens entre effort-forçage vocal et posture dans Lagier et al., 2010 [36].

L'autre intérêt majeur de l'EGG présenté dans Fourcin et al., 2002 [10] est sa capacité à effectuer de façon fiable la détection des cycles vibratoires sur de la parole continue. En effet, ce signal étant ciblé sur la source glottique, il est dépourvu de bruits de friction présents dans le signal acoustique et sa forme n'est pas modulée par les résonances supra-glottiques. Il est donc aisé d'obtenir un repérage des cycles vibratoires précis et de fournir des histogrammes sur une grande quantité de valeurs avec notamment la possibilité de détecter simultanément les capacités tonales et les irrégularités vibratoires [10].

4. L'évaluation vocale par aérophonométrie

4.1. Qu'est-ce que l'aérophonométrie ?

L'aérophonométrie est directement héritée des travaux de Marey et Rousselot (§ Instruments ou perception : une vieille histoire). Le principe est de capter les flux d'air mis en jeu dans la production de parole. Dès Teston, 1976 [59], des versions modernes d'aérophonomètres ont été développées. En France, une première utilisation en milieu clinique de ces techniques a été publiée dans Giovanni et al., 1991 [24]. Les avancées successives ont donné naissance par la suite au dispositif d'Evaluation Vocale Assistée (EVA©), développé par le LPL et SQLab (www.sqlab.fr). Cet appareillage permet d'enregistrer simultanément des signaux acoustiques et aérodynamiques par l'utilisation d'une pièce à main (Figure 13) contenant un microphone, un sonomètre, un pneumotachographe mesurant le débit d'air oral et nasal [17] ainsi que deux capteurs de pression (intra-orale par exemple). D'autres capteurs, comme l'électroglottographie, peuvent aussi être enregistrés en parallèle.

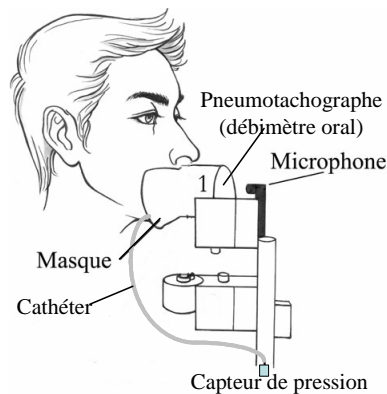


Figure 13 : Pièce à main du dispositif EVA©, permettant la mesure simultanée notamment de la voix, du débit d'air oral et de pressions (dessin réalisé par A. Michaud)

4.2. L'intérêt de l'aérophonométrie

La parole est un phénomène acoustique mais la plupart des mécanismes de production de la parole sont aérodynamiques : respiration, pression sous-glottique, transformation aéro-acoustique au niveau glottique, obstruction du flux d'air lors de l'articulation, passage vélo-pharyngé... Rappelons que le signal acoustique capté aux lèvres est le résultat d'un filtrage du signal glottique par les cavités orales et nasales, signal duquel il n'est pas toujours aisé d'identifier les mécanismes ayant donné lieu à ce résultat. Par exemple, les techniques de filtrage inverse qui consistent à retrouver la forme du signal glottique à partir du signal de parole restent controversées. En revanche, capter les flux d'air ou les pressions permet d'être plus direct. Ainsi, mesurer le débit d'air oral et mettre en évidence une fuite glottique importante dans une paralysie laryngée est plus direct que de mesurer un rapport signal/bruit faible, révélateur de la fuite. D'autre part, cela permet une certaine sélectivité, par exemple dans la distinction entre mécanisme oral vs nasal. Globalement, le débit d'air oral informe sur la quantité d'air pulsé par les poumons à travers la glotte (phonation) ou la fermeture/ouverture du conduit vocal (articulation). Le débit d'air nasal informe sur la fermeture/ouverture du voile du palais, fonction intéressante à observer dans certaines atteintes neurologiques [48]. Il permet aussi d'objectiver d'éventuelles fentes palatines ou fistules [41]. La pression sous-glottique renseigne sur la quantité d'énergie utilisée par le locuteur au niveau de sa source et sur l'éventuel hyper/hypo fonctionnement pneumophonatoire. Rappelons que la pression sous-glottique peut être estimée de façon indirecte par la "airway interrupted method" de Smitheran & Hixon, 1981 [53] 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 13). 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, bien plus facile à obtenir.

Une utilisation originale des techniques aérodynamiques a été proposée par Garrel et al., 2002 [13] : la tussométrie. En effet, la toux est un geste physiologique nécessitant une parfaite compétence laryngée. Elle peut permettre de quantifier la capacité d'occlusion du larynx notamment par la mesure du temps entre le début de la toux et le pic du débit d'air buccal (temps de montée). Les résultats de cette étude qui a porté sur des paralysies laryngées ont mis en évidence la réduction nette du temps de montée après injection de graisse dans les cordes vocales, mesure qui traduit une meilleure dynamique laryngée.

D'autres détails sur l'aérophonométrie peuvent être consultés dans Teston, 2007 [55]. Nous présentons par la suite l'utilisation conjointe de différentes méthodes.

5. L'évaluation vocale multiparamétrique

Cette partie décrit une démarche suivie depuis plusieurs années consistant à mettre en correspondance un ensemble de mesures instrumentales obtenues sur des patients dysphoniques, évalués parallèlement selon une méthode perceptive.

5.1. Historique

Plusieurs équipes ont abordé cette double évaluation perceptive et instrumentale. Dans Wolfe et al., 1995 [61], les auteurs ont étudié la combinaison de 4 mesures acoustiques (Fo 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. A l'aide d'une analyse en régression, les auteurs ont montré que 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 [62], 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. A partir de ces 4 mesures sélectionnées (F0 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 linéaire⁵. 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. En revanche, dans les études exposées par Piccirillo et al., 1998a & b [43, 44], 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 de la Timone à Marseille depuis 1991. 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) de la dysphonie est exploité. Parallèlement, des mesures instrumentales sont effectuées sur les mêmes patients à l'aide du dispositif EVA© (Figure 13) qui permet d'obtenir des mesures acoustiques primaires (F0, intensité en dB SPL), des mesures de stabilité laryngée (jitter, shimmer, coefficient de Lyapounov⁶), 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 pas 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.

La Table 3 synthétise les principaux résultats relatifs à ce travail. Plusieurs phénomènes sont observables :

- la progression globale de la concordance entre le jury d'écoute et les mesures instrumentales, amélioration que l'on peut expliquer par la prise en compte de mesures instrumentales de plus en plus pertinentes ;
- un effet de saturation des performances autour de 85% de concordance ;
- l'importance de la méthode d'évaluation perceptive, observable dans Yu et al 2002 [66] 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.

5 $DSI = 0.13 * TMP (s) + 0.0053 * Fomax (Hz) - 0.26 * Imin (dB) - 1.18 * Jitter (%) + 12.4$ (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".

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

Table 3 : Travaux effectués au CHU de la Timone sur la concordance entre jugement perceptif et mesures instrumentales

Publication	Patients	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 $f > 1\text{kHz}$ - 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		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%

5.2. Etat actuel

Yu et al. 2007 [64] détaillent les résultats les plus récents des expériences relatives à cette activité.

5.2.1. Matériels et méthode

449 enregistrements sonores ont été sélectionnés dans les données du service ORL du CHU 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).

Les instructions données aux sujets étaient de lire un texte standardisé (paragraphe de "la chèvre de M. Seguin, 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 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 [66]. 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 a montré une amélioration des performances du juge en réduisant la variabilité inter juge et en renforçant la concordance avec les mesures instrumentales.

Les mesures instrumentales ont été réalisées par l'intermédiaire du dispositif EVA©. Les instructions données aux sujets étaient de produire 3 voyelles tenues /a/ consécutives sur lesquelles étaient mesurées la F0 (en Hz), l'intensité SPL¹⁰ (en dB), le jitter factor (en %), le shimmer, le rapport signal/bruit, le débit d'air oral (en

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).

dm³/s). La pression sous-glottique était estimée de façon indirecte par la "airway interrupted method" (§ L'intérêt de l'aérophonométrie) sur des séquences /papapa/. Enfin, l'étendue vocale était mesurée en recueillant la F0 la plus haute et la F0 la plus basse que le locuteur pouvait produire après avoir reçu cette consigne de performance. De même, le temps maximal de phonation était mesuré après avoir demandé au locuteur de produire un /a/ le plus long possible suite à une longue inspiration.

5.2.2. Résultats

La pertinence des mesures et les résultats détaillés des analyses discriminantes peuvent être consultés dans Yu et al. 2007 [64]. 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 ont été introduites puis écartées une à une selon leur importance relative dans le modèle, ont été identifiés 7 paramètres pertinents pour les femmes et 6 pour les hommes dans le but de 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 Figure fournit les moyennes et écart-types des mesures instrumentales décrites précédemment en fonction du grade de dysphonie (G0 à G3) et du genre (H/F).

5.2.2.1. Cohérence et interprétation des mesures instrumentales

Nous observons une évolution cohérente des mesures instrumentales: plus la dysphonie est jugée sévère

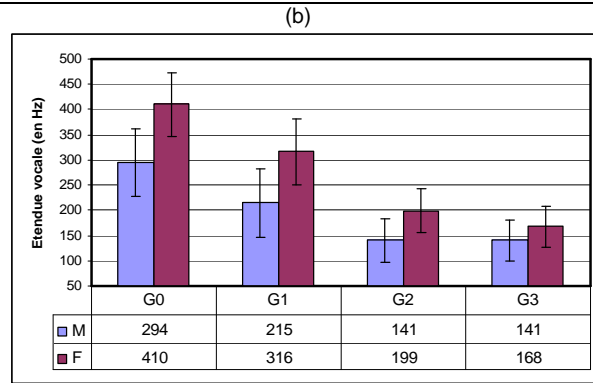
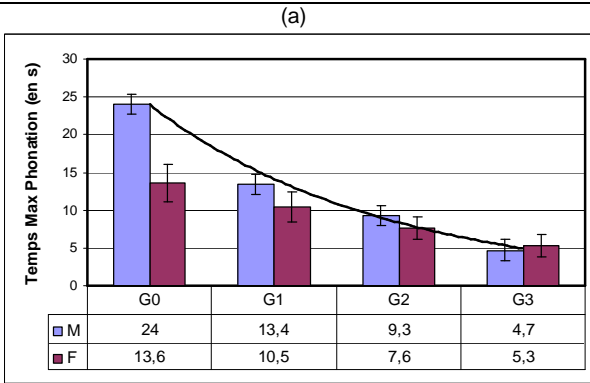
- plus le temps maximal de phonation (TMP) diminue (Figure 14.a)
- plus l'étendue vocale (Voice Range ou VR) se réduit (Figure 14.b)
- plus le débit d'air buccal (DAB) augmente, révélateur de fuite glottique (Figure 14.c)
- plus la pression sous-glottique (PSGE) augmente, révélatrice de forçage (Figure 14.d)
- plus la F0 moyenne augmente pour les hommes et diminue pour les femmes (Figure 14.e), évolution analogue aux courbes de vieillissement de la voix (Figure 5)
- plus le jitter augmente, révélateur d'instabilité laryngée (Figure 14.f)
- plus le taux d'énergie harmonique (Sr) diminue, révélateur de pauvreté harmonique et de présence de bruit (Figure 14.g)
- plus le plus grand exposant de Lyapunov (PGEL) augmente, révélateur de comportement chaotique du vibrateur laryngé (Figure 14.h)

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 :

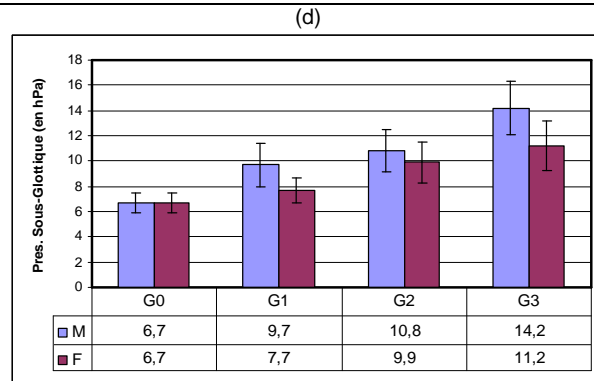
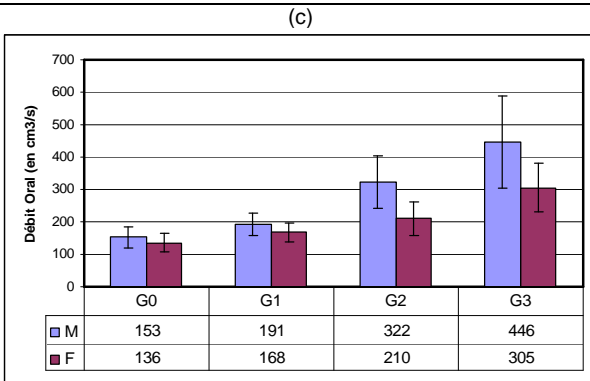
- les limites de l'espace de fonctionnement du système sont explorées par le TMP et le VR. Cet espace se réduit avec la sévérité de la dysphonie.
- l'hyperfonctionnement du système est exploré par la mesure de la PSGE. L'augmentation de cette dimension est probablement liée aux types de pathologies rencontrées induisant un comportement hyperfonctionnel.
- le contrôle de la fréquence de vibration est exploré par la mesure du jitter et du PGEL.
- le contrôle de l'amplitude de l'accolement est exploré par la mesure du DAB et du Sr.

Il s'agit d'une intensité calibrée analogue à celle fournie par un sonomètre. La comparaison d'intensité pour différents enregistrements est possible.

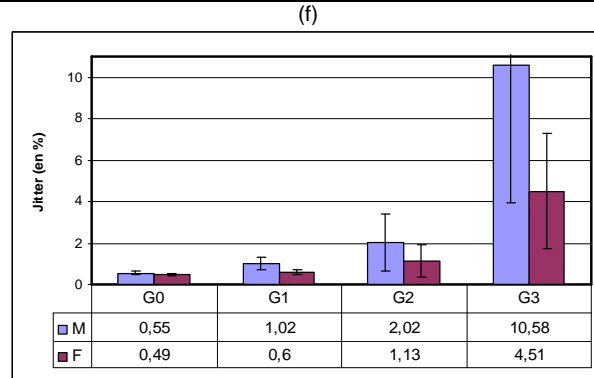
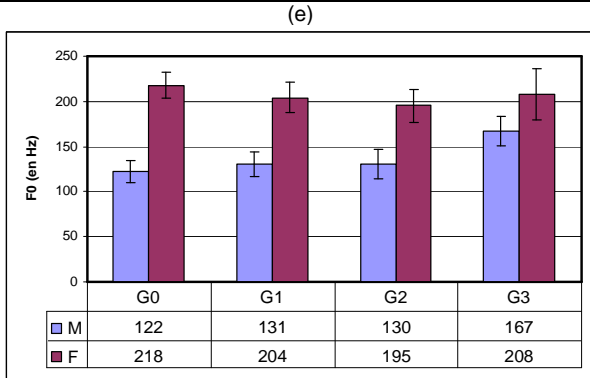
Performances Pneumophonatoires



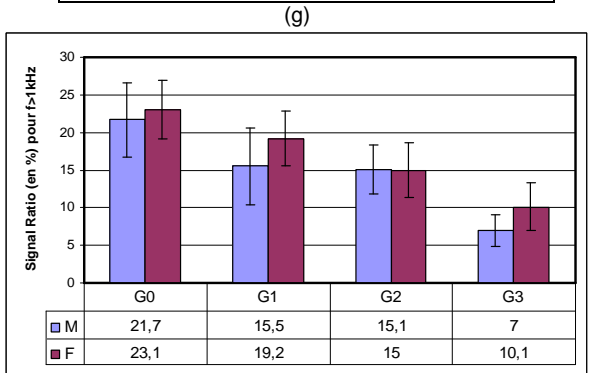
Mécanismes aérodynamiques



Mesures acoustiques classiques



Mesures de signal/bruit



Analyse acoustique non-linéaire

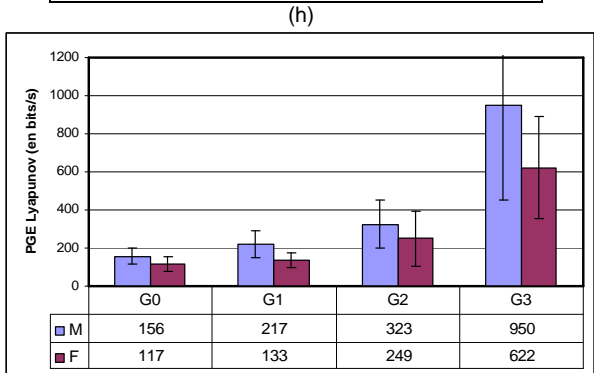


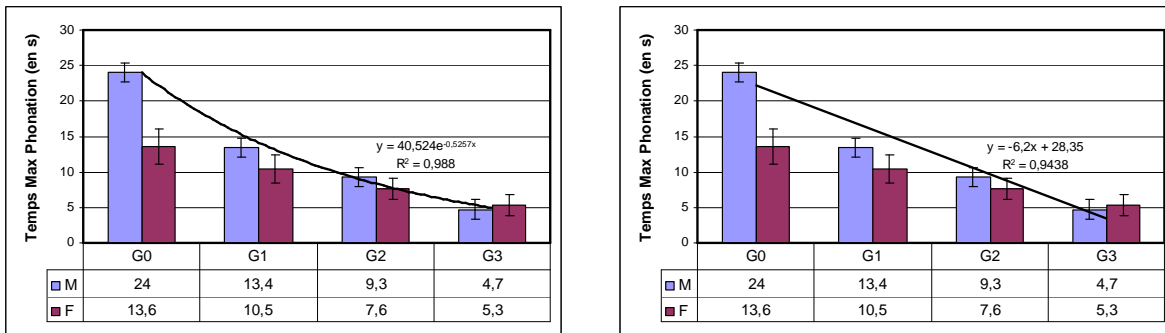
Figure 14 : Mesures instrumentales en fonction du grade de dysphonie (G0 à G3) et du genre

5.2.2.2. Vers un modèle prédictif de dysfonctionnement laryngé

La mesure de la F0 moyenne (Figure 14.e) met en évidence un comportement opposé entre les locuteurs hommes et femmes, la F0 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 ce point précédent, l'observation de la Figure 14 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 (Figure 14.a), le jitter (Figure 14.f) ou le plus grand exposant de Lyapunov (Figure 14.h) 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 [22] où les auteurs avaient effectué leur analyse statistique sur le Log(jitter).

Ces 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 [62]. En effet, cet index⁵ est une simple combinaison linéaire 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 Figure 15a) est la suivante : $TMP = 40.524 \times e^{-0.5257(G+1)}$ avec un coefficient de détermination de $R^2=0.988$.



(a) Courbe de tendance exponentielle

(b) Courbe de tendance linéaire

Figure 15 : 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 : $G_{\text{prédictif}} = \frac{3.7 - \ln(TMP)}{0.5257} - 1$

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

Les simulations exposées en Table 4 montrent qu'une modélisation non linéaire permet une meilleure détermination et prédiction.

Table 4 : Effet de la modélisation statistique de la variable 'Temps Maximal de Phonation' par rapport au grade G de dysphonie. La modélisation non linéaire permet une meilleure détermination et prédiction.

Grade	0	1	2	3	
TMP mesuré (en s)	24	13,4	9,3	4,7	
TMP avec modèle linéaire (Figure 15b)	22,2	16,0	9,8	3,6	$TMP = 28.35 - 6.2 \times (G + 1)$
Ecart modèle/mesure	-8%	19%	5%	-24%	
Grade prédit avec modèle linéaire	-0,3	1,4	2,1	2,8	$G_{\text{prédicatif}} = \frac{28.35 - TMP}{6.2} - 1$
Ecart prédiction/grade réel	-0,30	0,41	0,07	-0,19	
TMP avec modèle expon. (Figure 15a)	24,0	14,2	8,4	4,9	$TMP = 40.524 \times e^{-0.5257(G+1)}$
Ecart modèle/mesure	0%	6%	-10%	5%	
Grade prédit avec modèle expon.	0,0	1,1	1,8	3,1	$G_{\text{prédicatif}} = \frac{3.7 - \ln(TMP)}{0.5257} - 1$
Ecart prédiction/grade réel	-0,01	0,10	-0,20	0,09	

5.2.3. 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, 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 5) 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 5. Le résultat global montre que dans 82% des cas, le locuteur est classé de façon concordante entre le jugement perceptif et les mesures instrumentales.

Table 5 : Matrice de confusion entre prédiction du grade de dysphonie à partir des mesures instrumentales et jugement perceptif

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

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. 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'erreur où l'analyse instrumentale sous-estime le grade de dysphonie Yu et al., 2007 [64].

6. Bilan et perspectives de l'utilisation conjointe d'évaluation instrumentale et perceptive de la voix

6.1. Instruments et mesures

La marge de manœuvre liée aux mesures instrumentales nous semble plutôt restreinte, le matériel étant à présent utilisé de façon intensive, 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 celles décrites en § Les nouvelles tendances.

6.2. Contexte de production vocale

Il est souvent reproché aux méthodes instrumentales la nécessité de productions phonatoires "artificielles" comme des /a/ tenus, des séquences /papapa/, des glissandos. Il est souvent précisé que ce choix est lié à l'incapacité de 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 est à 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 d'instabilité laryngée (cf § L'ambiguïté contradictoire des variations de F_0).

De la même façon, une consigne de production de séquences /papapa/, de temps maximal de phonation, de glissandos 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 glissandos... Par conséquent, du fait de la non concordance des angles d'observation, il est légitime de constater une non correspondance exacte des deux évaluations.

6.3. Une approche méthodologique à revoir

La méthodologie actuelle qui consiste à considérer le jugement perceptif comme le '*Gold Standard*' et à confronter les mesures instrumentales à cette référence a probablement atteint ses limites. La restriction évidente 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 (voir § Les décalages entre mesures et perception : qui a raison ?, p.4).

L'évaluation instrumentale 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. 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 mais dans une approche non pas analytique descendante comme prévu mais globale, montante et aveugle. **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). **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. **Aveugle** ou en du moins opaque dans la mesure où les cliniciens ont demandé aux appareils 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. 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.

6.4. Bilan

Même si sa définition peut être sujette à controverse comme indiqué dans Le Huche et al., 1997 [37], la dysphonie est communément décrite comme une perturbation de l'émission du son laryngé. Nous nous sommes placés volontairement dans le cadre de cette définition "acoustique" restreinte. L'évaluation instrumentale de la dysphonie est clairement tournée vers une détermination des caractéristiques "mécaniques" du système phonatoire. En revanche, le jugement perceptif se situe 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. Nous pensons notamment à des cas de "difficultés vocales sans traduction acoustique" [37 où les instruments de mesure aérodynamique, comme celle de la pression sous-glottique estimée, peuvent mettre en évidence ces difficultés (par exemple lié à un forçage) qui ne se traduit pas nécessairement au niveau acoustique, ce phénomène pouvant passer 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 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 importantes différences entre ces approches et les échantillons vocaux utilisés, on peut finalement être surpris d'obtenir 80% de concordance 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: 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 transitoires (phase d'initiation/extinction de la vibration). Or, ces dysfonctionnements, potentiellement captés par l'évaluation instrumentale, 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 non parfaite, mais finalement assez bonne. 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 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 "sourdisse" 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 sourdisse actionnée de façon permanente ne pourra que fournir une prestation médiocre.

6.5. Perspectives

Nos perspectives font écho au paragraphe « Instruments ou perception : une vieille histoire. La phonétique moderne et nos connaissances sur les mécanismes de production de la parole ont émergé par l'utilisation de techniques instrumentales ayant permis d'obtenir des données objectives et ce, indépendamment de la perception des dits phénomènes. De nos jours, aucun phonéticien ne remettrait en cause une telle

affirmation. Parallèlement, les études sur la perception de la parole ont permis aussi de comprendre les mécanismes psycho-acoustiques et cognitifs mis en jeu dans le décodage de la parole. Ces deux activités de recherche ont évolué en parallèle et ce de façon quasi indépendante. L'approche actuelle dans l'étude de la communication parlée converge vers une approche intégrée des deux phénomènes, certains mécanismes de production ne pouvant être expliqués que par des contraintes perceptives.

Dans le cadre de l'évaluation des dysphonies, les perspectives pourraient résider dans l'utilisation de méthodes instrumentales non pas pour "confirmer" les impressions auditives laissées par le signal vocal mais pour obtenir une observation "brute" des mécanismes mis en jeu. Toute une expertise est à développer dans ce domaine avec la difficulté d'être confronté à des mesures inexplicables d'un point de vue perceptif, comme cela a été le cas en phonétique pendant des décennies. L'obtention de données normatives à grande échelle s'avère aussi primordiale. Si Yu et al., 2007 [64] ont apporté des éléments en ce sens (§ Résultats), il est nécessaire d'obtenir des valeurs normatives pour certaines tranches de locuteurs, notamment plus âgés. En effet, dans les troubles de la voix d'origine neurologique, la population est souvent âgée de plus de 50 ans et nous savons que dans ce cas, les caractéristiques vocales ou pneumophonatoires sont différentes de populations plus jeunes (Figure 5). Diverses études comme celle de Jayakumar et al. 2010 [30] montrent la nécessité d'adaptation des valeurs prototypiques en fonction des populations observées.

Parallèlement, l'étude de la perception des dysfonctionnements de la voix et de la parole est un enjeu majeur dans la mesure où la perception reste le point d'entrée principal pour considérer comme pathologique une production vocale. En l'état actuel, nous ne savons que peu de choses sur les liens entre un signal vocal perturbé et la perception de cette perturbation. Il s'agit bien d'étudier la perception d'un signal perturbé et non l'utilisation de la perception pour évaluer la perturbation, ce qui est l'approche actuelle dans le domaine. L'enjeu est complexe dans la mesure où les variations dans la voix et la parole dues à des troubles pathologiques se superposent à d'autres formes de variations normales comme des effets de contexte phonétique, prosodique mais aussi de situation de communication, de caractéristiques sociolinguistiques, ainsi que des mécanismes de compensation peu prévisible chez certains patients.

BIBLIOGRAPHIE

1. Alpan A., Maryn Y., Kacha A., Grenz F., Schoentgen J., Multi-band dysperiodicity analyses of disordered connected speech. *Speech Communication*, 2011 Vol. 53, 1, 131-141
2. Baken R.J., Orlikoff R.F., Clinical Measurement of Speech and Voice. San Diego, CA, USA : Singular Publ. Group, Inc., 2e édition, 2000, 604 p.
3. Bele I, Reliability in Perceptual Analysis of Voice Quality. *J Voice*. 2005; 19(4), 555-573.
4. Condamines R., Acoustique psycho-physique, Ed. Masson, 1985, 276 p.
5. Crevier-Buchman L, Monfrais-Pfauwadel MC, Laccourreye O., Jouffre V., Brasnu D., Laccourreye H., "La Laryngostroboscopie", *Ann Otolaryngol Chir Cervicofac*, 1993; 110: 355-357.
6. De Bodt MS, Wuyts FL, Van De Heyning PH, Crowe C., Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*, 1997; 11(1), 74-80.
7. De Jonckere P, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van de Heyning P, Remacle M, Woisard V. "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques.", *Eur Arch Otorhinolaryngol*, 2001 258:77-82
8. De Looze, C. Analyse et interprétation de l'empan temporel des variations prosodiques en français et en anglais contemporain. Thèse de Doctorat, Université de Provence, 2010.
9. Dubuisson, T., Dutoit, T., Gosselin, B., Remacle, M. On the Use of the Correlation between Acoustic Descriptors for the Normal/Pathological Voices Discrimination. *EURASIP J. Adv. Sig. Proc.*, 2009.
10. Fourcin A., McGlashan J., Blowes R.. Measuring voice in the clinic - Laryngograph® Speech Studio analyses, 6th Voice Symposium of Australia. Adelaide, 2002.
11. Fredouille C.; Pouchoulin G.; Ghio A.; Revis J.; Bonastre J.-F.; Giovanni A.. Back-and-Forth Methodology for Objective Voice Quality Assessment: From/to Expert Knowledge to/from Automatic

- Classification of Dysphonia. *EURASIP Journal on Advances in Signal Processing*, vol. 2009. 2009, 1-13.
12. Gaillard P, Billières M, Magnen C, La surdit  phonologique illustr e par une  tude de cat gorisation des voyelles fran aises per ues par les hispanophones. In: Proc. Percepci n y Realidad., Valladolid, Spain, 2007; 187-196.
 13. Garrel R.; Amy de la Bret que B.; Ghio A.; Giovanni A.; Guerrier B.. Int r t de la tussom trie dans les paralysies laryng es unilat rales. Congr s de la Soci t  Fran aise de Phoniatrie (2002, octobre : Paris, FRANCE). *Revue de laryngologie, otologie, rhinologie*, 2002, vol. 123, no. 5, 303-306.
 14. Ghio A.. Mod lisation du conduit vocal. In Auzou P.; Rolland V.; Pinto S. ; Ozsancak C. (eds.) Les dysarthries. ISBN 978-2-35327-021-7. Marseille: Solal. 2007, 140-156.
 15. Ghio , A., Merienne, S., Giovanni, A. "Influence du contexte d' coute dans la perception des dysphonies : effets de la connaissance de l' tat th rapeutique du patient sur le jugement de sa qualit  vocale", *Revue de Laryngologie*, 2011, volume 132.1, 9-17.
 16. Ghio A., Dufour S., Rouaze M., Bokanowski V., Pouchoulin G., R vis J., Giovanni A., Mise au point et  valuation d'un protocole d'apprentissage de jugement perceptif de la s v rit  de dysphonies sur de la parole naturelle, *Revue de Laryngologie*, 2011, vol 132.1, 19-27.
 17. Ghio A.; Teston B. Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. Proceedings of International Conference on Voice Physiology and Biomechanics (2004 ao t 18-20 : Marseille, FRANCE). . Marseille: Univ. M diterran e. 2004, 55-58.
 18. Giovanni A.; Revis J.; Ghio A.. Le phoniatre face   l' valuation instrumentale des dysphonies. *Revue de laryngologie, otologie, rhinologie*, 2009, vol. 130, no. 1., 3-10.
 19. Giovanni A.; Assaiaante Ch.; Galmiche A.; Vaugoyeau M.; Ouaknine M. ; Le Huche F, For age vocal et posture :  tudes exp rimentales chez le sujet sain, *Revue de laryngologie, d'otologie et de rhinologie*, 2006, vol. 127, no 5, 285-291
 20. Giovanni A, Ouaknine M, Triglia JM. Determination of the largest Lyapunov exponents of vocal signal. Applications to unilateral laryngeal paralysis. *J Voice* 1999; 13: 341-354
 21. Giovanni A, Robert D, Estublier N, Teston B, Zanaret M, Cannoni M., Objective evaluation of dysphonia: Preliminary results of a device allowing simultaneous acoustic and aerodynamic measures, *Folia Phoniatr Logop.*, 1996, 48:175-185.
 22. Giovanni A., Robert D., Teston B., Guarella M.D., Zanaret M., Etude pr liminaire des param tres acoustiques et a rodynamiques apr s laryngectomies frontales ant rieures de Tucker, *Ann. Otolaryngol. Chir. Cervicofac.*, 1996, 113, 277-284.
 23. Giovanni A., Estublier N., Robert D., Teston B., Zanaret M., Cannoni M., Evaluation vocale objective des dysphonies par la mesure simultan e des param tres acoustiques et a rodynamiques   l'aide de l'appareillage EVA, *Ann. Otolaryngol. Chir. Cervicofac.*, 1995, 112, 85-90.
 24. Giovanni A.; Molines V.; Nguyen N.; Teston B., L' valuation objective de la dysphonie: une m thode multiparam trique. Proceedings of International Congress of Phonetic Sciences (ICPhS) (12 : 1991 ao t 19-24 : Aix-en-Provence, FRANCE). 1991, p. 274-277.
 25. Godino-Llorente J.I., G mez-Vilda P., S enz-Lech n N., Blanco-Velasco M., Cruz-Rold n F., Ferrer-Ballester, M., Support Vector Machines Applied to the Detection of Voice Disorders. NOLISP 2005: 219-230
 26. Godino-Llorente J.I., Fraile R, S enz-Lech n, N, Osma-Ruiz, V., G mez-Vilda, P, Automatic detection of voice impairments from text-dependent running speech, *Biomedical Signal Processing and Control*, July 2009, 4 (3). 176-182
 27. Henrich N., d'Alessandro C., Castellengo M. and Doval B., On the use of the derivative of electroglottographic signals for characterization of nonpathological voice phonation, *Journal of the Acoustical Society of America*, 2004, 115(3), pp. 1321-1332
 28. Hess W., Pitch Determination of Speech Signals, Algorithms and Devices, Springer Series in Information Sciences, 1983, 698 p.
 29. Hiraoka, N., Kitazoe Y., Ueta H., « Harmonic-intensity analysis of normal and hoarse voices », *Journal of the Acoustical Society of America*, 1984, 76(6), 1648-1651.

30. Jayakumar T., Savithri S. R., "Effect of Geographical and Ethnic Variation on Dysphonia Severity Index: A Study of Indian Population", *J Voice*, in Press, 2010.
31. Hixon T.J.; Weismer G.; Hoit J.D. Preclinical speech science : anatomy physiology acoustics perception. San Diego, CA, USA : Plural Publishing, 2008, XII, 642 p.
32. Kacha A.; Grenez F; Schoentgen J., Estimation of dysperiodicities in disordered speech, *Speech Communication*, 2006, 48,1365-1378.
32. Kasuya H., Ogawa S., Mashima K., Ebihara S., Normalized noise energy as an acoustic measure to evaluate pathologic voice, *Journal of the Acoustical Society of America*, 1986, 80(5), 1329-1334.
33. Kent R., Ball M., Voice Quality Measurement. San Diego, CA, USA : Singular Publ. Group, Inc., 2000, 492 p.
34. Kreiman J, Gerrat B-R,, Precoda K, Berke GS, Individual differences in voice quality perception. *J Speech Hear Res.* 1992; 35(3), 512-520.
35. Kreiman J, Gerrat BR, Kempster GB, Erman A, Berke GS, Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res.* 1993; 36(1), 21-40.
36. Lagier A.; Vaugoyeau M.; Ghio A.; Legou T.; Giovanni A.; Assaiante C.. Coordination between posture and phonation in vocal effort behaviour. *Folia Phoniatica et Logopaedica*, 2010, vol. 62, 195-202.
37. Le Huche F., Allali A., La voix, Collection Phoniatrie, 1997, Masson, Paris, 199p.
38. Mc Gurk H., Mc Donald J., Hearing lips and seeing voices, *Nature*, 1976, Vol. 264, No. 5588, 746–748.
39. Meunier C., Floccia, C., Syllabe ou mot : quelle unité permet d'identifier les catégories phonétiques?, Actes des 2èmes Journées d'Etudes Linguistiques, "Syllabes", Nantes, 1999, 87-92.
40. Morsomme D., Estienne F., Bilan Vocal, In Les Bilans de Langage et de Voix, Fondements théoriques et pratiques, Masson, 2006, 266p.
41. Pech C, Lognos C., Cassassolles S., Bardot J., Magalon G., Résultats phoniatriques après pharyngoplastie, Actes du congrès sur le Voile Pathologique, Société Française de phoniatrie, Lyon, 1995,.43-51.
42. Peterson, G. E., Barney, H. L., Control methods used in a study of the vowels. *Journal of the Acoustic Society of America*, 1952, 24, 175–184.
43. Piccirello J, Colin P, Dennis F, Frederickson J. Multivariate analysis of objective vocal function. *Ann Otol Rhinol Laryngol.* 1998;107:107-112.
44. Piccirillo J, Colin P, Dennis F. Assessment of two objective voice function indices. *Ann Otol Rhinol Laryngol.* 1998;107:396-400.
45. Scherer R., Vail V. et Rockwell B., Examination of the laryngeal Adduction Measure EGGW, NCVS status and Progress Report, 1993, 73-82.
46. Revis J., L'analyse perceptive des dysphonies" In Giovanni A., "Le bilan d'une dysphonie : État actuel et perspectives, Solal, Marseille, 2004, 244 p.
47. Revis J, Giovanni A, Wuyts F, Triglia J, Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop.* 1999; 51(3), 108-116.
48. Robert D., Sangla I., Azulay J.P., Giovanni A., Cannoni M., Pouget J., Diagnostic et suivi de l'insuffisance vélaire dans les formes bulbaires des maladies du motoneurone, Actes du congrès sur le Voile Pathologique, Société Française de phoniatrie, Lyon, 1995, 63-74 .
49. Rossignol C., Classifications internationales des altérations corporelles, dysfonctionnements et handicaps. Pour une clarification des concepts. Actes, Entretiens de Bichat, Orthophonie (2007 septembre 28-29 : Paris, FRANCE). In Entretiens de Bichat, Entretiens d'orthophonie. ISBN 2-7046-1725-2. Paris: Expansion Formation Editions. 2007, 62-69.
50. Roy N, Bless DM, Heisey D. Personality and voice disorders: a multitrail – multidisorder analysis. *J Voice*, 2000, 14 : 521-548.

51. Saenz-Lechon, N., Godino-Llorente, J., Osma-Ruiz, V., Gomez-Vilda, P. Methodological issues in the development of automatic systems for voice pathology detection. *Biomed. Signal Process. Contr.*, 2006, 1 (2), 120-128.
52. Shrivastav R, Multidimensional Scaling of Breathly Voice Quality: Individual Differences in Perception. *Journal of Voice*. 2006; 20(2), 211-222.
53. Smitheran J., Hixon T., A Clinical method for estimating laryngeal airway resistance during vowel production, *J.Speech Hear Dis*, 1981, 46, 138-146.
54. Titze I.R., Liang H., Comparison of F0 Extraction Methods for High-Precision Voice Perturbation Measurements, *J Speech Hear Res*, 1983, 36:1120-1133.
55. Teston B., L'étude instrumentale des gestes dans la production de la parole: Importance de l'aérophonométrie. In P. Auzou; V. Rolland-Monnoury, S. Pinto, C. Ozsancak (ed.) *Les Dysarthries*. Marseille: SOLAL. 2007, p. 115-117.
56. Teston B., L'évaluation instrumentale des dysphonies : État actuel et perspectives In Giovanni A., "Le bilan d'une dysphonie : État actuel et perspectives", Solal, Marseille, 2004, 244 p.
57. Teston B., L'œuvre d'Etienne-Jules Marey et sa contribution à l'émergence de la phonétique dans les sciences du langage, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)* 23 (2004) 237-266.
58. Teston, B., Galindo B., A diagnostic and rehabilitation aid workstation for speech and voice pathologies. Proceedings of European Conference on Speech Communication and Technology (Eurospeech 1995) (4 : 1995 : Madrid, Spain), 1995, p. 1883-1886.
59. Teston B., Etude et réalisation d'un système d'analyse des paramètres articulatoires : le polyphonomètre. *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA)*, 1976, vol. 3, p. 151-207.
60. Woisard V., Bodin S., Puech M. The Voice Handicap Index : impact of the translation in French on the validation, *Rev Laryngol Otol Rhinol*, 2004, 125(5), 307-312.
61. Wolfe V, Fitch J, Cornell R. Acoustic prediction of severity in commonly occurring voice problems *J Speech Hear Res* 1995; 38:273-279.
62. Wuyts FL, De Bodt MS, Molenberghs G, Remacle M, Heylen L, Millet B, Lierde KV, Raes J, Van De Heyning PH., The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach, *J Speech Hear Res.*, 2000, 43:796-809.
63. Wuyts FL, De Bodt MS, Van De Heyning PH, Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice*. 1999; 13(4), 508-517.
64. Yu P., Garrel R., Nicollas R., Ouaknine M., Giovanni A., Objective voice analysis in dysphonic patients. New data including non linear measurements, *Folia Phoniatr et Logop*, 2007, 59:20-30.
65. Yu P., Ouaknine M., Revis J., Giovanni A, Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements, *J Voice*, 2001, Vol.15 (4), p.529-542.
66. Yu P, Revis J, Wuyts FL, Zanaret M, Giovanni A., Correlation of instrumental voice evaluation with perceptual analysis using a modified visual analogic scale., *Folia Phoniatr Logop* 2002; 54: 271-281.
67. Yu P, Ouaknine M, Giovanni A. Clinical significance of calculating the coefficients of Lyapunov in the objective assessment of dysphonia. *Rev Laryngol Otol Rhinol*. 2000;121:301-305.
68. Yumoto E, W.Gould, Harmonics-to-noise ratio as an index of the degree of hoarseness, *J.A.S.A*, 1982, 71(6), 1544-1550.