

Joana REVIS
Orthophoniste

*L'analyse perceptive des dysphonies :
approche phonétique de l'évaluation vocale*

Thèse de l'Université de la Méditerranée
Sous la direction de : P^r Antoine GIOVANNI

Laboratoire d'Audio-Phonologie Clinique
(UPRES-EA 2668)
UNIVERSITE DE LA MEDITERRANEE
Marseille, 2004

UNIVERSITÉ de la MÉDITERRANÉE – AIX MARSEILLE II
Faculté de médecine

THESE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE DE LA MEDITERRANEE

Spécialité : Pathologies du langage

*L'analyse perceptive des dysphonies :
approche phonétique de l'évaluation vocale*

Présentée et soutenue publiquement par

Joana Revis

Le Mardi 26 Octobre 2004

Jury :

Monsieur le Professeur Antoine GIOVANNI, Université de la Méditerranée, Directeur
Monsieur le Professeur Jean-Michel TRIGLIA, Université de la Méditerranée, Président
Monsieur le Professeur Philippe DEJONCKERE, Université d'Utrecht, Rapporteur
Monsieur le Professeur Noël NGUYEN, Université de Provence, Rapporteur
Madame le Docteur Lise CREVIER-BUCHMAN, Hôpital Georges Pompidou

Table des matières

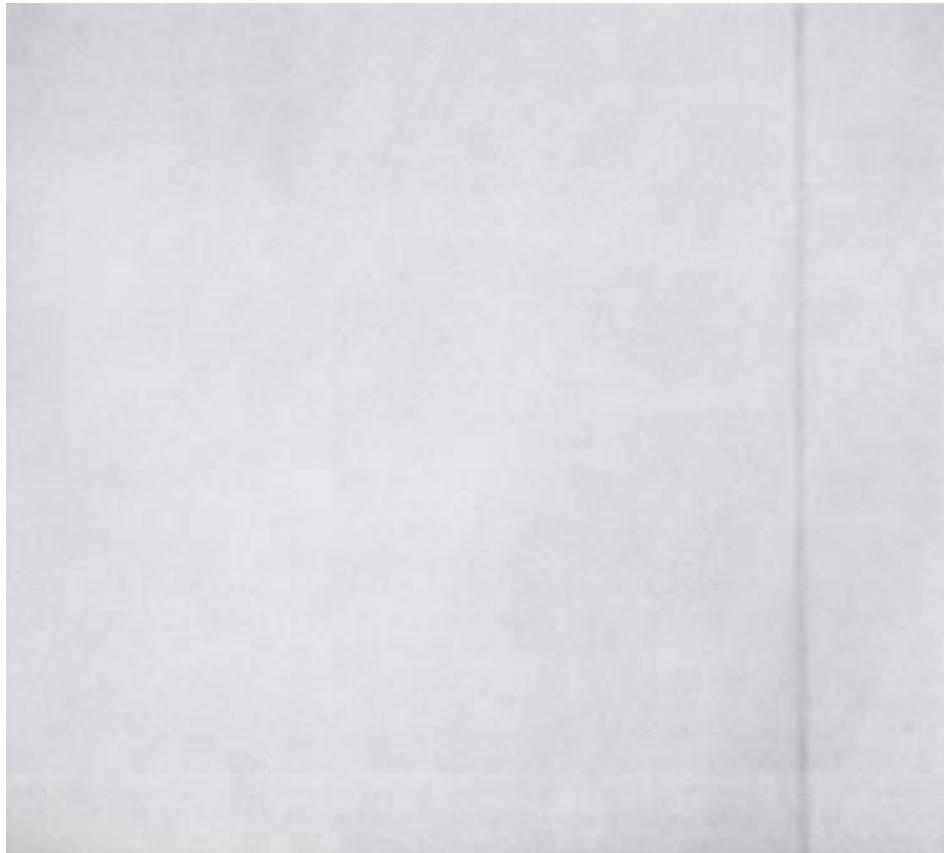
Table des matières

Préambule	9
Introduction générale	12
La philosophie de la perception	17
1. Perception et sensation	19
2. Crédibilité de la perception	20
3. Fiabilité de la perception	24
4. Perception et jugement	27
La physiologie de la perception	32
1. Anatomie de l'oreille	34
2. Physiologie de l'oreille	36
3. Les voies nerveuses de la perception auditive	38
4. Le codage de la fréquence	40
5. Perception des sons simples	41
6. Perception des sons complexes	43
La cognition de la perception	46
1. L'attention sélective	47
2. L'effet « cocktail party »	48
3. Cortex de « voix sélective »	50
La physiologie de la phonation	55
1. Anatomie du larynx	57
2. La vibration cordale	61
3. Le réglage de la hauteur	65
4. Le réglage de l'intensité	66
5. Production vocale et dysfonctionnement	67
6. La transformation du son en voyelles	68
7. Le contrôle nerveux de la production vocale	70

La perception de la dysphonie	72
1. Principes et méthodes de l'analyse perceptive	74
1.1. Les dimensions qualitatives de la perception de la voix	74
1.2. Les échelles quantitatives d'évaluation	76
1.3. Les différentes méthodes d'évaluation perceptive : premiers essais de standardisation	78
2. Le jury d'écoute	82
2.1. Le choix du jury	82
2.2. Les modèles perceptifs	84
2.3. Les références auditives du jury	85
2.4. L'ancrage auditif	87
2.5. La formation du jury d'écoute	89
2.6. L'ancrage par comparaison	92
3. Déroulement des sessions d'écoute	92
3.1. Nombre d'évaluations	93
3.2. Mode de passation	94
3.3. Présentation des stimuli sonores	95
4. Le matériau phonétique	96
4.1. La parole	97
4.2. La voyelle	97
4.3. Choix du matériau phonétique	98
 Section expérimentale	102
1. Protocole expérimental	103
1.1. Les sujets étudiés	103
1.2. L'enregistrement de la voix	103
1.3. Le matériau phonétique	104
1.3.1. Echantillon de parole	104
1.3.2. La voyelle tenue	104
1.4. L'analyse perceptive	105
2. Présentation des travaux personnels	113
a. <i>Etude préliminaire de la pertinence des mesures acoustiques et aérodynamiques pour l'évaluation de la voix après microchirurgie laryngée : à propos de 27 cas</i>	115
b. <i>Mesures objectives acoustiques et aérodynamiques de l'évolution de la voix après microchirurgie laryngée</i>	120

c. <i>Elaboration d'un protocole d'entraînement à l'analyse perceptive pour un jury inexpérimenté</i>	128
d. <i>Logiciel de CLassement de l'Analyse Perceptive (CLAP) : validité clinique et influence sur la variabilité du jury</i>	143
e. <i>Analyse objective de la dysphonie : protocole multiparamétrique incluant des mesures acoustiques et aérodynamiques</i>	152
f. <i>Corrélation entre l'analyse instrumentale et l'évaluation perceptive : utilisation d'une échelle visuelle analogique modifiée</i>	163
g. <i>Comparaison de différents matériaux phonétiques pour l'analyse perceptive des dysphonies</i>	176
h. <i>Influence de l'attaque sur l'analyse perceptive des dysphonies</i>	186
i. <i>Définition d'une mesure temporelle de l'attaque vocale</i>	191
j. « <i>Phonetic labelling</i> » de la dysphonie : une nouvelle perspective dans l'analyse perceptive	202
k. <i>Reproductibilité et validation de la méthode de «<i>Phonetic labelling</i>»</i>	217
l. <i>Influence de l'attaque dans la perception de la dysphonie au cours de la parole</i>	227
Discussion générale et perspectives	238
Conclusion générale	264
Bibliographie	272

« The Voice » Barnett Newman -1950. New York, Museum of Modern Art



« En 1950, lors d'une exposition, Barnett Newman fait la connaissance d'une baronne Suédoise qui l'impressionne par son récit d'un voyage en Laponie. Le paysage enneigé est nu et blanc, et le silence est total, lorsqu'une nuée d'oiseaux sauvages traverse le ciel remplissant l'espace de son cri. La baronne rapporte à Newman que le zip dans ses tableaux lui donne le même sentiment intense que ces sons criards, et Newman trouve cette réflexion extrêmement pertinente. Le zip chez Newman a certainement une fonction musicale : il donne une qualité rythmique au tableau et court à travers le paysage comme une ligne mélodique par expansion et contraction pulsionnelle. "The Voice" (1950) est un tableau monochrome blanc sur blanc. La sonorité y triomphe maintenant et le zip, quasi invisible comme "le bord d'un lac gelé", organise cette sonorité blanchissime s'intensifiant dans le filet mince d'une voix. Voix blanche, voix colorée si le blanc est une couleur. »

Herman Parret*

* Parret H. La voix et son temps. Bruxelles : De Boeck Université, 2002

Préambule

La voix *est* parce qu'elle est émise et perçue. Emise par le locuteur, la voix révèle l'identité de l'individu. Perçue par l'auditeur, elle est soumise au jugement qualitatif. Ce critère de qualité interroge la pensée dans l'Histoire. Philosophes, médecins, scientifiques, personnalités politiques, artistes, pédagogues de la voix, linguistes, phonéticiens et autres, se sont penchés sur les caractéristiques de la voix. Leur approche a permis de définir un continuum couvrant l'étendue de l'esthétique vocale, de la voix idéale à la voix pathologique. La connaissance de ce continuum s'est construite à travers l'élaboration des données normatives [Laver, 1980, 1981 ; Kreiman, 1992 ; Crevier-Buchman, 2002 ; Munoz, 2003] et la définition d'un cadre méthodologique de l'évaluation de la voix [Hirano, 1981 ; Dejonckere, 1996, 1998 ; Cordes, 1994].

De la même façon qu'on ne sépare pas voix et qualité de la voix, on ne peut séparer le concept de voix idéale du concept du discours idéal [Russ, 2000]. Dans la civilisation grecque (IV^{ème} siècle avant J-C), le concept de voix idéale opposait Platon attaché à la substance du discours, aux sophistes partisans de la rhétorique. Platon, fervent disciple de Socrate, dénonçait les mécanismes de la rhétorique où se confondaient opinion et vérité. Ses dialogues étaient l'occasion de véritables joutes oratoires d'une ironie corrosive dirigée contre les sophistes. « Le Protagoras » en donne une illustration évocatrice. C'est le concept du discours idéal.

Dans un ouvrage synthétique sur l'évaluation de la voix, Kent et Ball rappellent que le débat est toujours d'actualité en évoquant la pensée de Hoffman [Kent, 2000]. Celui-ci attribue la qualité du bon discours à la qualité de la voix et soutient qu'un bon orateur doit être un homme de qualité [Hoffman, 1940]. Plus récemment, Hollien décrit les techniques complexes qui sous-tendent chez le chanteur la voix idéale qu'il qualifie de voix « en or » [Hollien, 1993]. Dans le domaine artistique de la voix chantée, la technique et l'esthétique sont intimement liées. Or il existe autant de conceptions du Beau dans la voix que de spécialistes de la voix. La définition de la voix idéale ne fait donc pas l'objet d'un consensus.

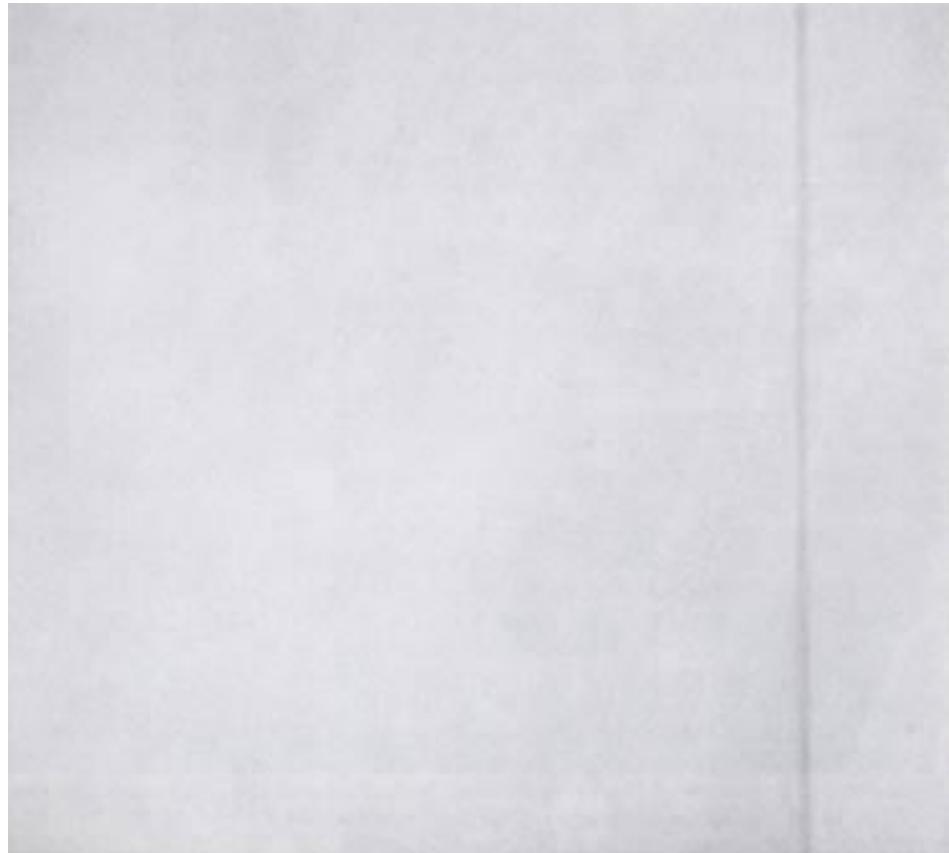
Dans la première moitié du XX^{ème} siècle, les spécialistes de la voix préfèrent le terme de voix supérieure à celui de voix idéale. Selon Anderson, la voix « supérieure » possède « *une intensité adaptée au milieu, un timbre clair et pur, une fréquence fondamentale efficace et plaisante, de la qualité vibratoire, de l'aisance et de la clarté dans la diction, (...)* » [Anderson, 1942]. La recherche s'est développée dans la qualification et la quantification de ces termes [Black, 1942 ; Cowan, 1936 ; Lynch, 1934 ; Murray, 1934 ; Pronovost, 1943 ; Snidecor, 1942 ; Talley, 1937]. En 1934, Lynch présente une étude dont l'objectif est de définir les caractéristiques de la voix supérieure [Lynch, 1934]. Il compare les mesures de fréquence fondamentale, étendue vocale, et débit de parole chez

deux groupes de locuteurs : entraînés (comédiens) et candides. Les résultats montrent que les locuteurs entraînés présentent une fréquence fondamentale plus élevée, une étendue vocale plus large, et un débit articulatoire ralenti. Lynch ne précise pas si le groupe de sujets entraînés est perçu, à l'oreille, comme supérieur par rapport au groupe candide. Une étude menée par Pronovost vient confirmer ces résultats [Pronovost, 1943]. Cowan ajoute que les voix supérieures ont une intensité plus élevée et un schéma prosodique plus marqué par rapport aux locuteurs ordinaires [Cowan, 1936]. Murray abonde dans le même sens [Murray, 1934]. Au total, la voix supérieure est caractérisée par une fréquence et une intensité élevées, un débit de parole ralenti, et une prosodie accentuée.

Ces vingt dernières années, les concepts de voix idéales ou supérieures ont été abandonnés. La recherche aborde la question de la voix sous deux angles complémentaires : la définition de normes (physiques, acoustiques ou perceptives) et l'étude de la voix pathologique, *la dysphonie* [Kent, 1987]. Mathieson définit la voix normale comme étant adaptée, confortable et efficace [Mathieson, 2000]. La voix doit correspondre au sexe et à l'âge du locuteur, permettre l'expression des émotions et conserver efficacité et confort de production quelles que soient les circonstances. Laver fournit une description complète des registres (modal, creak, falsetto, ...) et différents timbres (chuchotement, raucité, souffle, diplophonie ...) qui définissent la voix normale ou pathologique [Laver, 1980]. Il structure le tableau descriptif de la voix normale et le confronte aux données fournies par l'étude des voix pathologiques [Laver, 1981]. Les caractéristiques phonétiques ainsi établies ouvrent la porte sur une recherche fructueuse et abondante, qui vont servir de base à la conception d'une évaluation perceptive de la dysphonie [Hirano, 1981 ; Dejonckere, 1993, 2001 ; Laver, 1981].

Le travail que nous présentons s'inscrit dans la lignée de plusieurs siècles de recherches menées sur la voix, en tant que support du discours, véhicule du message et des émotions.

« The Voice » Barnett Newman -1950. New York, Museum of Modern Art



Introduction générale

Introduction générale

La voix est le support privilégié de la communication. Chaque structure communicationnelle fonctionne selon le principe de l'émetteur et du récepteur. Dans la communication orale, l'émetteur est le locuteur qui émet un message sonore par l'intermédiaire de sa voix. Le récepteur est l'auditeur qui reçoit le message par l'intermédiaire de l'oreille. Parret [Parret, 2002] souligne qu' « *Il n'y aurait pas de voix sans qu'il y ait oreille. La voix, en toute sonorité, n'est que dans la saisie par l'oreille.* » En effet, en l'absence d'une éducation spécifique, les patients sourds, atteints de déficits auditifs sévères, privilégient d'instinct la communication gestuelle. La voix est donc indissociable de l'oreille si l'on considère qu'en cas de défaillance de l'appareil auditif, la boucle audio-phonatoire est remplacée spontanément par une boucle visuo-gestuelle : en l'absence d'oreille, la voix est remplacée par le geste. La voix est donc, par essence, un phénomène perceptif ; la perception auditive est l'outil premier de l'évaluation de la voix.

La dysphonie est une altération de la voix qui résulte d'une dysfonction de production ou d'une lésion organique. La plainte du patient dysphonique est à la fois fonctionnelle (forçage vocal, douleurs musculaires, picotements laryngés, etc.) et esthétique (enrouement, voix aggravée, périodes d'aphonie, etc.). Mais c'est avant tout un trouble de la communication qui incite les patients à consulter un spécialiste. En effet, les sujets dysphoniques souffrent d'un manque d'efficacité dans la communication et se sentent diminués dans leurs capacités d'interaction sociale. A l'heure où la communication devient l'instrument indispensable d'une société basée sur l'échange de l'information, la dysphonie est une source de marginalisation. La prise en charge de la dysphonie peut être une thérapeutique sociale (certains patients éprouvent de graves difficultés dans leur travail : professeurs, standardistes, commerciaux, etc.) mais intervient également dans le cadre de maladies neurologiques lourdes comme la SLA où la communication des patients est en grave péril. Cette prise en charge fait intervenir des praticiens spécialistes de la voix : les ORL, les phoniatres et les orthophonistes. Leur choix thérapeutique est guidé par des évaluations du trouble réalisées au moment de la prise en charge, puis régulièrement au cours du traitement.

La perception auditive est l'outil premier de l'évaluation de la dysphonie. La pathologie laryngée se manifeste par une altération de la voix qui est entendue par le patient et son entourage. La perception est donc devenue tout naturellement un outil d'analyse pour les thérapeutes de la voix. En pratique clinique, l'analyse perceptive permet d'évaluer les résultats de la thérapeutique. En recherche, elle est utilisée comme la référence contre laquelle les mesures instrumentales sont testées. La perception auditive est donc considérée de fait comme un outil de mesure scientifique.

Que peut-on espérer d'une mesure établie à partir de nos perceptions ? Peut-on accorder une validité scientifique à une telle évaluation ? Dans la première section de ce travail, intitulée « **la philosophie de la perception** », nous tenterons de répondre, sous l'éclairage de la pensée des philosophes, aux questions fondamentales qui conditionnent la légitimité de la perception en tant qu'outil de mesure scientifique. Le jugement scientifique s'appuie sur du quantifiable. Son principe se résume en trois mots : mesurabilité, reproductibilité, objectivité. La perception est la capacité d'interprétation des sensations fournies par nos cinq sens. Ses caractéristiques sont strictement opposées au dictat scientifique : la perception est interprétative, variable et subjective. Cependant, la perception est pour l'individu l'unique moyen d'accéder au monde et d'en avoir une représentation consciente. Accorder notre confiance à la perception n'est donc pas un choix mais une nécessité. Nous verrons que l'apport conceptuel des théories philosophiques permet d'affirmer la légitimité de la perception.

La perception est l'interprétation des sensations. Dans l'étude des sensations, la question de la subjectivité ne se pose pas, mais la question de la physiologie. Les sensations sont le résultat d'un fonctionnement bio-mécanique. La sensation auditive est transmise par l'intermédiaire de l'oreille, dont le fonctionnement complexe et astucieux fascine depuis les développements jusqu'aux audiologues. La section suivante, « **la physiologie de la perception** » s'intéresse à l'anatomie et à la physiologie de l'oreille depuis l'oreille externe jusqu'aux voies nerveuses qui conduisent l'information jusqu'au cerveau où la sensation auditive devient perception des sons. Nous aborderons également les concepts physiques du traitement des informations auditives par l'étude de la perception des caractéristiques acoustiques des sons simples et complexes.

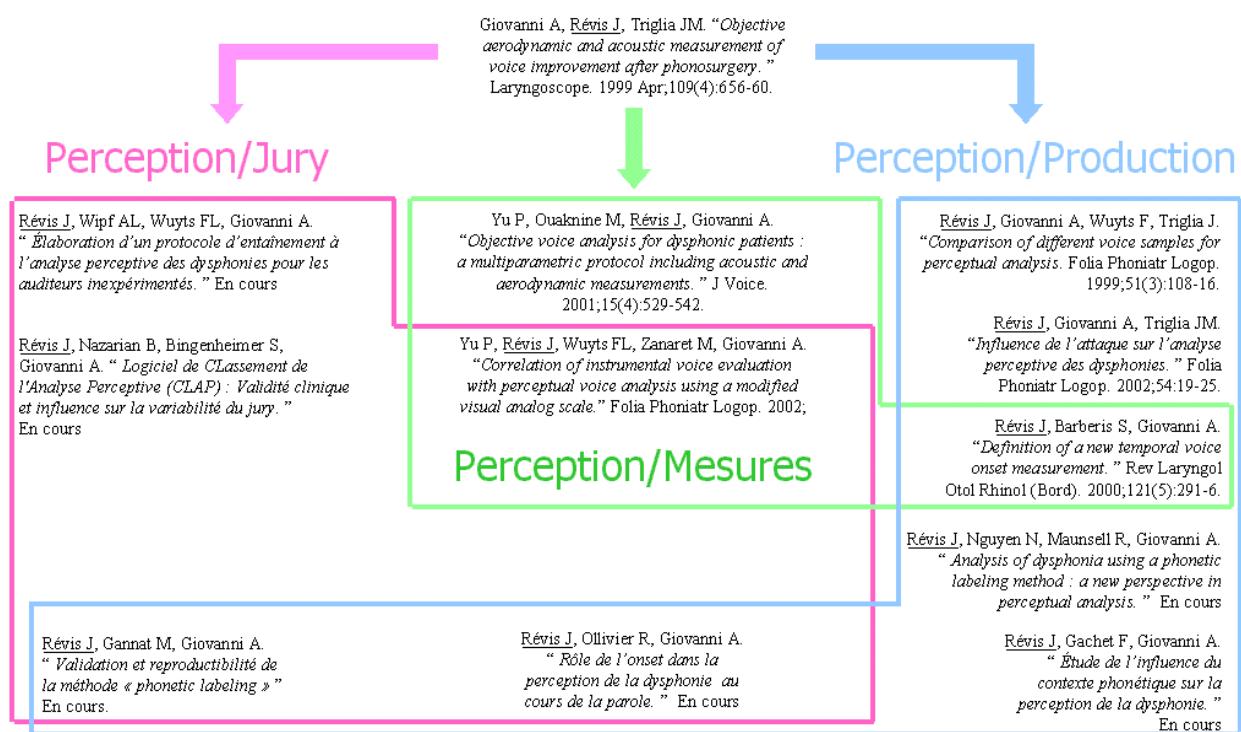
Comment l'interprétation des sons s'organise-t-elle au niveau cérébral ? Dans la section suivante, « **la cognition de la perception** », nous aborderons les processus intellectuels et les structures cérébrales qui sous-tendent la perception consciente de notre environnement sonore. En effet, rien ne semble plus naturel et simple que « d'entendre ». Néanmoins, l'immense richesse et la quantité colossale de stimulations sonores auxquelles nous sommes soumis, témoignent de la complexité des tâches cognitives qui permettent le traitement de l'information auditive.

Enfin, nous aborderons « **la physiologie de la phonation** » afin de clarifier le fonctionnement du larynx, en tant que « vibrateur » responsable de la production de la voix. Nous verrons que perception et production sont deux notions intimement liées, au point de s'influencer l'une l'autre par l'intermédiaire de la boucle audio-phonatoire.

Ces quatre sections préliminaires ont pour but d'explorer les différents axes d'étude permettant d'aborder la perception de la dysphonie. Notre objectif est d'ouvrir la réflexion sur la complexité de la perception, et de fournir des réponses préalables aux questions fondamentales pouvant survenir dans

les chapitres suivants. La section suivante, « la perception de la dysphonie » aborde plus directement les mécanismes et les méthodes de la perception de la dysphonie. Avant d'aborder la section expérimentale, il semble nécessaire de faire le point sur les connaissances théoriques qui sous-tendent la perception de la dysphonie et d'évoquer les différentes méthodologies mises en place. Cette section constitue donc une revue de la littérature sur les principes et les procédures de l'analyse perceptive des dysphonies.

La « section expérimentale », regroupe nos différentes recherches, publiées, soumises à publication ou en cours, dont l'organigramme suivant donne un aperçu.



L'organisation de la présentation des travaux prendra un départ chronologique, puisque notre première publication « Mesures acoustiques et aérodynamiques de l'amélioration vocale après microchirurgie laryngée »¹ a soulevé un certain nombre de questions qui ont guidé la suite de nos recherches. Ces questions peuvent être résumées en trois axes : comment le jury perçoit-il la dysphonie ? (axe rose) ; comment optimiser l'analyse perceptive pour tester les mesures instrumentales ? (axe vert) ; comment la production influence-t-elle la perception de la dysphonie ? (axe bleu). Les travaux suivants seront donc présentés en fonction de leur(s) axe(s) privilégié(s).

¹ Giovanni A, Révis J, Triglia JM. "Objective aerodynamic and acoustic measurement of voice improvement after phonosurgery." Laryngoscope. 1999 Apr;109(4):656-60

Enfin, en dernière section de l'ouvrage, nous proposons une « **discussion générale** » des méthodes et des concepts manipulés tout au long de ce travail. Notre dernière étude en cours permettra de boucler notre parcours scientifique, montrant le lien indissociable existant entre la perception et la production, tout en ouvrant de nombreuses perspectives de recherche.

La perception est un axe incontournable dans la démarche clinico-expérimentale de l'évaluation des dysphonies. Son étude est particulièrement passionnante, car elle renvoie, au delà de l'approche scientifique qui est notre objet, à des concepts aussi variés que la philosophie, la physiologie, la cognition et même la psychologie. On entend souvent dire que la voix est le reflet de l'âme. Depuis la bible, où l'étude de la nature de la voix permet la distinction entre l'énonciation divine et l'énonciation adamique, jusqu'à la biographie de Maria Callas², où l'on fait état de la technicité artistique, en passant par l'art graphique contemporain, la voix est une composante intégrale du monde culturel philosophique, artistique, littéraire, sur lequel elle entretient une fascination ambiguë.

Pour clore cette introduction, cette citation de Luria³ illustre la richesse et la complexité de la perception ainsi que le côté « instinctif » de la qualification de la voix :

« Quelle voix « farineuse », jaune vous avez ! Vous savez qu'il y a des gens qui ont plusieurs voix, dont les voix semblent une composition entière, un bouquet. Le vieux Eisenstein avait une telle voix : en l'écoutant, c'était comme si une flamme aux fibres saillantes avançait dans ma direction. J'étais si intensément intéressé à sa voix que je ne savais plus ce qu'il disait... D'ailleurs, je ne peux éviter de voir des couleurs quand j'entends des sons, et ce qui me frappe avant tout c'est la couleur de la voix de quelqu'un. »

² PJ Remy. Callas, une vie. Paris : Ramsay, 1978

³ AR Luria. The mind of a mnemonist. New York : Basic Books, 1968



« Le Penseur » A Rodin

La philosophie de la perception

La philosophie de la perception

La *perception* permet l'interprétation et la représentation consciente du monde d'après les informations fournies par nos cinq sens : la vue, l'ouïe, l'odorat, le toucher et le goût. Par l'intermédiaire des organes perceptifs (respectivement l'œil, l'oreille, le nez, la peau et la langue), le cerveau perçoit consciemment, analyse et interprète les objets ou les phénomènes extérieurs. La perception fait donc appel aux sensations.

Dans la culture populaire, le terme de sensation est souvent associé à la subjectivité. La perception semble individuelle et susceptible de varier en fonction de la personnalité de chacun. En effet, nous ne sommes et ne pouvons jamais être sûrs d'entendre ou de voir les objets, de les ressentir de manière strictement identique à autrui. Nous nous construisons donc au sein d'un monde commun qui nous demeure personnel. Or, en tant que lien unique et particulier entre l'individu et le monde, la perception se pose en référence de toute analyse et de tout concept.

En ce qui nous concerne, l'audition est le canal spécifique de la sensation sonore et donc par extension de la voix et de la pathologie dysphonique. En tant que telle, la perception auditive représente l'outil d'analyse de la dysphonie par excellence, et reste considérée comme la référence contre laquelle les mesures instrumentales sont testées. Or, notre perception est-elle compétente en tant qu'outil d'analyse ? Quelle confiance peut-on accorder à nos sens ? Le jugement perceptif qualitatif est-il légitime ? Peut-on concilier science et perception ? Ces interrogations sous-tendent et influencent la conception de protocoles expérimentaux d'analyse perceptive. Ce chapitre constitue une réflexion globale qui nous a paru intéressante en préambule à nos travaux centrés sur l'analyse perceptive de la dysphonie.

« *Nous croyons très bien savoir ce que c'est que "voir", "entendre", "sentir", parce que depuis longtemps la perception nous a donné des objets colorés ou sonores. Quand nous voulons l'analyser, nous transportons ces objets dans la conscience. Nous commettons ce que les psychologues appellent "l'experience error", c'est à dire que nous supposons d'emblée dans notre conscience des choses que nous savons être dans les choses. Nous faisons de la perception avec du perçu. Et comme le perçu lui-même n'est évidemment accessible qu'à travers la perception, nous ne comprenons finalement ni l'un ni l'autre. »*

[Merleau-Ponty, 1945]

1. Perception et sensation

En commençant l'étude de la perception, nous trouvons dans le langage la notion de sensation qui paraît immédiate et claire : je sens du chaud, du froid, du rouge, du bleu. Nous pourrions d'abord entendre par sensation la manière dont nous sommes affectés et le constat d'un état de nous-même. Par exemple, le sombre des yeux fermés qui nous entoure sans distance ou les sons du demi-sommeil que l'on entend sans écouter indiquerait ce que peut être le pur « sentir ». Mais ce serait considérer que la sensation est indépendante de tout concept objectif. Au lieu de ça, la sensation s'applique au monde et ses objets, qui ont des caractéristiques physiques. Voir, c'est considérer des couleurs et des lumières, entendre c'est considérer des sons avec leur fréquence et leur intensité, et, d'une manière générale, sentir, c'est considérer des qualités.

« Lorsque je flaire une rose, il y a à la fois sensation et perception dans cette opération. L'odeur agréable que je sens, considérée en elle-même, et sans aucune relation à la rose, est la sensation proprement dite ; elle m'affecte d'une certaine manière, et cette affection n'a rien de commun avec la rose, ni avec aucun autre objet. La sensation est uniquement ce que j'éprouve ; son essence consiste à être sentie ; quand elle cesse de l'être, elle n'est plus ; en un mot, il n'y a aucune différence entre la sensation et ce que sent l'esprit qui en est affecté. C'est pour cela (...) que le fait de sensation ne renferme point d'objet distinct de l'acte de l'esprit qui sent, et cela est vrai de toutes les sensations possibles. Examinons maintenant la perception que nous avons en flairant une rose. L'expérience m'apprend que la sensation est excitée par la présence de la rose, et qu'elle s'évanouit lorsque la rose s'éloigne ; les principes de la nature me font conclure de là qu'il y a dans la rose une qualité qui est la cause de la sensation. Cette qualité de la rose est l'objet perçu, et l'acte de mon esprit par lequel je crois à son existence, est ce que nous appelons ici perception. » [Reid (1785-1828) in Nicolas, 2002]

Perception et sensation sont deux notions intimement liées dans la réflexion des philosophes [Russ, 2000 ; Nicolas, 1998]. Pour autant, on note dans cet extrait, et contrairement à la croyance populaire, que ces termes ne sont pas synonymes. La sensation précède la perception en ce sens qu'elle renvoie à la caractéristique de l'objet qui est perçu. La sensation correspond alors à l'excitation olfactive (le parfum) et la perception correspond à l'interprétation de l'objet en tant que réalité (la rose). Cette dichotomie illustre la possibilité de reconnaître la rose à son parfum, sans même la voir. En d'autres termes, la perception est une représentation consciente, élaborée à partir des sensations (fournies par nos cinq sens).

Pour Merleau-Ponty [Merleau-Ponty, 1945], « *Le visible est ce que l'on saisit avec les yeux* (Perception), *le sensible est ce que l'on saisit par les sens* (Sensation) ». La sensation relève alors de la physiologie des organes sensitifs alors que la perception tient de l'interprétation psychologique des

faits. Pour le physiologiste, « *La perception commence par admettre un trajet anatomique qui conduit d'un récepteur déterminé, par un transmetteur défini, à un poste enregistreur spécialisé. Le monde objectif étant donné, on admet qu'il confie aux organes de sens des messages qui doivent être portés, puis déchiffrés, de manière à reproduire en nous le texte original.* » [Merleau-Ponty, 1945]. Pour le psychologue, la conscience est placée dans le corps au milieu du monde et la série stimulus-sensation-perception est une suite d'évènements à l'issue desquels la perception commence. Enfin, comme le souligne Helmholtz [Helmholtz (1866) in Nicolas, 2002] « *Les perceptions d'objets extérieurs sont des représentations, et les représentations sont toujours le résultats de notre activité psychique, les perceptions ne peuvent se produire qu'à l'aide de l'activité psychique ; l'étude des perceptions appartient donc, à proprement parler, à la psychologie, en tant qu'il s'agit de rechercher la nature et les lois de l'intervention de l'âme dans la production des perceptions.* » Perception et sensation se retrouvent donc au centre d'un double paradoxe : à la fois indissociables et distincts, ces concepts se trouvent à la croisée des chemins d'approches elles-mêmes indépendantes et complémentaires : la physiologie et la psychologie.

Pour [Kant, 1790], la sensation permet la représentation d'une chose par les sens. Les sens fournissent le détail des caractéristiques de l'objet, qui a une forme, une couleur, un poids, une texture, cet objet peut être sonore, avoir un goût, une odeur. Les sensations renvoient aux propriétés physiques de l'objet, et en cela, Kant les considère objectives. L'objectivité est définie comme la qualité de ce qui est conforme à la réalité, de ce qui est décrit avec exactitude. Reconnaissant ce caractère exact à la sensation qu'il dit objective, Kant confère aux sensations une appartenance à la faculté de connaître. Or, la connaissance renvoie à la notion de certitude et de vérité acquise. Ce point de vue amène la question suivante : peut-on se fier à nos sens ?

2. Crédibilité de la perception

Pour Descartes [Descartes in Nicolas, 2002], les perceptions et les sensations ne peuvent en aucun cas être considérées comme des sources de nos connaissances. Pour lui, elles sont une source infinie d'erreurs, qui ne réside pas dans nos perceptions mais dans leur interprétation. Nicolas cite Descartes : « *Puisque, à cause de nos erreurs de jugement sur le monde extérieur, l'expérience sensible du réel n'apporte aucune connaissance certaine, nous devons la chercher en nous-même, dans notre âme.* » Il précise que pour Descartes, l'âme n'est constituée que d'idées essentielles, les essences, que Dieu a mises en nous, claires et distinctes : les unes sont les actions de l'âme (volontés), les autres sont ses connaissances (idées innées et passions). Les sensations et les perceptions étant d'origine corporelle, elles n'entrent pas dans l'essence de l'âme. Dès le XVII^e siècle, cette philosophie cartésienne sera réfutée avec le développement d'une théorie de la connaissance plus sensualiste qui mettra en avant la

sensation en tant qu'intermédiaire entre le sujet et le monde comme un chemin vers la connaissance [Russ, 2000]. Merleau-Ponty [Merleau-Ponty, 1945], commente cette théorie sensualiste : « *Je vise et je perçois un monde. Si je disais avec le sensualisme qu'il n'y a là que des "états de conscience" et si je cherchais à distinguer mes perceptions de mes rêves par des "critères", je manquerais le phénomène du monde. Car si je peux parler de "rêves" et de "réalité", m'interroger sur la distinction de l'imaginaire et du réel, mettre en doute le "réel", c'est que cette distinction est déjà faite par moi avant l'analyse, c'est que j'ai une expérience du réel comme de l'imaginaire, et le problème est alors non pas de rechercher comment la pensée critique peut se donner des équivalents secondaires de cette distinction, mais d'expliciter notre savoir primordial du "réel", de décrire la perception du monde comme ce qui fonde pour toujours notre idée de la vérité. Il ne faut donc pas demander si nous percevons vraiment le monde, il faut dire au contraire : le monde est cela que nous percevons.* » Merleau-Ponty considère que si les perceptions nous amènent à la conception du monde, elles permettent également les rêves. Le rêve est une production psychique qui n'est pas reliée à la réalité physique des choses. On peut, par exemple, rêver d'un oiseau, l'entendre, le voir, le toucher en songe sans que cet oiseau existe. Or, nous sommes pourtant capables de distinguer cette perception du réel. En envisageant la perception comme un acte permettant à la fois d'intégrer les données du monde et de l'imaginaire, il fait abstraction de la notion d'illusion : « *le monde est cela que nous percevons.* » Or, dans ce que nous percevons du monde, il existe des phénomènes qui « trompent » la perception. Dans cette idée que la perception constitue la représentation du monde, on peut s'interroger sur la place de l'illusion. L'illusion est une interprétation erronée d'une donnée sensorielle. Tout sujet est sensible à l'illusion, et de ce point de vue, elle est l'illustration directe de la question que nous nous posons : peut-on se fier à nos sens ?

Si l'on est capable de parler d'illusion, c'est que nous sommes conscients du caractère artificiel de notre perception par rapport à l'objet réel. L'exemple le plus concret est l'illusion d'optique. Lorsque l'on regarde le dessin suivant (figure I.1.), nous voyons tous une juxtaposition de carrés noirs espacés de quelques millimètres sur un fond blanc.

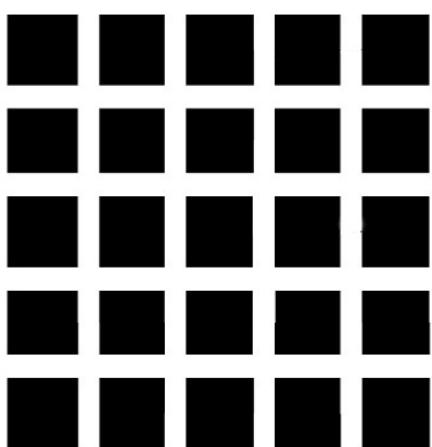


Figure I.1. : Illusion de L'Hermand (1870)

Si l'on fixe l'image quelques secondes, nous voyons apparaître des taches d'un gris sombre à chaque intersection. Ces taches ne correspondent pas à la réalité. Nous comprenons aujourd'hui, avec les progrès de la physiologie, que l'œil, dans sa gestion innée des contrastes, a tendance à remplir l'espace. Ainsi, dans un premier temps, nous percevons l'illusion (nous voyons des taches), dans un second temps nous éprouvons un doute et la

dénonçons grâce à notre connaissance de l'objet (il n'y a que des carrés noirs sur fond blanc), et dans une dernière étape, nos connaissances physiologiques nous permettent d'accéder à la compréhension de l'illusion. Au terme de cette expérience, nous sommes conscients de l'absence de réalité des tâches, témoignant que notre perception, sans l'intervention de nos connaissances, pourrait nous déraciner de la réalité.

L'illusion de Müller-Lyer (figure I.2.) illustre l'effet du contexte et du point de vue. Dans le dessin suivant, la perception nous incite à considérer la droite située en haut plus longue que l'autre. Cependant, la réalité est que ces droites sont toutes deux de longueur identique.

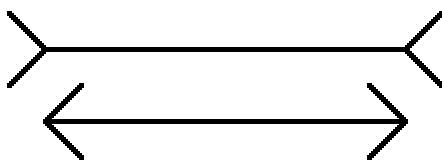


Figure I.2. : Illusion de Müller-Lyer (1889)

Pour Merleau-Ponty [Merleau-Ponty, 1945], « *Les deux segments de droite ne sont ni égaux ni inégaux, c'est dans le monde objectif que cette alternative s'impose.* » Ces deux figures sont géométriquement différentes. Dans un cas, les extrémités de la droite sont des flèches ouvrantes et dans l'autre cas ce sont des flèches fermantes. L'instinct inciterait à comparer des figures géométriques équivalentes, ne différant que par une caractéristique particulière (deux droites avec des flèches fermantes mais de longueur différente). Ce n'est pas le cas ici. Il n'y a donc pas dans l'absolu d'intérêt particulier et instinctif à comparer ces droites. C'est la construction de cette figure par son auteur qui nous incite à la comparaison : elles sont présentées l'une au dessus de l'autre. Pour Merleau-Ponty, l'interprétation de ressemblance ou de la dissemblance de ces objets est rendue délicate par le contexte de la consigne et le point de vue dans lequel la consigne place l'observateur : la tâche de comparaison serait aisée dans un contexte de réalité quotidienne (comparer des pommes rouges et des pommes vertes par exemple), cependant, ici, les objets sont saisis chacun dans leur contexte privé comme s'ils n'appartenaient pas au même univers et la comparaison n'a pas de repère (comparer des pommes et des poires). Merleau-Ponty résume cette explication ainsi : « *Dans l'illusion de Müller-Lyer, l'une des lignes cesse d'être égale à l'autre sans devenir "inégale" : elle devient "autre". C'est à dire qu'une ligne objective isolée et la même ligne prise dans une figure cessent d'être pour la perception "la même".* » La confiance que l'on peut accorder à la perception dépend donc du contexte dans lequel la tâche est effectuée. L'illusion peut être induite par une erreur d'interprétation de la consigne et la question posée doit tenir compte de cette notion. Par extension, « *la perception est le résultat d'une certaine relation entre l'individu et son milieu. Le sujet perçoit "son" monde, prend*

en compte les propriétés structurales de la stimulation, mais aussi de son milieu. » [Ames in Nicolas, 2002]

Pour Nicolas [Nicolas, 2002], « *Les situations ambiguës du point de vue du sujet montrent que la perception d'un individu correspond à ce qui est le plus plausible compte tenu de la situation.* » La figure I.3. montre une photographie (à gauche), réalisée sans trucage. La construction architecturale de la pièce (à droite), le positionnement des fenêtres et le carrelage qui n'est en réalité pas carré (à gauche), induisent une modification des perspectives. Ainsi, on a l'impression que les deux enfants sont à une échelle différente, alors que la pièce dans laquelle ils se trouvent n'est pas cubique et ne comporte aucun angle droit.

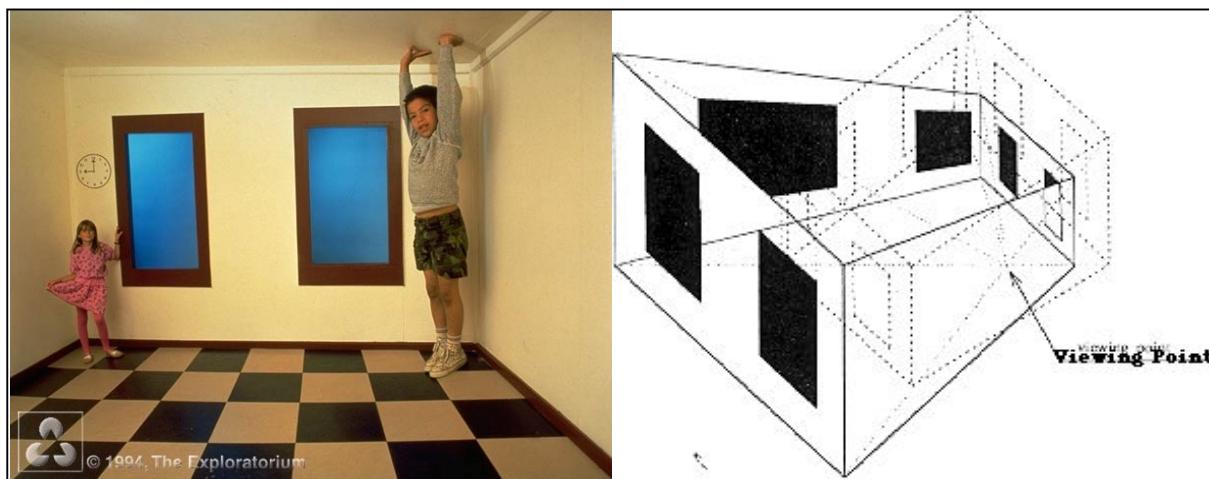


Figure I.3. : La chambre de Ames

Face à une situation énigmatique, le sujet émet des hypothèses (enfant géant / enfant minuscule), puis les réfute en fonction de ses connaissances de la réalité (impossible). Il procède ainsi par essai / erreur jusqu'à obtenir un compromis perception / réalité plausible. Ainsi, pour Nicolas [Nicolas, 2002] « *Notre perception anticipe à chaque instant pour guider notre comportement. Ces hypothèses perceptives se révèlent correctes lorsque nous disposons à la fois d'indices perceptifs et d'une connaissance antérieure des objets perçus.* »

Au total, suite aux différents points de vue que nous venons d'exposer, à la question « Peut-on se fier à nos sens ? », on pourrait répondre ceci : la sensation en tant que telle n'est pas discutable : elle correspond à l'excitation de nos organes sensitifs et ne conduit pas à un jugement de valeur. La sensation existe. Qu'elle soit issue d'un objet réel ou extraite des rêves, la sensation existe. Nous pouvons réunir plusieurs sujets et leur montrer une couleur. Tous (si l'on exclut la pathologie) seront capables de la nommer de manière identique. Ils pourront même la définir, affiner ses caractéristiques propres et conserver le consensus. En cela, oui, nous pouvons nous fier à nos sens. A l'inverse, la

perception peut être abusée. Le contexte de présentation de l'objet peut conduire à une interprétation inadéquate de la réalité de cet objet. Sans l'intervention des connaissances, l'interprétation des données fournies par les sens peuvent donc conduire à une conclusion erronée, qui ne correspond pas à la réalité. En cela, nous ne pouvons pas nous fier à nos sens.

Enfin, comme le souligne Merleau-Ponty [Merleau-Ponty, 1945] « *Chercher l'essence de la perception c'est déclarer que la perception est non pas présumée vraie, mais définie pour nous comme un accès à la vérité.* » Notre accès au monde dépendant du lien unique et privilégié qui unit objet et sensation, nous ne pouvons réfuter l'indubitable nécessité de nous fier à nos sens. Ce n'est pas un choix. En revanche, dans le cas de « la chambre de Ames », nos connaissances semblent prendre le pas sur nos sensations. On peut donc en conclure que dans ce cas précis nous rejetons les sensations en tant qu'outil fiable. Nous émettons un jugement. Là où la légitimité de la perception peut être remise en cause, c'est lorsqu'elle constitue la base d'un jugement. Peut-on, par l'intermédiaire de la perception émettre un jugement fiable ? Les points de vue développés dans les deux précédents chapitres amènent la question suivante : la perception est-elle compétente en tant qu'outil d'analyse ?

3. Fiabilité de la perception

Nous arrivons ici à une question cruciale pour le thème de nos travaux. Dans nos travaux, la perception auditive permet l'analyse de la qualité de la voix. Un jury d'écoute constitué de plusieurs auditeurs a pour tâche d'évaluer, à l'oreille, la quantité de dysphonie et de la caractériser. Les jugements perceptifs recueillis constituent une expertise des stimuli sonores. Le principal reproche opposé à ce type de procédure est son manque de fiabilité, lié à la subjectivité des jugements perceptifs. Dans ce chapitre, et avant de nous intéresser au problème du jugement perceptif, nous nous interrogerons sur les sources d'influence de la perception, et nous essaierons d'en tirer des conséquences pour nos travaux.

La perception est subjective. La subjectivité de la perception est évoquée dans la pensée kantienne comme un attribut propre des sensations. Kant [Kant, 1790] distingue deux types de sensations : la sensation qui permet la représentation d'une chose par les sens et qui est rapportée à l'objet, donc objective ; et la sensation qui permet la détermination du sentiment de plaisir ou de rejet, qui est rapportée au sujet, et donc subjective. La sensation subjective ne peut d'aucune manière constituer une représentation d'un objet. Kant fournit l'exemple suivant : « *La couleur verte des prés est une sensation objective, en tant que perception d'un objet des sens ; son caractère agréable est une sensation subjective par laquelle aucun objet n'est représenté.* » Or, est-il possible de dissocier les sensations objectives des sensations subjectives ?

Dans l'exemple de la rose de Reid (section I, chapitre 1, § 2), la sensation olfactive est décrite en tant que sensation objective. Cependant, on peut se demander dans quelle mesure la sensation subjective du parfum de la rose (agréable) intervient dans la perception finale de l'objet. La sensation subjective pourrait être partie intégrante de la perception au même titre que la sensation objective en tant que caractéristique propre de l'objet. Si l'on se place dans le cadre du jugement, en tant qu'estimation de la qualité de l'objet ou du sentiment qu'il procure, la sensation subjective intervient nécessairement de façon couplée avec la sensation objective. Dans nos travaux basés sur l'analyse perceptive de la voix, le jugement objectif du degré de dysphonie subit l'influence du jugement subjectif propre à chaque auditeur : le caractère esthétique ou dysharmonieux de la voix pathologique dépend aussi du goût de chacun. On comprend donc que la perception en général, et dans l'évaluation de la voix en particulier, soit soumis à la variabilité.

La perception est variable. La variabilité caractérise le changement, la fluctuation des phénomènes. La variabilité de la perception indique qu'il est possible de ressentir et d'interpréter les objets différemment, d'un individu à l'autre ou d'un moment à l'autre. Pour le psychologue, la perception peut varier en fonction de l'état du sujet. L'humeur, par exemple, va conduire à un changement d'interprétation des sensations. Et le linguiste de rebondir sur cette notion : « voir tout en noir », « voir la vie en rose », etc. La perception, résultant d'une ou d'un ensemble de sensations, peut être changeante en fonction du contexte : un même stimulus, qui provoquera en tant que tel les mêmes sensations, sera interprété - donc perçu - différemment en fonction de l'état psychologique, variable dans le temps, de l'individu. Par exemple, montrez la photo d'un chaton à un sujet qui vient d'en adopter un, cela le fera sourire. Montrez la même photo au même sujet qui vient de perdre le sien, cela le rendra triste. L'état d'esprit, du sujet peut expliquer que le jugement d'un individu puisse changer dans le temps, en fonction de son contexte psychologique : c'est la variabilité intra-individuelle. Cette variabilité se retrouve dans les travaux d'analyse perceptive de la dysphonie lorsqu'un auditeur n'attribue pas la même note à un même stimulus d'une semaine sur l'autre.

D'autre part, l'interprétation de la sensation est influencée par notre connaissance du monde et des propriétés physiques des objets (section I, chapitre 2, § 5). L'être humain n'aime pas les mystères. Pour Merleau-Ponty [Merleau-Ponty, 1945], « *La perception est une interprétation des signes que fournit la sensibilité, une hypothèse que l'esprit fait pour s'expliquer ses impressions.* » Nos connaissances associées à notre perception recherchent une conclusion plausible. Par exemple : lorsque nous fixons la mer, au loin, notre perception nous inciterait à voir une ligne, correspondant à une limite au delà de laquelle il n'y a plus rien. Avant Copernic, la seule conclusion possible était que la terre était plate. Depuis le XVI^e siècle, nous savons qu'il s'agit de la ligne d'horizon qui est due à la courbure de la terre qui est sphérique. Aujourd'hui, cette connaissance est acquise de tous et la ligne d'horizon ne nous trompe plus. Ainsi, le développement des connaissances influence notre perception.

Cela étant, l'étendue et le domaine de connaissance peut être différent d'un sujet à l'autre, et la même sensation peut donc aboutir à une interprétation perceptive différente en fonction de l'individu. Dans ce cas, on peut penser que la connaissance est un des facteurs d'influence de la perception subjective décrite par Kant. Cela pourrait expliquer la variabilité inter-individuelle de la perception, c'est à dire une interprétation différente d'un individu à l'autre. Dans la situation d'une tâche de jugement, c'est à dire en ce qui nous concerne l'évaluation de la qualité de la voix, la connaissance du phénomène dysphonique peut influencer le jugement de l'auditeur. Cela confirme la nécessité de formation du jury d'écoute. En effet, l'expérience des auditeurs est considérée comme un facteur permettant d'augmenter la fiabilité des résultats.

Perception – Expérience – Connaissances. L'expérience perceptive éduque la perception. L'expérience se définit comme tout ce qui est appréhendé par les sens et constitue la matière de la connaissance humaine. Elle se situe donc à la croisée exacte des chemins de la perception et de la connaissance. L'expérience est une transition entre perception et connaissances, qui permet un va et vient permanent et systématique entre l'un et l'autre (figure I.4.).

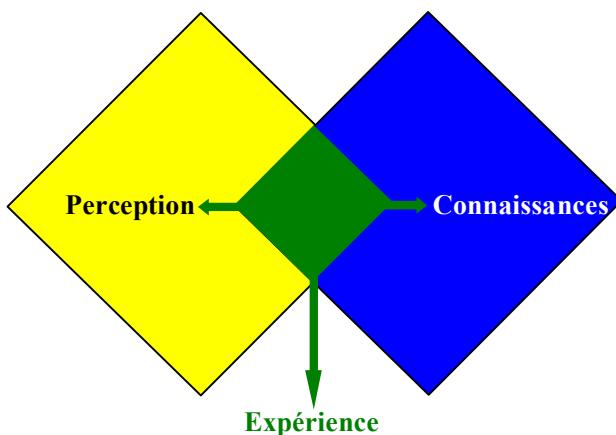


Figure I.4. : L'expérience constitue la transition qui relie perception et connaissances

La perception permet, au travers de l'expérience, d'acquérir des connaissances. Pour Merleau-Ponty [Merleau-Ponty, 1945], « *La fonction essentielle de la perception est de fonder ou d'inaugurer la connaissance.* » C'est un point que nous avons largement développé dans le chapitre précédent. Cependant, il faut considérer également que les connaissances apportent une expérience permettant d'améliorer les perceptions. Dans ce sens, l'expérience façonne notre perception. L'expérience est donc une zone transitoire permettant un échange permanent entre les perceptions et les connaissances (figure I.4.).

Le mécanisme de perception évolue grâce l'expérience : nous savons par exemple que le bébé est quasiment aveugle à la naissance et que sa vue se développe au cours des premières semaines. Le

contexte, c'est à dire le monde qui nous entoure et notre connaissance de ce monde, joue un rôle prépondérant sur le développement de la perception : les pygmées vivant dans des zones forestières très denses possèdent une mauvaise vue de loin. Les esquimaux possèdent dans leur langage une dizaine de mots pour qualifier les nuances de blanc, ce qui signifie qu'ils sont eux-mêmes capables de distinguer ces nuances. Enfin, des animaux dont les yeux sont exposés continuellement à une seule couleur dès la naissance deviennent incapables de discerner les couleurs : ils sont aveugles. Dans le même ordre d'idées, si on interdit la possibilité d'apprentissage de la pince fine à un enfant pendant trop longtemps, ce geste deviendra impossible à effectuer par l'adulte. Pour terminer cette série d'exemple, citons Lorena A Hickok¹ qui écrit en 1958 l'histoire d'Helen Keller, une enfant née aveugle et sourde : « *La petite fille, privée du sens de l'ouïe et de la vue, avait développé d'une façon extraordinaire son sens du toucher, ainsi que ceux de l'odorat et de goût.* » Par conséquent, il semble que la diversité des expériences sensorielles permet de modeler les capacités liées à ce sens.

On peut conclure de cela que perception, expérience et connaissance sont fortement liées. Dans notre exploitation de la perception en tant qu'outil d'analyse, nous retiendrons ceci : pour que le jugement perceptif soit le plus juste possible, il semble nécessaire que le jury recruté possède non seulement un niveau de connaissances le plus poussé possible dans le domaine de la pathologie vocale, et une expérience de l'écoute de la dysphonie et de la tâche d'évaluation afin de réduire la variabilité intra-individuelle, mais aussi que ces connaissances et expériences soient communes et de niveau équivalent entre tous les auditeurs afin de réduire la variabilité inter-individuelle. Cependant, la subjectivité et la variabilité de la perception posent le problème de la validité de la perception lorsqu'elle doit servir de base à un jugement. En d'autres termes, est-il légitime d'utiliser la perception pour émettre un jugement ? Cette interrogation fera l'objet du chapitre suivant.

4. Perception et jugement

Pour Nicolas [Nicolas, 1998], « *La perception du monde extérieur est le résultat d'expériences et de raisonnements ou de jugements faits sur les sensations.* » L'accès au monde se fait par la perception. Dans notre interprétation consciente du monde, nous procédons à des raisonnements et à des jugements qui nous permettent d'analyser et de comprendre le monde, pas seulement de le recevoir. Le jugement est une faculté qui permet de prendre une décision, d'estimer une valeur, de porter une appréciation. Le jugement fait donc partie intégrante de l'individu et possède une fonction essentielle dans l'interprétation du monde.

¹ Hickok LA. L'histoire d'Helen Keller. Paris : Robert Laffont, 1958

Pour Merleau-Ponty [Merleau-Ponty, 1945], « *Le jugement est souvent introduit comme ce qui manque à la sensation pour rendre possible la perception.* » Ainsi, la faculté jugement naît de l'ambiguïté des perceptions : c'est parce que la perception peut être abusée que l'esprit compense son incertitude par une prise de décision. Le jugement permet de reconstruire ce qui nous manque pour représenter l'objet. Descartes donne une illustration de cette idée : « *Je ne manque pas de dire que je vois des hommes (...) ; et cependant, que vois-je de cette fenêtre sinon des chapeaux et des manteaux qui peuvent couvrir des spectres ou des hommes feints qui ne se remuent que par ressort ? Mais je juge que ce sont de vrais hommes.* » [Descartes in Merleau-Ponty, 1945]. Il n'est donc pas nécessaire d'avoir une sensation détaillée de l'objet pour l'identifier. En cela, pour Merleau-Ponty, « *La perception est un jugement.* » Le jugement est donc une faculté interne de l'esprit qui « juge comme il respire » sans même s'en apercevoir afin de compléter ses sensations et d'affiner sa perception du monde. De ce point de vue, la légitimité du jugement perceptif ne ce discute pas ; le jugement perceptif existe et fonctionne.

De cette faculté de juger, l'individu tire la capacité d'évaluation qualitative ou quantitative d'un objet ou d'un phénomène. Le jugement permet de trancher entre deux couleurs la plus foncée (jugement qualitatif), ou entre deux arbres le plus haut (jugement quantitatif). Ce jugement est une tâche quotidienne si familière qu'elle est inconsciente. Il permet de s'adapter au monde et à ses objets et de prendre des décisions : le jugement qualitatif de l'aspect du ciel permet de déterminer si l'on a besoin d'un parapluie, le jugement quantitatif du volume des courses permet de décider si l'on prend un chariot ou un panier, etc. Il s'agit là de jugements pratiques, qui déterminent notre comportement dans la résolution d'une tâche ou d'un problème. Ils influencent donc le sujet dans sa gestion du quotidien, mais ne l'affectent pas. Les jugements qualitatifs et quantitatifs seraient donc du domaine de l'objectif, à la différence des jugements de valeur, pour lesquels interviennent nécessairement les notions de goût et d'esthétique, d'agréable.

Le goût, c'est la faculté de juger le beau ou le bon. C'est le sens intuitif des valeurs esthétiques. L'analyse des jugements de goût doit dégager ce qui est nécessaire pour dire qu'un objet est beau. Kant recherche, en suivant les fonctions logiques du jugement, les critères généraux qui permettent le jugement de goût. Il examine ainsi le critère de la qualité parce que , selon lui, « *(...) c'est celui que le jugement esthétique sur le beau considère en premier lieu.* » Il considère donc pour définir le jugement esthétique, la qualité en premier lieu, ainsi que la quantité, la relation et la modalité. Malgré les tentatives de description et de généralisation de la faculté de jugement, il demeure que « *Le jugement du goût n'est pas (...) logique, mais esthétique ; esthétique signifie : ce dont le principe déterminant ne peut être que subjectif.* » [Kant, 1790]. Le goût est donc individuel. C'est l'individu qui seul énonce son jugement de goût. Pour Kant, le lien qui unit l'objet au sentiment de plaisir ou de peine ne représente pas l'objet en tant que tel mais la manière dont le sujet en est affecté, le ressentit.

Le jugement esthétique est un jugement affectif, qui fait intervenir les sentiments. En cela, il est subjectif. La sensation de satisfaction que l'on peut éprouver en présence d'un objet est donc liée à nous-même davantage qu'à l'objet. Pour reprendre l'exemple des couleurs, il ne s'agirait plus de déterminer laquelle est la plus foncée mais laquelle est la plus belle, ce qui renvoie à l'appréciation du sujet lui-même et pas à la couleur. Toutefois, Kant souligne que le jugement porté par le moyen de représentations esthétiques est logique si celles-ci ne sont rapportée qu'à l'objet. Par exemple, le sentiment de plaisir éprouvé lors de la dégustation d'une fraise renvoie aux qualités intrinsèques de la fraise, plus qu'au sentiment éprouvé lorsqu'on la mange. En cela, le jugement esthétique est « *logique* » car il est rattaché à l'objet. Inversement, « (...) même si les représentations données étaient rationnelles, le jugement serait toujours esthétique dans la mesure où les représentations ne seraient en celui-ci rapportées qu'au sujet (à son sentiment). » Cela signifie que même si les indices sensitifs revoient à la qualité physique propre de l'objet, il suffit de s'attacher en particulier à la façon dont nous en sommes affecté pour que le jugement soit esthétique. Kant tendrait ainsi à dire que le jugement de goût est un jugement esthétique qu'il soit rattaché à l'individu ou à l'objet.

Les jugements perceptifs de la qualité de la voix sont des jugements qualitatifs et quantitatifs, dans lesquels intervient nécessairement la notion d'esthétique. Notre tâche d'évaluation de la dysphonie fait intervenir la caractérisation qualitative et la quantification de celle-ci. Le jugement qualitatif qui s'attache à déterminer la présence ou l'absence d'un caractère pathologique dans la voix serait un jugement rattaché à l'objet voix et donc en ce sens, objectif. Le critère de quantification de la présence de cette qualité pathologique serait un jugement rattaché au critère plus qu'à la voix, et serait donc soumis davantage à la subjectivité. Le jugement de goût permettant l'évaluation globale interviendrait en amont et serait rattaché au sentiment esthétique général procuré par l'écoute de la voix. Le jugement de goût serait donc subjectif. Notons que dans la tâche qui est la notre, la notion d'esthétique n'est pas présente en tant que telle ; on ne juge pas ce qui est « le plus beau » mais « le moins beau ». Il s'agit donc d'un jugement de « l'inesthétique » qui fonctionne de la même façon.

La voix porte à ressentir et à s'émouvoir. Pour Parret [Parret, 2002], « *La saisie de la qualité de la voix signifie toujours un certain abandon à la séduction. Les sirènes séduisent et les muses inspirent par le timbre de leur voix. (...) Le timbre de la voix, c'est le territoire des poètes et des amoureux plutôt que des phonéticiens.* » Les qualificatifs décrivant la voix sont nombreux et divers, introduisant des notions acoustiques aussi bien que sentimentales ou esthétiques. La voix peut être douce, mélodieuse, agréable, chantante, fluette, moelleuse, rassurante, caressante, timbrée, enjôleuse, coquine, enfantine, etc. Mais elle peut aussi être grinçante, rauque, nasillarde, dure, entêtante, assourdisante, criarde, aigre, d'outre-tombe, effrayante, cassée, etc. La voix est par essence un phénomène perceptif. La voix est émise pour être entendue. En tant que support de la communication ou en tant que support de l'art chanté, elle est soumise au jugement esthétique. Le jugement esthétique

de la voix ne peut être que perceptif, auditif. Il ne faut donc pas se demander si l'on peut utiliser la perception pour évaluer la voix, mais si l'on peut émettre un jugement sans faire intervenir la perception. Car la voix n'a de raison d'être que par la perception, et les approches scientifiques (mesures instrumentales) de quantification de la voix n'ont elles même d'intérêt que si elles reflètent le phénomène perçu.

Cette première partie nous a permis d'éclairer les interrogations basiques sur la perception grâce à l'introduction de la pensée philosophique. Au terme de cette réflexion, nous comprenons que la perception est l'incontournable lien entre nous et le monde. En dédicace du livre qui lui est consacré, Helen Keller, enfant née aveugle et sourde, évoque ce lien unique : « *A la mémoire de Maîtresse qui entraîna une petite fille hors des ténèbres et lui donna le monde...* » En l'absence des sens de la vue et de l'audition, la petite fille n'était pas dans le monde, mais enfermée dans le noir et le silence. L'accès au monde lui était – avant une éducation particulière et spécialisée – refusé. Merleau-Ponty disait : « *Il ne faut donc pas demander si nous percevons vraiment le monde, il faut dire au contraire : le monde est cela que nous percevons.* » En d'autres termes, il est vain de discuter la fiabilité, la crédibilité ou même la légitimité de la perception, puisqu'elle existe, fonctionne et que rien ne peut la suppléer.

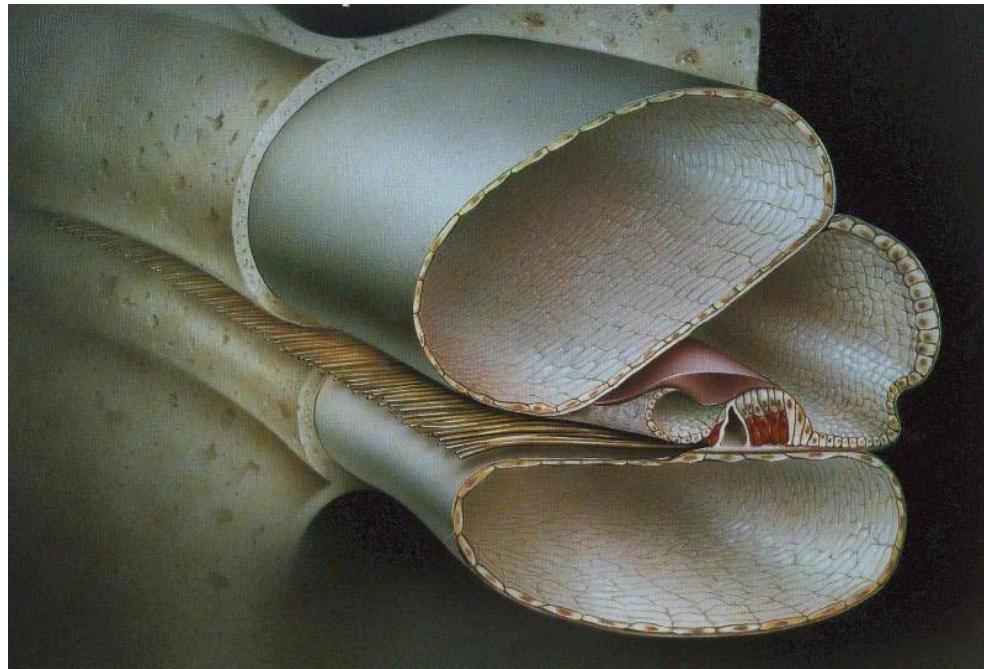
Notre besoin d'éclaircissement sur la légitimité de la perception était venu du fait que nos travaux, basés sur le jugement perceptifs, avaient un objectif de description scientifique. Il nous semblait alors douteux de relier la science qui est objective à la perception qui est subjective. La science a pour mission de décrire le monde, et en aucun cas de s'y substituer. Mais la science déjà, basée sur l'observation du monde dépend de la perception des évènements qui nous entourent. « *Tout l'univers de la science est construit sur le monde vécu et si nous voulons penser la science elle-même avec rigueur, en apprécier exactement le sens et la portée, il nous faut réveiller d'abord cette expérience du monde dont elle est l'expression seconde. La science n'a pas et n'aura jamais le sens d'être le monde perçu, pour la simple raison qu'elle en est une détermination ou une explication.* » [Merleau-Ponty, 1945]. Tout ce que nous savons du monde, même par science, nous le savons à partir d'une vue qui nous est propre ou d'une expérience du monde sans laquelle les symboles de la science ne voudraient rien dire. Toute notre connaissance est donc à la base la résultante d'une perception. Il n'existe donc pas de litige entre science et perception puisqu'au fond la science n'est que l'outil qui explique ce que nous percevons.

L'analyse perceptive est donc légitime dans le cadre des travaux scientifiques qui cherchent à décrire la voix et sa pathologie, même si le besoin de description qui est le propre de la science conduit à

concevoir des appareillages de mesure qui pourraient permettre de se substituer à la perception. « *Il est inévitable que dans son effort général d'objectivation la science en vienne à se représenter l'organisme humain comme un système physique en présence de stimuli définis eux-mêmes par leurs propriétés physico-chimiques, cherche à construire sur cette base la perception effective et à fermer le cycle de la connaissance scientifique en découvrant les lois selon lesquelles se produit la connaissance elle-même, en fondant une science objective de la subjectivité.* » [Merleau-Ponty, 1945].

Notre démarche scientifique est un effort pour rallier les mesures objectives au jugement perceptif. Partant du principe que la perception auditive est l'accès primitif à la production vocale, elle est l'outil d'analyse originel et incontestable de la qualité de la voix. La recherche et l'élaboration d'appareillages de mesure a donc pour but de décrire au mieux ce que nous percevons de ce phénomène, et il est donc inévitable de considérer le jugement perceptif, non seulement comme un outil d'analyse mais aussi comme la référence vers laquelle les mesures doivent tendre.

Section d'un tour de la Cochlée



La physiologie de la perception

La physiologie de la perception

La perception auditive, qui est une activité d'interprétation des données sensorielles fournies par l'ouïe, répond avant tout aux lois physiologiques et acoustiques de l'oreille. Afin de comprendre ce qui se passe pendant l'audition, nous avons choisi de reprendre brièvement l'anatomie et la physiologie de l'oreille.

L'oreille est un organe particulièrement ingénieux et complexe. La fascination et la perplexité que l'oreille exerce sur les développementistes, suggère que la conception de cet organe ne peut-être que le fruit d'une création divine, comme le souligne René Barjavel dans ce paragraphe aussi dubitatif que passionné.

« *Qui a conçu l'oreille ?*

Il faut être singulièrement facile à contenter pour accepter de voir dans la simplicité harmonieuse de son aménagement général, le raffinement de ses détails, la diversité de son fonctionnement mécanique, acoustique, électrique, chimique, sérieux, sanguin, conjonctif, osseux, musculaire, nerveux, liquide, solide, gazeux, et nous en oublions, et nous en ignorons, et dans la coordination immédiate et parfaite de cette multiple subtilité, le résultat chanceux de mutations hasardeuses.

Nous admettons volontiers le système de la sélection du mieux armé et du mieux adapté. L'animé qui avait une oreille a survécu à celui qui n'en avait pas. D'accord. Mais qui a donné son oreille à celui qui l'avait ?

Ce n'est pas si simple, dit-on. Il y a eu d'abord une cellule qui était vaguement sensible aux vibrations, puis...

D'accord.

Mais comment cette cellule vaguement sensible a-t-elle transformé cette vibration en une sensation auditive ? Comment s'est-elle adjoint d'autres cellules ? Comment se sont-elles fait pousser des cils sélectifs, se sont-elles enfermées dans le limaçon, le limaçon dans le labyrinthe ? Comment se sont-elles fait précéder d'un système amplificateur ? Comment ont-elles fait émerger et fleurir le pavillon ? Comment ? Comment ? Qui a voulu ces perfectionnements successifs ?

Est-ce l'individu ?

Si c'était possible, tous les hommes se seraient depuis longtemps fait pousser des ailes et des yeux derrière la tête.

Est-ce l'espèce ? La matière vivante elle-même ?

Qui ?

L'oreille ne s'est pas faite par l'inavraisemblable hasard de millions de mutations favorables.

L'oreille est un ensemble conçu, architecturé, organisé. Le hasard ne conçoit pas, n'ajuste pas, n'organise pas. Le hasard ne fait que de la bouillie.

Même si on tient compte du facteur temps, on ne peut pas accepter l'explication du hasard. Je connais l'argument du singe et de la machine à écrire : si on place un singe devant une machine à écrire et qu'il tape au hasard sur le clavier pendant l'éternité, comme il tapera une infinité de combinaisons de lettres, il finira par taper le texte de la Bible.

*Je n'accepte pas cet argument. Il est faux. Il confond la quantité et la qualité. Le singe ne tapera pas la Bible, pas même *La Cigale et la Fourmi*. Il tapera pendant l'éternité un cafouillis lettriste, jusqu'à la fin des temps.*

Vous pouvez lancer un dé pendant l'éternité, vous n'obtiendrez jamais une série de 1000 six. Or il faudrait une accumulation de mutations favorables autrement extraordinaire qu'une série de 1000 six pour fabriquer une oreille, ou une marguerite ou un petit chat.

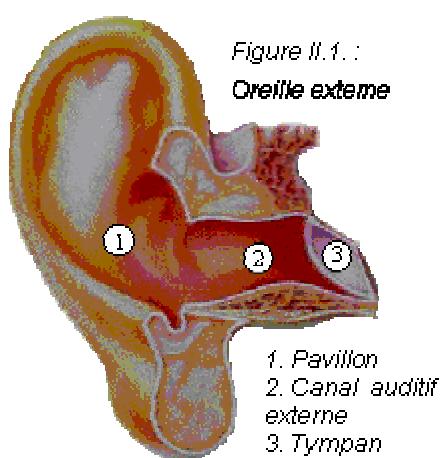
Mais d'où viennent l'oreille et la marguerite ?

IL Y A QUELQU'UN !...

Il y a quelqu'un sous le lit, dans l'armoire ! Il y a quelqu'un dans notre vie, dans notre chair. Quelqu'un qui nous a faits et qui fait de nous ce qu'il veut. »

Barjavel, 1966¹

1. Anatomie de l'oreille



L'oreille externe. Le rôle de l'oreille externe est de recueillir les ondes sonores. Le pavillon de l'oreille fonctionne comme un cornet acoustique. Il est prolongé du conduit auditif externe, qui correspond à une cavité de résonance. Ce conduit quasi cylindrique est long d'environ 25 mm et son diamètre est d'environ 8 mm. Il prolonge le pavillon de l'oreille en s'enfonçant obliquement dans l'os temporal, et se termine en dedans par la membrane du tympan qui constitue la frontière avec l'oreille moyenne. L'oreille externe est un milieu aérien dont la pression est équilibrée avec la pression atmosphérique, puisqu'elle est ouverte sur l'extérieur.

¹ Barjavel R. La faim du tigre. Paris : Presses Pocket, 1984

L'oreille moyenne. La limite externe de l'oreille moyenne est constituée par le tympan. Le tympan est une membrane élastique très résistante concave en dehors. Son diamètre est d'environ 10 mm, et son épaisseur de 1/10^{ème} de millimètre. L'oreille moyenne est de forme cubique et correspond à la caisse du tympan. Comme l'oreille externe, l'oreille moyenne est un milieu aérien. La pression à l'intérieur de cette cavité est gérée par la trompe d'Eustache qui relie l'oreille moyenne au pharynx et permet la circulation de l'air entre les deux, donc l'équilibration des pressions.

L'oreille moyenne, contenue dans la caisse tympanique, est constituée des osselets : le marteau, l'enclume et l'étrier. Les noms de ces petits os, issus du langage courant, proviennent de leur analogie morphologique avec ces objets. Les osselets sont articulés entre eux à partir du tympan jusqu'à la frontière de l'oreille interne. La grande apophyse du marteau (ou manche) est insérée dans la membrane tympanique. La tête du marteau repose sur l'enclume, laquelle se termine par un renflement, l'apophyse lenticulaire, qui s'articule avec l'étrier. La chaîne des osselets constitue un système oscillant qui permet de transmettre la vibration sonore depuis le tympan jusqu'à l'oreille interne. La variation de pression exercée sur le tympan entraîne un mouvement synchrone du marteau, répercute sur l'enclume, puis sur l'étrier. La platine de l'étrier vient s'enfoncer dans la fenêtre ovale qui est située sur la paroi interne de la caisse tympanique, et transmet alors la vibration de l'oreille moyenne au vestibule de l'oreille interne.

L'oreille interne. L'oreille interne est constituée de la cochlée en avant, et du système vestibulo-semi-circulaire en arrière. L'ensemble de ces cavités est rempli de liquide : le liquide périlabyrinthique. Contrairement à l'oreille externe et à l'oreille moyenne, l'oreille interne n'est donc pas un milieu aérien mais liquidiens.

La cochlée est formée d'un tube osseux enroulé autour du nerf auditif en une spirale de 2 tours ½. Ce tube est divisé en deux rampes par une cloison osseuse dans sa partie axiale (lame spirale), et membraneuse dans sa partie distale (membrane basilaire). Ces rampes sont en contact avec l'oreille moyenne par la présence de deux fenêtres : la fenêtre ovale qui reçoit l'impulsion vibratoire par l'intermédiaire de l'étrier, et la fenêtre ronde qui, fermée par une membrane élastique appelée tympan secondaire, sert de soupape de décompression. La membrane basilaire a une longueur de 32 mm. Sa

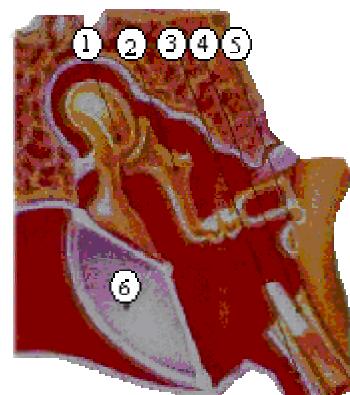


Figure II.2. : Oreille moyenne

1. Marteau
2. Enclume
3. Trompe d'Eustache
4. Tendon du tympan
5. Etrier
6. Tympan

largeur est de 0,04 mm à la base et croît progressivement jusqu'à 0,36 mm au sommet alors que la lame spirale osseuse diminue.

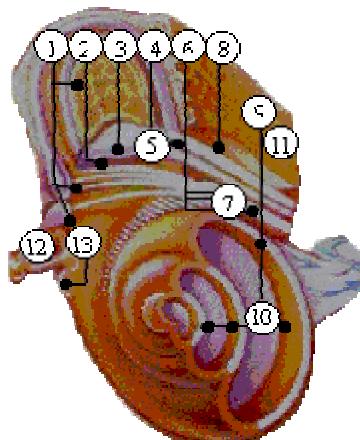


Figure II.3. : Oreille interne

1. Canal semi-circulaire
2. Nerf facial
3. Ganglion géniculaire
4. Nerf facial
5. Branche vestibulaire du nerf de la cochlée
6. Ganglion vestibulaire
7. Nerf vestibulaire
8. Marteau interne
9. Cochlée
10. Organe de Corti
11. Os temporal
12. Etrier
13. Fenêtre ronde

Le limacon membraneux est séparé de la rampe vestibulaire par la membrane de Reissner et repose sur la membrane basilaire. Il s'agit d'un fin canal de section triangulaire de structure très complexe. A la partie centrale de sa base d'implantation sur la membrane basilaire, son épithélium subit une haute différenciation cellulaire qui constitue un renflement : l'organe de Corti. L'organe de Corti présente au centre les piliers formant 5 000 arcades qui forment le tunnel de Corti. De part et d'autre de celui-ci sont implantées les cellules sensorielles, les cellules ciliées. Il existe une seule rangée en dedans du tunnel de Corti, les cellules ciliées internes qui sont au nombre de 3 500 et 3 à 4 rangées en dehors du tunnel, les



Figure II.4. : La cochlée

cellules ciliées externes, au nombre de 20 000 environ. Au contact de la base des cellules ciliées se terminent les ramifications du nerf cochléaire, 30 000 fibres environ. C'est donc à ce niveau que la vibration jusqu'alors mécanique se transforme en influx neuro-sensoriel. L'organe de Corti est recouvert de la membrane tectoriale qui s'insère à la face supérieure de la lame spirale en dedans et en contact très étroit avec les cellules ciliées internes et externes qui la pénètrent par leurs cils.

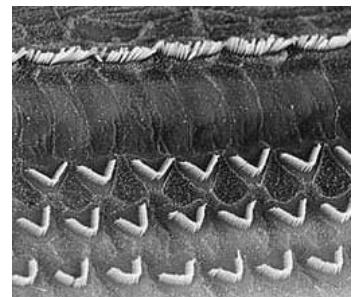


Figure II.5. : cils vibratiles

2. Physiologie de l'oreille

Transmission des mouvements vibratoires. Les ondes sonores sont collectées par le pavillon de l'oreille, amplifiées par le conduit auditif externe et atteignent le tympan qui se met à vibrer et

transmet le mouvement à la chaîne ossiculaire située dans l'oreille moyenne. L'apophyse lenticulaire de l'enclume entraîne l'étrier qui vient s'enfoncer dans la fenêtre ovale, qui constitue la frontière avec l'oreille moyenne. Les mouvements de la platine de l'étrier engendrent un déplacement de liquide périlymphatique dans la rampe vestibulaire. Comme tout liquide, le liquide périlymphatique est incompressible, la décompression du système est donc assurée par la fenêtre ronde qui sert de soupape. Les mouvements de la fenêtre ovale et la fenêtre ronde sont en opposition de phase : lorsque la fenêtre ovale est enfoncée, la fenêtre ronde se gonfle, et inversement. Ce bloc oscillant fonctionne librement lorsque les sons ont une faible intensité, mais dans le cas de fortes intensités, les articulations intermédiaires absorbent une partie de l'énergie afin de protéger l'oreille interne : c'est le réflexe stapédiens. Le bloc tympano-ossiculaire possède donc un rôle multiple : transmettre et amplifier les ondes sonores, et protéger l'oreille interne.

Au niveau de l'oreille interne, les déplacements du liquide périlymphatique dus aux mouvements des fenêtres, entraînent une activation de la membrane basilaire. Celle-ci est parcourue d'une « onde propagée » qui naît à la base de la cochlée et se déplace de la base vers l'apex. La position maxima de déplacement dépend de la fréquence du son : près de la base pour les fréquences graves, près de l'apex pour les fréquences aiguës. Puis l'onde s'amortit rapidement après ce déplacement maximum. Les mouvements des cellules ciliées sont provoqués par les déplacements de la membrane basilaire. Les cils sont maintenus rigides à leur base par une lame réticulaire et sont solidaires de la membrane tectoriale dans la face inférieure de laquelle s'incruste leur extrémité supérieure. Quand la membrane basilaire oscille, l'ensemble formé par le tunnel de Corti et la lame réticulaire glissent l'un par rapport à l'autre, et il en résulte une inclinaison des cils qui sont soumis à un effet de cisaillement. Les oscillations du canal cochléaire aboutissent ainsi aux vibrations des cils des cellules sensorielles. C'est là le phénomène mécanique fondamental de l'audition : les vibrations mettent en mouvement la membrane basilaire et entraînent une vibration alternative des cellules sensorielles.

Le nerf auditif est constitué de 40 000 fibres nerveuses qui coiffent la base des cellules ciliées internes et externes. Le nerf est déroulé comme un éventail en spirale. Pour un son pur, schématiquement, une seule fibre va être mise en action. Pour un son complexe, plusieurs fibres. Pour une intensité faible, les influx sont peu nombreux. Ils augmentent jusqu'à 1 000 par seconde quand l'intensité augmente, mais les fibres nerveuses ne peuvent émettre plus d'influx par seconde. Quand l'intensité augmente encore, les fibres voisines entrent en jeu même si elles ne correspondent pas exactement à la fréquence incitative. La spécificité de la perception diminue alors.

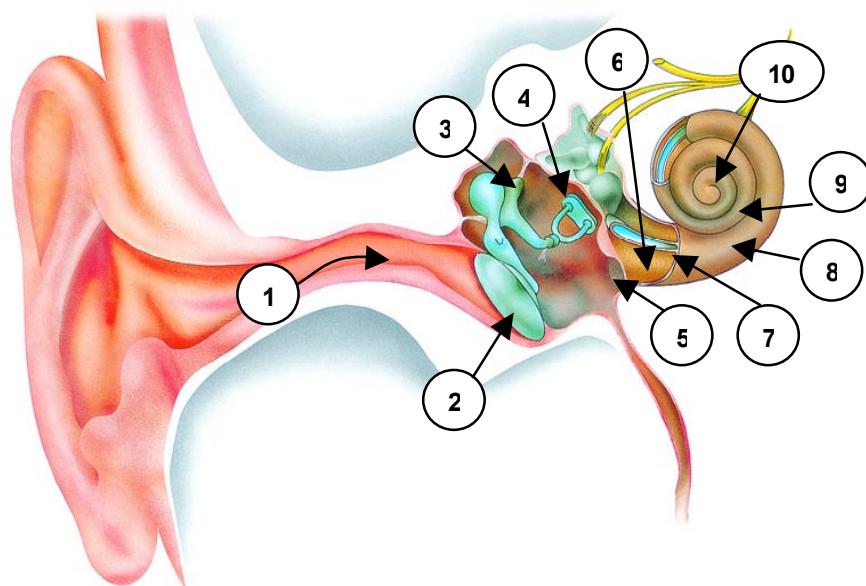


Figure II.6. : Physiologie de l'Audition

(1), les vibrations sonores sont renforcées par le CAE. (2), le tympan vibre, entraînant les (3) osselets dans la caisse du tympan. (4), la platine de l'étrier s'enfonce dans la fenêtre ovale et transmet les vibrations au liquide périlymphatique de la rampe vestibulaire. (5), le tympan secondaire de la fenêtre ronde vibre en opposition de phase avec la platine. (6), les fréquences hautes font vibrer la membrane basilaire au niveau de la base de la cochlée. (7), les cellules de Corti (sensorielles) transforment les vibrations mécaniques en potentiels électriques (8), qui sont transmis aux centres par le nerf auditif. (9), les fréquences basses agissent sur la partie haute de la cochlée. (10), les ondes cheminent ensuite dans la rampe tympanique en passant pas l'hélicotréma au sommet de la cochlée.

3. Les voies nerveuses de la perception auditive

Nerf auditif. Les influx nerveux qui prennent naissance au contact des cellules auditives sont conduits au cerveau par le nerf cochléaire. Le nerf cochléaire et le nerf vestibulaire sont étroitement liés et constituent le nerf auditif (VIII^e paire crânienne). Une fois constitué, le nerf cochléaire quitte le massif pétreux par le conduit auditif interne qui correspond à la base de la columelle. Il chemine alors, intimement accolé aux nerfs vestibulaire et facial, formant le paquet acoustico-facial, dans un espace libre endo-crânien (espace ponto-cérébelleux) pendant 2 à 3 cm, avant de pénétrer dans le bulbe.

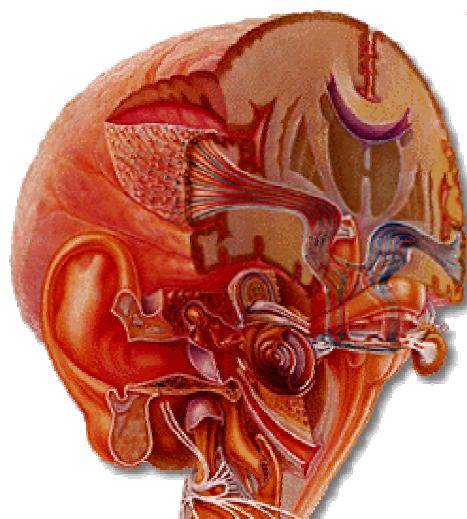


Figure II.7. : Trajet du nerf auditif, de la cochlée au cortex

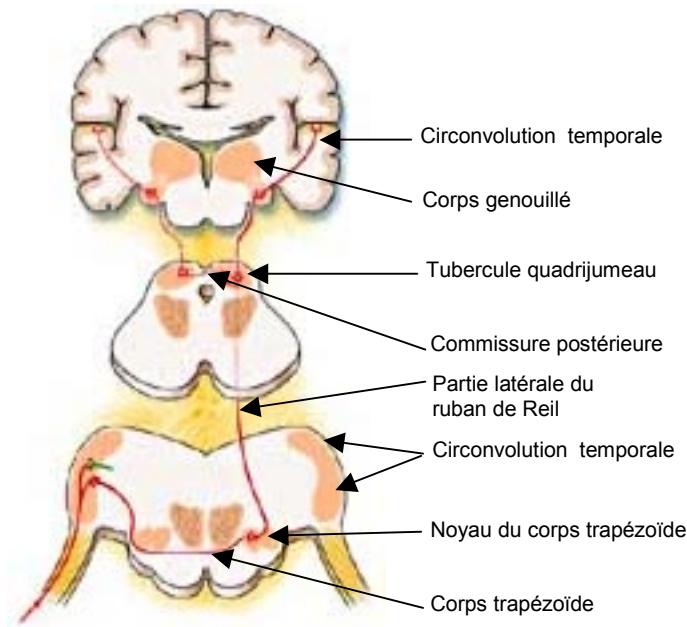


Figure II.8. : Voie auditive primaire

Aires auditives corticales. Les fibres de la voie cochléaire centrale se terminent dans la circonvolution transverse de Heschl située à la face supérieure du lobe temporal ; c'est l'aire primaire ou aire 41. Autour de cette aire primaire, existent deux aires para-auditives, les aires 42 et 22 (aires secondaires et tertiaires).

L'influx nerveux est transmis au cerveau depuis la cochlée par le nerf auditif, à une vitesse de 80 m/s sous forme d'un code traduisant l'intensité et la fréquence.

Pour un son pur de fréquence basse ou moyenne, la fréquence du potentiel d'action est identique à celle du son (jusqu'à 3 000 Hz). L'intensité met en jeu une augmentation des fibres activées. Au total, le codage de la fréquence et de l'intensité est fonction de la répartition des influx nerveux dans l'espace (nombre de fibres) et le temps (nombre d'influx sur chaque fibre). Dans les noyaux cochléaires, les neurones sont disposés en ordre pour les fréquences aiguës et graves, le nombre de neurones activés augmentant avec l'intensité du son comme le nombre de fibres. Mais le message auditif devient plus complexe dans l'identification de la fréquence et de l'intensité au fur et à mesure que l'on se rapproche du cerveau.

Noyaux et relais des voies auditives centrales. Le premier relais est bulbaire : 2 noyaux bulbaires, dorsal et ventral, renferment les deuxièmes cellules de la voie auditive. De là partent des fibres, qui, soit directes, soit croisées, montent dans l'axe cérébral par les rubans de Reil médian et latéral. À côté de ces fibres afférentes, il existe des fibres efférentes qui prennent naissance dans le tronc cérébral, se terminent dans l'organe de Corti, au niveau des cellules ciliées externes.

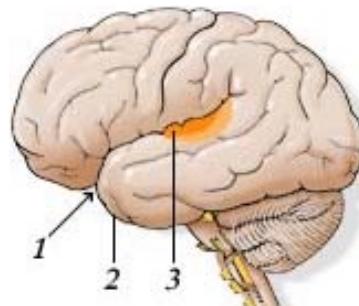


Figure II.9. : Cortex auditif (3) dans l'aire temporaire (2) chez l'homme. L'aire auditive au fond du sillon latéral (1 = sillon de Sylvius) est représentée en transparence

Temps cortical – Perception – Identification. On a pu trouver chez l'animal au niveau de l'aire corticale primaire, une présentation point par point de la cochlée, chaque zone corticale étant affectée à la perception d'une fréquence bien définie ; il existe donc une véritable carte tonale du cortex comme il existe une carte tonale de la cochlée. Du fait de la bilatéralité des voies, chaque cochlée se projette sur les deux cortex auditifs. La notion d'hémisphère dominant n'intervient donc pas pour le cortex auditif. Cela explique que l'audition ne soit pas ou peu perturbée après la destruction des aires auditives d'un seul côté. Par analogie avec les autres centres de projection corticale, on admet que les aires para-auditives jouent un rôle dans l'identification de la perception. L'aire auditive 41 de Brodmann correspondrait à la projection corticale des segments de la cochlée. Elle constitue la zone d'enregistrement des sons. L'aire 42 serait une zone de perception des sons permettant l'identification des mots. L'aire 22 joue un rôle capital dans la compréhension. Le fibres efférentes exercent sans doute une action inhibitrice et trophique au niveau de l'organe de Corti.

4. Codage de la fréquence

De façon simplifiée, on peut considérer qu'un stimulus sonore est caractérisé par son intensité et sa fréquence. L'intensité du son correspond à la sensation de « fort » ou « faible ». La fréquence correspond à la sensation de « grave » ou « aigu ». Ces deux notions sont liées aux caractéristiques de la période : l'amplitude de la période renvoie à l'intensité du son, et le nombre de périodes par intervalle de temps renvoie à la fréquence du son. Pour un son à 100 Hz, il y a 100 vibrations, donc 100 périodes par secondes. Un son à 100 Hz est un son relativement grave (en moyenne, la fréquence fondamentale des voix masculines parlées se situent entre 100 et 150 Hz). L'objectif de ce chapitre est de comprendre comment nous percevons la fréquence du son.

Code spatial des fréquences. La cochlée est le premier organe d'analyse des fréquences. Chaque fréquence correspond à une localisation particulière spécialisée au niveau de la cochlée. On sait que les fréquences hautes font vibrer la membrane basilaire au niveau de la base de la cochlée, et que les fréquences basses agissent sur la partie haute de la cochlée. En réalité, l'impulsion sonore provoque un mouvement d'ensemble de la membrane basilaire, mais chaque fréquence atteint sa plus grande amplitude en un endroit différent de la membrane. Ce codage spatial est répercuté le long des voies auditives nerveuses qui conduisent l'information jusqu'au cerveau. Ainsi les fréquences sont non seulement localisées au niveau de la cochlée mais correspondent aussi à des groupes spécifiques de neurones. C'est ce que l'on appelle la « tonotopie », qui permet une représentation cartographique de toutes les fréquences audibles en fonction des groupes cellulaires concernés. Cette théorie permet

d'expliquer la surdité sélective des métiers bruyants, limitée aux fréquences des bruits dominants. De la même façon, on explique l'effet de masque au cours duquel un bruit gêne la perception d'un son, d'autant plus que leurs fréquences sont proches. Cependant, la représentation spatiale des fréquences graves est assez peu spécifique, c'est à dire qu'elle est « étalée » sur la membrane basilaire. La tonotopie ne permet donc pas une représentation correcte de la hauteur pour les fréquences graves. D'où l'émergence d'une théorie prenant en compte le code temporel de la fréquence.

Code temporel des fréquences. Dans le codage temporel de la fréquence, la fréquence de la stimulation sonore est représentée par la durée des intervalles qui séparent des potentiels d'action successifs. La fréquence de décharge d'une fibre ou d'une cellule est liée à la fréquence du stimulus, le potentiel d'action est donc synchrone au son. Le nerf cochléaire est capable de suivre la fréquence sonore jusqu'à plus de 4,5 kHz. Or l'activité individuelle d'une fibre présente des potentiels d'action de fréquence nettement inférieure du fait de l'existence d'une période réfractaire d'environ 1 ms. En réalité, pour un son de fréquence élevée, toutes les fibres ne seraient pas mises en jeu simultanément, mais répondraient alternativement par groupes, une fois toutes les 2 ou 3 périodes de la stimulation sonore. Ce mécanisme est appelé « principe de la volée ». Le principe de la volée accepte donc la théorie de la localisation pour les fréquences élevées. En revanche, pour les basses fréquences, cette théorie explique la perception de la hauteur par la fréquence des influx nerveux.

Schématiquement, on peut donc considérer que les deux théories de codage de la fréquence s'opposent. La théorie de la localisation considère que le codage fréquentiel s'effectue entièrement au niveau de l'oreille interne, alors que la théorie temporelle dite de la volée situe ce décodage au niveau cérébral pour les basses fréquences.

5. Perception des sons simples

Hauteur subjective : mels. L'oreille humaine est capable de percevoir les sons dont la fréquence est comprise entre 16 Hz et 16000 Hz. A l'oreille, la sensation de différence de hauteur entre 250Hz et 500Hz est la même qu'entre 500Hz et 1000Hz. Autrement dit, la différence entre une vibration et la vibration deux fois plus rapide est perçue comme le même intervalle, quelle que soit la fréquence de départ ; c'est la notion d'octave. L'échelle des fréquences audibles (de 16Hz à 16000Hz) peut être décomposée en 10 octaves (échelle logarithmique de base 2) : 16-32 ; 32-64 ; 64-128 ; 125-250 ; 250-500 ; 500-1000 ; 1000-2000 ; 2000-4000 ; 4000-8000 ; et 8000-16000 Hz. Cependant, cette échelle n'est pas instinctive. Pour les musiciens, elle nécessite un apprentissage, car un son à l'octave est perçu comme la même note à une hauteur différente, mais pas comme un son deux fois plus aigu. En

1940, Stevens et Volkman [Stevens, 1940] font écouter à une population de jeunes adultes normo-entendants un son de 1000Hz et leur demandent de régler la fréquence du son suivant afin que sa hauteur soit le double de la hauteur du son de référence. En moyenne, les résultats montrent que les sujets régulent la fréquence du son à 3120 Hz. A partir de cette expérience, Stevens et Volkman proposent l'échelle des mels qui est une échelle des sensations tonales à partir de la perception subjective (figure II.10.).

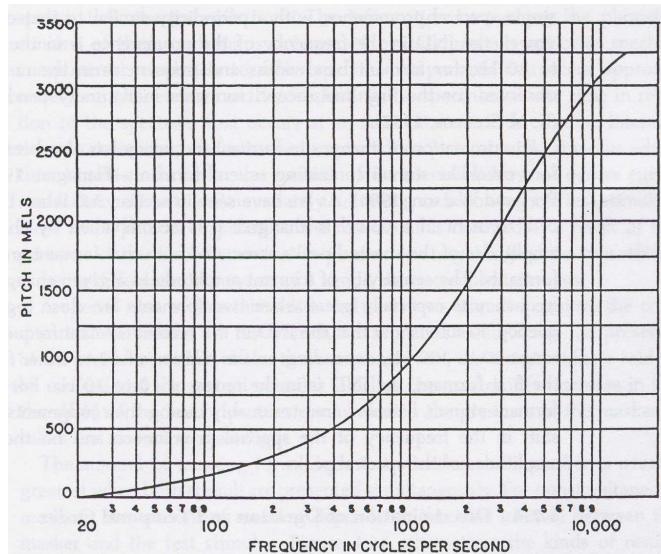


Figure II.10. : relation entre la fréquence (en Hz) et la hauteur (en mels) d'après les travaux de Stevens et Volkman [Stevens, 1940]

Le seuil de perception différentiel de la fréquence, c'est à dire le plus petit changement de fréquence que l'oreille humaine puisse percevoir, est de 3 Hz en dessous de 1000 Hz. Cela veut dire que, en dessous de 1000 Hz, toute différence de +/- 3 Hz est perçue et décodée par l'oreille comme un son différent, quelle que soit la fréquence du son de départ. Au delà de 1000 Hz, le seuil est de 3 %. En d'autres termes, au-delà de 1000 Hz, le seuil de perception différentielle varie en fonction de la hauteur : il est de 3 Hz à 1000 Hz ; 4,5 Hz à 1500 Hz ; 6 Hz à 2000 Hz, etc. On considère que dans le champ de l'audition, l'oreille est capable de discerner 1400 hauteurs différentes.

Intensité subjective. Le seuil de douleur provoquée par l'intensité maximale perçue à l'oreille se situe aux environs de 140 dB. L'intensité minimale quant à elle dépend de la fréquence du son (figure II.11.).

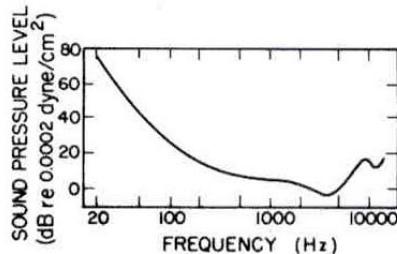


Figure II.11. : seuil d'intensité minimale en fonction de la fréquence du son

Les fréquences nécessitant le moins d'intensité pour être perçues sont comprises entre 1000 Hz et 4000 Hz. A 100-125Hz, l'oreille commence à percevoir le son à partir de 40 dB, soit une intensité 10000 fois plus importante qu'à 1000Hz.

De la même façon, l'intensité peut influencer à son tour la fréquence : pour des sons inférieurs à 1000Hz, la sensation de hauteur décroît avec l'augmentation de l'intensité, c'est à dire que pour une fréquence fixe, plus le son est fort, plus il paraît grave. A l'inverse, dans les fréquences élevées (supérieures à 4000Hz), plus le son est fort, plus il paraît aigu. Ces variations de hauteur peuvent atteindre 10% pour les fréquences extrêmes, mais ce phénomène est de moindre importance dans les hauteurs de gamme moyenne, où la sensibilité de l'oreille est la plus fine.

Le seuil différentiel d'intensité est ainsi fortement lié à la fréquence, mais aussi à l'intensité de départ : pour un son de 1000Hz, à 5dB au-dessus du seuil absolu, la plus petite variation d'intensité sera perçue à 3dB ; alors que pour le même son à 100dB au-dessus du seuil absolu, une différence sera perçue dès 0,25dB.

6. Perception des sons complexes

Effet de battement. Si on présente simultanément à l'oreille deux sons de fréquences très rapprochées, elle les confond en un son d'une seule hauteur dont l'intensité varie périodiquement. C'est l'effet de battement. La fréquence du battement est alors égale à la différence de fréquence entre les deux sons. Par exemple, l'émission d'un son à 1000Hz et d'un son à 1002Hz, produira un effet de battement à 2Hz, c'est à dire que la variation en intensité se produira deux fois par seconde. C'est la variation de phase des deux sons qui entraîne la variation périodique de l'intensité responsable de l'effet de battement. En musique, l'effet de battement est utilisé pour accorder deux instruments : lorsque l'effet de battement disparaît, cela signifie que les instruments sont strictement à la même fréquence.

Effet de masque. Un son de faible intensité n'est pas perçu en présence d'un son plus intense. Par exemple, la sonnerie d'un téléphone portable peut être couverte par le bruit de l'aspirateur. Dans ce cas, le bruit de l'aspirateur correspond au son masquant et la sonnerie de téléphone correspond au son masqué. Le seuil masqué correspond à l'intensité nécessaire pour que le son masqué devienne audible en présence du son masquant. Evidemment, ce seuil dépend de la fréquence des deux sons (figure II.12.).

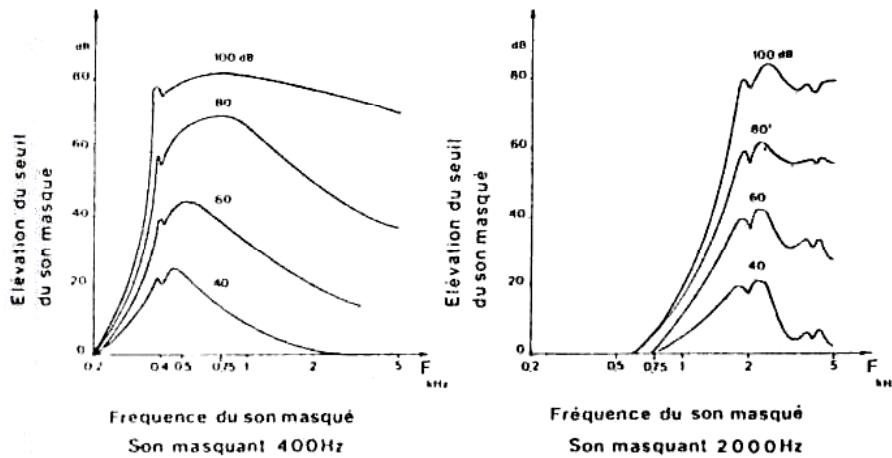


Figure II.12. : effet de masque d'un son pur sur un autre son en fonction de l'intensité du son masquant.

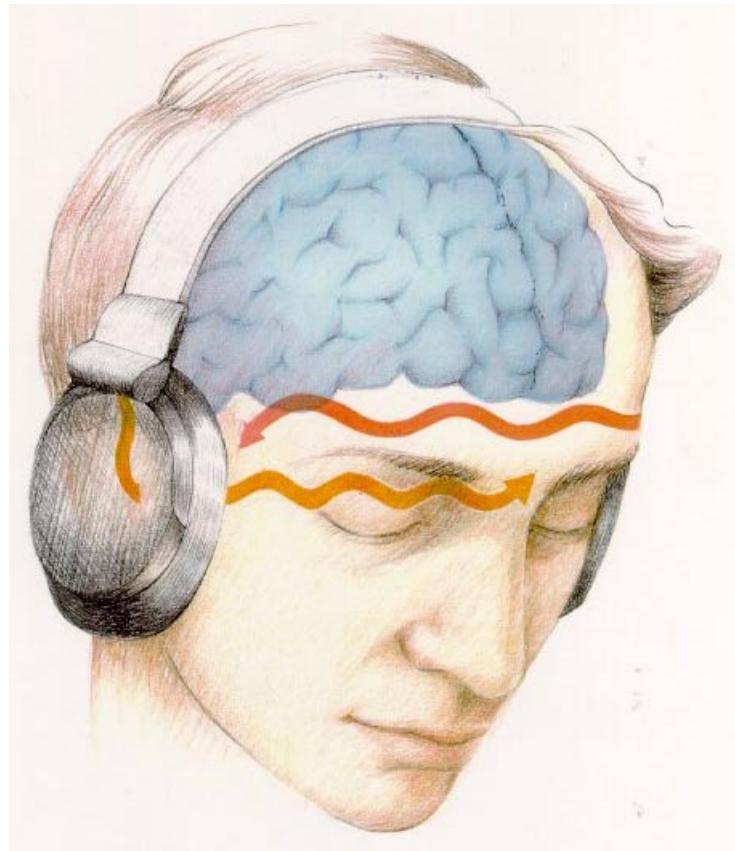
Hauteur des sons complexes. La hauteur d'un son complexe est déterminée par sa fréquence fondamentale. La fréquence fondamentale peut être physiquement absente du signal acoustique, elle détermine néanmoins la sensation de hauteur. En fait, la seule présence des harmoniques supérieurs permet d'en déterminer la hauteur. Cette perception est possible grâce à la périodicité de l'onde globale. Dans ce cas, la hauteur est intégrée par une analyse temporelle et non fréquentielle. Par ailleurs, la perception de la hauteur des sons complexes est moins dépendante de l'intensité que pour les sons simples.

Intensité et timbre. Suivant la théorie de la localisation, l'oreille effectue une analyse spectrale des sons complexes qu'elle reçoit. Cependant, le spectre physique des sons doit être pondéré en fonction des caractéristiques propres de l'oreille, notamment en ce qui concerne la non-linéarité subjective en intensité et en fréquence, ainsi que les effets de masque. Le spectre de sonie permet de tenir compte de ces particularités. Le spectre de sonie est moins redondant que le spectre réel. Il permet de rendre compte des principaux paramètres spectraux qui influencent l'audition et traduit le caractère global de la perception du timbre.

Au total, il faut considérer que les correspondances entre les paramètres acoustiques du son et le fonctionnement de la perception auditive sont particulièrement complexes. Les fondements acoustico-perceptifs développés dans ce chapitre sont basés sur l'étude de sons artificiels. Ainsi, même lorsque

l'on dispose de stimuli stables dont les paramètres sont connus et maîtrisés, il reste extrêmement ardu d'établir des lois permettant d'expliquer le processus de leur perception. Or, l'analyse spectrale des sons de la voix représente une analyse ponctuelle d'éléments noyés dans l'ensemble variable en continu qu'est la parole. Le signal vocal (et à fortiori de dysphonie) est loin d'être aussi simple que les sons étudiés ici. Il faut considérer que la perception de la voix résulte de la combinaison de facteurs multiples, captés de façon globale. Ce chapitre présente donc un survol du processus perceptif depuis l'anatomie de l'oreille jusqu'au fonctionnement acoustique de l'audition, mais nous devons conserver à l'esprit qu'une large part théorique sur la perception de la voix en soi, échappe encore à nos connaissances. D'une manière délibérée, nous avons choisi de ne pas développer les détails de ce qui est connu de la perception de la parole, la parole étant le support de notre champ thématique –la voix– mais pas spécifiquement notre objet d'étude.

La fonction auditive répond à une anatomie, une physiologie, et une physique particulièrement complexes, et se limiter à cette description, c'est se contenter d'observer comment se produit la sensation (voir Section I : La philosophie de la perception). Néanmoins, l'étude de la perception nécessite la compréhension de cette activité sensorielle. Nous retenons de ce chapitre que la perception auditive renvoie à des notions physiques qui sont à l'heure actuelle encore mal connues, mais qui constituent une base objective du « comment ça fonctionne ? ». La perception de la dysphonie renvoie, elle, au message implicite de qualité de la voix que l'on déduit de ce processus. Il ne s'agit plus là d'entendre mais d'écouter, et donc de comprendre l'activité d'interprétation de la sensation auditive produite par une voix. L'interprétation de cette sensation, la perception, développée dans le chapitre suivant : l'étude cognitive de la perception de la voix.



La cognition de la perception

La cognition de la perception

« *L'audition révèle un paradoxe similaire à celui rencontré dans tous les domaines de la perception : rien ne semble plus simple que de percevoir les sons de notre environnement et pourtant il s'agit là d'un phénomène particulièrement récalcitrant à l'analyse scientifique. Quelle difficulté par exemple à reconnaître son nom dans une conversation, à différencier le bruit d'une voiture de celui d'un avion à hélice, à percevoir le rythme entraînant d'un rock de Bill Haley, à reconnaître la voix de son enfant ou les bruits de pas d'une personne familière ? Il semble suffisant d'ouvrir grand ses oreilles pour cela.*

Imaginez cependant un seul instant, la quantité d'informations et le nombre de procédures qu'il serait nécessaire de donner à un ordinateur pour le rendre capable de distinguer un violon d'une flûte dans une polyphonie, de détecter un signal d'alarme dans le bruit de fond environnant, de saisir une relation entre un thème et ses variations, de repérer une anomalie de moteur à la seule écoute du bruit d'une voiture, de détecter un bruit inhabituel dans les battements de cœur d'un patient etc.. Malgré le très haut degré de sophistication technologique actuel, il est fort probable que ce type de question ne pourra être résolu qu'après de nombreuses années d'études en intelligence artificielle et en traitement du signal, tant la quantité d'informations nécessaire est grande et la façon dont elle doit être combinée complexe. »

Mac Adams et Bigand, 1994

L'oreille est l'organe de la vigilance : elle est perpétuellement en éveil. Pourtant, il existe une différence entre « entendre » et « écouter ». Les commandes cérébrales de l'audition permettent donc d'inhiber la perception des sons lors de la phase de sommeil par exemple, ou pour se concentrer sur un son en particulier.

1. L'attention sélective

L'attention sélective permet la perception consciente. C'est une focalisation perceptive qui amène à la prise de conscience. D'un point de vue cognitif, l'attention sélective est rendue possible par l'association des voies primaires (qui apportent l'information), et non spécifiques (qui permettent l'interprétation) [Boissacq-Schepens, 2000].

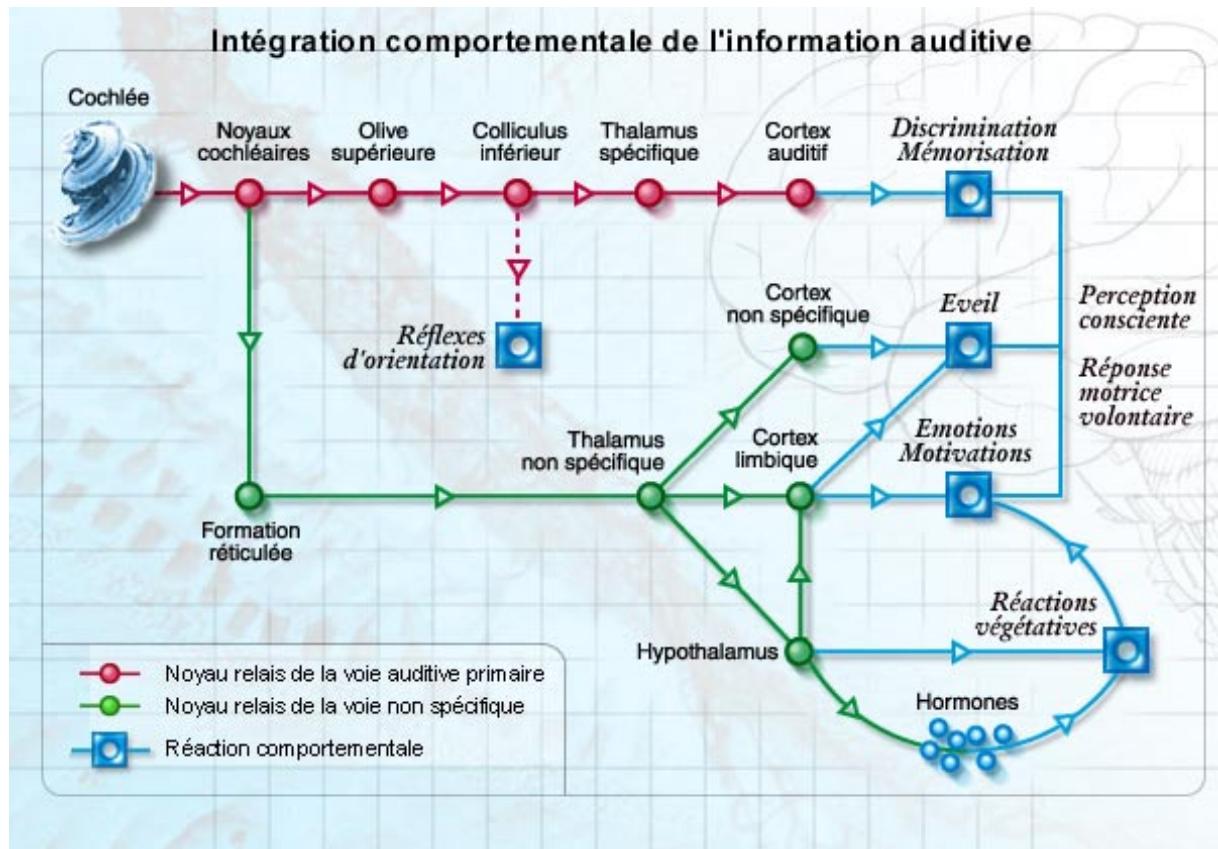


Figure III.1. : Intégration comportementale de l'information auditive.

L'intégralité et le bon fonctionnement des voies primaires et non primaires sont nécessaires à la perception consciente [Feldmeyer, 2002]. Par exemple, au cours du sommeil, la voie primaire fonctionne normalement (les sensations auditives sont décodées), mais cela n'est pas perçu conscientement, la liaison réticulée-centres de l'éveil n'étant plus assurée. De même, une pathologie affectant le cortex (coma dépassé) va supprimer la perception auditive tout en laissant s'exprimer les réactions réflexes et végétatives au son. L'audition permet donc le décodage simple et passif des informations sonores (c'est l'action d'entendre, voies primaires), mais aussi de se focaliser sur un message particulier (c'est l'action d'écouter, voies primaires + voies non-spécifiques).

2. L'effet « cocktail party »

L'attention sélective permet de percevoir le signal qui nous intéresse, parmi des distracteurs. En 1953, le psychologue EC Cherry a décrit l'effet « *cocktail party* » (figure III.2.). Dans le brouhaha ambiant d'une réception, nous sommes capables de nous focaliser sur une seule conversation. L'explication selon laquelle cette conversation est possible ne repose pas sur la proximité de l'interlocuteur, puisque l'expérience nous apprend que ce n'est pas toujours le message le plus fort qui est le mieux perçu. En réalité, l'objectif de l'auditeur est d'écartier les émissions sonores les plus intenses (comme la musique

ou les éclats des conversations voisines) pour se focaliser sur la conversation qui l'intéresse. L'effet « *cocktail party* », met en évidence un processus de traitement des stimulus par étape au cours duquel les stimuli sont triés en fonction de l'intérêt qu'on y porte. Les messages sonores sont donc sélectionnés ou rejetés par les processus attentionnels internes. Si nous n'avons aucune conscience des conversations que nous ignorons, c'est parce que nous filtrons les entrées non pertinentes avant qu'elles atteignent notre

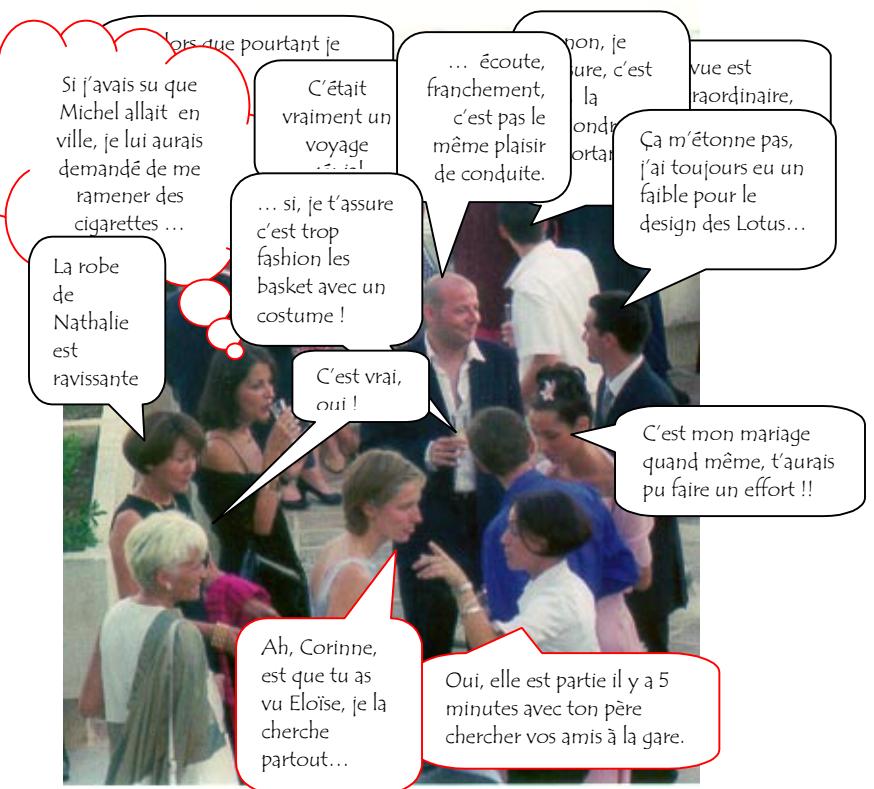


Figure III.2 : L'effet cocktail party

perception consciente. On distingue à ce niveau les systèmes de sélection précoce et tardifs. Dans les modèles de sélection précoce, il est possible de sélectionner les entrées avant que les caractéristiques spécifiques des stimulus aient été complètement analysées. Le concept de sélection précoce laisse donc entendre que l'attention (ou plutôt l'inattention) portée sur les informations négligées est capable d'altérer nos perceptions en filtrant les entrées ignorées à un niveau très bas des traitements auditifs. Inversement, la notion de sélection tardive presuppose que les entrées auxquelles il est ou non porté attention, font l'objet de traitements équivalents par le système perceptif et parviennent à l'étape d'encodage et d'analyse sémantique (signification). Ce n'est qu'après que se fait une sélection pour des traitements supplémentaires et pour une représentation au niveau conscient. Le concept de sélection tardive implique alors que les processus attentionnels n'affectent pas notre perception des stimulus.

Hillyard [Hillyard in Gazzaniga , 1998] a étudié l'attention sélective de la perception auditive en examinant les potentiels évoqués par des sons délivrés rapidement par des écouteurs stéréophoniques dans une situation du type « *cocktail party* ». Le but était d'exclure des effets non-spécifiques de vigilance et d'éveil. Les résultats ont montré que lorsque le sujet fait attention aux sons délivrés à une oreille et ignore ceux qui arrivent à l'autre oreille, les composantes à latence courte apparaissant 50 à 60 ms après le début du stimulus, sont fortement augmentées dans les potentiels évoqués par les sons envers lesquels l'attention est portée. Dans la lignée de cette étude, Woldorff a étudié conjointement l'enregistrement neuro-magnétique et la localisation des sources par IRM. Les résultats ont montré que

l'attention augmente l'activité sensorielle évoquée dans le cortex auditif primaire dès les 20 à 50 ms qui suivent le début du stimulus. Ces résultats confirment la théorie de l'attention précoce : les signaux sur lesquels se porte notre attention sont facilités, tandis que les sons d'environnement sont supprimés très tôt dans la tâche de traitement, avant que les caractéristiques complexes des stimulus soient analysées dans les aires corticales associatives.

3. Cortex de « voix sélective »

Dans la tâche spécifique de l'évaluation perceptive de la voix, le message qui sert de support à l'émission vocale est perçu par l'auditeur. Cependant, l'attention maximale est portée sur les caractères pathologiques du timbre davantage que sur le contenu sémantique. La voix humaine contient dans sa structure une grande richesse d'informations sur l'identité du sujet et son état émotionnel que nous percevons avec une facilité et une acuité remarquables. Bien que la perception des caractéristiques de la voix du locuteur joue un rôle majeur dans la communication humaine, on connaît peu de choses sur ses bases neurologiques. En 2000, Belin fait la couverture de *Nature* avec une étude sur le fonctionnement cérébral de la perception de la voix [Belin, 2000]. Son hypothèse est qu'il existe des zones cérébrales de « voix sélective » dans le cortex auditif humain, de la même façon qu'il existe des zones de « visage sélectif » dans le cortex visuel. Pour cela, il procède à trois expériences. L'expérience 1 a pour but d'identifier les régions de cerveau montrant une activité plus importante pendant la stimulation auditive avec les stimulus vocaux qu'avec les stimulus non-vocaux en utilisant un dispositif d'IRMf adapté pour les présentations auditives. 8 adultes droitiers ont été observés durant le silence et pendant l'écoute passive des stimulus des 2 catégories (figure III.3a.) : (1) des sons vocaux produits par plusieurs locuteurs de sexe et d'âge différents, incluant de la parole (par exemple : mots isolés, parole articulée dans plusieurs langues) et de la non-parole (rires, soupirs, toux) ; et (2) des sons non-vocaux d'énergie équivalente (sons de la nature, cris d'animaux, sons mécaniques) d'une grande variété de sources environnementales.

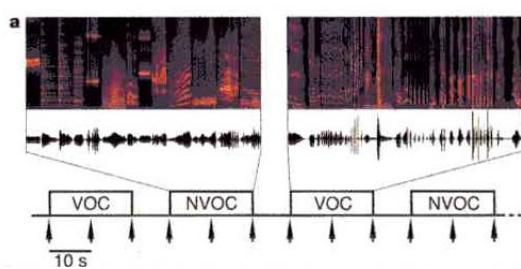


Figure III.3a. : spectrogrammes (0-4kHz) et signal acoustique des stimuli auditifs. Les stimuli vocaux (VOC) et non-vocaux (NVOC) sont présentés en blocs de 20-s avec 10-s de silence intercalé pendant que l'acquisition (flèches) intervient toutes les 10-s [Belin, 2000].

Chez chacun des 8 sujets, les sons vocaux ont montré de façon significative une activation plus importante que pour les sons non-vocaux, de façon bilatérale, dans plusieurs régions du cortex auditif

non-primaire. L'activation maximale de sensibilité à la voix est localisée le long du bord supérieur de la partie centrale du sulcus temporal supérieur chez 7 des 8 sujets (figure III.3b.). En moyenne dans le groupe de sujets, l'activité voix-sensible apparaît plus forte dans l'hémisphère droit, et est distribuée dans trois clusters bilatéraux le long du sulcus temporal supérieur (figure III.3c.) : un dans la portion antérieure proche du pole temporal, un dans la portion centrale approximativement au niveau de l'extension antérieure du gyrus de Heschl, et un dans le sulcus temporal supérieur, portion postérieure du gyrus de Heschl, dont les portions dorsales et caudales s'étendent au planum temporale dans le gyrus temporal supérieur. Il est important de noter qu'il n'existe aucune région d'activation plus importante pour les sons non-vocaux que pour les stimuli vocaux.

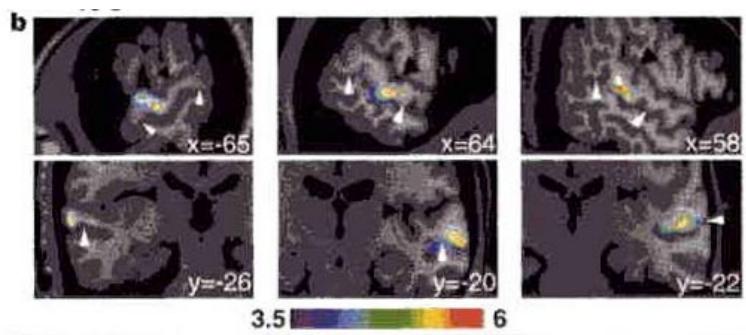


Figure III.3b. : activations individuelles « voix sensitive » chez trois sujets sur les images anatomiques sagittales (en haut) et coronales (en bas). Les flèches blanches indiquent la limite du sulcus temporal supérieur [Belin, 2000].

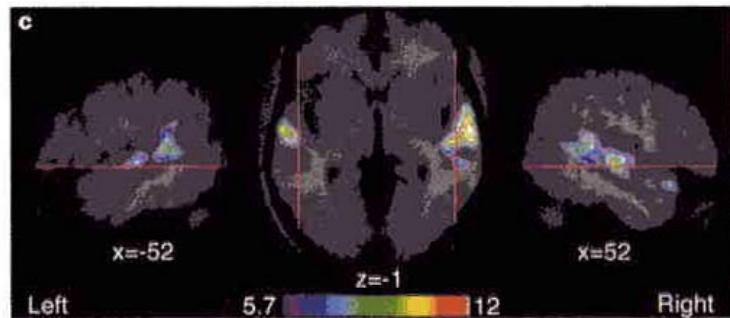


Figure III.3c. : activation « voix sensitive », moyenne du groupe. Régions fournissant une réponse significativement plus importante ($p < 0,01$) aux voix humaines qu'aux stimuli non-vocaux [Belin, 2000].

Ainsi, l'expérience 1 montre que le cerveau contient plusieurs régions qui sont sensibles à la voix, bien qu'il n'y ait pas encore d'évidence claire montrant qu'elles sont activées de manière sélective par la voix. Dans l'expérience 1, les stimuli vocaux et non vocaux étaient différents dans un certain nombre de caractéristiques acoustiques de bas niveau et les aires activées pourraient simplement fournir une réponse à quelques composantes acoustiques plus fortement présentes dans les stimuli vocaux. Cette remarque conduit Belin à proposer une seconde expérience avec le même groupe de sujets afin de déterminer s'il est possible de trouver marque d'activation sélective lors de la comparaison de stimuli vocaux avec des sons contrôle plus proches. Belin propose 4 types de stimuli : (1) l'enregistrement de différentes cloches, afin de vérifier si les mêmes régions sont activées par la présentation d'exemplaires d'une même catégorie, autre que la voix ; (2) des sons humains non-vocaux (par exemple claquements de doigts, applaudissements) afin d'examiner la possibilité qu'une aire supérieure réponde spécifiquement aux sons humains, quelle qu'en soit l'origine ; (3) un bruit blanc modulé avec la même enveloppe d'amplitude que les sons vocaux ; et (4) des voix brouillées,

préservant l'enveloppe d'amplitude des sons vocaux mais ne ressemblant plus à de la voix (figures 4a, b, c).

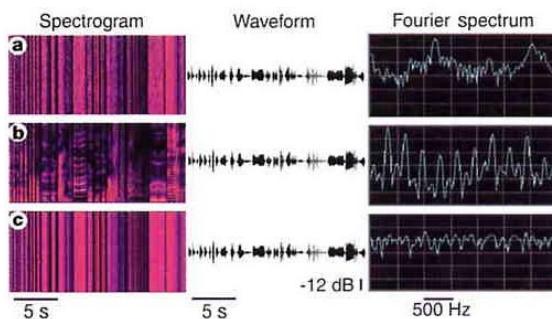
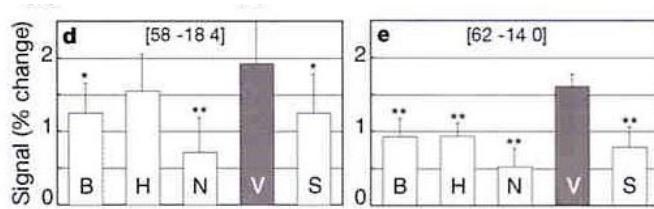


Figure III.4a., b., c. : stimuli de contrôle. Les sons provenant de voix brouillées (a) et de bruits blancs modulés en amplitude (c) conservent la même amplitude que les stimuli vocaux (b), mais diffèrent dans leurs caractéristiques spectrales [Belin, 2000]

Les résultats montrent qu'effectivement, les régions fournissant la plus forte réponse aux stimuli vocaux sont à nouveau distribuées le long du sulcus temporal supérieur ($p<0,01$), comme dans l'expérience 1. Les réponses neuronales sont différentes en fonction des classes de stimuli (figure III.4d., e.). Belin note que les réponses correspondant aux stimuli vocaux contenant de la parole ou ne contenant pas de parole obtiennent les mêmes réponses, indiquant que la réponse de voix-sensitive n'est pas entièrement due à la présence de parole dans le stimulus vocal. L'expérience 2 montre que les aires de voix-sensitive ne répondent pas simplement à la présentation de sons d'une même catégorie (cloches), ou aux sons d'origine humaine, puisque les réponses spécifiques aux stimuli vocaux sont plus importantes, particulièrement dans l'hémisphère droit. De plus, la structure fréquentielle joue un rôle plus proéminent dans l'activation des aires de voix sensitive que l'enveloppe d'amplitude, puisque les stimuli de bruit (3) et de voix brouillées (4) présentent des activation de moindre importance par rapport aux stimulus vocaux. Or ces stimuli 3 et 4 sont similaires à la voix en amplitude mais différents dans leur structure spectrale.



*Figure III.4d., e. : activation maximale des aires de voix-sensitive en fonction des 5 classes de stimuli dans l'hémisphère gauche (d) et droit (e). B = cloches ; H = sons humains non-vocaux ; N = bruit modulé en amplitude ; V = sons vocaux ; S = voix brouillées. * = $p<0,05$; ** = $p<0,01$ [Belin, 2000].*

Enfin, Belin propose une troisième expérience à partir d'un autre groupe de sujets afin de (1) proposer un contrôle de la distribution spectrale des stimuli vocaux et non-vocaux, dans le but de vérifier que les résultats observés dans les expériences 1 et 2 ne peuvent être attribués à une différence quelconque affectant ces variables, et (2) afin d'examiner comment l'activité des aires de voix sensitives (de même que la performance des sujets dans la tâche de perception de la voix) serait affectée par la modification de la structure spectrale des sons vocaux. Pendant l'acquisition, Belin présente aux sujets des corpus de stimuli vocaux et non-vocaux, chacun dans leur forme originelle, mais aussi après un filtrage spectral permettant de supprimer leurs hautes et basses fréquences. Les

auteurs s'assurent que la distribution spectrale des stimuli reste comparable pour les stimuli vocaux et non-vocaux (figure III.5a.). Après l'acquisition, les sujets sont soumis à une tâche perceptive lors de laquelle ils doivent désigner les sons vocaux ou non-vocaux ainsi que reconnaître le sexe du locuteur sur les mêmes stimuli.

Les résultats correspondant aux sons non-filtrés sont similaires que lors des expériences 1 et 2, et de même localisation. La figure III.5b. montre que l'intensité de l'activation de ces zones reste plus importante pour les sons vocaux ($p < 0,01$), mais décroît significativement avec le filtrage fréquentiel sans différence hémisphérique. Parallèlement, la performance des sujets dans leur deux tâches perceptive est décroissante avec le filtrage.

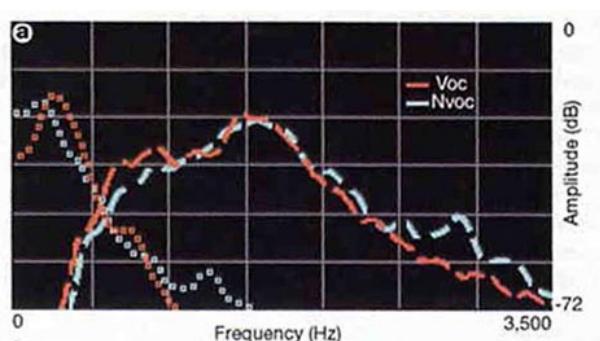


Figure III.5a. : distribution fréquentielle des stimuli filtrés en fréquence ; lignes tirets filtre passe-bande centré sur 1600Hz ; lignes pointillés filtre passe-bande centré sur 200Hz [Belin, 2000]

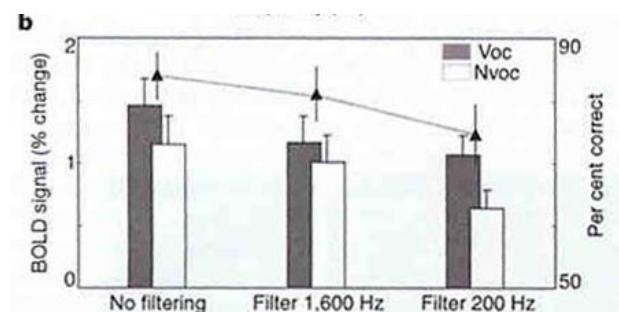
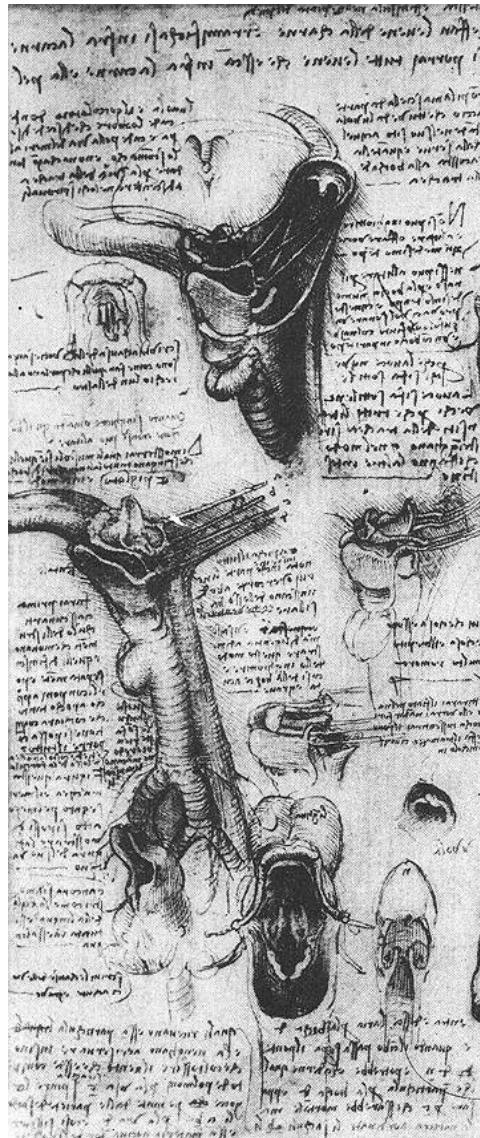


Figure III.5b. : activation maximale des aires de voix-sensitive. Cette activation reste meilleure pour les sons vocaux quelle que soit la condition de filtrage. Au dessus, les triangles noirs montrent les performances des sujets dans la tâche de reconnaissances perceptive. On note que l'évolution des performances est strictement parallèle à l'intensité des activations sous les différentes conditions [Belin, 2002]

Au total, ces trois expériences montrent que des régions de voix-sélective peuvent être trouvées de façon bilatérale le long du bord supérieur du sulcus temporal supérieur (STS). Ces régions montrent une activité neuronale plus importante lorsque les sujets écoutent passivement des sons vocaux, qu'il s'agisse de parole ou non, que pour les sons environnementaux non-vocaux. Les régions centrales STS montrent également un haut degré de sélectivité en répondant davantage aux sons vocaux qu'aux stimuli de contrôle appariés, incluant des voix brouillées et des bruits de modulation d'amplitude. De plus, la réponse aux stimuli dégradés par un filtrage de la fréquence est corrélée à la performance du comportement des sujets dans les tâches de perception de la voix qui utilisent ces stimulus. Les aires de voix sélective dans le STS pourraient donc représenter le pendant des aires de visage sélectif dans le cortex visuel humain.

Au total, il semble que l'écoute de la voix réponde à des activités neuronales spécifiques. La perception de la voix est un phénomène complexe qui n'a d'égal que la complexité de sa production. Afin de bien comprendre la tâche d'évaluation perceptive de la voix, il semble nécessaire de comprendre et d'étudier la physiologie de la phonation elle-même : quels sont les phénomènes intervenant dans la production, qui sous-tendent notre perception de la voix ? Quelles sont les défaillances du système en cas de dysphonie ? Comment modifient-elles notre perception ? Dans la section suivante, nous proposons une description du fonctionnement laryngé lors de la phonation.



« Anatomie du Larynx » L De Vinci

La physiologie de la phonation

La physiologie de la phonation

*« Tel se préforme au col de la cithare
 Un son naissant comme à l'embouchure
 d'un chalumeau le souffle s'introduit,
 pareillement ce murmure de l'aigle
 parut remonter sans retard ni relâche
 le long du col comme s'il était creux ;
 il y devint une voix, puis jaillit
 hors de son bec en forme de paroles
 que mon cœur attendait pour les inscrire. »*

Dante, 1316

Le terme de phonation décrit les processus physiologiques et physiques correspondant à l'apparition d'une vibration sonore au niveau des cordes vocales. L'énergie aérodynamique est fournie par la pression de l'air trachéal (« soufflerie pulmonaire »). La vibration sonore qui est la transformation de l'énergie aérodynamique en énergie acoustique apparaît au niveau des cordes vocales lorsqu'elles sont rapprochées sur la ligne médiane (« vibrateur »). Le son produit au niveau des cordes est aussitôt modifié, filtré par les cavités aériennes situées au dessus des cordes vocales jusqu'au niveau des lèvres (« résonateurs bucco-pharyngés »).

L'émission de la voix est un phénomène d'une grande variabilité non seulement entre individus mais encore chez un même individu selon les circonstances dans lesquelles il prend la parole et qui influencent les mécanismes corporels qui permettent l'apparition du son. Il existe en effet une définition étroite de la voix qui correspond à la vibration des cordes et une signification plus large qui correspond au message que la voix véhicule. La voix est en effet le support de la communication parlée. Elle permet à l'homme d'exprimer ce qu'il veut, ce qu'il pense, ce qu'il ressent. Le message peut être lui-même verbal (c'est-à-dire les mots prononcés) mais aussi non verbal. On sait que d'une manière générale, l'expression de l'émotion est largement vocale (peur, colère, tristesse, surprise, etc) et la voix donne des indications sur la personnalité. Par exemple, l'autorité est souvent associée à une voix forte et grave. De même, certains traits de la personnalité comme la gentillesse ou encore le caractère charmeur ou timide sont souvent associés à certaines caractéristiques vocales même si les études acoustiques spécifiques sont incomplètes. Plus encore, la voix est une véritable fenêtre sur la santé et la forme du locuteur. Un enrouement est ainsi considéré comme l'indicateur d'une affection

rhino-pharyngée et toute dysphonie traînant plus de quinze jours impose un examen des cordes vocales par un spécialiste, un tremblement de la voix peut être le premier signe d'une affection neurologique.

1. Anatomie du larynx

Le larynx correspond à la partie supérieure de la trachée qu'il surplombe. On peut considérer qu'il existe des éléments de soutien (cartilage thyroïde, cartilage cricoïde) dont la principale caractéristique est la rigidité pour maintenir ouverte la filière aérienne. Les autres éléments sont au contraire des éléments mobiles dont la fonction est de fermer le larynx. Dans ce groupe entrent les cartilages arytenoïdes, l'épiglotte et les muscles endolaryngés.

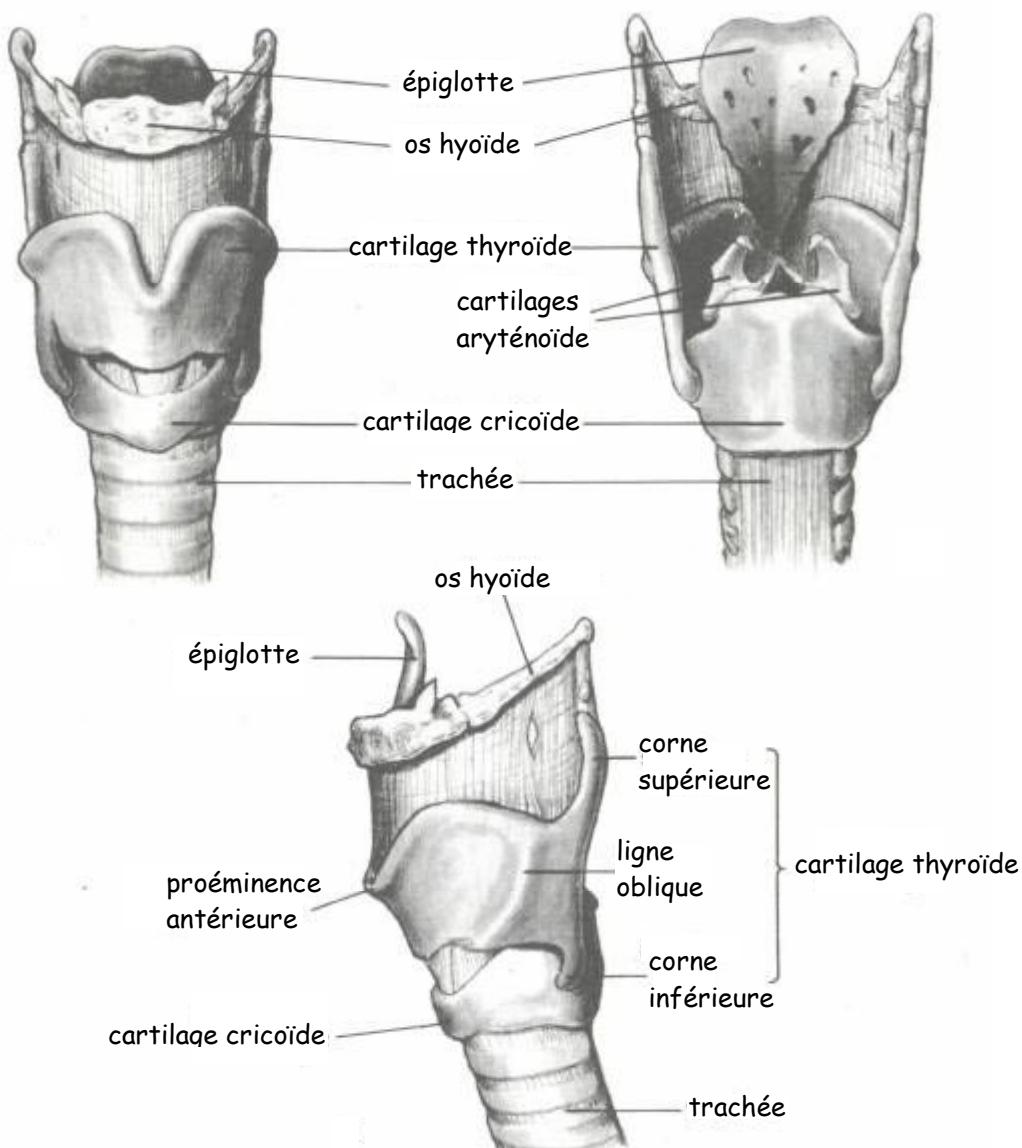


Figure IV.1. : les cartilages du larynx

Les muscles intrinsèques : Les muscles intrinsèques sont responsables de l'abduction (écartement), de l'adduction (rapprochement) et du réglage de la tension des cordes vocales.

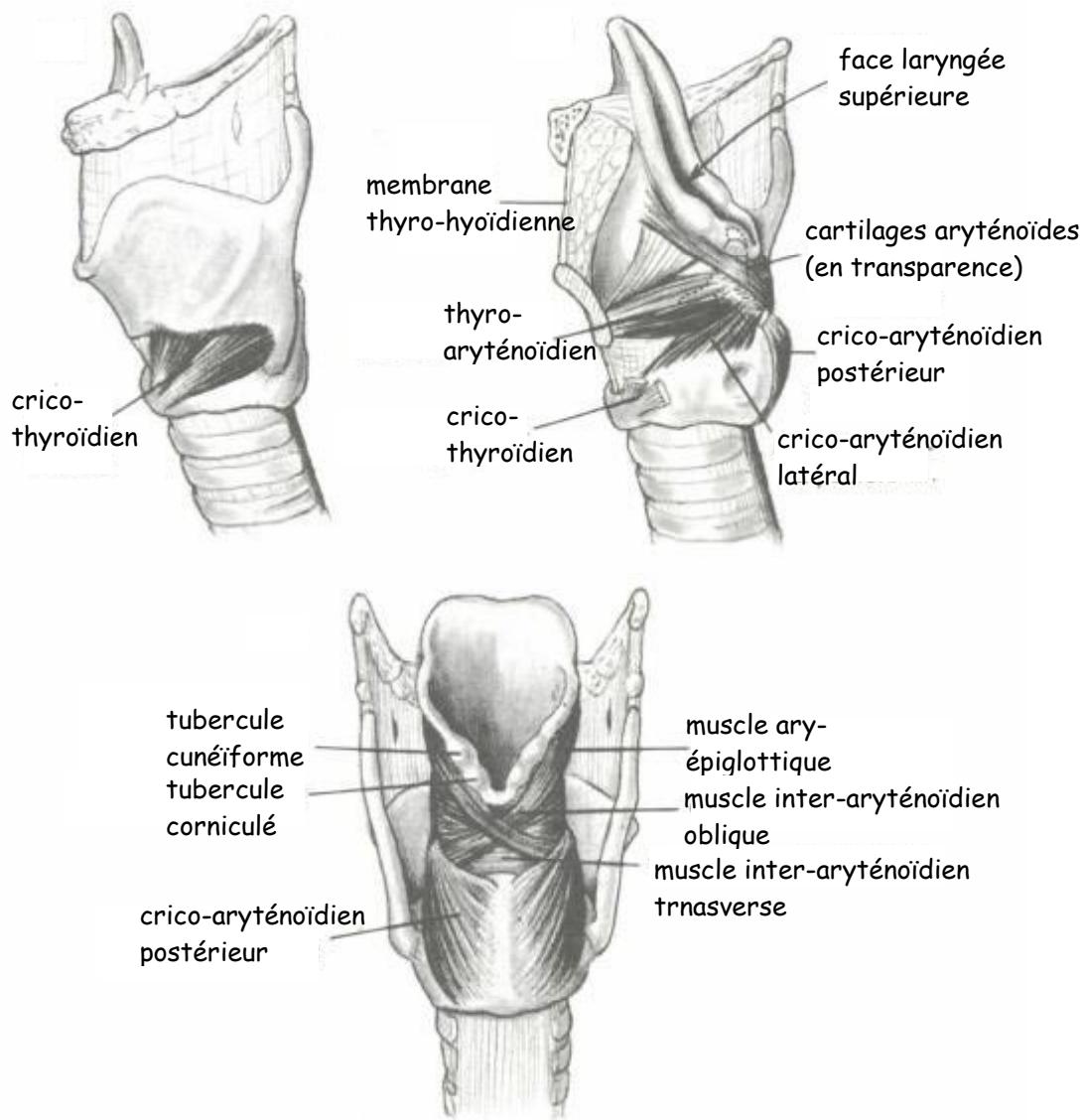


Figure IV.2. : les muscles intrinsèques du larynx

- Le muscle thyro-aryténoidien (TA) est innervé par le nerf laryngé inférieur (NLI). En se contractant, il raccourcit le corps de la corde vocale en l'épaississant. Il contribue à augmenter la raideur de la corde et à augmenter l'adduction, en particulier de la portion membraneuse de la corde.
- Le muscle crico-aryténoidien latéral (CAL) est innervé par le NLI. Il est responsable d'une rotation en dehors de l'apophyse musculaire et donc d'une rotation en dedans de l'extrémité antérieure de l'apophyse vocale. C'est un muscle adducteur qui abaisse, allonge et affine la corde vocale. Toutes les couches sont ainsi raidies et le bord libre de la corde prend une forme plus triangulaire.

- Le muscle inter-aryténoïdien (IA) est constitué de fibres transverses et de fibres obliques. Il est innervé par le NLI mais pourrait recevoir quelques fibres motrices du nerf laryngé supérieur (NLS). Il est responsable de l'adduction de la partie cartilagineuse des cordes vocales. Il est donc particulièrement important pour la fermeture de la partie postérieure de la glotte. Il a relativement peu d'effets sur la tension de la corde vocale elle-même.
- Le muscle crico-aryténoïdien postérieur (CAP) reçoit son innervation motrice du NLI. Son rôle est l'abduction des cordes vocales. Il est responsable de leur élévation et de leur allongement du fait d'une rotation du cartilage aryténoïde latéralement et vers l'arrière. Son action augmente la tension de toutes les couches de la corde.
- Le muscle crico-thyroïdien (CT) est innervé par le NLS. Sa contraction est responsable d'un abaissement d'une élongation et d'une mise en tension de la corde vocale. Cette contraction contribue à l'adduction des cordes vocales en position para-médiane ainsi qu'à une augmentation de la tension longitudinale de toutes les couches simultanément. Le muscle crico-thyroïdien rend ainsi plus fin le bord libre de la corde vocale.

Les muscles extrinsèques : Il s'agit d'un ensemble de lanières qui maintiennent la position du larynx dans le cou pour permettre une bonne efficacité de l'action des muscles endolaryngés intrinsèques.

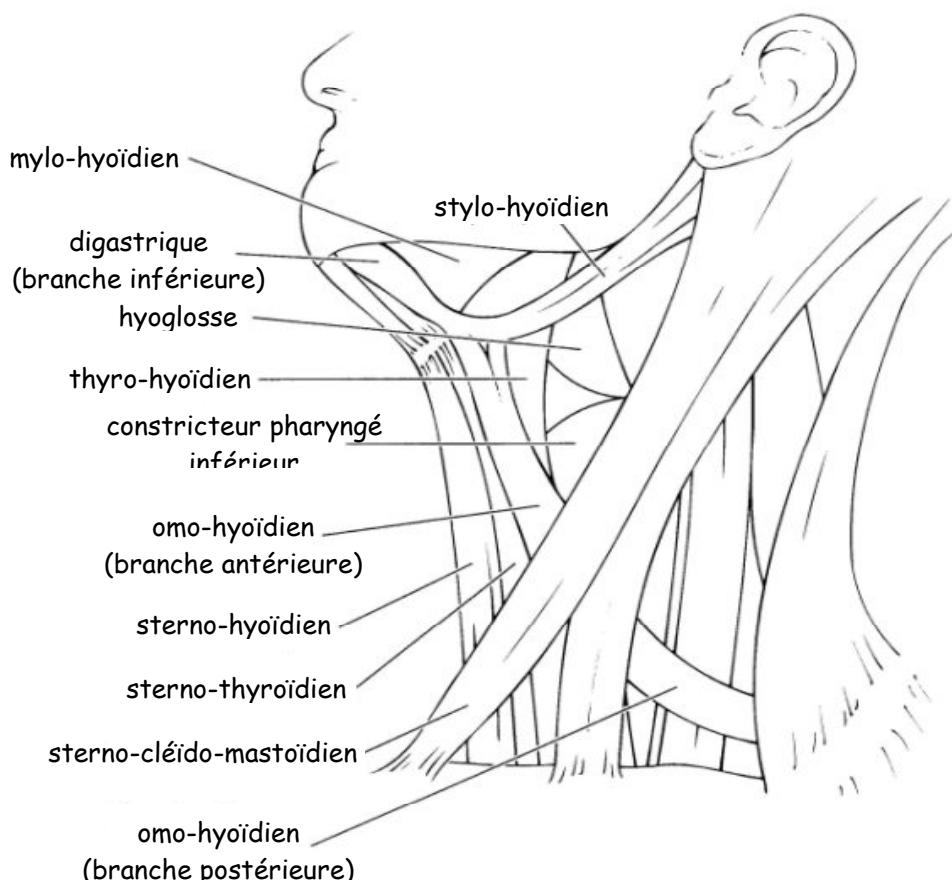


Figure IV.3. : muscles extrinsèques du larynx

Les muscles sous-hyoïdiens comprennent le thyro-hyoïdien, le sterno-thyroïdien, le sterno-cleïdo-hyoïdien et l'omo-hyoïdien. L'innervation motrice de tous les muscles sous-hyoïdiens est due aux branches descendantes de l'hypoglosse (XII nerf crânien).

- le thyro-hyoïdien : sa contraction rapproche l'os hyoïde et le cartilage thyroïde surtout en avant.
- le sterno-thyroïdien : sa contraction abaisse le cartilage thyroïde.
- le sterno-cleïdo-hyoïdien et l'omo-hyoïdien permettent d'abaisser l'os hyoïde.

Les muscles sus-hyoïdiens comprennent le digastrique, le mylo-hyoïdien, le génio-hyoïdien et le stylo-hyoïdien.

- la contraction du ventre postérieur du digastrique (sous le contrôle du nerf facial, VII^{ème} paire crânienne) élève l'os hyoïde vers l'arrière, la contraction du ventre antérieur (sous le contrôle du nerf trijumeau, V^{ème} paire crânienne) l'élève au contraire vers l'avant.
- le mylo-hyoïdien et le génio-hyoïdien : leur contraction élève l'os hyoïde et l'attire en avant.
- le stylo-hyoïdien : son action élève l'os hyoïde et l'attire en arrière.

On voit qu'une coordination précise entre tous ces muscles est nécessaire pour un positionnement précis du larynx dans le cou.

La corde vocale : elle correspond au muscle thyro-aryténoïdien, à son tissu fibreux de recouvrement et à la muqueuse en regard. Il s'agit d'une structure particulière dite « feuilletée » qui n'existe en réalité qu'au niveau de la partie dite « ligamentaire » de la corde. On peut décrire la muqueuse, l'espace décollable de Reinke et le ligament vocal qui recouvre le muscle vocal thyro-aryténoïdien. Les structures sous-muqueuses spécifiques de la corde vocale sont appelées lamina propria.

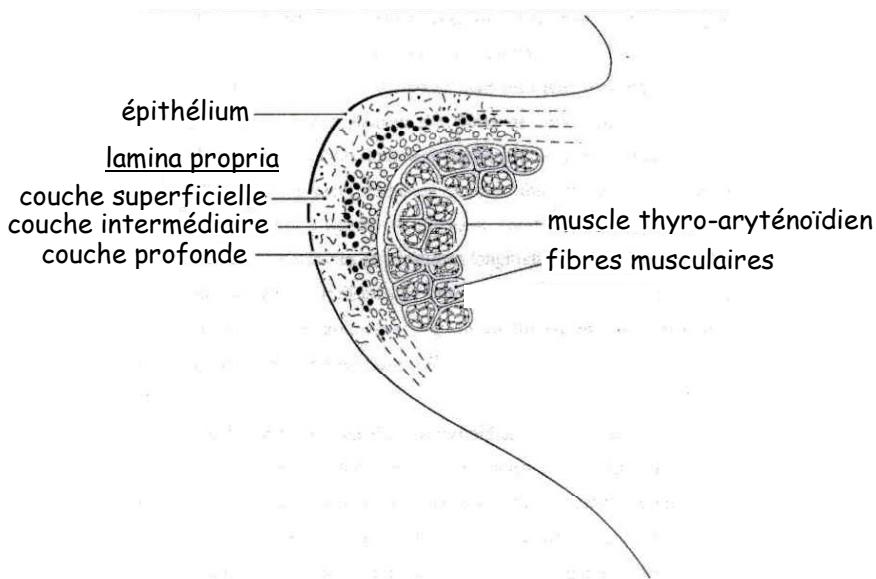


Figure IV.4. : structure en couches de la corde vocale

2. La vibration vocale

Les théories et modèles actuels dérivent tous plus ou moins de la théorie myo-élastique de Ewald. Dans un premier temps, les cordes vocales doivent être correctement positionnées de part et d'autre de la ligne médiane (position fermée pré-phonatoire) avec une tension appropriée. Elles opposent ainsi une certaine résistance à l'écoulement de l'air. La vibration commence avec la compression de l'air sous-glottique sous l'action des muscles expiratoires. Lorsque la pression de l'air sous les cordes vocales dépasse le seuil de cette résistance, le mécanisme cyclique de la vibration se met en route: l'air peut s'échapper à travers les cordes ce qui fait diminuer la pression sous-glottique. Dès lors, les cordes se referment à la fois sous l'effet de leur élasticité propre et grâce à l'effet Bernoulli. Lorsque les cordes vocales sont refermées le processus peut recommencer (figure IV.5.).

La production du son correspond donc à la transformation d'un mouvement continu (l'expiration active et la mise en position phonatoire des cordes vocales) en un phénomène périodique. Le mécanisme de base correspond à la résolution du conflit de forces entre les forces de fermeture des cordes vocales (contraction musculaire, élasticité) et les forces qui tendent à les écarter (pression sous-glottique).

La vibration glottique : C'est la vibration du bord libre de la corde vocale sous l'influence de l'air expiratoire (appelé ici air phonatoire). Les éléments-clés en sont d'une part les phénomènes de démarrage (notion de seuil phonatoire, c'est-à-dire de niveau minimum d'énergie pour que la vibration commence) et d'autre part les phénomènes d'auto entretien. Dans tous les cas, la vibration laryngée est conçue ici sous l'angle de la théorie myo-élastique qui fait de la vibration elle-même un phénomène passif dont le contrôle est obtenu par contrôle musculaire des forces mécaniques et aérodynamique de l'appareil vocal.

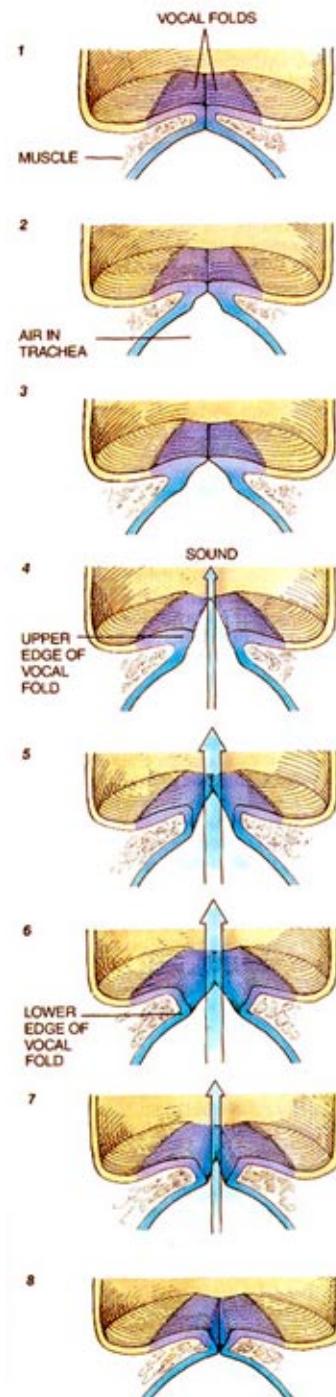


Figure IV.5. : la vibration des cordes vocales correspond à la résolution du conflit élastique entre la pression de l'air et la force de fermeture des cordes.

L'auto-entretien de l'oscillation à partir de la mise en route a fait l'objet d'une très abondante littérature. Le modèle le plus simple est le modèle dit à 1 masse qui fait référence aux mouvements observés au niveau du tiers moyen des cordes vocales (figure IV.6.). Dans ce modèle les cordes vocales sont ouvertes ou fermées; il s'agit d'un oscillateur faiblement amorti et dont l'oscillation est mise en jeu par un petit ébranlement de départ.

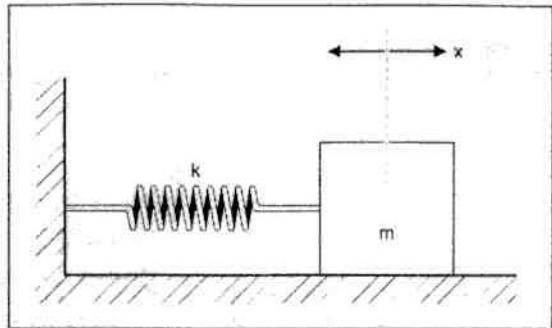


Figure IV.6. : le modèle à 1 masse

L'oscillation démarre lorsque la pression sous-glottique donne l'impulsion pour le début de l'oscillation. Lors de chaque cycle, l'ouverture glottique entraîne la compression de la masse musculaire c'est-à-dire du ressort du modèle à 1 masse. Cette compression met en jeu son élasticité de façon non-linéaire puisqu'elle augmente sa raideur et cette élasticité entraîne une force de rappel (*recoil*) qui provoque la fermeture glottique. Par ailleurs, le passage de l'air à travers le plan glottique entraîne un effet Bernoulli au niveau du bord libre des cordes vocales et qui tend à les refermer. Enfin l'asymétrie entre la pression sus et sous-glottique contribue à entretenir ce cycle : lorsque l'air a franchi le plan glottique et que le phénomène de Bernoulli a entraîné la fermeture des cordes, la pression immédiatement sus-glottique diminue du fait que l'air continue son chemin à travers le tractus vocal. Pendant ce temps la pression sous-glottique augmente à nouveau et le cycle peut se reproduire à la fois du fait de la « sur »pression sous-glottique et de la « dé »pression sus-glottique. Il existe des relations entre la pression intra glottique c'est-à-dire la pression observée au niveau des cordes vocales et le débit d'air. Lorsque la pression intra-glottique est positive les cordes vocales sont écartées et l'air passe, le débit augmente. Lorsque les cordes se referment, la pression intra glottique baisse à cause de l'inertie de la colonne d'air et le débit d'air chute. L'asymétrie nécessaire pour que la vibration se produise est donc une asymétrie de phase correspondant à l'inertie de la colonne d'air.

Les observations du mouvement des cordes vocales en cinématographie ultra-rapide et plus simplement la stroboscopie ont révélé au niveau des cordes vocales des mouvements trop complexes pour que le modèle à 1 masse en rende compte complètement : il existe en réalité une ondulation de la muqueuse, une vague qui progresse de bas en haut. La fermeture commence à la partie inférieure sous-glottique des cordes vocales et se propage vers le haut. La fermeture commence à la partie inférieure sous-glottique des cordes vocales et se propage vers le haut. Il en résulte que le mouvement observable à la surface des cordes doit être schématisé avec deux oscillateurs en

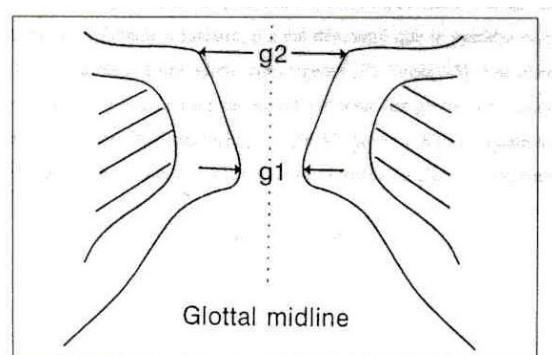


Figure IV.7. : l'ondulation de la muqueuse

opposition de phase situés l'un au dessus de l'autre (modèle à 2 masses de Ishizaka et Flanagan). Le mouvement d'ouverture fermeture est ainsi complété par un mouvement vertical (composante verticale de la vibration). En fait le modèle le plus couramment utilisé est le modèle à 3 masses, encore appelé *body-cover model* de Titze décrit initialement pour modéliser le réglage de la hauteur et qui complète le modèle à 2 masses.

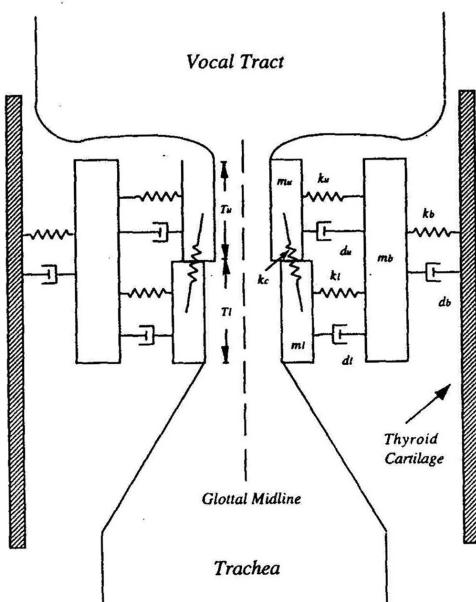


Figure IV.8. : le modèle à 3 masses dit modèle "body-cover"

Sur le plan pratique il en résulte que la forme de la corde vocale, son profil, son épaisseur sont d'une grande importance pour la mise en route et l'auto entretien de cette oscillation. Lorsque les cordes vocales sont convergentes c'est-à-dire que la lèvre supérieure est fermée et la lèvre inférieure ouverte, la pression intra-glottique augmente et tend à ouvrir les cordes vocales. A l'inverse lorsque la lèvre inférieure est fermée et la lèvre supérieure ouverte, la glotte est dite divergente et la pression intra-glottique diminue tendant à faire refermer les cordes vocales.

Certaines études particulièrement dans des situations pathologiques ont montré l'existence de modes vibratoires complexes dans lesquels la totalité de la corde vocale ne

vibre pas de la même manière. Cependant ces degrés de liberté ne s'expriment qu'à la faveur d'asymétries ou d'anomalies plus ou moins importantes de la corde vocale ou de la muqueuse. Des études plus récentes font état de modèles « continus » encore plus complexes permettant de reproduire avec fidélité la complexité des mouvements observés en clinique et particulièrement en pathologie.

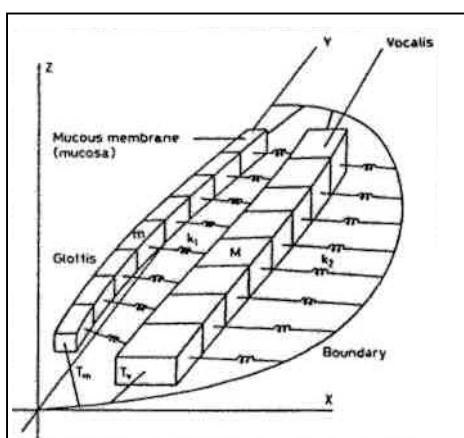


Figure IV.9. : Les modes complexes de vibration

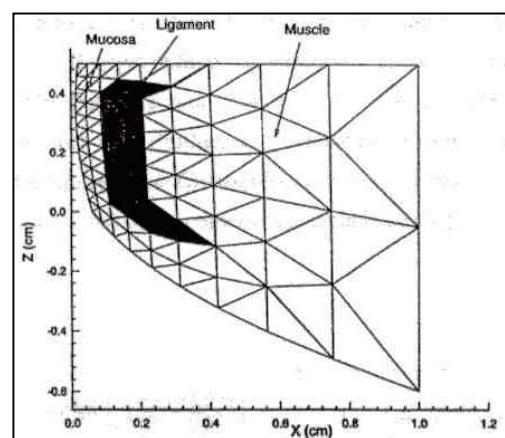


Figure IV.10. : les modèles à continuum de masses

La notion de seuil phonatoire : Le seuil de pression phonatoire serait selon Titze le « chaînon manquant » de notre connaissance de la physiologie des cordes vocales. Le seuil à partir duquel les phénomènes cycliques de la vibration de produisent, dépend de plusieurs facteurs :

- la raideur de la partie vibrante de la corde vocale,
- sa viscosité,
- l'épaisseur du bord libre,
- la largeur de la fente glottique pré-phonatoire,
- la différence de pression transglottique.

Ce seuil est plus élevé dans un certain nombre de circonstances. On sait que l'élévation de la hauteur de la voix correspond à une élévation de la tension de la corde vocale. On conçoit donc que le seuil de pression sous-glottique soit plus élevé. De même si la pression sus-glottique est élevée, la pression sous-glottique nécessaire sera importante. Enfin dans les cas de pathologie comme les nodules ou les polypes des cordes vocales, l'augmentation de la raideur de la muqueuse est responsable d'une augmentation du seuil de pression phonatoire. Lorsqu'il existe une paralysie laryngée unilatérale en position latérale, la fente glottique pré-phonatoire est trop large et il est nécessaire d'augmenter la pression sous-glottique pour faire démarrer la vibration glottique. En pratique le seuil de pression phonatoire normal est de l'ordre de 2 à 4 hPa et la pression sous-glottique usuelle est typiquement de 7 hPa. Cette pression est plus élevée dans les situations pathologiques. Sa valeur est un indicateur assez précis du degré de forçage vocal.

Les moyens de diminuer le seuil de pression phonatoire sont nombreux. La diminution de la vélocité des tissus en général peut être obtenue par une amélioration de l'hydratation permettant de diminuer la viscosité des tissus. La diminution de la vélocité de l'ondulation muqueuse est un autre moyen de diminuer le seuil de pression phonatoire. Cette diminution peut être obtenue en diminuant la tension de surface (voix plus grave) ou encore en hydratant le mucus de surface. La diminution de la largeur glottique pré-phonatoire peut être obtenue par une augmentation modérée de la fermeture des cordes vocales en augmentant un peu le serrage musculaire. Dans le cadre des paralysies laryngées c'est pour diminuer la fente glottique pré-phonatoire que sont réalisées les médialisations. Enfin il peut être utile d'augmenter l'épaisseur de la corde vocale par exemple en utilisant une voix plus grave ou dans certains cas en changeant de mécanisme vocal (lourd, léger).

3. Le réglage de la hauteur

Le réglage de la hauteur de la voix correspond au réglage de la fréquence de la vibration. Il peut s'agir du réglage de la masse et/ou du réglage de la tension. L'épaisseur de la corde vocale a été démontrée comme ayant un effet sur la hauteur de la voix : si l'épaisseur diminue, le ton de la voix (pitch) est plus aigu, aussi bien chez l'homme que chez la femme. En réalité ce qui compte n'est pas la masse totale de la corde vocale mais sa masse vibrante qui peut n'être qu'une partie de la corde en fonction de la hauteur et en fonction de l'intensité.

La fréquence vibratoire de base : La hauteur de la voix est différente chez l'homme en rapport avec les caractéristiques de taille et de poids des cordes vocales en fonction de l'âge comme le montre le tableau ci-dessous.

	Poids	Taille	Longueur Corde	Longueur Aryténoides	Fo
Nouveau-né	3,5 kg	0,50 m	2 mm	2 mm	500 Hz
Enfant 8 ans	30 kg	1,2 m	6 mm	3 mm	300 Hz
Femme adulte	60 kg	1,60 m	10 mm	4 mm	200 Hz
Homme adulte	75 kg	1,80 m	16 mm	4 mm	125 Hz

Tableau IV.1. : caractéristiques vocales en fonction de l'âge

Les mécanismes de la modification de la fréquence : D'une manière générale la fréquence de vibration d'un dispositif quel qu'il soit est inversement proportionnel à sa taille. La Fo est donc plus basse si l'instrument est plus gros. La masse de la corde vocale correspond à la quantité de matériel effectivement en vibration. Cette quantité est une fonction assez complexe de la géométrie glottique et de la contraction musculaire.

La raideur est largement fonction de la longueur mais uniquement s'il s'agit d'un tissu inerte. Dans ce cas, si les fibres sont étirées, elles sont plus tendues et plus raides. Dans le cas du muscle vocal, un muscle en action (raccourci) a une raideur plus grande. De plus, la contraction du muscle antagoniste peut entraîner un changement de la tension sans changement de la longueur (contraction isométrique). Ainsi si le CT est contracté et le TA relâché, la totalité de la longueur de la corde est augmentée, sa raideur est globalement augmentée dans toutes ses couches, et la Fo est augmentée. Inversement si le TA est augmenté et le CT désactivé, la rigidité de la masse musculaire augmente et la Fo augmente malgré la diminution de la longueur de la corde. Il existe donc un contrôle différentiel de la Fo par ces

deux muscles. En fait les capacités élastiques des cordes (module élastique) sont élevées. Il faut donc une augmentation importante de la longueur pour que la tension soit significativement modifiée.

La tension dans le cover peut également être augmentée par l'amincissement de la tranche de section de la vibration. La profondeur de la vibration peut être régulée par l'activité du TA. Le relâchement du TA fait porter au ligament vocal la majeure partie de l'étirement. Le ligament vocal permet d'augmenter considérablement l'étirement dans la corde vocale sans entraîner de stress effectif dans la muqueuse. Dans les hautes fréquences le ligament vocal assume donc la plus grande partie de l'étirement, la muqueuse restant relativement laxe. Ainsi la muqueuse de recouvrement peut être libre de propager une ondulation muqueuse dont la vitesse relativement basse peut être utile pour réguler le transfert de l'énergie aérodynamique vers la vibration glottique. Un ligament vocal tendu et une muqueuse « libre » et souple semblent donc la configuration idéale pour la phonation dans l'aigu. Dans les basses fréquences et dans les fréquences intermédiaires, la portion musculaire du body peut participer à l'étirement et à la mise en tension. Ainsi la muqueuse et le ligament vocal peuvent rester laxus tous les deux et permettre une plus grande « profondeur » de la vibration dans le registre modal.

La prédictibilité de la Fo en relation avec l'activité du TA reste faible. Cette imprédicibilité peut être attribuée au fait qu'une portion significative du body (cad du corps du TA) soit en vibration dans les fréquences graves et à forte intensité en particulier.

4. Le réglage de l'intensité

Il correspond au réglage de l'amplitude de la vibration par la combinaison des réglages de la pression sous-glottique, de la géométrie glottique, de la force d'adduction des cordes vocales et de la géométrie du tractus vocal dans son ensemble. L'intensité est mesurée en dB et fait référence à la pression sonore (SPL pour Sound Pressure Level), proportionnelle au carré de la distance entre la source et le dispositif de recueil.

La pression sous-glottique dépend de la pression pulmonaire qui est en rapport avec la pression imposée par les forces qui régissent l'appareil respiratoire. Il existe une relation quasi-linéaire entre la pression pulmonaire et trachéale et l'intensité du son. En fait si on augmente la pression sous-glottique, toutes choses étant égales par ailleurs, la fréquence augmente en même temps que la pression. Pour compenser cette élévation de la fréquence, il est nécessaire d'augmenter la force d'adduction des cordes et d'augmenter ainsi le temps de contact entre les cordes. L'augmentation du temps de contact est un facteur qui s'oppose à l'augmentation de fréquence et qui permet la transformation en augmentation de l'intensité vocale. Que ce soit chez les hommes ou chez les

femmes, l'augmentation de l'intensité va de pair avec la diminution du temps pendant lequel les cordes vocales restent ouvertes.

Plus l'adduction est importante et plus le temps de contact est important, plus l'onde glottique est à pente abrupte. La pente de fermeture est d'autant plus abrupte que les conditions de phonation sont plus intenses. La forme de la courbe de fermeture de l'onde glottique peut être définie comme le taux de déclinaison maximum du flux TDMF. Une diminution du quotient ouvert et une augmentation du TDMF correspondent à une augmentation de la force d'adduction c'est-à-dire schématiquement à une augmentation de l'intensité des conditions phonatoires (forçage vocal).

5. Production vocale et dysfonctionnement

On peut se demander s'il existe un degré d'adduction idéal : si les cordes vocales ne se touchent pas, la voix n'est ni forte ni même bonne ; si elles sont trop serrées (force d'adduction trop importante), la voix est « serrée », « étranglée » et il s'agit d'un effort vocal. Une configuration idéale semble correspondre à une situation où les cordes vocales sont presque en contact avant la mise en phonation (diminution de la largeur glottique pré-phonatoire). Cette configuration assure un fonctionnement presque libre des cordes qui sont ainsi capables d'exprimer leurs modes de vibration. En pratique, le signal produit est presque sinusoïdal. Ce type de fonctionnement correspond à ce que les pédagogues du chant appellent « la voix qui flotte sur le souffle » ou une image mentale voisine. On peut supposer que l'image de parler ou de chanter sur le souffle est utilisée pour ajuster la résistance glottique de manière à assurer le meilleur rendement possible de la conversion de l'énergie aérodynamique en énergie acoustique en modifiant le moins possible la vibration des cordes vocales.

Lorsqu'il est nécessaire d'augmenter l'intensité, la forme de l'onde glottique change, ce qui correspond à un mode de fonctionnement glottique plus proche d'un modèle « ouvert-fermé » du type oscillateur à relaxation. Ce mode de fonctionnement a un rendement plus faible et une énergie importante est dissipée au niveau des cordes vocales sous forme de frottements qui sont susceptibles d'entraîner une inflammation locale et même des lésions des cordes. Ces lésions, appelées, lésions dysfonctionnelles, siègent préférentiellement au niveau de la zone des cordes vocales sur laquelle les phénomènes d'accolement sont les plus importants, c'est-à-dire le tiers moyen.

Dans la parole l'augmentation des tensions dans l'appareil vocal correspond à un ensemble comportemental de la part du locuteur qui est appelé « forçage vocal » lorsqu'il tend à se pérenniser. En réalité, les configurations glottiques correspondent au-delà du fonctionnement des cordes vocales à tout l'ensemble des phénomènes physiologiques qui sous-tendent la communication. Dans une communication parfaitement « détendue », les phénomènes de tension musculaire sont au minimum.

Dans une communication dont l'objectif est de modifier le comportement de l'interlocuteur (voix projetée), une tension musculaire plus grande est requise pour produire une voix plus forte et plus « efficace ». En réalité, il existe une augmentation de la tension à tous les niveaux de l'organisme : il a été démontré que tous les muscles, y compris les muscles de la posture voient leur activité augmenter. Au niveau comportemental, on observe en général une attitude de redressement du corps avec un regard tourné vers l'interlocuteur. Au niveau respiratoire, on observe une inspiration plus ample (anticipation pré-phonatoire) pour faire face à l'augmentation prévue de la pression sous-glottique. Il est à noter que certains sujets peuvent avoir des difficultés à obtenir le relâchement musculaire nécessaire à une inspiration profonde suivie d'une expiration contrôlée par les muscles de la paroi abdominale. Au niveau postural, on observe également une augmentation de la raideur de tous les muscles, qu'il s'agisse des muscles du cou et du larynx ou de muscles situés plus à distance comme les muscles des mollets et du dos. Les modifications de l'activité musculaire en rapport avec l'augmentation de l'intensité sont donc génératrices d'une dépense d'énergie. Normalement le sujet compense cet excès d'énergie par un repos réparateur. Mais dans certains cas, le sujet n'observe pas ce repos et risque l'apparition de complications à type de laryngopathies dysfonctionnelles (malmenage vocal). On comprend que la prise en charge rééducative des patients présentant un malmenage vocal ne puisse se résumer à la modification de la configuration glottique même si elle est nécessaire et que la prise en charge d'aspects aussi divers que la tension musculaire générale, le niveau de stress, la posture, la respiration pré-phonatoire soient des éléments indispensables d'une rééducation.

6. La transformation du son en voyelles

Toutes les réalisations vocales humaines nécessitent une configuration du tractus vocal. L'idée générale de la théorie « source-filtre » est que le son source provient du débit d'air pulsé au niveau glottique. Ce son contient de nombreuses fréquences. Puis le son est filtré par le tractus vocal qui sélectionne un certain nombre de ces fréquences pour qu'elles irradient au niveau de la bouche. Il s'agit des fréquences qui « résonnent » avec le tractus vocal.

La notion de résonance dans un tube correspond à des interférences entre des ondes soumises à de multiples réflexions. Le tractus vocal humain comme le corps d'un instrument à vent résonne pour certaines fréquences de la source. Ces fréquences dépendent de la forme du tractus vocal ce qui détermine l'apparition des sons du langage (phonèmes).

Un formant est une résonance du tractus vocal. Les voyelles sont perçues et classifiées sur la base des deux formants les plus graves appelés F1 et F2. Pour schématiser, le formant grave correspond au pharynx, et le formant aigu à la cavité buccale. Ainsi du point de vue de la perception des voyelles, le

processus de filtrage aboutit à une simplification du code qui doit être perçu par l'auditeur. Dans la phonétique articulatoire classique, les voyelles sont décrites en fonction de la position haute ou basse de la langue ainsi que de sa position reculée ou avancée : pour le /a/ la langue est basse et reculée ; pour le /i/, elle est haute et avancée ; pour le /u/, la langue est haute et reculée. Ces trois voyelles déterminent donc sur une représentation orthonormée F1 et F2, le triangle vocalique.

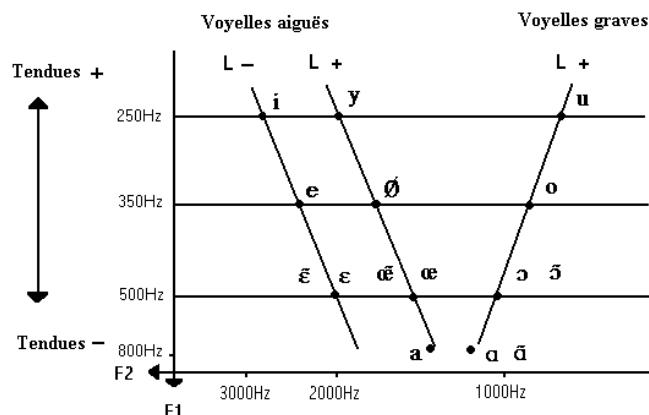


Figure IV.11. : le triangle vocalique

La forme du tractus vocal pour le son /ø/ est approximativement tubulaire avec une section à peu près uniforme du fait d'une position neutre de la langue. La position de F1 est environ 500 Hz et de F2 à 1500 Hz. On peut voir sur le même schéma la position de F1 et de F2 lorsque la position de la langue varie. En pratique on peut modéliser le tractus vocal dans /a/ avec un tube étroit pour le pharynx et un tube plus large pour la cavité buccale.

Les formants peuvent être modifiés par quelques manœuvres articulatoires. La première règle est que tous les formants diminuent uniformément au fur et à mesure que la longueur du tube s'allonge. Ces modifications de longueur peuvent être obtenues par un abaissement du larynx ainsi que par une projection ou au contraire une rétraction des lèvres. Ces modifications entraînent donc un glissement global des fréquences sans modification de l'intervalle entre les formants, donc sans modification de l'identification de la voyelle. La deuxième façon de modifier les voyelles est d'arrondir les lèvres, ce qui correspond schématiquement à obturer légèrement la bouche comme font parfois certains instrumentistes à vent. Cette « couverture » aboutit à l'équivalent acoustique de l'augmentation de la longueur du tube et donc à un glissement des fréquences vers le grave. La combinaison entre la hauteur du larynx et la position et la forme des lèvres permet effectivement d'éclaircir ou d'assombrir le timbre de la voix. L'abaissement de la mâchoire peut également être utilisé. En particulier F1 peut être abaissé par un abaissement de la mâchoire, ce qui est utilisé par certaines sopranos pour amener leur F1 au contact de F0 et ainsi augmenter leur puissance acoustique. Par ailleurs, la contraction de la bouche abaisse F1 et élève F2, ce qui donne une voyelle au spectre plus large avec un étalement de la

puissance acoustique comme dans /i/. Inversement la contraction du pharynx élève F1 et diminue F2 ce qui rend la voyelle plus compacte comme /a/.

7. Le contrôle nerveux de la production vocale

La production vocale est le résultat d'une coordination neuromotrice des muscles de tous les organes impliqués dans la phonation depuis les muscles de la posture et de la respiration jusqu'aux muscles du larynx et de l'appareil articulatoire pharyngo-bucco-labial.

L'innervation sensitive du larynx : Elle est assurée surtout par le nerf laryngé supérieur qui reçoit des fibres sensitives issues du vestibule laryngé et de la margelle. Ces fibres rejoignent le nerf vague au niveau du ganglion inférieur du vague. Elle est assurée également, en ce qui concerne l'innervation de la corde vocale et de la région sous-glottique, par des fibres qui rejoignent le nerf laryngé inférieur. Il existe des récepteurs muqueux sensibles au contact (mécanorécepteurs) et qui entraînent, lorsqu'ils sont stimulés, le réflexe de toux. Ceux-ci sont situés principalement au niveau de l'étage vestibulaire. Il existe également des mécanorécepteurs articulaires et intra-musculaires dans les muscles intrinsèques et extrinsèques de plusieurs types (corpusculaires, fuseaux neuro-musculaires, spiralés) qui renseignent les centres nerveux sur les phénomènes proprioceptifs de tension et d'étirement présents au niveau des cordes. Ces récepteurs, plus particulièrement situés au niveau des cordes, transmettent des messages qui sont acheminés par le nerf récurrent.

Les fibres pénètrent dans le bulbe avec le nerf vague et se dirigent vers le noyau du faisceau solitaire où certaines fibres issues du larynx se terminent. Les autres fibres continuent leur chemin à travers le tronc cérébral vers le noyau rond. A partir de celui-ci, des deutoneuronies rejoignent le lemniscus médian (ruban de Reil) et se projettent dans le noyau ventral postérieur du thalamus. A partir du noyau du faisceau solitaire, il existe des projections thalamiques et de nombreuses projections dans le noyau ambigu.

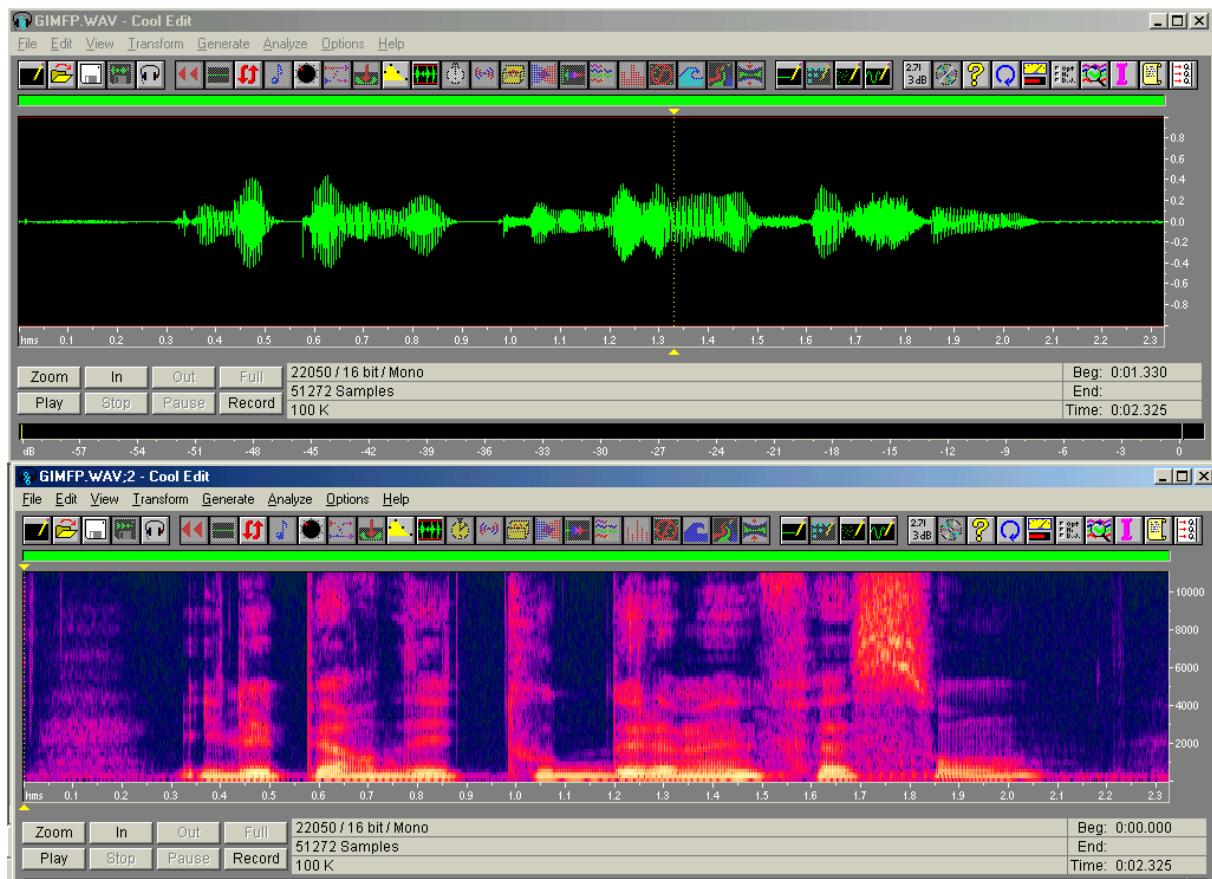
Les centres nerveux : La zone cérébrale motrice du pharyngolarynx est au niveau de la partie basse de la circonvolution frontale ascendante (ou gyrus précentral) des deux hémisphères. Il existe une autre zone, mise en évidence expérimentalement, qui se trouve à la partie postérieure de la première circonvolution frontale, empiétant sur la face interne de l'hémisphère. Lors de la stimulation de tout ou partie de ces zones, on observe une réponse laryngée globale avec vocalisation, inhibition du muscle crico-aryténoïdien postérieur et activation de un ou plusieurs muscles adducteurs de façon bilatérale (des lésions cérébrales à ce niveau n'entraînent pas d'immobilité unilatérale).

De très nombreuses connexions cérébrales existent en particulier avec les centres impliqués dans le langage. On citera en particulier les voies associatives entre les régions motrices pharyngolaryngées et les aires auditives corticales et sous-corticales.

Contrôles réflexes : Des ajustements articulatoires surviennent dans le processus de phonation à deux niveaux : lors de l'ajustement pré-phonatoire et lors de l'émission sonore. L'ajustement préphonatoire est indépendant du contrôle audio-phonatoire. Ainsi en voix chantée, il est possible d'émettre un son qui soit d'emblée à la hauteur et à l'intensité voulue. Ce réglage pré-phonatoire d'origine corticale dépend notamment des informations fournies par les mécanorécepteurs laryngés qui donnent aux centres des informations sur l'état de tension et la position des différents muscles et des différentes articulations. En cours de phonation, ces informations permettent les ajustements instantanés nécessités par le maintien d'une configuration glottique donnée. D'autres arcs réflexes d'origine abdominale thoracique, cervicale, linguale, etc contribuent probablement aux ajustements permanents agissant par feed-back sur le larynx au cours de la phonation.

La rapidité du flux de la parole et la précision des points d'articulation nécessitent des adaptations à la fois très rapides et très fines, ces points d'articulation pouvant être très proches les uns des autres. Ainsi de la degré de constriction du conduit vocal est déterminé par l'espace entre différentes structures telles que la langue et le palais. Ce degré de constriction définit à son tour le type d'écoulement de l'air (laminaire ou turbulent). De très faibles variations de l'ordre du millimètre de la distance entre langue et palais permettent donc de produire des sons aussi différents qu'une voyelle, une consonne voisée ou encore un consonne occlusive. Les adaptations sont assurées grâce à des arcs réflexes rapides reposant sur la sensibilité tactile et proprioceptive de la langue, du palais et des lèvres.

Contrôle audio-phonatoire : Le réglage de la voix est sous le contrôle des centres auditifs comme le montre la voix perturbée et non modulée des sourds profonds. Ce contrôle audio-phonatoire semble comporter un processus automatique involontaire et un processus volontaire. Il existe probablement des commandes induites volontairement par les voies cortico-bulbaires en réponse aux informations acoustiques parvenues jusqu'au cortex auditif de même qu'un ensemble de réflexes acoustico-laryngés. Ce contrôle s'exerce cependant en synergie avec le contrôle proprioceptif qui permet l'anticipation pré-phonatoire (c'est ce qui permet à un chanteur d'émettre d'emblée une note à la hauteur voulue sans attendre le contrôle auditif). Cet ensemble proprioceptif explique la voix relativement peu altérée des sourds récents.



La perception de la dysphonie

La perception de la dysphonie

« *La voix vous fût connue (et chère ?)*

Mais à présent, elle est voilée

Comme une veuve désolée

Pourtant, comme elle est encore fière »

Verlaine, 1878

La méthode d'évaluation de la qualité de la voix la plus ancienne et probablement la plus répandue est l'évaluation perceptive. Elle est définie par Wirz en 1995 [in De Bodt, 1996] comme « *les procédures qui font appel à l'habileté du clinicien sans dépendre d'une évaluation standardisée ou de mesures instrumentales* ». Son principe est d'évaluer, à l'oreille, la quantité de dysphonie. Elle est communément utilisée en clinique pour contrôler les résultats d'une thérapeutique, malgré son caractère subjectif.

Les jugements des auditeurs sont parallèlement considérés comme le standard contre lequel les mesures instrumentales de la voix sont validées [Askenfelt, 1986 ; Bielamowicz, 1996 ; Giovanni, 1995, 1996, 1999 ; Hillenbrand, 1994, 1996 ; Martin, 1995 ; Wolfe, 1991, 1995 ; Wuyts, 1996, 1997, 1999, 2000 ; Yu, 2000, 2001, 2002]. Ainsi, la fiabilité de tels jugements est un point central dans l'étude de la qualité de la voix et des dysphonies. L'analyse perceptive nécessite une standardisation définissant la méthode de mesure utilisée, le choix des membres du jury, le matériau phonétique, etc.

Pour Kreiman [Kreiman, 93], « *la recherche dans ce domaine s'est poursuivie sans référence à un modèle de perception de la voix. Par conséquent, les auteurs n'ont eu aucune base théorique particulière pour sélectionner l'un ou l'autre des nombreux protocoles et analyses disponibles pour leurs études* ». En effet, les méthodologies varient largement d'une étude à l'autre, et malgré la définition de plusieurs protocoles standardisés d'évaluation perceptive, de nombreux problèmes freinent encore leur développement : la définition sémantique des paramètres qualitatifs de la voix, la mise au point de méthodes d'évaluation standardisées, la question du référent auditif, la question de l'expérience du jury, le matériau phonétique...

1. Principes et méthodes de l'évaluation perceptive

1.1. Les dimensions qualitatives de la perception de la voix

Dans la pratique clinique, un ensemble de termes est utilisé pour décrire la qualité de la voix. Ces termes sont fondés sur l'impression auditive et les jugements perceptifs [Hammarberg, 1980]. Pour Childers [Childers, 1991], « *il n'existe pas de définition généralement acceptée de la qualité de la voix, et le terme a été utilisé dans différents contextes : le phonéticien utilise le terme de qualité dans un contexte de différences articulatoires ; un chanteur pourrait référer à la qualité des registres vocaux, qui sont liés à la vibration des cordes vocales ; et la qualité peut être utilisée pour décrire certains types de voix comme le souffle, la raucité, ou le forçage*une voix contenant clairement des composantes de bruit, et qui peut être considérée comme rauque et soufflée (c'est à dire éléments de bruit dans la source associés à des bruit de friction) ; sa hauteur perçue tend à varier substantiellement ; des termes courants décrivant cette qualité de voix sont "bruitée", "dure", "grasse". »

L'évaluation perceptive n'est pas l'objet d'un consensus dans la définition des paramètres décrivant la qualité de la voix. Par exemple, certains auteurs semblent employer le terme "éraillement" ("hoarseness") en tant que "raucité", alors que d'autres emploient le terme "éraillement" dans un sens très étendu qui signifie en fait "voix pathologique", "voix dysphonique", au sens large. De très nombreux termes sont ainsi employés pour décrire la qualité de la voix, avec parfois des nuances difficiles à discerner, comme par exemple "éraillement", "raucité", "enrouement", ou "voix rocallieuse".

Lundälv & Nygren (1974) [in Hammarberg, 1980] ont montré que les orthophonistes et les étudiants en orthophonie utilisaient 88 termes différents pour la description de la qualité de la voix et de la hauteur pour un corpus de 10 patients. Devant une telle masse de termes qualitatifs disponibles pour

décrire la voix et un tel flou sémantique entre les différentes équipes, un besoin de vocabulaire commun à tous s'est fait ressentir.

Hammarberg [Hammarberg, 1980] a présenté un travail sur la définition et la description des caractères régissant la qualité de la voix ainsi que leurs techniques d'évaluation. Le but de cette étude était d'analyser les corrélations entre les variables décrivant la voix afin de structurer la terminologie et de donner un ensemble précis de références pour la perception de la voix grâce à la méthode statistique de l'analyse factorielle. Elle a proposé un ensemble de 28 critères de jugement de la voix (tableau V.1.). Ces 28 paramètres avaient été sélectionnés parmi les 50 termes les plus fréquemment utilisés par les cliniciens de l'Hôpital Universitaire de Huddinge (Suède) pour décrire un corpus de 115 voix. 23 termes décrivaient la qualité de la voix, 2 la hauteur, et 3 le registre.

<i>Terme professionnel Suédois</i>	<i>Equivalent Anglais</i>	<i>Equivalent Français</i>
1. läckande	breathy	soufflée
2. skrovlig	rough, raucous	rauque
3. diplofoni	diplophonia	diplophonique
4. knarr	creaky, vocal fry	voix craquée, grinçante
5. halsig	guttural, throaty	gutturale
6. återhållen	repressed / restrained	retenue
7. grov	coarse	rude
8. pressad	strained	forcée
9. strypta frasslut	strangle phrase ending	fin de rhète étranglée
10. klangfattning	lack of timbre	pauvreté du timbre
11. afoniska inslag / afoni	moments of aphonia / aphonia	moments d'aphonie / aphonie
12. väsande	wheezing, hissing	voix faible, asthénique
13. bitonalitet	bitonality	bitonalité
14. hyperkinetisk	hyperfunctional	hyperfonctionnelle
15. monoton	monotonous	monotone
16. beslöjad	husky	enrouée
17. klangförändring utan brott	quality changes without voice, breaks	dévoisement, cassures
18. fladder	flutter	tremblement, chevrottement
19. skrap	grating	érailement
20. registerbrott / registerbrottstendens	voice breaks / tendency towards	cassures dans la voix / tendance aux cassures
21. hårliga ansatser	voice breaks	
22. instabil klang	hard glottal attacks	attaques glottiques dures
23. sträv	unstable quality	instabilité
24. bröstregister	harsh	enrouement, rugueux
25. mellan register	chest register	registre lourd
26. falsettregister	middle register	registre moyen
27. rösläge	head (falsetto) register	registre de tête, falsetto
28. instabil rösläge	pitch	hauteur
	unstable pitch	instabilité de la fréquence

Tableau V.1. : Paramètres décrivant la voix en Suédois et leur équivalent en Anglais [Hammarberg, 1980] et en Français.

20 voix dysphoniques ont été jugées par un jury de 14 auditeurs d'après ces 28 paramètres. Une analyse en composantes principales a permis d'extraire de cette masse de termes 5 facteurs bipolaires expliquant 85.3% de la variance (tableau V.2.).

Factor	1. Unstable-Steady	2. Breathy-Overtight	3. Hyper-Hypofunctional	4. Coarse-Light	5. Head-Chest register
Variance Explained	30.0%	27.3%	13.5%	10.1%	4.4%

Tableau V.2. : Proportion de la variance expliquée par les 5 facteurs [Hammarberg, 1980]

Ainsi, pour Hammarberg, 85.3% de l'impression auditive de dysphonie pourrait être contenue dans ces 5 facteurs bipolaires résumant à eux seuls les 28 termes proposés : stable / instable, soufflé / serré, hyper- / hypo-fonctionnel, rude / léger, et registre de tête / de poitrine.

Ces tentatives de définition de termes qualifiant la qualité de la voix ont constitué une première étape dans la définition de protocoles standardisés d'analyse perceptive de la voix.

1.2. Les échelles quantitatives d'évaluation

La définition d'un protocole d'analyse perceptive de la qualité de la voix implique plusieurs choix :

- un choix qualitatif (paramètres perceptifs que l'on veut étudier)
- et un choix quantitatif (échelle de mesure)

Les différents paramètres perceptifs sont contenus dans des proportions différentes d'une voix à l'autre. Il s'agit de déterminer quelles sont ces proportions. Plusieurs échelles quantitatives ont été décrites :

- *les échelles bipolaires* : elles fonctionnent sur le mode oui / non, c'est à dire présence / absence du paramètre étudié. Hammarberg utilise ce genre d'échelle pour opposer deux dimensions contraires (stable / instable ; hyper / hypofonctionnel...) [Hammarberg, 1980]. L'échelle bipolaire induit donc l'idée que la présence d'un paramètre exclut la présence de l'autre (figure V.1.).

Vous diriez que cette voix est plutôt : (cocher une seule case par ligne)

<input type="checkbox"/> Stable	ou	<input checked="" type="checkbox"/> Instable
<input checked="" type="checkbox"/> Soufflée	ou	<input type="checkbox"/> Serrée
<input type="checkbox"/> Hyper-fonctionnelle	ou	<input checked="" type="checkbox"/> Hypo-fonctionnelle
<input type="checkbox"/> Rude	ou	<input checked="" type="checkbox"/> Légère
<input checked="" type="checkbox"/> Registre de tête	ou	<input type="checkbox"/> Registre de poitrine

Figure V.1. : Echelle bipolaire sur le modèle de Hammarberg.

- *les échelles de classe* : plusieurs niveaux sont prédéfinis afin de quantifier la présence d'un ou de plusieurs paramètres séparément. Le plus souvent, 4 sont proposés, de 0 à 3, qui peuvent être interprétés de la façon suivante : 0 = voix normale, 1 = dysphonie légère, 2 = dysphonie moyenne, 3 = dysphonie sévère. Néanmoins, on retrouve quelquefois des échelles à 7 niveaux, de 1 à 7, ce qui revient simplement à rajouter des « demi » (figure V.2.). En général, le nombre de niveaux utilisés dépend d'études préalables sur les capacités de discernement du jury. Ces échelles sont couramment appelées “échelles à points d’intervalles équidistants.” (EAI = Equal Appearing Interval).

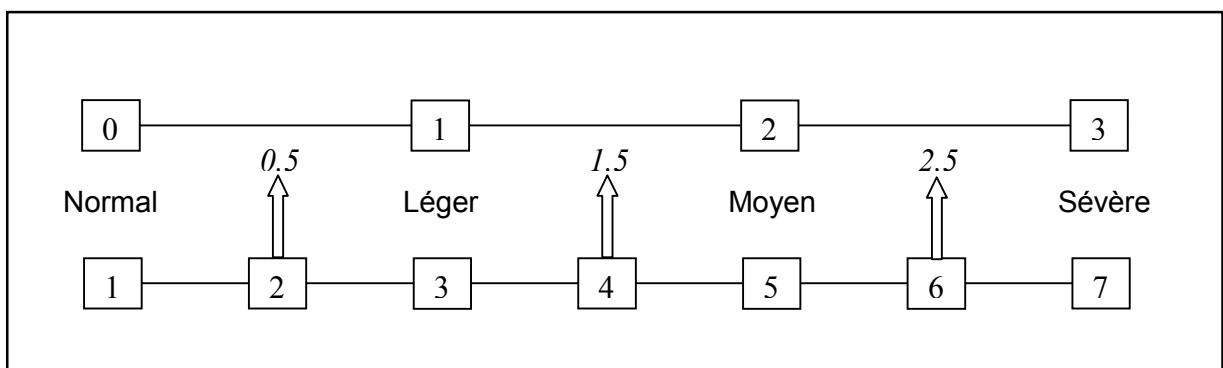


Figure V.2. : Echelles de classe à 4 (en haut) ou 7 points (en bas). Dans l'échelle à 7 points, les niveaux 2, 4 et 6 correspondent à des $\frac{1}{2}$ points, permettant de quantifier une dysphonie qui serait « plus que légère, mais pas tout à fait moyenne » : classe 4 (correspondant à 1,5).

- *les échelles ouvertes* : ce sont des échelles numériques où n’importe quelle valeur peut être attribuée pour quantifier la présence d’un paramètre, de 0 (exclu) à l’infini. Chaque voix est alors évaluée par rapport à la précédente. L’inconvénient de ces échelle est que aucune consigne d’ordre de grandeur n’étant donnée aux auditeurs, certains pourront évaluer les voix entre 50 et 200 (figure V.3.), d’autres entre 3 et 26, par exemple, etc. Ces échelles nécessitent donc une standardisation des données afin de ramener tous les jugements sur un même ordre de valeurs. Ces échelles sont couramment appelées “estimation directe de magnitude” (DME = Direct Magnitude Estimation). Elles permettent une discrimination plus fine que les EAI pour les différents niveaux de dysphonie.

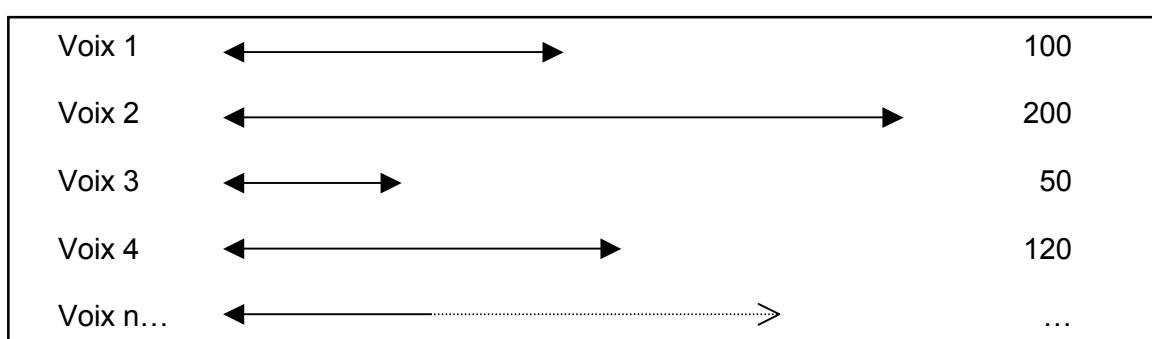
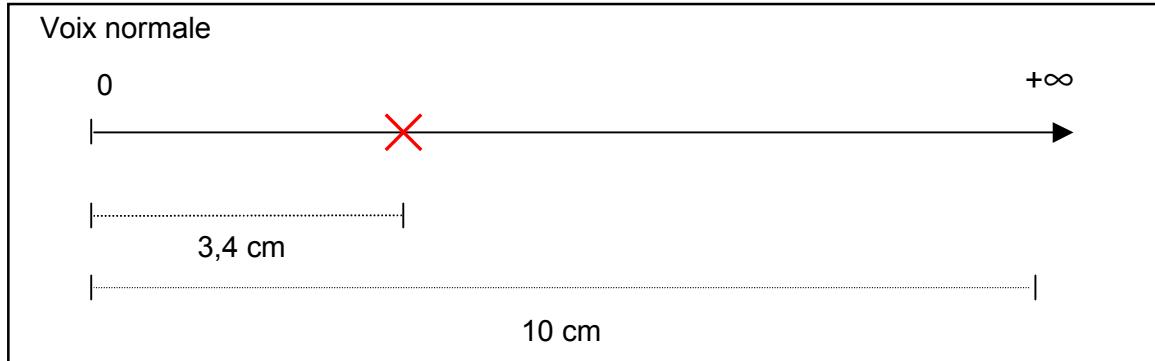


Figure V.3. : Echelles ouvertes. N’importe quelle valeur peut être attribuée de 0 (exclu) à l’infini.

- *les échelles visuelles analogiques* : permettent d'appréhender la quantité de façon visuelle, spatiale. Ce sont des échelles semi-ouvertes : l'auditeur porte son évaluation sur un axe de 10 cm (le plus souvent) par une marque (figure V.4.). Le chiffre attribué à la voix correspond à la distance en cm entre le 0 et la croix. Cette échelle combine les avantages de l'échelle ouverte (toutes les valeurs intermédiaires sont disponibles) et d'une échelle de classe (les valeurs sont standardisées puisque l'échelle est de dimension prédéfinie, ainsi que l'unité de mesure). Il s'agit donc d'une échelle semi-ouverte.



Le choix de l'échelle de mesure dépend de ce que l'on veut mesurer. Les échelles de classe permettent de ranger les voix dans des cases prédéfinies : c'est une classification des voix en grands groupes. Les échelles ouvertes et semi-ouvertes permettent au contraire une discrimination plus fine et plus précise des voix mais sont aussi davantage soumises à la variabilité.

Dans la pratique clinique, les échelles de classe sont le plus souvent utilisées. Elles sont simples et rapides à mettre en place. Les cliniciens d'une même équipe, formés à une échelle de classe, peuvent échanger des données comparables en ce qui concerne leurs patients. C'est sans doute la raison pour laquelle les différentes méthodes standardisées d'analyse perceptive utilisent ce genre d'échelles à 4, 5, ou 7 points d'intervalles équidistants.

1.3. Les différentes méthodes d'évaluation perceptive : premiers essais de standardisation

Dans le but d'harmoniser les résultats de l'analyse perceptive des dysphonies, de nombreuses tentatives de standardisation des méthodes d'évaluation ont été rapportées. Ces classifications ont toutes fait l'objet d'études spécifiques et proposent d'évaluer la qualité de la voix d'après différents critères qualitatifs et quantitatifs.

Pour De Bodt [De Bodt, 1996], s'il existe un consensus sur la nécessité d'une évaluation perceptive pour le suivi de patients dysphoniques, il n'en va pas de même pour la méthode ou l'échelle utilisée. Dans un travail publié en 1996, il donne le profil des acquis historiques et présente l'étude des méthodes de jugements perceptifs contemporaines. Il propose un tableau comparatif de quatre méthodes d'évaluation perceptive parmi les plus fréquemment utilisées :

	BVP	Hammarberg Scheme	VPAS	GRBAS
Auteur	<i>Wilson</i>	<i>Hammarberg</i>	<i>Laver</i>	<i>Hirano</i>
Niveaux Etudiés	<i>laryngé</i> <i>supra-laryngé</i> <i>comportement vocal</i>	<i>laryngé</i>	<i>laryngé</i> <i>supra-laryngé</i>	<i>laryngé</i>
Echelle Quantitative	<i>5 niveaux</i>	<i>5 niveaux</i>	<i>6 niveaux</i>	<i>4 niveaux</i>
Echantillons	<i>parole spontanée</i> <i>comptage</i> <i>voyelle tenue</i>	<i>lecture</i>	<i>lecture</i> <i>parole spontanée</i>	<i>parole spontanée</i>
Durée de Passation	<i>non-spécifiée</i>	<i>10-15 mn</i>	<i>10-15 mn</i>	<i>< 5 mn</i>
Entraînement	<i>non-spécifié</i>	<i>non-spécifié</i>	<i>cassettes audio</i>	<i>cassettes audio</i>
Age de passation	<i>enfants</i>	<i>tous âge</i>	<i>tous âges</i>	<i>tous âges</i>

Tableau V.3. : Comparaison des 4 échelles perceptives [De Bodt, 1996]. Les "Niveaux Etudiés" correspondent aux lieux de dysfonctionnement physiologique et "l'Entraînement" correspond au mode de formation prévu par l'auteur pour le jury d'écoute.

- *The Buffalo Voice Profile Systeme* (BVP) de Wilson (1987) [in De Bodt, 1996] est une méthode d'analyse perceptive à 12 paramètres (le timbre laryngé, la hauteur, l'intensité, la résonance nasale, la résonance orale, le volume respiratoire, le tonus musculaire, le forçage, le débit, l'anxiété perçue dans la voix, l'intelligibilité, le grade global de la dysphonie). Ces 12 paramètres doivent être évalués par le jury selon cinq niveaux (0 = normal; 1 = peu sévère; 2 = moyen; 3 = sévère; 4 = très sévère). Les échantillons de voix comprennent de la voix parlée, de la lecture, une épreuve de comptage, et des voyelles tenues. Pour De Bodt, le BVP fournit trois éléments fondamentaux pour l'évaluation de la voix qui sont : la sévérité du trouble de la voix, l'impact du trouble sur la communication, l'indication pour l'intervention. Les fiabilités intra- et inter-juges ne sont pas garanties. Le BVP est largement utilisé aux Etats-Unis [Hammarberg, 1980].
- Dans l'optique de tester la fiabilité des mesures perceptives, Hammarberg [Hammarberg, 1980] a analysé la voix à partir d'un texte standardisé de 92 mots, lu par 17 patients dysphoniques présentant des pathologies variées. Ces voix ont été présentées à un jury d'expert qui devait les juger sur une échelle de 5 points selon 28 termes (voir tableaux V.1. & 2.) fréquemment utilisés pour décrire les voix. Une analyse en composantes principales a permis d'extraire 5 facteurs (stable / instable; soufflé / serré; hyper- / hypofonctionnel; sombre / lumineux; registre lourd / registre léger) qui semblent les plus pertinents et qui semblent recouvrir plus particulièrement les notions de « stabilité » et de « souffle ».

- Présenté plus tard, *The Hammarberg Scheme* [Hammarberg, 1986] est une méthode d'évaluation perceptive à 12 paramètres.

Paramètre de qualité vocale	Définition
Aphonique / Aphonique intermittent	La voix manque constamment ou quelquefois de phonation ; il y a des moments de chuchotement ou de perte de voix
Voix soufflée	Fuite audible d'air à travers la glotte à cause d'une fermeture glottique insuffisante
Hyperfonctionnel / Tendu	Voix tendue, comme si les cordes vocales étaient comprimées pendant la phonation
Hypofonctionnel / Lâche	Contraire à hyperfonctionnel, tension insuffisante des cordes vocales, résultat : une voix faible et lâche
Voix grinçante (fry)	Série rapide de coups ; comme quand on ouvre une vieille porte rouillée ; vibration périodique de basse fréquence
Voix rauque	Bruit non-périodique avec basse fréquence ; probablement en rapport avec un certain type de vibrations irrégulières des cordes vocales
Voix éraillée	Bruit non-périodique avec haute fréquence ; probablement en rapport avec un certain type de vibrations irrégulières des cordes vocales
Ruptures vocales	Ruptures intermittentes de fréquence
Diplophonique / Bitonalité	Deux fréquences différentes perçues simultanément
Registre	Mécanismes de phonation : lourd (poitrine) ; léger (tête)
Hauteur	Corrélation auditive principale de la fréquence fondamentale
Intensité	Corrélation auditive principale du niveau sonore

Tableau V.4. : Définition des paramètres vocaux utilisés par Hammarberg [Hammarberg, 1986]

Ces 12 paramètres sont évalués selon 5 niveaux (0 = normal; 4 = très sévère). Les échantillons de voix sont composés de la lecture d'un texte d'une durée d'environ 40 secondes. *The Hammarberg Scheme* est une échelle spécifiquement adaptée à la langue Suédoise et malgré les efforts de standardisation, elle n'a jamais été adoptée par les autres pays.

- *The Vocal Profile Analysis Scheme* (VPAS) de Laver (1980) [in De Bodt, 1996] est une échelle à 17 paramètres divisés en 3 catégories :

- la qualité de la voix au sens strict (positionnement des lèvres, de la mandibule, de la langue, le fonctionnement vélopharyngé et pharyngé, le positionnement du larynx, le type de phonation, les tensions musculaires supra-laryngées et laryngées).
- des caractéristiques prosodiques (hauteur et intensité)
- les caractéristiques temporelles (le rythme respiratoire, la continuité, le débit, la cadence)

Ces 17 paramètres sont à évaluer selon 6 niveaux (1 = présence d'un paramètre à peine audible ; 6 = effet auditif le plus mauvais dont est capable un appareil vocal). Les échantillons de voix sont composés de voix de lecture et de discours spontané. Laver s'intéresse plus spécialement à la voix non-pathologique. Il cherche à définir la qualité de la voix au sens large. Il est précisé que 12 heures de formation sur cassettes audio sont nécessaires pour atteindre une fidélité de jugement de 65-75%. Le VPAS est surtout utilisé en Grande-Bretagne et en Australie [Hammarberg, 1986].

- L'échelle GRBAS Hirano (*GRBAS Scale*) [De Bodt, 1997 ; Hirano, 1981], décrite en 1969, et inspirée de Isshiki (1966) [Isshiki, 1969], est une échelle à 5 paramètres :

INITIALE	TERME	SIGNIFICATION	DÉFINITION
G	<i>Grade</i>	<i>Grade global de dysphonie</i>	<i>Impression globale du degré d'anormalité de la voix</i>
R	<i>Rough</i>	<i>Raucité</i>	<i>Impression d'irrégularité des vibrations de la corde vocale qui correspond aux fluctuations irrégulières de la Fo et/ou à l'amplitude du son glottique</i>
B	<i>Breathy</i>	<i>Caractère soufflé</i>	<i>Impression d'une fuite d'air assez importante à travers les cordes vocales, relative à des turbulences</i>
A	<i>Asthenic</i>	<i>Asthénie</i>	<i>Manque de puissance de la voix relatif à une intensité faible du son et/ou un manque des harmoniques élevés</i>
S	<i>Strained</i>	<i>Forçage</i>	<i>Impression d'un état hyperfonctionnel de phonation relatif à une Fo anormalement haute (bruit dans les hautes fréquences et/ou richesse en harmoniques dans les hautes fréquences)</i>

Tableau V.5. : Définition des paramètres de l'échelle GRBAS

Ces 5 paramètres sont évalués selon 4 niveaux (0 = voix normale ; 3 = altération sévère). Les échantillons de voix sont composés de voix conversationnelle. L'échelle GRBAS est probablement la plus compacte des méthodes d'évaluation perceptive, elle a été conçue pour un usage clinique quotidien. Elle présente pour avantage de fournir à l'examineur une cassette audio de formation, afin de s'assurer d'un terrain sémantiquement commun. Malheureusement, il est à noter que cette cassette n'est disponible qu'en version originale (japonaise). Cette classification est actuellement la plus utilisée. [De Bodt, 1996 ; De Krom, 1994 ; Dejonckere, 1993 ; Giovanni, 1995, 1996, 1999 ; Wolfe, 1995 ; Wuyts, 1996, 1999 ; Yu, 2001]

- En 1993, un travail visant à compléter l'échelle GRBAS en adaptant un “profil vocal” avec des évaluations complémentaires a été présenté [Dejonckere, 1993] pour constituer un bilan vocal “complet”. Il comprend, en plus des évaluations classique du GRBAS, des mesures instrumentales et physiologiques : le temps maximum de phonation (TMP), l'étendue vocale, la puissance vocale maximale (sur un /a/ tenu), une épreuve de comptage maximal de 1 à 25, et un test d'intelligibilité. L'échelle d'évaluation peut s'utiliser de différentes façons : graphique (en réunissant par des lignes les divers paramètres), descriptive (à la manière de l'échelle GRBAS), statistique (moyenne, écart-type, etc...). Toutefois, les paramètres présentés et étudiés ressemblent davantage à une présentation des moyens d'évaluer les voix qu'à un choix réfléchi d'un protocole utilisable en clinique, avec en particulier le problème du “poids” relatif de chaque paramètre qui n'est pas évoqué.

De la même façon que le choix de l'échelle quantitative dépend de ce que l'on veut mesurer, l'échelle GRBAS est souvent modifiée en fonction des objectifs de l'étude présentée. On retrouve ainsi fréquemment les seuls paramètres GRB sur une échelle de classe ou visuelle analogique [Anders, 1988 ; Martin, 1996 ; De Bodt, 1997 ; Millet, 1998 ; Kreiman, 1998 ; Chan, 2002 ; Yamaguchi, 2003]. En effet, on retiendra que les critères de grade global (G), raucité (R) et souffle (B) sont évalués de façon plus sûre et sont moins soumis à la variabilité du jury que l'asthénie (A) et le forçage (S) [Dejonckere, 1993 ; Millet, 1998].

Enfin, Kreiman [Kreiman, 1993], dans une revue de la littérature de 57 articles de 1951 à 1990 a montré qu'aucune des nombreuses échelles, procédures et statistiques n'a démontré sa supériorité pour mesurer la qualité de la voix. Sur toutes les études analysées par Kreiman, les niveaux de reproductibilité et de fiabilité variaient de 18% à 100% selon les auteurs avec des protocoles et des méthodes d'analyse statistiques si variées que toute conclusion est impossible. En réalité, un cadre théorique est nécessaire pour définir un modèle d'étude et servir de base à la recherche. Toutes les étapes de l'évaluation perceptive doivent être contrôlées pour permettre des comparaisons valables entre les études publiées et repérer les biais créés par des incertitudes méthodologiques, en particulier au niveau de l'expérience et de la sensibilité de l'auditeur, ainsi que le choix du matériau phonétique et de l'échelle utilisée.

2. Le jury d'écoute

2.1. Le choix du jury

Le choix des auditeurs composant un jury d'écoute est fait dans le but d'obtenir un jugement représentatif, c'est à dire fiable et constant, de la qualité de chaque voix présentée. Dans cette optique, le choix des membres d'un jury doit prendre en compte différentes notions, comme l'expérience des auditeurs dans l'évaluation de la voix, leurs stratégies perceptives, et l'homogénéité du groupe obtenu. Ces différents critères de recrutement constituent autant de problèmes soulevés par plusieurs auteurs [Anders, 1988 ; Askenfelt, 1986 ; De Bodt, 1997 ; Gerratt, 1993 ; Kreiman, 1992 ; Millet, 1998 ; Chan, 2002 ; Goldstone, 1998 ; Oates, 1998 ; Yamaguchi, 2003 ; Wolfe 2000]. La plupart du temps, la qualité d'un jury est évaluée en termes de fiabilité. De manière générale, les auteurs considèrent qu'un jury est fiable si les jugements qu'il porte sur les voix sont reproductibles : un auditeur donné est supposé évaluer une même voix de manière identique à chaque fois qu'il l'entend, c'est ce que l'on appelle l'agrément intra-auditeur, ou la constance du jury. De même, une même voix est sensée produire la même impression perceptive sur les différents auditeurs, c'est l'agrément inter-auditeur, ou la cohérence du jury. Kreiman [Kreiman, 1993] a publié une revue de la littérature sur le thème de l'évaluation perceptive. Elle a ainsi observé que la fiabilité inter- et intra-auditeurs varie grandement d'une étude à l'autre.

En première analyse, on pourrait penser qu'une façon d'obtenir un jury plus fiable, est de multiplier l'effectif des auditeurs. En augmentant le nombre des auditeurs, la tendance générale se fortifie, les différences s'estompent, et on obtient statistiquement davantage de fiabilité, malgré la variabilité individuelle des sujets qui reste identique. Kreiman [Kreiman, 1993] observe que le nombre de

membres du jury dans les publications étudiées varie de 1 à 461. La distribution de l'effectif des groupes est globalement trimodale avec des groupes de 5, 8 et 11 membres. Il n'existe aucun argument pour justifier le choix d'un nombre ou d'un autre.

Anders [Anders, 1988] admet qu'une diversité de recul professionnel individuel, d'expérience dans l'écoute de voix dysphoniques, peut induire de larges variations dans le jugement des voix. Pour Kreiman [Kreiman, 1993], les auditeurs peuvent être classés en trois catégories :

- aucun entraînement (mais les auditeurs candides tendent à fonder leurs impressions sur la hauteur et la puissance davantage que sur les notions de souffle et de raucité [Hammarberg, 1980; Kreiman, 1992])
- un entraînement succinct qui consiste en la définition des termes qui constituent les différents paramètres de l'échelle, des exemples et/ou un nombre limité d'essais d'évaluation d'échantillons vocaux (cassettes audio de formation à certaines échelles d'évaluation dont le GRBAS)
- un entraînement intensif, c'est à dire que les membres du jury sont amenés à évaluer des voix dans la pratique quotidienne de leur profession et/ou ont suivi une formation particulière (Askenfelt [Askenfelt, 1986] utilise un jury constitué de 6 orthophonistes professionnels ayant participé à un entraînement systématique et intensif à l'évaluation de la voix durant 4 ans).

A partir de ce classement, Kreiman montre que les auditeurs qui ont le plus d'expérience sont plus fiables que ceux qui en ont moins. Pour elle, les auditeurs expérimentés sont plus constants. Elle ne précise pas si les “experts” sont aussi plus précis dans leur jugement que les “candides”. De même, Anders [Anders, 1988] conclut qu'il existe une influence de l'expérience des auditeurs dans la fiabilité d'un jury.

A l'inverse, De Bodt [De Bodt, 1997] a démontré que le niveau d'expérience n'influence en rien la fiabilité test-retest du jury. Son protocole d'étude comprenait 9 voix dysphoniques, évaluées au cours de 2 sessions séparées de 14 jours par un jury de 23 auditeurs : 12 auditeurs expérimentés (médecins ORL et orthophonistes) confrontés à la pathologie vocale depuis plus de 3 ans) et 11 juges inexpérimentés. L'échelle utilisée était la GRBAS Scale (5 paramètres et 4 niveaux). De Bodt calculait la fiabilité test-retest grâce au coefficient de Kappa-Cohen. Les résultats mettaient en évidence une faible corrélation entre le niveau d'expérience et la fiabilité intra- et inter-individuelle du jury (tableaux V.6. et 7.).

	I	E	ENT	S/LP	All
G	0.50	0.70	0.62	0.58	0.60
R	0.29	0.40	0.27	0.43	0.35
B	0.29	0.45	0.30	0.46	0.38
A	0.51	0.28	0.42	0.36	0.39
S	0.30	0.36	0.32	0.35	0.34

Tableau V.6. : [De Bodt, 1997] fiabilité intra-individuelle. Valeurs du coefficient Kappa entre la session 1 et la session 2 pour les auditeurs inexpérimentés (I), expérimentés (E), ORL (ENT), orthophonistes (S/LP), et tous ensemble (All) ; sur l'échelle GRBAS.

	I	E	ENT	S/LP	All
G	0.39	0.49	0.45	0.46	0.44
R	0.16	0.20	0.14	0.25	0.17
B	0.21	0.20	0.19	0.31	0.21
A	0.33	0.27	0.26	0.28	0.27
S	0.11	0.17	0.12	0.12	0.10

Tableau V.7. : [De Bodt, 1997] agrément inter-auditeurs en fonction de leur expérience. Valeurs du coefficient Kappa entre les auditeurs d'un même groupe.

Ces résultats montrent que la fiabilité test-retest de l'échelle GRBAS est modérée. Il n'y a pas de différences significatives entre les 4 classes d'auditeurs, même si l'on peut observer une tendance selon laquelle les auditeurs expérimentés seraient plus constants et plus fiables. Deux conclusions sont envisageables : soit le niveau d'expérience n'influence pas la fiabilité du jury, soit l'échelle GRBAS ne permet pas d'objectiver cette influence. Selon Kreiman [Kreiman, 1993], pour qu'une évaluation perceptive soit significative, l'échelle de mesure doit être choisie avec soin en fonction de ce que l'on veut mesurer et les auditeurs doivent l'utiliser de manière cohérente. Il est probable que la fiabilité du jury est différente selon le paramètre évalué. Les résultats de De Bodt (tableaux V.6. et 7.) montrent que les résultats sont plus fiables, quelle que soit l'expérience du jury, sur le grade global de dysphonie (paramètre G) que sur les autres paramètres (RBAS).

Enfin, dans sa revue de la littérature, Kreiman [Kreiman, 1993] note que seulement 10,5% des études utilisent un jury constitué d'auditeurs non-expérimentés (contre 50,9% avec des auditeurs expérimentés). La plupart des auteurs considèrent l'expérience du jury comme un facteur essentiel de l'évaluation perceptive.

En réalité, l'étude des stratégies d'écoute des auditeurs permet d'aller plus loin dans la définition du choix des membres du jury et des protocole expérimentaux en proposant notamment des méthodes d'entraînement du jury.

2.2. Les modèles perceptifs

Pour Kreiman, la littérature manque d'un modèle théorique clair de la perception et des jugements. Bien que dans certains cas, des traits caractéristiques aient été décrits pour des études particulières (Bassich & Ludlow, 1986) [in Kreiman, 1992], dans la majorité des cas le choix entre les différents

types d'échelles, la population des auditeurs, la taille des échantillons, les approches statistiques, n'est pas basé sur des buts clairement établis ou des considérations théoriques [Kreiman, 1992].

Les modèles de perception de la qualité de la voix ont longtemps consisté à décrire les paramètres perceptifs qui paraissaient prégnants à priori pour l'auditeur. Ainsi, pour Kempster [Kempster, 1991], il apparaît que la plupart des auditeurs appliquent une stratégie d'écoute semblable en utilisant trois dimensions principales : intensité, fréquence fondamentale et perturbations. L'étude de Murry, en 1977 [in Giovanni, 1995] avait également démontré l'importance de la sensation de hauteur de la voix. Isshiki [Isshiki, 1969] déterminait des classes perceptives en fonction des pathologies et des données physiologiques en combinant divers paramètres. Hammarberg [Hammarberg, 1980] suggère que les caractéristiques perceptives des voix humaines peuvent être interprétées comme des changements brusques de certains traits ou caractéristiques plus ou moins reliés les uns aux autres. Ces observations venaient étayer le principe d'analyse perceptive en répondant à la question : « que voulons-nous mesurer ? », mais étaient peu utiles dans l'élaboration d'un modèle perceptif nécessaire à la définition de protocoles standardisés et reproductibles : « comment le mesurer ? ».

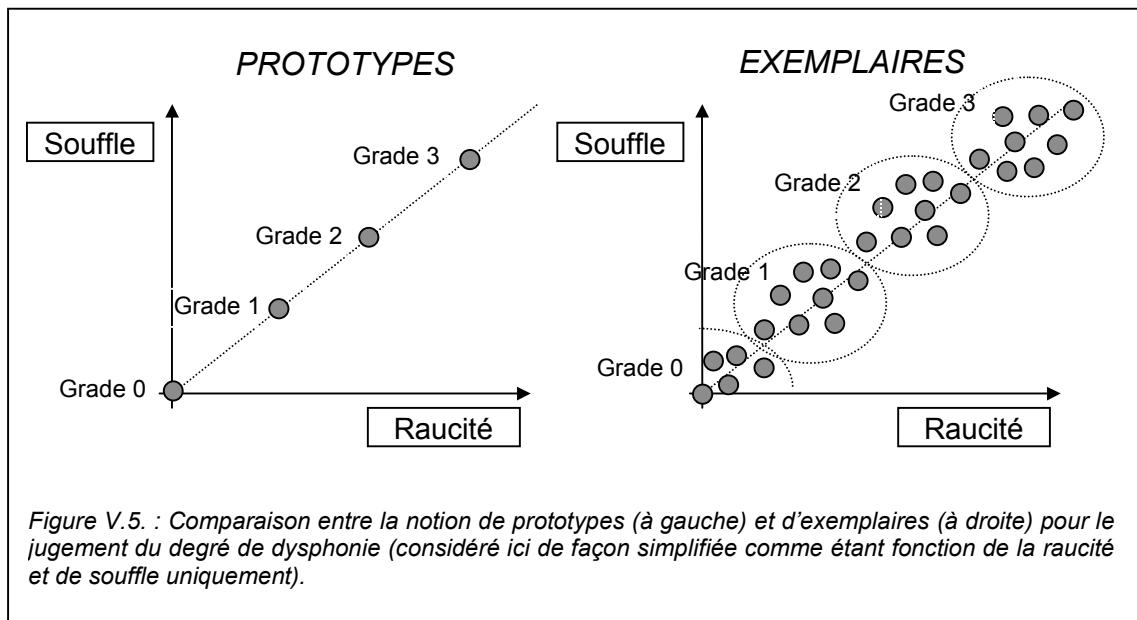
Kreiman [Kreiman, 1993] propose une structure théorique qui attribue la variabilité des jugements à plusieurs sources (incluant l'expérience des auditeurs, les méthodes utilisées pour recueillir les jugements, les interactions entre les auditeurs et ces méthodes, et les erreurs au hasard).

La principale évolution dans les modèles de perception de la dysphonie provient des expérimentations explorant les références auditives utilisées par le jury pour évaluer une voix.

2.3. Les références auditives du jury

Lorsque des auditeurs évaluant la dysphonie, ils comparent intuitivement le stimulus présenté à un référent auditif [Kreiman, 1992]. Il s'agit d'un standard interne (ou prototype) de qualité de la voix contre lequel est comparé la voix que le jury doit évaluer. Plus simplement, c'est "l'idée que l'auditeur se fait" de ce qu'est une voix dysphonique. La distance entre ce référent et la qualité du stimulus proposé permet d'évaluer la dysphonie et d'attribuer un grade. Kreiman donne la définition suivante : «*les standards internes (ou "prototypes") sont développés indépendamment de l'expérience de l'auditeur en matière de qualité de voix* ». Le prototype est donc un spécimen symbolique. Des prototypes ont été proposés dans la recherche sur la parole et sur la voix et sont utilisés pour la reconnaissance de voix non-familier [Kreiman, 1991 ; Papcun, 1988].

A cette notion de prototype s'oppose la notion d'exemplaires : les exemplaires sont des ensembles de possibilités correspondant à une catégorie particulière (le grade), alors qu'un prototype correspond à un seul spécimen symbolique identifié à cette catégorie (figure V.5.). La question est de savoir si c'est l'exemplaire ou le prototype qui fournit une référence interne à l'auditeur pour l'évaluation.



Dans le modèle défini par Kreiman [Kreiman, 1992], les prototypes sont utilisables pour différents niveaux de qualité (par exemple : normal, moyen, extrême), mais « *le nombre précis de niveaux représentés n'est pas connu, et probablement variable d'un auditeur à l'autre* ». Pour Goldstone [Goldstone, 1998], « *l'empreinte est liée aux modèles internes et se forme par une exposition répétée à un stimulus* ». En d'autres termes, plus on a l'habitude d'effectuer une tâche perceptive, plus le modèle interne est solide. Puisque tous les auditeurs (expérimentés ou non) ont une expérience considérable et approximativement identique des voix normales (que l'on est tous habitué à entendre dans la vie quotidienne), tous les auditeurs ont des prototypes similaires et relativement stables pour les voix normales. Le prototype “voix normale” est donc solide, indépendamment du niveau d'expérience. D'ailleurs, les jugements portés sur les voix normales montrent de petites variations intra- et inter-auditeurs [Anders, 1988; Askenfelt, 1986; De Bodt, 1997; Kreiman, 1992, 1993]. Pour Rabinov [Rabinov, 1995], la fiabilité des auditeurs augmente avec la sévérité de la pathologie : il suggère que plus la voix est dysphonique, plus la distance qui la sépare du prototype “voix normale” de l'auditeur est grande, plus elle heurte l'oreille du jury et est reconnue comme telle. Ainsi, il est relativement aisé de faire la distinction bipolaire voix normale / voix anormale. Selon Anders [Anders, 1988], « *la dysphonie est une qualité phonatoire qu'il est facile d'identifier (peut-être même de façon universelle) et les groupes expérimentés et inexpérimentés peuvent identifier de manière fiable la*

présence d'une dysphonie ». Ainsi, la distinction voix normale / dysphonie sévère est réalisée de façon constante et fiable par les jury d'écoute.

Dans cette logique, De Bodt [De Bodt, 1997] n'a pas inclus de sujets témoins (voix normales) dans son étude, car les auditeurs sont moins fiables dans leurs jugements quand il s'agit de voix pathologiques que lorsqu'il s'agit de voix saines [Kreiman, 1993, 1994]. Ainsi, pour eux, inclure des voix saines dans une étude, induit un biais qui améliorerait artificiellement la fiabilité du jury.

Comme nous venons de le voir, la différence entre une voix normale et une voix très dysphonique (par exemple entre une voix notée 0 et une voix notée 3 sur le paramètre G de l'échelle GRBAS) est évidente. Par contre, il est plus difficile de discriminer les nuances plus fines :

- voix normale (G = 0) / voix légèrement dysphonique (G = 1) ;
- voix légèrement dysphonique (G = 1) / voix moyennement dysphonique (G = 2) ;
- voix moyennement dysphonique (G = 2) / voix sévèrement dysphonique (G = 3).

En effet, il n'existe pas d'échelle de valeur interne permettant d'établir avec précision et de manière fiable une hiérarchie dans la dysphonie : les auditeurs expérimentés diffèrent davantage dans les tâche d'évaluation des voix pathologiques, ce qui tendrait à prouver que les prototypes de niveaux pathologiques varient d'un auditeur à l'autre. Les travaux de Kreiman suggèrent que les auditeurs "candides", qui ne subissent pas d'exposition particulière aux voix pathologiques, manquent de standards internes pour les qualités pathologiques. Leurs stratégies perceptives sont alors pratiquement similaires à celles des auditeurs expérimentés dont les jugements varient largement. C'est dans le but de contrôler cette inexpérience dans la hiérarchisation des troubles, que certaines méthodes standardisées d'analyse perceptive (dont le GRBAS) proposent une cassette de formation à l'attention du jury.

2.4. L'ancre auditif

Les standards internes de qualité de voix sont par définition instables et pourraient être influencés par d'autres facteurs que les caractéristiques intrinsèques du stimulus proposé. Par exemple, une question de « goût individuel » peut être prise en compte, certains juges ayant une « préférence » pour un type de voix pathologique ou seuil de tolérance supérieur pour une voix de type soufflé ou pour une voix de type rauque. On peut également prendre en compte le facteur culturel ou linguistique qui peut inciter l'auditeur occidental à entendre de la raucité chez un locuteur japonais, alors qu'il s'agit d'un trait phonologique constitutif de la langue [Yamaguchi, 2003]. D'une manière générale, on considère d'ailleurs que les auditeurs ne sont en mesure d'évaluer seulement les voix produites dans leur langue

natale. Au delà de ces facteurs, des effets de contexte pourraient causer des biais dans le jugement des auditeurs en influençant le standard interne sur lequel se basent les jugements. Parmi ces facteurs, on peut considérer la présentation des consignes, qui peut être exprimée différemment d'une session à l'autre, ou l' « humeur » de l'auditeur, son état de stress ou de détente, de disponibilité, qui peut être différent d'une session à l'autre et influencer son jugement.

Gerratt [Gerratt, 1993] a émis l'hypothèse selon laquelle « *substituer un standard explicite externe à ces référents internes instables pourrait améliorer la fiabilité des auditeurs.* » En effet, on peut considérer qu'un apprentissage par “ancrage auditif externe” de ce qui est discriminant dans la dysphonie optimiserait la fiabilité des évaluations perceptives. De la même façon, le nourrisson apprend à acquérir le système phonologique d'une langue par discrimination des phonèmes pertinents. Un protocole utilisant des voix de référence fixes pourrait contrôler les différences entre les auditeurs dans leurs standards internes pour les différentes qualités de voix, en fournissant à chaque juge un panel constant de référents (c'est à dire des exemples de chaque niveau pathologique pour chaque critère). Il serait alors possible de limiter la variabilité liée au contexte puisque les standards externes demeurent constants d'une session à l'autre et pour chaque auditeur.

Dans une étude intitulée « *Comparing internal and external standards in voice quality judgments* » [Gerratt, 1993], les jugements perceptifs étaient recueillis en utilisant deux protocoles. Le premier était une échelle traditionnelle de 5 points d'intervalles équidistants (EAI) où 1 représentait une voix normale et 5 une raucité sévère. Les auditeurs devaient concentrer leur jugement sur la raucité et ignorer tout autre aspect pathologique présent dans la voix. Le jury pouvait réécouter la voix autant de fois que nécessaire afin d'établir son évaluation. Le second protocole présentait une échelle de 5 points d'un nouveau type, basée sur l'ancrage auditif. Chaque valeur de l'échelle était représentée par un échantillon de voix synthétique. L'échantillon #1 représentait une voix normale et le #5 une raucité sévère. Cette fois, les auditeurs devaient sélectionner l'ancrage, le référent, qui représentait au mieux le degré de raucité présenté par le stimulus de voix. Les résultats obtenus par Gerratt et présentés dans le tableau 8 ci-dessous, montrent qu'un protocole d'ancrage auditif permet une amélioration potentiellement significative de la fiabilité des évaluations perceptives de la voix dans les pratiques cliniques et expérimentales.

	intra-rater reliability		inter-rater reliability	
	EAI scale	Anchored scale	EAI scale	Anchored scale
Mean % exact agreement	54.2	75.0	50.2	68.4
SD	9.57	12.33	8.88	12.54
Range	40.9-68.2	50-90.9	27.3-70.5	36.4-88.6
Mean % ± 1 scale value	94.0	99.6	92.8	98.7
SD	7.07	11.30	5.22	2.43
Range	81.8-100	95.5-100	77.3-100	90.9-100
Mean Pearson's r	.86	.95	.85	.93
SD	.08	.03	.06	.03
Range	.71-.95	.89-.98	.71-.94	.84-.98

Tableau V.8. : Fiabilité des jugements pour le EAI et le protocole d'ancrage [Gerratt, 1993]

L'ancrage auditif semble être une solution possible pour améliorer la fiabilité du jury mais les auteurs se heurtent actuellement encore aux moyens techniques permettant de créer les ancrages. Les voix synthétiques gardent un aspect artificiel qui rend difficile la comparaison avec les voix naturelles. Il reste extrêmement ardu de simuler les différents paramètres de la dysphonie (considérés comme prégnants par les échelles perceptives) par des échantillons de voix synthétique et d'établir un intervalle équidistant entre ces échantillons afin de produire un exemple pour chacun des 5 points de l'EAI.

2.5. La formation du jury d'écoute

Dans la littérature, de nombreux auteurs soulignent l'importance du niveau d'expérience des auditeurs pour la fiabilité des évaluations perceptives [Kreiman, 1992, 1993 ; De Bodt, 1996, 1997 ; Gerratt, 1993 ; Dejonckere, 1993 ; Anders, 1988 ; Bassich, 1986 ; Chan, 2002 ; Goldstone, 1998]. De fait, l'expérience est le critère primordial de recrutement des auditeurs pour un jury d'écoute. En effet, plusieurs années sont nécessaires pour développer un modèle stable des critères impliqués dans l'analyse perceptive des dysphonies [Kreiman, 1993 ; Goldstone, 1998]. Si De Bodt considère qu'un jury expert doit justifier d'un recul professionnel de 3 ans avec au moins 1/3 de ce temps consacré à l'évaluation vocale [De Bodt, 1997], les caractéristiques retrouvées le plus fréquemment dans la littérature parlent d'un minimum de 2 ans d'expérience professionnelle de la dysphonie [Kreiman, 1993 ; Rabinov, 1995]. Pour Goldstone [Goldstone, 1998], « *l'apprentissage perceptif implique des changements de longue durée pour le système perceptif d'un organisme qui améliore sa capacité à répondre à l'environnement et aux stimulations causées par cet environnement* ».

Par ailleurs, la composition d'un jury nécessite le recrutement de plusieurs auditeurs expérimentés, afin de former un groupe homogène. S'il n'existe pas de règle concernant l'effectif d'un jury d'écoute, la fiabilité des jugements réalisés par un seul auditeur, même très expérimenté, est fortement suspecte.

Ainsi, la revue de la littérature met en évidence les difficultés de composition d'un jury d'écoute : les critères déterminant le niveau d'expérience des auditeurs sont mal définis, le recrutement de plusieurs professionnels de la dysphonie présentant un niveau d'expérience équivalent est donc difficile. Dans ce contexte, la formation du jury d'écoute présente deux avantages manifestes : d'abord, le suivi de la formation peut être un critère objectif de niveau d'expérience ; ensuite, la formation à l'analyse perceptive peut permettre d'augmenter le nombre d'auditeurs susceptibles d'être recrutés.

Les procédures de formation proposées dans la littérature sont variées :

- l'utilisation de la cassette de formation est proposée dès 1981 par Hirano afin de familiariser les auditeurs à l'échelle GRBAS [Hirano, 1981]. Cet enregistrement est en réalité une sorte de « notice d'utilisation » et n'est disponible qu'en Japonais.
- l'utilisation d'ancrages permet de fournir aux auditeurs un ensemble d'exemples étalonnés représentant les différents paramètres de la dysphonie à différents niveaux de gravité [Gerratt, 1993 ; Chan, 2002]. Cette procédure implique l'utilisation de voix synthétique puisque le caractère multidimensionnel de la dysphonie exclut la possibilité de créer des continuums purs avec des voix naturelles. Ainsi, Chan [Chan, 2002] a recours à un sujet sain qui simule certains degrés de raucité ou de souffle, ce qui semble introduire un biais.
- l'évaluation de quelques voix non-inclues dans le corpus permet au jury de se familiariser avec l'échelle [Wolfe, 2000]. Il s'agit d'un entraînement à la méthode, mais pas d'une formation à l'écoute de la dysphonie.
- la correction par expertise a pour principe de faire évaluer une voix par un sujet « naïf » puis de lui fournir la réponse donnée par les experts [Chan, 2002].
- le processus de feedback permet de proposer de nouveau le stimulus au sujet naïf après une correction par expertise. Cette méthode permet d'accélérer l'apprentissage perceptif [Goldstone, 1998].

En 1996, Martin et Wolfe [Martin, 1996] proposent d'étudier les effets d'un entraînement perceptif basé sur des signaux de voix synthétiques. Leur objectif est double : 1- faciliter le développement de références internes chez un jury inexpérimenté, et 2- sélectionner un nombre limité de dimensions rapidement utilisables en entraînement et applicables à la voix pathologique. 28 étudiants ont participé à ce travail. Leur tâche consistait en l'évaluation de 60 /a/ tenus synthétiques selon les paramètres GRB, avant et après entraînement. La procédure d'entraînement était réalisée en trois étapes : d'abord, on leur faisait écouter un continuum de voix de normal à sévère pour chaque paramètre étudié ;

ensuite, on leur présentait des paires de voix avec pour consigne de désigner la plus dysphonique ; enfin, on leur présentait à nouveau les mêmes paires de voix en leur indiquant la « réponse juste ». Cette procédure d'entraînement incluait 101 items et durait environ 40 minutes. Les résultats ont montré que les étudiants présentaient avant entraînement un niveau de compétence assez disparate et que donc, les effets de la formation étaient plus bénéfiques pour certains que pour d'autres. Après entraînement, les auditeurs montraient une amélioration de leur capacité à discriminer les caractéristiques de la voix synthétique pathologique. Néanmoins, il faut préciser que cette étude menée sur voix synthétiques n'a jamais été adaptée aux voix naturelles.

Les travaux de Chan [Chan, 2002] se sont intéressés à l'effet d'ancrage auditif couplé à la notion de formation du jury d'écoute. Les auteurs ont proposé 4 protocoles d'évaluation perceptives : 28 voix d'hommes et de femmes ont été évaluées par un jury d'écoute formé d'auditeurs candides. La première évaluation s'est déroulée de façon classique, sur une échelle visuelle analogique. Puis deux sessions ont été proposées : une avec des voix d'ancrage synthétiques puis la seconde avec des voix d'ancrage naturelles. Ensuite, une session de formation a été proposée au jury d'écoute, utilisant un paradigme "stimulus/réponse/feedback/stimulus". Enfin deux sessions d'évaluation ont été de nouveau réalisées avec des voix d'ancrage synthétiques et naturelles. Les résultats ont montré que l'utilisation couplée de l'ancrage et de l'entraînement permettait d'améliorer la fiabilité des jugements perceptifs, en particulier pour les voix d'hommes. Malheureusement, l'association des deux méthodes et leur application conjointe ne permet pas de déterminer leur efficacité propre, ni le poids relatif de chacune d'elle dans les résultats. De plus, il semblait apparaître que les ancrages synthétiques étaient plus efficaces que les ancrages de voix naturelles. Cependant, il est à noter que, faute d'exemples prototypiques naturels pour certains niveaux pathologiques, les auteurs ont eu recours à une simulation par un spécialiste ayant une voix normale. En comparaison avec les exemples sonores synthétiques parfaitement calibré, il semble que la simulation présente un biais qui a pu diminuer les performances des ancrages naturels de façon artificielle. Néanmoins, ce travail illustre deux points essentiels : l'expérience du jury d'écoute est une nécessité pour la tâche d'évaluation perceptive et le choix du matériau d'ancrage présente des difficultés à l'heure actuelle quasi insurmontables : d'une part, les voix synthétiques sont trop artificielles pour être comparées aux voix naturelles et d'autre part, les voix naturelles ne sont pas suffisamment calibrées pour constituer un exemple représentatif d'un niveau et d'un type pathologique.

L'étude de la formation des auditeurs à l'écoute de voix pathologiques doit donc être poursuivie afin de proposer un protocole de formation permettant 1- d'améliorer la fiabilité des jugements perceptifs, 2- de proposer un enseignement spécifique aux étudiants en orthophonie, 3- de standardiser les compétences du jury d'écoute et 4- d'en augmenter les effectifs.

2.6. L'ancrage par comparaison

Si les protocoles d'ancrage auditif externe demeurent encore à l'état expérimental, des logiciels d'aide à l'analyse perceptive ont été développés ces dernières années dans un objectif essentiellement clinique. Leur principe est basé non plus sur l'évaluation de la dysphonie dans l'absolu, mais sur la comparaison de différentes voix. Il s'agit en fait de classer les voix proposées par ordre croissant de dysphonie. Dans ce cas, les voix proposées représentent un ancrage pour chacune des autres. L'avantage de cette méthode est d'être instinctive, facile à manipuler et relativement fiable, l'inconvénient est que les capacités de discrimination des auditeurs limitent le nombre de voix étudiées : au delà de 6 à 8 voix, la tâche de comparaison n'est plus possible. Cette approche est donc réservée à la pratique clinique puisque les études expérimentales nécessitent des corpus beaucoup plus importants.

Le premier logiciel d'assistance à l'analyse perceptive a été présenté par Hammarberg en 1997 [Hammarberg, 1997]. Dans le but d'améliorer la performance des sujets dans la tâche d'évaluation perceptive, Hammarberg a proposé un programme informatique basé sur l'utilisation d'une échelle visuelle analogique : le VIvisual SOrt and Rate (logiciel VISOR). Le principe était de présenter sous forme d'icônes une série de 5 voix synthétiques présentant divers degrés de pathologie. L'évaluation était réalisé par le classement des stimuli par ordre croissant du niveau pathologique, avec la possibilité de réécouter les voix aussi souvent que nécessaire pour effectuer la comparaison. Les résultats ont montré que cette méthode permettait d'améliorer la fiabilité des évaluations perceptives de façon significative. Cependant, ce logiciel conçu spécifiquement pour la langue suédoise n'est à l'heure actuelle toujours pas disponible.

3. Déroulement des sessions d'écoute

L'organisation et le déroulement des évaluations par le jury d'écoute doit être défini dans le protocole expérimental et doit faire l'objet d'une justification méthodologique par rapport à la question posée dans l'étude. En effet, le nombre d'évaluations organisées en sessions d'écoute, le mode d'évaluation, qu'il soit en individuel, en groupe ou par consensus, et l'ordre de présentation des stimuli sonores, sont susceptibles de modifier le comportement des auditeurs dans leur tâche d'évaluation et d'influencer les résultats.

3.1. Nombre d'évaluations

D'une manière générale, tenant compte du caractère variable des jugements perceptifs, il est préférable d'organiser plusieurs sessions d'écoute. En effet, la mesure de la variabilité intra-individuelle des auditeurs ne peut être évaluée que d'une écoute sur l'autre au cours de différentes sessions. D'autre part, la variabilité inter-individuelle est susceptible de varier d'une fois sur l'autre et il est intéressant de suivre son évolution. Une revue de la littérature montre que les auteurs accomplissent le plus souvent 3 sessions d'écoutes distantes d'une semaine, au cours desquelles le jury exécute la même tâche d'évaluation selon le même protocole expérimental [Crevier-Buchman, 1997; De Bodt, 1996, 1997; De Krom, 1994; Dejonckere, 1993; Giovanni, 1995, 1996, 1999; Wolfe, 1995; Yu, 2001, 2002]. En effet, si plusieurs sessions sont nécessaires, il faut tenir compte du fait que la réunion périodique des experts est difficile à mettre en place. Il est donc impossible en pratique de multiplier les sessions. Le nombre de 3 sessions d'écoute semble donc constituer un compromis convenable entre les nécessités méthodologiques (évaluation des variabilités du jury) et l'organisation pratique de l'étude (réunion des experts, recueil des données).

Les résultats recueillis au cours de 3 sessions distinctes peuvent être exploités de différentes façon :

- les données font l'objet d'une étude de variabilité des auditeurs. La variabilité intra-auditeur est décrite auditeur par auditeur sur l'ensemble des 3 sessions. La variabilité inter-individuelle est décrite session par session sur l'ensemble des auditeurs.
- les résultats des trois sessions d'écoutes peuvent être conservés et exploités ensemble pour le traitement statistique afin de conserver toute la richesse de l'information fournie par le jury.
- les résultats recueillis au cours des 3 sessions d'écoute peuvent faire l'objet d'une moyenne des évaluation de chaque auditeur afin de renforcer et de stabiliser le jugement d'expert.

Notons que la présence de 3 sessions d'écoute ne constitue pas de facteur d'influence quant à une éventuelle mémorisation des voix par le jury. En effet, il a été démontré qu'il n'existe pas de phénomène d'apprentissage en deçà de 5 sessions [Kreiman, 1993], et ce, quel que soit l'effectif du corpus étudié.

Notons également que l'évaluation des dysphonie est une tâche difficile pour les experts. Elle se révèle très coûteuse en temps et en concentration et l'examinateur doit s'assurer de la motivation des auditeurs lors du recrutement de son jury d'écoute. Certains auteurs mettent en place un système de rémunération sensé intéresser les auditeurs à l'étude [De Krom, 1994 ; De Bodt, 1996]. Notons qu'une session d'écoute ne devrait pas dépasser 45 minutes afin de conserver toutes les capacités de

concentration des auditeurs. Certains auteurs mettent en place des « pauses » durant lesquelles le jury peut se reposer et se détendre afin d'enchaîner 2 séries de 45 minutes [Dejonckere, 1993].

3.2. Mode de passation

Le déroulement de chaque session d'écoute, pour être valide, doit être strictement identique. La consigne est donnée aux auditeurs à chaque début de session. Elle doit être détaillée et rigoureuse. Certains auteurs choisissent d'en faire un enregistrement afin de standardiser les explications et les informations fournies [Hirano, 1981]. L'examinateur prend le temps de répondre aux questions méthodologiques éventuelles avant le début de la session.

Aucune information sur l'objectif de l'étude ne doit être donnée au jury. En effet, les auditeurs peuvent être influencés dans leur jugement par leur désir de « coller » aux résultats attendus. D'autre part, s'il est nécessaire d'un point de vue déontologique, de conserver l'anonymat des patients enregistrés, il faut garder à l'esprit que les auditeurs (souvent praticiens dans le même service) peuvent être influencés dans leur évaluation par la reconnaissance d'un sujet. Aucune identification des sujets participant à l'étude ne doit donc filtrer. Pour l'analyse perceptive, comme dans toute recherche, les travaux doivent être menés en aveugle.

Dans la littérature, on retrouve essentiellement 3 modes de passation :

- sessions individuelles [Giovanni, 1995 ; De Bodt, 1997 ; Dejonckere, 1993] : lors de sessions individuelles chaque auditeur procède à son évaluation indépendamment. Ce mode d'évaluation peut être nécessaire lorsque la réponse doit être donnée oralement. Il présente l'avantage de supprimer le risque de concertation et d'influence des auditeurs entre eux, mais alourdit considérablement l'organisation des sessions d'écoute. En effet, pour un jury composé de 7 auditeurs par exemple et 3 sessions pour chacun d'eux, cela nécessite 21 sessions pour l'examinateur.
- sessions de groupe [Yu, 2001 ; Wolfe, 1995] : tous les auditeurs membres du jury d'écoute effectuent leur évaluation de façon simultanée, dans la même pièce. Ce mode de passation est impossible lorsque la réponse doit être donnée oralement. Il présente pour avantage d'organiser seulement 3 sessions effectives, cependant, il augmente le risque de concertation et d'influence des auditeurs entre eux.
- sessions par consensus [Yu, 2002] : certains travaux n'ont pas pour objectif d'étudier la variabilité du jury d'écoute, mais de fournir une expertise du niveau de dysphonie des sujets de l'étude. Dans ce cas, le jury procède à une évaluation dite « par consensus » au cours de laquelle les auditeurs discutent leur jugement jusqu'à commun accord. Les sujets

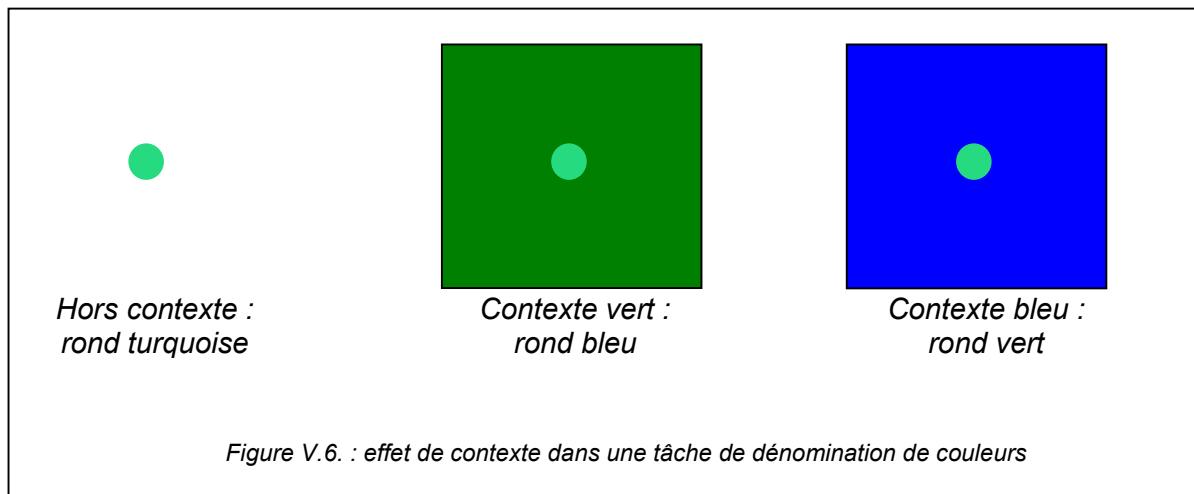
pour lesquels le jury ne parvient pas à l'unanimité sont alors exclus de l'étude. Ce mode de passation permet d'obtenir des évaluations solides qui ne sont plus discutables. Il est particulièrement intéressant pour les études visant à fournir une expertise pour tester la fiabilité des dispositifs instrumentaux de mesures objectives de la voix.

Le choix du mode de passation doit être défini dans le protocole expérimental avant de commencer les expérimentations. Il doit nécessairement tenir compte des objectifs de l'étude et de ses contraintes expérimentales propres.

3.3. Présentation des stimuli sonores

L'ordre de présentation des stimuli sonores doit tenir compte de la possibilité de caractéristiques vocales communes entre les patients. Par exemple, il ne peut être établi en fonction des pathologies présentées par les patients. En effet, il faut considérer qu'un groupe pathologique peut induire un type de dysphonie commun aux patients appartenant à ce groupe. Par exemple, la paralysie laryngée provoque des dysphonies où le timbre est à la fois soufflé et plus aigu. A l'inverse, le polype provoque une dysphonie de type rauque et aggravée. Ainsi, un « classement » des stimuli par groupes pathologiques est un facteur d'influence sur l'évaluation du jury. Il en va de même, pour un classement en fonction de l'âge des sujets, ou en fonction du sexe, etc. D'une manière générale, il est considéré que la présentation des stimuli sonores doit faire l'objet d'un ordre aléatoire, afin d'éviter tout biais lié à une possibilité de lien logique entre les voix [Kreiman, 1992 ; De Bodt, 1997]. Notons que certains auteurs admettent que l'ordre alphabétique peut être considéré comme aléatoire [Goldstone, 1998] : il n'y a pas de raison particulière qu'un lien logique existe en fonction du nom des patients. Cependant, en utilisant l'ordre alphabétique, deux enregistrements d'un même patient (pré et post thérapeutique par exemple) les deux stimuli seraient présentées successivement. Il est donc préférable de générer un ordre aléatoire au moyen d'un logiciel de calcul mathématique.

D'autre part, l'ordre de présentation des stimuli sonores revêt un caractère essentiel lorsqu'on sait que l'effet de contexte est un facteur d'influence du jury d'écoute [Goldstone, 1998]. En effet, lors d'une tâche d'évaluation perceptive, quelle qu'elle soit, le jury va être influencé par le stimulus précédent. Par exemple, dans une tâche visuelle de dénomination de couleur, la perception d'un même stimulus varie en fonction du contexte. Ainsi, un turquoise peut être sembler plutôt bleu en contexte vert ou vert en contexte bleu (figure V.6.).



De la même façon, dans une tâche auditive d'évaluation de la dysphonie, la voix précédente engendre un effet de contexte qui influence l'impression dysphonique du stimulus présenté. Par exemple, une voix présentant un niveau moyen de dysphonie paraîtra plus altérée si elle est présentée après une voix normale et pourra être désignée comme dysphonie sévère. Cette même voix (dysphonie moyenne) paraîtra moins altérée si elle est présentée après une dysphonie sévère et pourra être désignée comme dysphonie légère. Le seul moyen de neutraliser l'effet de contexte réside dans la présentation des stimuli dans un ordre aléatoire différent au cours des 3 sessions d'écoute.

En conclusion, dans le but de limiter les facteurs d'influence liés à la méthode, l'ordre de présentation des stimuli sonores doit tenir compte de ces observations et répondre à la double exigence suivante : un ordre aléatoire, différent à chaque session d'écoute.

4. Le matériau phonétique

Il s'agit d'un problème central des études concernant la voix. Le matériau phonétique est le “support vocal” qui est analysé par les appareillages de mesures objective ou par un jury d'écoute. Il doit être à la fois standardisé et représentatif de la voix d'un patient. Deux types de matériau phonétique sont actuellement utilisés : la parole (discours spontané [Anders, 1988], répétition de mots ou d'une phrase type [Chevrie-Muller in Giovanni, 1995] ; lecture d'un texte de référence [Askenfelt, 1986 ; Hammarberg, 1980]), ou la voyelle tenue [Gerratt, 1993 ; Kreiman, 1993].

4.1. La parole

Le matériau le plus représentatif de la voix conversationnelle d'un patient est la parole. Tout le monde a une grande expérience d'écoute de la parole. La parole est sans aucun doute le matériau clinique par excellence : c'est sur de la parole que le clinicien et le patient se forgent une opinion. C'est aussi, pour le jury d'écoute le matériau le plus intuitif et le plus facile à évaluer, du fait de son habitude à écouter le discours d'un dysphonique, et de la quantité considérable d'informations contenues dans la parole : outre la production des phonèmes voisés qui témoignent de la vibration des cordes vocales, elle comprend des transitoires (attaque / sortie ; passage entre phonème non-voisé / voisé), les phénomènes de coarticulation, les transition formantiques, hauteur, intensité, intonation, mélodie, etc. Les fragments de parole constituent le matériau phonétique privilégié pour les études perceptives. Il peuvent être constitués de répétition de mots ou de phrases, mais le plus souvent, il s'agit de la lecture d'un texte de référence phonétiquement équilibré (c'est à dire qui tend à contenir tous les phonèmes de la langue dans toutes les combinaisons possibles). Le « *Rainbow Passage* » est le plus utilisé pour les anglophones, en français, on retrouve fréquemment le 1^{er} paragraphe de « *La petite chèvre de Monsieur Seguin* » d'Alphonse Daudet. L'utilisation systématique d'un même texte permet la standardisation et la comparaison des données.

4.2. La voyelle tenue

Parmi les voyelles tenues, la voyelle la plus utilisée est le /a/ mais d'autres choix peuvent être faits comme par exemple, le /i/, le /u/, ou des comparaisons de voyelles. Le /a/ tenu est intéressant pour sa simplicité de mise en action dans la pratique clinique courante. Il s'agit d'une voyelle commune à de nombreuses langues, ce qui simplifie la question de l'accent et rend possible la constitution de banques de données internationales [Beckman, 1995]. Elle est peu complexe à produire, facilement reproductible et peu perturbée par l'émotion du sujet. Le /a/ est la voyelle la plus « ouverte », et donc la moins affectée par les phénomènes engendrés par le tractus vocal supra-glottique. Les possibilités de contrôle et de standardisation par l'examineur semblent correctes. Selon De Krom [De Krom, 1994], le fait que les voyelles tenues soient moins affectées par la confusion articulatoire (influence des phonèmes proches) et les contextes prosodiques (accents régionaux, etc.) permet d'aider l'auditeur à se concentrer sur les caractéristiques laryngées.

La voyelle tenue est le matériau préféré par la majorité des auteurs qui travaillent sur l'analyse objective de la voix [Baken, 1987 ; De Krom, 1995 ; Titze, 1995 ; Wuyts, 1997 ; entre autres]. La plupart des techniques d'analyse acoustique ne se révèlent fiables que si le signal laryngé à analyser

est relativement stable, comme lors de la production des voyelles tenues. Les variations du signal comme les intervalles entre les segments voisés et non-voisés, les variations de hauteur et les transitions formantiques rapides (typiques de la parole) ne peuvent pas, la plupart du temps, être traitées.

C'est la raison pour laquelle la plupart des études objectives de la voix utilisent la voyelle tenue et plus particulièrement la partie stable de la voyelle tenue. Les transitoires d'attaque et de sortie du son correspondent à la mise en vibration des cordes vocales. Pour les cas de voix pathologiques, cette mise en vibration se révèle à la fois beaucoup plus longue que dans les cas de voix normales, et très perturbée ce qui augmente les mesures d'instabilité de manière disproportionnée. La plupart des études objectives excluent donc les transitoires pour n'explorer que la partie stable du /a/ tenu, c'est à dire la portion du signal où la vibration des cordes vocales atteint son maximum de régularité. Certains auteurs proposent une mesure de l'attaque calculée comme le temps mis par la vibration à atteindre une stabilité [Giovanni, 1995].

4.3. Choix du matériau phonétique

En théorie, différents types de fragments de voix peuvent être utilisés pour l'analyse perceptive de la qualité de la voix. Cependant, en pratique, les études employant les voyelles tenues sont plus nombreuses que celles utilisant la voix conversationnelle, et ce, principalement pour des raisons de facilité de mise en place du protocole.

Pourtant, beaucoup de chercheurs s'accordent à dire que si la parole est représentative de la voix conversationnelle, les voyelles tenues sont davantage caractéristiques de la voix chantée. La voix humaine est avant tout utilisée pour l'expression de la communication au quotidien, donc pour la parole. La validité de l'évaluation de la qualité de la voix obtenue à partir de voyelles tenues est donc discutable [De Krom, 1994].

Hammarberg [Hammarberg, 1980] considère que, « *d'un point de vue clinique, il n'est pas satisfaisant d'évaluer la qualité de la voix d'après une voyelle tenue isolée. Les changements survenant au cours de la parole articulée comme les transitoires d'attaque et de sortie, les cassures dans la phonation, etc. sont des indices cruciaux de qualité de la voix, qui ne peuvent pas apparaître lors d'une voyelle tenue.* » En effet, la production de fragments de parole implique une activité musculaire laryngée plus complexe et plus dynamique que la production de voyelles tenues. Pour produire une voyelle tenue, les cordes vocales n'ont besoin d'être mises en adduction qu'une seule fois

alors que la parole est caractérisée par une série coordonnée d'adduction et d'abduction des cordes vocales. Du fait de sa complexité physiologique, la parole met la voix dans une situation d'utilisation complète qui correspond à une situation d'effort pour une voix pathologique. Il ne semble pas illogique d'imaginer que les différents aspects pathologiques d'une voix seront mieux mis en évidence et donc évalués de façon plus fiable à partir de fragments de parole qu'à partir de voyelles tenues, de la même façon qu'une faiblesse cardiaque est explorée par un test d'effort.

Les arguments en faveur de la voyelle tenue sont donc plus pragmatiques et concernent plutôt les mesures objectives, tandis que les arguments en faveur de la parole sont plus théoriques et concernent davantage l'évaluation perceptive par un jury d'écoute. Dans ce sens, Giovanni [Giovanni, 1995], dans la définition d'un protocole d'analyse objective de la voix, réalise les mesures objectives sur le /a/ tenu, alors que son évaluation perceptive, considérée comme le standard de référence est obtenue sur de la parole. Cette dichotomie illustre la diversité des points de vue présentée dans la littérature. Utiliser la parole pour l'analyse perceptive et la voyelle tenue pour l'analyse objective est un compromis nécessaire si l'on considère d'une part que les appareillages de mesures ne sont performants que sur des voyelles tenues, et d'autre part que la parole articulée demeure pour le jury le matériau phonétique le plus représentatif de la voix d'un patient. Ce compromis permet de mettre ces deux systèmes d'analyse à leur meilleur niveau de performance. Mais la comparaison entre les données objectives et les données perceptives est rendue difficile par cette différence de matériau [Giovanni, 1995] : « *la dissociation entre les données perceptives et les mesures objectives semble correspondre à des cas pour lesquels il existe une difficulté lors de la mise en vibration des cordes vocales suivie d'une stabilisation du signal par un processus physiologique. Lors de l'analyse perceptive de la parole articulée, le jury ne pouvait interpréter qu'une série d'attaques vocaliques sans jamais pouvoir juger de voyelle stable. Le sujet était donc classé dans un grade de dysphonie élevé. En revanche, les mesures objectives étaient réalisées sur la partie la plus stable de la voyelle tenue et le sujet obtenait donc des valeurs correspondant à un grade de dysphonie moins élevées.* » Par ailleurs, Schoentgen [Schoentgen, 1989] a mesuré le jitter sur des fragments de parole et sur des voyelles tenues produits par des sujets sains et des sujets dysphoniques. Bien que Schoentgen n'ait pas trouvé que le jitter mesuré à partir des segments de parole soit plus discriminant que le jitter obtenu à partir des voyelles tenues, les valeurs du jitter étaient plus hautes pendant les transitoires d'attaque et de sortie. Koike [in De Krom, 1994] a démontré que le jitter était plus important pendant les transitoires que dans la partie stable des voyelles tenues. Les expériences de Koike et de Schoentgen montrent que l'attaque et la sortie de la voyelle tenue pourraient fournir plus d'informations concernant la dysphonie que la partie stable de la voyelle tenue durant laquelle les cordes vocales restent en position statique d'adduction. Ces exemples illustrent la nécessité de définir le matériau phonétique le plus pertinent pour l'analyse de la voix.

Comme nous l'avons vu, la plupart des études actuelles fondent leur protocole de recherche sur un matériau phonétique : le /a/ tenu. Il présente pour avantages d'être facilement reproductible et interprétable par les logiciels de mesures. Cependant, on peut lui reprocher de ne pas être suffisamment représentatif de la voix usuelle. En effet, c'est sur de la parole continue que le clinicien (ou le jury) juge la voix d'un patient. Wolfe [Wolfe, 1995] décrit une expérience perceptive dont l'objectif est de déterminer dans quelle mesure la voyelle est représentative de la voix usuelle. Des échantillons de voix étaient enregistrés pour 20 sujets normaux et 60 patients présentant des pathologies courantes (nodules, paralysie unilatérale, et dysphonies fonctionnelles). Les échantillons de voix étaient évalués par 25 auditeurs sur la base de 12 échelles bipolaires sémantiquement différentes. Wolfe observe que le /a/ tenu est souvent jugé plus sévèrement que la parole continue. Cependant, elle obtient une corrélation de 0,78 entre les grades attribués par un jury d'écoute sur le /a/ tenu et sur la parole articulée, ce qui laisse supposer que le /a/ tenu est un matériau convenablement représentatif. Pour De Krom [De Krom, 1994], il apparaît que les phonèmes en coarticulation ne sont pas nécessairement préférables à une voyelle tenue pour une analyse perceptive du grade global de dysphonie, de la raucité et du souffle.

Le plus souvent, lors des mesures effectuées sur le /a/ tenu, l'attaque (qui est la période de mise en vibration des cordes vocales), est exclue à cause des difficultés d'interprétation que cela pose aux logiciels de mesure. La fenêtre de mesure est placée sur la partie stable du signal (c'est à dire sur la partie où la vibration atteint son maximum de régularité). Mais l'attaque est porteuse d'information sur la qualité de la voix. On sait qu'elle peut être altérée, en particulier dans sa longueur. Lors de la production de parole articulée, elle fait partie intégrante de la voix. De Krom a supposé que les transitoires d'attaque contiennent une information perceptive nécessaire à l'évaluation par le jury. Il a spécifiquement étudié la fiabilité d'un jury en fonction du matériau : il décrit une expérience perceptive dans laquelle un jury de 6 auditeurs doit évaluer des fragments de voix (grade de dysphonie, souffle, raucité) recueillis chez 78 sujets (57 dysphoniques et 21 sains) en utilisant l'échelle GRBAS, sur 10 niveaux. 4 différents types de stimulus étaient présentés au jury.

- 1 type de stimulus était basé sur des fragments de parole (en coarticulation)
- les 3 autres étaient basés sur différents fragments de voyelle tenue (sur 200 ms à l'attaque; 200 ms dans la partie stable; et 1000 ms sur toute la durée du stimulus).

Les analyses portaient sur la constance et la fiabilité du grade (G), de la raucité (R), et de l'aspect soufflé (B). Les fragments de voyelle stable étaient jugés de façon moins fiable, que la voyelle complète ou la parole.

Stimulus types	Voice quality aspects		
	Grade	Breathiness	Roughness
Connected speech	0.94	0.88	0.88
Whole vowel	0.93	0.93	0.88
Post-onset vowel	0.90	0.88	0.79

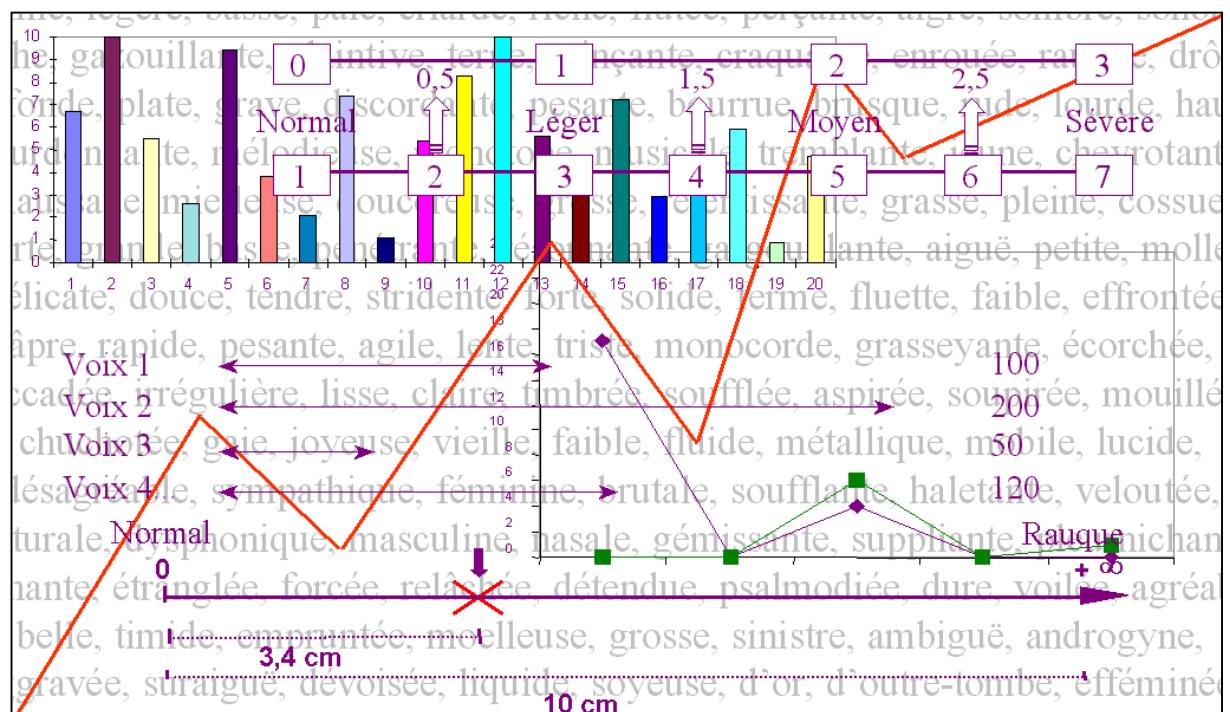
Tableau V.9. : Fiabilité des jugements (coefficients ρ_r) pour les différents types de stimulus et pour les paramètres GBR

De façon inattendue, le jugement de la partie stable du /a/ tenu n'était pas davantage perturbé sur R que sur B, alors qu'on aurait supposé que l'attaque contenait surtout des informations d'instabilité, donc de raucité. De Krom a montré que les voyelles tenues étaient un matériau pertinent permettant au jury d'être constant et fiable, si ces voyelles contenaient l'information de l'attaque. De Krom concluait que l'analyse des voyelles isolées pouvait être proposée en remplacement de la parole articulée dans les études visant à tester les méthodes objectives ; le /a/ tenu demeurant quand même un matériau plus simple à analyser et plus économique.

Au total, la perception de la dysphonie est une tâche complexe qui ne peut être effectuée intuitivement. Ces dernières années, les recherches menées sur ce thème ont permis d'améliorer les protocoles mis en place, que ce soit dans des objectifs expérimentaux perceptifs ou instrumentaux. Néanmoins, l'évaluation perceptive reste considérée comme « subjective ». La variabilité des jugements perceptifs, combinée à la diversité des protocoles recensés dans la littérature, rendent extrêmement difficile l'interprétation des résultats. Le manque de fiabilité des résultats demeure un biais majeur, qui, à l'heure actuelle, n'est pas résolu.

Dans la section suivante, nous présentons l'ensemble des travaux personnels poursuivis sur le thème de l'analyse perceptive. Ce travaux sont répartis selon trois grands axes :

- perception/jury : influence de la méthodologie mise en œuvre sur les résultats de l'analyse perceptive
- perception/mesures : optimiser le protocole expérimental d'analyse perceptive conçu spécifiquement pour la comparaison avec les données instrumentales
- perception/phonétique : influence des contextes de production sur la manifestation dysphonique et sa perception



Section expérimentale

1. Protocole expérimental

Les travaux que nous présentons ont tous suivi les mêmes procédures de sélection des sujets et d'enregistrement de la voix. Afin d'éviter les redondances, nous présentons ici le mode de recrutement des sujets dysphoniques, les consignes et les matériels utilisés pour le stockage des échantillons.

1.1. Les sujets étudiés

Le diagnostic a été posé par laryngoscopie indirecte, réalisée soit au moyen d'un optique rigide (dans la plupart des cas), soit au moyen d'un fibroscopie (dans les cas présentant un réflexe nauséux important) lors de l'examen ORL. L'examen stroboscopique a été utilisé chez les patients au diagnostic délicat (par exemple, kyste intra-cordal). Les enregistrements ont été effectués avant tout traitement médicamenteux ou chirurgical et avant toute prise en charge rééducative (sauf pour les études concernant le résultat post-opératoire)

Les pathologies rencontrées correspondaient soit à une anomalie morphologiques des cordes vocales (lésions d'origine dysfonctionnelle ou lésions d'origine organique) soit à une anomalie fonctionnelle de la mobilité des cordes vocales (paralysie laryngée unilatérale).

1.2. L'enregistrement de la voix

Les enregistrements vocaux ont été réalisés dans une pièce insonorisée à l'aide d'un magnétophone DAT (Tascam DA-20). Le microphone (Sennheiser) était placé à 20 cm de la bouche du patient, légèrement décalé sur le côté afin d'éviter les bruits de souffle. L'examineur veillait à ce que le gain soit suffisant tout en évitant la saturation afin d'utiliser au mieux toute la dynamique du système. Les sujets étudiés avaient pour tâche de lire un texte et de produire une série de voyelles tenues sur plusieurs secondes. Les consignes étaient données à chaque sujet de manière identique : "*Parlez à hauteur et intensité confortable, le plus naturellement possible*". Plusieurs essais étaient effectués jusqu'à ce que la qualité des enregistrement et des productions soient jugées satisfaisantes par l'examineur, c'est-à-dire correspondant perceptivement à la voix spontanée conversationnelle du sujet.

1.3. Le matériel phonétique

1.3.1. Echantillon de parole

Le texte utilisé était le plus souvent le premier paragraphe de « La petite chèvre de Monsieur Seguin » d'Alphonse Daudet : « *Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon. Un beau matin, elles cassaient leur corde, s'en allaient dans la montagne et là-haut, le loup les mangeait. Ni les caresses de leur maître, ni la peur du loup, rien ne les retenait.* » Pour quelques études basées sur l'approche phonétique de la perception de la dysphonie, ce texte a pu être remplacé par des phrases conçues spécifiquement. Les changements de texte seront donc précisés au fur et à mesure de la présentation des travaux, si nécessaire.

L'intégralité du texte enregistré était numérisée par ordinateur (fréquence d'échantillonnage à 22050 Hz, résolution en amplitude à 16 bits) grâce au logiciel Cool Edit®. Seule une phrase était retenue pour l'analyse perceptive : « *Il les perdait toutes de la même façon.* »

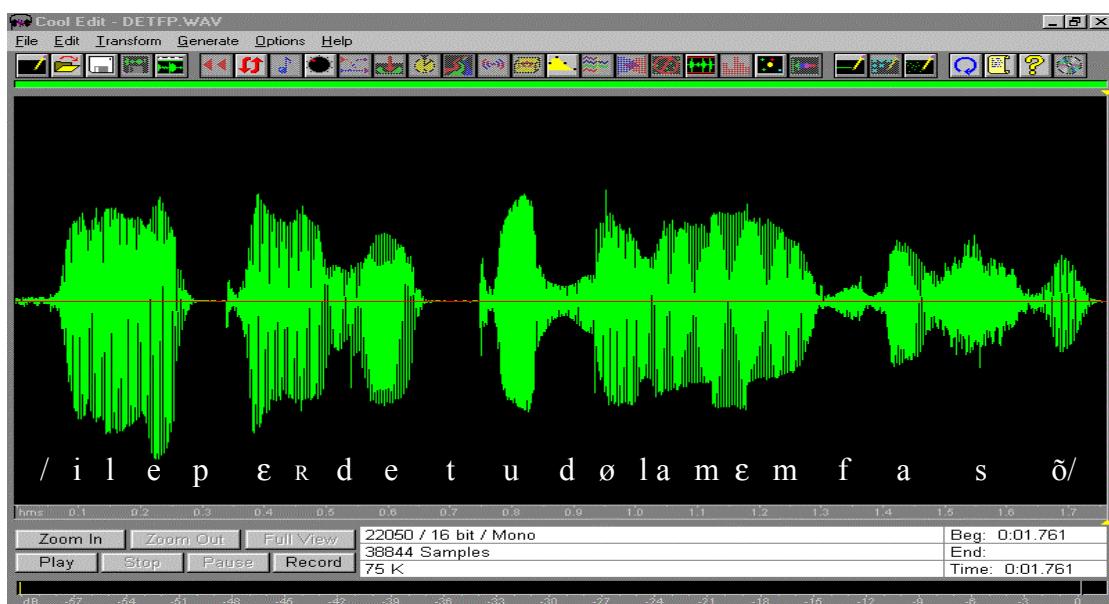


Figure VI.1. : Présentation de la phrase issue de « *La petite chèvre de Monsieur Seguin* » retenue pour les analyses perceptives et numérisée grâce au logiciel CoolEdit ® : « *Il les perdait toutes de la même façon.* »

1.3.2. La voyelle tenue

Conformément aux données de la littérature et dans la logique des protocoles d'analyse instrumentale, notre choix s'est porté sur la voyelle /a/ (voir section V., chapitre 4.2., p 99). Les sujets étudiés devaient produire une série de 5 /a/ tenus sur plusieurs secondes. Seules les 3 productions médianes

étaient retenues pour l'étude : le premier /a/ était supprimé afin d'éviter l'effet d'inhibition et de stress dû à la situation d'examen ; le dernier était supprimé afin d'éviter les risques de fatigabilité de la voix, fréquents chez les dysphonies sévères. Nous considérons que les 3 productions médianes étaient les plus représentatives d'une situation « normale ». De la même façon que pour le texte enregistré, la consigne était « *Parlez à hauteur et intensité confortables, le plus naturellement possible.* » Plusieurs essais étaient effectués jusqu'à ce que l'examinateur soit satisfait des productions et de la qualité de l'enregistrement. Les /a/ tenus étaient ensuite numérisés grâce au logiciel CoolEdit ® (fréquence d'échantillonnage à 22050 Hz, résolution en amplitude à 16 bits).

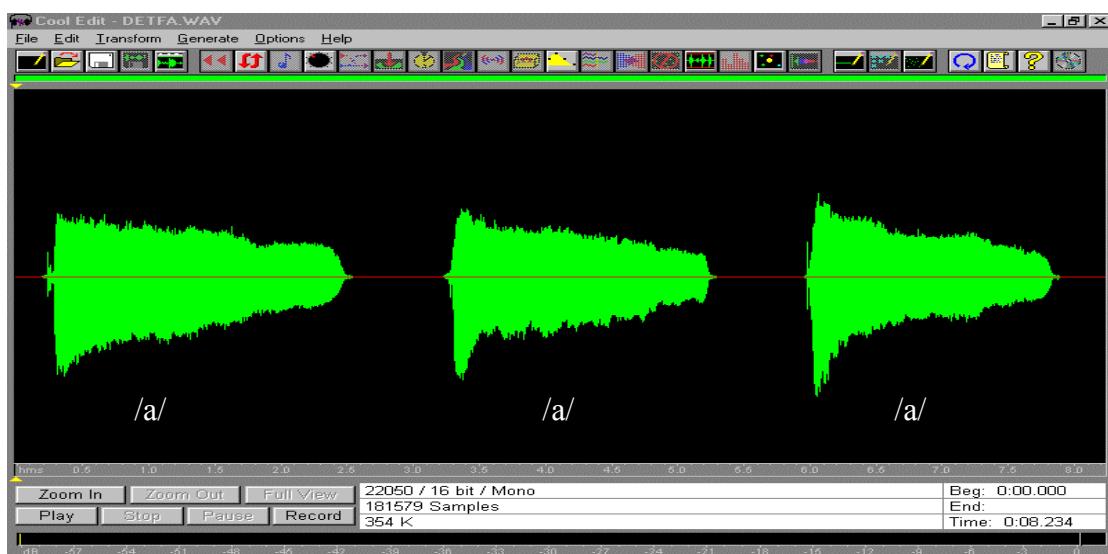


Figure VI.2. : Présentation des 3 /a/ tenus médians retenus pour les analyses perceptives et numérisés grâce au logiciel CoolEdit ®.

Pour certaines études, s'intéressant spécifiquement à l'attaque, les échantillons de /a/ tenu ont fait l'objet d'une manipulation informatique permettant de supprimer l'attaque et/ou la finale afin de ne conserver que la partie stable de la voyelle. La technique de traitement du signal mise en place pour ces études sera décrite dans la partie « protocole expérimental » des travaux concernés.

1.4. L'analyse perceptive

Les échantillons de voix étaient analysés par un jury d'écoute dont la composition a pu varier au cours des différentes études que nous présentons. Ce point précis sera donc détaillé au fur et à mesure de la présentation des travaux.

Afin de faciliter les écoutes par le jury et de standardiser les procédures et les consignes d'analyse, nous avons procédé à la conception et à la réalisation d'une station d'écoute avec le logiciel

FileMakerPro 3 ®. Ce programme pouvait être modifié en fonction de l'échelle utilisée (ici, GRBAS sur échelle ordinaire à 4 niveaux) et ne convenait que dans le cas d'un mode de passation individuel. L'arborescence du programme se présente de la façon suivante :

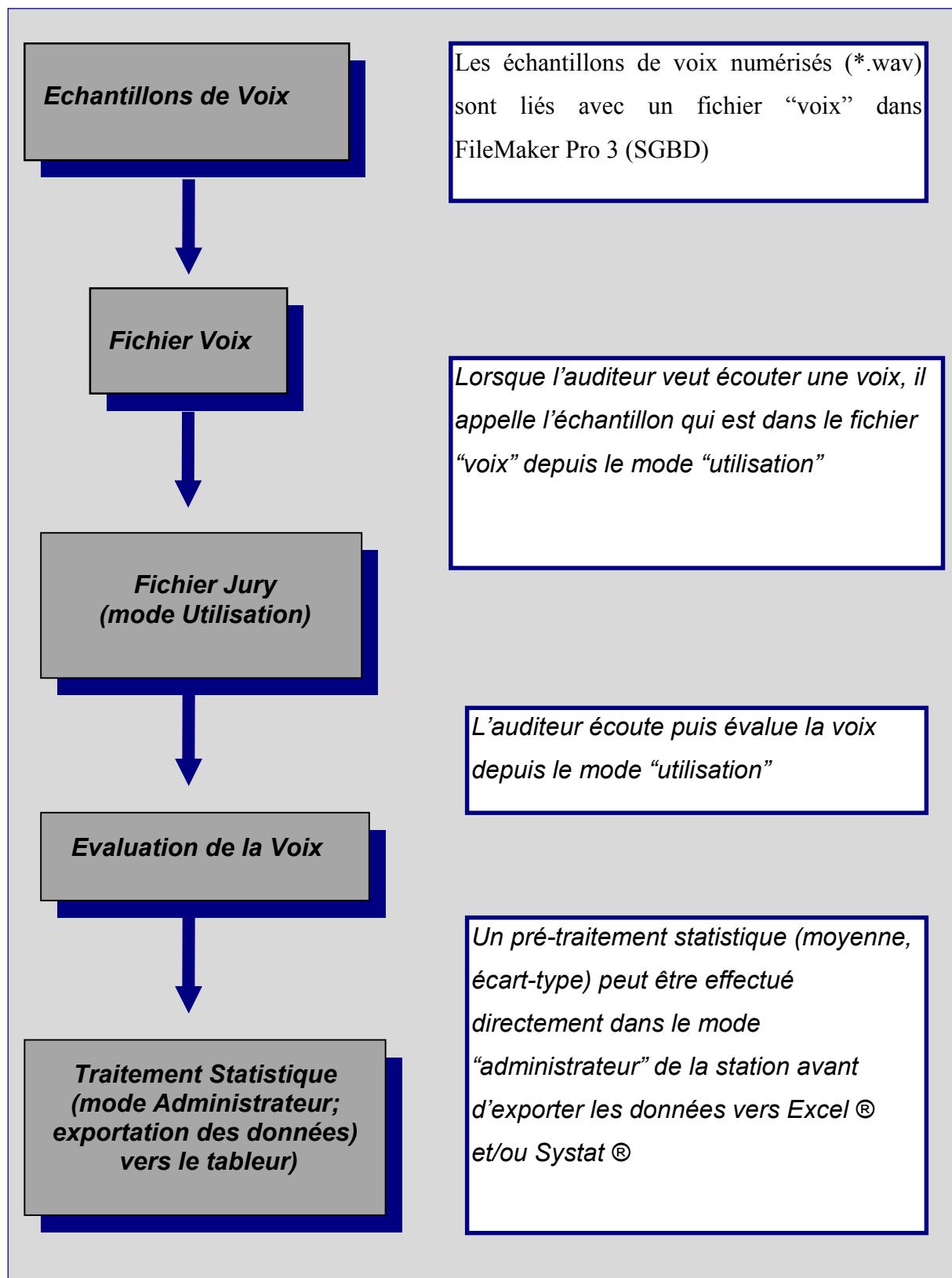


Figure VI.3. : Arborescence de la Station d'Ecoute réalisée dans le logiciel FileMaker Pro 3 ®.

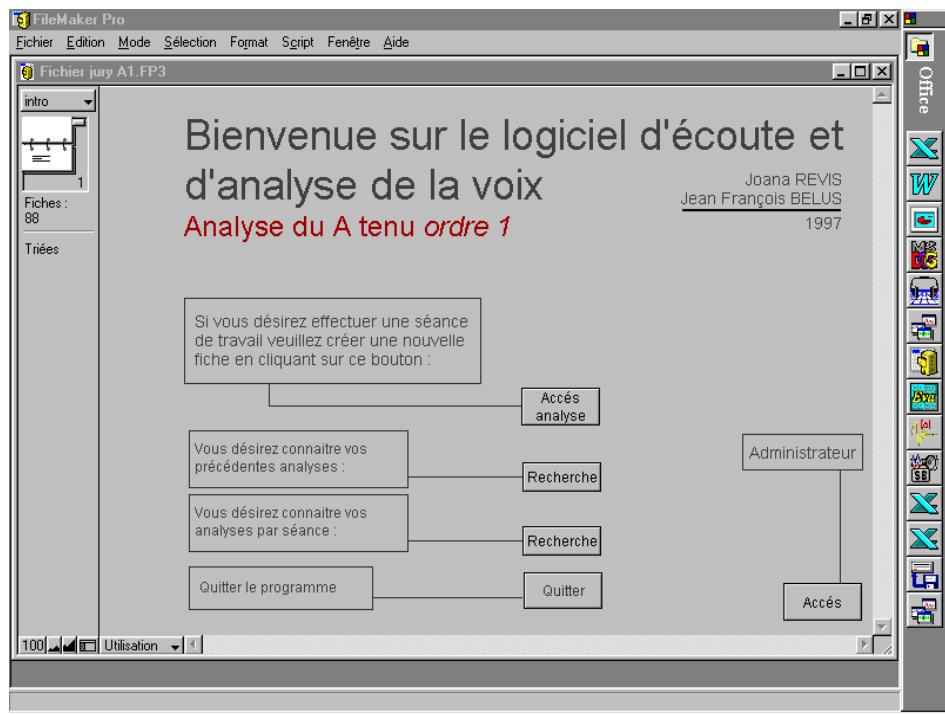


Figure VI.4. : Fenêtre de menu principal de la Station d'Ecoute réalisée grâce au logiciel FileMaker Pro 3 ®.

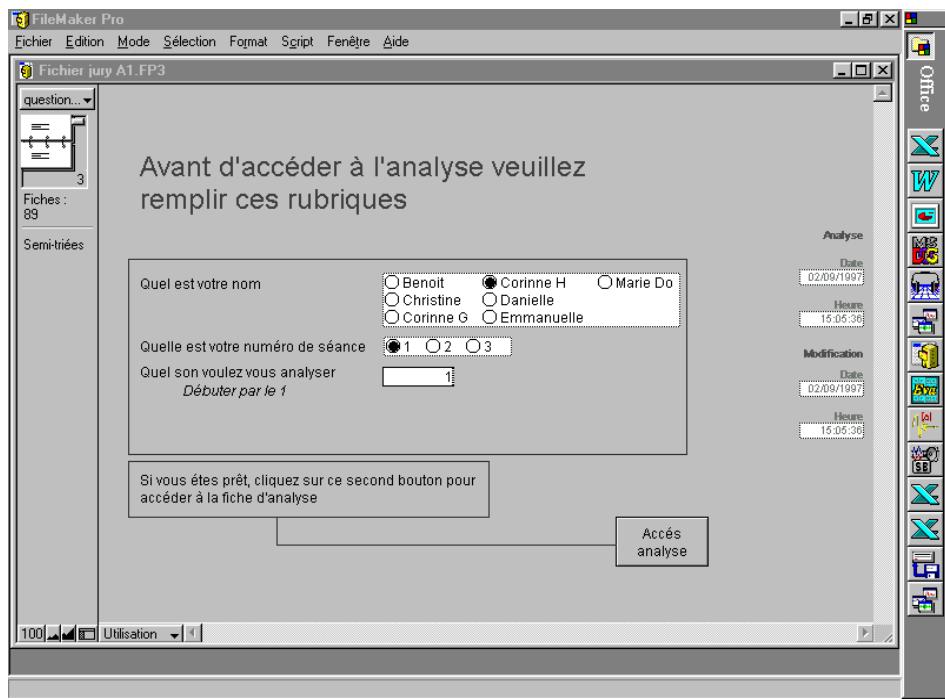


Figure VI.5. : Menu d'identification du jury de la Station d'Ecoute

Les échantillons de voix numérisés sont liés avec un fichier appelé fichier « voix » dans FileMaker Pro 3 ®. Ce fichier voix est lui-même lié à un fichier “utilisation” dans lequel le jury écoute et évalue les voix.

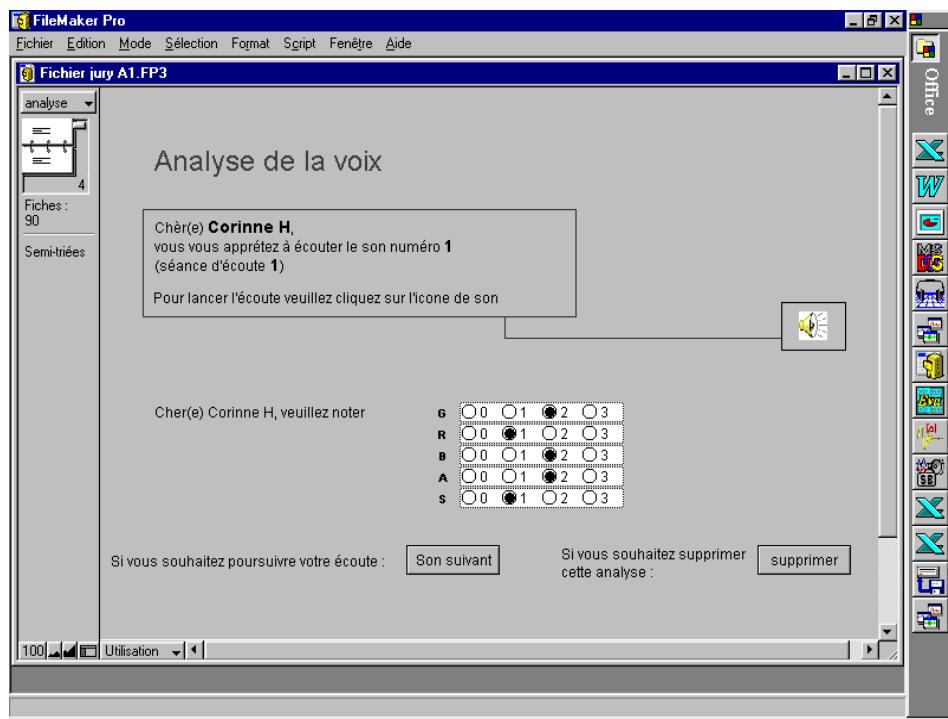


Figure VI.6. : Menu “Utilisation” de la Station d’Ecoute.

Chaque auditeur a la possibilité d’écouter le stimulus plusieurs fois en cliquant sur l’icône et de modifier sa cotation en cochant la note choisie, s’il le juge nécessaire. Aussitôt qu’il est satisfait, il passe à la voix suivante qui est activée automatiquement en cliquant sur le bouton « Son Suivant ».

Seul l’examinateur a la possibilité d’accéder au mode « Administrateur », protégé par mot de passe.

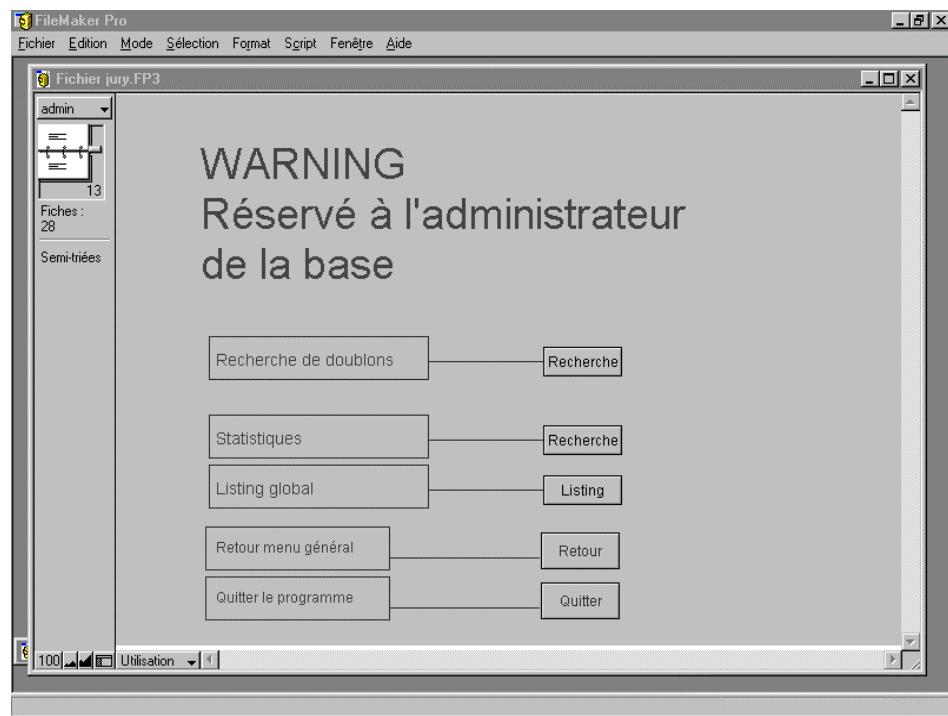


Figure VI.7. : Menu “Administrateur” de la Station d’Ecoute .

Ce module permet d'accéder à une liste complète de toutes les évaluations effectuées par le jury (« Listing global »), de procéder à une recherche des erreurs (présence de doublons ou de manques) ainsi que de réaliser un pré-traitement statistique (moyennes et écarts-types), avant d'exporter les données vers un tableur de type Excel ®. C'est également à partir du mode « Administrateur » que le responsable de la recherche a la possibilité de modifier la présentation des fiches afin de changer les paramètres qualitatifs ou l'échelle quantitative proposés au jury d'écoute.

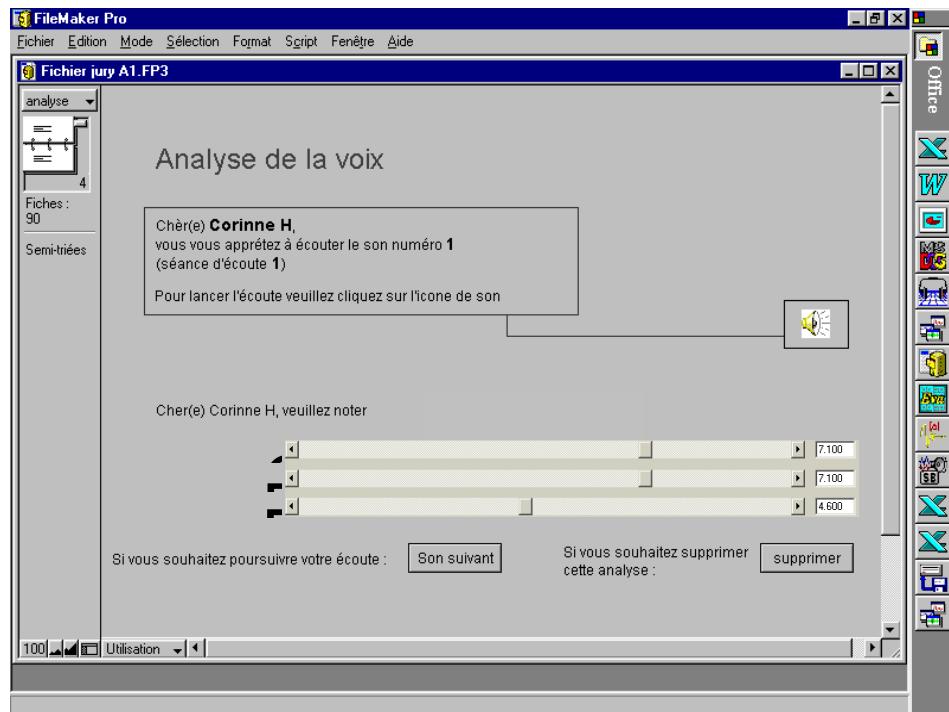


Figure VI.8. : Modification des paramètres des fiches d'évaluation : paramètres qualitatifs GRB seulement, échelle quantitative visuelle analogique.

Cette station d'écoute a été utilisée pour tous les travaux nécessitant une analyse individuelle par chacun des membres du jury d'écoute séparément. L'examinateur était toujours présent afin de pouvoir résoudre des problèmes liés à la manipulation du programme (si par exemple un auditeur cliquait par erreur sur « Son Suivant » avant d'avoir coché les réponses). Néanmoins, l'utilisation de ce programme présente deux avantages primordiaux : les consignes sont présentées systématiquement de la même façon au jury puisqu'elles sont fournies par le programme lui-même, et surtout le transfert automatique des résultats vers le tableur permet de s'affranchir complètement des risques d'erreur contemporains du recueil des données.

Les travaux nécessitant une évaluation du jury par groupe ont été réalisés sur le même ordinateur avec la même carte son et les mêmes enceintes. Dans ce cas, l'examinateur procédait au lancement des

fichiers son à partir de l' « Explorateur Windows », et utilisait le simple magnétophone proposé par Windows afin d'effectuer la lecture des stimulus.

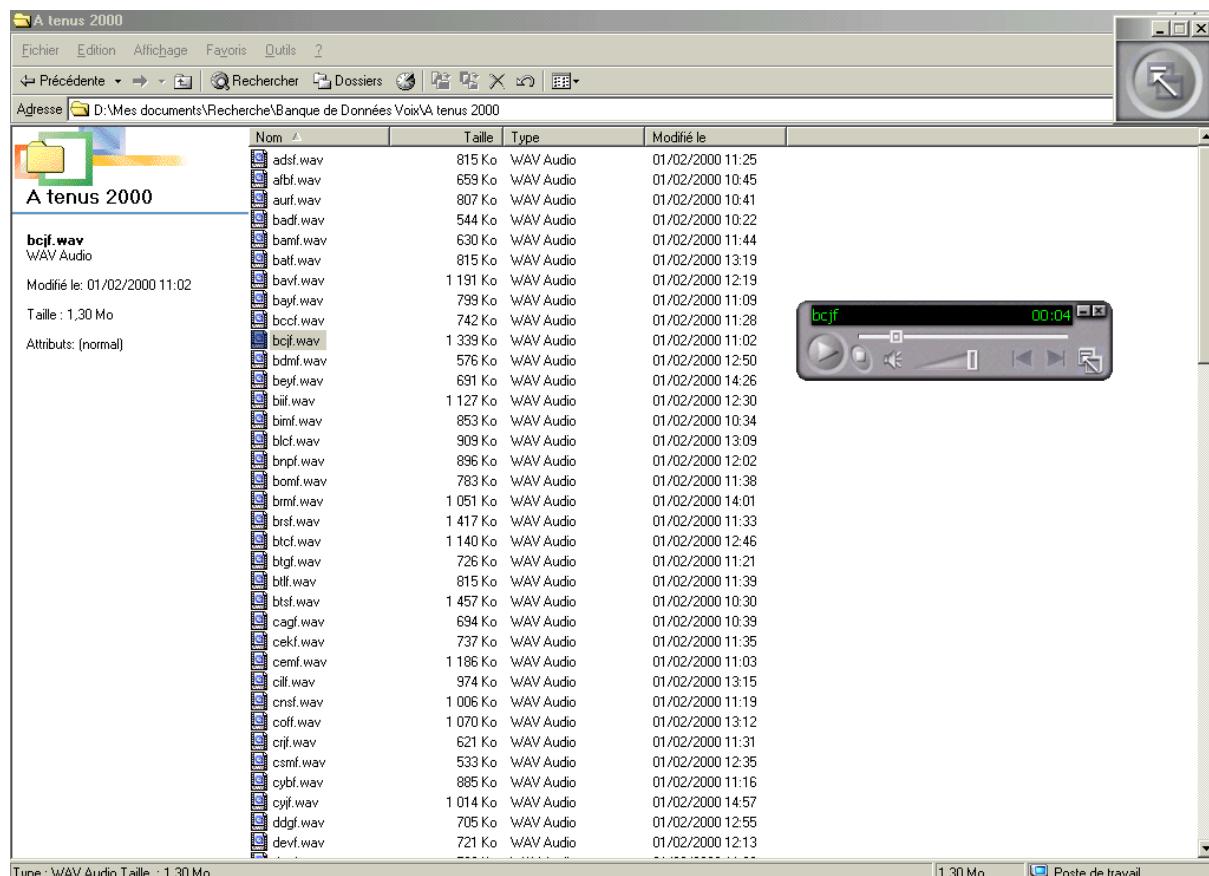


Figure VI.9. : Affichage des stimuli et présentation des fichiers audio depuis l' « Explorateur Windows » avec le logiciel « Windows Media Player ® », dans le cas d'évaluations par groupe.

Le magnétophone Windows était préféré à tout autre éditeur de signal pour essentiellement deux raisons : d'abord, il s'agit du programme exécuté par défaut, son ouverture est automatique lors du double-clic, cela permet donc un gain de temps considérable ; d'autre part, il permet d'écouter la voix sans afficher le signal acoustique dont le visionnage pourrait fournir une information autre qu'auditive à notre jury d'écoute, habitué à effectuer des tâches de traitement du signal de voix pathologique.

Toujours dans le cas d'évaluations en groupes, les auditeurs disposaient de fiches d'évaluation conçues pour la présentation des échelles choisies, qu'ils remplissaient à la main au moyen d'une croix. Ces fiches pouvaient se présenter de deux manières différentes essentiellement : soit sous forme de tableaux à double entrée permettant une cotation de plusieurs paramètres sur une échelle ordinaire (figure VI.10.), soit sous forme de droites calibrées de 10 cm représentatives d'une échelle visuelle analogique (figure VI.11.).

Nom :	Date :				
Prénom :	Evaluation n° :				
Voix n° 1	0 1 2 3	Voix n° 2	0 1 2 3	Voix n° 3	0 1 2 3
G		G		G	
R	✓	R		R	
B		B		B	
A	✓	A		A	
S	✓	S		S	
Voix n° 4	0 1 2 3	Voix n° 5	0 1 2 3	Voix n° 6	0 1 2 3
G		G		G	
R		R		R	
B		B		B	
A		A		A	
S		S		S	
Voix n° 7	0 1 2 3	Voix n° 8	0 1 2 3	Voix n° 9	0 1 2 3
G		G		G	
R		R		R	
B		B		B	
A		A		A	
S		S		S	
Voix n° 10		Voix n° 11		Voix n° 12	

Figure VI.10. : fiche de cotation fournie à chacun des membres du jury d'écoute pour une évaluation de groupe sur échelle ordinaire

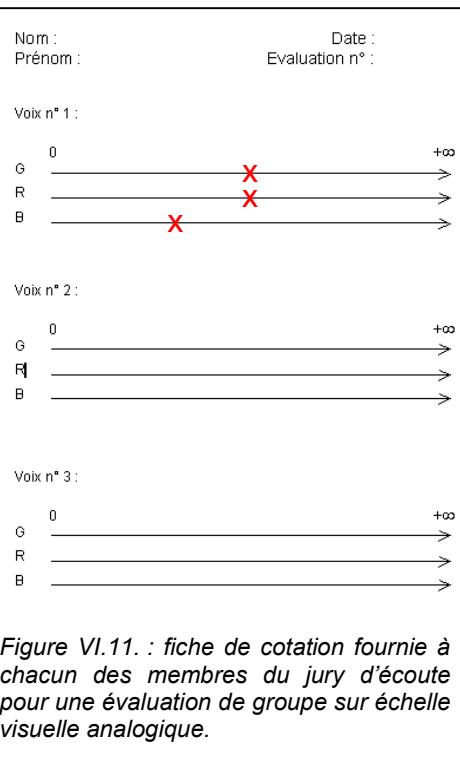


Figure VI.11. : fiche de cotation fournie à chacun des membres du jury d'écoute pour une évaluation de groupe sur échelle visuelle analogique.

Dans ce cas, les résultats sont recueillis à posteriori par l'examinateur dans un tableur (Excel ®) : soit en recopiant directement le grade attribué par l'auditeur avec l'échelle ordinaire, soit en relevant la valeur correspondant à la distance en cm entre le 0 et la marque sur l'échelle visuelle analogique. Ce principe d'évaluation par groupe permet un gain de temps considérable dans le cas d'un jury composé de nombreux auditeurs puisque cela ne nécessite la présence de l'examinateur qu'une fois par session. Néanmoins, la manipulation des données « à la main » par l'examinateur, augmente drastiquement le risque d'erreur lors du recueil des résultats et nécessite donc une attention et un soin accrus.

Enfin, dernier cas de figure, le responsable de recherche a pu avoir recours à une évaluation par consensus. Dans ce cas, un jury d'experts composé d'un nombre d'auditeurs impair était réuni dans la salle d'évaluation et chaque voix était discutée jusqu'à commun accord sur son évaluation qualitative et quantitative. L'examinateur se charge de diffuser les voix de la même façon que pour les évaluations en groupe et recueille lui-même les résultats après discussion sur une feuille de cotation « échelle ordinaire ». En effet, les évaluations par consensus ne peuvent être effectuées que d'après les échelles ordinaires. Ensuite, le recueil des données vers un tableur est fait « à la main », comme nous l'avons vu précédemment.

Pour conclure, les travaux que nous présentons ici sont basés sur l'application d'une procédure expérimentale telle que nous l'avons décrite ici. Les changements survenus selon les études dans le choix du jury d'écoute, du matériau phonétique ou du mode de passation seront spécifiés au fur et à mesure de la présentation des travaux. De même, les choix expérimentaux concernant les mesures instrumentales effectuées avec l'appareillage EVA® seront décrits dans la section perception / mesures.

2. Présentation des travaux personnels

Comme nous l'avons vu dans les chapitres précédents, la perception est un phénomène complexe qui intéresse des disciplines aussi diverses que la philosophie, la physiologie, la cognition ou encore la physique acoustique. Mais avant tout, la perception auditive reste le moyen immédiatement disponible afin de qualifier la voix et de quantifier les différents aspects de son timbre.

En pratique clinique, la perception est le seul outil « accessible » permettant de caractériser la dysphonie. En effet, les spécialistes de la dysphonie (ORL, Phoniatres, Orthophonistes) se trouvent face à une double nécessité d'évaluation : l'évaluation de première intention, lors du bilan vocal, qui va permettre de diagnostic et la conception du projet thérapeutique ; et l'évaluation de contrôle, en fin de prise en charge, qui va permettre de vérifier les résultats du traitement. Comme nous avons pu le voir dans les chapitres concernant la philosophie de la perception et les méthodologies de l'analyse perceptive, l'évaluation auditive reste soumise à des difficultés inhérentes à son principe même : la subjectivité et la variabilité. Ce mode d'appréciation demeure néanmoins le plus répandu en pratique clinique et l'analyse perceptive est actuellement toujours considérée comme la référence. Pourtant, les thérapeutes de la voix restent dans l'attente d'un outil « objectif » d'assistance à l'évaluation de la dysphonie, qui leur permettrait d'étayer leurs impressions cliniques en fournissant des mesures précises de la pathologie vocale et de son évolution.

Dans ce but, notre équipe a contribué à la réalisation d'un appareillage permettant la mesure simultanée de paramètres acoustiques (la voix) et de paramètres aérodynamiques (le souffle de la voix). Cet appareillage, réalisé en collaboration avec le Laboratoire CNRS Parole et Langage (appareillage EVA®), est actuellement le seul matériel de ce type ayant la norme CE et commercialisé en Europe.

Notre premier travail de recherche avait pour but de vérifier les capacités « cliniques » de l'appareillage EVA® par la confrontation des mesures obtenues chez 27 patients avant et après microchirurgie laryngée, avec l'évaluation du changement de qualité vocale par un jury d'experts. Cette étude a en réalité été menée en deux étapes. La première étape que nous présentons ici n'a pu être proposée pour une publication. En effet, les résultats obtenus par l'analyse perceptive montraient des variations si importantes, que la comparaison avec les données instrumentales n'a pu être établie. Néanmoins, ce travail nous a permis d'alimenter notre réflexion concernant l'analyse perceptive des dysphonies et les résultats ambigus que nous avons observés ont servi de base à toute notre démarche scientifique ultérieure. C'est pourquoi nous avons jugé nécessaire d'en rendre compte ici. Les biais

expérimentaux présents dans ce travail ont ensuite fait l'objet d'une exploration systématique qui nous a permis de reprendre cette étude, publiée dans une seconde étape.

a. Etude préliminaire de la pertinence des mesures acoustiques et aérodynamiques pour l'évaluation de l'évolution de la voix après microchirurgie laryngée : à propos de 27 cas.

Giovanni A, Revis J, Triglia JM.

(Non publié)

Essentiellement deux modèles physiopathologiques sous-tendent les indications cliniques de microchirurgie laryngée. Le premier modèle concerne les pathologies de type nodules ou kystes, qui induisent une asymétrie des cordes vocales à l'origine d'un comportement vibratoire irrégulier. Le second modèle concerne les pathologies de type paralysie laryngée unilatérale qui induisent un mauvais accrolement glottique à l'origine d'une fuite d'air. Traditionnellement, on considère que l'irrégularité vibratoire produit une voix « rauque », et que la fuite d'air produit une voix « soufflée ». Le but de la microchirurgie laryngée est de restaurer une anatomie normale. Ce faisant, le comportement vibratoire des cordes vocales est rétabli, et la voix rauque ou soufflée avant l'intervention tend à redevenir normale.

Ces dernières années, de nombreuses études ont décrit l'évaluation « objective » de la voix par l'intermédiaire d'appareillages de mesure [Hillenbrand, 1994 ; Giovanni, 1996 ; Titze, 1995 ; Wolfe, 1995], mais peu se sont attachées à comparer les résultats avant et après microchirurgie laryngée [Bouchayer, 1992]. Les indices acoustiques de stabilité vibratoire, comme le jitter ou le shimmer, peuvent être mesurés par de nombreux outils informatiques ou logiciels. Ces algorithmes sont fréquemment utilisés pour l'évaluation de la voix, malgré l'importance des variations inter- et intra-individuelles de leurs mesures [Bielałowicz, 1996 ; Schoentgen, 1989]. L'évaluation de paramètres aérodynamiques est moins courante. La pertinence clinique de la mesure du débit d'air buccal (DAB) fournie par l'appareillage EVA® a été démontrée [Giovanni, 1996]. En effet, l'appareillage EVA® permet de fournir simultanément des mesures acoustiques (représentatives de la régularité vibratoire) et des mesures aérodynamiques (représentatives de la fuite glottique). Le protocole clinique de cet appareillage est conçu, précisément, dans le but d'être représentatif des deux modèles physiopathologiques à l'origine des dysphonies. Le but de cette étude est de démontrer la validité d'un protocole instrumental associant les mesures acoustiques et aérodynamiques pour l'évaluation de l'évolution de la voix après microchirurgie laryngée.

Protocole expérimental : 27 patients consécutifs, candidats à la microchirurgie laryngée, ont été étudiés. Il s'agissait de 17 femmes, 7 hommes et 3 enfants. Les indications chirurgicales étaient

nodules dans 9 cas, polype dans 11 cas, oedème de Reinke dans 5 cas et kyste épidermique dans 2 cas. Les enregistrements de voix étaient effectués la veille de l'intervention, puis à nouveau 3 mois après, dans le cadre de la surveillance médicale, afin de respecter la période de cicatrisation. Les patients n'ont pas été sélectionnés en fonction de la qualité du résultat. Les examens et enregistrements étaient effectués comme décrit dans la section « Protocole Expérimental ». Le matériau phonétique enregistré était le premier paragraphe de « *La petite chèvre de Monsieur Seguin* », d'Alphonse Daudet.

Dans le même temps, les enregistrements numériques destinés à l'analyse instrumentale étaient réalisés avec l'appareillage EVA® qui permet la mesure simultanée de paramètres acoustiques et aérodynamiques. Un masque en silicone, contenant un microphone calibré et une grille de pneumotachographe était placé sur la bouche du patient et l'examineur vérifiait l'étanchéité du système. Le patient devait prononcer la voyelle /a/ pendant plusieurs secondes, à hauteur et intensité normales. Plusieurs essais étaient réalisés jusqu'à ce que l'examineur estime que la production était représentative de la voix usuelle du sujet. La durée totale de l'enregistrement ne dépassait jamais 10 mn. Le traitement du signal était réalisé grâce au logiciel Phonedit® (SQ-Lab, Marseille). L'analyse était effectuée sur un segment de 2 secondes (fenêtre d'analyse) localisé 100 ms après le début du signal afin d'éliminer l'instabilité liée à l'initiation de la vibration. Le débit d'air buccal (DAB) était mesuré en dm^3/s . Le jitter et le shimmer étaient calculés en pourcentage d'après les valeurs moyennes de Fo et d'intensité, respectivement. Les valeurs sont rapportées selon la formule : (valeurs post-opératoires – valeurs pré-opératoires) / valeurs pré-opératoires.

Les enregistrements de parole (texte lu) ont été évalués par un jury d'écoute constitué de 8 auditeurs : 2 médecins phoniatres, 2 orthophonistes expérimentées et 6 étudiantes en dernière année d'orthophonie. Les enregistrements pré-opératoires et post-opératoires de chaque patient étaient présentés appariés deux à deux et dans l'ordre chronologique. La tâche du jury était d'évaluer la quantité de changement des paramètres GRBAS (G : grade global, R : raucité, B : souffle, A : asthénie, S : forçage) au moyen d'une échelle visuelle analogique conçue spécialement pour l'étude afin de prendre en compte l'éventualité d'une détérioration de la voix (figure VI.a.1.).

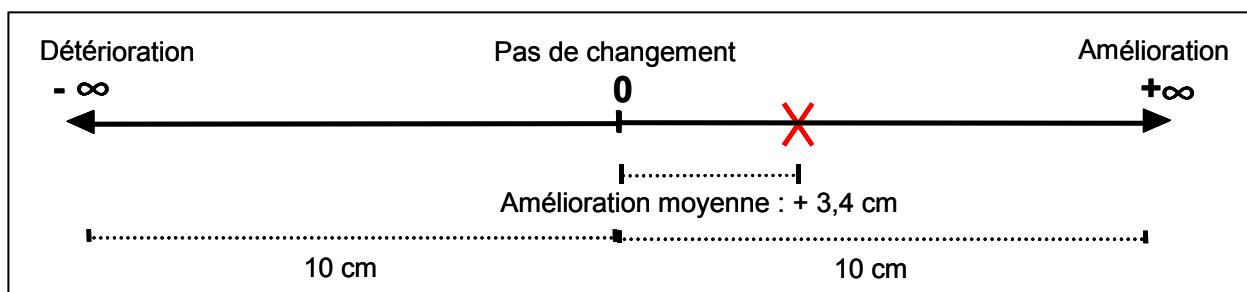


Figure VI.a.1. : échelle visuelle analogique. La quantité de changement de la voix est appréciée par la distance mesurée en cm entre le 0, au centre de la droite (aucun changement), et la croix portée par l'auditeur. A droite du 0, le jury quantifie l'amélioration, à gauche la détérioration. Dans l'exemple ci-dessus, la voix montre une amélioration moyenne, avec une valeur d'évolution de la voix après chirurgie de +3,4.

Les valeurs d'évolution objective obtenues par l'équation ([valeur post-opératoire – valeur pré-opératoire] / valeur pré-opératoire) ont été comparées à la quantité perceptive d'évolution de la voix déterminée par le jury grâce au test statistique non-paramétrique de Mann-Whitney.

Résultats et discussion : Avant de procéder à l'analyse statistique de l'ensemble des données recueillies, nous avons procédé à une observation systématique et minutieuse des résultats de l'analyse perceptive (figure VI.a.2.).

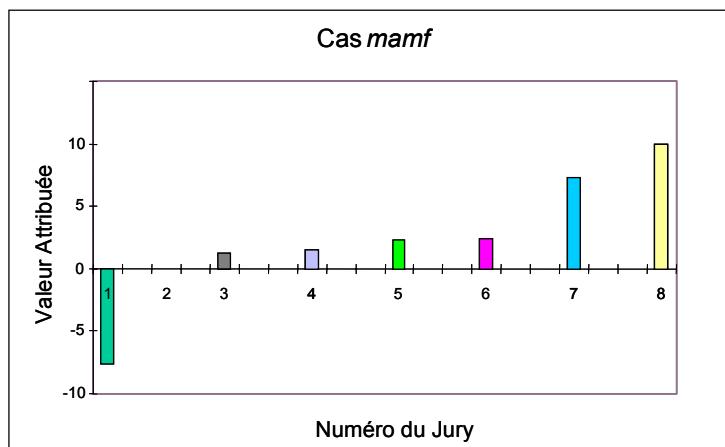


Figure VI.a.2. : exemple de résultats obtenus pour le patient « mamf » par l'évaluation perceptive. Ce tableau montre l'évaluation de l'évolution de la voix d'un patient par le jury d'écoute. L'axe horizontal désigne chaque auditeur membre du jury d'écoute d'après son numéro d'identification. L'axe vertical représente la valeur attribuée à la quantité de changement de la voix après microchirurgie laryngée. Au dessus de 0, les valeurs positives montrent l'amélioration de la voix, en dessous de 0, les valeurs négatives montrent la détérioration.

Sur la figure VI.a.2., les 8 auditeurs constitutifs de notre jury d'écoute ont évalué la quantité d'évolution de la voix du patient « mamf » après l'intervention chirurgicale. Pour l'auditeur n°1, la voix est nettement détériorée après le traitement, puisqu'elle obtient un score négatif de -7,5. Pour l'auditeur n°2, la voix reste inchangée puisque le score est de 0. Pour les auditeurs n°s 7 et 8, la voix est largement améliorée par la chirurgie, puisqu'elle obtient des valeurs positives de +7,6 et +10, respectivement. Enfin, pour les auditeurs n°s 3 à 6, il existe une amélioration modérée près l'intervention (+2,8 en moyenne). On observe ainsi un désaccord flagrant entre les auditeurs concernant l'évaluation de cette voix. Ce schéma, bien que parfois plus modéré, était retrouvé pour de nombreux patients et nous a empêché de poursuivre l'étude par la comparaison des données perceptives et instrumentales initialement prévue.

Néanmoins, ces résultats ont alimenté notre réflexion et l'interprétation de ce graphique a ouvert des pistes expérimentales. Nous avons relevé deux biais dans notre protocole d'étude, susceptibles d'avoir augmenté la variabilité du jury d'écoute.

Le premier biais que nous avons relevé concerne l'utilisation de l'échelle visuelle analogique bipolaire. Il semble que la consigne d'utilisation de l'échelle ait été comprise de manière différente par nos auditeurs. Habituellement, l'échelle visuelle analogique se présente sous la forme suivante (figure VI.a.3.).

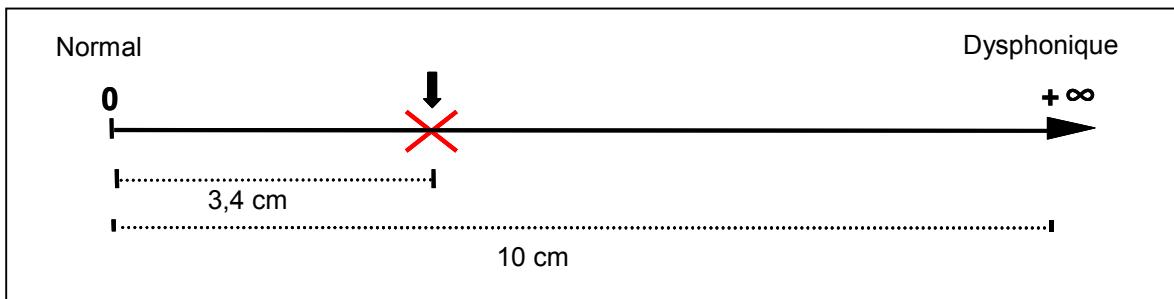


Figure VI.a.3. : Echelle visuelle analogique « classique »

Le principe de l'échelle visuelle analogique est de quantifier la dysphonie ou la présence d'un critère dysphonique. Cette échelle permet donc une mesure de la qualité de la voix. Dans notre travail, l'échelle visuelle analogique mise en place ne permettait pas d'évaluer la voix elle-même, mais de mesurer une quantité d'évolution. Ainsi, dans l'échelle visuelle analogique classique, l'écart de la croix à la droite du zéro mesure une quantité de pathologie, il s'agit donc d'une quantité péjorative (figure VI.a.3.). A l'inverse, avec l'échelle de mesure mise en place pour notre étude, l'écart de la croix à la droite du zéro mesure la quantité d'amélioration, il s'agit donc d'une quantité laudative (figure VI.a.1.). Habitué à manipuler les échelles classiques, notre jury d'écoute était perturbé par le fait que cette fois-ci, plus la croix était portée à la droite du zéro, plus l'évaluation était positive. Certains auditeurs ont d'ailleurs fait la remarque que la mesure était moins instinctive et demandait une réflexion consciente systématique. Notre jury, habitué à l'évaluation de la dysphonie en soi, était donc embarrassé face à l'évaluation d'une quantité de changement. Il semble donc que ces tâches d'évaluation soient franchement différentes. Par ailleurs, le fait que notre échelle soit bipolaire, intégrant la possibilité d'une détérioration de la voix après chirurgie, ait introduit un biais supplémentaire : nos auditeurs, habitués à travailler avec le chirurgien ayant opéré nos sujets d'étude, se sentaient implicitement en situation de « noter » son travail, et n'ont utilisé la partie gauche de l'échelle que très exceptionnellement, et pour certains, pas du tout. Souvent, les voix détériorées se trouvaient évaluées entre 0 et +1, ce qui correspondait à « aucun changement » ou « amélioration pauvre ». Au total, il semble que la manipulation de cette échelle originale ait été un facteur de perturbation du jury d'écoute et que les évaluations fournies soient biaisées. Une solution possible serait d'évaluer la quantité de dysphonie de chaque voix séparément et de laisser à l'expérimentateur la tâche de faire la différence entre les évaluations portées sur les voix avant traitement, puis après microchirurgie laryngée.

La deuxième remarque renvoie au recrutement des auditeurs. En effet, considérant le volume de l'effectif comme une garantie de solidité des évaluations, nous avons recruté, en plus de nos 4 experts habituels (2 médecins phoniatres, 2 orthophonistes du service spécialistes de la voix), 4 étudiantes en dernière année d'orthophonie. Nous considérons que les étudiantes, sensibilisées à la dysphonie au cours de leurs études, seraient compétentes en tant que jury expérimental. En réalité, il semble que leurs évaluation aient été largement discordantes. En effet, dans le graphique 2, les auditeurs 1 et 2 d'une part, et 7 et 8 d'autre part, montrent des résultats très hétérogènes, allant de grave détérioration (auditeur n°1) à complète amélioration (auditeurs n°s 7 et 8), en passant par aucune évolution (auditeur n°2). La vérification à posteriori nous a permis de constater que ces auditeurs correspondaient justement aux 4 étudiantes recrutées. En revanche, les évaluations réalisées par nos 4 auditeurs experts (3, 4, 5 et 6) étaient tout à fait homogènes. Il semble donc que l'expérience du jury d'écoute soit une source d'influence pour la variabilité des évaluations et qu'il soit préférable de conserver un jury plus réduit mais plus expérimenté.

Ces différentes réflexions nous ont conduit à recommencer ce travail en tenant compte des biais relevés dans le protocole expérimental. Pour cela, nous avons conservé les sujets recrutés et les enregistrements réalisés, le protocole de mesures instrumental est également demeuré identique. En revanche, toute la partie perceptive a été revue. C'est le travail que nous présentons maintenant.

b. Mesures objectives acoustiques et aérodynamiques de l'évolution de la voix après microchirurgie laryngée.¹

Giovanni A, Revis J, Triglia JM.
(Laryngoscope, 1999)

Nous reprenons ici ce travail par la présentation des changements survenus dans le protocole expérimental. Afin d'éviter les redondances, nous ne présenterons pas à nouveau l'introduction, ni le choix des sujets de l'étude ou le recueil des mesures instrumentales.

Protocole expérimental modifié : Les échantillons de parole ont été évalués par un jury expérimenté constitué de 2 médecins phoniatres et 2 orthophonistes du service spécialistes de la dysphonie. Ces 4 auditeurs ont été recruté sur la base de leurs années d'expérience, mais aussi sur leur habitude de travailler ensemble, dans la même équipe. Les enregistrements de voix étaient analysés selon les paramètres G (grade global), R (raucité), B (souffle) de l'échelle GRBAS, sur une échelle ordinaire à 4 niveaux (0 = normal ; 3 = anormalité sévère). Les voix pré-opératoire et post-opératoires étaient évaluées dans l'ordre chronologique, selon la méthode du consensus : les auditeurs discutaient de leurs impressions et réécoutaient la voix jusqu'à commun accord. Théoriquement, la méthode de consensus implique la suppression des cas pour lesquels le jury ne parvient pas à se mettre d'accord, mais ça n'a pas été le cas ici. Un score perceptif était ensuite calculé par l'expérimentateur comme la différence entre les évaluation pré- et post-opératoires. Si le paramètre considéré (G, R ou B) était amélioré d'1 point ou plus, le patient était considéré comme « Bon Résultat Perceptif » (BRP). Les patients présentant moins de 1 point d'amélioration étaient considérés comme « Faible Résultats Perceptif » (FRP).

Les valeurs d'évolution objective obtenues par l'équation ([valeur post-opératoire – valeur pré-opératoire] / valeur pré-opératoire) on été comparées à la quantité perceptive d'évolution de la voix déterminée par le jury grâce au test statistique non-paramétrique de Mann-Whitney.

Résultats : Les données recueillies sont présentées dans le tableau VI.b.1.

¹ Giovanni A, Revis J, Triglia JM. « Objective aerodynamic and acoustic measurement of voice improvement after phonosurgery. » Laryngoscope. 1999 Apr;109(4):656-60.

#	G1	R1	B1	G2	R2	B2	DAB 1	DAB 2	DIF DAB	JITTER1	JITTER2	DIF JIT	SHIM1	SHIM2	DIF SHIM
1	1	0	1	0	0	0	0,30	0,19	-0,37	0,59	0,43	-0,27	0,62	0,61	-0,01
2	2	1	2	1	0	1	0,47	0,24	-0,50	0,50	0,33	-0,34	0,52	0,42	-0,05
3	3	3	3	3	2	3	0,43	0,39	-0,10	20,00	2,80	-0,86	2,00	0,92	-0,54
4	2	2	1	2	1	1	0,19	0,25	0,29	0,48	0,56	0,17	1,00	0,52	-0,24
5	2	2	2	0	0	0	0,13	0,15	0,14	1,00	0,33	-0,67	0,33	0,78	0,23
6	2	2	2	2	2	2	0,12	0,09	-0,23	0,46	0,47	0,02	0,22	0,45	0,12
7	3	3	3	3	3	3	0,22	0,31	0,42	26,00	0,94	-0,96	1,90	0,65	-0,63
8	2	2	2	1	0	1	0,30	0,20	-0,34	0,40	0,40	0,00	1,20	0,57	-0,32
9	2	0	2	2	1	1	0,22	0,12	-0,45	0,70	0,38	-0,46	0,78	0,47	-0,16
10	1	1	1	1	0	1	0,15	0,20	0,32	0,56	0,44	-0,21	1,00	0,80	-0,10
11	2	1	2	2	1	2	0,13	0,08	-0,33	0,61	0,84	0,38	0,40	0,35	-0,03
12	2	1	2	2	2	1	0,23	0,10	-0,56	0,55	2,70	3,91	0,50	0,40	-0,05
13	2	1	2	1	1	1	0,17	0,12	-0,28	0,76	0,48	-0,37	0,29	0,37	0,04
14	1	0	1	0	0	0	0,16	0,07	-0,58	0,38	0,51	0,34	0,59	0,35	-0,12
15	2	2	2	2	2	2	0,08	0,18	1,12	3,70	0,90	-0,76	0,61	0,35	-0,13
16	2	2	2	0	0	0	0,19	0,09	-0,52	0,54	0,37	-0,31	0,82	0,56	-0,13
17	2	2	1	2	1	1	0,26	0,23	-0,13	0,48	0,41	-0,15	0,71	0,55	-0,08
18	2	1	3	1	1	1	0,33	0,08	-0,77	0,61	0,64	0,05	0,33	0,37	0,02
19	2	1	2	2	1	2	0,11	0,17	0,50	0,48	0,94	0,96	0,46	0,23	-0,12
20	2	2	1	2	2	2	0,10	0,21	1,16	0,35	0,48	0,37	0,88	0,58	-0,15
21	2	1	2	1	1	1	0,14	0,10	-0,31	0,62	0,35	-0,44	0,62	0,06	-0,28
22	1	0	1	0	0	0	0,09	0,10	0,12	0,30	0,42	0,40	0,77	0,12	-0,33
23	3	3	3	2	1	1	0,19	0,13	-0,29	0,93	0,55	-0,41	0,84	0,07	-0,39
24	3	3	3	3	3	2	0,15	0,15	-0,01	0,43	0,59	0,37	0,83	0,72	-0,06
25	2	2	2	1	1	1	0,20	0,20	0,01	0,51	0,47	-0,08	0,61	0,40	-0,11
26	2	1	2	2	1	2	0,24	0,10	-0,59	0,54	0,46	-0,15	0,31	0,38	0,04
27	3	3	3	0	0	0	0,31	0,12	-0,62	2,10	0,39	-0,81	1,20	0,11	-0,55

Tableau VI.b.1. : résultats recueillis en pré-opératoire (*1) puis en post-opératoire (*2) pour chacun des sujets de l'étude (#). Calcul du score d'évolution pour les données instrumentales (DIF*).

Les moyennes et écarts-types sont présentés dans les tableaux VI.b.2, 3 et 4 pour chacun des paramètres G,R,B, ainsi que la comparaison entre le résultat perceptif et le score objectif.

Grade G	DIF DAB (%)	DIF Jitter (%)	DIF Shimmer (%)
BRP n=13	-33 [28]	-22 [35]	-27 [63]
FRP n=14	10 [50]	19 [120]	-21 [42]
P (Mann-Whitney test)	0.033	0.357 (NS)	0.627 (NS)

Tableau VI.b.2. : moyenne et [écart-type] des scores objectifs et comparaison avec le grade perceptif G (test non-paramétrique de Mann-Whitney)

Grade R	DIF DAB (%)	DIF Jitter (%)	DIF Shimmer (%)
BRP n=11	-15 [32]	-33 [33]	-29 [60]
FRP n=16	-7 [59]	21 [110]	-21 [48]
P (Mann-Whitney test)	0.921 (NS)	0.289 (NS)	0.093 (NS)

Tableau VI.b.3. : moyenne et [écart-type] des scores objectifs et comparaison avec le grade perceptif R (test non-paramétrique de Mann-Whitney)

Grade B	DIF DAB (%)	DIF Jitter (%)	DIF Shimmer (%)
BRP n=11	-33 [27]	-6 [108]	-27 [57]
FRP n=16	-22 [57]	-11 [58]	-20 [48]
P (Mann-Whitney test)	0.07	0.805 (NS)	0.902 (NS)

Tableau VI.b.4. : moyenne et [écart-type] des scores objectifs et comparaison avec le grade perceptif B (test non-paramétrique de Mann-Whitney)

Les tableaux VI.b.2. et 4. montrent que l'amélioration des valeurs de débit d'air buccal (DAB) en post-opératoire était significativement plus importante pour le groupe de Bon Résultat Perceptif (BRP) en ce qui concerne les paramètres de grade global (G) et de souffle (B). La moyenne de diminution du DAB était de 0.079 dm³ (33%) chez les patients BRP pour le paramètre G, alors que l'on n'observe pas de changement chez les patients FRP (Faible Résultat Perceptif), pour le paramètre G. Inversement, du fait de la grande dispersion des valeurs mesurées, il n'existe pas de différence significative entre les groupes BRP et FRP concernant les mesures acoustiques (Jitter et Shimmer) pour le grade G.

La matrice de corrélation présentée dans le tableau VI.b.5. montre qu'il n'existe pas de redondance entre les différentes mesures objectives mises en place dans cette étude.

	DIF OAF	DIF Jitter	DIF Shimmer
DIF OAF	1	-	-
DIF Jitter	-0.112	1	-
DIF Shimmer	-0.082	0.020	1

Tableau VI.b.5. : matrice de corrélation entre les différentes mesures objectives.

Les résultats de l'analyse discriminante sont présentés dans les tableaux VI.b.6., 7. et 8.

Grade G	BRO	FRO	% correct
BRP	10	3	77
FRP	5	9	64
	-	-	70

Wilk's lambda p : 0.047

Canonical discriminant function

OAF : 2.134

Jitter : 0.677

Shimmer : 0.407

Tableau VI.b.6. : résultats de l'analyse discriminante pour le grade G entre les différents groupes déterminés par la qualité des résultats perceptifs (BRP = bon résultat perceptif ; FRP = faible résultat perceptif) et objectifs (BRO = bon résultat objectif ; FRO = faible résultat objectif).

Grade R	GOR	POR	% correct
GPR	7	4	64
PPR	9	7	44
	-	-	52

Wilk's lambda p : 0.413

Canonical discriminant function

OAF : 0.809

Jitter : 1.069

Shimmer : 0.545

Tableau VI.b.7. : résultats de l'analyse discriminante pour le grade R entre les différents groupes déterminés par la qualité des résultats perceptifs (BRP = bon résultat perceptif ; FRP = faible résultat perceptif) et objectifs (BRO = bon résultat objectif ; FRO = faible résultat objectif).

Grade B	GOR	POR	% correct
GPR	12	4	75
PPR	4	7	64
	-	-	70

Wilk's lambda p : 0.024

Canonical discriminant function

OAF : 2.402

Jitter : -0.070

Shimmer : 0.449

Tableau VI.b.8. : résultats de l'analyse discriminante pour le grade B entre les différents groupes déterminés par la qualité des résultats perceptifs (BRP = bon résultat perceptif ; FRP = faible résultat perceptif) et objectifs (BRO = bon résultat objectif ; FRO = faible résultat objectif).

Utilisant les 3 mesures objectives ensemble, il était possible de prédire les résultats perceptifs de G et de B avec une précision de 70% ($p<0,05$).

Discussion : Cette étude confirme l'intérêt de la mesure de débit d'air buccal, davantage que les mesures acoustiques, pour l'évaluation de l'amélioration de la voix après microchirurgie laryngée.

L'analyse perceptive de la parole par un jury d'écoute est la référence contre laquelle les mesures objectives sont testées et éventuellement validées [Kreiman, 1993 ; Hammarberg, 1980]. Or, pour des raisons techniques inhérentes aux appareillages et aux algorithmes utilisés pour le calcul des indices, les mesures objectives doivent être réalisées sur des matériaux phonétiques stables (les voyelles tenues) et pas sur de la parole. Cela soulève la question suivante : l'évaluation d'une seule voyelle tenue est elle comparable à l'analyse perceptive de la parole ? De précédentes études perceptives portant sur des voyelles isolées ou de la parole n'ont pas montré de différences [De Krom, 1994]. La validité de cette approche est également étayée par l'analogie avec l'audiométrie qui utilise des sons purs (absents dans la nature) pour déterminer les capacités auditives.

La fiabilité et la reproductibilité de la méthode dépendent de l'expérience des auditeurs et de l'homogénéité de leurs références pour l'analyse perceptive de la qualité de la voix [Kreiman, 1993 ; De Bodt, 1997 ; Gerratt, 1993]. Plusieurs échelles quantitatives ont été développées dans un effort de standardisation des résultats de l'analyse perceptive, afin de réduire la variabilité inter- et intra-individuelle [Kreiman, 1993 ; Hammarberg, 1980 ; De Bodt, 1996 ; Hirano, 1981]. L'échelle la plus fréquemment mise en œuvre est la méthode GRBAS [De Bodt, 1997 ; Hirano, 1981], et notre équipe possède une grande expérience dans son utilisation. Dans cette étude, nous avons fait le choix de n'évaluer que les critères G, R et B dont l'application est plus intuitive et naturelle, donc moins sujette à la subjectivité que les critères A et S [De Krom, 1994 ; De Bodt, 1997 ; Dejonckere, 1993]. A ce sujet, Hammarberg relève que les auditeurs pourraient avoir des difficultés à séparer les différents critères d'une échelle [Hammarberg, 1980]. Pourtant, pour un jury expérimenté, composé d'auditeurs habitués à travailler ensemble, ce risque semble plus hypothétique que réel.

La « performance » relativement pauvre des mesures acoustiques de la stabilité vibratoire pourrait être reliée à la forte dispersion inter-individuelle des valeurs observées dans notre étude comme dans la plupart des travaux publiés. Cette dispersion est à l'origine d'un recouplement significatif entre les groupes BRP et FRP. Nous supposons que ces paramètres acoustiques seraient mieux corrélés avec le critère R, ce qui n'est pas le cas dans nos résultats. Pour l'instant, le jitter ou le shimmer semblent insuffisants pour prédire à eux seuls le résultat de la microchirurgie laryngée dans la pratique clinique quotidienne. Cette observation confirme que la perception de la raucité ne se résume pas au jitter et au shimmer.

L'utilisation du débit d'air buccal montre des résultats plus intéressants. La mesure du DAB sur une voyelle tenue était significativement différente pour les groupes BRP et FRP selon les critères G et B évalués par le jury d'écoute. Il est nécessaire de préciser ici que le critère de souffle (B) est habituellement relié davantage à la turbulence qu'à la fuite [Hillenbrand, 1994]. Néanmoins, dans notre étude, la diminution de la fuite glottique, chez les patients présentant une dysphonie pré-opératoire importante, apparaissait comme un indicateur efficace de bon résultat perceptif pour les critères G et B. Ainsi, il semble que la diminution post-opératoire du DAB chez les patients présentant des valeurs élevées avant l'intervention, soit un bon critère objectif pour l'évaluation de l'amélioration de la voix après microchirurgie laryngée.

Les résultats de l'analyse factorielle discriminante ont montré une relativement bonne prédiction du jugement perceptif pour les critères G et B (70% ; $p<0,05$). Dans ces cas la fonction canonique discriminante des mesures de DAB était plus importante que les mesures acoustiques (tableaux 6, 7, 8), soulignant l'importance du DAB comme indicateur du souffle et du grade global de dysphonie.

Conclusion : Notre protocole d'évaluation objective de la voix est extrêmement rapide et simple. La mesure du DAB sur une voyelle tenue est un paramètre fiable pour l'évaluation clinique quotidienne de l'amélioration de la voix après microchirurgie laryngée et peut être utilisée en association avec les mesures acoustiques plus conventionnelles comme le jitter et le shimmer.

A l'issue de cette étude, de nombreux questionnements concernant l'analyse perceptive ont alimenté notre réflexion selon 3 grands axes :

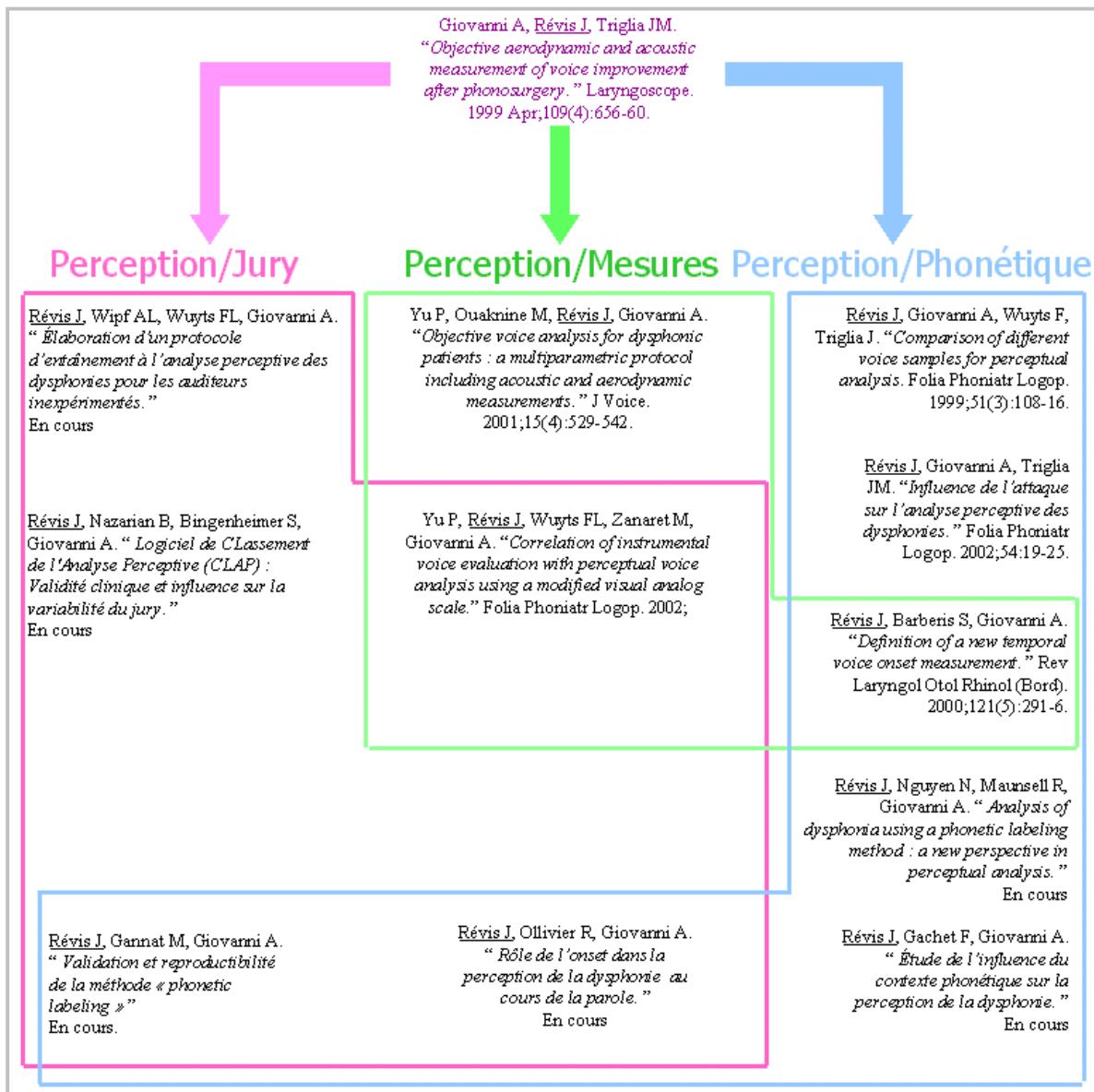
Axe Perception / Jury: L'expérience des auditeurs semble être un facteur d'influence pour la fiabilité des mesures perceptives. La formation des auditeurs permet-elle d'améliorer les performances du jury d'écoute ? Comment optimiser les protocoles d'évaluation perceptive afin de limiter la variabilité des auditeurs ? Dans quelle mesure les échelles qualitatives et quantitatives influencent-elles la sensibilité du jury ? Quelles sont les autres sources d'influence potentielles ?

Axe Perception / Mesures: L'analyse perceptive est considérée comme la référence contre laquelle les mesures instrumentales, dites « objectives », sont testées. En effet, les mesures instrumentales n'ont de valeur que si elles reflètent la perception que le patient a de sa propre voix. Or, l'analyse perceptive, dite « subjective », est soumise à la variabilité. Comment améliorer les protocoles d'analyse perceptive dans le but d'obtenir une meilleure adéquation

avec les mesures instrumentales ? Les échelles utilisées actuellement sont-elles adaptées pour la comparaison avec les mesures objectives ? Les paramètres qualitatifs et les échelles quantitatives peuvent-ils être optimisés dans le but de favoriser cette comparaison ?

Axe Perception / Phonétique : A l'heure actuelle, pour des raisons techniques, les auteurs comparent les mesures instrumentales réalisées sur une voyelle tenue avec l'analyse perceptive effectuée sur de la parole. Cette dichotomie est justifiée par le fait que ce procédé permet de mettre les deux systèmes de mesure à leur meilleur niveau de performance. Néanmoins, on peut se poser la question de savoir si cette comparaison est valide, dans la mesure où les dispositifs instrumentaux et perceptifs n' « écoutent » pas la même chose. Quelles sont les différences fondamentales entre les différents matériaux phonétiques ? Comment influencent-ils la perception du jury d'écoute ?

Ces trois grands axes de réflexion nous ont permis de développer un ensemble d'études cohérent allant dans le sens de la conceptualisation de l'analyse perceptive et de la recherche de méthodes d'investigations plus pertinentes. Ces différents travaux nous ont conduit à aborder l'analyse perceptive selon une vision nouvelle et originale pour aboutir à une conception phonétique du fonctionnement perceptif. Notre question de départ était : quel modèle conceptuel régit la perception de la dysphonie et comment optimiser son application clinique ? Chemin faisant, notre interrogation s'est orienté de la façon suivante : quels sont les indices que nous entendons et qui nous font percevoir une voix comme dysphonique ? Notre démarche, qui s'est voulue relativement exhaustive nous a conduit vers l'exploration des 3 axes présentés ci-dessus. L'ensemble de ces travaux est présenté dans l'organigramme suivant afin de faciliter la compréhension de la démarche globale. Certaines études développées dans notre approche entrent toutefois dans le cadre de plusieurs axes, c'est pourquoi nous avons mis en place un code couleur permettant de situer les articles dans le concept plus général.



Nous avons choisi de présenter nos études en fonction des thématiques abordées dans notre démarche. A la présentation des titres, un cadre coloré rappellera l'axe ou les axes de recherche à l'origine de l'étude. Le premier ensemble de travaux que nous présentons regroupe les études consacrées aux caractéristiques du jury d'écoute. C'est l'axe **Perception / Jury**.

L'article suivant est une étude expérimentale prospective conçue dans le but d'améliorer la fiabilité d'un jury d'écoute novice par une formation à l'analyse perceptive.

c. Elaboration d'un protocole d'entraînement à l'analyse perceptive pour un jury inexpérimenté.

Revis J, Wipf AL, Wuyts FL, Giovanni A.

(Rev Otol Rhinol Laryngol, soumis à publication)

Lors de notre première étude, nous avons été confrontés à des cas de discordance majeure entre nos différents auditeurs, concernant l'évaluation de l'amélioration de la voix d'un patient avant et après microchirurgie laryngée. Entre autres sources de variabilité, nous avons observé l'effet probable du manque d'homogénéité de notre jury d'écoute. En effet, nous estimions que la solidité de nos évaluations serait renforcée par l'augmentation de l'effectif du jury, et pour cela, nous avions recruté 4 étudiantes en dernière année d'orthophonie, en plus de nos 4 auditeurs expérimentés. Or, l'observation des résultats nous a révélé que les plus forts niveaux de discordance étaient produits par les auditeurs inexpérimentés, alors que les jugements des experts semblaient davantage cohérents. Cette réflexion s'est avérée conforme aux données de la littérature qui indiquent que l'expérience est un facteur primordial pour la fiabilité des jugements perceptifs (Section V, Chapitre 2.1., p 84).

Nous nous sommes donc posé la question suivante : est-il possible d'améliorer les compétences d'évaluation de la dysphonie d'un jury « naïf » par une formation à l'analyse perceptive ? La mise en place de cette étude ouvre des perspectives majeures dans les conceptions méthodologiques de l'analyse perceptive. Cela pourrait permettre 1- d'améliorer la fiabilité des jugements perceptifs, 2- de proposer un enseignement spécifique aux étudiants en orthophonie, 3- de standardiser les compétences du jury d'écoute et 4- d'en augmenter les effectifs.

L'objectif de notre travail est de proposer un protocole de formation à l'analyse perceptive, destiné aux étudiant de 2^{ème} année d'orthophonie. Ce travail se déroulera en trois étapes :

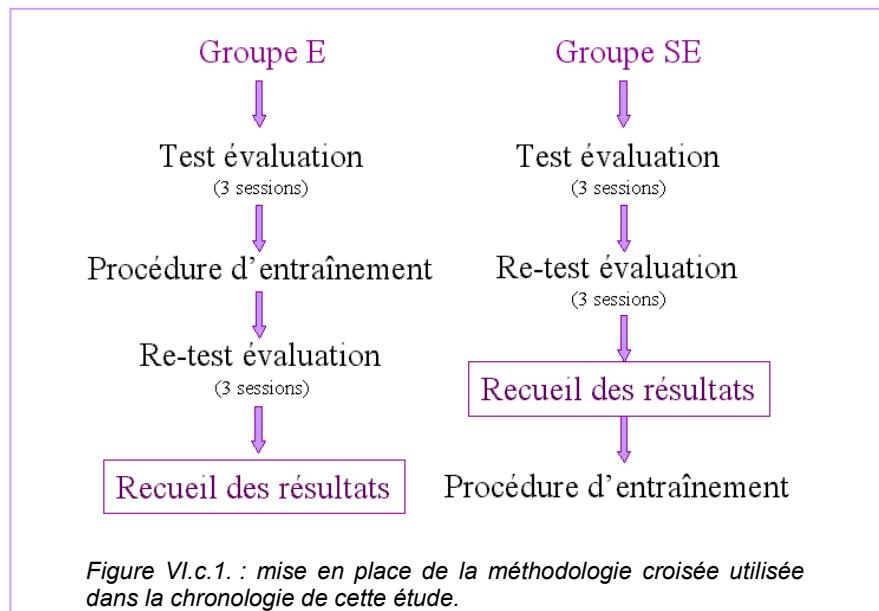
- évaluation du corpus par les étudiants afin de définir leurs compétences avant formation (3 sessions)
- formation individuelle des étudiants selon le paradigme : stimulus / réponse / correction par expertise / feedback proposé par Goldstone [Goldstone, 1998] et repris par Chan [Chan, 2002]
- évaluation du corpus par les étudiants afin de vérifier l'amélioration de leur compétence après formation (3 sessions).

Notre hypothèse est que les variabilités intra- et inter-individuelles des étudiants avant formation seront bien supérieures à celles du jury d'expert, et que leurs valeurs se rapprocheront de celles de l'expertise après formation.

Protocole expérimental : 78 voix ont été sélectionnées dans la banque de données de notre laboratoire. Ces voix avaient déjà été expertisées par notre jury d'écoute selon la méthode GRBAS, sur une échelle ordinaire à 4 niveaux. Le jury d'expert était composé de 7 auditeurs : 4 phoniatres et 3 orthophonistes travaillant ensemble dans le même service et membres permanents du jury d'écoute de voix dysphoniques depuis plus de 5 ans. Leurs évaluations de GRB ont été conservées afin de proposer aux auditeurs naïfs la « correction par expertise » lors de la procédure de formation. Les 78 voix ont été choisies en fonction de leur niveau de dysphonie. Nous avons retenu 22 grades 0 (voix normales) ; 22 grades 1 (dysphonies légères) ; 19 grades 2 (dysphonies moyennes), et 15 grades 3 (dysphonies sévères). Les voix grade zéro correspondaient à des sujets sains sans antécédents de dysphonie ; les voix pathologiques correspondaient à des dysphonies dysfonctionnelles. Parmi ces voix, 50 étaient utilisées pour les évaluations (16 G0 ; 13 G1 ; 12 G2 ; 9 G3) et 28 pour la formation (6 G0 ; 9 G1 ; 7 G2 et 6 G3). Les voix avaient été enregistrées puis numérisées selon la méthodologie habituelle. Le matériau phonétique était constitué de la phrase « *il les perdait toutes de la même façon* » extraite de « *La petite chèvre de Monsieur Seguin* » d'Alphonse Daudet.

Notre jury inexpérimenté était constitué de 18 étudiants en 2^{ème} année d'orthophonie, tous volontaires. A cette période de l'année, les cours de pathologie de la voix avaient déjà été dispensés. Les étudiants avaient donc connaissance des lésions responsables des dysphonies dysfonctionnelles, mais n'avaient encore jamais été confrontés à l'écoute massive de voix pathologiques. De ce fait, nous pouvons les considérer comme candides. Afin de mesurer l'évolution de leurs compétences, estimées d'après leurs variabilité intra- et inter-individuelle, nous leur avons fait évaluer les 50 voix de test lors de 3 sessions d'écoute avant la formation. A chaque session, les voix étaient présentées dans un ordre aléatoire différent. Puis nous avons effectué un re-test selon les mêmes modalités après la formation.

Afin de pouvoir étudier les effets spécifiques de la formation proposée, nous avons divisé les étudiants en 2 groupes de 9 ce qui nous a permis d'avoir un groupe contrôle. Afin de ne pénaliser personne et que chaque étudiant puisse bénéficier de la formation, nous avons eu recours à une méthodologie croisée (figure VI.b.1.). Les 2 groupes ont donc suivi l'intégralité du protocole mais dans un ordre différent. Le groupe entraîné (groupe E) bénéficiait de la formation immédiatement après la fin des 3 sessions d'évaluation initiale, et le groupe non-entraîné (groupe NE) ne bénéficiait de la formation qu'après avoir effectué les 3 sessions d'évaluation test et les 3 sessions d'évaluation re-test.



Pour les sessions de test, tous les étudiants étaient présents dans une même salle. L'examinateur commençait la session par une explication de l'échelle GRB à quatre niveaux et prenait tout le temps nécessaire pour répondre aux questions des étudiants. Les évaluations étaient ensuite effectuées de façon individuelle sur une fiche nominative. L'examinateur s'assurait de l'absence de concertation entre les auditeurs. Les sessions de re-test, respectant le calendrier de la méthodologie croisée, se sont déroulées de la même façon, mais en groupe : d'abord le groupe SE (qui n'avait pas encore bénéficié de la formation) qui a effectué le re-test 5 semaines après le test ; puis le groupe E, qui a effectué le re-test une semaine après la formation, soit 9 semaines après le test.

La formation des auditeurs s'est déroulée de façon individuelle. Chaque séance durait entre 60 et 90 minutes, au cours de laquelle les voix d'entraînement étaient présentées à l'étudiant sur un ordinateur dédié. En cliquant sur l'icône « son », l'auditeur écoutait la voix (figure VI.c.2.).

Lancer l'écoute du stimulus n° 5 :				
Evaluation :				
	0 (normal)	1 (léger)	2 (moyen)	3 (sévère)
G (grade global)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
R (raucité)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
B (souffle)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Valider				

Figure VI.c.2. : fenêtre d'évaluation du logiciel de formation à l'analyse perceptive.

Puis il effectuait son évaluation en cliquant sur les ronds à cocher. Une fois l'évaluation réalisée, l'auditeur cliquait sur le bouton « valider » (figure VI.c.3).

Lancer l'écoute du stimulus n° 5 :

Evaluation :

	0 (normal)	1 (léger)	2 (moyen)	3 (sévère)
G (grade global)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
R (raucité)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
B (souffle)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Valider

Figure VI.c.3. : exemple d'évaluation.

S'affichait alors la fenêtre de « correction par expertise », récapitulant l'évaluation de l'auditeur et montrant les résultats obtenus par le jury d'experts (figure VI.c.4.).

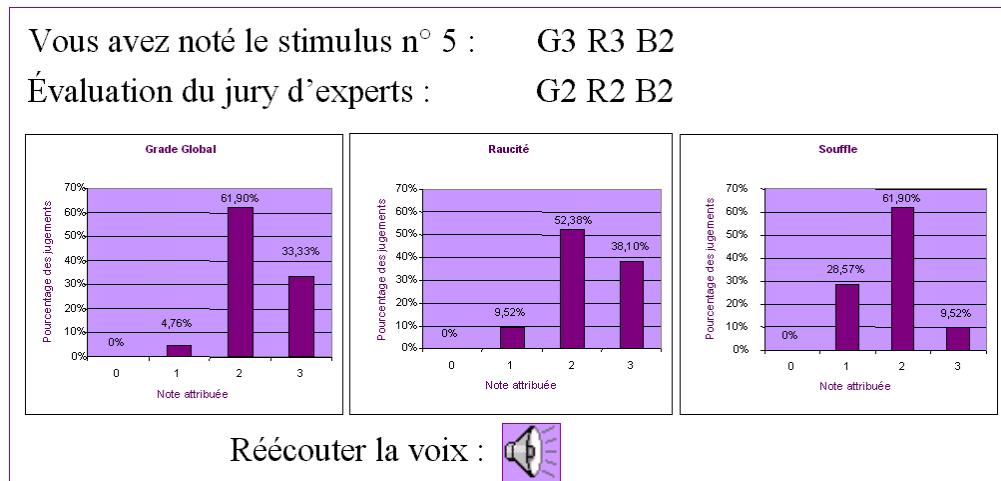


Figure VI.c.4. : fenêtre de « correction par expertise » et feedback.

L'analyse par expertise était présentée sous forme graphique indiquant le pourcentage de réponses données par les experts. Le jugement final d'expertise correspondait au grade présentant le plus fort pourcentage de réponses identiques. Cette présentation permettait à l'étudiant de confronter ses choix avec l'évaluation des experts, tout en lui faisant comprendre qu'il n'existe pas de réponse « juste ». Lors de cette étape, la voix était réécoutée plusieurs fois, et l'examinateur prenait le temps de commenter les caractéristiques de la voix ainsi que les choix des experts. Il répondait également aux

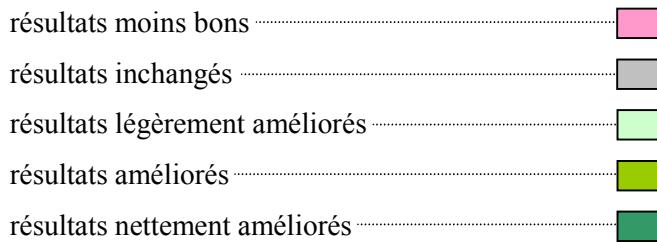
questions posées par l'étudiant. C'est pourquoi les séances de formations duraient aussi longtemps pour seulement 28 voix.

Résultats : L'outil statistique utilisé est le coefficient Kappa Cohen (κ) qui permet la mesure de la variabilité pour des données ordinaires. Afin de faciliter l'interprétation des résultats, nous proposons une table des valeurs coefficients Kappa [De Bodt, 1997] (tableau VI.c.1.).

Valeur du coefficient Kappa	Degré de fiabilité
<0.20	Faible
0.21 – 0.40	Raisonnables
0.41 – 0.60	Assez bon
0.61 – 0.80	Bon
0.81 – 1.00	Très bon

Tableau VI.c.1. : Table d'interprétation des coefficients Kappa.

On note que plus la valeur de κ est faible, plus la variabilité est importante. Ainsi, plus la valeur de κ est élevée, plus les évaluations sont fiables. Le coefficient κ est accompagné de la valeur de « standard error » (se) qui permet de comparer les résultats : si deux valeurs de κ sont distantes de $\pm 2 \text{ se}$, alors la différence est significative ($p < 0.05$). C'est de cette façon que nous avons évalué l'évolution des compétences de notre jury candide. Afin de faciliter la lecture des tableaux, nous avons également mis en place un code couleur, mettant en valeur l'évolution des résultats entre les sessions de test et de re-test :



Variabilité inter-individuelle :

La variabilité inter-individuelle explore les différences de jugements retrouvées entre les auditeurs sur chaque voix. Le tableau VI.c.2. montre l'évolution des performances du groupe Non-Entraîné (NE), pour chacun des paramètres G,R,B, entre les sessions de test et de re-test.

Groupe Non-Entraîné (NE)								
Critères	Sessions Test			Evolution	Sessions Re-test			Interprétation
	κ	se	Interprétation		κ	se	Interprétation	
G	0,438	0,017	assez bon	ns	0,411	0,018	assez bon	
R	0,369	0,016	raisonnable	ns	0,385	0,015	raisonnable	
B	0,334	0,017	raisonnable	ns	0,364	0,024	raisonnable	

Tableau VI.c.2. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité inter-individuelle chez le groupe NE

Les performances inter-individuelles des auditeurs candides en test/re-test sans formation sont moyennes et ne montrent aucune évolution significative. On retrouve une variabilité plus faible sur G que sur R et B, conformément aux données de la littérature.

Le tableau VI.c.3. montre l'évolution des performances du groupe Entraîné (E), pour chacun des paramètres G,R,B, entre les sessions de test et de re-test.

Groupe Entraîné (E)								
Critères	Sessions Test			Evolution	Sessions Re-test			Interprétation
	κ	se	Interprétation		κ	se	Interprétation	
G	0,410	0,015	assez bon	ns	0,407	0,02	raisonnable	
R	0,345	0,015	raisonnable	p<0,05	0,249	0,016	raisonnable	
B	0,253	0,015	raisonnable	p<0,05	0,304	0,017	raisonnable	

Tableau VI.c.3. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité inter-individuelle chez le groupe E

En ce qui concerne les auditeurs candides, la variabilité inter-individuelle sur le grade G ne montre aucune amélioration significative après formation, bien qu'ils passent de « assez bon » à « raisonnable ». En revanche l'évolution sur R et B est significative ($p<0.05$). On note pour le grade R une augmentation significative de la variabilité après entraînement, alors que les performances sur le grade B sont améliorées.

Variabilité intra-individuelle :

La variabilité intra-individuelle explore les différences entre les jugements portés par un même auditeur au cours de différentes sessions. Le tableau VI.c.4. montre l'évolution des performances de chaque auditeur du groupe Non-Entraîné (NE), pour le paramètre G, entre les sessions de test et de re-test.

Groupe Non-Entraîné (NE)							
G	Sessions Test			Evolution	Sessions Re-test		
	κ	se	Interprétation		κ	se	Interprétation
Auditeur a	0,569	0,052	assez bon	ns	0,550	0,052	assez bon
Auditeur b	0,391	0,053	raisonnable	ns	0,440	0,052	assez bon
Auditeur c	0,480	0,048	assez bon	ns	0,456	0,048	assez bon
Auditeur d	0,461	0,053	assez bon	ns	0,463	0,053	assez bon
Auditeur e	0,552	0,048	assez bon	ns	0,548	0,052	assez bon
Auditeur f	0,432	0,048	assez bon	ns	0,405	0,049	faible
Auditeur g	0,463	0,048	assez bon	ns	0,512	0,057	assez bon
Auditeur h	0,602	0,064	assez bon	ns	0,617	0,062	bon
Auditeur i	0,413	0,060	assez bon	p<0,05	0,608	0,074	assez bon

Tableau VI.c.4. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité intra-individuelle chez le groupe NE, pour le grade G.

On note, entre les sessions de test et de re-test, une évolution des performances des auditeurs assez disparate, empirées chez les auditeurs a,c,f ; inchangées chez les auditeurs d,e,h ; et améliorées chez les auditeurs b et g, le tout de manière non significative. Seul l'auditeur i présente une amélioration significative de la reproductibilité de ses jugements avec un coefficient Kappa qui passe de 0.413 lors des sessions de test à 0.608 lors des sessions de re-test, sans avoir bénéficié de la formation entre les deux.

Le tableau VI.c.5. montre l'évolution des performances de chaque auditeur du groupe Entraîné (E), pour le paramètre G, entre les sessions de test et de re-test.

Groupe Entraîné (E)							
G	Sessions Test			Evolution	Sessions Re-test		
	κ	se	Interprétation		κ	se	Interprétation
Auditeur 1	0,521	0,058	assez bon	ns	0,533	0,062	assez bon
Auditeur 2	0,328	0,050	pauvre	p<0,05	0,653	0,051	bon
Auditeur 3	0,483	0,057	assez bon	ns	0,506	0,051	assez bon
Auditeur 4	0,380	0,049	pauvre	p<0,05	0,555	0,053	assez bon
Auditeur 5	0,371	0,049	pauvre	ns	0,439	0,047	assez bon
Auditeur 6	0,452	0,056	assez bon	p<0,05	0,575	0,069	assez bon
Auditeur 7	0,476	0,054	assez bon	ns	0,505	0,061	assez bon
Auditeur 8	0,476	0,049	assez bon	p<0,05	0,645	0,05	bon
Auditeur 9	0,576	0,062	assez bon	ns	0,477	0,071	assez bon

Tableau VI.c.5. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité intra-individuelle chez le groupe E, pour le grade G.

Sur les 9 auditeurs du groupe E, seul le sujet 9 présente des performances altérées lors du re-test, mais de manière non-significative. Tous les autres présentent une amélioration après entraînement, avec une évolution positive significative pour 4 d'entre eux.

Le tableau VI.c.6. montre l'évolution des performances de chaque auditeur du groupe Non-Entraîné (NE), pour le paramètre R, entre les sessions de test et de re-test.

Groupe Entraîné (NE)								
R	Sessions Test			Evolution	Sessions Re-test			Interprétation
	κ	se	Interprétation		κ	se	Interprétation	
Auditeur a	0,420	0,052	assez bon	p<0,05	0,647	0,061	bon	
Auditeur b	0,505	0,054	assez bon	ns	0,417	0,049	assez bon	
Auditeur c	0,355	0,050	raisonnable	p<0,05	0,464	0,052	assez bon	
Auditeur d	0,476	0,050	assez bon	p<0,05	0,596	0,048	assez bon	
Auditeur e	0,516	0,048	assez bon	ns	0,542	0,051	assez bon	
Auditeur f	0,393	0,048	raisonnable	ns	0,460	0,051	assez bon	
Auditeur g	0,402	0,050	raisonnable	ns	0,501	0,052	assez bon	
Auditeur h	0,466	0,057	assez bon	ns	0,503	0,059	assez bon	
Auditeur i	0,404	0,051	raisonnable	p<0,05	0,647	0,057	bon	

Tableau VI.c.6. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité intra-individuelle chez le groupe NE, pour le grade R.

Dans l'ensemble, on peut noter que les performances du groupe non-entraînés sur l'évaluation du paramètre R sont majoritairement améliorées entre les sessions de test et de re-test, avec une évolution significative pour 4 d'entre eux. On note que l'auditeur i présente une forte amélioration significative avec des résultats passant de raisonnables à bons.

Le tableau VI.c.7. montre l'évolution des performances de chaque auditeur du groupe Entraîné (E), pour le paramètre R, entre les sessions de test et de re-test.

Groupe Entraîné (E)								
R	Sessions Test			Evolution	Sessions Re-test			Interprétation
	κ	se	Interprétation		κ	se	Interprétation	
Auditeur 1	0,428	0,050	assez bon	ns	0,500	0,052	assez bon	
Auditeur 2	0,403	0,050	raisonnable	p<0,05	0,659	0,052	bon	
Auditeur 3	0,309	0,052	raisonnable	p<0,05	0,578	0,049	assez bon	
Auditeur 4	0,329	0,048	raisonnable	ns	0,406	0,051	raisonnable	
Auditeur 5	0,386	0,049	raisonnable	ns	0,409	0,048	raisonnable	
Auditeur 6	0,342	0,047	raisonnable	p<0,05	0,562	0,053	assez bon	
Auditeur 7	0,420	0,050	assez bon	ns	0,363	0,052	raisonnable	
Auditeur 8	0,426	0,051	assez bon	p<0,05	0,555	0,05	assez bon	
Auditeur 9	0,287	0,049	faible	p<0,05	0,492	0,068	assez bon	

Tableau VI.c.7. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité intra-individuelle chez le groupe E, pour le grade R.

En ce qui concerne le groupe entraîné, les performances après formation sont améliorées dans 8 cas sur 9, dont 5 évolutions positives significatives.

Le tableau VI.c.8. montre l'évolution des performances de chaque auditeur du groupe Non-Entraîné (NE), pour le paramètre B, entre les sessions de test et de re-test.

Groupe Entraîné (NE)							
B	Sessions Test			Evolution	Sessions Re-test		
	κ	se	Interprétation		κ	se	Interprétation
Auditeur a	0,390	0,058	raisonnable	ns	0,439	0,074	assez bon
Auditeur b	0,429	0,081	raisonnable	ns	0,438	0,087	assez bon
Auditeur c	0,304	0,050	raisonnable	p<0,05	0,443	0,056	assez bon
Auditeur d	0,367	0,048	raisonnable	ns	0,410	0,05	assez bon
Auditeur e	0,446	0,050	assez bon	ns	0,484	0,062	assez bon
Auditeur f	0,325	0,060	raisonnable	ns	0,399	0,069	raisonnable
Auditeur g	0,366	0,074	raisonnable	ns	0,427	0,069	assez bon
Auditeur h	0,381	0,054	raisonnable	ns	0,443	0,064	assez bon
Auditeur i	0,314	0,052	raisonnable	p<0,05	0,438	0,064	assez bon

Tableau VI.c.8. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité intra-individuelle chez le groupe NE, pour le grade B.

Dans l'ensemble, toutes les performances montrent une tendance à l'amélioration. Cette tendance est significative pour deux auditores parmi les 8, dont l'auditeur i.

Le tableau VI.c.9. montre l'évolution des performances de chaque auditeur du groupe Entraîné (E), pour le paramètre B, entre les sessions de test et de re-test.

Groupe Entraîné (E)							
B	Sessions Test			Evolution	Sessions Re-test		
	κ	se	Interprétation		κ	se	Interprétation
Auditeur 1	0,521	0,066	assez bon	ns	0,586	0,067	assez bon
Auditeur 2	0,380	0,049	raisonnable	p<0,05	0,573	0,053	assez bon
Auditeur 3	0,334	0,049	raisonnable	ns	0,286	0,052	raisonnable
Auditeur 4	0,105	0,060	faible	p<0,05	0,359	0,066	raisonnable
Auditeur 5	0,356	0,048	raisonnable	ns	0,401	0,057	raisonnable
Auditeur 6	0,324	0,053	raisonnable	p<0,05	0,566	0,056	assez bon
Auditeur 7	0,369	0,049	raisonnable	ns	0,393	0,05	raisonnable
Auditeur 8	0,381	0,055	raisonnable	p<0,05	0,600	0,057	assez bon
Auditeur 9	0,315	0,053	raisonnable	p<0,05	0,451	0,056	assez bon

Tableau VI.c.9. : résultats coefficient κ , se, et interprétation de l'évolution de la variabilité intra-individuelle chez le groupe E, pour le grade B.

En ce qui concerne le groupe ayant bénéficié de la formation, 5 auditores présentent une amélioration significative de leurs performances.

Discussion : L'objectif de cette étude était de proposer une méthodologie d'entraînement selon un paradigme stimulus/réponse/feedback/stimulus pour la formation d'auditeurs candides à l'analyse perceptive de la dysphonie. Notre objectif était de vérifier les effets de ce protocole sur les mesures de variabilité intra- et inter-individuelle (coefficient Kappa de Cohen) chez les auditeurs inexpérimentés. Notre hypothèse était que la concordance et la constance des évaluations portées sur un panel de 50 voix dysphoniques et normales seraient améliorées après la formation.

Variabilité inter-individuelle :

En 1997, De Bodt présente une étude sur la fiabilité de l'échelle GRBAS en fonction de l'expérience des auditeurs [De Bodt, 1997]. Le tableau suivant (tableau VI.c.10.) montre les résultats des coefficients Kappa qu'il a obtenus sur les paramètres GRB, en fonction du niveau d'expérience des auditeurs.

	I	E
G	0,39	0,49
R	0,16	0,20
B	0,21	0,20

Tableau VI.c.10. [De Bodt, 1997] : Agrément inter-auditeurs (coefficient Kappa) en fonction de leur niveau d'expérience pour les paramètres GRB. I représente le groupe d'auditeurs inexpérimentés, E représente le groupe d'auditeurs expérimentés.

Dans notre étude, les résultats montrent que les valeurs de concordance inter-individuelle avant formation de nos 18 auditeurs (tableaux VI.c.2. et 3.), tous inexpérimentés, sont proches des valeurs obtenues par De Bodt pour les auditeurs expérimentés (tableau VI.c.10.). En effet, lors de la première série d'écoutes (test), l'interprétation des valeurs de κ pour nos deux groupes renvoie à la catégorie « assez bon » pour le paramètre G (Groupe E=0.438 ; groupe SE=0.410). Pour les paramètres R et B, nos résultats montrent des coefficients Kappa « raisonnables », alors que les auditeurs expérimentés de De Bodt obtiennent des résultats entrant dans la catégorie « faible ». Ainsi, pour le paramètre G, nos auditeurs montrent une concordance inter-individuelle, c'est à dire un agrément entre eux, proche de celle obtenue par les auditeurs expérimentés de De Bodt. Pour les grades R et B, cette concordance semble même meilleure, bien que demeurant classée dans la catégorie « « raisonnable ». Ces résultats pourraient évoquer un biais de recrutement de notre jury d'écoute inexpérimenté. Pourtant, comme décrit dans notre protocole d'étude, tous nos auditeurs sont des étudiants de 2^{ème} année d'orthophonie. Ils n'ont pas reçu de formation particulière et ne possèdent aucune expérience dans la tâche d'évaluation de la dysphonie. D'ailleurs les valeurs du coefficient κ obtenues par nos deux groupes sont modérées, puisque leur interprétation varie de la catégorie « raisonnable » à la catégorie « assez bon ». Ces résultats modestes semblent être en adéquation avec l'absence d'expérience de nos

auditeurs. En revanche, le classement des auditeurs expérimentés relevé par De Bodt semble limite pour des professionnels de l'évaluation de la voix. En d'autres termes, on ne peut pas considérer que notre jury présente des résultats étonnamment bons, mais que les auditeurs expérimentés de De Bodt présentent des résultats en deçà de leurs capacités attendues. Il ne semble donc pas exister de biais de recrutement de notre jury. Cependant, il faut considérer que le fait que nos auditeurs présentent une concordance inter-individuelle proche de celle relevée dans la littérature pour les auditeurs expérimentés [De Bodt, 1997], induit une faible marge de progression pour notre groupe bénéficiant de la formation.

En effet, on constate que le groupes des auditeurs ayant bénéficié de la formation (groupe E) ne montre pas d'évolution significative des performances inter-individuelles sur le critère G, après la séance d'entraînement (tableau VI.c.3.). Pour le critère de souffle (B), les résultats sont significativement meilleurs après la formation, mais demeurent néanmoins « raisonnables ». En revanche, les résultats de variabilité inter-individuelle sur la raucité (R) montrent une détérioration significative après la formation, bien qu'il demeurent dans la même classe d'interprétation (« raisonnable »). Ce point particulier illustre la possibilité d'une hétérogénéité de la compréhension du caractère rauque entre les différents auditeurs. Notons que la valeur obtenue par De Bodt chez les auditeurs inexpérimentés est également la plus faible ($\kappa=0.16$), pour un classement en catégorie « faible ». Il semble que l'apprentissage de la perception de la raucité soit délicate et résistante à la formation proposée, en ce qui concerne l'agrément entre les auditeurs.

Au total, la fiabilité inter-individuelle de nos auditeurs est classée de « raisonnable » à « assez bon ». Le jugement attribué à un même stimulus est donc assez variable d'un auditeur à l'autre. Ces résultats s'expliquent classiquement par l'hétérogénéité de la sensibilité des individus membres d'un jury d'écoute [Kreiman, 1993 ; De Bodt, 1997]. En effet, les caractéristiques utilisées pour décrire la voix pathologiques (souffle, raucité) sont les mêmes que les critères de « voix de charme », décrits par Laver [Laver, 1981]. Certains auditeurs peuvent donc par exemple trouver dans un caractère rauque une particularité esthétique et tendre à sous évaluer son trait pathologique. Nous pouvons également considérer le facteur « sympathie du locuteur » auquel les auditeurs candides sont particulièrement sensibles. En effet, lors des séances individuelles d'entraînement, nous avons pu constater que certains auditeurs pouvaient se laisser « séduire » ou montrer une forme de rejet selon le stimulus proposé. Par exemple, une voix a été qualifiée à plusieurs reprises de voix « d'outre-tombe » et notée sévèrement alors qu'elle n'était cotée que grade 1 par le jury d'expert. A l'inverse, la voix d'un « papi » a été prise en sympathie par certains sujets. Ces exemples illustrent la difficulté des sujets candides à se concentrer sur la qualité vibratoire sans se laisser influencer par les critères prosodique, linguistiques et psychologiques véhiculés par la voix du locuteur. Notre jury, composé d'étudiants, a quelquefois eu du mal à se soustraire de l'aspect « encouragement » ou « sanction » contenu implicitement dans la

tâche de « noter ». Cela révèle une difficulté à comprendre et à appliquer la tâche d'évaluation puisque le but est de qualifier et de quantifier la dysphonie, et pas de révéler quelles sont les voix que l'on préfère ou que l'on n'aime pas. Ce facteur d'influence, largement décrit dans la littérature [Kreiman, 1993 ; De Bodt, 1997 ; Millet, 1998 ; Goldstone, 1998] a affecté notre jury de façon hétérogène : certains ont mieux compris et appliqué la méthode que d'autres. Notre hypothèse était que la formation que nous proposions permettrait de normaliser l'application de la méthode entre les étudiants constituant le jury inexpérimenté. Il semble que notre protocole d'entraînement soit insuffisant pour réduire la variabilité inter-individuelle d'un jury d'écoute inexpérimenté. Il est possible d'envisager qu'une formation complémentaire, plus axée sur une explication de la tâche à effectuer et des objectifs de la méthode, permette d'améliorer l'efficacité des sessions de formations.

Variabilité intra-individuelle :

Le tableau suivant (tableau VI.c.11.) montre la fiabilité intra-individuelle (coefficients Kappa de Cohen) obtenue par De Bodt sur les grades GRB, en fonction du niveau d'expérience des auditeurs [De Bodt, 1997].

	I	E
G	0,50	0,70
R	0,29	0,40
B	0,29	0,45

Tableau VI.c.11. [De Bodt, 1997] : Agrément intra-auditeurs (coefficient Kappa) en fonction de leur niveau d'expérience pour les paramètres GRB. I représente le groupe d'auditeurs inexpérimentés, E représente le groupe d'auditeurs expérimentés.

En ce qui concerne la fiabilité intra-individuelle, c'est à dire la reproductibilité des jugements d'un même auditeur d'une session à l'autre, nous obtenons, tout comme De Bodt, des résultats largement meilleurs que pour la fiabilité inter-individuelle. Ce phénomène est largement décrit dans la littérature [Kreiman, 1993 ; Hammarberg, 1980 ; De Bodt, 1997 ; Dejonckere, 1993]. D'une manière générale, on considère que les jugements effectués à plusieurs reprises par un même auditeur sont moins variables que les jugements prononcés d'un auditeur à l'autre.

Pour le grade global (G), l'ensemble des valeurs de κ que nous obtenons sont proches des résultats de De Bodt pour les auditeurs inexpérimentés. On peut donc considérer que la tendance d'amélioration entre la fiabilité inter- et intra-individuelle est conservée dans notre étude bien que l'on ne retrouve pas l'analogie avec le groupe expérimenté que nous observions pour la fiabilité inter-individuelle. En revanche, en ce qui concerne les paramètres de raucité (R) et de souffle (B) (tableaux VI.c.6. à 9.),

l'ensemble de nos auditeurs obtiennent, avant formation, des résultats équivalents, voire meilleurs que le groupe d'auditeurs expérimentés décrit par De Bodt.

En ce qui concerne l'effet de la formation sur l'évolution de la fiabilité intra-individuelle du grade global de dysphonie (G), on constate qu'après l'entraînement, tous les auditeurs du groupe E sont améliorés, sauf un (tableau VI.c.5.). Sur les 8 auditeurs qui présentent une amélioration, 4 montrent une évolution significative ($p<0.05$). En particuliers les auditeurs 2 et 8 qui passent respectivement d'une reproductibilité classée « raisonnable » à « bon », et « assez bon » à « bon ». Seul l'auditeur 9 semble n'avoir pas tiré profit de la formation puisque sa reproductibilité semble légèrement détériorée bien que de manière non-significative. Malgré cette amélioration globale des performances intra-individuelles de notre jury, la reproductibilité n'atteint pas les résultats décrits par De Bodt pour un jury expérimenté [De Bodt, 1997]. Néanmoins, les scores obtenus sont largement supérieurs à ceux de son jury inexpérimenté. Notre protocole semble donc favoriser l'amélioration de la reproductibilité intra-individuelle du paramètre G chez les auditeurs candides, bien que rien ne remplace les années d'expérience professionnelle. Enfin, en ce qui concerne le groupe NE, qui n'a pas bénéficié de la formation (tableau VI.c.4.), seul l'auditeur i présente une amélioration significative ($p<0.05$). Pour tous les autres, il n'existe pas d'évolution. Il semble donc que cet auditeur ait su tirer partie de l'effet d'apprentissage lié à l'exposition répétée aux voix pathologiques de notre corpus.

Pour le critère de raucité (R) (tableau VI.c.7.), la reproductibilité intra-individuelle est améliorée pour tous les auditeurs ayant bénéficié de la formation (groupe E), sauf un. Parmi ces 8 auditeurs montrant une évolution positive, 5 présentent une amélioration significative ($p<0.05$), avec en particulier l'auditeur 2 dont les performances passent du classement « raisonnable » à « bon ». D'autre part, on constate qu'après la formation, les performances de ce groupe sur l'évaluation de la raucité sont largement meilleures que les résultats obtenus par De Bodt sur les auditeurs expérimentés (tableau VI.c.11.). Les performances de ce groupe étaient déjà proches du groupe expérimenté de De Bodt avant formation, pourtant, notre protocole d'entraînement montre un effet très positif sur la reproductibilité des évaluations de raucité. Cependant, lorsqu'on regarde les résultats obtenus par le groupe non-entraîné (NE), on constate globalement la même évolution. Notons l'évolution de l'auditeur i, qui ici encore se détache du lot par une amélioration significative qui lui permet de passer d'un classement en « raisonnable » à un classement en « bon ». Il est donc difficile de conclure que cette amélioration soit due exclusivement à l'entraînement. Une explication de ce phénomène peut provenir de la possibilité d'un effet d'apprentissage. En effet, dans le but de pouvoir mesurer la fiabilité intra-individuelle, et conformément aux méthodologies mises en place dans la littérature [Kreiman, 1992 ; Dejonckere, 1993 ; Millet, 1998], nous avons proposé à nos deux groupes 2 séries de 3 sessions d'écoute de notre corpus, c'est à dire 3 sessions test et 3 sessions re-test. Ainsi, lors de la première session de la deuxième série (re-test), nos auditeurs évaluaient déjà les mêmes voix pour la

quatrième fois. Au total, toutes nos voix auront été évaluées 6 fois par nos 18 auditeurs. Il est donc possible que l'amélioration des performances lors de la série re-test soit due en partie à l'effet d'apprentissage par exposition répétée aux différents stimuli. Une étude ultérieure pourrait s'attacher à étudier spécifiquement cet effet d'apprentissage.

Pour le critère de souffle (B) (tableaux VI.c.8. et 9.), 8 auditeurs sur 9 ont amélioré leurs résultats, et ce, dans les 2 groupes. 2 auditeurs du groupe NE obtiennent des résultats significatifs, contre 5 dans le groupe E ($p<0.05$). L'effet de l'entraînement semble donc supérieur à l'effet d'apprentissage.

D'une manière générale, on constate que les auditeurs qui ont obtenu une amélioration significative de leurs résultats sur le grade global montrent aussi une évolution positive sur les autres critères (RB). C'est le cas de 5 auditeurs du groupe E (2, 4, 6, 8, 9), et de 1 auditeur du groupe NE (i). L'auditeur 3 du groupe E et les auditeurs a et d du groupe NE montrent une amélioration significative uniquement sur le critère R. Ceci va dans le sens de Dejonckere qui estime que le critère R est le plus facile à repérer dans la dysphonie [Dejonckere, 1995]. Une autre hypothèse est que ces auditeurs peuvent être plus sensibles à la raucité. Si on compare l'évolution des résultats au sein des deux groupes, l'amélioration est significative et plus importante pour le groupe E. On peut donc penser que notre protocole d'entraînement a contribué à l'amélioration des résultats de ce groupe. Précisons que le groupe NE obtenait initialement de meilleurs résultats que le groupe E. Il semble que la formation soit d'autant plus efficace que les performances initiales sont faibles. En d'autres termes, plus les auditeurs sont variables à priori, plus la formation leur est bénéfique. Enfin, les effets de l'entraînement semblent positifs et efficaces, mais la différence des résultats obtenus entre les groupe E et NE est insuffisante pour affirmer l'efficacité spécifique de la formation : on peut penser que les sessions d'évaluation ont constitué un apprentissage pour tous nos auditeurs, leur permettant d'affiner leur perception des paramètres de la dysphonie et de les stabiliser.

Conclusion : Notre étude visait deux objectifs : 1- obtenir une diminution de la variabilité inter- et intra-individuelle chez les auditeurs inexpérimentés, et de façon plus importante pour le groupe ayant bénéficié de la formation et 2- voir le niveau de variabilité après entraînement se rapprocher des valeurs obtenues chez les experts. Nos résultats montrent une amélioration plus grande pour la fiabilité intra-individuelle que pour la fiabilité inter-individuelle. Cependant, les valeurs de fiabilité inter-individuelle chez nos auditeurs inexpérimentés sont déjà proches des résultats relevés dans la littérature concernant les auditeurs expérimentés, pour chacun des critères GRB. Notre protocole a permis d'améliorer la variabilité intra-individuelle de notre jury inexpérimenté. En fin d'expérimentation, les performances de notre jury candide étaient toujours supérieures aux performances de jurys inexpérimentés relevées dans la littérature, et se rapprochaient voire dépassaient des performances de jurys expérimentés. Nous pouvons donc considérer que nos hypothèses sont

validées. Néanmoins, la question de l'apprentissage par exposition aux stimuli reste posée. Des travaux ultérieurs pourraient permettre de quantifier la part de l'effet d'apprentissage dans l'amélioration des résultats chez les auditeurs inexpérimentés.

Parallèlement aux travaux présentés dans le but d'améliorer les performances de l'analyse perceptive en milieu expérimental, notre réflexion s'orientait vers une démarche plus clinique, permettant aux praticiens d'optimiser leur évaluation dans la pratique quotidienne de leur profession. Cette démarche a abouti en 1999 à la conception et à la réalisation du logiciel CLAP ® (Classement de l'Analyse Perceptive) destiné à fournir aux ORL, Phoniatres et Orthophonistes, un outil d'aide à l'analyse perceptive. L'étude que nous présentons maintenant a pour but de valider le logiciel CLAP ® en tant que méthodologie d'analyse perceptive de la dysphonie pour thérapeutes de la voix.

d. Logiciel de CLassement de l'Analyse Perceptive (CLAP) :
Validité clinique et influence sur la variabilité du jury

Revis J, Nazarian B, Bingenheimer S, Giovanni A

(Glossa, soumis à publication)

Introduction : L'analyse perceptive est utilisée de façon systématique par les différents praticiens intervenant auprès des patients dysphoniques. Cette procédure, qui fait appel à la discrimination auditive, est souvent mise en place de façon empirique sans tenir compte d'une méthodologie rigoureuse. Or, le caractère « subjectif » du jugement perceptif nécessite la mise en place d'un protocole d'évaluation strict pour que les résultats obtenus soient valides. Notre expérience dans la formation des orthophonistes à l'analyse perceptive (animation d'ateliers en congrès) nous a montré que peu de précautions allant dans ce sens étaient prises pour ce type d'évaluation en pratique clinique : la plupart du temps, les praticiens effectuent leur évaluation du patient « en direct » sur la parole spontanée, sans avoir recours à un enregistrement, donc sans la possibilité de comparer l'état de la voix en début de prise en charge et en fin de traitement. Cette réflexion nous a conduits à concevoir et développer un logiciel d'aide à l'analyse perceptive destiné aux thérapeutes de la voix, tenant compte des avancées de la recherche dans les méthodes permettant de fiabiliser le jugement perceptif des dysphonies.

Le premier logiciel d'assistance à l'analyse perceptive a été présenté par Hammarberg [Hammarberg, 1997] (voir section V., chapitre 2.6., p 93). Dans le but d'améliorer la performance des sujets dans la tâche d'évaluation perceptive, Hammarberg a proposé un programme informatique basé sur l'utilisation d'une échelle visuelle analogique : le VIIsual SOrt and Rate (logiciel VISOR). Le principe était de présenter sous forme d'icônes une série de 5 voix synthétiques présentant divers degrés de pathologie. L'évaluation était réalisée par le classement des stimuli par ordre croissant de niveau pathologique, avec la possibilité de réécouter les voix aussi souvent que nécessaire pour effectuer la comparaison. Les résultats ont montré que cette méthode permettait d'améliorer la fiabilité des évaluations perceptives de façon significative. Cependant, ce logiciel - conçu spécifiquement pour la langue suédoise - n'est à l'heure actuelle toujours pas disponible.

Notre objectif était de concevoir et de tester, sur les bases du logiciel VISOR et en tenant compte des plus récentes publications sur les méthodologies perceptives, un logiciel d'assistance à l'analyse perceptive des dysphonies. Le logiciel de CLassement de l'Analyse Perceptive (CLAP) est né de la

collaboration entre une orthophoniste spécialiste de l'analyse perceptive et de l'ingénieur de notre laboratoire. Nous l'avons voulu complet et facile à utiliser. CLAP comprend donc une partie didactique destinée à la formation du jury d'écoute grâce à une banque de données riche d'une centaine de voix pathologiques, et combine deux possibilités d'évaluation : une évaluation individuelle avec ancrage auditif externe naturel permettant de procéder à l'écoute d'un nombre infini de stimuli ; et une évaluation par comparaison de 8 stimuli simultanément. Ces écoutes sont réalisées en aveugle, et les résultats sont transférés automatiquement dans un tableur. Le but de ce travail est de tester la validité clinique de cette seconde méthode par comparaison, notre hypothèse étant que le logiciel CLAP permet de diminuer la variabilité des jugements perceptifs.

Protocole expérimental :

1- Présentation du logiciel.

Le logiciel CLAP permet d'évaluer la dysphonie par le classement des stimuli par ordre croissant de leur niveau pathologique. Dans l'écran « affectation des voix » (figure VI.d.1.), le manipulateur est invité à attribuer une couleur à chacun des stimuli utilisés pour l'évaluation.

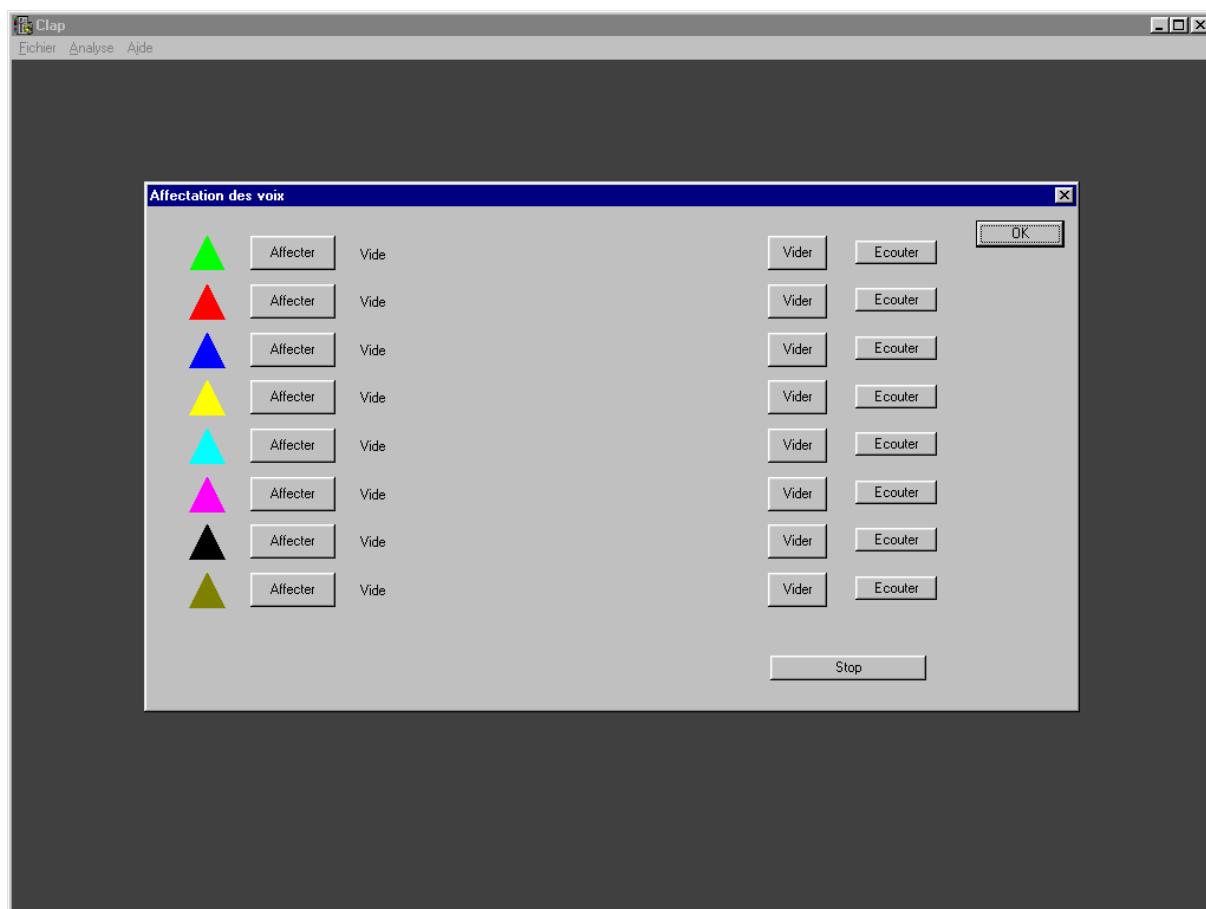


Figure VI.d.1. : fenêtre d'affectation des voix dans le logiciel CLAP.

Chaque voix est alors représentée par une couleur que l'on retrouve dans le module d'évaluation. Cette procédure permet de conserver l'anonymat des stimuli et d'effectuer par conséquent l'évaluation en aveugle. CLAP fournit la possibilité d'évaluer par comparaison jusqu'à 8 voix simultanément.

Une fois l'affectation réalisée hors de la présence de l'auditeur, le manipulateur passe à l'écran d'analyse (figure VI.d.2.). L'auditeur a alors la possibilité d'écouter chaque voix en cliquant sur le bouton « écouter » aussi souvent que nécessaire afin d'établir un classement par ordre croissant de dysphonie. L'évaluation de chaque voix se fait en déplaçant un curseur le long d'une échelle visuelle analogique, l'extrême gauche représentant la normalité et l'extrême droite le niveau maximal de pathologie. En fonction de la position du curseur, une représentation 3-D située à droite de l'écran, sous forme de disques de couleur se déplaçant le long d'un axe, permet à l'auditeur d'avoir une vue d'ensemble simplifiée du résultat de son classement afin de pouvoir éventuellement réécouter les voix dans l'ordre qu'il a choisi (le bas de l'axe représente la normalité, et le haut le niveau pathologique maximal). Le manipulateur a la possibilité de faire apparaître ou non la valeur chiffrée correspondant à la place du curseur sur l'échelle visuelle analogique, puis de l'exporter vers un tableau (figure VI.d.3.). Le traitement des résultats peut alors se faire en valeurs relative (ordre de classement) ou en valeur absolue (note de l'échelle visuelle analogique).

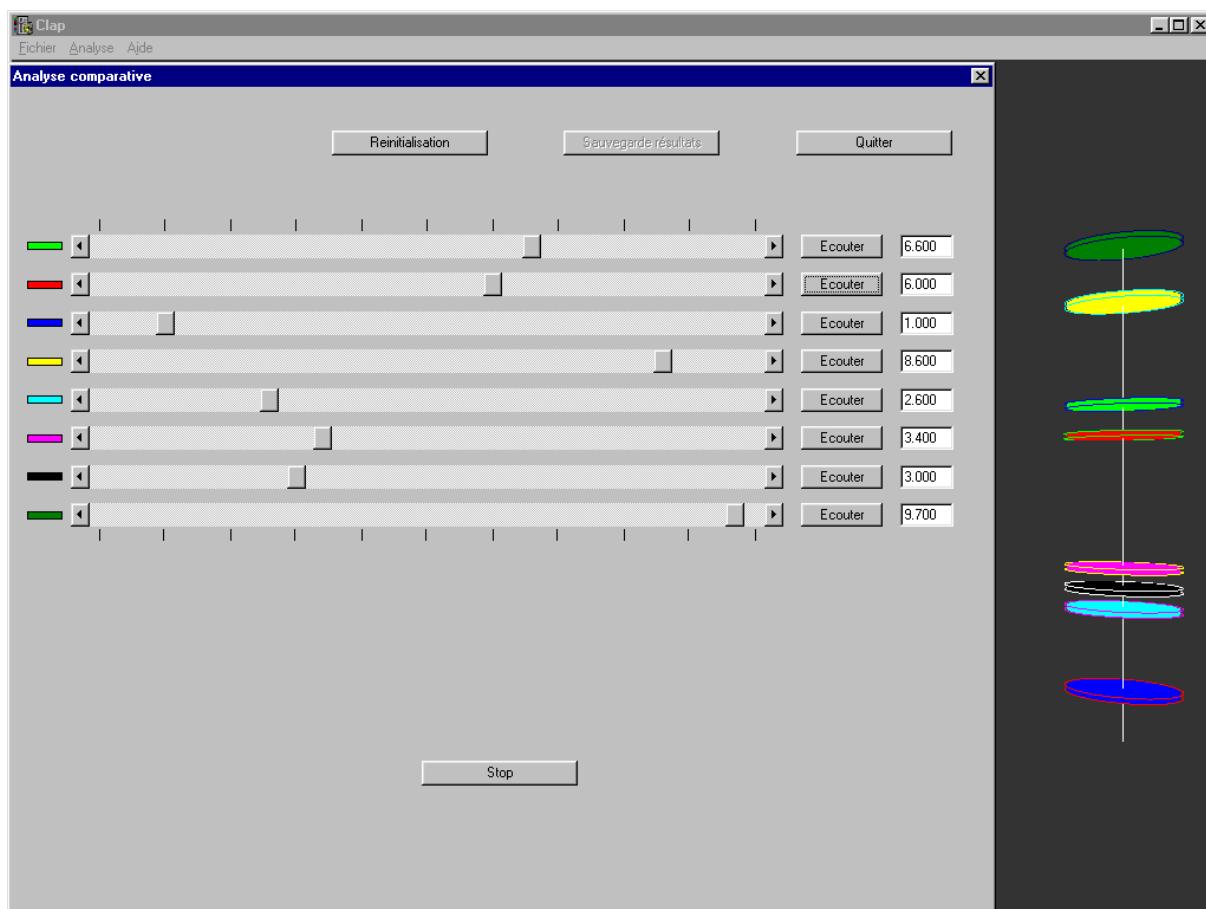


Figure VI.d.2. : fenêtre d'analyse comparative.

Couleur	Code Patient	Note	Ordre
Vert	gdmf	6,60	6
Red	rsjf	6,00	5
Blue	godf	1,00	1
Yellow	aymf	8,60	7
Cyan	abqf	2,60	2
Magenta	gimf	3,40	4
Black	flcf	3,00	3
Green	dnsf	9,70	8

Figure VI.d.3. : recueil et présentation des données dans le tableau

2- Choix des sujets

Les voix que nous avons utilisées pour cette étude ont été choisies dans la banque de données de notre laboratoire en fonction de leur niveau pathologique déterminé par jury d'experts lors de précédentes études. Nous avons choisi 24 voix réparties en 6 voix grade 0 (voix normale), 6 voix grade 1 (dysphonie légère), 6 voix grade 2 (dysphonie moyenne) et 6 voix grade 3 (dysphonie sévère). La moyenne d'âge était de 49.9 ans et la répartition en pathologies est donnée en figure VI.d.4.

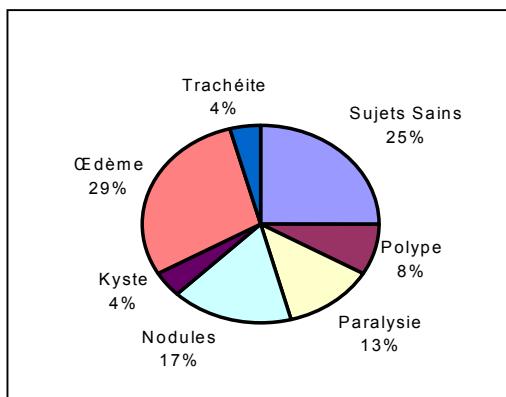


Figure VI.d.4. : répartition en pathologies.

Les patients choisis ont ensuite répartis par tirage au sort en 3 groupes équivalents comprenant chacun 2 voix grade 0, 2 voix grade 1, 2 voix grade 2, et 2 voix grade 3 pour l'évaluation avec le logiciel CLAP.

3- Choix du jury d'écoute

Le jury d'écoute était constitué de 3 cliniciens expérimentés (1 médecin phoniatre et 2 orthophonistes) ayant participé de façon systématique à toutes nos études d'analyse perceptive de la dysphonie.

4- Déroulement des sessions d'écoute

Les voix ont été évaluées selon 2 protocoles d'analyse. Le premier protocole consistait en l'évaluation individuelle du grade global de dysphonie (grade G de l'échelle GRBAS) sur la totalité des voix recueillies divisées en 3 groupes équivalents, comprenant chacun 2 voix grade 0, 2 voix grade 1, 2 voix grade 2 et 2 voix grade 3, soit 8 voix. A l'intérieur de ces 3 groupes les 8 voix étaient présentées dans un ordre aléatoire. L'échelle quantitative utilisée était une échelle visuelle analogique. Les auditeurs pouvaient réécouter les voix aussi souvent que nécessaire afin d'effectuer leur évaluation. Ces évaluations ont été renouvelées 3 fois à une semaine d'intervalle afin de mesurer la variabilité intra- et inter-individuelle de notre jury. Lors de chaque session, les voix étaient présentées dans un ordre aléatoire différent tout en respectant les 3 groupes prédefinis afin d'éviter tout effet de contexte. Nous avons relevé les valeurs absolues de notes attribuées à chaque voix, à partir desquelles nous avons déterminé le classement des voix par ordre croissant de dysphonie en fonction des 3 groupe de 8 voix prédefinis. Pour le traitement statistique, nous avons conservé l'ordre de classement des voix.

Le second protocole consistait en l'évaluation perceptive individuelle de la totalité du corpus divisé en 3 groupes équivalents, comprenant chacun 2 voix grade 0, 2 voix grade 1, 2 voix grade 2 et 2 voix grade 3, soit 8 voix. Cette évaluation était réalisée avec le logiciel CLAP et la consigne était de classer les voix par ordre croissant de niveau de dysphonie. Les auditeurs pouvaient réécouter les voix aussi souvent que nécessaire afin d'effectuer leur comparaison et leur classement. Chaque membre du jury a renouvelé son évaluation 3 fois à une semaine d'intervalle, afin que nous puissions mesurer la variabilité intra- et inter-individuelle. Lors de chaque session, les groupes étaient constitués des mêmes voix, mais une couleur différente leur était attribuée par tri aléatoire afin d'éviter tout effet d'apprentissage. Pour le traitement statistique, nous avons conservé l'ordre de classement des voix.

5- Le traitement statistique

Dans un premier temps, nous avons calculé la variabilité intra- et inter-individuelle de notre jury sur le premier protocole d'évaluation (échelle visuelle analogique) par le calcul du pourcentage de classement identique, puis nous avons utilisé la même méthode pour le calcul des variabilités sur le second protocole (logiciel CLAP). Nous avons ensuite comparé les résultats obtenus selon ces deux méthodes.

Notre hypothèse était que le logiciel CLAP permettait d'améliorer la fiabilité des évaluations perceptives.

Résultats : Le tableau VI.d.1. montre le pourcentage de classements identiques obtenu sur l'ensemble des évaluations effectuées par le jury avec l'échelle visuelle analogique (VAS) d'une part, et avec le logiciel CLAP d'autre part, en fonction des 3 groupe d'évaluation.

VARIABILITE INTER-INDIVIDUELLE		
% Classement Identique	VAS	CLAP
Groupe 1	49,95%	65,21%
Groupe 2	65,24%	74,95%
Groupe 3	52,73%	72,30%
Total	55,97%	70,82%

Tableau VI.d.1. : expression de la variabilité inter-individuelle par le pourcentage de classement identique des 3 membres du jury selon les 3 groupes de 8 voix.

Les résultats montrent que la reproductibilité du classement d'un auditeur à l'autre est meilleure avec le logiciel CLAP qu'avec une méthodologie VAS classique, puisqu'il permet de passer de 57,97% de classements identiques à 70,82%.

Le tableau VI.d.2. montre le pourcentage de classements identiques obtenu par chacun des membres du jury avec l'échelle visuelle analogique (VAS), en fonction des 3 groupe d'évaluation.

VARIABILITE INTRA-INDIVIDUELLE VAS			
% Classement Identique	Jury 1 VAS	Jury 2 VAS	Jury 3 VAS
Groupe 1	70,83%	66,66%	70,83%
Groupe 2	66,66%	70,83%	58,33%
Groupe 3	83,33%	62,50%	62,50%
Total	73,61%	66,66%	63,88%

Tableau VI.d.2. : expression de la variabilité intra-individuelle avec la méthodologie VAS classique par le pourcentage de classement identique de chaque membre du jury selon les 3 groupes de 8 voix.

Le tableau VI.d.3. montre le pourcentage de classements identiques obtenu par chacun des membres du jury avec le logiciel CLAP, en fonction des 3 groupe d'évaluation.

VARIABILITE INTRA-INDIVIDUELLE CLAP			
% Classement Identique	Jury 1 CLAP	Jury 2 CLAP	Jury 3 CLAP
Groupe 1	79,16%	91,66%	79,16%
Groupe 2	75,00%	100,00%	83,33%
Groupe 3	91,66%	83,33%	70,83%
Total	81,94%	91,66%	77,77%

Tableau VI.d.3. : expression de la variabilité intra-individuelle avec le logiciel CLAP par le pourcentage de classement identique de chaque membre du jury selon les 3 groupes de 8 voix.

Conformément aux données relevées dans la littérature concernant les mesures de variabilité du jury d'écoute [Kreiman, 1993 ; De Bodt, 1996], on observe que la reproductibilité intra-individuelle est meilleure que la reproductibilité inter-individuelle quelle que soit la méthode utilisée (tableaux VI.d.1., 2 . et 3.). Néanmoins, on constate que l'utilisation du logiciel CLAP permet une grande amélioration des résultats de chaque auditeur pour tous les groupes de voix étudiés (tableaux VI.d.2. et 3.). On remarque que l'auditeur 2 passe de 70.83% de reproductibilité avec la méthode VAS classique, à 100% avec le logiciel CLAP pour l'évaluation des voix du groupe 2.

Discussion : Notre hypothèse était que l'utilisation du logiciel CLAP permet d'améliorer la fiabilité des jugements perceptifs. Les résultats que nous avons obtenus permettent de confirmer cette hypothèse puisque tous les pourcentages de classement identiques sont meilleurs avec le logiciel CLAP qu'avec la méthodologie VAS classique.

Lors de la conception du logiciel CLAP, notre objectif était de proposer un logiciel d'assistance à l'analyse perceptive en situation clinique. Notre cahier des charges était que cet outil informatique permette :

- Une utilisation simple et rapide, adaptée à la pratique clinique.

Les remarques des auditeurs ayant participé à l'étude ont confirmé le côté facile et ludique de l'utilisation du logiciel. L'évaluation par comparaison est particulièrement favorable à l'utilisation clinique où l'enjeu principal est de vérifier l'évolution de la voix d'un même patient au cours du traitement, davantage que de comparer la gravité des troubles de plusieurs patients dysphoniques. De ce point de vue, CLAP offre la possibilité de confronter jusqu'à 8 enregistrements de la voix d'un même patient afin d'observer l'efficacité du traitement rééducatif. Pour ce travail, nous avons pris parti de proposer aux auditeurs un ensemble d'enregistrements provenant de différents locuteurs afin de pouvoir comparer les résultats de la variabilité du jury avec les données disponibles dans la littérature. Cette démarche nous a permis de confirmer la fiabilité des évaluations réalisées avec ce logiciel. Néanmoins, une autre étude proposant les enregistrements de l'évolution de la voix d'un même locuteur pourrait permettre de valider l'intérêt thérapeutique de l'utilisation de CLAP.

- De proposer un cadre méthodologique.

Pour la conception du logiciel CLAP, inspiré du logiciel VISOR présenté par Hammarberg, nous avons tenu compte des avancées proposées dans la littérature concernant l'analyse perceptive des dysphonies. Dans une démarche résolument clinique, notre choix s'est porté vers l'échelle quantitative visuelle analogique, particulièrement adaptée à la discrimination de différences fines. En effet, si l'échelle ordinaire en 4 classe (de 0 à 3) est sans doute plus instinctive pour l'évaluation de la dysphonie « dans l'absolu », elle est inadaptée à la comparaison des voix entre elles. En pratique

clinique, l'amélioration de la voix d'un patient peut être probante, sans nécessairement conduire à un changement de classe. L'échelle visuelle analogique peut permettre d'objectiver cette évolution qui passerait inaperçue avec une échelle ordinaire. En ce qui concerne le paramètre choisi (grade global), il semble qu'il soit le plus souvent utilisé en pratique clinique et qu'il reflète la demande du patient (amélioration générale du timbre de la voix). Néanmoins, rien n'empêche le praticien de renouveler l'opération 3 fois s'il désire tenir compte également des paramètres de raucité et de souffle. Enfin, l'idée de classement par comparaison permet de bénéficier de la méthodologie inspirée de l'ancrage auditif externe [Gerratt, 1994] puisque chaque voix proposée fournit un référent pour chacune des autres.

- D'imposer à l'auditeur d'effectuer ses évaluations d'une manière systématique et rigoureusement identique

L'utilisation du logiciel limite volontairement les degrés de liberté de l'auditeur afin de le cantonner dans une méthodologie rigoureuse. Les conditions d'écoute sont nécessairement identiques (matériel utilisé, échelle quantitative, conditions en aveugle). CLAP permet donc de standardiser les évaluations effectuées en pratique clinique grâce à une procédure imposée par un cadre relativement rigide.

Conclusion : Le logiciel CLAP, conçu pour l'analyse perceptive des dysphonies en pratique clinique, permet de fiabiliser l'évaluation de la voix. Le taux de reproductibilité des auditeurs dans la tâche de classement par ordre croissant de dysphonie est largement amélioré par rapport à l'utilisation d'une échelle visuelle analogique classique, et supérieur aux données relevées dans la littérature. Le logiciel permet une standardisation des procédures d'évaluation tenant compte des toutes dernières avancées de la recherche, qui permet d'optimiser la fiabilité des évaluations. Conçu par une orthophoniste, il est adapté aux besoins des praticiens libéraux, développé par un ingénieur de recherche informaticien il est simple à utiliser. Une extension est en cours de développement afin de permettre l'enregistrement des voix directement dans le logiciel avec un simple microphone d'ordinateur. Enfin, une nouvelle étude basée sur l'évaluation de l'évolution de la voix d'un même auditeur en conditions expérimentales permettra de confirmer l'intérêt résolument clinique du logiciel CLAP.

Le travail de validation du logiciel CLAP est actuellement poursuivi. Dans un même temps, en collaboration avec l'équipe « mesures instrumentales », notre objectif était d'optimiser les protocoles d'analyse perceptive pour la comparaison avec les mesures obtenues grâce à l'appareillage EVA. Les études suivantes montrent les avancées que notre participation a permis d'apporter dans ce domaine.

e. Analyse objective de la dysphonie : protocole multiparamétrique incluant des mesures acoustiques et aérodynamiques.²

Yu P, Ouaknine M, Revis J, Giovanni A.

(Journal of Voice, 2001)

Ce travail s'inscrit dans la thématique « mesures instrumentales » du Laboratoire d'Audio-Phonologie Clinique. Les précédents travaux réalisés par notre équipes dans ce domaine ont permis le développement d'un protocole d'analyse instrumentale de plus en plus complet et fiable [Giovanni, 1992 ; Giovanni, 1995 ; Giovanni, 1996(a) ; Giovanni, 1996(b) ; Giovanni, 1998 ; Giovanni, 1999 ; Giovanni, 2000 ; Yu, 2000]. De grand progrès ont notamment été réalisés grâce à l'introduction de mesures de la pression sous-glottique estimée (PSGE) [Giovanni, 2000] pour le versant aérodynamique, ainsi que le développement d'un indice d'instabilité issue des mathématiques non linéaires, le coefficient de Lyapunov [Giovanni, 1999(a) ; Giovanni, 1999(b) ; Yu, 2000], pour le versant acoustique.

Dans un travail non publié, Alfonsi-Guarella avait réalisé une analyse multiparamétrique des dysphonies en introduisant les mesures de la PSGE [Alfonsi-Guarella, 1998]. Elle avait choisi un ensemble de mesures acoustiques réalisées pendant l'émission d'une voyelle tenue (Fo, intensité, shimmer, rapport signal/bruit), de mesures aérodynamiques (DAB, PSGE), associant l'étude de l'étendue vocale et celle du temps maximum de phonation (TMP). Elle avait limité l'étude aux dysphonies chez les femmes. A l'aide de ce protocole, il avait été possible de reclasser correctement 72% des sujets par rapport au grade perceptif G de l'échelle GRBAS.

Plus récemment, une étude réalisée par le Dr Yu intégrait un ensemble de 8 mesures instrumentales pour l'évaluation de la voix de 308 voix de femmes [Yu, 2001]. Ce protocole comprenait des mesures acoustiques (Fo, Jitter, Shimmer, coefficient de Lyapunov), des mesures aérodynamiques (Débit d'air buccal, PSGE), et des mesures physiologiques (Temps maximum de phonation, étendue vocale). L'analyse perceptive de ce corpus était réalisée pendant la consultation par un seul expert (orthophoniste ou médecin phoniatre) qui avait pour tâche d'évaluer le grade global de dysphonie sur une échelle ordinaire à 4 points. Les résultats de l'analyse discriminante montraient 58% d'adéquation entre l'évaluation instrumentale et l'analyse perceptive. En d'autres termes, dans seulement 58% des cas, l'appareillage donnait le même résultat que l'évaluation perceptive, ce qui est non seulement

² 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;15(4):529-542.

insuffisant, mais également en deçà des résultats obtenus dans l'étude précédente avec un protocole moins complet. Pour le Dr Yu, « *ce mauvais résultat est susceptible de provenir de la fiabilité de l'analyse perceptive. Rappelons-nous que dans ce travail, l'évaluation perceptive était faite par un seul auditeur, pas systématiquement le même, au cours de l'examen. La fiabilité de l'analyse perceptive par cette méthode doit être mise en question. Il apparaît de bon sens que l'erreur d'appréciation d'un clinicien isolé est nécessairement plus probable que celle d'un groupe d'experts utilisant une méthodologie rigoureuse d'analyse.*

 »

Dans son travail suivant, le Dr Yu avait pour objectif de vérifier que la combinaison des mesures acoustiques et aérodynamique permettrait de retrouver les résultats d'une analyse de la voix des patients effectuée par un jury d'experts, chez une population masculine. Tenant compte des difficultés rencontrées dans son précédent travail, le Dr Yu nous a confié la charge de la conception et de la réalisation de la partie perceptive du protocole expérimental.

Le but principal de ce travail était de valider un protocole instrumental. En cela, il était nécessaire de construire la méthodologie perceptive dans le sens d'une expertise, c'est à dire de fournir une analyse solide, s'affranchissant des problèmes de variabilité du jury d'écoute. En effet, l'évaluation de la fiabilité des jugements perceptifs eux-mêmes (étude des variabilités intra- et inter-individuelles) n'était pas l'objet ici. Nous avons donc proposé une analyse perceptive par consensus d'un jury composé de 6 experts.

Protocole expérimental : 63 sujets dysphoniques de sexe masculin ont été choisis dans la banque de données du service ORL du CHU Timone. L'étiologie de la dysphonie était dans 22 cas des polypes, dans 4 cas des nodules, dans 5 cas un œdème de Reinke, dans 3 cas des sulcus, dans 1 cas un kyste épidermique, dans 6 cas des papillomatoses, dans 7 cas une dysplasie, dans 11 cas une paralysie laryngée unilatérale, dans 4 cas une dysphonie dysfonctionnelle. La moyenne d'âge était 49 ans (23 à 75 ans). Nous avons également enregistré la voix de 21 sujets témoins masculin présentant une voix jugée normale par l'examinateur responsable des enregistrements et ne présentant aucun antécédent de troubles vocaux. Ces sujets étaient volontaires et ont été choisis dans l'environnement du laboratoire. La moyenne d'âge était de 36 ans (22 à 56 ans). Notons que certains de ces sujets ont rejoint le groupe de « dysphonies légères » après évaluation par le jury. Les examens des patients et enregistrements vocaux étaient effectués selon les procédures décrites dans la section « Protocole Expérimental ». Le matériau phonétique était constitué du premier paragraphe de « *La petite chèvre de Monsieur Seguin* » d'Alphonse Daudet.

Dans cette étude, nous avons recruté un jury d'écoute composé de six auditeurs expérimentés (2 médecins phoniatres et 4 orthophonistes) afin d'améliorer la fiabilité de l'évaluation perceptive. Il était

demandé au jury d'évaluer les échantillons de voix de lecture selon le paramètre G (Grade G : grade global de dysphonie), tiré de l'échelle GRBAS [Hirano, 1981] (voir section V, chapitre 1.3., p 83). Le paramètre G était jugé sur une échelle ordinaire à 4 niveaux (voir section V, chapitre 1.2., p 79). L'analyse était réalisée par consensus des différents membres du jury pour éviter les effets de la variabilité inter-individuelle (voir section V, chapitre 3.2., p 96). Cette analyse a fait l'objet d'une session d'écoute unique pour éviter les effets de la variabilité intra-individuelle. Les échantillons de voix étaient présentés selon un ordre aléatoire déterminé par un programme informatique et la note affectée était celle pour laquelle 4 membres du jury ou plus étaient d'accord.

Les mesures objectives ont été effectuées avec l'appareillage EVA® (SQ-Lab, Marseille) présenté en 1996 par Giovanni et Teston [Giovanni, 1996a]. Nous avons disposé de la version 1 de l'appareillage. Ce système permet des mesures simultanées acoustiques et aérodynamiques. L'ensemble des capteurs est regroupé au niveau d'une embouchure spécifique placée sur le visage du patient. La capture des signaux acoustiques était réalisée par le microphone coaxial AKG C409, contenu dans l'embouchure, placé à 5 cm de la bouche, en aval de la grille de tranquillisation du pneumotachographe. La capture des signaux aérodynamiques était effectuée avec le capteur du débit d'air constitué par un pneumotachographe. L'examineur demandait au sujet d'émettre une voyelle /a/ à une intensité confortable et à une hauteur la plus proche possible de sa hauteur de voix habituelle. Plusieurs essais étaient effectués jusqu'à ce que l'examineur retienne une émission qui lui semblait correspondre à la voix spontanée du sujet, observée pendant l'enregistrement de la voix de lecture. Cette émission était alors enregistrée en tant que fichier informatique. Systématiquement toutes les mesures ont été répétées trois fois et nous avons utilisé la valeur moyenne des trois résultats. Les données étaient recueillies à partir d'une fenêtre normalisée, d'une durée de deux secondes, placée 200 ms après l'attaque vocalique afin d'éliminer les phénomènes liés à celle-ci. Pour chacun des sujets, 9 mesures étaient retenues : Fo, Jitter, Shimmer, H/Nr, Coefficient de Lyapunov, DAB, PSGE, Etendue vocale, Temps Maximum de Phonation.

Toutes les données ont été retranscrites dans le logiciel Excel® puis transférées dans Systat®. Dans une première étape, nous avons réalisé une série d'analyses uni variées. Dans une seconde étape, nous avons réalisé une analyse multivariée à l'aide d'une méthode d'analyse discriminante. Elle nous a permis de retrouver les grades perceptifs attribués par le jury d'écoute (groupes) par une combinaison des paramètres objectifs (caractéristiques). A l'aide de Systat® 7.0 for Windows®, nous avons pu obtenir des coefficients de fonctions discriminantes de chaque groupe et établir une équation de discrimination pour chaque groupe.

Toutes les équations pouvaient être résumées dans la matrice suivante :

$$f_n = \begin{bmatrix} 1 & V_1 & V_2 & \dots & V_i \end{bmatrix} \begin{bmatrix} 2A_0 & B_1 & B_2 & \dots & B_j \\ 2C_{11} & C_{12} & \dots & C_{1j} \\ 2C_{21} & C_{22} & \dots & C_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ 2C_{ij} \end{bmatrix} \begin{bmatrix} 1 \\ V_1 \\ V_2 \\ \vdots \\ V_i \end{bmatrix}$$

Figure VI.e.1. : Matrice des coefficients de fonctions discriminantes. n = grade 0, 1, 2, 3 ; V : variables ; A : constante ; B et C : coefficients des fonctions discriminantes.

Résultats : La répartition des pathologies vocales en fonction du grade G attribué par le jury est donné dans le tableau VI.e.1.. Il est à noter que le groupe témoin, composé de 21 sujets, comprend 2 sujets qui ont été classés G1 par le jury. De même, trois sujets présentant une dysphonie dysfonctionnelle et un sujet présentant des nodules ont été classés G0 par le jury.

	Grade 0	Grade 1	Grade 2	Grade 3	Total
Témoins	19	2	-	-	21
Nodules	1	1	2	-	4
Polypes	-	12	9	1	22
Oedème de Reincke	-	1	4	-	5
Kystes	-	1	-	-	1
Sulcus	-	1	1	1	3
Papillomatose	-	2	4	-	6
Dysplasie	-	1	4	2	7
Dysphonie dysfonctionnelle	3	1	-	-	4
Paralysie laryngée	-	2	3	6	11
Total	23	24	27	10	84

Tableau VI.e.1. : Répartition des grades perceptifs affecté par le jury en fonction des groupes pathologiques.

Analyse uni-variée :

Le tableau VI.e.2. présente les résultats des mesures objectives en fonction du grade perceptif G. Nous avons calculé la moyenne et l'écart type.

	Grade 0	Grade 1	Grade 2	Grade 3
Fo (Hz)				
Moyenne	119	122	128	175
SD	20	29	28	19
Intensité (dB)				
Moyenne	90	94	91	89
SD	4.6	7.3	6.4	5.9
Jitter (%)				
Moyenne	1.0	2.07	3.44	10.70
SD	0.45	1.92	3.05	4.04
CL (bits/s)				
Moyenne	140	190	459	1626
SD	82	108	472	1320
SR _{f>0} (%)				
Moyenne	50.7	42.6	45.4	23.9
SD	16.5	11.3	11.7	12.8
SR _{f>1kHz} (%)				
Moyenne	11.2	10.3	8.8	3.8
SD	3.0	4.9	4.8	1.4
DAB (cm ³ /s)				
Moyenne	157	205	295	385
SD	45	109	110	202
PSGE (hPa)				
Moyenne	6.1	9.4	12.7	14.4
SD	1.3	3.1	4.2	5.5
Etendue (Hz)				
Moyenne	427	194	119	179
SD	168	129	74	82
TMP (s)				
Moyenne	21.0	13.5	9.2	4.3
SD	6.8	5.9	5.4	2.9

Tableau VI.e.2. : Valeurs des moyennes et des écart types (SD) en fonction des grades.

La répartition des variables en fonction du grade, est présentée sous forme de box-plot (figure VI.e.2.). Les graphiques mettent en évidence les zones de chevauchement plus ou moins importantes des valeurs en fonction des variables.

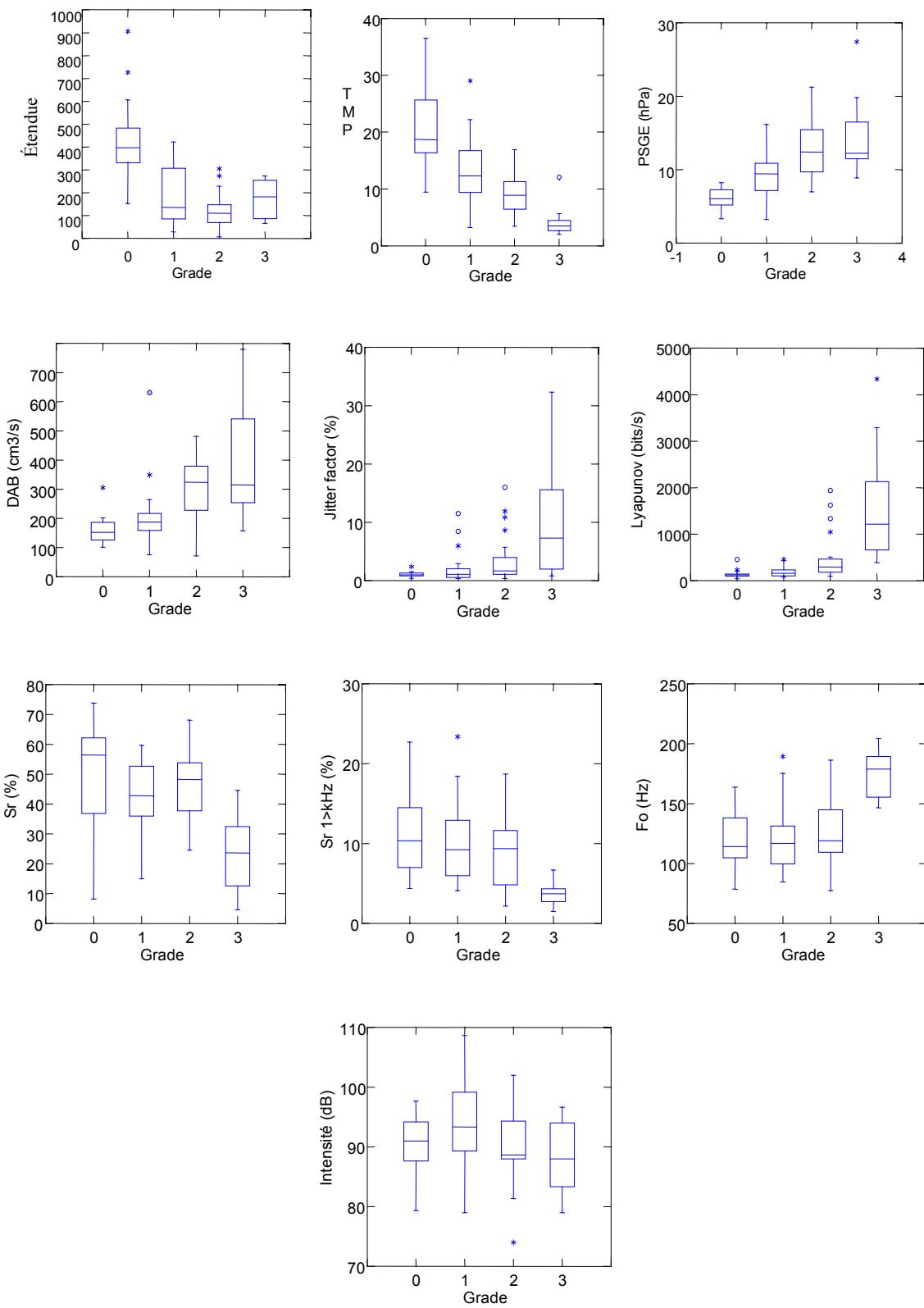


Figure VI.e.2. : Répartition des valeurs de 10 paramètres en fonction du grade.

Le tableau VI.e.3. montre le résultat des comparaisons entre les grades consécutifs (G0 contre G1, G1 contre G2, G2 contre G3). On observe que tous les paramètres, sauf l'intensité, sont significativement différents dans au moins une des études comparatives entre les grades adjacents.

	G0/G1	G1/G2	G2/G3
Jitter	0.509 (ns)	0.063 (ns)	0.031
SR _{f>0}	0.028	0.3931(ns)	0.001
SR _{f>1kHz}	0.509 (ns)	0.391 (ns)	0.007
DAB	0.038	0.001	0.374 (ns)
PSGE	0.001	0.005	0.549 (ns)
Etendue	0.001	0.050	0.054 (ns)
TMP	0.001	0.006	0.001
CL	0.115 (ns)	0.002	0.001
Fo	0.865 (ns)	0.322 (ns)	0.001
Intensité	0.142 (ns)	0.131 (ns)	0.304 (ns)

Tableau VI.e.3. : Valeur de p (test non-paramètre de Mann-Whitney comparant deux à deux les groupes perceptifs). "ns" signifie non-significatif.

Les résultats de l'analyse de la matrice de corrélation sont montrés dans le tableau VI.e.4. Nous avons conservé la valeur au carré du coefficient de Pearson (r^2), les valeurs de p sont entre-parenthèses. Tous les paramètres, à l'exception de l'intensité, sont significativement corrélés avec le grade perceptif.

Grade.	Fo	Etendu	TMP	Jitter	Intensité	DAB	PSGE	SR _{f>0}	SR _{f>1kHz}	CL
Fo	0.224 (0.002)	1.000								
Etendue	0.354 (0.001)	0.002 (1.000)	1.000							
TMP	0.531 (0.001)	0.097 (0.220)	0.227 (0.001)	1.000						
Jitter	0.220 (0.001)	0.069 (0.873)	0.050 (1.000)	0.142 (0.022)	1.000					
Int	0.012 (1.000)	0.109 (0.120)	0.003 (1.000)	0.004 (1.000)	0.089 (0.323)	1.000				
DAB	0.320 (0.001)	0.158 (0.010)	0.052 (1.000)	0.349 (0.001)	0.116 (0.085)	0.000 (1.000)	1.000			
PSGE	0.417 (0.001)	0.135 (0.001)	0.155 (0.011)	0.257 (0.001)	0.078 (0.557)	0.047 (1.000)	0.191 (0.002)	1.000		
SR _{f>0}	0.160 (0.009)	0.272 (0.001)	0.073 (0.697)	0.032 (0.176)	0.249 (0.001)	0.018 (1.000)	0.013 (1.000)	0.153 (0.013)	1.000	
SR _{f>1kHz}	0.153 (0.013)	0.054 (1.000)	0.092 (0.273)	0.097 (0.115)	0.085 (0.394)	0.071 (0.794)	0.061 (1.000)	0.085 (0.402)	0.292 (0.001)	1.000
CL	0.297 (0.001)	0.138 (0.028)	0.055 (1.000)	0.203 (0.001)	0.578 (0.001)	0.192 (0.002)	0.134 (0.035)	0.076 (0.628)	0.318 (0.001)	0.189 (0.002)

Tableau VI.e.4. : Matrice des corrélations entre les paramètres mesurés et le grade de dysphonie affecté par le jury. La valeur de p est donnée entre-parenthèses.

Analyse discriminante :

L'analyse discriminante a porté sur les 9 paramètres représentatifs du niveau de dysphonie (tableau VI.e.3.). La technique « pas à pas », a permis de faire émerger de l'ensemble 6 paramètres « utiles » : l'étendue vocale, le CL, le TMP, la PSGE, le rapport signal sur bruit total ($SR_{f>0}$) et la Fo. A partir de la combinaison de ces 6 paramètres, nous avons donc réalisé une analyse discriminante quadratique car certains paramètres ne présentaient pas une distribution « normale » et la covariance n'était pas identique pour tous les groupes. Au total, 86% des cas prédis par ce modèle de 6 paramètres étaient classés de la même façon par l'analyse perceptive (tableau VI.e.5.).

	Groupe 0	Groupe 1	Groupe 2	Groupe 3	% correct
Grade 0	22	1	0	0	96
Grade 1	2	20	2	0	83
Grade 2	0	5	20	2	74
Grade 3	0	0	0	10	100
Total	24	26	22	12	86

Tableau VI.e.5. : Table de prédiction des grades de dysphonie affectés par le jury. Les cas bien classés sont en caractère gras.

Le tableau VI.e.6. donne les valeurs de F obtenues en fonction des différents paramètres, c'est-à-dire leur contribution relative à la prédiction du grade (tableau VI.e.5.). Le CL est en deuxième place seulement après l'étendue vocale, ce qui confirme son importance dans l'analyse instrumentale des dysphonies.

Variables	F
Etendue	10.44
CL	8.54
PSGE	5.36
TMP	5.39
Fo	3.31
$SR_{f>0}$	2.07

Tableau VI.e.6. : Valeurs de F pour l'analyse discriminante.

Conformément à la méthode choisie (6 paramètres), nous avons établi 4 équation correspondant respectivement à chacun des niveaux dysphoniques (0 à 3) dans l'analyse discriminante. Ces quatre équations ont été appliquées à chaque patient, obtenant donc quatre valeurs correspondant aux 4 grades. Le patient a été classé dans le grade dont la valeur est la plus élevée. A partir de ces quatre valeurs absolues, nous avons calculé quatre valeurs de probabilité. Ces quatre chiffres indiquent, pour chaque patient, la probabilité d'être classé dans tel ou tel groupe. Le tableau VI.e.7., montre un exemple des résultats obtenus pour les voix de grade 2.

Patient	Grade perceptif	Probabilité				Grade objectif	Classement
		Grade 0	Grade 1	Grade 2	Grade 3		
1	2	0.01 (15.53)	0.37 (17.25)	0.62 (17.50)	0.00 (-26.30)	2	B
2	2	0.00 (12.67)	0.83 (17.41)	0.17 (16.62)	0.00 (-82.86)	1	M
3	2	0.00 (-5.49)	0.06 (14.49)	0.94 (15.90)	0.00 (9.41)	2	B
4	2	0.00 (7.85)	0.00 (11.6)	1.00 (14.82)	0.00 (-17.37)	2	B
5	2	0.00 (-22.4)	0.00 (10.43)	1.00 (15.14)	0.00 (-47.01)	2	B
6	2	0.00 (12.13)	0.28 (16.69)	0.72 (17.17)	0.00 (4.89)	2	B
7	2	0.00 (13.74)	0.26 (16.71)	0.74 (17.24)	0.00 (-2.24)	2	B
8	2	0.00 (10.24)	0.14 (16.34)	0.86 (17.24)	0.00 (-37.46)	2	B
9	2	0.00 (11.84)	0.09 (15.99)	0.91 (17.16)	0.00 (-30.66)	2	B
10	2	0.00 (4.92)	0.01 (15.35)	0.99 (17.66)	0.00 (-27.19)	2	B
11	2	0.09 (16.65)	0.78 (17.72)	0.13 (16.63)	0.00 (-98.54)	1	M

Tableau VI.e.7. : Classification d'analyse discriminante à l'aide d'une modèle à 6 paramètres objectifs (les chiffre entre-parenthèses sont les valeurs de chaque groupe). Sur la colonne de classement, B signifie "cas bien classé", M signifie "cas mal classé". Ce tableau ne concerne que les sujets G2.

Discussion : Les résultats obtenus dans ce travail confirment que les paramètres objectifs choisis varient de façon cohérente avec les paramètres perceptifs et qu'il est possible de retrouver le classement perceptif du jury à partir des mesures objectives.

A la suite d'une série d'analyses uni-variées, nous avons isolé les 9 paramètres les plus pertinents parmi les paramètres mesurés : d'une part, ils sont significativement différents au moins dans une des études comparatives entre grades adjacents ($p<0.05$) ; d'autre part, la matrice de corrélation montre que ces 9 paramètres sont en corrélation significative ($p<0.01$) avec le grade perceptif de dysphonie. Au total, chaque variable donne une indication du grade perceptif.

A partir de ces 9 paramètres, nous avons réalisé une analyse discriminante quadratique. L'analyse discriminante permet de déterminer plusieurs modèles de classification en combinant des mesures acoustiques et aérodynamiques. Cependant, il faut considérer qu'un modèle représentatif doit nécessairement répondre au compromis solidité de l'analyse statistique / concept physiopathologique : un modèle statistiquement parfait mais illogique sur le plan physiopathologique n'aurait aucun sens. Dans notre étude, nous avons utilisé 6 paramètres « utiles » : l'étendue vocale, le CL, la PSGE, le

TMP, le rapport signal sur bruit total ($SR_{f>0}$) et la Fo. L'association de ces 6 paramètres a permis d'obtenir 86% de concordance avec le jugement du jury d'écoute.

En comparaison avec les résultats obtenus chez les femmes par [Alfonsi-Guarella, non publié] (72% concordance entre les mesures objectives et perceptives), ce travail montre une importante amélioration de la corrélation entre l'analyse objective et perceptive (86% concordance entre ces deux méthodes). Cette observation peut être expliquée de deux façons : (1) l'introduction d'un nouvel indice acoustique - le coefficient de Lyapunov (CL) ; (2) l'utilisation d'une analyse discriminante originale : l'analyse discriminante de type quadratique.

Il peut être utile de souligner que dans notre protocole d'analyse statistique, le DAB et le jitter n'ont pas été considérés comme des paramètres utiles relatifs à l'analyse discriminante. L'introduction de ces deux paramètres dans le modèle de l'analyse discriminante n'améliore pas les résultats. Or, dans la littérature, le jitter et le DAB sont des deux paramètres largement utilisés pour décrire la voix pathologique, bien corrélés à l'intensité de la dysphonie [Hirano, 1988 ; Dejonckere, 1990 ; Hillenbrand, 1994]. Une analyse détaillée des rapports entre les paramètres mesurés (tableau VI.e.4.) permet d'envisager l'explication suivante : le DAB montre une forte corrélation avec le TMP et la PSGE. De même, le jitter avec le CL et le rapport signal sur bruit total ($SR_{f>0}$). On peut ainsi supposer que, dans notre approche de l'analyse discriminante, l'information « contenue » dans le DAB est représentée par le TMP et la PSGE, alors que l'information contenue dans le jitter est représentée par le CL et le rapport signal sur le bruit total. Les mesures du DAB et du jitter sont donc redondantes dans cette étude, ce qui ne remet évidemment pas en cause leur validité clinique.

Le caractère innovant de ce travail réside dans l'utilisation des probabilités obtenues par l'analyse discriminante pour différencier des dysphonies classées dans le même grade. En effet, les caractéristiques dysphoniques de deux voix classées dans le même grade G peuvent être différentes. Cela peut être lié à l'utilisation de l'échelle à 4 niveaux. Cette échelle, relativement grossière, n'est pas adaptée pour la discrimination de différences fines dans le degré de dysphonie. L'introduction de la probabilité relative à l'analyse discriminante permet une classification plus précise de la voix dysphonique : les patients du même groupe n'ont pas les mêmes valeurs de probabilité, leur degré de dysphonie est donc différent.

Les résultats que nous avons obtenus confirment l'intérêt d'un tel protocole multiparamétrique dans l'analyse des dysphonies. Cependant, dans ce travail, il existe encore certaines limites : (1) l'étude n'a pas inclus les patients après traitement (par exemple après la phonochirurgie) et nous ne savons pas si les paramètres objectifs sont aussi sensibles à l'appréciation de l'efficacité du traitement ; (2) nous

avons observé une difficulté relative à l'analyse discriminante pour les grade 1 et 2. Cette difficulté peut être liée à la variabilité observée dans l'analyse perceptive considérée comme modèle de référence. Dans la littérature, plusieurs études ont déjà montré que le jury d'écoute est davantage variable sur les dysphonies intermédiaires [Kreiman, 1992, 1993].

Au total, il semble que les modifications du protocole d'analyse perceptive que nous avions proposées au Dr Yu aient largement amélioré l'adéquation entre les mesures instrumentales et perceptives, puisque nous avons pu atteindre un score de 86% de variance commune. Néanmoins, à l'issue de ce travail, il nous semble encore possible d'améliorer les performances de l'analyse perceptive. En effet, la tâche d'évaluation par le jury doit tenir compte des exigences induites par la comparaison avec les mesures instrumentale et doit être conçue spécifiquement pour cela. Notamment, les mesures obtenues doivent être aussi précise que celles obtenues par une échelle visuelle analogique, tout en étant aussi reproductible et synthétique que celles obtenues par une échelle ordinaire. Cette réflexion commune nous a conduits à proposer une nouvelle étude permettant de tester une échelle de mesure perceptive originale, spécifiquement adaptée à la comparaison des mesures perceptives et instrumentales. Notre idée est qu'une telle échelle doit combiner les avantages de l'échelle visuelle analogique (permettant une discriminante fine et précise car toutes les valeurs intermédiaires sont disponibles) et d'une échelle ordinaire (moins soumise à la variabilité du jugement).

**f. Corrélation entre l'analyse instrumentale et l'évaluation perceptive :
utilisation d'une échelle visuelle analogique modifiée³**

Yu P, Revis J, Wuyts FL, Giovanni A.
(Folia Phoniatrica, 2002)

L'analyse de la voix par un jury d'experts est la méthode de référence universellement admise. La fiabilité des jugements d'auditeurs est donc d'une grande importance. Les résultats obtenus dans le travail précédent par analyse discriminante des mesures instrumentales, montraient toujours une insuffisance de corrélation avec l'analyse perceptive. Cette insuffisance était essentiellement liée à une difficulté de discrimination des grades 1 et 2. Ces résultats ont ouvert le débat sur les méthodologies de l'analyse perceptive et ses sources de variabilité, avec en particulier le choix de l'échelle quantitative. En effet, il semble qu'une inadéquation entre l'échelle utilisée et l'acuité perceptive des auditeurs peut être responsable de ces difficultés de discrimination. Le but de ce travail est de concevoir une échelle quantitative d'évaluation perceptive spécifiquement adaptée à la comparaison avec des mesures instrumentales. Nous proposons une échelle mixte (modified Visual Analogic Scale « mVAS ») réalisée d'après une combinaison originale l'échelle visuelle analogique en une échelle de classe à 4 niveau. Notre objectif est de déterminer quelle est la méthodologie perceptive la plus performante dans la comparaison avec les mesures instrumentales. Pour cela, nous proposons de comparer les possibilités de corrélation entre les mesures objectives d'une part, et l'échelle perceptive d'autre part (échelle de classe et échelle visuelle analogique modifiée).

Protocole expérimental : L'étude a été menée sur 74 échantillons de voix féminines correspondant à 68 patientes dysphoniques et 6 sujets témoins. Leur moyenne d'âge était de 40 ans (21 à 68 ans). Les pathologies vocales étaient des nodules dans 26 cas, un polype dans 11 cas, un œdème de Reinke dans 13 cas, une paralysie laryngée unilatérale dans 6 cas, des kystes dans 6 cas, une dysphonie dysfonctionnelle simple dans 4 cas et 2 cas post phonochirurgie, ainsi que 6 sujets témoins féminins présentant une voix jugée normale par l'examinateur et aucune plainte vocale ni aucun antécédent de dysphonie. L'examen des patientes et l'enregistrement des voix étaient effectués selon les procédures décrites dans la section « Protocole expérimental » (section V, chapitre 1, p 105-114). Le matériau phonétique choisi était les deux premières phrases du premier paragraphe de « La petite chèvre de Monsieur Seguin » d'Alphonse Daudet.

³ Yu P, Revis J, Wuyts FL, Zanaret M, Giovanni A. « Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale. » Folia Phoniatr Logop. 2002;54(6):274-281.

Le jury d'écoute était composé de 4 auditeurs expérimentés (un médecin phoniatre et 3 orthophonistes), amenés à évaluer des voix dans la pratique quotidienne de leur profession et ayant déjà participé à des études perceptives de la voix.

Nous avons demandé au jury d'écouter d'évaluer la voix selon les paramètres G (grade global), R (raucité) et B (souffle) de l'échelle GRBAS. Dans ce travail, nous avons retenu seulement le paramètre G. Les jugements perceptifs étaient recueillis selon deux méthodologies distinctes, visant à tester l'échelle quantitative utilisée :

- Le premier protocole était réalisé selon une échelle traditionnelle de classe à 4 points d'intervalles équidistants (EAI) où 0 représentait une voix normale et 3 une dysphonie sévère. L'évaluation était faite par consensus des différents membres du jury pour éviter les effets de la variabilité inter-individuelle. La note affectée était celle pour laquelle 3 membres du jury au moins étaient d'accord. L'évaluation des voix réalisée au moyen de ce premier protocole n'a fait l'objet que d'une session d'écoute. En effet, les évaluations par consensus constituent une expertise, supposée indiscutable, puisque les différents experts argumentent leurs choix jusqu'à commun accord. Compte tenu de l'objectif de cet étude, il n'est donc pas nécessaire de tester la variabilité inter-individuelle.
- Le second protocole était réalisé selon une échelle visuelle analogique. Les feuillets de cotation présentaient une droite de 10 cm. Le 0 (à gauche) représentait la normalité et l'extrême de droite le niveau pathologique maximum. Les auditeurs avaient pour tâche de représenter au moyen d'une marque sur la droite, le niveau pathologique correspondant à chacune des voix. Les résultats étaient recueillis grâce à un double décimètre, la note attribuée correspondant à la distance (en millimètres) entre le 0 et la marque apposée par l'auditeur. Les résultats étaient donc compris entre 0 et 100.

Les évaluations perceptives selon ce second protocole ont fait l'objet de trois sessions d'écoute séparées d'une semaine. Les 74 enregistrements de voix étaient présentés dans un ordre aléatoire différent lors de chaque session afin d'évaluer la variabilité intra-individuelle des auditeurs et de limiter l'effet de contexte. Au total, 12 notes étaient donc attribuées à chaque voix (4 auditeurs*3 sessions = 12). Dans le cas d'une variable quantitative, la moyenne est très sensible à un résultat de mesure extravagant. Pour éviter l'influence sur la moyenne de l'émergence de valeurs extrêmes, nous avons enlevé la valeur la plus basse et la valeur la plus élevée (cette méthode est largement utilisée dans plusieurs domaines, par exemple le concours de chant ou compétition sportive), et calculé la moyenne sur les 10 notes restantes pour chaque voix afin d'obtenir une tendance générale.

Puis nous avons calculé le résultat de mVAS : l'échelle mixte. Cette échelle mixte nous permettait de combiner les avantages de l'échelle visuelle analogique (finesse de discrimination) avec ceux de l'échelle de classes (résultats compacts). Nous avons donc regroupé les voix en 4 classes, en fonction de la moyenne de chaque voix calculée sur 10 évaluations, en utilisant la méthode présentée ci-dessous :

Echelle visuelle analogique (cm)	Echelle mixte
0.0 - 0.9	Grade 0
1.0 - 5.0	Grade 1
5.1 - 9.0	Grade 2
9.1 – 10	Grade 3

Tableau VI.f.1. : transformation d'une échelle visuelle analogique en une échelle de classe à 4 niveaux.

Les mesures objectives ont été effectuées avec l'appareillage EVA®. La technique des enregistrements de la voix était strictement identiques à la technique utilisée précédemment.

Le traitement statistique de notre étude a été réalisé en trois étapes :

- Validité du jury d'écoute sur l'échelle mixte : elle est explorée par une étude de la variabilité intra- et inter-auditeur de notre jury. L'outil statistique utilisé est le coefficient Kappa Cohen qui permet la meilleure mesure de variabilité pour des données ordinaires. Afin de faciliter l'interprétation des résultats, nous proposons une table des valeurs coefficients Kappa [De Bodt, 1997].

Valeur du coefficient Kappa	Degré de fiabilité
<0.20	Faible
0.21 – 0.40	Raisonnables
0.41 – 0.60	Assez bon
0.61 – 0.80	Bon
0.81 – 1.00	Très bon

Tableau VI.f.2. : table d'interprétation des coefficients Kappa.

- Etude de la variabilité des jugements perceptifs, réalisés sur l'échelle visuelle analogique, à l'aide de calcul de la moyenne des écarts types en fonction du niveau de dysphonie.
- Comparaison des adéquations obtenues entre les résultats de l'analyse objective et les résultats de l'analyse perceptive par échelle de classe d'une part, et entre les résultats de l'analyse objective et les résultats de l'analyse perceptive par échelle mixte d'autre part, à l'aide d'une série de méthodes univariées et de méthodes multivariées.

Résultats : Validité du jury d'écoute : le tableau VI.f.3. et le tableau VI.f.4. montrent respectivement les résultats de la variabilité intra-auditeur et de la variabilité inter-auditeurs pour l'échelle mixte. Notons que toutes les valeurs de coefficient Kappa sont rangées de assez bon à bon (d'après tableau VI.f.2.).

Kappa ± SE	Echelle mixte
Juge1	0.518 ± 0.047
Juge2	0.524 ± 0.046
Juge3	0.655 ± 0.045
Juge 4	0.530 ± 0.044

Tableau VI.F.3. : Valeurs de coefficient Kappa intra-juge sur le paramètre G pour l'échelle mixte.

Kappa ± SE	Echelle mixte
Session 1	0.560 ± 0.031
Session 2	0.598 ± 0.033
Session 3	0.591 ± 0.036

Tableau VI.f.4. : Valeurs de coefficient Kappa inter-juge sur le paramètre G pour l'échelle mixte.

Validité du jugement perceptif en fonction du niveau de dysphonie : afin d'étudier la variabilité des jugements perceptifs sur l'échelle mixte, nous avons d'abord calculé la moyenne des 10 notes de chaque voix et son écart type, puis nous avons calculé la moyenne des écarts types en fonction du niveau de dysphonie (tableau VI.f.5.). Notons que les valeurs recueillies sur les grades 1 et 2 montrent une variabilité plus importante que celles recueillies sur les grades 0 et 3.

Echelle mixte	Moyenne de SD
Grade 0	0.338
Grade 1	1.106
Grade 2	0.891
Grade 3	0.399

Tableau VI.f.5. : valeurs moyennes d'écart type en fonction de grade perceptif.

Analyse objective : les tableaux VI.f.6. et 7. présentent les résultats de l'analyse descriptive en fonction du grade perceptif, respectivement en fonction de l'échelle mixte (tableau VI.f.6.) et de l'échelle traditionnelle de classe (tableau VI.f.7.). Nous observons que tous les paramètres, sauf le paramètre de Fo et d'intensité, manifestent une tendance cohérente par rapport au grade de dysphonie.

	Grade 0 (n:11)	Grade 1 (n:30)	Grade 2 (n:25)	Grade 3 (n:8)
Fo (Hz)				
Moyenne	217	209	193	201
SD	38.8	23.5	40	47.7
Intensité (dB)				
Moyenne	89	88	89	90
SD	5.5	4.0	4.6	4.0
Jitter (%)				
Moyenne	0.40	0.52	1.1	4.8
SD	0.06	0.1	1.35	7.6
CL (bits/s)				
Moyenne	80	113	279	636
SD	47	48	234	555
SR _{f>0} (%)				
Moyenne	60.8	59	55.6	35.6
SD	11.4	13.7	17.4	18.4
SR _{f>1kHz} (%)				
Moyenne	24.6	17.9	16.5	6.6
SD	6.5	8.5	7.9	2.8
DAB (cm ³ /s)				
Moyenne	175	185	221	318
SD	75	84	132	172
PSGE (hPa)				
Moyenne	7.1	7.3	10.0	13.1
SD	1.5	2.2	3.7	3.9
Etendue (Hz)				
Moyenne	481	329	186	184
SD	178	108	65	73
TMP (s)				
Moyenne	14.4	9.8	7.6	4.6
SD	6.2	4.0	2.7	1.9

Tableau VI.f.6.. Valeurs des moyennes et des écarts types (entre parenthèse) en fonction des grades perceptifs pour l'échelle mixte.

	Grade 0 (n:11)	Grade 1 (n:30)	Grade 2 (n:25)	Grade 3 (n:8)
Fo (Hz)				
Moyenne	211	212	196	200
SD	33	26	36	50
Intensité (dB)				
Moyenne	91	88	89	90
SD	5.6	3.6	4.7	4.0
Jitter (%)				
Moyenne	0.42	0.51	0.88	3.7
SD	0.08	0.16	1.2	6.37
CL (bits/s)				
Moyenne	103	110	223	525
SD	42	55	225	486
SR _{f>0} (%)				
Moyenne	55.3	61.0	57.1	41.6
SD	7.7	14.8	16.0	20.5
SR _{f>1kHz} (%)				
Moyenne	22.6	19.0	17.0	9.9
SD	6.8	8.4	8.5	7.5
DAB (cm ³ /s)				
Moyenne	152	199	184	332
SD	73	83	108	157
PSGE (hPa)				
Moyenne	7.1	7.9	8.9	12.0
SD	1.7	2.6	3.7	4.0
Etendue (Hz)				
Moyenne	420	287	215	170
SD	205	130	85	71
TMP (s)				
Moyenne	13.6	9.9	8.8	5.0
SD	6.8	4.3	3.0	2.6

Tableau VI.f.7. : Valeurs des moyennes et des écarts types (entre parenthèse) en fonction des grades perceptifs pour l'échelle traditionnelle de classe.

Nous avons réalisé une étude comparative des résultats obtenus avec nos 2 échelles perceptives avec analyse objective. Les comparaisons ont été faites entre deux grades perceptifs contigus : G0 contre G1, G1 contre G2, G2 contre G3 en utilisant le test non paramétrique de *Mann-Withney*, la valeur critique de rejet de l'hypothèse nulle était fixée à $p < 0.05$. Nous observons que la répartition des paramètres significatifs est meilleure pour l'échelle mixte (tableau VI.f.8.), que pour l'échelle traditionnelle de classe (tableau VI.f.9.).

	G0/G1	G1/G2	G2/G3
Jitter	0.01	0.007	0.034
SR _{f>0}	0.648 (ns)	0.521 (ns)	0.017
SR _{f>1kHz}	0.017	0.636 (ns)	0.001
DAB	0.566 (ns)	0.648 (ns)	0.141 (ns)
PSGE	1.000 (ns)	0.001	0.048
Etendue	0.002	0.000	0.966 (ns)
TMP	0.008	0.062 (ns)	0.010
CL	0.032	0.000	0.071 (ns)
Fo	0.537 (ns)	0.072 (ns)	0.933 (ns)
Intensité	0.546 (ns)	0.600 (ns)	0.303 (ns)

Tableau VI.f.8. : Valeurs de p de l'échelle mixte. "ns" signifie le paramètre étant non significatif entre deux grades perceptifs.

	G0/G1	G1/G2	G2/G3
Jitter	0.072 (ns)	0.037	0.001
SR _{f>0}	0.199 (ns)	0.373 (ns)	0.031
SR _{f>1kHz}	0.226 (ns)	0.378 (ns)	0.004
DAB	0.140 (ns)	0.255 (ns)	0.008
PSGE	0.496 (ns)	0.312 (ns)	0.019
Etendue	0.019	0.000	0.226 (ns)
TMP	0.051 (ns)	0.563 (ns)	0.001
CL	0.650 (ns)	0.047	0.006
Fo	0.734 (ns)	0.196 (ns)	0.637 (ns)
Intensité	0.248 (ns)	0.707 (ns)	0.261 (ns)

Tableau VI.f.9. : Valeurs de p de l'échelle traditionnelle de classe "ns" signifie le paramètre étant non significatif entre deux grades perceptifs.

Une analyse discriminante a été réalisée sur les 8 variables les plus représentatives du niveau de dysphonie : étendue vocale, PSGE, CL, jitter, TMP, DAB, SR_{f>0} et SR_{f>1kHz}. Nous avons réalisé deux analyses discriminantes respectivement pour l'échelle mixte et l'échelle traditionnelle de classe. Nous avons trouvé 6 variables pertinentes et communes : étendue vocale, PSGE, CL, TMP, DAB et SR_{f>1kHz}. La combinaison de ces 6 variables permet de reclasser 88% des sujets dans le grade affecté par le jury avec l'échelle mixte (tableau VI.f.10.), et 64 % des sujets dans le grade affecté par le jury avec l'échelle traditionnelle de classe (tableau VI.f.11.). La validité de l'analyse statistique est confirmée par le test *Lambda de Wilk* ($p < 0.001$).

	Groupe 0	Groupe 1	Groupe 2	Groupe 3	% correct
Grade 0	10	1	0	0	91
Grade 1	3	25	2	0	83
Grade 2	0	3	22	0	88
Grade 3	0	0	0	8	100
Total	13	29	24	8	88

Tableau VI.f.10. : Table de prédiction des grades de dysphonie affectés par le jury avec l'échelle mixte. Les cas bien classés sont en caractère gras.

	Groupe 0	Groupe 1	Groupe 2	Groupe 3	% correct
Grade 0	8	2	0	0	80
Grade 1	8	15	1	0	63
Grade 2	2	10	13	3	46
Grade 3	0	1	0	11	92
Total	18	28	14	14	64

Tableau VI.f.11. : Table de prédiction des grades de dysphonie affectés par le jury avec l'échelle traditionnelle de classe. Les cas bien classés sont en caractère gras.

Discussion : Les résultats de cette étude montrent que la corrélation entre les mesures objectives et l'échelle perceptive est meilleure avec l'échelle mixte qu'avec l'échelle de classe.

L'observation des résultats indique que la principale faiblesse des protocoles d'évaluation concerne surtout la discrimination des dysphonies intermédiaires : G1 et G2. Cette imprécision pourrait être attribuée au protocole d'analyse perceptive, en faisant référence au modèle interne décrit par l'équipe de UCLA [Kreiman, 1993 ; Gerratt, 1993] : dans la pratique clinique quotidienne, tous les auditeurs (expérimentés ou non) ont une expérience considérable et approximativement identique des voix normales (que l'on est tous habitué à entendre dans la vie quotidienne). Le prototype « voix normale » est donc stable indépendamment du niveau d'expérience [Askenfelt, 1986 ; Anders, 1988 ; De Bodt, 1997 ; Kreiman, 1992 et 1993]. Dans le cas de dysphonie, la fiabilité des auditeurs augmente avec la sévérité de la pathologie [Rabinov, 1995 ; Anders, 1988]. Une dysphonie sévère correspond à la présence pathologique maximale, il est donc facile de l'identifier et peut être même de façon universelle. Ainsi, la distinction voix normale/voix anormale ne présente pas de difficulté et n'est pas dépendante du niveau d'expérience des auditeurs. En revanche, toute la difficulté de l'analyse perceptive réside dans la classification des voix en différents niveaux pathologiques et en particulier dans la discrimination des dysphonies intermédiaires : dysphonies légères (G1) et dysphonies moyennes (G2). Dans ces cas, les évaluations des auditeurs, même expérimentés, varient largement. Cette variabilité d'évaluation perceptive est liée aux standards internes qui sont par définition instables.

Il n'existe pas d'échelle de valeur interne permettant d'établir avec précision et de manière fiable une hiérarchie pour discriminer les différentes types dysphonies. Pourtant, la définition d'un protocole d'analyse perceptive nécessite de faire un choix quant à la quantification de la dysphonie. A l'heure actuelle, aucune des deux méthodes classiques de quantification (l'échelle de classe et l'échelle visuelle analogique) ne semble résoudre cette difficulté de discrimination des dysphonies intermédiaires. C'est pourquoi nous avons tenté d'introduire cette échelle mixte d'évaluation perceptive.

C'est sur la base de cette réflexion et sur l'observation de notre jury d'expert que nous avons conçu notre échelle mixte. En effet, plusieurs solutions étaient envisageables en ce qui concerne le découpage des résultats obtenus avec l'échelle visuelle analogique en 4 classes distinctes. Une solution était de faire un classement équidistant, par une division arithmétique par 4. Le découpage se faisant alors de la façon suivante : toutes les notes comprises entre 0 et 2.5 devenaient grade 0, les notes comprises entre 2.5 et 5 devenaient grade 1, les notes comprises entre 5 et 7.5 devenaient grade 2 et les notes comprises entre 7.5 et 10 devenaient grade 3. C'est la solution choisie par Wuyts dans une étude publiée récemment sur la fiabilité du jugement perceptif [Wuyts, 1999]. Il a demandé à un jury d'écoute composé de 29 auditeurs, de faire une évaluation perceptive en utilisant deux échelles : une échelle de classe à 4 niveau et une échelle visuelle analogique. Il a ensuite transformé l'échelle visuelle analogique en deux échelles de classes distinctes : une échelle de classes à 6 niveaux et une échelle de classes à 4 niveaux, toutes deux présentant des points d'intervalle égaux. Il a comparé la fiabilité de ces trois échelles : échelle de classe traditionnelle, échelle mixte à 4 niveaux et échelle mixte à 6 niveaux, en calculant les coefficients Kappa propres à chaque échelle. Les résultats montraient que, globalement, la variabilité inter-individuelle était un peu moins importante pour l'échelle de classe que pour les échelles mixtes. Sa conclusion était que la variabilité inter-individuelle des auditeurs augmente avec les degrés de liberté fournis par l'échelle quantitative. En d'autre termes, plus le nombre de niveaux pathologiques proposés par l'échelle quantitative est grand, plus la variabilité du jury augmente. Ceci vient confirmer les résultats obtenus par Gerratt [Gerratt, 1993]. D'autre part, les résultats de Wuyts ne permettaient pas de confirmer l'intérêt d'une échelle mixte à niveaux équidistants.

L'originalité de notre méthode vient de la définition de classes non équidistantes sous tendue par le modèle de référent interne décrit par Kreiman [Kreiman, 1993]. Notre conception du regroupement des valeurs de l'échelle visuelle analogique en 4 classes répondait à deux observations : (1) le jury identifie facilement les deux extrêmes de l'échelle, voix normale (G0) / dysphonie sévère (G3), et (2) présente une imprécision dans la discrimination des valeurs intermédiaires. Ces observations nous ont conduit à penser que les intervalles correspondant aux niveaux extrêmes (G0 et G3) devaient être réduits puisque le jury les discrimine de façon très précise, par rapport aux intervalles correspondant

aux niveaux intermédiaires (G1 et G2) qui devaient proposer un intervalle de confiance plus large puisque le jury présente davantage d'imprécision. Nous avons donc construit notre échelle mixte de la façon suivante : toutes les notes comprises entre 0 et 1 devenaient grade 0., les notes comprises entre 1 et 5 devenaient grade 1., les notes comprises entre 5 et 9 devenaient grade 2 et les notes comprises entre 9 et 10 devenaient grade 3. Ainsi, sur la base du caractère instable du standard interne et d'une importante variabilité du jugement perceptif concernant les dysphonies intermédiaires, il nous a semblé judicieux de restreindre l'échelle de choix pour le grade 0 et grade 3, tandis que nous avons agrandi l'échelle de choix pour le grade 1 et le grade 2 (tableau VI.f.1.). Nous avons ensuite étudié la fiabilité de l'échelle mixte obtenue. Nous avons calculé le coefficient Kappa qui est l'outil statistique le plus performant pour la mesure de variabilité des données ordinales. Les résultats montrent que les constances intra- et inter-auditeurs sont satisfaisantes (tableau VI.f.3. et 4.) : toutes les valeurs de coefficient Kappa sont classées de assez bon à bon. Cette vérification nous a permis de poursuivre notre étude par la comparaison entre l'échelle mixte et l'échelle de classe utilisées comme référence pour tester l'analyse objective.

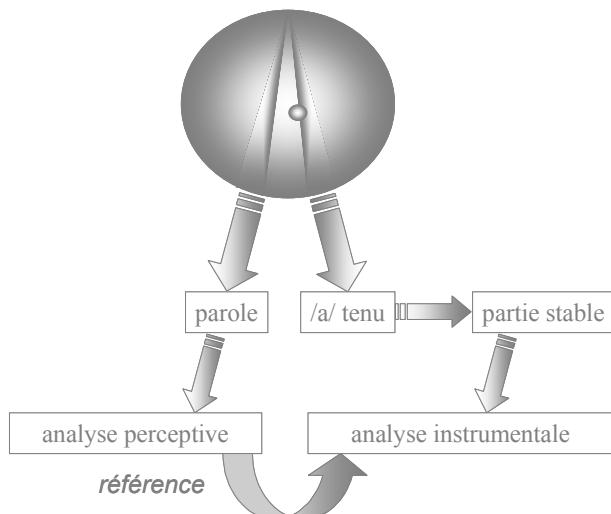
L'analyse comparative des résultats de l'analyse instrumentale avec une échelle de classes traditionnelle d'une part, et des résultats de l'analyse instrumentale avec notre échelle mixte d'autre part, confirment l'intérêt de l'utilisation d'une échelle mixte. En effet, grâce à l'utilisation de l'échelle mixte les paramètres objectifs deviennent plus performants qu'avec l'échelle traditionnelle de classe. A l'aide d'une combinaison de 6 paramètre pertinents, nous avons retrouvé le même classement des voix dysphoniques dans 88% cas avec l'échelle mixte, contre 64% avec l'échelle de classe. De plus, les résultats indiquent que l'échelle mixte permet une discrimination plus précise que l'échelle de classe, surtout pour les grades 1 et 2.

Au total, l'utilisation de l'échelle mixte semble améliorer à la fois les performances des paramètres instrumentaux, la discrimination spécifique des dysphonies intermédiaires et l'adéquation entre analyse perceptive et analyse instrumentale. L'utilisation de l'échelle mixte semble donc mettre les deux systèmes de mesure à leur meilleur niveau de performance. La méthode d'échelle mixte semble donc apporter une solution pour améliorer l'adéquation entre les mesures perceptives et objectives.

Au total, les travaux de la section **perception / mesures** que nous venons de présenter montrent l'importance du soin accordé à la méthodologie perceptive pour l'adéquation avec les mesures instrumentales. La collaboration entre les différentes équipes du laboratoire d'Audio-Phonologie Clinique a abouti à la définition d'un protocole mixte permettant d'obtenir 88% de variance commune entre les deux systèmes de mesure. Le cahier des charges de l'appareillage EVA ® défini conjointement avec le laboratoire Parole et Langage (CNRS – Aix en Provence) était de parvenir à une adéquation de 90% environ. Néanmoins, il faut considérer que les calculs statistiques permettent de définir la corrélation à posteriori à partir d'un corpus de patients défini à l'avance et que, pour autant, cela ne signifie pas que les résultats obtenus soient aussi performants pour les nouveaux patients inclus dans le groupe expérimental. La recherche dans ce domaine doit continuer afin que notre appareillage puisse acquérir le statut d'outil clinique, permettant de fournir des résultats fiables pour chaque patient, pris indépendamment. C'est l'objectif poursuivi par l'équipe « mesures instrumentales » de notre laboratoire, dirigée par le professeur Giovanni.

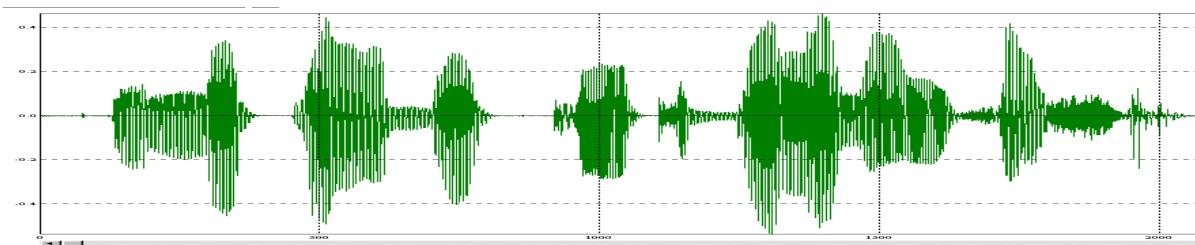
En ce qui concerne la démarche perceptive, il semble que notre protocole d'étude soit optimisé pour la comparaison avec les données instrumentales. Néanmoins, à ce stade, le mode d'enregistrement des voix et particulièrement le matériau phonétique utilisé dans la démarche analytique, reste un point obscur, tant dans le développement méthodologique que dans la compréhension du phénomène dysphonique. En effet, comme nous l'avons souligné dans la section « Méthodologies de l'analyse perceptive » (section V, chapitre 4, p 98-103), les exigences techniques des appareillages de mesure imposent l'utilisation d'un /a/ tenu, alors que les jurys d'écoute évaluent la dysphonie sur de la parole. Bien que produite par le même patient, la voix proposée aux deux systèmes d'évaluation diffère tant au niveau conceptuel que dans la quantité d'information constitutive de l'échantillon. Cette dichotomie nous a poussés à développer une réflexion concernant les dimensions phonétiques de la dysphonie que nous présentons maintenant dans la section **Perception / Phonétique**.

A l'heure actuelle, l'exploration de la dysphonie suit le schéma suivant : la voix est produite par un patient qui possède un larynx « pathologique », symbolisé ici par un nodule sur la corde vocale droite. L'examineur demande au sujet de produire de la parole (texte lu) et un /a/ tenu afin de procéder à l'enregistrement de la voix. L'échantillon de parole est mis à la disposition d'un jury d'écoute pour l'analyse perceptive. Le /a/ tenu est traité



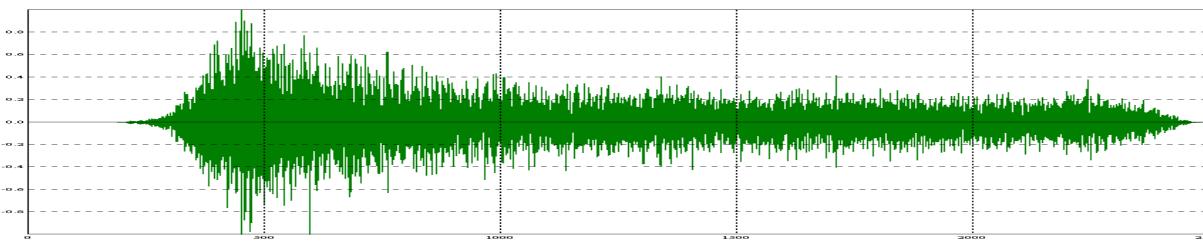
par informatique afin d'en extraire la partie stable qui va être analysée par un appareillage de mesures. L'analyse perceptive est la référence pour la comparaison des résultats obtenus par les deux systèmes de mesure. Or, même si l'on exclut la charge sémantique des productions, la quantité d'information contenue dans ces matériaux phonétiques est largement inéquitable :

- La parole est le matériau phonétique le plus riche en informations. Il s'agit d'un signal complexe qui est représentatif de tous les aspects de la voix en tant qu'outil de communication. La parole rend compte des phénomènes phonétiques (coarticulation, transitions formantiques, attaques, interruptions vibratoires...), et linguistiques (prosodie, rythme, intention...). La succession des phonèmes est extrêmement rapide, de l'ordre de quelques millisecondes, c'est pourquoi il est souvent considéré qu'il n'y a pas de parties stables (par opposition à la voyelle tenue sur plusieurs secondes). Pour Hammarberg [Hammarberg, 2000], « *les variations survenant dans la parole, comme l'attaque vocale, l'arrêt vibratoire, les cassures de la phonation, etc., sont des éléments cruciaux de la qualité de la voix.* »



Signal acoustique de parole (Phonédit ®)

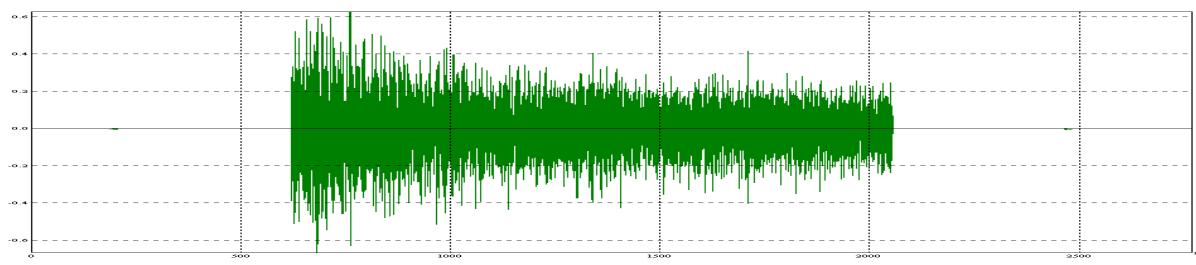
- Le /a/ tenu est un matériau phonétique « naturel », c'est à dire produit par le patient sans intervention de traitement informatique du signal. Cet échantillon vocal est réduit à la production d'un voisement dans une position articulatoire relativement neutre. Il contient une information phonétique réduite : une attaque, une partie stable, et une interruption vibratoire. Il ne contient aucune information linguistique.



Signal acoustique de /a/ tenu (Phonédit ®)

- La partie stable du /a/ tenu peut être considéré comme un échantillon vocal « artificiel » dans la mesure il est nécessaire de procéder à un traitement informatique du signal afin de supprimer

l'attaque et l'interruption vibratoire du /a/ produit par le patient. La partie stable du /a/ tenu contient une information phonétique pauvre, réduite au régime vibratoire stable.



Signal acoustique de partie stable d'un /a/ tenu (Phonédit ®)

Malgré ces différences, il faut considérer que chacun de ces trois matériaux présentent leurs avantages. La parole, bien sûr, est la plus représentative de la plainte du patient. Cependant, le /a/ tenu permet d'évaluer la stabilité du signal dans la durée et de déterminer la stratégie d'attaque du patient (attaque normale, dure ou soufflée). Il est d'ailleurs couramment utilisé comme examen complémentaire en pratique clinique. Pour Parsa [Parsa, 2001], « *les voyelles tenues sont représentatives de la voix chantée.* » Enfin la partie stable de la voyelle tenue est le seul matériau toléré par les appareillages de mesure, précisément pour ses caractéristiques épurées. Si les spécialistes des mesures instrumentales se retrouvent pour dire qu'il serait idéal de pouvoir utiliser la parole pour les mesures objectives, aucun appareillage n'est en mesure à l'heure actuelle de proposer cette possibilité.

En revanche, le jury d'écoute semble plus flexible. Dans le souci de parvenir à uniformiser le matériau phonétique, et en l'absence de la possibilité d'utiliser la parole pour l'analyse instrumentale, nous nous sommes demandé si le jury d'écoute ne serait pas capable, lui, de pratiquer son évaluation sur la partie stable du /a/ tenu, ou à défaut, sur le /a/ tenu complet. Cette réflexion a fait l'objet de notre première publication sur le thème **perception / phonétique**.

*g. Comparaison de différents matériaux phonétiques pour l'analyse perceptive des dysphonies.*⁴

Revis J, Giovanni A, Wuyts FL, Triglia JM
 (Folia Phoniatri Logop, 1999)

En 1994, De Krom [De Krom, 1994] a rapporté une expérience d'analyse perceptive visant à mettre en évidence que la voyelle tenue peut être un matériau pertinent à condition d'y inclure le transitoire d'attaque (voyelle tenue «complète»). Un jury de 6 auditeurs devait évaluer le grade global de dysphonie, la sensation de souffle, et la raucité des fragments de voix recueillis chez 78 sujets sur une échelle ordinaire à 10 niveaux. Il a ainsi montré que la constance d'un jury était plus grande sur la parole ou la voyelle tenue complète, que sur la partie stable de la voyelle tenue.

Le travail prospectif que nous présentons a pour but d'analyser l'influence du choix de l'échantillon phonétique sur les évaluations du jury. Le premier objectif est de vérifier si nous retrouvons les résultats de De Krom [De Krom, 1994], car l'application à la langue française de cette étude menée en langue néerlandaise nécessite une validation expérimentale du fait des différences culturelles susceptibles d'interférer avec les résultats [Hammarberg, 1980]. Le deuxième objectif est de démontrer que, en cas de dysphonie, les jugements portés sur la partie stable des voyelles tenues sont moins sévères que ceux portés sur la parole. En effet, nous supposons que certaines informations acoustiques présentes dans les transitoires d'attaque sont manquantes sur la partie stable de la voyelle et que ce type d'échantillon fait courir le risque d'une sous-évaluation de la dysphonie.

Protocole expérimental : L'étude a été menée sur 80 échantillons de voix correspondant à 60 patients dysphoniques et 20 sujets témoins. Il s'agissait de 26 hommes et 54 femmes âgés de 19 à 83 ans. Les pathologies présentées par les patients étaient des nodules dans 11 cas, un polype angiomeux dans 10 cas, un kyste épidermique dans 3 cas, un sulcus glottidis dans 4 cas, une paralysie laryngée unilatérale dans 10 cas, un oedème de Reinke dans 9 cas, un cancer du tiers moyen d'une corde dans 3 cas, une laryngite dans 9 cas. Les sujets témoins ne présentaient aucune plainte vocale ni aucun antécédent de dysphonie. Les examens proposés aux patients et les procédures d'enregistrement étaient effectuée selon la méthodologie décrite dans la section « Protocole

⁴ Revis J, Giovanni A, Wuyts FL, Triglia JM. Comparison of different voice samples for perceptual analysis. Folia Phoniatri Logop, 1999;51(3):108-116

Expérimental » (section VI, chapitre 1.2., p 105-106). Le matériau phonétique enregistré était composé du premier paragraphe de « *La petite chèvre de Monsieur Seguin* » d'Alphonse Daudet, ainsi que de 5 /a/ tenus consécutifs. Le jury était composé de 7 auditeurs expérimentés, ayant déjà participé à des études perceptives de la voix, choisis parmi les médecins et les orthophonistes du laboratoire.

A partir des enregistrements effectués, 3 échantillons différents ont été conservés pour chaque sujet :

- L'échantillon de parole (P) était la phrase: « *il les perdait toutes de la même façon* », extraite du texte lu (figure VI.g.1.)



Figure VI.g.1. : Exemple de signal acoustique du stimulus de parole articulée « *il les perdait toutes de la même façon* »

- L'échantillon de voyelles tenues complètes (VTC) correspondait aux 2°, 3° et 4° voyelles tenues.

Nous avons éliminé le premier et dernier échantillons de voyelles tenues (figure VI.g.2.).

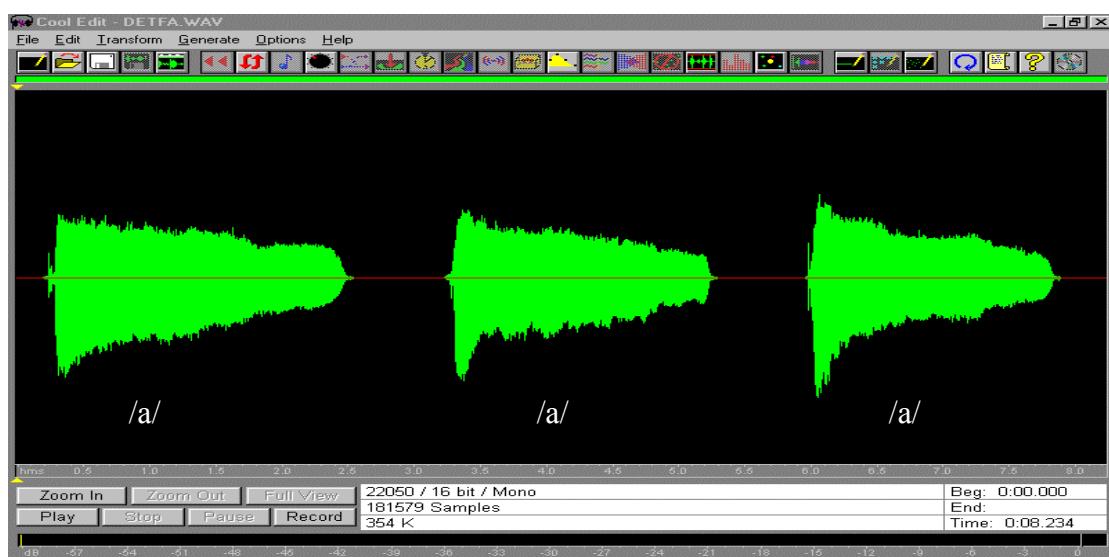


Figure VI.g.2. : Exemple de signal acoustique du stimulus de /a/ tenus.

- L'échantillon de parties stables des voyelles tenues (PSVT) était obtenu de la manière suivante: une portion de 200 ms était supprimée et une pente linéaire 100 msec était ensuite appliquée au début et à la fin du signal pour éviter la sensation auditive de clic (figures VI.g.3., 4., 5.).

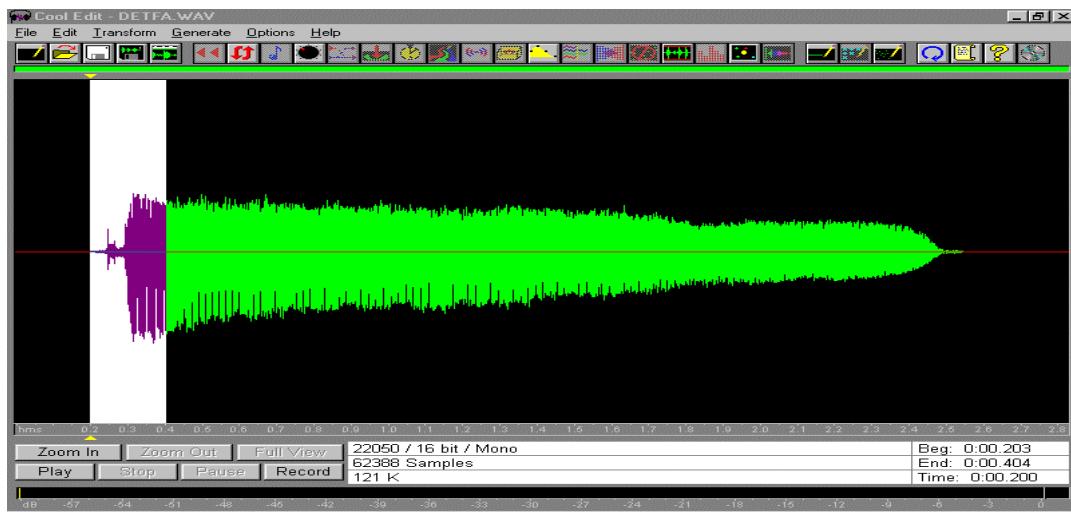


Figure VI.g.3. : Sélection de 200 msec au début du signal grâce aux systèmes de zoom et de compteur de durée de la fenêtre du logiciel Cool Edit ®.

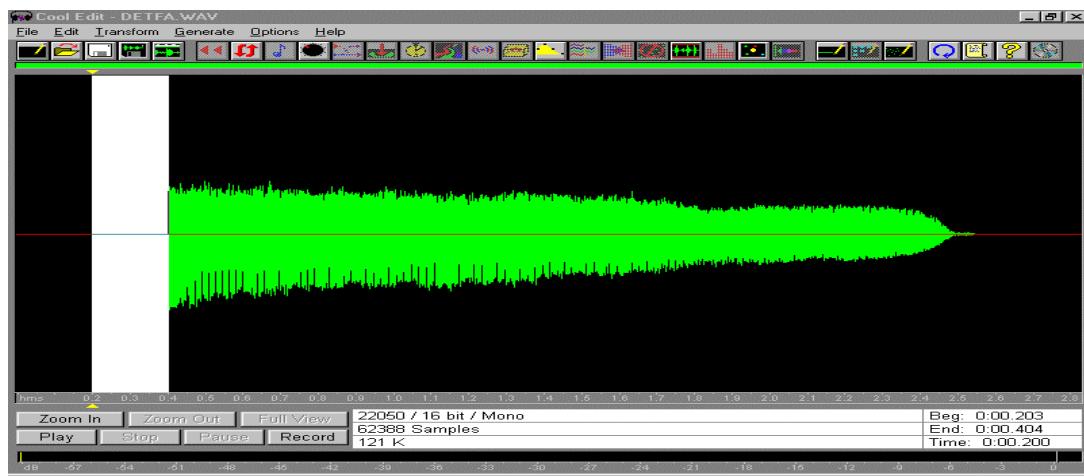


Figure VI.g.4. : Suppression du contenu de la fenêtre de 200 msec.

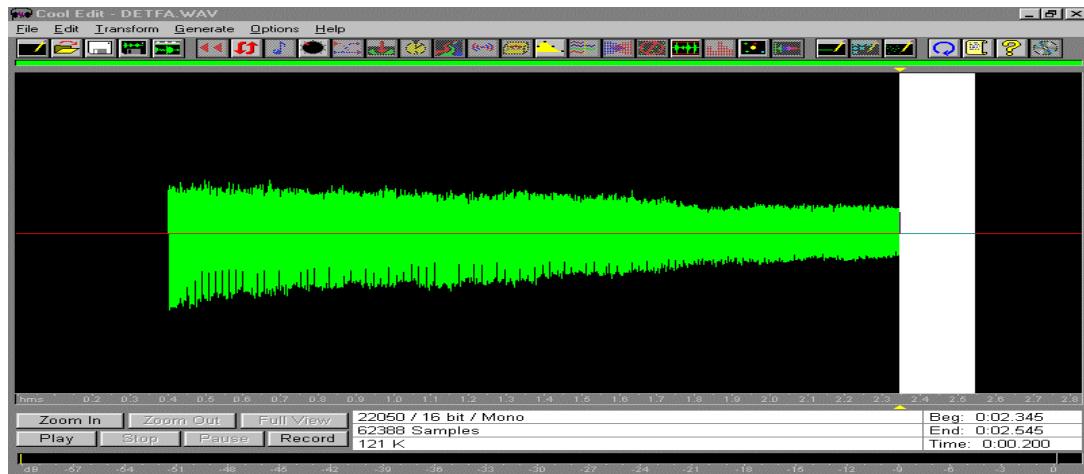


Figure VI.g.5. : La même opération est réalisée à la fin du signal.

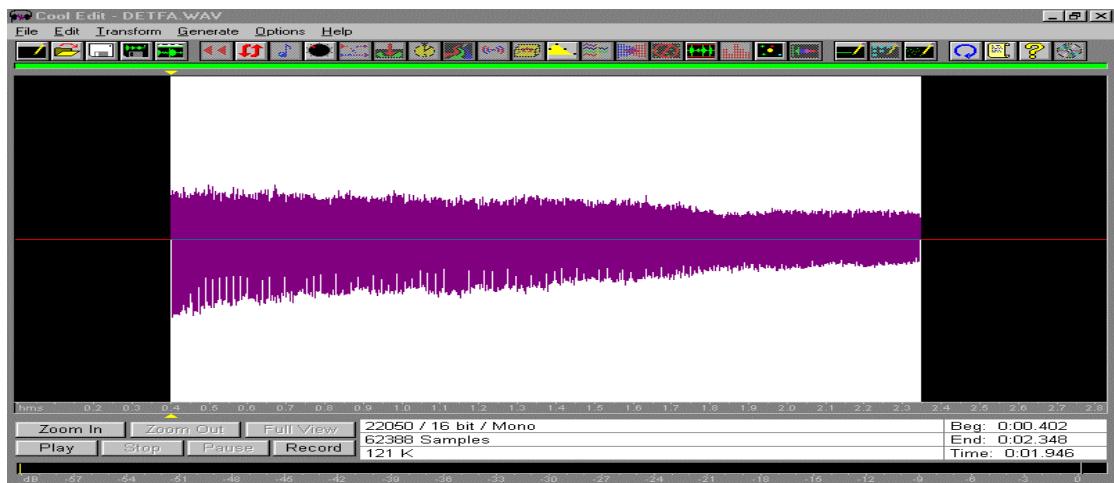


Figure VI.g.6. : Sélection de la totalité de la “partie stable” du signal.

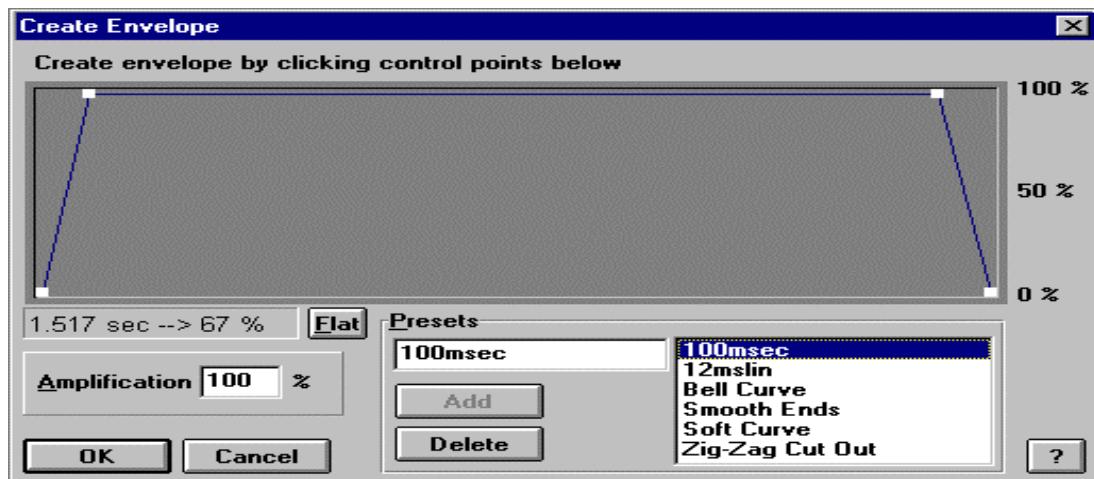


Figure VI.g.7. : Sélection de la pente linéaire de 100 msec grâce à la fonction “enveloppe” du logiciel Cool Edit ®.

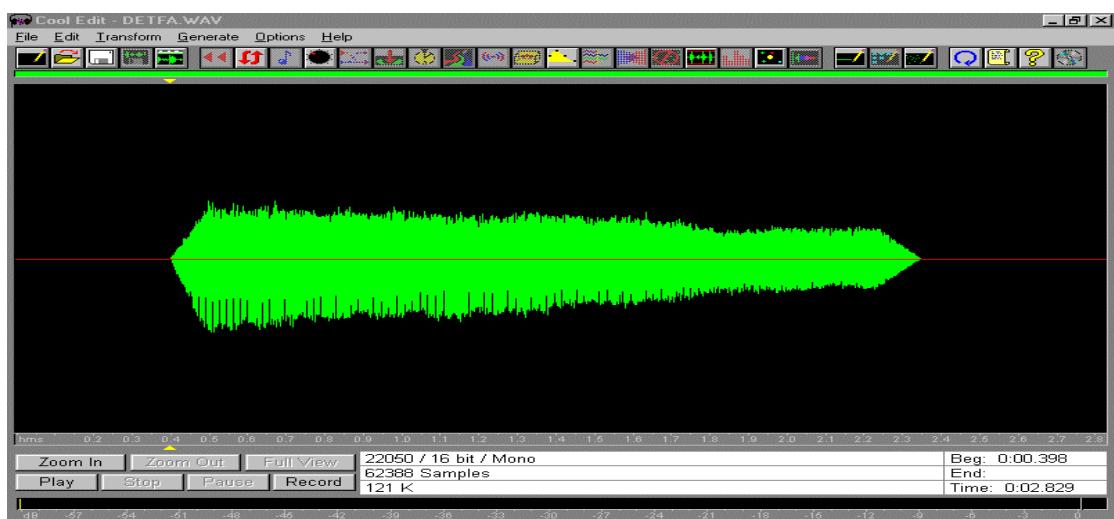


Figure VI.g.8. : Application de la pente au signal sélectionné et obtention du stimulus de /a/ tenu sans transitoires (PSVT).

Les auditeurs ont participé individuellement à 3 sessions séparées d'une semaine pour l'étude de chacun des 3 matériaux phonétiques, soit 9 sessions. Les échantillons de voix des 80 sujets étaient présentés dans un ordre aléatoire différent au cours de chaque session. La méthode d'évaluation choisie a été la méthode GRBAS comprenant 5 paramètres G (grade global de dysphonie), R (raucité), B (souffle), A (asthénie) et S (voix forcée) notés sur une échelle ordinaire à 4 niveaux (0 : voix normale; 1 : trouble léger, 2 : trouble moyen, 3 : trouble sévère). Chaque session durait entre 30 et 45 mn, les membres du jury étaient tous volontaires et ne recevaient aucune rétribution pour les sessions de jugement.

La constance du jury a été appréciée par le pourcentage de réponses identiques lors des 3 présentations différentes de chaque type d'échantillon. Ce test était rendu possible par l'utilisation de l'échelle ordinaire de jugement (0 à 3). Le pourcentage était interprété, comme le proposaient De Bodt et Wuyts [De Bodt, 1997] par le coefficient κ de Cohen qui compare le pourcentage brut obtenu au pourcentage qui pourrait être dû au hasard. Nous avons utilisé pour l'interprétation des résultats la table fournie par De Bodt (tableau VI.g.1.).

Valeur du coefficient Kappa	Degré de fiabilité
<0.20	faible
0.21 - 0.40	raisonnable
0.41 - 0.60	assez bon
0.61 - 0.80	bon
0.81 - 1.00	très bon

Tableau VI.g.1. : Table d'interprétation des coefficients κ proposée par De Bodt et Wuyts [De Bodt, 1997]

L'étude de l'effet du type d'échantillon sur la sévérité du jugement par le jury a reposé sur : 1- l'étude du coefficient de corrélation r de Pearson entre d'une part la parole (variable de référence) et, d'autre part, les échantillons PSVT et VTC (variables testées) et 2- la comparaison des graphiques de régression linéaire.

Résultats : Tous les membres du jury ont déclaré avoir ressenti plus de difficultés à juger l'échantillon PSVT qui leur apparaissait artificiel et désagréable.

La constance du jury pour chaque paramètre est présentée dans les tableaux VI.g.2a. (constance intra-individuelle) et VI.g.2b. (constance inter-individuelle).

2a	Parole	/a/ Tenu	/a/ Tronqué	2b	Parole	/a/ Tenu	/a/ Tronqué
G	0.553	0.554	0.595	G	0.633	0.564	0.511
R	0.475	0.516	0.548	R	0.482	0.501	0.454
B	0.458	0.486	0.507	B	0.462	0.456	0.456
A	0.443	0.485	0.515	A	0.451	0.396	0.458
S	0.456	0.463	0.512	S	0.352	0.301	0.254

Tableau VI.g.2. : constance intra individuelle (2a) et inter individuelle (2b) analysée par le coefficient κ .

Quelque soit le type d'échantillon utilisé, le paramètre G montre le coefficient κ le plus élevé (parole : 0.633, voyelle tenue : 0.564, partie stable de la voyelle : 0.511), tandis que les paramètres A et S présentent des coefficients plus bas, témoignant d'une plus grande variabilité. Les résultats présentés ne montrent pas de différence entre les différents échantillons du point de vue de la constance du jury.

Les coefficients de corrélation entre la parole et les deux échantillons testés (PSVT et VTC) sont présentés sur les tableaux VI.g.3a. (parole et PSVT) et VI.g.3b. (parole et VTC). Les résultats montrent, dans les deux cas, une corrélation hautement significative pour tous les paramètres ($p < 0.001$).

3a	r	coefficient	p	3b	R	coefficient	p
G	0.961	1.221	< 0.001	G	0.920	0.973	< 0.001
R	0.901	1.097	< 0.001	R	0.914	0.870	< 0.001
B	0.926	1.128	< 0.001	B	0.653	0.959	< 0.001
A	0.895	1.091	< 0.001	A	0.884	0.896	< 0.001
S	0.751	0.609	< 0.001	S	0.696	0.545	< 0.001

Tableau VI.g.3. : Corrélation entre les échantillons parole et PSVT (3a) et entre parole et VTC (3b)

L'analyse graphique a révélé que les jugements portés sur la partie stable des voyelles tenues étaient moins sévères que ceux portés sur la parole, d'autant plus que la dysphonie était sévère (figure VI.g.2.).

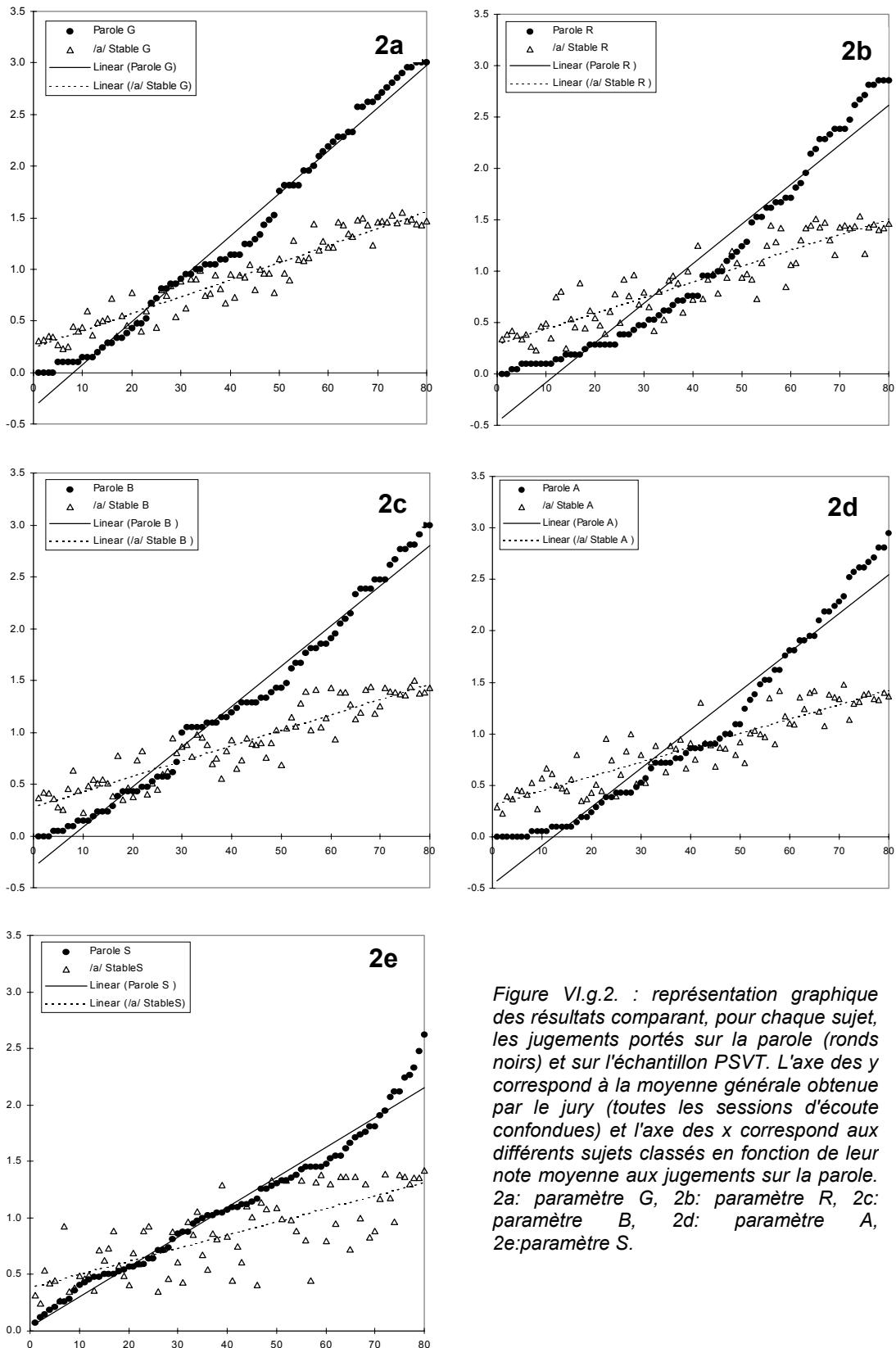


Figure VI.g.2. : représentation graphique des résultats comparant, pour chaque sujet, les jugements portés sur la parole (ronds noirs) et sur l'échantillon PSVT. L'axe des y correspond à la moyenne générale obtenue par le jury (toutes les sessions d'écoute confondues) et l'axe des x correspond aux différents sujets classés en fonction de leur note moyenne aux jugements sur la parole.
 2a: paramètre G, 2b: paramètre R, 2c: paramètre B, 2d: paramètre A,
 2e: paramètre S.

Cette différence n'apparaissait pas sur le graphique comparant la parole aux voyelles tenues complètes (figure VI.g.3.).

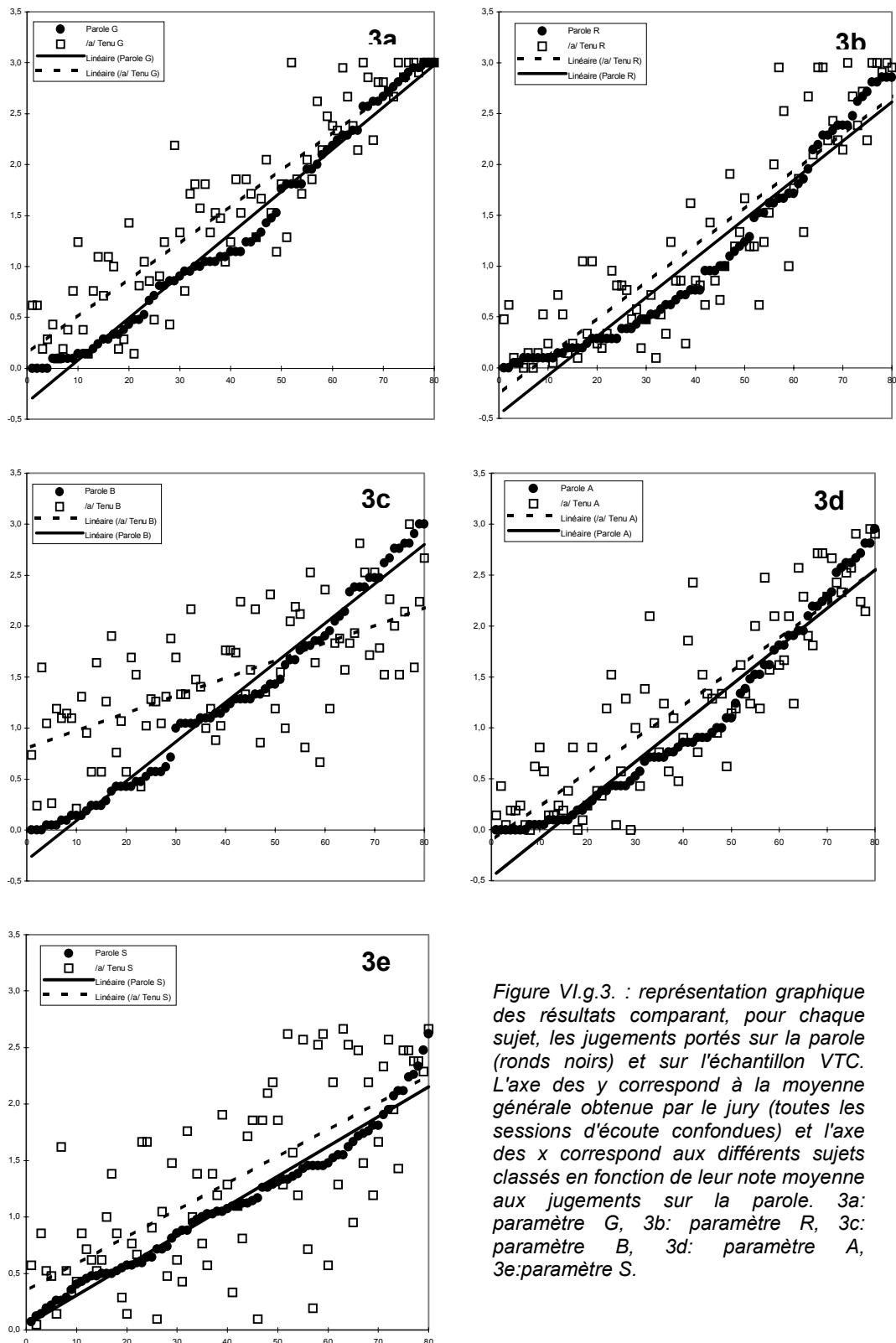


Figure VI.g.3. : représentation graphique des résultats comparant, pour chaque sujet, les jugements portés sur la parole (ronds noirs) et sur l'échantillon VTC. L'axe des y correspond à la moyenne générale obtenue par le jury (toutes les sessions d'écoute confondues) et l'axe des x correspond aux différents sujets classés en fonction de leur note moyenne aux jugements sur la parole. 3a: paramètre G, 3b: paramètre R, 3c: paramètre B, 3d: paramètre A, 3e: paramètre S.

Discussion : Notre travail avait pour objectif de tester l'influence du choix du matériau phonétique sur les jugements perceptifs effectués par un jury. Nos résultats montrent que les jugements portés sur la partie stable d'une voyelle tenue sont moins sévères que ceux portés sur la parole, alors que les jugements portés sur des voyelles tenues comprenant l'attaque sont similaires à ceux portés sur la parole.

La variabilité intra et inter individuelle des jugements portés par notre jury apparaît satisfaisante quand on les compare aux données de la littérature, notamment l'étude de De Bodt et Wuyts [De Bodt, 1997] qui retrouve des coefficients kappa entre 0.28 et 0.70. La comparaison avec les résultats d'autres études est plus difficile car il n'existe pas d'outil statistique normalisé permettant de prendre en compte des résultats issus d'échelles différentes. Pour De Krom [De Krom, 1994], l'analyse de la variabilité est faite en utilisant le coefficient σ de Spearman, plus adapté pour l'analyse d'une échelle ordinaire à 7 niveaux. Notre étude ne retrouve pas de différence entre les différents matériaux phonétiques, confirmant que le choix de celui-ci n'influence pas la stabilité des évaluations du jury. Ce résultat est en accord avec De Krom [De Krom, 1994]. A l'inverse, nos résultats montrent une variabilité différente en fonction du paramètre étudié. La meilleure constance est obtenue avec le paramètre G alors que la moins bonne est obtenue pour les paramètres A et S. Ce type de résultat est retrouvé par Dejonckere [Dejonckere, 1993] qui les attribue au fait que les dimensions perceptives des dimensions A et S sont moins bien établies dans les références perceptives des auditeurs. Cela confirme l'intérêt des méthodes d'ancre externe proposées par Gerratt [Gerratt, 1993].

La comparaison graphique (figure VI.g.2.) a confirmé que, pour un même sujet, les jugements portés sur la partie stable des voyelles tenues sont moins sévères que ceux portés sur la parole. La sous-évaluation est d'autant plus importante que la dysphonie est plus sévère, ce qu'exprime la pente supérieure à 1 de la fonction de corrélation (tableau VI.g.3a.). Ce résultat est en contradiction avec les résultats obtenus par De Krom [De Krom, 1994]. Pour lui, dans les cas de dysphonies sévères, le vibrateur laryngé ne parvient jamais à la stabilité et les transitoires d'attaque apportent une information redondante avec la partie médiane du signal. Nos résultats semblent au contraire montrer que l'information acoustique contenue dans l'attaque n'est jamais redondante : dans les cas de dysphonie les plus sévères (partie droite des figures VI.g.2.), la moyenne des jugements portés par le jury ne dépasse jamais 1,6 lorsque le jugement est porté sur la partie stable des voyelles alors que la valeur moyenne maximum de 3 est atteinte pour plusieurs sujets sur les échantillons de parole. Il existe une forte corrélation, hautement significative (tableau VI.g.3a.) entre les deux types de jugement et une sorte de coefficient de « correction » équivalent à la pente de la corrélation pourrait être envisagé. Cependant, des études expérimentales doivent être menées avant de conclure pour savoir si ce type de coefficient pourrait être valide.

Les jugements portés sur les voyelles tenues complètes évoluent en revanche parallèlement à ceux portés sur la parole. Les courbes de la figure VI.g.3. montrent que les droites de régression linéaire sont presque confondues, ce que confirme le coefficient de la fonction de corrélation proche de 1 (tableau VI.g.3b.). Cependant, il existe une plus grande dispersion des jugements portés sur l'échantillon VTC que ce qui était observé avec l'échantillon PSVT. Il existe donc des informations perceptives dans l'attaque de voyelles tenues différentes de celles présentes dans les attaques « normales » présentes dans la parole. Wolfe [Wolfe, 1995] a relevé que les notes attribuées sur une voyelle /a/ tenue sont légèrement mais significativement plus sévères que celles attribuées sur la parole. Pour elle, la production d'une voyelle tenue, en dehors d'un contexte articulatoire normal favoriserait l'apparition d'anomalies vocales même chez les sujets sains. Il est possible que, dans notre étude, certains sujets, aient modifié dans un sens ou dans un autre leur production vocale lorsqu'ils ont produit les voyelles tenues.

Au total, nos résultats peuvent être interprétés de la façon suivante. L'évaluation perceptive sous-évalue la dysphonie lorsqu'elle est réalisée sur la partie stable des voyelles tenues. L'importance de l'analyse des transitoires d'attaque est donc confirmée. Cependant, la partie stable des voyelles tenues fournit une information pertinente sur la qualité de la voix. Il s'agit d'un matériau artificiel qu'un jury peut trouver désagréable à analyser mais qui peut être raisonnablement proposé aux dispositifs d'analyse instrumentale. L'évaluation perceptive de voyelles tenues complètes comprenant les transitoires d'attaque ne sous-évalue pas la dysphonie mais il est possible que les conditions mêmes de la production vocale introduisent un biais. Il n'est donc pas démontré que les voyelles tenues complètes sont un matériau phonétique pertinent pour l'analyse perceptive et elles ne doivent donc être proposées qu'avec prudence pour remplacer, dans une étude expérimentale, les fragments de parole « normale ».

h. Influence de l'attaque sur l'analyse perceptive des dysphonies.⁵

Revis J, Giovanni A, Triglia JM

(Folia Phoniatri Logop, 2002)

Dans notre travail précédent [Revis, 1999], les résultats concernant la partie stable du /a/ tenu nous ont permis d'invalider ce matériau phonétique pour l'analyse perceptive des dysphonies. Cependant, nous ne pouvions pas tirer de conclusions quant à la nature de l'influence de l'attaque : 1- nous utilisions 3 /a/ consécutifs, nous ne pouvions donc pas savoir sur lequel de ces 3 /a/ se produisait l'effet de sous-évaluation ; 2- le traitement du signal défini par notre protocole supprimait à la fois l'attaque et la finale, nous ne pouvions donc pas savoir sur quel transitoire se produisait l'effet ; et 3- les écoutes ne se déroulaient pas en aveugle, le jury savait qu'il analysait des parties stables de /a/ tenu et a pu être influencé dans sa cotation.

Notre nouvelle étude a pour but de caractériser la nature de l'influence perceptive de la mise en vibration des cordes vocales : l'attaque. Notre hypothèse est que l'absence d'attaque influence la perception de la dysphonie par un jury d'écoute expérimenté. Notre protocole d'évaluation tiendra compte des 3 points mis en évidence ci-dessus : 1- évaluation du même /a/ présenté 3 fois ; 2- suppression de l'attaque seulement ; 3- évaluation en aveugle.

Protocole expérimental : Nous avons choisi 40 sujets dans le corpus de notre précédente étude en fonction du niveau d'agrément entre les auditeurs : nous n'avons retenu que les sujets pour lesquels notre jury n'avait pas ou peu été variable sur l'évaluation du /a/ tenu. Cette évaluation est considérée comme la référence contre laquelle nous avons testé nos nouveaux résultats. Nous avons retenu 10 Grades 0 (voix normale), 10 Grades 1 (dysphonie légère), 10 Grades 2 (dysphonie moyenne) et 10 Grades 3 (dysphonie sévère), répartis ainsi : femmes n=28 ; hommes n=12 ; âge compris entre 19 et 76 ; pathologies : cancer n=2 ; laryngite n=3 ; nodules n=6 ; œdème de Reinke n=8 ; paralysie n=5 ; polypes n=4 ; kyste n=2 ; et 10 sujets sains non-tabagiques et sans antécédent de dysphonie.

Les enregistrements de voix étaient effectués de la manière habituelle. 2 échantillons différents ont été conservés pour chaque sujet. L'échantillon de voyelles tenues complètes correspondait à la 3^{ème} voyelle tenue. Nous n'avons conservé que l'échantillon médian. Nous avons ensuite dupliqué cet

⁵ Revis J, Giovanni A, Triglia JM. Influence de l'attaque sur l'analyse perceptive des dysphonies. Folia Phoniatri Logop. 2002;54:19-25

échantillon, afin que le jury puisse écouter le même échantillon 3 fois de suite. L'échantillon voyelle tenue sans l'attaque était obtenu de la manière suivante: une portion de 200 ms était supprimée au début du signal et une pente linéaire 100 msec était ensuite appliquée au début signal pour éviter la sensation auditive de clic (figure VI.h.1.). Aucun traitement du signal n'était appliqué sur le transitoire de sortie du signal acoustique. De la même façon que pour le /a/ complet, l'échantillon était ensuite dupliqué 3 fois.

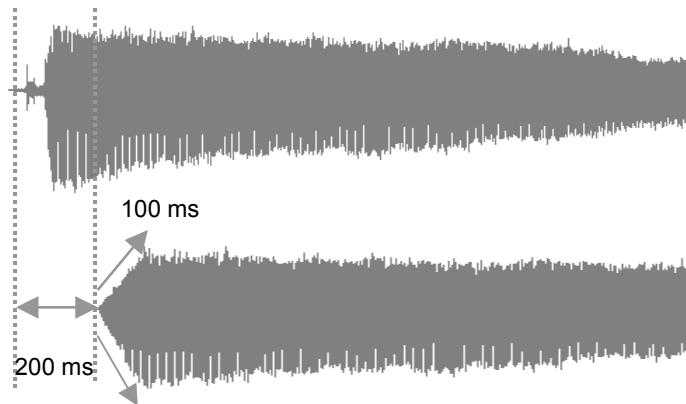


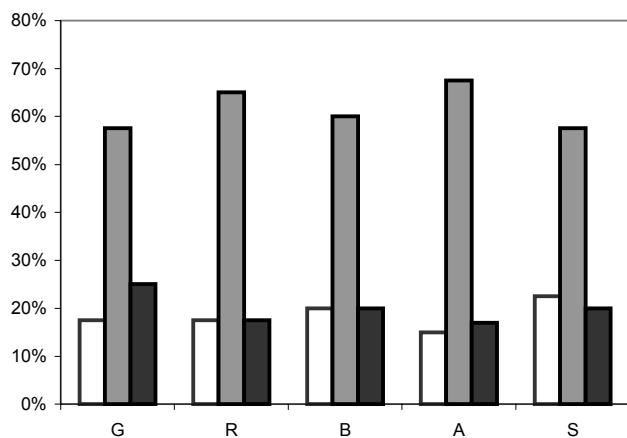
Figure VI.h.1. : traitement de l'échantillon de voyelle pour parvenir à l'échantillon de /a/ sans attaque : suppression des 200 premières ms, puis application d'une pente linéaire de 100 ms à l'initiation du son.

3 auditores expérimentés ont participé ensemble à 1 session d'écoute par la méthode du consensus. Les enregistrements de voix des 40 sujets, comprenant les /a/ tenus et les /a/ tenus sans attaque, c'est à dire 80 stimuli, étaient présentés dans un ordre aléatoire à l'aide d'une station informatique dédiée. L'évaluation s'est déroulée en aveugle : le jury d'écoute ne savait pas que des /a/ sans attaque étaient inclus dans le corpus. La méthode d'évaluation choisie a été la méthode GRBAS [Wuyts, 1999 ; Hirano, 1981 ; De Bodt, 1997] comprenant 5 paramètres : G (grade global de dysphonie), R (raucité), B (souffle), A (asthénie) et S (voix forcée) notés sur une échelle ordinaire à 7 niveaux (0 : voix normale; 1 : dysphonie légère, 2 : dysphonie moyenne, 3 : dysphonie sévère, en utilisant les $\frac{1}{2}$ valeurs pour les cas limitrophes: 0.5; 1.5; 2.5). La session a duré environ 60 mn, les membres du jury étaient tous volontaires et ne recevaient aucune rétribution pour la session de jugements.

L'étude de l'influence de l'attaque sur la sévérité du jugement par le jury a reposé sur le pourcentage de réponses identiques retrouvé sur les deux matériaux : /a/ tenu et /a/ tenu sans l'attaque.

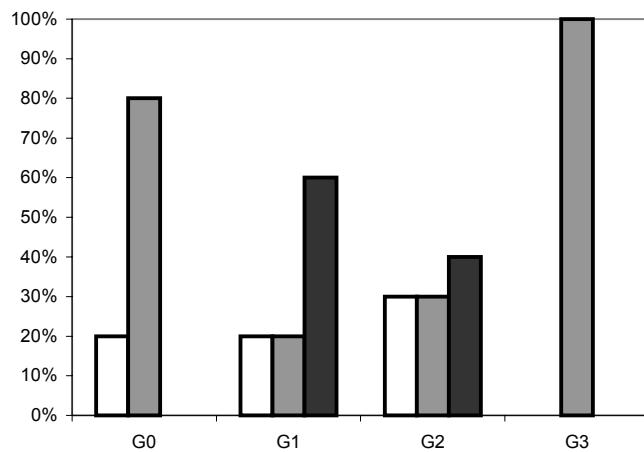
Résultats : La comparaison /a/ tenu complet versus /a/ tenu sans attaque a fait l'objet d'une représentation graphique en histogrammes. Les résultats sont exprimés par le pourcentage de réponses identiques.

1- Nous avons comparé les résultats obtenus sur les /a/ sans attaque à ceux obtenus sur les /a/ tenus en fonction du paramètre de l'échelle : G, R, B, A, S (figure VI.h.2.)



En fonction du paramètre (figure VI.h.2.) : on retrouve environ 60% de réponses identiques quel que soit le paramètre étudié. Dans 21,5% des cas, le /a/ tenu est plus dysphonique que le /a/ sans l'attaque. Dans 18,5% des cas, le /a/ tenu est moins dysphonique que le /a/ sans attaque.

2- Nous avons comparé les résultats obtenus sur les /a/ sans attaque à ceux obtenus sur les /a/ tenus en fonction du niveau de dysphonie : 0, 1, 2, 3 (figure VI.h.3.)



En fonction du grade de dysphonie (figure VI.h.3.) : on retrouve environ 80% de réponses identiques pour le grade 0 (voix normale). Dans 100% des différences que l'on retrouve, le /a/ sans attaque est évalué plus sévèrement que la /a/ tenu. Pour le grade 3 (dysphonie sévère), on retrouve 100% de réponses identiques : la même note a été attribuée sur le /a/ tenu et sur le /a/ sans attaque. Pour ces 2 grades (voix normale et dysphonie sévère), on retrouve donc un niveau élevé de réponses identiques. En revanche, pour le grade 1 (dysphonie légère) on ne retrouve que 20% de réponses identiques. Dans 75% des cas, le /a/ tenu est sous-évalué par rapport au /a/ sans attaque. Et pour le grade 2 (dysphonie moyenne), on ne retrouve que 30% de réponses identiques. Dans 57% des cas, le /a/ tenu est sous-évalué par rapport au /a/ sans attaque.

Discussion : Notre précédente étude avait démontré que le /a/ tenu était un substitut convenable à la parole pour l'évaluation perceptive des dysphonies [Revis, 1999]. Ce résultat nous a permis d'utiliser le /a/ tenu comme méthode de référence pour ce travail. Ainsi, nous avons pu faire notre expérimentation sur 2 matériaux ne différant qu'en 1 seul point : l'attaque.

Le choix de l'évaluation perceptive par la méthode du consensus a été fait en fonction de notre hypothèse [Kreiman, 1993]. L'objet de notre travail était d'étudier l'influence de l'attaque sur la sensation auditive de dysphonie. Une évaluation perceptive classique en trois sessions distinctes et indépendantes aurait mis en évidence une variabilité de notre jury. Cette variabilité existe et a été démontrée au cours de nombreux travaux [Revis, 1999 ; Hammarberg, 1980 ; Kreiman, 1992 ; Wuyts, 1999 ; Dejonckere, 1993 ; De Krom, 1994 ; De Bodt, 1997 ; Kreiman, 1993 ; Gerratt, 1993]. Cependant, aucune solution n'est actuellement disponible afin de réduire cette variabilité [De Bodt, 1997]. Notre objectif n'était pas d'étudier la variabilité du jury. Nous avons donc choisi la méthode du consensus qui permet d'obtenir une évaluation solide, puisqu'elle résulte de la concertation de plusieurs auditeurs expérimentés, et ne tient pas compte de la variabilité.

L'attaque correspond à la mise en vibration des cordes vocales. Une attaque normale est caractérisée par une fermeture des cordes vocales simultanée à la mise en fonctionnement de la colonne d'air. L'attaque pathologique est caractérisée essentiellement par deux fonctionnements perceptibles à l'oreille : l'attaque soufflée et l'attaque dure [Titze, 1994 ; Cooke, 1997 ; Köster, 1999 ; Decoster, 1997]. Dans les deux cas, la pathologie correspond à une perte de la simultanéité de la fermeture et du souffle. L'attaque soufflée correspond à une fuite d'air lors de la mise en vibration des cordes vocales. C'est une attaque longue. La colonne d'air est mise en fonctionnement avant la fermeture des cordes vocales. Une attaque soufflée pourrait amplifier la perception du souffle. On pourrait donc s'attendre à ce que le paramètre de souffle (B) soit sous-évalué sur un /a/ sans attaque. L'attaque dure correspond à un coup de glotte. C'est une attaque brève. Les cordes vocales sont fermées avant la mise en fonctionnement de la colonne d'air. Ce type d'attaque entre dans le cadre de la perception du forçage vocal. On pourrait donc s'attendre à ce que le paramètre de forçage (S) soit sous-évalué sur un /a/ sans attaque [Cooke, 1997 ; Köster, 1999]. Or nos résultats montrent que la différence entre les évaluations portées sur le /a/ tenu et celles portées sur le /a/ sans attaque évolue au hasard quel que soit le paramètre étudié : dans environ 50% des cas le /a/ sans attaque sous-évalue la dysphonie et dans environ 50% des cas le /a/ sans attaque surévalue la dysphonie. L'importance de l'attaque ne s'exprime pas de façon significative en fonction du paramètre exploré.

L'influence de l'attaque, en revanche, varie en fonction du niveau de dysphonie. Pour les grades 0 (voix normales) on retrouve une importance faible de l'attaque (80% de réponses identiques sur les 2

matériaux). Une voix normale reste normale avec ou sans transitoire. L'attaque porte peu d'information perceptive sur la qualité de la voix normale, puisque les cordes vocales entrent immédiatement en vibration synchrone. Pour les grades 3 (dysphonies sévères), l'attaque a une importance négligeable (100% de réponses identiques sur les 2 matériaux). Dans les dysphonies sévères, la vibration n'atteint jamais la stabilité. L'attaque, qui est caractérisée par une instabilité porte donc peu d'information perceptive sur la dysphonie. Ces résultats sont en adéquation avec ceux présentés par De Krom [De Krom, 1994]. En ce qui concerne les grades 1 et 2, l'attaque possède un véritable rôle perceptif : pour les grades 1, dans 75% des cas, le /a/ sans attaque est sous-évalué ; pour les grades 2, dans 57% des cas, le /a/ sans attaque est sous-évalué. L'attaque pourrait donc avoir un rôle de discrimination des dysphonie légères et moyennes. De plus, de nombreuses études ont montré que le jury d'écoute est davantage variable sur les grades 1 et 2. Il est possible de penser que l'attaque joue un rôle dans cette variabilité.

Conclusion : Notre hypothèse était que l'absence d'attaque influence la perception de la dysphonie par un jury d'écoute expérimenté. Nos résultats ont montré que cette influence ne porte pas sur le paramètre étudié : quelle que soit la dimension dysphonique étudiée (GRBAS), l'absence d'attaque n'a pas eu d'effet significatif sur l'évaluation du jury. En revanche, nous avons montré que l'influence de l'attaque varie en fonction du niveau de dysphonie : négligeable pour les voix normales et dysphonies sévères, cette influence est importante pour les cas de dysphonie légère et moyenne. En d'autres termes, on peut penser que la perception d'une irrégularité vibratoire durant l'attaque dépend de la perception globale de l'irrégularité vibratoire sur l'ensemble du signal. De nouveaux travaux sur la physiologie phonatoire pendant l'attaque permettront de déterminer si cette perception correspond à une réalité vibratoire en définissant un rapport « irrégularités sur l'attaque / irrégularités sur la partie stable » selon une méthode assimilable à la technique du rapport signal/ bruit.

i. Définition d'une mesure temporelle de l'attaque vocale⁶

Revis J, Barberis S, Giovanni A, Triglia JM

(Rev Otol Rhinol Laryngol, 2000)

Nos précédents travaux montrent que l'attaque est porteuse d'une information sur la dysphonie, que cette information est essentiellement de type quantitatif, et qu'elle est perçue par le jury d'écoute. Or, cette information n'est pas prise en compte par les mesures instrumentales puisque celles-ci s'effectuent sur la seule partie stable d'un /a/ tenu. D'ailleurs, en tant que clinicien, nous avons été confrontés à des patients présentant indiscutablement une dysphonie lors de l'analyse perceptive, mais classés « sujets sains » par l'analyse instrumentale. Cette constatation, appuyée par les travaux présentés ci-dessus, nous a conduit à réfléchir à un moyen de « mesurer » les phénomènes contemporains de l'attaque.

En effet, de nombreux travaux centrés sur l'analyse instrumentale des dysphonies ont soulevé la question de l'analyse de la mise en vibration des cordes vocales sans proposer de méthodologie de mesure de l'attaque [Giovanni, 1996]. Pour Koike [Koike, 1973] « *L'attaque est caractérisée par une élévation significative du jitter par rapport à la partie stable* ». En d'autres termes, la mise en vibration des cordes vocales est caractérisée par une grande instabilité fréquentielle. Pour Giovanni [Giovanni, 1996] « *l'attaque pourrait contenir une information capitale qui n'est pas explorée par les méthodes de mesure actuelles* ». Au total, ces travaux mettent en évidence la nécessité de développer une mesure de l'attaque.

En réalité, il faut considérer qu'une mesure de l'instabilité vibratoire lors de la mise en vibration des cordes vocales ne peut être envisagée que si l'on sait déterminer précisément les limites temporelles de l'attaque. En d'autres termes, il est nécessaire de savoir isoler l'attaque en définissant des critères temporels de début et de fin. S'il est relativement aisé de définir le début de l'attaque par l'émergence du signal glottique, comment déterminer le point d'infexion à partir duquel on n'est plus dans l'instabilité contemporaine de la mise en vibration des cordes vocales mais où on passe à une vibration stable. Autrement dit, quand passe-t-on de l'attaque à la partie stable de la voyelle tenue. La mesure utilisée par les phonéticiens, le VOT (Voice Onset Time) [Decoster, 1997], considère que le temps d'attaque correspond à la durée comprise entre l'explosion de la consonne occlusive et l'émergence de la première onde glottique contenant les 2 premiers formants. Ceci correspond à un critère articulatoire

⁶ Revis J, Barberis S, Giovanni A. Definition of a new temporal voice onset measurement. Rev Laryngol Otol Rhinol (Bord). 2000;121(5):291-6.

tenant compte de la transition entre les phonèmes et non de la mise en vibration laryngée à proprement parler. Baken [Communication Personnelle], dans une approche plus phonologique, considère que la durée de l'attaque correspond au décalage temporel compris entre l'émergence du signal acoustique et du signal électroglottographique. En effet, si l'on considère que le signal acoustique capté par un microphone est détecté dès le début de l'émission vocale, le signal électroglottographique, lui, n'émerge que lors de la synchronisation des cordes vocales, soit à la stabilisation du signal. En fait, il apparaît que l'électroglottographe détecte la vibration avant qu'elle ne se stabilise, ce qui est normal puisque l'électroglottographe est un outil de visualisation du signal dysphonique, donc instable. Il existe donc bien un décalage entre l'émergence des deux signaux, mais ce temps de latence est sous estimé par rapport à la réelle stabilisation du signal.

Le but de ce travail est de mettre au point une méthode de mesure standardisée de la durée de l'attaque. Considérons que le signal vocal présente les mêmes caractéristiques que tout autre signal : un temps de montée, une partie stable, et un temps de descente. Le temps de montée d'un signal (l'attaque) peut être calculé par la méthode dite des "10%-90%". Si l'on considère que l'amplitude maximale d'un signal (100%) correspond à l'obtention d'un régime stable, il a été défini que le temps de montée de ce signal correspond à la durée comprise entre le point où le signal atteint 10% de cette amplitude (début de l'attaque) et le point où le signal atteint 90% de cette amplitude (fin de l'attaque) (figure VI.i.1.).

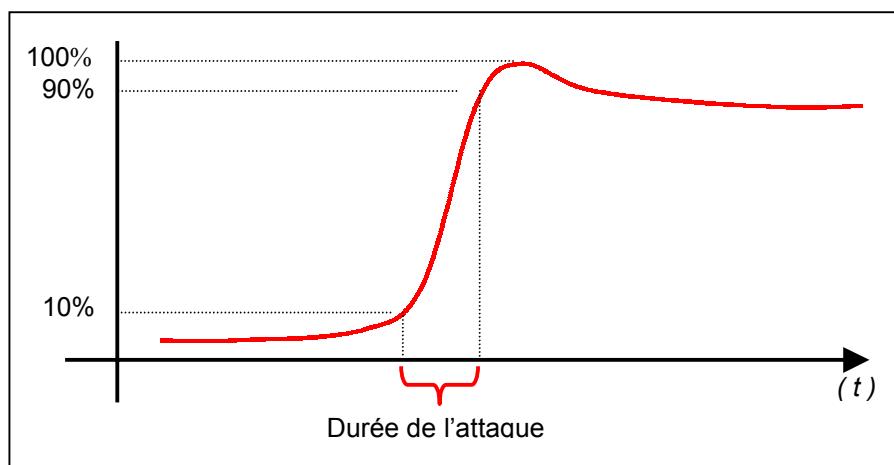


Figure VI.i.1. : Calcul du temps de montée d'un signal par la méthode « 10%-90% »

Cette méthode peut être appliquée au signal vocal en utilisant la courbe d'intensité. 100% d'amplitude correspondant au niveau d'intensité maximum du signal en dB (A_m), 10% correspondant à ($A_m - 20$ dB), et 90% correspondant à ($A_m - 1$ dB) (figure VI.i.2.).

$$A_0 = A_m * 10/100 = 0,1 * A_m$$

$$20 \log_{10} A_0 = 20 \log (0,1 * A_m)$$

$$A_0 \text{dB} = 20 \underbrace{\log_{10}(0,1)}_{-1} + 20 \underbrace{\log_{10} A_m}_{A_m \text{dB}}$$

$$A_0 \text{dB} = A_m \text{dB} - 20 \text{dB}$$

$$10\% = A \text{ max} - 20 \text{ dB}$$

$$90\% = A \text{ max} - 1 \text{ dB}$$

Figure VI.i.2. : Calcul du temps de montée appliquée à la mesure de l'intensité du signal vocal

Notre hypothèse est que la méthode des « 10%-90% », appliquée à la phonation, permet de mesurer le temps de mise en vibration des cordes vocales, l'attaque.

Protocole Expérimental : Nous avons testé la validité clinique de la mesure temporelle de l'attaque vocale sur un corpus de voix représentatif des différents niveaux de dysphonie. Nous avons choisi 80 sujets féminins dans la banque de données de voix du Laboratoire d'Audio-Phonologie Clinique en fonction de leur niveau de dysphonie évalué par notre jury d'experts. Nous avons retenu 20 Grades 0 (voix normale), 20 Grades 1 (dysphonie légère), 20 Grades 2 (dysphonie moyenne) et 20 Grades 3 (dysphonie sévère), répartis ainsi : femmes n=80 ; âge compris entre 19 et 76 ; pathologies : laryngite n=? ; nodules n=? ; œdème de Reinke n=? ; paralysie n=? ; polypes n=? ; kyste n=? ; et 20 sujets sains non-tabagiques et sans antécédent de dysphonie.

Les enregistrements de voix étaient effectués de la manière habituelle. Le matériau phonétique était constitué de 5 /a/ tenus consécutifs. Nous n'avons conservé pour notre étude que les 3 échantillons médians (nous avons supprimé le 1^{er} /a/ ainsi que le 5^{ème}). Puis nous avons effectué nos mesures sur chacun des 240 /a/ tenus que nous avions enregistrés.

Le calcul de la durée de l'attaque a été réalisé grâce au logiciel Phonédit © . Pour chacun des 240 /a/ tenus, nous avons affiché simultanément le signal acoustique et la courbe d'intensité (figure VI.i.3.). Puis nous avons relevé la valeur de l'intensité de référence (100%) sur la partie stable de la courbe d'intensité : curseur 0. Nous avons ensuite placé le curseur 1 à cette valeur – 20 dB (10%), puis le curseur 2 à cette valeur – 1 dB (90%). Puis nous avons calculé le temps compris entre les curseurs 1 et 2, en millisecondes.

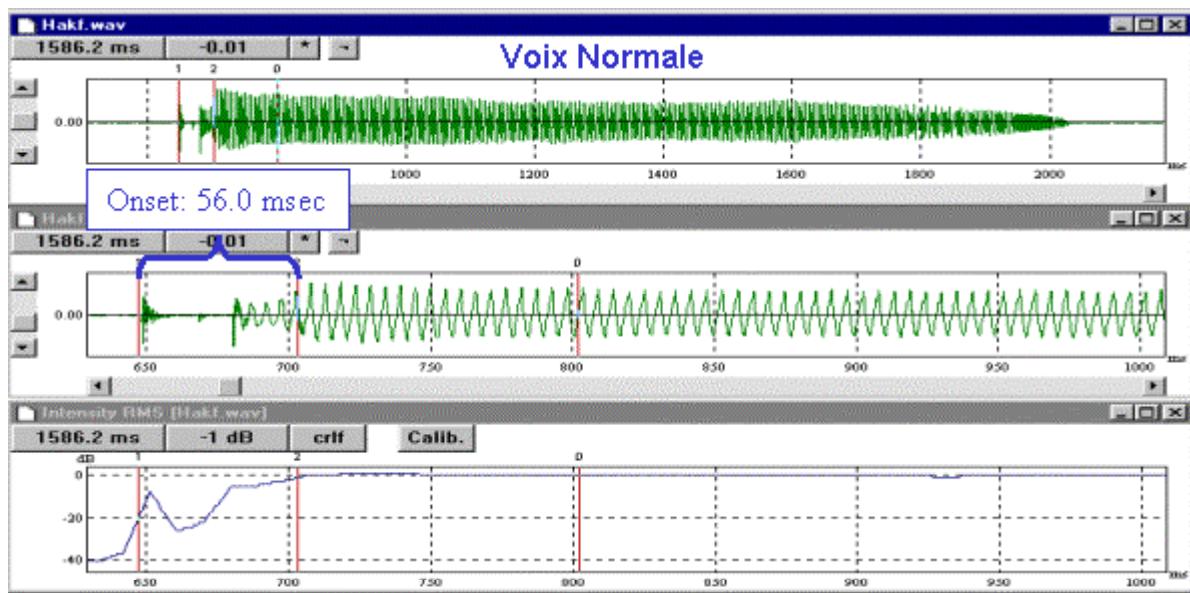


Figure VI.i.3. : Traitement du signal, calcul de la durée de l'attaque

Notre étude statistique a reposé sur la description de la répartition des valeurs en fonction de la durée de l'attaque, en fonction du grade de dysphonie, et en fonction de l'individu.

Résultats :

- La répartition des valeurs en fonction de la durée de l'attaque est représentée sur la figure VI.i.4.

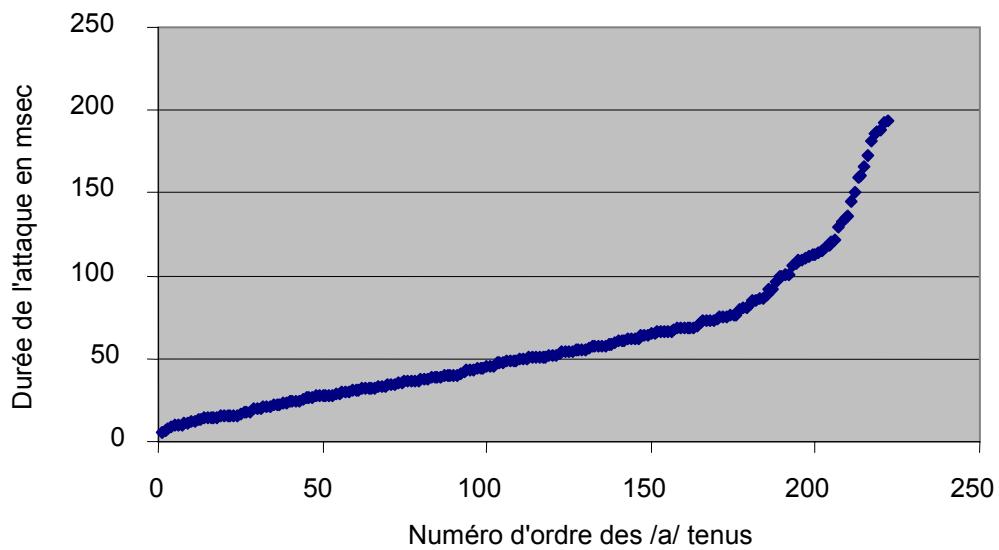


Figure VI.i.4. : Répartition des valeurs par ordre croissant de la durée de l'attaque.

La durée moyenne de l'attaque calculée sur les 240 /a/ tenus est de 51,8 msec. L'attaque la plus courte est de 5,4 msec et l'attaque la plus longue est de 193,3 msec.

- 2- La répartition des valeurs en fonction du grade de dysphonie est représentée sur la figure VI.i.5.
 G0 = voix normale ; G1 = dysphonie légère ; G2 = dysphonie moyenne ; G3 = dysphonie sévère.

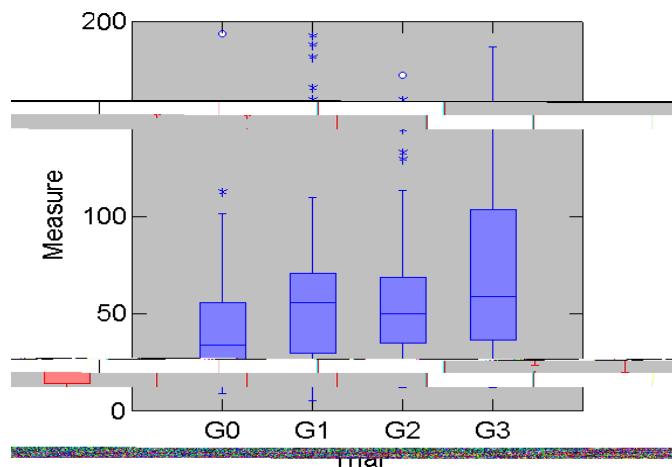


Figure VI.i.5. : Répartition des valeurs de la durée de l'attaque en fonction du grade de dysphonie

Il n'existe pas de différence significative de la durée de l'attaque en fonction du niveau de dysphonie.

- 3- La répartition des valeurs en fonction de l'ordre de production de la voyelle tenue est représentée sur la figure VI.i.6.

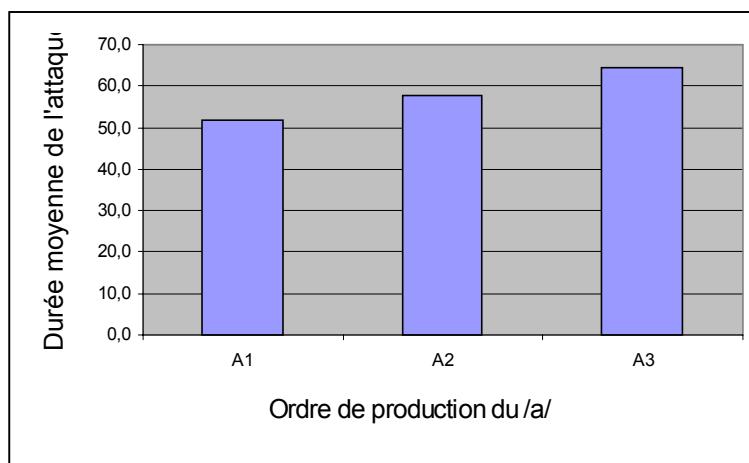


Figure VI.i.6. : Répartition des valeurs de la durée de l'attaque en fonction de l'ordre de production du /a/

La durée de l'attaque en fonction de l'ordre du production du /a/ par chaque sujet est sensiblement la même pour les 3 /a/ tenus avec un écart-type de 6,3 msec.

Discussion : La répartition des valeurs de la durée de l'attaque montre que la mise en vibration des cordes vocales n'excède pas 193,3 msec. Ces résultats confortent les travaux de De Krom [De Krom,

1994] basés sur le postulat qu'il est nécessaire et suffisant de supprimer les 200 premières millisecondes du signal pour supprimer l'attaque vocale.

Nos résultats montrent que la durée de l'attaque en elle-même ne renseigne pas sur le niveau de dysphonie. Notons que notre travail consiste à mesurer la durée de l'attaque afin de pouvoir l'extraire de l'ensemble du signal pour y appliquer ultérieurement une mesure qualitative. Aucune hypothèse n'étant formulée quant à la pertinence d'une mesure temporelle.

Les mesures de la durée de l'attaque sont sensiblement les mêmes en fonction de l'ordre de production des /a/ tenus : un sujet produisant une attaque longue, produit une attaque longue sur chacun de ses /a/ tenus à quelques millisecondes près. De la même façon, un sujet produisant une attaque courte, produit une attaque courte sur chacun de ses /a/ tenus à quelques millisecondes près. En d'autres termes, il semble exister une cohérence intra-individuelle de la stratégie d'attaque. Il est également intéressant de noter qu'il semble exister une fatigabilité de l'attaque vocale, puisque l'attaque mesurée sur le premier /a/ est sensiblement plus courte que celle mesurée sur le deuxième /a/, qui est sensiblement plus courte que celle mesurée sur le troisième /a/.

Enfin, nous avons isolé quelques cas particuliers, correspondant à des cas extrêmes de production vocale : les voix supra-normales (chanteurs) et certains cas de dysphonie sévère.

1- Dans les cas de voix supra-normales, nous sommes en présence de sujets possédant une bonne culture vocale, et qui produisent intentionnellement une attaque douce, progressive, évitant soigneusement le coup de glotte. Ceci donne des temps d'attaque rallongés à l'extrême (figure VI.i.7.).

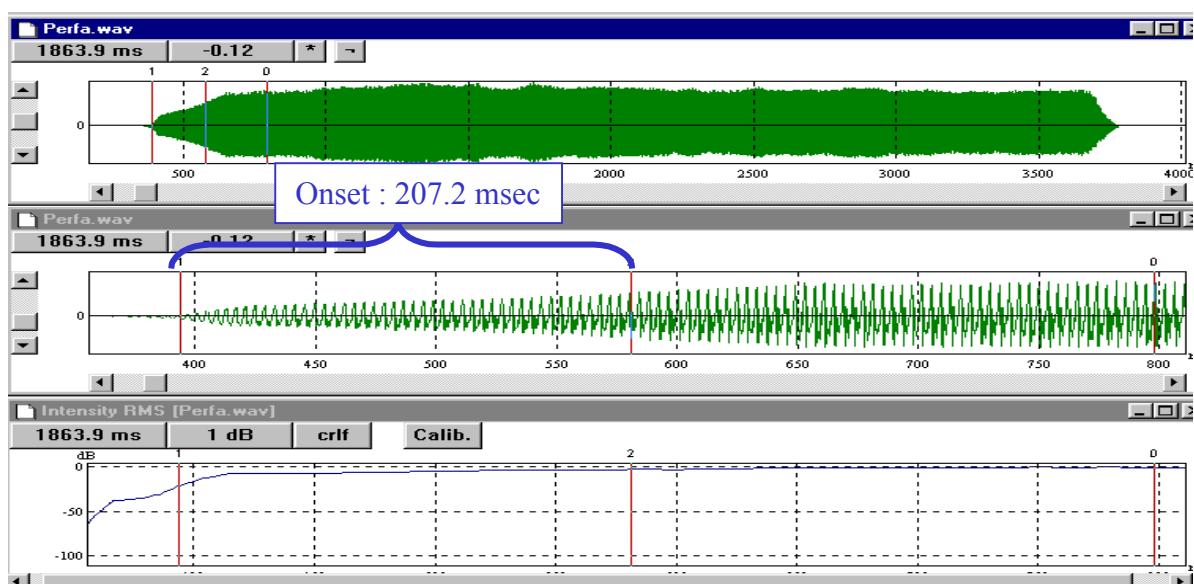


Figure VI.i.7. : Cas particulier: attaque progressive chez le sujet possédant une bonne culture vocale

2- Dans les cas de dysphonie sévère, nous avons distingué 2 sous groupes : les cas où le signal n'atteint jamais la stabilité et les cas où le signal présente une bifurcation.

a- La définition même de la durée de la mise en vibration des cordes vocales précise que l'attaque débute avec le signal vocal et finit lorsque le signal se stabilise. Dans les cas où le signal n'atteint jamais la stabilité, il est possible techniquement d'effectuer la mesure de la durée de l'attaque, mais cela n'a aucune valeur conceptuelle. (Figure VI.i.8.)

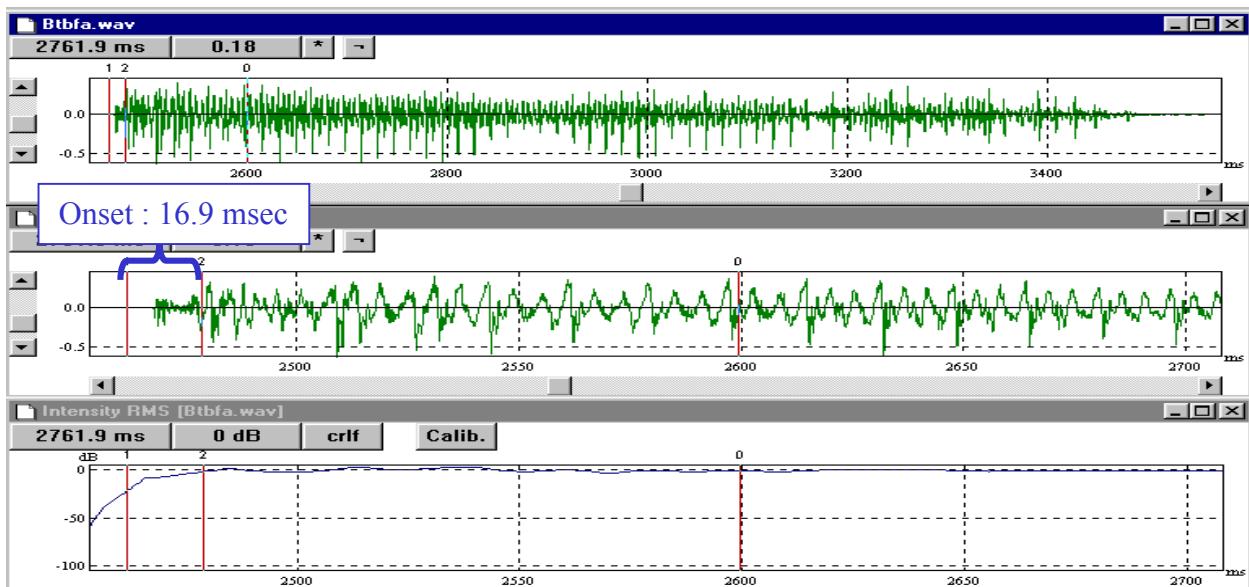


Figure VI.i.8. : Cas particulier: le signal n'atteint jamais la stabilité

b- Dans certains cas de dysphonie sévère, le signal est fortement instable sur toute sa durée, sauf quelques millisecondes de partie stable situées au milieu. Ce changement d'état correspond à une bifurcation dans le signal et non à la stabilisation de la vibration. Si l'on prend la valeur de l'intensité sur cette partie du signal, on obtient une durée de l'attaque démesurément longue, qui ne correspond pas à la mise en vibration. Si l'on prend la valeur d'intensité de référence avant la bifurcation, on retombe dans le cas particulier du signal instable. Mesurer la durée de l'attaque sur un signal comportant une bifurcation n'a pas de sens (figure VI.i.9.).

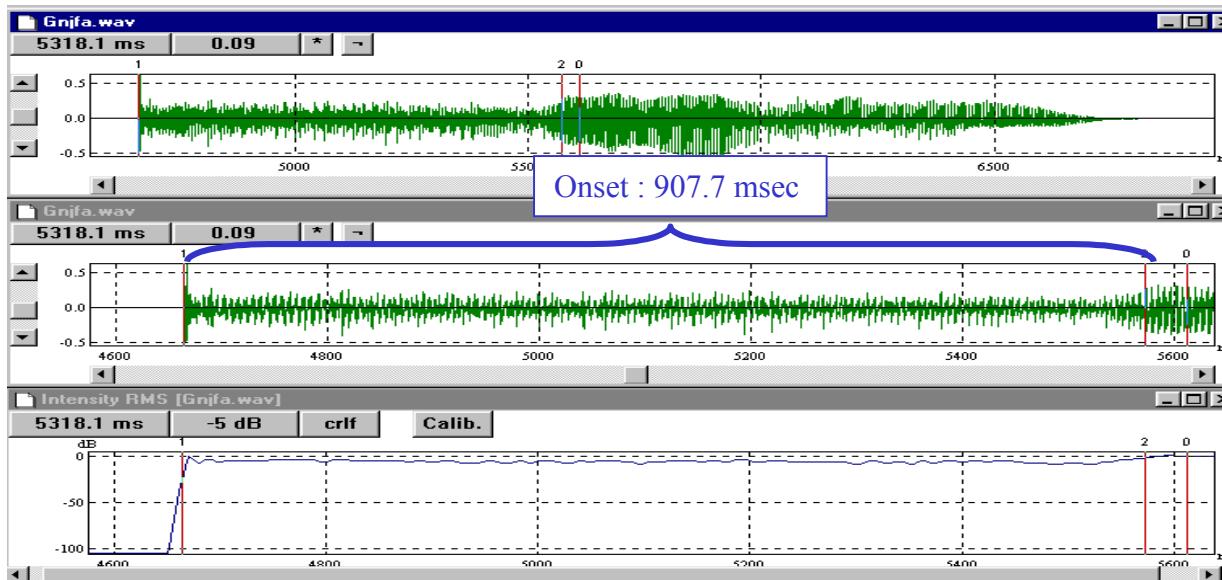


Figure VI.i.9. : Cas particulier: le signal comporte une bifurcation

Conclusion : La méthode des « 10%-90% » permet de mesurer l'attaque et d'en définir les contours. Cette possibilité d'isoler l'attaque ouvre la porte à des mesures qualitatives de l'instabilité contemporaines de l'attaque. Les indices classiques d'instabilité comme le jitter n'étant pas adaptés à une portion du signal trop instable, la mesure qualitative de la mise en vibration des cordes vocales serait donc une application aux méthodes issues des mathématiques non-linéaires et en particulier au coefficient de Lyapunov. L'intégration de cette mesure lors de l'attaque dans les protocoles multiparamétriques de mesures instrumentales pourrait améliorer l'adéquation entre les mesures instrumentales et les évaluations perceptives qui tiennent déjà compte de l'attaque et servent de référence.

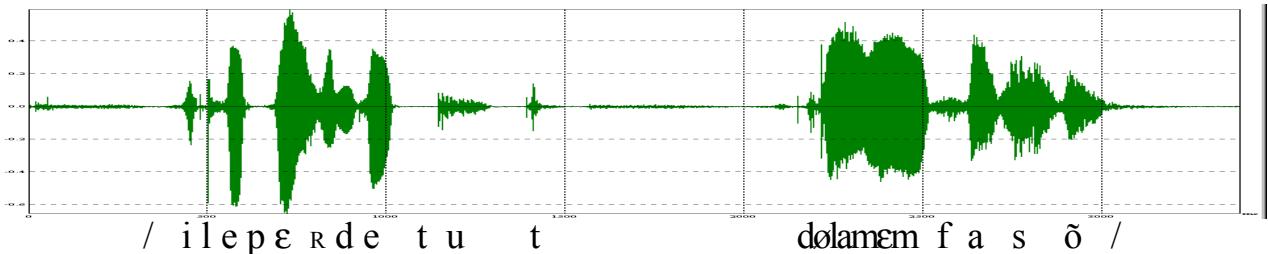
A la suite de ce travail, la méthodologie des « 10%-90% » était confiée à l'équipe « mesures instrumentales » de notre laboratoire afin de tenter une application concrète dans le protocole d'évaluation objective. Notre démarche perceptive ayant exploré successivement les méthodologies d'analyse perceptive, la comparaison avec les données instrumentales, et le rôle de l'attaque dans la détermination du grade pathologique, nous nous retrouvions en bout de course sans véritable résolution du problème posé au départ : quel est le modèle de fonctionnement de la perception de la dysphonie ? et comment améliorer la fiabilité de l'analyse perceptive ?

Nous avons alors fait un point sur l'état de nos connaissances. A ce stade, nous savions que l'oreille humaine possède de grandes capacités de discrimination des différents niveaux pathologiques. Le jury d'écoute peut utiliser toute la dynamique d'une échelle visuelle analogique et il est capable de déterminer des nuances fines entre les stimulus qui lui sont proposés. Néanmoins, dans ces tâches d'évaluation, les jugements effectués sont peu reproductibles et induisent une large variabilité inter- et intra-individuelle. Les résultats sont donc difficilement exploitables dans les démarches expérimentales visant à fournir une expertise des voix proposées. En revanche, il est possible de classifier les voix en grands groupes, avec une reproductibilité d'environ 65% [Kreiman, 1994]. On remarque également que les échelles ordinaires à 4 niveaux (normal, léger, moyen, sévère) sont très performantes sur les extrêmes (normal et sévère) mais que la discrimination des dysphonies intermédiaires (léger et moyen) reste plus aléatoire. Ceci va dans le sens de l'empreinte décrite par Goldstone [Goldstone, 1998] et des standards internes de Gerratt [Gerratt, 1993], selon lesquels, ayant tous une grande habitude d'entendre des voix normales, nous sommes donc tous capables, indépendamment du niveau d'expérience, de déterminer l'absence de pathologie ou sa présence flagrante. D'ailleurs, les travaux proposés sur la formation du jury d'écoute [Chan, 2000], se heurtent également aux difficultés de discrimination des dysphonies intermédiaires. De la même façon, les difficultés rencontrées dans l'approche instrumentale sont particulièrement flagrantes dans les cas de dysphonie légère et moyenne.

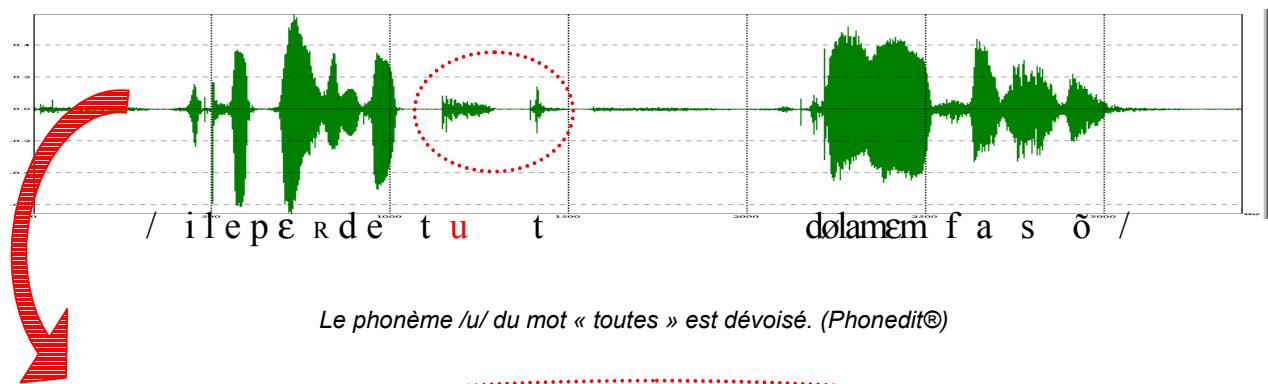
A l'inverse, l'étude spécifique des phénomènes contemporains de la mise en vibration des cordes vocales, l'attaque, a montré que cette portion du signal contenait une information dysphonique utile dans la discrimination des dysphonies intermédiaires (grades 1 et 2) [Revis, 2002]. De même, les cas particuliers que nous avons rencontrés dans l'étude précédente [Revis, 2000] concernaient les voix supra-normales (grade 0) et les voix très dysphoniques (grade 3) qui n'atteignaient jamais la stabilité ou de façon « accidentelle ». Notre méthodologie était donc particulièrement adaptée à la discrimination des dysphonies intermédiaires. Ainsi, pour résumer, toutes les différentes études menées sur le comportement du jury d'écoute se sont heurtées aux difficultés de discrimination fine, alors que les travaux s'intéressant à certaines particularités du signal de parole (l'attaque) ont permis d'ouvrir des perspectives sur la détermination des grades intermédiaires. Suite à ces réflexions, notre conviction était que notre recherche devait s'orienter vers une approche plus phonétique, au cœur du signal de parole.

Jusqu'à présent, nous envisagions la dysphonie comme un bruit intervenant au cours de la parole que nous évaluions à partir d'échantillons de voix de lecture. Le jugement obtenu tenait compte de l'impression pathologique globale éprouvée à l'écoute des stimulus. A cette étape, nous nous sommes posé la question suivante : quels sont les indices intervenant au cours de la parole qui nous font percevoir une voix comme dysphonique ? Nous avons alors procédé à une écoute systématique des

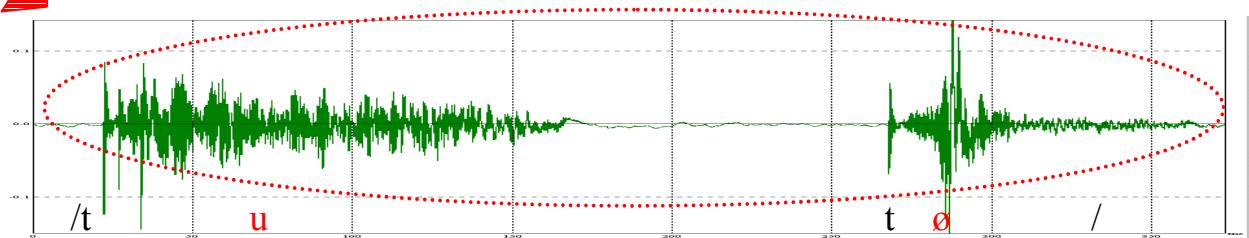
nombreuses voix constitutives de notre banque de données ainsi qu'à l'examen attentif des signaux acoustiques correspondants. Cette démarche a permis de mettre en évidence un événement récurrent chez certains de nos sujets dysphonique : de manière non systématique, mais suffisamment répétitive pour attirer notre attention, on retrouvait un dévoisement sur le mot « toutes » de la phrase « *il les perdait toutes de la même façon* », qui, à lui seul, justifiait l'impression dysphonique éprouvée sur la phrase (« *La petite chèvre de Monsieur Seguin* », §1, Alphonse Daudet).



Signal acoustique de parole de la phrase « il les perdait toutes de la même façon » (Phonedit ®)



Le phonème /u/ du mot « toutes » est dévoisé. (Phonedit®)



Zoom sur le mot « toutes ». On observe un signal très instable sur le /u/ qui correspond en fait au bruit de chuchotement de celui-ci. Le zoom permet également de remarquer que le /ø/ n'est pas omis comme nous l'avions pensé initialement, mais chuchoté brièvement, lui-aussi.

Le fait que cette observation intervienne de manière récurrente, sur le même phonème, à l'écoute de plusieurs voix distinctes, nous a conduit à penser que cela n'était pas dû au hasard, mais plus vraisemblablement à la possibilité d'une influence phonétique. Notre hypothèse était que la perception de la dysphonie fonctionne de la même façon que la perception d'un accent régional, avec des caractéristiques phonétiques (liées aux phonèmes) et linguistiques (liées à la courbe mélodique). En effet, de la même façon que certains accents sont caractérisés par l'allongement spécifique d'un phonème ou la nasalisation d'un son particulier, la dysphonie ne se manifeste pas par une modification

globale du signal (un bruit constant), mais par l'apparition d'évènements rattachés à certains phonèmes au cours de la parole. Ces évènements correspondraient à des occurrences pathologiques qui interviendraient en fonction du contexte phonétique et de la courbe mélodique. Pour Laver [Laver, 81], « *la qualité de la voix est une abstraction accumulative qui s'organise comme une sommation d'évènements pathologiques* ». L'évaluation globale du signal n'aurait alors plus de sens et l'analyse devrait s'orienter vers l'écoute spécifique et la caractérisation de chaque phonème. Ces réflexions inaugurent l'orientation de nos travaux vers un axe résolument phonétique. Dans notre étude suivante, nous nous sommes donc posé la question suivante : une nouvelle méthodologie basée sur l'étiquetage de chaque phonème d'une phrase permettrait-elle de décrire la dysphonie de manière plus précise que l'analyse perceptive classique ?

*j. Phonetic labeling of dysphonia :
a new perspective in perceptual analysis*

Revis J, Nguyen N, Giovanni A

(J Voice, soumis à publication)

Introduction : Le travail que nous présentons est une étude descriptive des caractéristiques pathologiques de chaque phonème intervenant au cours de la production de parole chez des patients dysphoniques. En 2002, Lorch [Lorch, np] a présenté une étude des caractéristiques perceptives de la dysphonie spasmodique en fonction de la langue maternelle des patients : français ou anglais. Elle a procédé à l'analyse perceptive de chacun des phonèmes constitutifs d'un échantillon de parole spontanée (100 syllabes) recueilli chez 6 patients. Les paramètres retenus étaient : 1) les breaks phonatoires, 2) la raucité, 3) le souffle, et 4) la voix soupirée, comprenant le dévoisement des fricatives, l'aspiration des phonèmes, et le silence articulatoire. Les résultats ont montré que certains contextes phonétiques déclenchaient l'apparition d'occurrences pathologiques. Pour toutes les voyelles étiquetées rauques, 74% étaient précédées d'une consonne voisée, 18% d'une consonne non-voisée, et 8% étaient isolées. Il existe donc une influence du contexte phonétique sur l'apparition de la raucité. Les résultats concernant les autres paramètres étudiés ne sont pas précisés dans l'article. La dysphonie spasmodique est une pathologie peu fréquente et possède un mode de fonctionnement marginal dans le champ des dysphonies. Nous avons adapté le protocole utilisé par Lorch à un vaste champ de pathologies dysphoniques, avec un matériau phonétique réduit et standardisé, chez un plus grand nombre de patients.

Protocole expérimental : L'étude a été menée sur 20 sujets sains (12 femmes et 8 hommes) et 60 sujets dysphoniques (42 femmes et 18 hommes) âgés de 19 à 83 ans (moyenne d'âge : 45,6 pour les femmes ; 49,4 pour les hommes). Les pathologies présentées par les patients étaient des nodules dans 24 cas (22 femmes et 2 hommes), un polype dans 9 cas (1 femme et 8 hommes), un œdème Reinke dans 11 cas (7 femmes et 4 hommes), une paralysie laryngée unilatérale dans 6 cas (3 femmes et 3 hommes), des kystes dans 6 cas (6 femmes), et 4 cas post phonochirurgie (3 femmes et 1 homme). L'examen des patients et l'enregistrement étaient réalisés selon les procédures habituelles.

Nous avons exploité une phrase tirée d'un texte utilisé habituellement pour les bilans de dysphonie (« La petite chèvre de Monsieur Seguin » d'Alphonse Daudet) : « Il les perdait toutes de la même façon. » Cette phrase a été choisie car elle était déjà disponible dans notre banque de données et

présentait des configurations contextuelles diverses : voyelle initiale, combinaisons de phonèmes voisés, occlusives voisées et non voisées, etc. Il ne s'agit en aucune manière d'un texte « phonétiquement équilibré ».

Phrase extraite du texte lu : « Il les perdait toutes de la même façon. »

Transcription phonétique : / i l e p ε R d e t u t d ø l a m ε m f a s ð /

Un jury d'écoute composé de 7 auditeurs expérimentés (3 phoniatres et 4 orthophonistes du service habitués à participer ensemble à des jury d'écoute) a procédé à une analyse perceptive classique, d'après les paramètres G (grade global) R (raucité) B (souffle) du GRBAS [Hirano, 1981] sur une échelle ordinaire à 4 niveaux (0 : normal ; 1 : altération légère ; 2 : altération moyenne ; 3 : altération sévère). Les écoutes se sont déroulées de façon individuelle, à 3 reprises avec une semaine d'intervalle. Les résultats servaient de référence pour l'étiquetage des phonèmes.

Nous avons retenu 5 paramètres pour l'étiquetage des phonèmes : 2 paramètres dysphoniques (raucité et souffle), 2 paramètres phonétiques (aspiration et creak) et un paramètre mixte (dévoisement). Les deux paramètres phonétiques ont été intégrés au protocole afin d'éviter une confusion possible entre des paramètres proches perceptivement : le creak (voix « craquée » en fin de phrase sur les fréquences les plus graves de la courbe intonative) qui est un phénomène normal et la raucité qui est un paramètre dysphonique ; et l'aspiration qui correspond à une brève expiration d'air et qui constitue un trait phonétique, et le souffle qui est un paramètre dysphonique. L'étiquetage a été réalisé par une orthophoniste formée aux techniques phonétiques, sur un ordinateur dédié. Le signal acoustique de la phrase était affiché grâce au logiciel Cool Edit © et chaque phonème faisait l'objet d'une écoute spécifique. Les résultats ont été recueillis sur un tableur (Excel ©).

La procédure, dite « en entonnoir », que nous avons utilisée a été proposée par le laboratoire de phonétique Parole et Langage (CNRS, Aix en Provence-France) afin de s'affranchir des difficultés liées à l'évaluation de la dysphonie sur un matériau aussi court qu'un phonème isolé. L'orthophoniste qui a réalisé l'étiquetage des phonèmes écoutait dans un premier temps l'intégralité de la phrase. A ce stade, elle effectuait la transcription phonétique de la production du sujet et se faisait une idée du niveau pathologique global. Ensuite, elle écoutait le premier mot qu'elle classait en normal ou pathologique. S'il était classé pathologique, elle recherchait les phonèmes atteints à l'intérieur de ce mot. Une fois les occurrences repérées, elle procédait à leur caractérisation en fonction des paramètres retenus. Cette opération était renouvelée pour chaque mot de la phrase et nécessitait au total environ 15 minutes par stimulus. La procédure était donc très longue et demandait une concentration intense. L'analyse de la totalité du corpus a nécessité 20 heures de travail, répartis en 5 jours à raison de 2 sessions de 2 heures par jour.

Résultats :

1- Description des productions recueillies chez les sujets

Le nombre de syllabes et de phonèmes produits par les sujets sur la phrase retenue était variable en fonction de la prononciation et de l'accent régional des patients. Nous avons observé les configurations suivantes :

Accent neutre : / i l e p ε R d e t u t d ø l a m ε m f a s ñ /

10 syllabes ; 22 phonèmes (9 patients)

Débit rapide : / i l e p ε R d e t u d ø l a m ε m f a s ñ /

10 syllabes ; 21 phonèmes (13 patients)

ou / i l e p ε R d e t u t d l a m ε m f a s ñ /

9 syllabes ; 20 phonèmes (1 patient)

Accent Marseillais : / i l e p ε R d e t u t ø d ø l a m ε m f a s ñ /

11 syllabes ; 23 phonèmes (26 patients)

ou / i l e p ε R d e t u t ø d ø l a m ε m ø f a s ñ /

12 syllabes ; 24 phonèmes (27 patients)

ou / i l ø l e p ε R d e t u t ø d ø l a m ε m ø f a s ñ /

13 syllabes ; 26 phonèmes (2 patients)

Erreurs de lecture : / i l e p ε R d e t u t ø l a d ø l a m ε m ø f a s ñ /

13 syllabes ; 26 phonèmes (1 patient)

ou / i l e p ε R d e t u t ø d ø l a z m ε m f a s ñ /

10 syllabes ; 24 phonèmes (1 patient)

Au total, nous avons observé une moyenne de 11,1 syllabes (écart-type de 0,9) et une moyenne de 22,9 phonèmes produits (écart-type de 1,2). Pour le recueil des résultats nous avons retenu la configuration la plus fréquente 24 phonèmes 12 syllabes, et les phonèmes non-prononcés par les patients « accent neutre / débit rapide » ont été noté « omission ». Les phonèmes ajoutés correspondant aux erreurs de lecture n'ont pas été pris en compte pour l'étude dans un souci d'uniformisation du matériau phonétique.

2- Analyse perceptive classique

L'analyse perceptive classique nous a permis de classer les sujets en 4 groupes : voix normale (grade 0), dysphonie légère (grade 1), dysphonie moyenne (grade 2) et dysphonie sévère (grade 3). Le grade attribué correspondait à la note attribuée de façon majoritaire par les 7 auditeurs sur les 3 sessions (soit un total de 21 jugements par voix), et nous avons tenu compte de l'indice de confiance : par exemple, un patient noté 15 fois grade 0 et 6 fois grade 1 intégrait le groupe voix normale avec un indice de confiance de 71,4% ; un patient noté 2 fois grade 0, 13 fois grade 1 et 6 fois grade 2 intégrait le groupe

dysphonie légère avec un indice de confiance de 61,9%. L'indice de confiance nous a permis de vérifier à posteriori les cas litigieux.

Nous avons mesuré la variabilité intra- et inter-individuelle de notre jury d'écoute grâce au coefficient Kappa de Cohen (tableau VI.j.1.). Les résultats montrent que les variabilités observées au sein de notre jury sont convenables et conformes à la littérature [De Bodt, 96]

	Inter	Intra
G	0.553	0.633
R	0.475	0.482
B	0.458	0.462

Tableau VI.j.1. : calcul du coefficient Kappa qui renseigne sur les variabilités inter et intra-individuelles du jury d'écoute dans l'analyse perceptive classique du grade global, de la raucité et du souffle.

3- Données quantitatives de l'étiquetage phonétique

Pour chaque patient, nous avons relevé :

- le nombre d'occurrences de Raucité
- le nombre d'occurrences de Souffle
- le nombre d'occurrences de Dévoisement
- le nombre d'occurrences d'Aspiration
- le nombre d'occurrences de Creak
- le nombre total d'occurrences tous paramètres confondus
- le nombre total de phonèmes atteints

Puis nous avons calculé le pourcentage d'occurrences par rapport au nombre total de phonèmes prononcés et par rapport au nombre de phonèmes voisés. Enfin, nous avons calculé le ratio x/y qui donne le nombre de critères par phonème atteint, puisqu'un phonème pouvait être par exemple à la fois rauque et soufflé. Les valeurs obtenues étaient comprises entre 1 (1 seul critère par phonème atteint) et 1,94. Ces résultats ont été utilisés pour l'analyse comparative entre l'analyse perceptive « classique » et l'étiquetage des phonèmes.

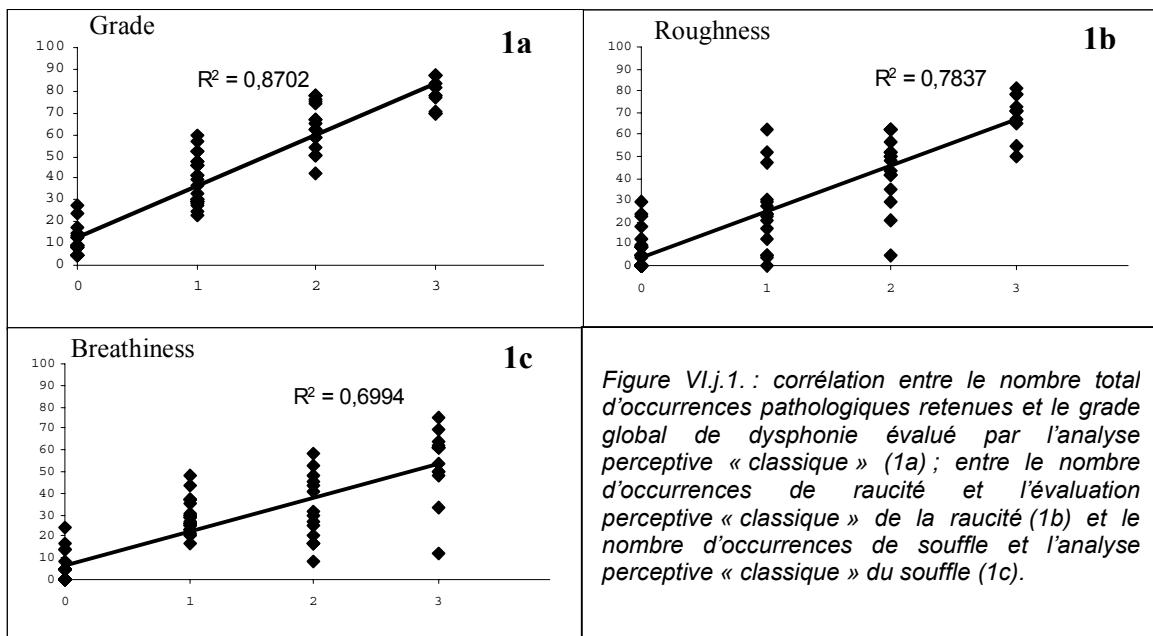
4- Analyse comparative étiquetage phonétique / analyse perceptive « classique »

Le tableau VI.j.2. montre la moyenne et l'écart-type des valeurs quantitatives de l'étiquetage phonétique en fonction des grades.

	G0	G1	G2	G3
R	0,78 (1,35)	3,00 (2,77)	8,88 (4,86)	15,07 (3,24)
R%	3,52 (6,22)	13,37 (12,65)	38,32 (21,25)	64,24 (12,44)
B	0,52 (0,79)	6,20 (2,45)	7,28 (3,24)	12,80 (3,80)
B%	2,32 (3,54)	26,71 (9,94)	31,89 (14,77)	54,46 (15,51)
D	0,04 (0,21)	0,36 (0,64)	2,24 (2,63)	2,20 (2,78)
D%	0,21 (0,99)	1,58 (2,82)	9,75 (11,40)	9,41 (11,67)
A	0,30 (0,63)	0,84 (1,03)	1,29 (0,77)	0,67 (1,05)
A%	1,37 (2,84)	3,35 (4,37)	5,60 (3,33)	2,86 (4,42)
C	0,61 (0,58)	0,20 (0,41)	0,47 (0,62)	0,13 (0,35)
C%	2,77 (2,69)	0,86 (1,77)	2,04 (2,71)	0,62 (1,65)
Ttl Phoneme	10,19 (6,10)	38,28 (10,69)	64,59 (11,18)	77,98 (7,51)
Pho Voisés	11,35 (7,12)	44,69 (11,52)	71,08 (12,02)	90,30 (11,84)

Tableau VI.j.2. : moyenne (écart-type) des valeurs quantitatives recueillies lors de l'étiquetage phonétique

La figure VI.j.1. montre la corrélation entre le nombre total d'occurrences pathologiques attribué par l'étiquetage phonétique et le grade obtenu par l'analyse perceptive classique.



Le tableau VI.j.3. montre la moyenne et l'écart-type du ratio x/y en fonction du grade. Le coefficient de détermination entre le ratio x/y et le grade global donné par l'analyse perceptive « classique » est de 0,68, ce qui montre une forte corrélation entre le niveau de dysphonie et le nombre de critères pathologiques par phonème atteint : plus la dysphonie est importante, plus on retrouve de phonèmes atteint par plusieurs critères. Pour les grades 0 (voix normales), le pourcentage de phonèmes présentant une occurrence est toujours inférieur à 25% et ils ne sont atteints que par un seul critère (ratio x/y = 1).

x/y	G0	G1	G2	G3
Moy	1	1,19	1,37	1,69
E-Type	0	0,18	0,19	0,2

Tableau VI.j.3. : le ratio x/y explore le nombre de critère par phonème atteint. Ce tableau présente la moyenne et l'écart-type du ratio x/y en fonction des grades déterminés par l'analyse perceptive « classique ».

5- Analyse descriptive chez les sujets sains

Les figures VI.j.2a. et 2b. montrent la répartition des occurrences pathologiques retrouvées chez les sujets témoins en fonction du sexe. La barre rouge représente la moyenne du nombre de phonèmes prononcés par les sujets (n=24).

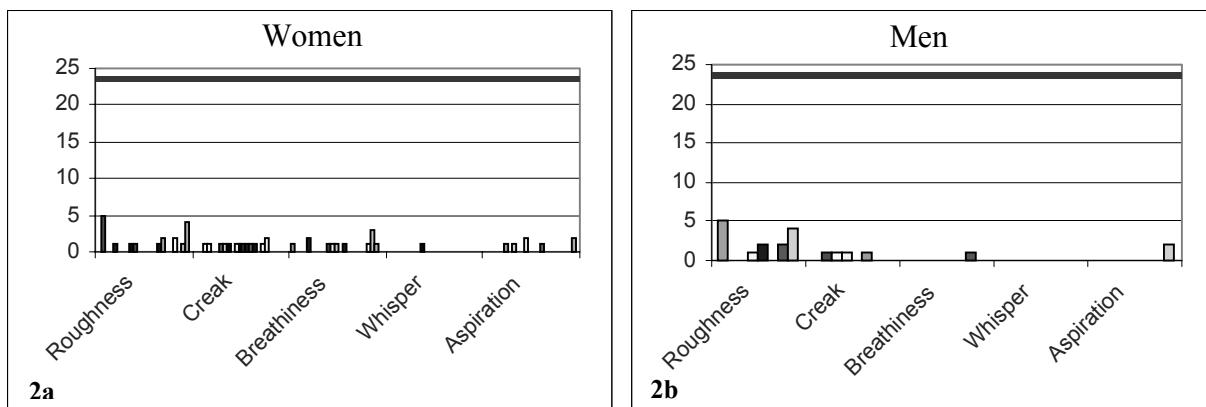
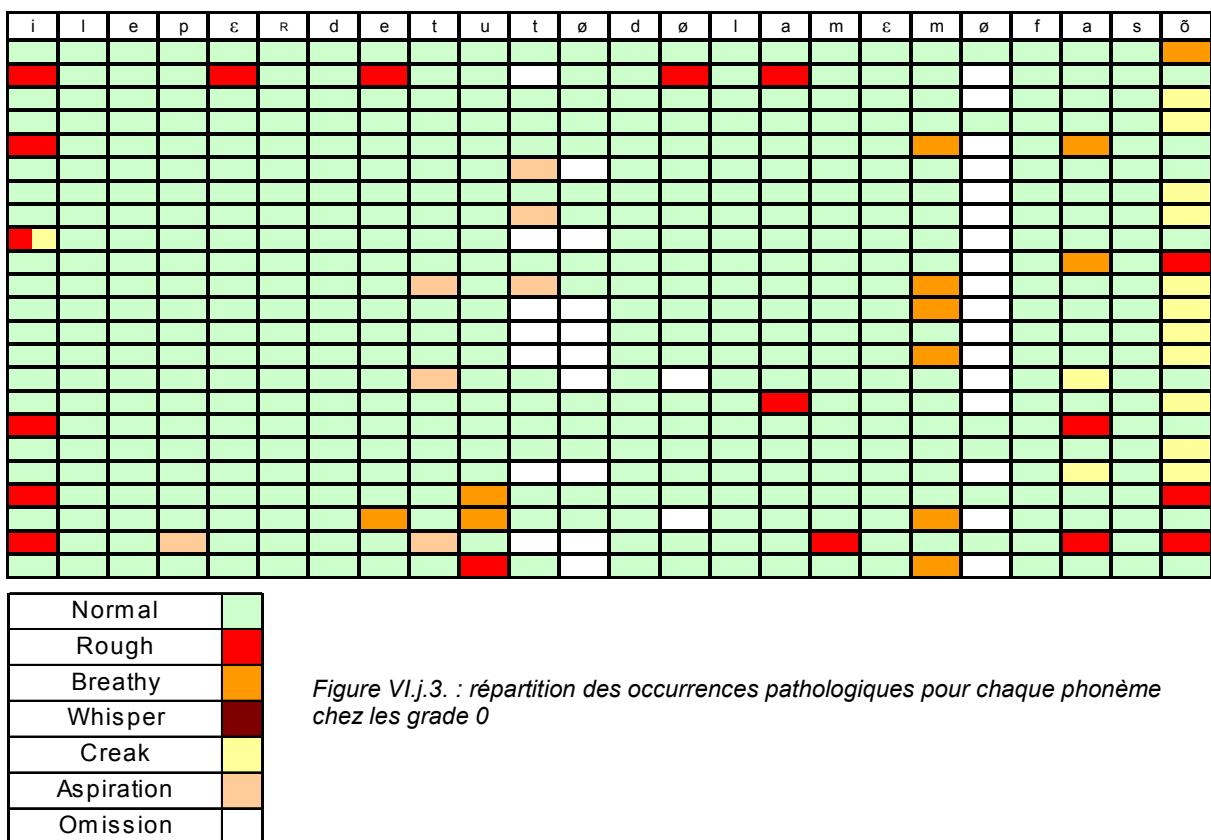


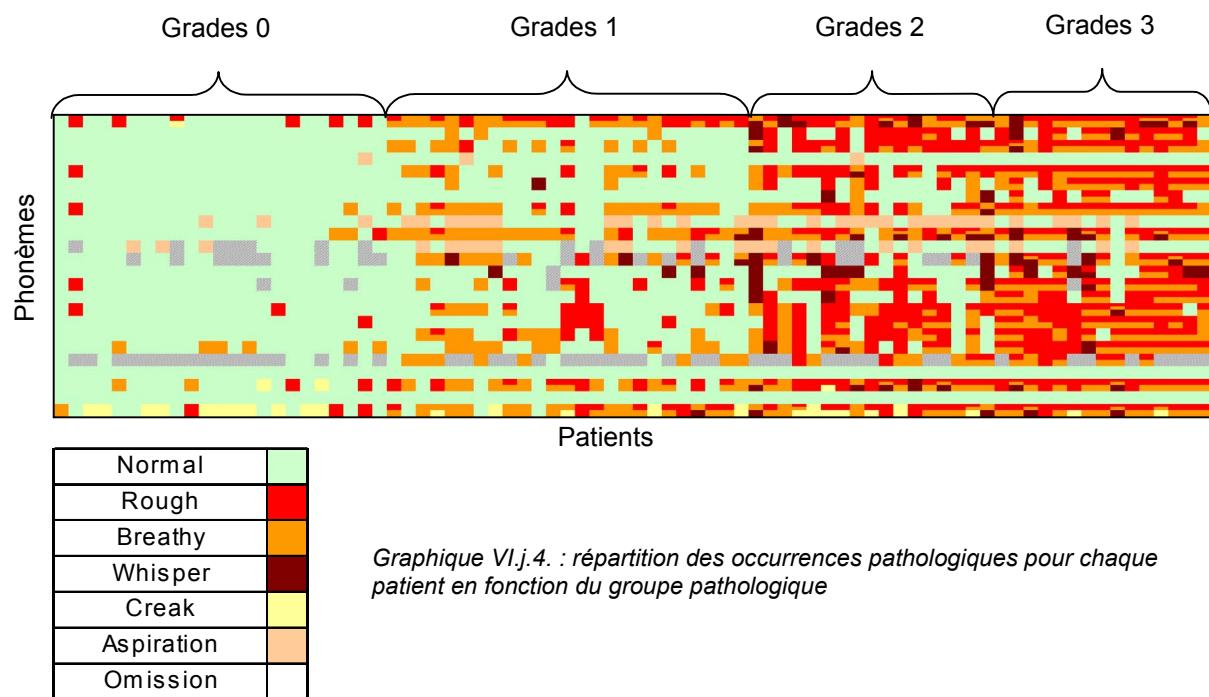
Figure VI.j.2. : répartition des occurrences pathologiques chez les sujets normaux : les femmes témoins (2a) et les hommes témoins (2b)

La figure VI.j.3. montre la répartition phonémique des occurrences pathologiques chez les sujets témoins. Une couleur arbitraire a été attribuée à chaque paramètre afin de faciliter l'apprehension du graphique. L'intensité de couleur n'indique en aucun cas un niveau d'intensité de la pathologie, mais permet de distinguer au premier coup d'œil le paramètre concerné. Comme on peut le voir dans la légende, le vert est attribué aux phonèmes normaux, les niveaux les plus clairs concernent les paramètres phonétiques (aspiration et creak) et les niveaux foncés concernent les paramètres dysphoniques (rauque, soufflé et dévoisé). Les hachures correspondent aux phonèmes qui n'ont pas été prononcés par les sujets.



6- Analyse descriptive chez les sujets pathologiques

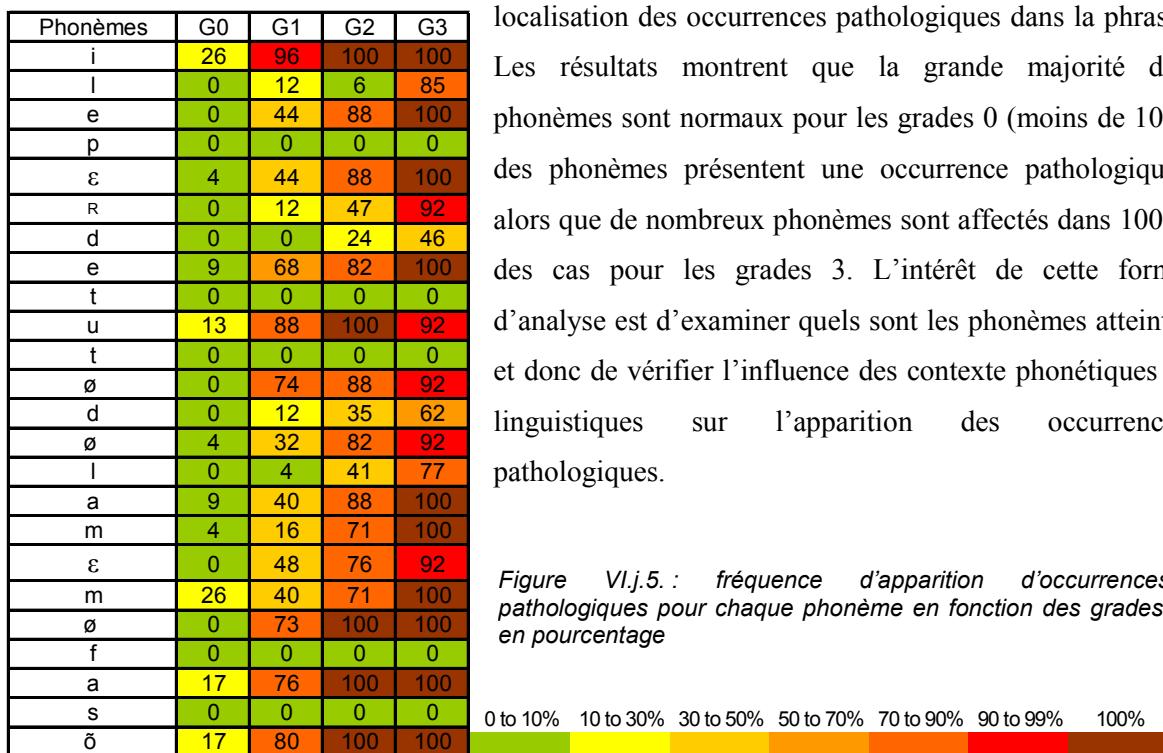
La figure VI.j.4. présente la répartition des occurrences pathologiques pour chaque phonème en fonction du grade. Les phonèmes produits sont présentés en ordonnée et chaque patient rangés par ordre croissant de dysphonie sont en abscisse.



La figure VI.j.5. montre la fréquence (en pourcentage) d'occurrences pathologiques, tous paramètres confondus, pour chaque phonèmes en fonction du grade. Ce graphique permet d'observer la

localisation des occurrences pathologiques dans la phrase. Les résultats montrent que la grande majorité des phonèmes sont normaux pour les grades 0 (moins de 10% des phonèmes présentent une occurrence pathologique) alors que de nombreux phonèmes sont affectés dans 100% des cas pour les grades 3. L'intérêt de cette forme d'analyse est d'examiner quels sont les phonèmes atteints, et donc de vérifier l'influence des contexte phonétiques et linguistiques sur l'apparition des occurrences pathologiques.

Figure VI.j.5. : fréquence d'apparition d'occurrences pathologiques pour chaque phonème en fonction des grades, en pourcentage



Discussion : L'analyse perceptive des dysphonie est habituellement réalisée sur de la parole continue, prenant en compte la globalité du signal. Dans une précédente étude [Revis, 1999], nous avions demandé à un jury d'écouter de réaliser la tâche d'évaluation sur des voyelles tenues afin de tester les possibilités de discrimination de niveaux pathologiques sur un matériau phonétique simplifié. Les résultats montraient qu'il est possible et valide d'utiliser la voyelle tenue pour analyser la dysphonie. Néanmoins, le jury s'était unanimement plaint du caractère artificiel de la voyelle isolée. Dans la parole continue, la durée des phonèmes est très réduite (de l'ordre de 5 à 50 millisecondes). Tenant compte de cette information, il paraissait très difficile de réaliser une analyse perceptive de chaque phonème séparément. La procédure « en entonnoir » que nous avons utilisée nous a permis de nous affranchir de ces difficultés.

Une seule évaluation a été retenue pour l'analyse des résultats. Cependant, l'orthophoniste qui a réalisé l'étiquetage phonétique avait effectué l'évaluation des voix selon la même méthode une première fois en pré-test afin de vérifier la faisabilité de la procédure. Une comparaison des résultats obtenus lors de ces deux sessions d'évaluation a montré une reproductibilité d'environ 75% pour les paramètres RBD ($r^2 = 0.748$) et d'environ 55% pour les paramètres phonétiques CA ($r^2 = 0.554$). Cependant, le pré-test a été réalisé dans des conditions non-expérimentales et les résultats ne fournissent qu'une idée de la reproductibilité de la méthode. En conséquence, il serait intéressant de

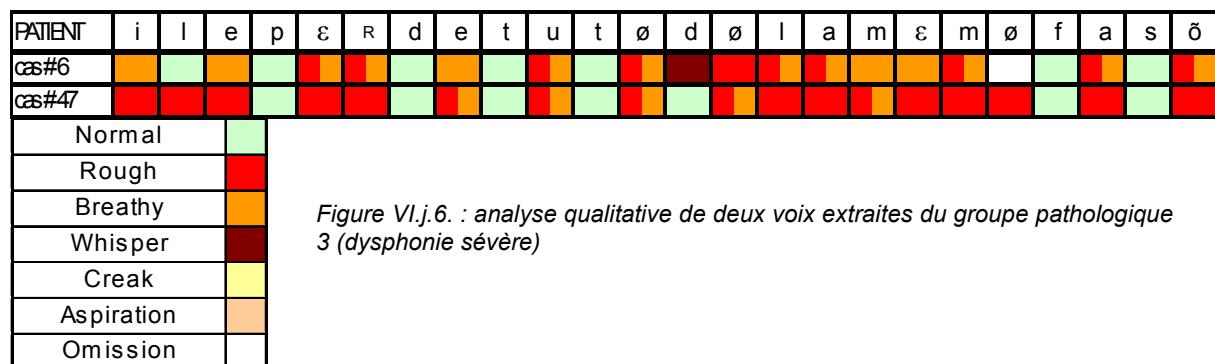
renouveler l'expérience avec 3 évaluations afin de vérifier la reproductibilité de notre procédure. Nous émettons l'hypothèse que la fiabilité de cette méthode d'analyse est meilleure que pour l'analyse perceptive « classique », dans la mesure où l'identification des caractéristiques pathologiques en présent / absent, sur chaque phonème est à la fois plus facile et plus représentative que l'analyse globale du signal. Ce point de vue fait l'objet d'une étude en cours qui vérifiera également la variabilité inter-individuelle de la méthode par le recrutement de plusieurs auditeurs supplémentaires.

Le nombre total de phonèmes atteints a été calculé en comptabilisant tous les paramètres retenus pour l'étude. Les paramètres phonétiques de creak et d'aspiration ne sont pas représentatifs en soi de la dysphonie, puisqu'ils sont décrits comme des caractéristiques de la voix normale au cours de la parole. Cependant, nous avons fait le choix de les intégrer à la mesure, partant du postulat que la pathologie vocale pouvait faire émerger des critères phonétiques. En effet, on peut supposer que les difficultés vibratoires des cordes vocales engendrent des phénomènes plus fréquents de creak qui peut correspondre à une stratégie de repos, ou d'aspiration par la présence audible de la fuite d'air. En pratique, nous avons observé que les paramètres phonétiques étaient plus présents dans les voix normales et les dysphonies légères, que dans les dysphonies plus importantes. Cela peut être dû au fait que l'intensité du niveau de présence des critères dysphoniques masque les paramètres plus discrets dans l'impression perceptive globale. Les critères de creak et d'aspiration n'étaient donc pas très représentatifs du niveau de dysphonie. Cependant, leur intégration dans le calcul du nombre total de phonèmes atteints a gonflé les chiffres, en particulier pour les voix normales et les dysphonies légères. Ainsi, le cas 22 (figure VI.j.7.) qui est classé en groupe 1, montre une atteinte de 50% des phonèmes, alors que si l'on supprime les occurrences des critères phonétiques, l'atteinte n'est plus que de 37,5%. Par ailleurs, au cours de la tâche d'étiquetage, certains phénomènes sont apparus dans la production des dysphonies moyennes et sévères, avec l'émergence d'une énergie excessive des fricatives non-voisées. Cette constatation s'est produite suffisamment souvent pour suggérer un effet dysphonique, cependant, elle n'a pas été retenue dans ce protocole. Ces observations illustrent la nécessité de réfléchir à nouveau au choix des critères que nous retiendrons pour nos prochains travaux.

L'analyse descriptive des sujets sains (figure VI.j.3.) montre que les productions vocales des témoins ne sont pas exemptes d'occurrences pathologiques. Aucun des cas examinés dans cette étude ne présentait un nombre d'occurrences égal à 0. Le taux maximum d'occurrences pathologique retrouvé dans les cas de voix normales était de 25% (Moyenne : 10,19% ; Ecart-type : 6,10) avec un ratio critères/phonèmes atteints toujours égal à 1, ce qui signifie que les phonèmes n'étaient atteints que par un seul critère. Si l'on regarde les résultats obtenus en fonction du sexe, on relève que 53% des femmes témoins présentaient des occurrences de souffle et 65% des hommes témoins des occurrences de raucité. Cela suggère que les femmes ont une tendance naturelle au souffle et les hommes à la raucité. Si les analyses instrumentales portant sur les différences en fonction du sexe ont révélé des

dissemblance quantitatives, en particulier dans la mesure du jitter [Schoentgen, 1989], l'analyse perceptive par étiquetage des phonèmes montre une dissemblance qualitative.

L'analyse descriptive de l'ensemble de notre corpus (figure VI.j.4.) montre que le nombre d'occurrences augmente très visiblement avec l'importance de la dysphonie. On observe que la couleur dominante des grades 0 (voix normales) est le vert, couleur qui était attribuée arbitrairement aux phonèmes normaux. Puis, progressivement, l'apparition de couleurs sombres (attribuées aux phonèmes pathologiques) s'intensifie, jusqu'à ce que le graphique devienne essentiellement sombre pour les grades 3 (dysphonies sévères). La figure VI.j.1. montre qu'il existe une forte corrélation entre les résultats obtenus par cette méthode d'étiquetage des phonèmes et l'analyse perceptive classique. Cela suggère que notre méthode est très représentative de la perception dysphonique globale, tout en fournissant un niveau de détail supérieur. Les difficultés de l'analyse perceptive rapportée par les différents membres de notre jury d'écoute, résident dans le classement au sein d'un même groupe de deux voix très différentes. La méthode d'étiquetage des phonèmes permet à la fois le classement en groupe pathologique et la nuance entre les qualités spécifiques des voix au sein du groupe (figure VI.j.6.).



La figure VI.j.6. montre que les deux cas exposés présentent un fort taux d'occurrences pathologiques (cas 6 : total phonèmes 69,57% ; phonèmes voisés 88,89% ; ratio x/y 1,56 ; et cas 47 : total phonème 70,83% ; phonèmes voisés 89,47% ; ratio x/y 1,24). Les deux sujets ont été classés groupe 3 par la procédure d'analyse perceptive classique. Néanmoins, l'observation de la représentation en couleurs, montre que si les deux voix sont équivalentes sur un plan quantitatif global, elles sont très différentes sur le plan qualitatif, avec une prévalence du souffle pour le cas 6 (60,87%) et une prévalence rauque pour le cas 47 (70,83%). La méthode d'étiquetage des phonèmes permet donc, outre une quantification globale fiable, une cartographie qualitative précise de la dysphonie.

Cette précision se révèle particulièrement utile dans l'analyse des cas mal explorés par l'analyse perceptive « classique ». Par exemple, le cas 22 a été classé dans le groupe 2 avec un indice de

confiance de 33,3%, ce qui veut dire que la voix de cette patiente a été classée G1 (dysphonie légère) 7 fois, G2 (dysphonie moyenne) 7 fois, et G3 (dysphonie sévère) 7 fois. Il s'agit d'un cas extrême de désaccord du jury. Dans ce cas, l'analyse perceptive classique ne permet de fournir aucune information valide. Le classement définitif en groupe 2 (qui correspond au jugement global moyen) est alors purement arbitraire et pas plus fondé qu'un classement en groupe 1 ou 3. Pour ce cas précisément, l'étiquetage phonétique fournit une description détaillée la voix de cette patiente (figure VI.j.7.) et permet de s'affranchir de l'approximation du jury. Le cas 22 présente un total de 14 occurrences pathologiques pour 12 phonèmes atteints, soit 50 % du nombre total de phonèmes produits, avec un ratio x/y (critères/phonème) de 1,17. Si l'on se réfère aux tables de moyennes et écart-type des données quantitatives (tableaux VI.j.2. et VI.j.3.), les valeurs du cas 22 correspondent à un grade 1. A posteriori, le cas 22 peut donc être reclassé en groupe 1.

PATIENT	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
cas#22	red	light green	red	light green	red	light green	red	orange	light orange	red	orange	dark red	light green	red	light green	light green	light green	light green	orange	light green	red	light green	light green	yellow
Normal		light green																						
Rough			red																					
Breathy				orange																				
Whisper					dark red																			
Creak						yellow																		
Aspiration						orange																		
Omission							white																	

Figure VI.j.7. : étiquetage phonétique du cas 22

En ce qui concerne la localisation des occurrences dans la phrase, on remarque une double influence, phonétique et linguistique. La considération phonétique permet d'observer à l'influence du contexte de la phrase sur l'apparition des occurrences. Pour cela (figure VI.j.5.), nous avons observé le nombre total d'occurrences pathologiques intervenant sur chaque phonème (en pourcentage). Dans 66% des cas, on retrouve une occurrence de souffle sur le dernier phonème placé avant le /f/. Cela pourrait indiquer qu'il existe un phénomène de fusion entre le dernier phonème voisé (/m/ ou /ø/) et la fricative non-voisée /f/. La fusion entre les phonèmes a été décrite par Laver [Laver, 1980, 1981, 1991] comme un phénomène d'« attraction » du premier phonème par le second. En effet, lors de la prononciation du /m/ ou du /ø/, le sujet prépare la consonne suivante /f/ qui est une fricative non-voisée. Il y a donc émergence d'un bruit de souffle qui peut être interprété comme une occurrence pathologique de souffle, tout en étant un phénomène de fusion normale. En effet, ce phénomène est observé chez le sujet témoin dans 22% des cas. Pour Lorch [Lorch, np], une présence anormale de souffle peut se manifester en Français dans une grande variété de contextes, comme l'aspiration des consonnes occlusives, ou la proximité d'une fricative non-voisée. Il est donc également possible que la préparation de la fricative non-voisée engendre une difficulté vibratoire et augmente la fuite glottique, ce qui pourrait mettre en valeur le souffle pathologique d'une voix dysphonique.

Certains contextes phonétiques semblent favoriser l'apparition d'occurrences comme dans le mot « toutes » (/tut/), où la voyelle fermée /u/ est située entre deux occlusives non-voisées. En effet, chez les sujets sains, le phonème /u/ est atteint dans 19% des cas, puis dans 82% des grades 1 (dysphonie légère), 100% des grades 2 (dysphonie moyenne) et 96% des grades 3 (dysphonie sévère). Ohala [Ohala, 1993] a montré que les voyelles fermées comprises entre deux occlusives non-voisées subissaient une tendance au dévoisement. Cette observation illustre à la fois l'effet de fusion entre les phonèmes et la difficulté de mise en vibration des cordes vocales. La présence fréquente d'occurrences pathologiques sur le phonème /u/ est donc la conséquence d'une difficulté vibratoire d'autant plus importante que le niveau dysphonique est élevé.

Dans un même ordre d'idées, il semble exister une influence de l'attaque sur l'apparition des occurrences. Tous les phonèmes ne présentent pas une attaque : lors d'une combinaison de plusieurs phonèmes voisés (/d ø l a m ε m /), l'attaque n'intervient que sur le premier phonème (/d/) puis la vibration des cordes vocales reste active sur les phonèmes suivants, sans interruption. Notre travail montre que les voyelles présentant une attaque (phonème initial /i/ et phonèmes situés après une consonne non-voisée) sont très fréquemment atteints (environ 15% chez les grades 0 ; 70% chez les grades 1 ; 90% chez les grades 2 et 100% chez les grades 3). Lors de précédents travaux, nous avons montré l'effet de l'attaque sur la perception de la dysphonie [Revis, 1999 ; Revis, 2000]. En effet, la mise en route des cordes vocales est rendue précaire par l'existence d'une lésion laryngée, et cette difficulté constitue une information pathologique qui est essentielle pour la détermination du niveau dysphonique global. Pour Rees (1958) [in Laver, 1980], le niveau de dysphonie sur les voyelles est davantage marqué pour les voyelles isolées ou présentant une attaque. Nos résultats confirment donc l'impact de la mise en vibration des cordes vocales sur la perception dysphonique.

Enfin, on observe que les phonèmes non-voisés ne sont pas affectés par les occurrences pathologiques dysphoniques (raucité, souffle, dévoisement) dans la mesure où la dysphonie apparaît exclusivement sur la vibration glottique. En revanche, ils peuvent être atteints spécifiquement par les critères phonétiques et particulièrement le paramètre d'aspiration.

La considération linguistique fait intervenir la courbe mélodique induite par les effets d'intonation et d'accentuation. Pour tous les groupes pathologiques, les phonèmes correspondant à une accentuation sur la courbe mélodique sont affectés (phonème initial /i/, phonème médian accentué /u/, et phonème final /on/). Cela confirme l'influence de la courbe mélodique et donc du niveau linguistique, sur l'apparition des occurrences pathologiques.

Enfin, on peut également envisager une application clinique immédiate de la méthode de « Phonetic Labeling », en particulier dans la comparaison de la qualité de la voix d'un même patient avant et

après thérapeutique. Nous avons appliqué cette méthode au cas d'une patiente qui présentait un kyste sur la corde vocale gauche et qui a été opérée dans notre service. Nous disposions de l'enregistrement réalisé la veille de l'intervention et nous avons enregistré sa voix à nouveau, selon les procédures habituelles 3 mois après la microchirurgie laryngée. Durant cette période, notre patiente avait également suivi une rééducation orthophonique. Le tableau VI.j.4. montre les résultats de l'analyse perceptive classique avant et après traitement.

	Grade	Roughness	Breathiness
BTBF before surgery	3	3	2
BTBF after surgery	1	0	1

Tableau VI.j.4. : résultats de l'analyse perceptive classique chez la patiente « btbf » avant et après microchirurgie laryngée.

On observe que la voix de cette patiente est largement améliorée par l'intervention puisque sa dysphonie passe de sévère à légère, avec disparition de la raucité qui était très importante avant l'opération (R3), et une diminution du souffle qui passe de moyen (B2) à léger (B1) en post-traitement. Nous avons alors procédé à l'analyse de sa voix par « Phonetic Labeling ». Le graphique du nombre d'occurrences pathologiques par paramètre (figure VI.j.8.) permet de retrouver les résultats obtenus par analyse perceptive classique.

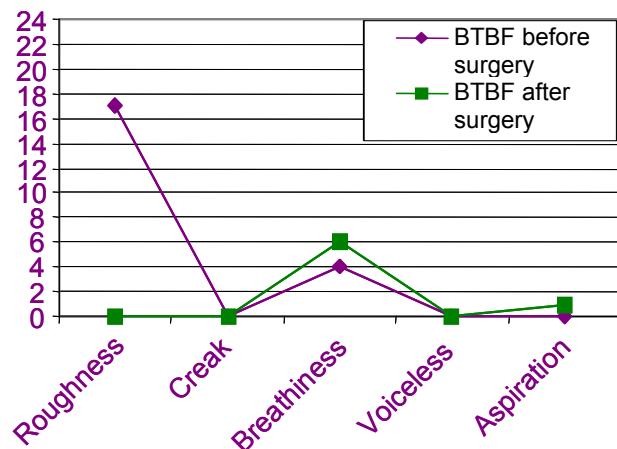


Figure VI.j.8. : évolution du nombre d'occurrence pathologiques en fonction des paramètres étudiés avant et après microchirurgie laryngée.

On peut constater que le nombre global d'occurrences pathologique est diminué de façon drastique après l'intervention, avec une disparition des occurrences de raucité et une persistance du souffle. L'étude du nombre d'occurrences pathologiques en fonction du paramètre étudié permet donc de retrouver les résultats de l'analyse perceptive classique. Enfin, l'analyse graphique de la répartition des occurrences en fonction des phonèmes permet de fournir un niveau d'informations supérieur (figure VI.j.9.).

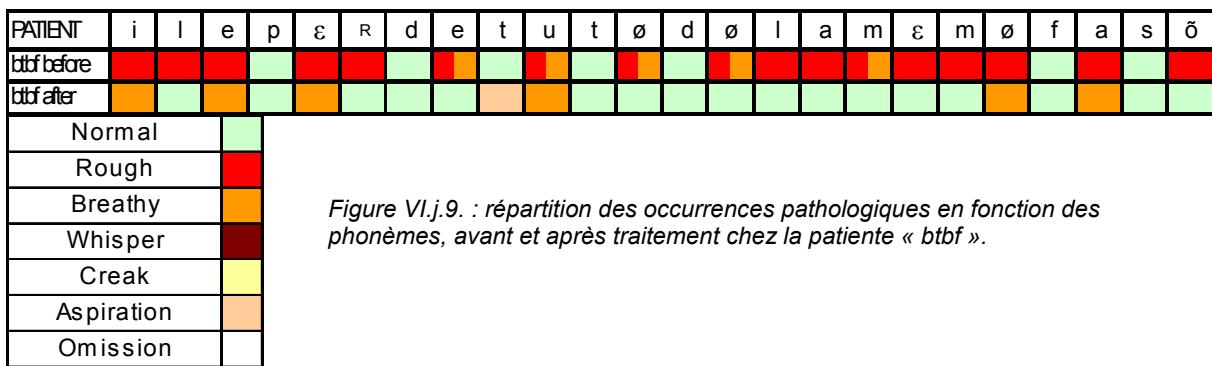


Figure VI.j.9. : répartition des occurrences pathologiques en fonction des phonèmes, avant et après traitement chez la patiente « btbf ».

Cette figure permet de visualiser au premier coup d’œil l’évolution de la voix de cette patiente : on note l’apparition de nombreux phonèmes normaux (en vert), la disparition des occurrences de raucité et la persistance du souffle. De plus, on constate que la plupart des occurrences pathologiques résiduelles apparaissent après un silence (phonème initial et phonèmes précédés d’un phonème non-voisé). En d’autres termes, les phonèmes touchés sont ceux qui présentent une attaque vocale. Cette constatation peut fournir une information clinique intéressante pour l’orientation de la rééducation de cette patiente, vers un travail spécifique de l’amorce du son.

Conclusion : Les apports potentiels de la méthode que nous proposons sont nombreux. Tout d’abord, les voix normales ne sont pas exemptes d’occurrences pathologiques. Lorsque le nombre de ces occurrences reste inférieur à 25% des phonèmes prononcés et que le ratio x/y (critère par phonème) est égal à 1, la voix peut être considérée comme normale. Nous avons également remarqué que les sujets témoins masculins avaient une tendance naturelle à la raucité, alors que pour les sujets témoins féminins, cette tendance était plutôt soufflée. Les résultats montrent que le nombre d’occurrences est fortement corrélé avec le niveau dysphonique déterminé par analyse perceptive « classique », ce qui suggère que la perception fonctionne comme la sommation d’événement dysphoniques intervenant au cours de la parole. Concernant l’impact linguistique, la fréquence d’apparition des occurrences sur le phonème initial, le phonème accentué médian et le phonème final, montre l’influence de la courbe mélodique. Néanmoins, dans la phrase que nous avons choisi, le phonème accentué médian correspond également à un contexte phonétique favorisant le dévoisement. Le contexte phonétique a une influence majeure sur l’apparition des occurrences pathologiques. Enfin, cette étude met en évidence l’influence de l’onset sur l’apparition des occurrences comme sur la perception dysphonique globale.

La conception et l'application de la méthode de « phonetic labelling » a ouvert de nombreux axes de réflexion sur la manifestation phonétique de la dysphonie au cours de la parole. Néanmoins, cette étude réalisée avec l'intervention d'un seul auditeur, particulièrement sensibilisé à l'écoute des voix dysphoniques et au problème de l'analyse perceptive, courrait le risque que la méthode soit « manipulateur dépendant », c'est à dire que seul cet auditeur permette de retrouver ces résultats. Il était donc absolument nécessaire de reprendre ce travail avec plusieurs auditeur afin de vérifier la reproductibilité des mesures et de valider la méthode. C'est l'objectif du travail que nous présentons maintenant.

k. Reproductibilité et validation de la méthode de « Phonetic labeling »

Revis J, Gannat M, Giovanni A

(J Voice, soumis à publication)

Introduction : Les nombreux intérêts et les perspectives ouvertes par la méthode de « phonetic labelling » ont été présentés dans l'article précédent à l'issue duquel nous avions conclu sur la nécessité d'étudier la reproductibilité des résultats. Le travail que nous présentons a pour objectif de vérifier la faisabilité de la méthode sur un jury d'écoute formé de plusieurs auditeurs. Pour cela, nous avons choisi de faire écouter un nombre limité de voix dysphonique (8) à un maximum d'auditeurs (20). En effet, l'étude portant spécifiquement sur la méthode et non sur la dysphonie en soi, nous avons privilégié l'effectif du jury par rapport au nombre de voix à évaluer. Les auditeurs ont effectué leur analyse trois fois à une semaine d'intervalle afin que nous puissions observer les reproductibilités intra-individuelles aussi bien qu'inter-individuelles. Notre hypothèse est que la méthode de phonétic labelling est davantage reproductive que l'évaluation perceptive classique dans la mesure où l'analyse est à la fois plus précise qualitativement (chaque phonème) et plus facile à quantifier (présent/absent).

Protocole expérimental : Les voix dysphoniques ont été sélectionnées dans notre banque de données. Comme dans l'étude précédente, le matériau phonétique était constitué de la phrase « *il les perdait toutes de la même façon* », extrait du premier paragraphe de « La petite chèvre de Monsieur Seguin » d'Alphonse Daudet. Nous avons retenu 8 enregistrements expertisés par consensus de 3 orthophonistes expérimentées selon le grade G de l'échelle de Hirano [Hirano, 1981] : 2 voix normales, 2 dysphonies légères, 2 dysphonies moyennes et 2 dysphonies sévères. Tous les locuteurs étaient féminins afin de constituer un échantillon homogène à partir duquel les auditeurs puissent élaborer des modèles internes les plus stables possibles.

Le jury d'écoute expérimental était constitué de 20 auditeurs recrutés dans le service ORL du CHU Timone : 5 praticiens spécialistes de la dysphonie (2 médecins phoniatres et 3 orthophonistes) et 15 étudiants orthophonistes en dernière année de formation. Du fait du caractère original et novateur de la méthode, une séance individuelle de formation au phonetic labelling était donnée à chacun des auditeurs durant environ 90 minutes avant le début des analyses. Cette séance s'est déroulée dans une salle dédiée, avec le même matériel informatique que pour l'expérimentation afin que les auditeurs soient familiarisés avec le programme d'édition du signal et l'environnement (ordinateur Network MD9465, processeur Pentium III, 128 Mo de RAM, carte son ESS Allegro PCI Audio, casque Sony

MDR-013, logiciel CoolEdit® 96). L'entraînement était effectué sur deux voix dysphoniques recrutées dans la banque de données pour leur richesse en occurrences pathologiques. La voix d'exemple 1 était analysée par l'examineur en présence de l'auditeur avec tous les détails explicatifs pour chacune des procédures et chacun des phonèmes. Le détail de l'analyse était recueilli sur une fiche fournissant la transcription phonétique des deux voix d'entraînement (document VI.k.1.).

Entraînement

Document VI.k.1. : Fiche de notation pour la séance de formation.

La voix d'exemple 2 était analysée par l'auditeur sous la direction de l'examinateur qui confirmait ou corrigeait les manipulations informatiques et les interprétations. Les 20 membres du jury ont tous reçu les mêmes explications. A l'issue de la séance de formation, une notice explicative était donnée aux auditeurs afin de figer les informations acquises lors de l'entraînement (document VI.k.2.).

L'analyse perceptive par la méthode de « Phonetic Labelling »

Le logiciel de traitement du signal « CoolEdit @ » vous permet d'afficher le signal acoustique pour l'évaluation des voix.

– Quantification de la dysphonie :

La première écoute de l'intégralité de la phrase vous permet de vous faire une idée du niveau pathologique global. Cliquez sur « Play ».



– Ecoute spécifique isolée :

Pour écouter séparément chaque groupe vocalique, on sélectionne un groupe par un « clic and drag » maintenu sur le signal. Le groupe sélectionné apparaît alors sur fond turquoise. Ensuite, on clique sur « Zoom ».



Seul le groupe vocalique sélectionné apparaît alors sur l'écran.

– Identification des occurrences pathologiques :

À cette étape, posez vous la question : « est-ce que j'entends un événement pathologique dans ce groupe vocalique ? »

Si la réponse est oui, l'écoute doit être affinée en effectuant séparément chaque phonème grâce au « Zoom » jusqu'à déterminer quel(s) phonème(s) est(sont) affecté(s). Si la réponse est non, vous pouvez passer au groupe vocalique suivant.

Vous avez la possibilité de naviguer entre les différents niveaux possible (phrase/groupe vocalique/phonème) autant de fois que nécessaire afin d'être sûr(e) de la localisation de l'occurrence pathologique.

– Définition des paramètres qualitatifs :

Au fur et à mesure vous devez remplir la grille de cotation comme indiqué par l'examineur en inscrivant l'initial de des paramètres concernés sous chaque phonème affecté. Pour les phonèmes normaux, la case est laissée vide.

Les paramètres étudiés sont :

- Rauçité (R) impression d'irrégularité de la vibration des cordes vocales
- Souffle (S) impression d'un flux d'air audible entre les cordes vocales
- Dévoilement (D) le phonème est chuchoté, les cordes vocales ne vibrent pas
- Creak (C) impression de rugosité sur les fréquences les plus graves
- Aspiration (A) bruit de « h » aspiré à la fin d'une occlusive non voisine

Les 8 voix doivent être écoutées au casque, de façon consécutive, au cours de 3 sessions distinctes d'une semaine. Une pause de 20 minutes est autorisée au cours d'une session.

Nous vous remercions de votre collaboration et vous souhaitons bon travail.

Document VI.k.2. : Notice explicative de la méthode de phonetic labelling, mise à la disposition des auditeurs

Après cette formation, les 20 auditeurs participant à l'étude ont effectué leurs analyses individuellement, lors de 3 sessions distantes d'une semaine. A chaque session, les 8 voix devaient être analysées de façon consécutive, une pause de 20 minutes étant autorisée à mi-parcours du fait de la durée de la tâche (20 à 30 minutes par voix soit environ 2h30 par session). Au cours des différentes sessions, les voix étaient présentées dans un ordre aléatoire différent afin de diminuer les effets de contexte et d'apprentissage. Le signal acoustique était affiché grâce au logiciel CoolEdit 96 ® et chaque phonème était segmenté pour l'écoute isolée. Les résultats étaient recueillis individuellement par les auditeurs sur un feuillet spécifique à chacune des sessions (document VI.k.3.). Les paramètres de dysphonie retenus étaient : Raucité, Souffle et Dévoisement. L'initiale du ou des critères concernés étaient inscrits à la main par l'auditeur dans l'espace réservé en dessous du phonème atteint.

Session 1																							
Nom de l'auditeur :								Prénom :															
Voix 1 :								Grade :															
i	l	e	p	ɛ	r	d	e	t	u	+	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
Remarques :																							
Voix 2 :								Grade :															
i	l	e	p	ɛ	r	d	e	t	u	+	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø

Document VI.k.3. : Feuillet de cotation fourni aux auditeurs, reprenant la transcription phonétique de chaque voix proposée.

Puis l'examinateur recueillait tous les résultats à l'aide d'un tableur (Excel®).

Résultats : Les résultats expriment la reproductibilité intra- et inter-individuelle des auditeurs par le pourcentage de réponses identiques en fonction du niveau dysphonique global, des phonèmes prononcés et des paramètres dysphoniques RBD. Le tableau VI.k.1. montre les résultats obtenus sur les deux voix normales (G0) : « CXGF » et « LESF » concernant la raucité, le souffle et le dévoisement au cours des 3 sessions d'écoute. Les chiffres représentent le pourcentage des auditeurs ayant côté le phonème pathologique. La couleur représente l'agrément inter-individuel comme indiqué dans la légende. Les cases grisées indiquent les phonèmes non prononcés par le locuteur ; ils sont désignés comme « omis », leur présence correspondant à un accent régional. Ainsi, pour la voix CXGF session 2, le phonème initial a été côté rauque par 5% des auditeurs, ce qui signifie qu'il a été classé « non-rauque » par 95% des auditeurs. L'agrément inter-individuel sur ce phonème concernant le paramètre de raucité est donc de 95% lors de la 2^{ème} session d'écoute.

CXGF	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð	
Session																									
1	0	0	0	0	5	0	0	5	0	0	0	0	0	0	15	0	0	0	0	0	10	0	0	25	
2	5	0	0	0	0	0	0	5	0	0	0	0	10	0	15	0	0	0	0	0	15	0	0	30	
3	0	0	0	0	0	0	0	5	0	0	0	0	5	0	5	0	5	0	0	5	0	0	10	0	30
Session																									
1	15	0	5	0	5	0	0	10	0	10	0	0	10	0	10	0	10	0	5	5	5	0	20	0	25
2	10	0	0	0	10	0	0	10	0	15	0	0	5	5	0	10	5	5	5	0	15	0	0	5	
3	0	0	0	0	10	0	0	5	0	10	0	0	5	10	0	5	5	0	0	0	0	20	0	5	
Session																									
1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LESF																									
Session																									
1	0	0	0	0	0	0	0	10	0	0	0	0	5	0	0	0	5	0	0	15	0	20			
2	10	0	5	0	15	0	0	10	0	5			0	20	10	15	0	10	5		0	5	0	20	
3	5	0	5	0	0	0	0	0	0	0	0	0	5	0	20	0	10	0		0	20	0	15		
Session																									
1	5	0	15	0	5	0	0	30	0	15			0	0	0	10	0	5	5		0	30	0	10	
2	5	0	15	0	5	0	0	25	0	25			0	5	0	10	0	0	10		0	10	0	0	
3	5	0	0	0	5	0	0	10	0	0			0	0	0	5	0	5	0		0	5	0	0	
Session																									
1	0	0	0	0	0	0	0	0	0	5			0	0	0	0	0	0	0		0	0	0	5	
2	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0	0		0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0	0		0	0	0	0	

	Agrément inter-individuel
Omission	
< 69%	
70-84%	
85-100%	

Tableau VI.k.1. : agrément inter- et intra-individuel de la méthode de phonetic labelling sur les voix normales (grade 0). Résultats présentés pour chaque phonème en fonction des paramètres de raucité, souffle et dévoisement au cours des 3 sessions d'écoute.

Le tableau VI.k.1. montre que la reproductibilité des analyses par phonetic labelling sur les voix normales est supérieure à 85% dans une large majorité des cas, et jamais inférieur à 70%. D'une manière générale, on observe que les phonèmes sont perçus comme normaux par le jury d'écoute, puisque les chiffres relevés sont tous inférieurs ou égaux à 30%. Dans ce cas, on considère que seuls 30% des auditeurs considèrent le phonème atteint par un critère pathologique. Il s'agit donc d'une déviance par rapport à la majorité (70% des auditeurs considèrent le phonème normal) qui exprime la variabilité des analyses. Les résultats récapitulatifs des occurrences pathologiques sont montrés dans le tableau VI.k.2.

		VOIX NORMALE (G0) : C X G F																						
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
		VOIX NORMALE (G0) : L E S F																						
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð

Tableau VI.k.2. : récapitulatif de l'apparition et de la localisation des occurrences pathologiques pour les voix normales. Un phonème est considéré comme pathologique lorsque plus de 50% des auditeurs notent le paramètre comme présent.

Le tableau VI.k.2. montre que tous les phonèmes prononcés par les locuteurs sains sont considérés comme normaux.

Le tableau VI.k.3. montre les résultats obtenus sur les deux dysphonies légères (G1) : « LICF » et « PEJF » concernant la raucité, le souffle et le dévoisement au cours des 3 sessions d'écoute.

LICF	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Session																								
1	80	5	40	0	35	5	5	30	0	15	0		5	70	5	80	15	45	15		0	90	0	40
2	55	10	30	0	40	15	0	20	0	15	0		0	55	10	90	15	40	0		0	70	0	5
3	65	0	35	0	30	30	5	30	0	5	0		0	65	5	80	5	30	5		0	65	0	40
Session																								
1	35	0	45	0	45	5	0	90	0	95	0		5	10	5	35	5	65	20		0	40	0	50
2	55	15	35	0	40	5	0	85	0	90	0		0	25	5	45	0	45	20		0	70	0	70
3	50	10	40	0	40	5	10	75	0	95	0		0	30	5	30	5	50	10		0	45	0	60
Session																								
1	5	0	0	0	0	0	0	0	0	0	0		15	0	0	0	0	0	0		0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0		15	0	0	0	0	0	0		0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0		10	0	0	0	0	0	0		0	0	0	0
PEJF																								
1	30	0	30	0	50	10	0	50	0	20	0		0	45	5	60	5	75	20	15	5	55	0	35
2	50	5	35	0	60	5	5	20	0	20	0		0	45	20	60	5	60	40	20	0	55	0	40
3	15	0	10	0	65	0	0	20	0	20	0		0	40	0	65	10	50	25	20	0	10	0	40
Session																								
1	80	15	55	0	80	5	0	70	0	70	0		0	30	0	15	15	50	10	45	0	60	0	10
2	80	25	70	0	50	5	15	70	0	90	0		15	20	0	20	10	60	5	45	0	70	0	20
3	70	15	60	0	50	10	0	85	0	95	0		10	25	0	25	5	40	15	35	0	65	0	20
Session																								
1	0	0	5	0	0	0	0	0	0	15	0		15	0	0	0	5	30	0	0	0	0	0	0
2	0	0	0	0	0	5	0	0	0	15	0		10	0	0	0	0	0	0	5	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0		5	0	0	0	0	5	0	5	0	5	0	0

	Agrément inter-individuel
Omission	
< 69%	
70-84%	
85-100%	

Tableau VI.k.3. : agrément inter- et intra-individuel de la méthode de phonetic labelling sur les dysphonies légères (grade 1). Résultats présentés pour chaque phonème en fonction des paramètres de raucité, souffle et dévoisement au cours des 3 sessions d'écoute.

En ce qui concerne les dysphonies légères, on constate que les paramètres de raucité et de souffle sont plus variable que le dévoisement. L'observation des résultats sur la raucité montre que 68,1% des évaluations montrent un taux de reproductibilité supérieur à 85% ; 60,6% des évaluations portées sur le souffle montrent un taux de reproductibilité supérieur à 85% ; alors que 100% des évaluations du dévoisement sont reproductibles à plus de 85%. Par ailleurs, il semble également que les évaluations aient été davantage reproductibles pour la voix LICF que pour la voix PEJF.

Le tableau VI.k.4. montre la localisation des occurrences pathologiques retrouvées sur les dysphonies légères.

D Y S P H O N I E L E G E R E (G 1) : L I C F																								
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
D Y S P H O N I E L E G E R E (G 1) : P E J F																								
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð

Normal	
Omission	
Raucité	
Souffle	
Dévoismt	

Tableau VI.k.4. : récapitulatif de l'apparition et de la localisation des occurrences pathologiques pour les dysphonies légères. Un phonème est considéré comme pathologique lorsque plus de 50% des auditeurs notent le paramètre comme présent.

Les résultats montrent que pour les deux dysphonies légères, les voix sont affectées à la fois par la raucité et le souffle, avec une légère dominance du souffle. On n'observe aucune occurrence de dévoisement. Au total, environ 30% des phonèmes sont classés pathologiques pour les dysphonies légères.

Le tableau VI.k.5. montre les résultats obtenus pour les deux dysphonies moyennes (G2) : « FOJF » et « FRMF » concernant la raucité, le souffle et le dévoisement au cours des 3 sessions d'écoute.

FOJF	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Session	R A U C I T E (%)																							
1	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	10	0	10	0	0	10	0	0	0
2	5	0	10	0	10	5	0	0	0	0	0	0	0	5	0	20	5	15	0	0	25	0	10	0
3	20	5	20	0	20	5	0	10	0	0	0	0	0	10	10	20	0	20	0	0	25	0	20	0
Session	S O U F F L E (%)																							
1	90	20	70	0	90	45	20	85	0	15	0	0	15	50	15	70	25	55	25	0	85	0	90	
2	90	10	75	0	85	45	0	100	0	15	0	5	15	55	10	95	15	70	20	0	100	0	100	
3	95	10	75	0	90	30	5	85	0	25	0	20	5	55	10	90	5	75	25	0	95	0	95	
Session	D E V O I S E M E N T (%)																							
1	5	20	0	0	5	5	0	5	0	90	0	15	25	20	10	0	0	0	10	0	5	0	15	
2	5	30	5	0	0	0	5	0	0	85	0	30	40	15	10	0	0	0	0	0	5	0	5	
3	5	15	5	0	0	0	0	0	0	80	0	55	25	5	5	0	0	0	5	0	0	0	5	
FRMF	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Session	R A U C I T E (%)																							
1	30	15	50	0	35	0	5	30	0	35	0	15	10	10	60	5	30	0	15	0	90	0	35	
2	35	15	75	0	55	10	0	25	0	15	0	5	30	15	90	5	30	0	40	0	85	0	40	
3	30	10	60	0	45	0	10	15	0	30	0	5	20	10	85	15	35	10	25	0	85	0	40	
Session	S O U F F L E (%)																							
1	90	10	85	0	50	10	5	95	0	60	0	0	0	50	15	15	0	75	0	80	0	75	0	80
2	80	5	80	0	40	15	15	100	0	85	0	10	45	15	35	10	50	5	60	0	85	0	90	
3	80	0	85	0	45	25	20	100	0	65	0	15	55	15	25	5	75	5	80	0	80	0	90	
Session	D E V O I S E M E N T (%)																							
1	5	0	0	0	0	0	0	30	0	0	10	0	30	15	0	0	0	0	0	0	0	0	0	
2	5	0	0	0	0	0	0	25	10	0	10	0	15	15	0	0	0	0	0	0	0	0	10	
3	5	0	0	0	0	0	0	50	5	0	5	0	10	5	0	0	0	0	0	0	0	0	0	

Agrément inter-individuel	
Omission	
< 69%	
70-84%	
85-100%	

Tableau VI.k.5. : agrément inter- et intra-individuel de la méthode de phonetic labelling sur les dysphonies moyennes (grade 2). Résultats présentés pour chaque phonème en fonction des paramètres de raucité, souffle et dévoisement au cours des 3 sessions d'écoute.

Les résultats montrent une forte reproductibilité quel que soit le paramètre étudié. Cependant, on note un agrément plus fort concernant la voix FOJF (81% des jugements montrent un taux de reproductibilité supérieur à 85%) que pour la voix FRMF (76% des jugements montrent un taux de reproductibilité supérieur à 85%). Par ailleurs, il semble que le paramètre de raucité soit le plus reproductible en ce qui concerne la voix FOJF, contre le dévoisement pour la voix FRMF.

Le tableau VI.k.6. montre la localisation des occurrences pathologiques retrouvées sur les dysphonies moyennes.

	D Y S P H O N I E M O Y E N N E (G 2) : F O J F																							
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
	D Y S P H O N I E M O Y E N N E (G 2) : F R M F																							
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð

Normal	
Omission	
Raucité	
Souffle	
Dévoismt	

Tableau VI.k.6. : récapitulatif de l'apparition et de la localisation des occurrences pathologiques pour les dysphonies moyennes. Un phonème est considéré comme pathologique lorsque plus de 50% des auditeur notent le paramètre comme présent.

Les résultats montrent que les deux dysphonies moyennes sont affectées différemment par les occurrences pathologiques : la voix FOJF est exclusivement soufflée (39% des phonèmes) alors que la voix FRMF est simultanément rauque (17% des phonèmes) et soufflée (39% des phonèmes) avec une nette prépondérance du souffle. Au total, environ 45% des phonèmes sont classés pathologiques pour les dysphonies moyennes.

Le tableau 7 montre les résultats obtenus sur les deux dysphonies sévères (G3) : « GNGF » et « BTBF » concernant la raucité, le souffle et le dévoisement au cours des 3 sessions d'écoute.

BTBF	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ð
Session R A U C I T E (%)																								
1	100	10	80	0	100	45	10	90	0	95	5	80	20	90	25	85	60	95	50	70	0	90	0	85
2	90	25	90	5	95	55	5	95	0	95	5	80	20	90	25	85	55	95	50	85	5	95	0	85
3	95	35	100	5	100	80	10	100	0	95	0	75	15	95	55	100	70	100	65	85	5	100	0	85
Session S O U F F L E (%)																								
1	40	0	55	0	30	0	15	60	0	80	0	40	5	50	5	50	10	40	15	55	0	55	0	15
2	40	5	45	0	20	5	5	75	0	80	0	30	5	50	5	60	10	40	10	45	0	45	0	30
3	45	0	50	0	40	15	5	75	0	55	0	20	5	50	5	60	10	40	10	40	0	40	0	15
Session D E V O I S E M E N T (%)																								
1	0	5	0	0	5	0	0	0	0	0	0	5	5	5	0	15	5	0	5	5	0	5	0	0
2	0	0	0	0	5	0	0	0	0	0	0	5	5	0	0	5	0	0	0	5	0	0	0	0
3	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GNGF i l e p ɛ r d e t u t ø d ø l a m ɛ m ø f a s ð																								
Session R A U C I T E (%)																								
1	70	20	80	5	100	25	0	90	0	40	0	10	80	25	85	15	75	30	5	90	0	65		
2	85	30	90	0	100	35	10	100	0	60	5	20	75	50	100	65	85	85	5	100	0	75		
3	85	15	95	5	95	35	5	80	0	20	0	5	85	20	100	30	70	65	0	90	0	85		
Session S O U F F L E (%)																								
1	60	20	75	0	65	10	15	70	0	75	0	5	45	0	50	10	75	15	0	60	0	80		
2	60	30	80	0	55	15	10	80	0	80	0	10	55	0	40	15	70	5	0	60	0	75		
3	65	35	80	0	55	15	5	95	0	80	0	20	65	0	40	15	75	15	0	70	0	55		
Session D E V O I S E M E N T (%)																								
1	5	0	0	0	0	10	5	0	0	20	0	25	5	0	0	0	10	0	0	0	0	0	0	5
2	5	0	0	0	5	5	5	0	0	20	0	5	0	0	0	0	15	0	0	0	0	0	0	0
3	0	0	0	0	0	5	0	0	10	0	5	0	0	0	0	0	15	0	0	0	0	0	0	0

Agrément inter-individuel	
Omission	
< 69%	
70-84%	
85-100%	

Tableau VI.k.7. : agrément inter- et intra-individuel de la méthode de phonetic labelling sur les dysphonies sévères (grade 3). Résultats présentés pour chaque phonème en fonction des paramètres de raucité, souffle et dévoisement au cours des 3 sessions d'écoute.

Ici encore, les résultats montrent une forte reproductibilité quel que soit le paramètre étudié. Cependant, on note un agrément plus fort concernant la voix BTBF (75% des jugements montrent un

taux de reproductibilité supérieur à 85%) que pour la voix GNGF (70% des jugements montrent un taux de reproductibilité supérieur à 85%).

Le tableau VI.k.8. montre la localisation des occurrences pathologiques retrouvées sur les dysphonies moyennes.

	D Y S P H O N I E S E V E R E (G 3) :															B	T	B	F					
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
	D Y S P H O N I E S E V E R E (G 3) :															G N G F								
Raucité	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
Souffle	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
Dévoismt	i	l	e	p	ɛ	r	d	e	t	u	t	ø	d	ø	l	a	m	ɛ	m	ø	f	a	s	ø
Normal																								
Omission																								
Raucité																								
Souffle																								
Dévoismt																								

Tableau VI.k.8. : récapitulatif de l'apparition et de la localisation des occurrences pathologiques pour les dysphonies sévères. Un phonème est considéré comme pathologique lorsque plus de 50% des auditeurs notent le paramètre comme présent.

Les résultats montrent que les deux dysphonies sévères sont affectées simultanément par la raucité et le souffle. Au total, environ 70% des phonèmes voisés sont classés pathologiques pour les dysphonies sévères.

Discussion : Notre hypothèse était que les analyses par phonetic labelling seraient davantage reproductibles que les évaluations perceptives classiques tout en fournissant une niveau de détail supérieur. Habituellement, le niveau de reproductibilité des évaluations perceptives est d'environ 57% en inter-individuel et 62% en intra-individuel. Les résultats que nous avons obtenus montrent que l'analyse par phonetic labelling permet d'atteindre des taux d'agrément d'environ 80% en intra-individuel aussi bien qu'en inter-individuel, ce qui représente un haut niveau de fiabilité. Néanmoins, du fait de la longueur des analyses, de la finesse des manipulations informatiques pour le séquençage des phonèmes, et de l'attention soutenue nécessaire à l'application de cette tâche, il faut considérer que la méthode de phonetic labelling ne peut être appliquée à la pratique clinique. Il s'agit d'une démarche expérimentale permettant, par son haut niveau de détail, de comprendre le fonctionnement phonétique de la dysphonie. La faisabilité du phonetic labelling nécessitait d'être validée par un plus grand nombre d'auditeurs, il semble que le niveau de reproductibilité de cette analyse soit pleinement satisfaisant.

Les auditeurs ayant pris part à cette étude présentaient un niveau d'expérience hétérogène. En effet, le but de ce travail était de valider la méthode sur un grand nombre d'auditeurs. Ne disposant que de 5 auditeurs expérimentés, habitués à la tâche d'évaluation de la dysphonie, nous avons dû recruter 15 auditeurs supplémentaires dans la promotion des étudiants en dernière année d'orthophonie. Dans la mesure où c'est la méthode qui était évaluée et non la capacité à déterminer différents niveaux de

dysphonie, nous avons considéré que tous les auditeurs étaient inexpérimentés, puisque personne n'avait été formé au phonetic labelling auparavant. L'homogénéité des résultats que nous avons obtenus valide ce choix fait à priori.

La formation des auditeurs aux techniques de traitement du signal nécessaires à l'évaluation par phonetic labelling n'a pas posé de problème particulier. Néanmoins, les différentes manipulations de découpage du signal sont longues et alourdissent considérablement la procédure. De plus, la précision du découpage est essentielle à la pertinence de l'évaluation. Les manipulations individuelles augmentent le risque d'erreurs techniques, donc d'approximation, en particulier dans la localisation des occurrences. Nous avons imaginé proposer aux auditeurs un signal « pré-découpé », en insérant 200 msec de silence entre chaque phonème. Testée sur quelques voix, cette méthode permettait effectivement de s'assurer de la pertinence du découpage, mais rendait impossible la navigation entre les différents niveaux d'écoute : phrase entière / groupe vocalique / phonème. Or, il semble que cet aller-retour entre le plus global et le précis soit absolument nécessaire à la caractérisation des phonèmes et à la localisation des occurrences. Cette méthode a donc été abandonnée.

En ce qui concerne l'observation des contextes phonétiques, cette étude permet de confirmer les points mis en évidence dans notre précédent travail. En effet, il semble que la fiabilité des évaluations soit influencée par les caractéristiques intrinsèques des phonèmes ou par leur localisation contextuelle. Les plus forts taux de reproductibilités observés dans cette étude correspondent aux hypothèses pressenties dans notre étude préliminaire :

- les consonnes non-voisées ne sont pas affectées par les occurrences pathologiques, dans la mesure où la dysphonie apparaît sur la vibration des cordes vocales, donc exclusivement sur les phonèmes voisés. Nos résultats abondent dans ce sens : les occlusives non-voisées /p/ et /t/ et les constrictives non-voisées /s/ et /f/ ne présentent aucun critère de dysphonie. La reproductibilité inter et intra-individuelle pour ces phonèmes atteint 100%.
- certaines structures phonologiques, semblent favoriser l'apparition d'occurrences pathologiques. En particulier, l'enchaînement « occlusive non-voisée / voyelle fermée / occlusive non-voisée » que l'on retrouve dans le mot /tut/ (« toutes ») semble inciter au dévoisement de la voyelle. Dans ce travail, la présence du dévoisement est flagrante sur la voix « FOJF » (tableau VI.k.6.). La qualification et la localisation de l'occurrence pathologique présente une forte fiabilité avec 95% de reproductibilité en intra-individuel et plus de 80% en inter-individuel. D'une manière plus générale, le critère de dévoisement montre une fiabilité plus importante que les autres critères sur l'ensemble des voix. En effet, un phonème est voisé ou non, il n'y a pas de niveau. Le dévoisement est donc vraisemblablement plus facile à évaluer que la raucité ou le souffle qui peuvent être présents avec plus ou moins d'importance. Le dévoisement n'est pas affecté par la sensibilité des auditeurs, son évaluation est donc plus simple.

- l'attaque favorise l'émergence de la dysphonie. Les 5 phonèmes de notre corpus présentant une attaque sont présentés en rouge, ci-dessous.

/ i l e p **ε** r d e t **u** t d ø l a m **ɛ** m f **a** s **ɔ̃** /

« Il les perdait toutes de la même façon. »

Nos résultats montrent non seulement une fréquence d'apparition des occurrences pathologiques supérieure sur ces phonèmes, mais aussi une reproductibilité plus grande de leur évaluation (85% à 95% en inter- et intra-individuel).

Au total, la reproductibilité des mesures effectuées par phonetic labelling varie de 45% à 100% en fonction des critères explorés, des phonèmes et de leur localisation dans la phrase. La moyenne que nous obtenons est de 77% en inter-individuel et de 86% en intra-individuel.

Conclusion : L'analyse perceptive de la dysphonie par phonetic labelling est une méthode nouvelle et innovante de l'exploration de la voix. Son principe, basé sur l'évaluation de chaque phonème, permet de prendre en compte des éléments jusqu'alors négligés dans l'étude des dysphonies : l'influence des contraintes phonétiques et linguistiques. Issu d'une démarche essentiellement expérimentale, le phonetic labelling permet une caractérisation précise des troubles de la voix et de leur manifestation au cours de la parole. Cependant, cette méthode nécessite un haut niveau de technicité pour la manipulation du signal vocal. Décrise dans notre précédente étude, à partir de l'intervention d'un seul manipulateur, la méthode devait être reprise par plusieurs auditeurs afin de confirmer sa validité expérimentale. Les résultats que nous avons obtenus montrent que la procédure est hautement reproductible en intra- comme en inter-individuel. L'analyse des variations de la fiabilité du jury au cours de la phrase confirme l'influence du contexte phonétique sur la perception de la dysphonie. Notamment, les phonèmes présentant une attaque semblent particulièrement sensibles à l'apparition des occurrences pathologiques et leur caractérisation est largement reproductible. Une étude spécifique de l'attaque au cours de la parole devrait confirmer l'importance des informations contenues dans cette portion du signal pour la perception de la dysphonie.

I. Influence de l'attaque dans la perception de la dysphonie au cours de la parole

Revis J, Ollivier R, Giovanni A.

(Folia Phoniatri Logop, soumis à publication)

Introduction : Ces dix dernières années, plusieurs auteurs ont porté leur attention sur les phénomènes contemporains de la mise en vibration des cordes vocales [Decoster, 1997 ; Cooke, 1997 ; Köster, 1999], et leur influence sur la perception des dysphonies [De Krom, 1994 ; Revis, 1999, 2000, 2002]. Pour des raisons techniques, ces travaux portaient sur l'analyse de l'attaque au cours de la production d'une voyelle tenue. Cependant, la voyelle tenue présente plusieurs inconvénients pour la mesure de l'attaque : tout d'abord, elle ne contient qu'une seule attaque à l'initiation du son alors que la parole est constituée d'une succession d'attaques [Laver, 1980 ; Landercy, 1977] ; ensuite, la voyelle tenue est influencée par la « culture vocale » du locuteur qui peut, lorsqu'il est entraîné, contrôler la mise en vibration de ses cordes vocales [Revis, 2000]. Il semble donc que la voyelle tenue ne soit pas adaptée à l'étude de l'attaque.

La conception et la validation de la méthode de phonetic labelling présentées dans les travaux précédents a ouvert une porte sur les possibilités d'analyse perceptive au cours de la parole. L'écoute spécifique des phonèmes constitutifs de la phrase permet une analyse de la manifestation dysphonique au cours de la parole et permet donc d'observer l'évolution des occurrences pathologiques. L'objectif de ce travail est de vérifier l'influence de l'attaque au cours de la parole grâce à la méthode de phonetic labelling. Pour cela, nous avons proposé au jury d'écouter l'analyse de deux phrases prononcées par les mêmes locuteurs : une phrase « stable », constituée exclusivement de phonèmes voisés et présentant donc une seule attaque sur le phonème initial ; et une phrase « attaques » constituée d'un enchaînement de syllabes commençant par une occlusive non-voisée et présentant 7 attaques. Notre hypothèse était double : d'une part, la phrase « attaques » serait perçue comme plus dysphonique que la phrase « stable » ; d'autre part, occurrences pathologiques seraient plus nombreuses dans la phrase « attaque » et localisées spécifiquement sur les phonèmes présentant une attaque.

Protocole expérimental : Nous avons recruté et enregistré 30 sujets, tous féminins, selon les critères et la méthodologie habituels. Parmi cette population, nous avions 15 sujets témoins ne

présentant aucun antécédent de dysphonie (âgés de 22 à 40 ans) et 15 patientes dysphoniques (âgées de 26 à 73 ans) dont les pathologies sont présentées dans le tableau ci-dessous :

Nodules	n = 5
Œdème de Reinke	n = 4
Polype	n = 3
Granulôme	n = 1
Paralysie unilatérale	n = 1
Fuite glottique	n = 1

Tableau VI.I.1. : répartition des sujets dysphoniques en fonction de leur pathologie.

Lors de l'enregistrement, les sujets devaient prononcer deux phrases choisies en fonction du nombre d'attaques induit par l'enchaînement des phonèmes :

- la première phrase, phrase « stable », était constituée uniquement de phonèmes voisés (tableau VI.I.2.).

/ a m e l i a m i l o o m u l ē /

“Amélie a mis l'eau au moulin.”

/a/	Voyelle ouverte postérieure	Voisée
/m/	Occlusive nasale bi-labiale	Voisée
/e/	Voyelle semi-fermée antérieure	Voisée
/l/	Liquide latérale alvéolaire	Voisée
/i/	Voyelle fermée antérieure	Voisée
/a/	Voyelle ouverte postérieure	Voisée
/m/	Occlusive nasale bi-labiale	Voisée
/i/	Voyelle fermée antérieure	Voisée
/l/	Liquide latérale alvéolaire	Voisée
/o/	Voyelle semi-fermée postérieure	Voisée
/ɔ/	Voyelle semi-fermée postérieure	Voisée
/m/	Occlusive nasale bilabiale	Voisée
/u/	Voyelle fermée postérieure	Voisée
/l/	Liquide latérale alvéolaire	Voisée
/ɛ/	Voyelle semi-ouverte antérieure nasale	Voisée

Tableau VI.I.2. : description phonétique des phonèmes constitutifs de la phrase « stable » : « Amélie a mis l'eau au moulin. »

Ainsi, cette phrase ne présentait qu'une seule attaque sur le premier phonème et la vibration était entretenue sans interruption jusqu'à la fin de la production (figure VI.I.1.).

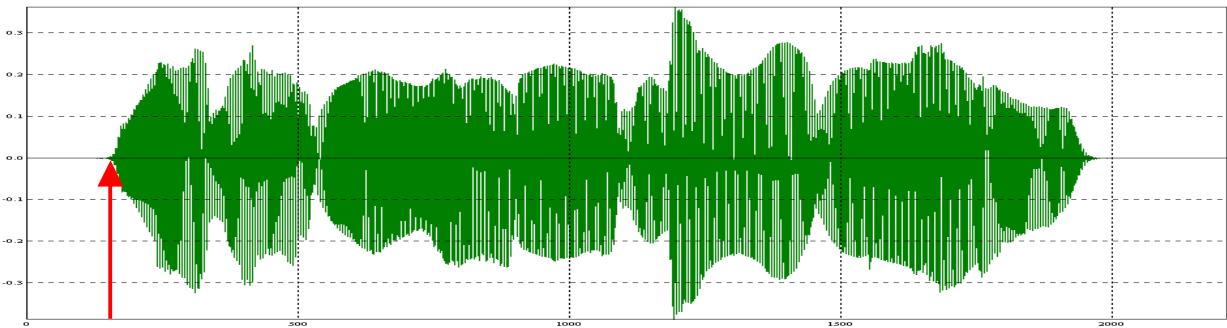


Figure VI.I.1. : signal acoustique de la phrase « Amélie a mis l'eau au moulin. » pendant laquelle on peut noter que la vibration ne s'arrête pas : le signal est continu, il n'est pas interrompu. Il existe une seule attaque sur le phonème initial (flèche rouge) comme on pourrait l'observer sur une voyelle tenue.

- La deuxième phrase, phrase « attaques », était conçue pour présenter plusieurs attaques. Elle était constituée de l'enchaînement récurrent occlusive non-voisée / voyelle (tableau VI.I.3.). Durant chaque occlusive non-voisée, la vibration était interrompue, la voyelle suivante présentait donc une attaque.

/t ð t e t a t i l o t e t a t u /
 « Ton thé t'a-t-il ôté ta toux. »

/t/	Occlusive apico-dentale	Non-voisée
/ð/	Voyelle semi-ouverte postérieure nasale	Voisée
/t/	Occlusive apico-dentale	Non-voisée
/e/	Voyelle semi-fermée antérieure	Voisée
/t/	Occlusive apico-dentale	Non-voisée
/a/	Voyelle ouverte postérieure	Voisée
/t/	Occlusive apico-dentale	Non-voisée
/i/	Voyelle fermée antérieure	Voisée
/l/	Liquide latérale alvéolaire	Voisée
/o/	Voyelle semi-fermée postérieure	Voisée
/t/	Occlusive apico-dentale	Non-voisée
/e/	Voyelle semi-fermée antérieure	Voisée
/t/	Occlusive apico-dentale	Non-voisée
/a/	Voyelle ouverte postérieure	Voisée
/t/	Occlusive apico-dentale	Non-voisée
/u/	Voyelle fermée postérieure	Voisée

Tableau VI.I.3. : description phonétique des phonèmes constitutifs de la phrase « attaques » : « Ton thé t'a-t-il ôté ta toux. »

Ainsi, cette phrase présentait une série de 7 attaques, sur chaque voyelle précédée de l'occlusive non-voisée /t/ dont la production induit une interruption de la vibration (figure VI.I.2.).

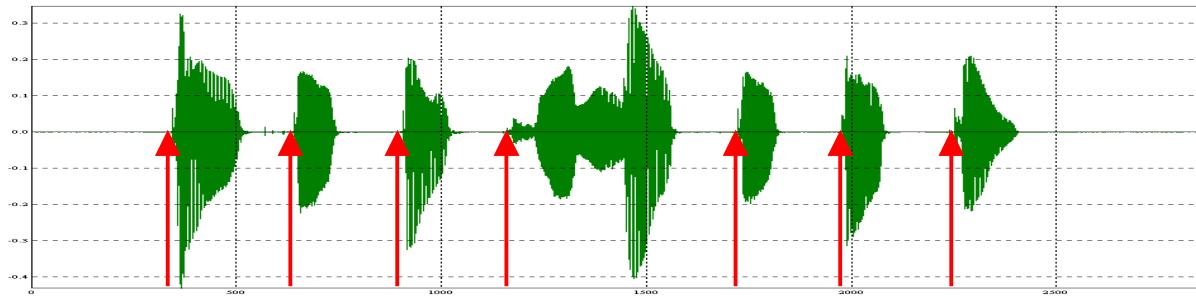


Figure VI.I.2. : signal acoustique de la phrase « Ton thé t'a-t-il ôté ta toux. » pendant laquelle on peut noter que la vibration est interrompue par la production de l'occlusive non-voisée /t/. On dénombre donc 7 attaques (flèches rouges) puisque la vibration est initiée à nouveau sur la voyelle suivante.

L'analyse perceptive des enregistrements s'est déroulée selon deux protocoles. Le premier protocole a nécessité le recrutement d'un jury d'écoute (un médecin phoniatre et deux orthophonistes). Les auditeurs ont effectué une analyse perceptive classique en aveugle selon la méthode du consensus. Leur tâche était d'évaluer les paramètres G (grade global), R (raucité) et B (souffle), sur une échelle ordinaire à 4 niveaux (0 : voix normale ; 1 : dysphonie légère ; 2 : dysphonie moyenne ; 3 : dysphonie sévère). Deux sessions d'évaluation distantes d'une semaine ont été organisées : lors de la première session, le jury évaluait la phrase « stable » et lors de la seconde, la phrase « attaques ». Lors des deux sessions, les phrases étaient présentées dans un ordre aléatoire différent. En effet, la condition en aveugle supposait que les auditeurs ne reconnaissent pas que les mêmes sujet prononçaient les deux corpus de phrase. Les deux sessions avaient été présentées au jury comme deux études distinctes et sans rapport.

Le second protocole consistait en l'évaluation du corpus par la méthode du phonetic labelling, selon les procédures décrites dans les articles précédents. L'analyse était effectuée en aveugle par une orthophoniste. En raison de la longueur des analyses par phonetic labelling, l'évaluation a été effectuée en plusieurs sessions : deux demi-journées avec des pauses de 15 minutes toutes les 90 minutes pour le corpus de phrases « stable », puis de nouveau deux demi-journées une semaine plus tard pour le corpus « attaques » selon les mêmes modalités. Les résultats étaient recueillis à la main sur la fiche de notation, puis transférés par l'examinateur sur un tableau. Les mêmes précautions que pour le jury d'écoute avaient été mises en place afin de s'assurer du respect de la condition en aveugle.

Le but de l'analyse perceptive classique était de vérifier que la phrase « attaques » tendait à augmenter l'impression auditive de dysphonie (hypothèse 1), et le phonetic labelling permettait d'observer à la fois le nombre d'occurrences pathologiques sur les deux corpus de phrase et leur localisation (hypothèse 2). L'analyse des résultats est présentée de façon descriptive sous forme de tableaux.

Résultats : Le tableau VI.I.4. montre les résultats de l'évaluation perceptive classique sur le corpus de phrases « stable » et le corpus de phrases « attaques ».

Voix	Phrase "stable"			Phrase "attaques"			
	G	R	B	G	R	B	
voix 01	0	0	0	0	0	0	
voix 02	0	0	0	0	0	0	
voix 03	0	0	0	0	0	0	
voix 04	0	0	0	0	0	0	
voix 05	0	0	0	0	0	0	
voix 06	0	0	0	0	0	0	
voix 07	0	0	0	0	0	0	
voix 08	0	0	0	0	0	0	
voix 09	0	0	0	0	0	1	
voix 10	0	0	0	0	0	1	
voix 11	0	0	0	0	1	1	
voix 12	0	0	0	1	1	0	
voix 13	1	0	1	0	0	0	
voix 14	1	1	1	1	0	1	
voix 15	1	1	1	1	0	1	
voix 16	1	0	1	1	0	1	
voix 17	1	1	0	1	1	0	
voix 18	1	0	1	1	0	1	
voix 19	1	0	1	1	0	1	
voix 20	1	1	0	1	1	1	
voix 21	1	1	0	1	1	1	
voix 22	1	0	1	2	1	1	
voix 23	1	1	1	2	1	2	
voix 24	2	1	2	1	0	1	
voix 25	2	2	2	2	2	2	
voix 26	2	3	2	3	3	2	
voix 27	2	2	2	3	3	3	
voix 28	3	3	3	3	3	3	
voix 29	3	2	2	3	2	3	
voix 30	3	2	3	3	3	3	

Résultats identiques entre les 2 phrases
Sous-évaluation sur la phrase "attaques" (-1)
Sur-évaluation sur la phrase "attaques" (+1)

Tableau VI.I.4. : résultats obtenus par analyse perceptive classique sur les corpus de phrases « stable » et de phrases « attaques ». Les trait horizontaux en gras marquent les limites des classes de grade de dysphonie. Les cases colorées correspondent à la légende ci-contre : en blanc, résultats identiques ; en bleu, sous-évaluation sur la phrase « attaques » ; en rose, sur-évaluation sur la phrase « attaques ».

Dans ce tableau, on constate qu'environ 50% des voix (14/30) sont évaluées de la même façon par l'analyse perceptive classique quel que soit la phrase prononcée. Les cas qui présentent une différence montrent une sur-évaluation de la phrase « attaques » dans 75% des cas (12/16), et une sous-évaluation de la phrase « attaques » dans 15% des cas (4/16).

Les tableaux suivants concernent les résultats de l'analyse par phonetic labelling. Le tableau VII.5. montre le pourcentage de phonèmes atteints par des occurrences pathologiques par rapport au nombre total de phonèmes voisés, sur chacun des deux corpus de phrases.

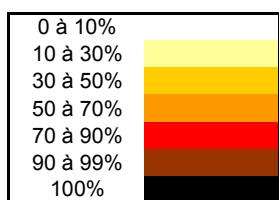
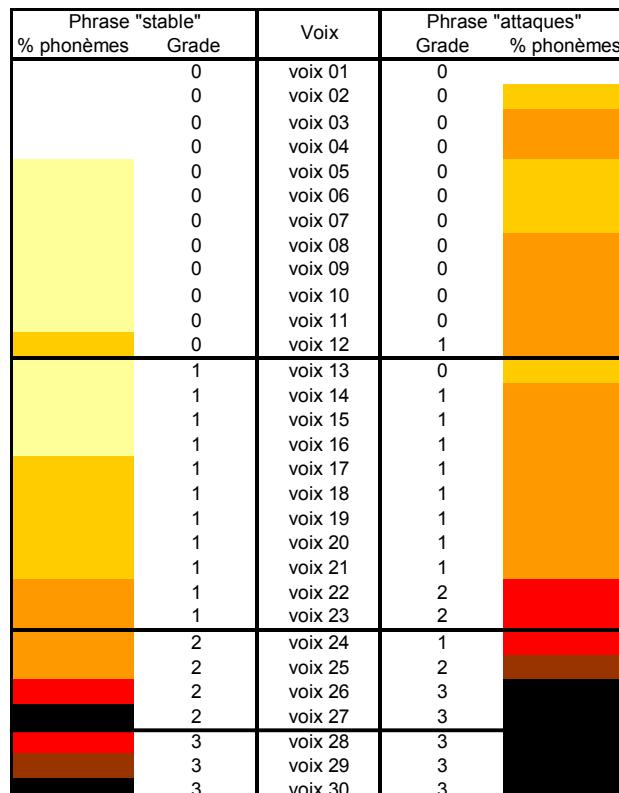


Tableau VII.5. : pourcentage de phonèmes atteints par une (ou plusieurs) occurrence(s) pathologique(s) par rapport au nombre total de phonèmes voisés. Pour faciliter la lecture du tableau, un code couleur a été mis en place selon la légende ci-dessus.



Le tableau VI.1.5. montre que le pourcentage de phonèmes atteints est systématiquement plus élevé sur la phrase « attaques » que sur la phrase « stable », sauf pour la voix 1 (voix normale) qui ne change pas de classe (0 à 10%). D'autre part, on constate que les voix qui ont été sur-évaluées sur la phrase « attaques » par l'analyse perceptive classique, c'est à dire celles qui sont passées au grade perceptif supérieur (voix 12, 22, 23 26 et 27), présentaient déjà un nombre d'occurrences pathologiques supérieur aux autres voix du même groupe sur la phrase « stable ».

Les tableaux VI.1.6. et VI.1.7. montrent la répartition phonémique des occurrences pathologiques sur les deux corpus de phrases. Afin de faciliter la comparaison entre les deux tableaux, nous ne présentons que les phonèmes voisés de la phrase « attaque » (les phonèmes non-voisés étant toujours exempts d'occurrences de dysphonie).

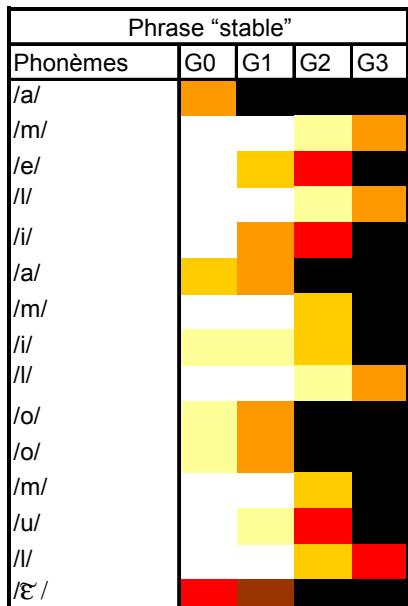


Tableau VI.1.6. : fréquence d'apparition des occurrences pathologiques sur chaque phonème de la phrase « stable », en fonction des grades.

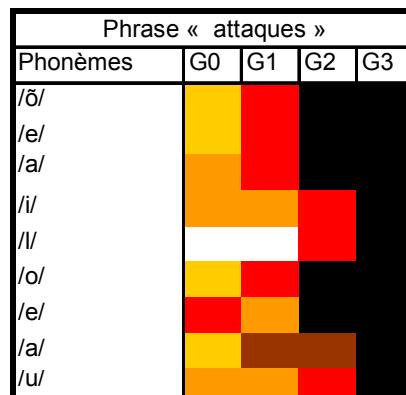
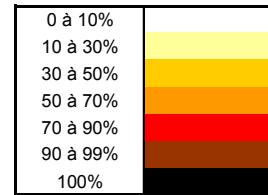


Tableau VI.1.7. : fréquence d'apparition des occurrences pathologiques sur chaque phonème de la phrase « attaques », en fonction des grades.

Sur la phrase « stable » (tableau VI.1.6.), on remarque que le phonème initial (/a/ de « Amélie ») est atteint d'occurrences pathologiques quel que soit le grade, et de façon systématique (dans 100% des cas) dès l'apparition de la dysphonie (grade 1). De même, le phonème final /ɛ/ semble fortement sensible aux occurrences pathologiques, puisqu'il est atteint dans 71% des voix normales, 88% des dysphonies légères et 100% des dysphonies moyennes et sévères. Les autres phonèmes sont atteints de manière progressive au fur et à mesure de l'augmentation de la dysphonie. Dans les grades 3 (dysphonies sévères), 73,3% des phonèmes sont atteints dans 100% des cas.

En ce qui concerne la phrase « attaques » (tableau VI.I.7.), on constate que les phonèmes atteints sont plus nombreux (tous sauf le /l/ pour les grades 0 et 1) et dans des proportions plus importantes. Dès le grade 0 (voix normales), 88% des phonèmes voisés sont atteints d'occurrences pathologiques dans au moins 30% des cas. Dans les grades 1 la fréquence d'apparition augmente largement (environ 70% en moyenne). Dans les grades 2, 100% des phonèmes sont atteints par les occurrences pathologiques dans des proportions importantes. Enfin, dans les grades 3, 100% des phonèmes voisés sont atteints dans 100% des cas.

Discussion : L'objectif de cette étude était d'évaluer l'influence de l'attaque au cours de la parole. Nos hypothèses étaient que 1- la phrase « attaques » est évaluée de façon plus sévère par l'analyse perceptive classique, 2- la phrase « attaques » contient plus d'occurrences pathologiques, et 3- les occurrences pathologiques sont localisées sur les phonèmes présentant une attaque.

Nos résultats montrent qu'environ 50% des sujets obtiennent la même évaluation des paramètres GRB sur les deux phrases (tableau VI.I.4.). Les voix normales (grade 0) et les dysphonies sévères (grade 3) semblent plus stables que les dysphonies légères et moyennes. En effet, lorsque la voix est normale, elle est peu influencée par la présence ou non d'attaques, une voix normale reste normale. De même, une dysphonie sévère, déjà très altérée, est assez peu aggravée par la présence d'attaques [Revis, 2002]. Notons que les grades 0 ne peuvent pas être sous-évalués, de la même façon que les grades 3 ne peuvent pas être sous-évalués. Dans les cas de différence, pour 75% des sujets, il existe une sur-évaluation de la phrase « attaques ». Les différences de classement sont toujours d'un seul grade : une même voix ne passe jamais d'un grade 1 à un grade 3 ; en revanche, elle peut passer d'un grade 1 à un grade 2. Lorsqu'on prend en compte les paramètres GRB, sur l'ensemble des 30 voix, on comptabilise 18 sur-évaluations sur la phrase « attaques » contre 7 sous-évaluations. La répartition en fonction du paramètre est la suivante : pour le grade global, 5 voix sont jugées plus sévères contre 2 moins sévères. Parmi ces deux voix jugées moins sévères, une était évaluée en grade 1 sur la phrase « stable » puis grade 0 sur la phrase « attaques ». Notons que nous avions enregistré 15 sujets témoins, mais que 3 d'entre eux ont été gradés dysphonie légère par le jury d'écoute. Le groupe témoin était constitué de sujets sains, néanmoins, 3 d'entre eux présentaient une légère déviance. Le sujet grade 1 sur la phrase « stable », déclassé en grade 0 sur la phrase « attaques » correspondait à un cas « limite ». En ce qui concerne le paramètre de raucité, 5 voix sont jugées plus sévères contre 3 moins sévères. Enfin, pour le souffle, 8 voix sont jugées plus sévères contre 2 moins sévères. La différence de jugement est donc plus flagrante sur le souffle que sur les autres paramètres. Il est possible que le contexte phonétique d'occlusives non-voisées sur la phrase « attaques » favorise l'émergence du souffle. Une étude ultérieure s'attachera à définir l'influence du contexte phonétique sur l'apparition des paramètres qualitatifs de dysphonie.

Cette première partie de l'étude concerne l'analyse perceptive classique par un jury d'écoute. On peut se demander si les différences d'évaluation observées sur les deux phrases ne sont pas simplement le résultats de la variabilité inhérente aux jugements perceptifs. Dans le but de nous affranchir des problèmes liés à la variabilité du jury, nous avions choisi une évaluation par consensus. Le principe est de réunir un nombre impair d'experts et que l'évaluation de chaque voix soit discutée jusqu'à commun accord. Dans la littérature, le jugement par consensus est considéré comme une expertise et n'est pas remis en cause. Néanmoins, en admettant qu'il puisse exister quand même une variabilité des jugements, les résultats que nous obtenons ne vont pas dans ce sens. En effet, dans 75% des cas, il existe une sur-évaluation. Si cette variation était due au hasard, nous aurions obtenu 50% de sur-évaluation et 50% de sous-évaluation. Il semble donc que l'effet de sur-évaluation que nous observons est bien dû à la présence d'attaques dans la phrase, et pas à la variabilité du jury.

L'analyse par phonetic labelling a ensuite été appliquée aux deux phrases afin de localiser et de qualifier les occurrences pathologiques. En ce qui concerne la phrase « stable », le découpage des phonèmes s'est avéré complexe du fait de la continuité du signal acoustique, puisque le voisement était constant. En effet, il était difficile de déterminer la fin d'un phonème et le début du suivant. En revanche, pour la phrase « attaques », l'alternance de phonèmes voisés / non-voisés a largement facilité l'opération. Le passage d'un phonème à l'autre était net, presque pré-découpé.

L'analyse par phonetic labelling a permis de révéler un biais dû à l'enregistrement. En effet, malgré l'insistance de l'examinateur qui donnait les consignes de lecture, 7 sujets ont eu des difficultés à prononcer la phrase « stable » d'une seule traite. Ils ont introduit des pauses phonatoires en découpant la phrase de la façon suivante : « Amélie / a mis l'eau / au moulin. » En introduisant des pauses là où il n'y en a théoriquement pas, de nouvelles attaques ont été créées : « **A**mélie / **a** mis l'**eau** / **au** moulin ». L'analyse par phonetic labelling a révélé que ces nouvelles attaques ont induit systématiquement des occurrences pathologiques. Cela a eu pour effet d'une part d'introduire « artificiellement » des occurrences qui ne seraient peut-être pas apparues sans cela, donc d'augmenter la proportion d'occurrences de la phrase « stable » ; et d'autre part, cela nous a permis de confirmer l'effet « pathogénique » de l'attaque.

Quel que soit le grade de dysphonie, on retrouve plus d'occurrences pathologiques sur la phrase « attaques » que sur la phrase « stable » (tableau VI.1.5.). 27 voix sur 30 montrent un pourcentage de phonèmes atteint plus élevé sur la phrase « attaques ». Pour un patient dysphonique, la difficulté de production réside dans l'initiation de la vibration des cordes vocales. Dans la phrase stable, il n'y a pas d'arrêt vibratoire. Une fois l'attaque produite, la vibration atteint un régime stable, elle est auto-entretenue. Le sujet dysphonique n'a donc pas de difficultés à émettre la suite du signal vocal. En revanche, dans la phrase « attaques », les cordes vocales ont à peine le temps d'entrer en vibration sur

l'émission de la voyelle, qu'elle sont immédiatement stoppées par l'occlusive non-voisée. Dans ce cas, la production correspond à un enchaînement de transitoires, il n'y a pas de partie stable [Revis, 2002]. Pour les sujets dysphoniques, la production de ces transitoires est perturbée et participe à l'impression auditive de dysphonie.

Les tableaux VI.I.6. et VI.I.7. donnent la fréquence (en pourcentage) des occurrences pathologiques, tous paramètres confondus, pour chaque phonème voisé en fonction du grade. Dans la phrase « stable » (tableau VI.I.6.), le phonème /a/ de « Amélie » est très atteint pour l'ensemble des voix : 60% d'occurrences pathologiques pour les voix normales, et 100% d'occurrences pathologiques pour les sujets dysphoniques. Cette observation va dans le sens de notre hypothèse puisque ce phonème, en position initiale, contient une attaque. Deux autres phonèmes sont également fréquemment atteints : le /a/ de « a mis l'eau » (40% d'occurrences pathologiques pour les G0, 60% pour les G1, et 100% pour les G2 et 3), et le /o/ de « au moulin » (20% d'occurrences pathologiques pour les G0, 60% pour les G1 et 100% pour les G2 et 3). Notons que le /o/ de « a mis l'eau » montre strictement les mêmes résultats, mais il faut considérer qu'il était impossible de déterminer la fin de l'un et le début de l'autre lors de l'analyse par phonetic labelling. Les deux phonèmes /o/ sont donc assimilés. Les phonèmes /a/ de « a mis l'eau » et /o/ de « au moulin » ne subissent pas l'influence de l'attaque, cependant, ils se situent aux points d'accentuation de la phrase qui sont des points clé de la production vocale. Il semble que la prosodie joue un rôle dans l'apparition des occurrences pathologique. Cette hypothèse fera l'objet de notre prochain travail. D'autre part, dans 5 cas, ces points clé correspondent aux emplacements des découpages inopportunus décrits précédemment : ils deviennent alors phonème initial du groupe vocalique et présentent une attaque.

En ce qui concerne la phrase « attaques » (tableau VI.I.7.), tous les phonèmes voisés montrent un taux important d'occurrences pathologiques à l'exception du /l/ de « il » qui reste normal dans les grades 0 et 1. Tous les phonèmes voisés de cette phrase présentent une attaque à l'exception de /l/ et /o/ (« Ton thé t'a-t-il ôté ta toux. »). Le fait que le /l/ soit relativement préservé va dans le sens de notre hypothèse. En revanche, le fort taux de fréquence du phonème /o/, qui ne contient pas d'attaque, nous conduit à poser la question de la finale. Le signal vocal est composé de deux transitoire : l'attaque qui correspond à la mise en vibration des cordes vocales, et la finale qui correspond à l'amortissement de la vibration jusqu'à son interruption. Si l'attaque a été largement étudiée ces dernières années [De Krom, 1994, 1995 ; Revis, 1999, 2000, 2002], la finale n'a pas fait l'objet de publications spécifiques. Pourtant, la finale qui est également caractérisée par une importante instabilité, peut elle aussi exercer une influence sur la perception de la dysphonie. Dans la phrase « stable », on retrouve une seule **attaque** sur le phonème initial, et une seule **finale** sur le dernier phonème :

/ **a** m e l i a m i l o o m u 1 **ɛ** /

Dans la phrase « attaques », tous les phonèmes voisés qui suivent une occlusive non-voisée contiennent une **attaque** :

/ t ð t e t a t i l o t e t a u /

De même, tous les phonèmes voisés qui précèdent une occlusive non-voisée, contiennent une **finale** :

/ t ð t e t a t i l o t e t a u /

Dans la phrase « attaques », le seul phonème qui ne subit ni l'influence de l'attaque, ni l'influence de la finale est le /l/ de « il ». On constate que ce phonème est le plus préservé dans la phrase. Au total, dans les deux corpus, seuls 2 phonèmes subissent l'influence exclusive de la finale : le /o/ de « Ton thé t'a-t-il ôté ta toux. » (ce qui explique le fort taux d'occurrences pathologiques retrouvé sur ce phonème), et le /ɛ/ de « Amélie a mis l'eau au moulin. », qui est le dernier phonème de la phrase et précède la reprise inspiratoire. Lors de la production de ce dernier phonème, il semble y avoir une sorte de relâchement. D'après l'auditeur de phonetic labelling, la fin de ce phonème (finale) apparaît comme soupirée, comme si le locuteur « démissionnait », ce qui induit une occurrence de souffle. Ainsi, nos résultats montrent que la finale, au même titre que l'attaque, semble favoriser l'apparition d'occurrences pathologiques : 60% pour les grades 0, 90% pour les grades 1 et 100% pour les grades 2 et 3. Cette observation ouvre une piste qu'il serait intéressant d'approfondir par l'étude spécifique de l'influence de la finale sur la perception de la dysphonie.

Conclusion : L'objectif de ce travail était d'évaluer l'influence de l'attaque sur la perception de la dysphonie au cours de la parole. Les résultats ont montré que la phrase composée de nombreuses attaques était fréquemment sur-évaluée par rapport à la phrase composée uniquement de phonèmes voisés par l'analyse perceptive classique. Par ailleurs, l'analyse par phonetic labelling a montré que la phrase « attaques » contenait davantage d'occurrences pathologiques que la phrase « stable ». Outre l'influence de l'attaque, il semble que la finale favorise également l'apparition d'occurrences pathologiques et particulièrement de souffle. Enfin, le contexte phonétique d'occlusives non-voisées tend à induire préférentiellement des occurrences de souffle, et le contexte linguistique d'accentuation semble également être une source d'influence de la manifestation dysphonique. Notre prochain travail s'attachera à décrire et à comprendre les modalités des influences phonétiques et linguistiques dans l'apparition des occurrences pathologiques de dysphonie.

Les trois travaux que nous venons de présenter montrent la validité et l'intérêt de l'analyse de la dysphonie par phonetic labelling. Cette méthode d'étiquetage des phonèmes permet une cartographie précise de la dysphonie. L'observation de l'influence des transitoires du signal vocal , mais aussi des contextes phonétiques et linguistiques suggère qu'il est possible, en fonction du matériau phonétique choisi, d'agir sur la production des patients dysphoniques. Notre réflexion s'oriente donc désormais vers la conception et l'élaboration d'un corpus de phrases qui permettrait, par l'introduction intentionnelle et maîtrisée de difficultés contextuelles, de mettre les épisodes dysphoniques «en valeur». Notre idée est de pouvoir prédire la localisation et la qualification d'une occurrence pathologique dans une phrase. En incitant l'auditeur expérimental ou le clinicien à concentrer son attention sur un seul phonème dans la phrase, correspondant à une occurrence prédictible, ce nouveau matériau phonétique pourrait permettre de faciliter et d'accélérer l'analyse par phonetic labelling et donc d'en « démocratiser » l'application.



Discussion générale et perspectives

Discussion générale et perspectives

« *La perception de la voix correspond à l'accumulation de sensations abstraites qui impliquent des caractéristiques laryngées et supra-laryngées. Les voix peuvent être décrites par des paramètres spécifiques que l'on peut distinguer à l'oreille.* »

Laver, 1980

Au total, les travaux que nous avons présentés vont dans le sens d'une amélioration effective de la fiabilité des évaluations perceptives de la dysphonie par un jury d'écoute en recherche ou par le praticien en pratique clinique.

Les travaux présentés dans l'axe **perception/jury** ont permis de révéler la sensibilité des auditeurs à un protocole de formation à l'analyse perceptive. L'entraînement d'un groupe d'auditeurs candides à l'écoute de voix pathologiques selon un paradigme « stimulus/réponse/feedback/stimulus » permet à la fois d'améliorer la pertinence et l'homogénéité des évaluations. Proposer une formation à un groupe d'auditeurs permet donc de s'assurer de la cohérence du jury et de son adéquation commune à la méthode. Il semble que ce type de procédure puisse être proposé, y compris avec les auditeurs expérimentés, afin de s'assurer d'une compréhension uniforme des termes qualitatifs descriptifs de la dysphonie et de l'utilisation de l'échelle quantitative mise en place pour l'étude. Par ailleurs, on peut envisager l'entraînement à l'analyse perceptive comme une partie intégrante de la formation clinique proposée aux professionnels de la voix et notamment aux orthophonistes. Actuellement, les élèves orthophonistes de l'école de Marseille ont la possibilité de participer à un enseignement spécifique de la perception de la dysphonie sous forme de TD en 4^{ème} année. Ces travaux ont également permis la conception d'un outil informatique (le logiciel CLAP®) permettant de fournir au clinicien une aide méthodologique structurelle dans l'évaluation de la voix de leurs patients. A l'heure actuelle, ce logiciel a été diffusé à une cinquantaine d'orthophonistes dont nous attendons les commentaires afin d'en améliorer les fonctionnalités.

Les travaux présentés dans l'axe **perception/mesures** ont permis une amélioration notable des performances des protocoles instrumentaux. En tant que référence de validation des mesures

instrumentales, l'analyse perceptive doit proposer des résultats à la fois fiables et précis. La conception de l'échelle visuelle analogique modifiée constitue notamment un réel progrès dans la compréhension du comportement de l'auditeur face à la tâche d'évaluation de la dysphonie. L'échelle *mVAS* a permis d'améliorer la corrélation entre les deux systèmes de mesure.

Enfin, les travaux présentés dans l'axe *perception/phonétique* ont permis d'aborder la perception de la dysphonie selon un axe original et novateur. En effet, la perception de la dysphonie était considérée jusqu'alors comme l'impression globale de perturbation contenue dans un signal de durée variable. Amorcée par l'étude du comportement des auditeurs sur une portion spécifique du signal (l'attaque), la recherche s'est orientée vers une approche phonétique de la perception de la dysphonie, entrant directement au cœur de la parole. La conception de la méthode de « Phonetic Labelling », a permis une approche radicalement différente de la perception de la dysphonie, franchissant le pas entre l'*évaluation* perceptive classique et une véritable *analyse* perceptive. En effet, le principe d'écoute et de qualification de chaque phonème au cours de la parole nous a permis de réaliser une véritable cartographie de la dysphonie et de mieux comprendre les influences sous-tendant l'apparition d'occurrences pathologiques. L'observation de l'influence des transitoires du signal vocal, mais aussi des contextes phonétiques et linguistiques suggère qu'il est possible, en fonction du matériau phonétique choisi, d'agir sur la production des patients dysphoniques. Il semble inévitable que les travaux s'orientent vers l'étude étroitement conjointe de la perception de la dysphonie et de la production de parole par le sujet dysphonique. Dès lors, on ne peut plus faire l'économie de l'exploration de l'influence des contraintes articulatoires et prosodiques imposées par le matériau phonétique choisi.

Plusieurs études montrent que la production d'une voyelle est affectée par son environnement consonantique et prosodique [Golb, 1993 ; De Jong, 1993 ; Hardcastle, 1999 ; Fujimoto, 2002 ; Revis, 2004]. Des contextes phonétiques et linguistiques semblent favoriser l'apparition d'occurrences pathologiques sur certains phonèmes. Au cours de la production de la parole, les phonèmes interagissent entre eux et l'environnement modifie leur articulation. Cet effet de contexte peut rendre précaire la vibration des cordes vocales, en particulier chez le sujet dysphonique qui éprouve déjà des difficultés à émettre sa voix. Notre hypothèse¹ actuelle est que l'utilisation d'un matériau phonétique spécifique, conçu d'après l'étude des interactions phonétiques et linguistiques, constituerait une sorte d'épreuve d'« effort » qui permettrait de révéler les épisodes dysphoniques. En effet, la dysphonie devrait se manifester de manière plus évidente sur des voyelles mises en contexte. Cette méthodologie pourrait permettre d'accroître la fiabilité du jury d'écoute, de mieux évaluer les caractéristiques intrinsèques de la dysphonie et de réduire la durée des sessions d'écoute. Notre objectif est de

¹ Revis J, Gachet F, Giovanni A. Influence des contextes phonétiques et linguistiques sur l'évaluation perceptive de la dysphonie. En cours.

sélectionner des contextes phonétiques spécifiques favorisant l'apparition des paramètres habituellement étudiés dans l'analyse par phonetic labelling : la raucité, le souffle, et le dévoisement.

La production d'un phonème nécessite la réalisation simultanée de plusieurs gestes phonatoires et articulatoires. La production de la parole est réalisée par l'intervention des muscles vocaux, mais aussi des configurations sus-glottiques (résonateurs). Le type de phonème produit (occlusives, constrictives, nasales, liquides) fait varier la pression de l'air dans la cavité pharyngo-bucale. La vibration des cordes vocales est affectée par ces variations de pression, et la qualité de la voix subit cette influence [De Jong, 1993]. Au cours de la parole, les phonèmes sont coarticulés et interagissent entre eux [Ohala, 1993]. Lors de la coarticulation, les gestes articulatoires et phonatoires s'enchaînent, ils peuvent se chevaucher et se mélanger. Apparaissent alors des phénomènes d'assimilation progressive ou régressive [Hardcastle, 1993]. L'assimilation progressive est l'influence de l'articulation d'un phonème sur le(s) suivant(s). L'assimilation régressive est l'influence de l'articulation d'un phonème sur le(s) précédent(s). Par exemple, lorsque l'on prononce la syllable /si/, la position articulatoire du /s/ est plus antérieure que pour la syllabe /su/. En effet, lors de la production du /s/ de /si/, les lèvres sont déjà étirées pour la production du /i/. Pour la production du /s/ de /su/, au contraire, les lèvres sont en protrusion. La consonne est donc produite avec un effet d'anticipation de la voyelle suivante. C'est une assimilation régressive de la voyelle, puisque le /i/ ou le /u/ influencent la production de la consonne précédente /s/. Ainsi, les caractéristiques propres des phonèmes (en l'occurrence, le /s/ de /si/ ou /su/) sont modifiées par leur environnement phonétique [Beckham, 1992].

Lors de la production d'un phonème voisé, les cordes vocales vibrent. À l'inverse, lors de la production d'un phonème non-voisé, le larynx est en position d'ouverture. La dysphonie, en tant qu'altération de la vibration des cordes vocales, se manifeste donc sur les phonèmes voisés. L'élaboration d'environnements phonétiques nous oriente ainsi plus particulièrement vers l'étude des caractéristiques de la voyelle. En effet, les auteurs s'accordent à dire que « *la voyelle porte la voix* » [Laver, 1980 ; Ohala, 1993]. C'est d'ailleurs l'origine étymologique de son nom. D'une manière simplifiée, lorsqu'une voyelle est mise en contexte entre deux consonnes, le groupe de phonèmes obtenu est noté CVC (Consonne / Voyelle / Consonne). La voyelle est le noyau du groupe phonétique. La conception de contextes phonétiques nécessite la création de groupes phonétiques de ce type. Toute la complexité du problème réside dans la compréhension des interactions entre les phonèmes. L'étude de ces interactions, d'après les données la littérature, rend possible la conception des combinaisons de consonnes et de voyelles favorisant l'émergence des critères de raucité, de souffle et de dévoisement.

Contexte favorisant la raucité : le critère de raucité est défini comme la « *perception d'une irrégularité, d'une apéridodicité de la vibration des cordes vocales lors de la phonation* » [Hirano, 1981].

- Choix de la voyelle : selon Rees, plus la voyelle est ouverte, plus son caractère rauque sera émergent [Rees in Laver, 1980]. L'ouverture des voyelles est définie par la position de la langue ; plus la langue est basse, plus la voyelle est ouverte. Les trois voyelles les plus ouvertes du français sont : le /ɛ/, le /a/ et le /ɔ/ (/o/ ouvert). Nous avons éliminé le /ɛ/ car il est fortement influencé par les accent régionaux (cette étude est menée à Marseille), et prononcé /e/ qui est une voyelle plus fermée. En 2002, Lorch a réalisé une étude des caractéristiques perceptives de la dysphonie spasmodique . Elle a procédé à une analyse perceptive de chaque phonème d'un échantillon de parole de 100 syllabes chez 6 patients. Le phonème /ɔ/ était perçu comme la voyelle la plus rauque [Lorch, np]. Sherman, dans une étude de l'influence du type de voyelle sur la raucité de la voix a également décrit le /ɔ/ comme plus favorable à l'émergence de la raucité que le /a/ [Sherman, 1954].

Ces travaux ont déterminé notre choix du /ɔ/ comme la voyelle qui pourrait favoriser les occurrences pathologiques de raucité chez les patients dysphoniques.

- Choix du contexte consonantique : pour Rees, la raucité d'une voyelle est plus importante dans un environnement voisé [Rees in Laver, 1980]. Lorch décrit également l'influence des consonnes voisées sur la perception de la raucité : 74% des voyelles perçues rauques sont précédées d'une consonne voisée [Lorch, np]. Il existe plusieurs modes d'articulation pour les consonnes : l'occlusion, la constriction ou la nasalisation. Lors de la production d'une occlusive, les cordes vocales sont plus raides que pour la production d'une constrictive [Stevens, 1998]. On peut donc penser qu'il existe un temps d'adaptation entre la vibration de l'occlusive voisée et la vibration plus souple de la voyelle qui suit. Cet ajustement est susceptible d'imposer une difficulté supplémentaire au sujet dysphonique, et de favoriser l'apparition d'une occurrence pathologique. En effet, chez les patients dysphoniques, la vibration des cordes vocales est rendue précaire par l'existence d'une lésion laryngée. En français, nous disposons de trois occlusives voisées : le /b/, le /d/ et le /g/. La position articulatoire du /g/ est très variable selon la voyelle qui lui est associée. Notre objectif est d'élaborer un contexte phonétique influençant la production de la voyelle qui suit et non l'inverse. Nous avons donc éliminé le /g/. Nous avons envisagé les contextes phonétiques suivants : « bord », « dogue », « bol ». Plusieurs critères ont été pris en compte pour le choix définitif : 1- la qualité de l'écoute nécessite que la voyelle ne soit pas influencée par la consonne suivante. Le /r/ de « bord » peut introduire une impression artificielle de raucité sur le /o/ ouvert. 2- l'accent régional des patients dysphoniques peut induire une variation du contexte d'accentuation linguistique. Ces variations accentuelles sont moindres sur les mots

monosyllabiques. Si l'accent doit être marqué, il est proté sur la voyelle, noyau du groupe CVC [Beckham, 1992].

Suite à ces observations, nous avons choisi le mot /bol/ (« bol ») comme contexte phonétique favorisant la raucité.

Contexte favorisant le souffle : le critère de souffle est défini comme la « *perception d'une importante fuite d'air lors de la phonation* » [Hirano, 1981].

- Choix de la voyelle : le mode de phonation « soufflé » est caractérisé par des cordes vocales en légère abduction avec une faible tension musculaire longitudinale [Laver, 1980]. Une voyelle a plus de probabilité d'être produite de manière soufflée lorsque les muscles thyro-arythénoidiens sont moins tendus. Or, selon l'hypothèse de Lehiste, la « tongue-pull hypothesis » [Lehiste, 1970], l'élévation de la langue tire le larynx vers le haut, favorisant l'augmentation de la tension des cordes vocales. En effet, plusieurs muscles linguaux s'insèrent sur l'os hyoïde, lui-même relié au larynx par les muscles thyro-hyoïdiens et la membrane thyro-hyoïdienne. Par le jeu de ces muscles, l'élévation de la langue a donc un effet tenseur des cordes vocales. Nous avons donc écarté toutes les voyelles fermées puisque leur production induit une position de la langue haute.

Nous avons sélectionné la voyelle la plus ouverte du français : le /a/.

- Choix du contexte consonantique : plusieurs études montrent que la perception du souffle sur les voyelles est influencée par leur coarticulation avec les consonnes [Ohala, 1993 ; Golb, 1993] et notamment par un effet d'assimilation régressive.

Choix de la consonne finale : Ohala évoque un effet d'assimilation régressive lorsqu'une voyelle est suivie d'une constrictive non-voisée : « *lors de la phonation, la in des voyelles peut être soufflée lorsqu'elles sont placées avant une fricative non-voisée.* » [Ohala, 1993]. Selon Golb, « *l'abduction précoce des cordes vocales est probablement une stratégie de facilitation de la production des constrictives non-voisées. Ainsi, le voisement est supprimé avant le début de la fricative et l'abduction glottique permet d'établir une pression d'air suffisante dans la cavité orale pour réaliser la constriction.* » [Golb, 1993]. Lorsqu'une voyelle est suivie d'une constrictive non-voisée, l'écartement des cordes vocales commence pendant la production de la voyelle. En effet, la perception d'une constrictive non-voisée, par exemple le /s/, provient exclusivement du bruit de souffle qui s'écoule entre le palais et la langue. La préparation d'une

constrictive non-voisée engendre une difficulté vibratoire et augmente la fuite glottique, ce qui pourrait mettre en évidence le souffle pathologique d'une voix dysphonique. Les travaux présentés précédemment vont dans le sens de cette hypothèse, puisque nous avions observé que la préparation du /s/ de « façon » favorisait l'apparition d'une occurrence de souffle sur le /a/.

Les trois constrictives non-voisées du français sont le /ʃ/, le /f/, et le /s/. Parmi ces trois phonèmes, nous avons exclu le /ʃ/, dont la position articulatoire est peu précise et qui induit un fort bruit de souffle, peu focalisé, susceptible gêner l'écoute de la voyelle précédente.

Choix de la consonne initiale : nous avons éliminé les contextes phonétiques où la voyelle est comprise entre deux constrictives non-voisées afin d'éviter un bruit de friction trop important qui pourrait masquer l'occurrence de souffle. Les occlusives non-voisées sont susceptibles d'entraîner un bruit d'aspiration sur le début de la voyelle et de fausser également l'interprétation. Les consonnes nasales (/m/ et /n/) ont également été écartées car la résonnance nasale risque également d'entraver la qualité de l'écoute. Nous avons conservé comme contextes phonétiques possibles les mots « basse » et « lasse ». Lors de la conception des phrases constituant le matériau phonétique, il s'est avéré que la locution /las/ était plus facile à employer.

Nous avons sélectionné /las/ comme contexte phonétique pouvant mettre en évidence des occurrences de souffle.

Contexte favorisant le dévoisement : le dévoisement est caractérisé par la désonorisation d'un phonème voisé. Ce phénomène se produit lorsque les cordes vocales ne parviennent pas à entrer en vibration [Laver, 1980].

- Choix de la voyelle : les voyelles sont sonores (ou voisées), elle sont émises lorsque les cordes vocales sont accolées. Sous l'effet du souffle, les cordes vocales vibrent. Le signal vocal est ensuite filtré par les cavités de résonance sus-glottiques [laver, 1980]. Plus la voyelle est courte, plus le temps d'accolement des cordes vocales est court. Les travaux de Sherman [Sherman, 1954] et Fujimoto [Fujimoto, 2002] montrent que les voyelles courtes sont plus fréquemment dévoisées. Chaque phonème possède des caractéristiques intrinsèques propres [Vilkman, 1991]. Les formants des voyelles du français sont définis par des intervalles de valeur fréquentielle [Malmberg, 1979]. En dehors de ces intervalles de fréquence, le phonème n'est plus identifié clairement. Les formants permettent de distinguer par exemple le /a/ du /ɛ/. Chaque phonème possède des formants propres et une durée propre. Les voyelles les plus hautes ont une durée d'émission plus courte [Sherman, 1954 ; Fujimoto, 2002]. Selon Fujimoto, « pour les voyelles

hautes, le passage d'air est réduit. Par conséquent, le débit diminue et la pression intra-orale augmente. Cette augmentation de la pression sus-glottique et trans-glottique ne favorise pas l'initiation de la vibration cordale. Le début de la vibration des cordes vocales (l'attaque) se fait moins facilement. » [Fujimoto, 2002]. Les trois voyelles les plus hautes du français sont le /i/, le /y/ et le /u/.

- Choix du contexte consonantique : lors de la production des occlusives non-voisées, l'air est accumulé derrière un organe buccal (par exemple les lèvres pour les bilabiales) qui va brusquement se relâcher. Le bruit acoustique des occlusives (similaire à une explosion) est le résultat de l'air sous pression qui est expulsé de manière soudaine [Stevens, 1998]. Pour que la pression d'air dans la bouche soit suffisamment élevée, il est nécessaire de mobiliser de l'air envoyé par les poumons. Les cordes vocales doivent être en position écartée pour permettre un passage rapide de l'air. Stevens confirme que les cordes vocales sont en position d'ouverture (abduction) maximale au moment de l'explosion des occlusives [Stevens, 1998]. Lors de la production des constrictives, le contexte non-voisé favorisera le dévoisement d'une voyelle [Ohala, 1993]. En effet, lors de la production de phonèmes non-voisés, il n'y a pas de vibration des cordes vocales. Pour Fujimoto et Ohala, une voyelle fermée entourée de deux occlusives non-voisées est plus souvent dévoisée [Fujimoto, 2002 ; Ohala, 1993]. Golb précise que lorsqu'une voyelle est suivie d'une occlusive non-voisée en français, il y a un arrêt rapide de la vibration glottique. L'arrêt de la voix résulte d'une synchronisation entre l'abduction glottique et le mouvement de fermeture orale [Golb, 1993]. Ce contexte phonétique semble favoriser les occurrences pathologiques de dévoisement. Nos études précédentes confirment que pour le mot « toutes » (voyelle fermée /u/ comprise entre deux occlusives non-voisées /t/), la voyelle est dévoisée dans 19% des cas de voix normales (grade 0), 82% des cas de dysphonie légère (grade 1), et 100% des cas de dysphonie moyenne (grade 2) et sévère (grade 3).

Choix de la consonne initiale : il y a trois occlusives non-voisées en français : /p/, /t/, et /k/. Golb a montré que le phonème /p/ induit un VOT (Voice Onset Time) plus court que le phonème /Tt/. Le VOT est une mesure de la durée entre l'ouverture glottique maximale (explosion de l'occlusive) et le moment où les cordes vocales commencent à vibrer (émergence de la voyelle contenant les deux premiers formants). Quand les cordes vocales sont en ouverture maximale, il faut prendre en compte un délai durant lequel elles rejoignent la ligne médiane, s'accroient et entrent en vibration. Ensuite, les cordes vocales s'ouvrent rapidement pour la production de l'occlusive suivante. Le temps d'ouverture glottique entre le /p/ et la voyelle suivante est plus court qu'entre le /t/ et la voyelle suivante [Golb, 1993]. Nous avons éliminé le phonème /k/ de notre contexte phonétique pour deux raisons : 1- comme pour le phonème /g/, la consonne /k/ est fortement influencée par la voyelle précédente. Sa position articulatoire est très

variable. 2- selon Cooper, « *la durée du dévoisement est plus importante lors de la réalisation du geste articulatoire du /t/ que du /k/.* » [Cooper in Golb, 1993]

Notre choix de consonne initiale se porte donc sur l'occlusive non-voisée /t/.

Choix de la consonne finale : le choix de la consonne finale est déterminé par rapport à la conception du matériau phonétique. Le mot doit être signifiant et la phrase doit être sémantiquement logique. Les seuls mots monosyllabiques signifiants et correspondant à nos critères avec la consonne /k/ en finales sont les mots « tic » ou « tique », qui ne sont pas aisés à introduire dans une phrase. Notre choix de consonne finale se porte donc sur /p/ ou /t/.

Les contextes phonétiques possibles sont donc : /tip/ (« type ») ou /tut/ (toute). Nous avons sélectionné le mot /tip/ comme contexte phonétique favorisant l'occurrence de dévoisement.

Nous avons donc trois contextes phonétiques susceptibles de mettre en évidence les occurrences pathologiques de raucité, souffle et dévoisement chez les patients dysphoniques. Notre choix s'est porté sur :

Raucité	⇒	/bɔl/
Souffle	⇒	/las/
Dévoisement	⇒	/tip/

Influence du contexte linguistique : la parole est un phénomène complexe qui ne peut être réduit qu'à son aspect phonétique. Les contextes phonétiques appartiennent à la partie segmentale de la parole. Dans la parole, il existe des unités segmentales et suprasegmentales. Les unités segmentales correspondent aux phonèmes et à leurs caractéristiques (durée, hauteur, position articulatoire). Les unités suprasegmentales correspondent à la prosodie. Elles véhiculent l'intention du message. Les unités suprasegmentales comprennent, entre autres choses, l'accentuation volontaire d'une syllabe ou la variation mélodique de la phrase. Les unités suprasegmentales constituent une source d'influence pour l'apparition des occurrences, au même titre que le contexte phonétique.

La parole est adressée habituellement à un interlocuteur, c'est une situation de communication. Pour deux phrases de même contenu phonétique, la variation mélodique permet de différencier une

affirmation d'une interrogation, par exemple : « *Il part demain.* » / « *Il part demain ?* ». A l'oral, seule l'intonation permet de distinguer les deux phrases. En cela, la prosodie porte l'intention du message.

L'accentuation permet de mettre en évidence le mot de la phrase qui porte le sens principal du message. C'est l'influence sémantique. Par exemple, reprenons la phrase « *Il part demain.* » dans deux contextes sémantiques différents :

« Il va à la plage demain ? – Non, non, tu n'as pas compris : *il part demain.* »

« Je l'emmène à l'aéroport Jeudi ? – Non, non, tu n'as pas compris : *il part demain.* »

Ici encore, les deux phrases « *il part demain.* » sont identiques sur le plan phonétiques. Cette fois, ce sont deux affirmatives. Cependant, l'influence linguistique, par l'intermédiaire de l'accentuation, intervient sur le mot qui contient l'information principale, en fonction du contexte sémantique. Ce mot est mis en valeur, il devient émergent dans la phrase, grâce à l'accentuation qui est ressentie comme une « *insistance* ». A l'oral, ces deux phrases ne sont donc pas prononcées de manière identique.

L'aspect suprasegmental doit être pris en compte pour deux raisons essentielles : 1- la plainte du patient dysphonique porte sur les difficultés éprouvées en situation de communication [Goldstone, 1998] ; 2- l'accentuation de la parole modifie les caractéristiques des phonèmes [De Jong, 1993]. En effet, l'accentuation d'une syllabe modifie la durée et l'intensité de la voyelle. L'accentuation est donc largement susceptible d'influencer la qualité de production de la voyelle en présence d'une dysphonie. Ainsi, il semble que l'effet des contextes phonétiques que nous avons élaborés subissent l'influence du contexte prosodique. Pour De Jong, « *la structure prosodique affecte les temps et les mouvements articulatoires, donc les possibilités de chevauchement entre les phonèmes.* » [De Jong, 1993]. Les caractéristiques de durée des phonèmes pourront être modifiées [Goldstein in De Jong, 1993]. Certains phonèmes vont être tronqués du fait du chevauchement des gestes articulatoires, d'autres, au contraire, seront hyper-articulés [De Jong, 1993]. Différents types d'accentuation sont décrits : au niveau du mot, au niveau de la proposition, au niveau de la phrase [De Jong, 1993]. Chaque mot possède une accentuation propre. Cette accentuation dépend de la longueur du mot (nombre de syllabes) et des règles d'accentuation de chaque langue. « *Lorsqu'une syllabe est accentuée, la voyelle de la syllabe se manifeste par des schémas de formants différents ainsi que pas des différences de durée et d'intensité. La voyelle accentuée est plus longue, plus intense.* » [De Jong, 1993]. S'il existe des accentuations secondaires (au niveau lexical), une seule syllabe par phrase énoncée porte l'accentuation principale. Cette accentuation dépend de l'intention du discours. De Jong la définit comme « *l'accent nucléaire* » de la phrase. Cette syllabe accentuée est la plus proéminente dans une phrase. Lorsqu'on écoute une phrase, la syllabe accentuée se distingue des autres parce qu'elle est perçue comme émergente. L'accent principal porte sur le mot qui a le plus grand contenu informatif dans une phrase [De Jong, 1993]. Ce mot est la cible du discours, le support d'une communication informative : « Qui est cette personne assise sur le banc ? – C'est mon *fils* assis sur le banc. » Dans la deuxième phrase, le mot

« fils » porte l’accentuation principale de la phrase puisqu’il donne l’information que l’on attend en réponse à la question posée. Dans le cas d’un mot monosyllabique de type CVC, l’accent nucléaire influence forcément les caractéristiques de la voyelle.

Nous choisissons donc de placer les contextes phonétiques que nous avons sélectionnés dans des phrases, en position accentuée et inaccentuée pour évaluer l’importance du contexte prosodique sur l’apparition d’occurrences pathologiques. Dans le but de l’assurer de l’émergence de l’accentuation sur le mot précis que nous avons choisi, chaque phrase sera introduite par une phrase interrogative portant sur le mot contenant le contexte phonétique que nous avons choisi. Lorsque la question porte spécifiquement sur le mot sélectionné, le contexte phonétique est en position accentuée ; lorsque la question porte sur autre chose, le contexte phonétique est en position inaccentuée :

R
A
U
C
I
T
E

- Préférez vous boire votre café dans une tasse ou dans un bol ?
 C'est dans un **bol** que je préfère le boire. ⇒ accentué
- Vous aimez les bols bleus, n'est-ce pas ?
 Je préfère les jolis **bols** rouges. ⇒ inaccentué

S
O
U
F
F
L
E

- Comment se sent-elle aujourd’hui ?
 Elle se sent **lasse**, mais ça va un peu mieux qu’hier. ⇒ accentué
- Tu as été prise pour ce nouveau travail ?
 Non, hélas je n’ai pas été prise. ⇒ inaccentué

D
E
V
O
I
S
E
M
E
N
T

- Quelles sont les voitures que tu préfères ?
 Le **type** de voitures que je préfère, c'est vraiment les voitures de sport. ⇒ accentué
- Que penses-tu de ce garçon ?
 C'est vraiment un sale **type**. ⇒ inaccentué

Ce corpus de phrases constitue un nouveau matériau phonétique conçu spécifiquement pour l'analyse perceptive. L'expérimentation que nous présentons maintenant est une étude préliminaire qui a pour but de tester ces phrases sur un groupe de sujets normaux et dysphoniques. Nos hypothèses sont :

- 1- les contextes phonétiques favorisent l'émergence des critères concernés
- 2- le contexte linguistique d'accentuation favorise l'émergence des critères concernés
- 3- l'influence des contextes est d'importance moindre pour les voix normales (qui restent normales) et pour les dysphonies sévères (qui présentent une atteinte sur l'ensemble des phonèmes).

Protocole expérimental : Les sujets dysphoniques de l'étude ont été recrutés et enregistrés lors de la consultation ORL du CHU Timone, selon les techniques habituelles. Nos critères d'inclusions étaient : 1- présence d'une dysphonie ; 2- un larynx fonctionnel (les sujets ne doivent pas être entravés dans leurs mouvements layngés d'adduction et d'abduction, nous avons donc exclus les sujets atteints de paralysie laryngée) ; 3- pas d'antécédent de rééducation vocale. Nos critères d'exclusion étaient : 1- tout antécédent chirurgical susceptible d'entraîner une modification du geste vocal ; 2- présence d'une dysarthrie.

Nous avons enregistré tous les patients correspondant à ces critères dans les délais impartis pour la réalisation de ce travail. Sur les 13 sujets dysphoniques enregistrés, la répartition était de 12 femmes pour 1 homme. Pour plus de cohérence dans l'analyse des résultats, nous n'avons conservé que les sujets féminins. Parmi ces 12 femmes, un jury de pré-sélection composé de deux orthophonistes a retenu 5 patientes en fonction du respect du schéma prosodique (seuls les sujets respectant les schémas prosodiques imposés sur chaque phrase ont été conservés) et du niveau de dysphonie (répartition en grade de dysphonie : léger, moyen, sévère). Nous avons également inclus deux sujets témoins enregistrés parmi les étudiantes de l'école d'orthophonie de Marseille et que nous avons soumis aux mêmes critères de sélection.

	Age	Pathologie	Grade
Sujet 1	25	Œdème de Reinke	1
Sujet 2	62	Œdème de Reinke	3
Sujet 3	23	Sain	0
Sujet 4	33	Granulôme	1
Sujet 5	44	Kyste + Œdème controlatéral	2
Sujet 6	40	Sain	0
Sujet 7	74	Polype	2

Tableau VI.1. : répartition des sujets de l'étude -tous féminins- en âge, pathologie, et niveau dysphonique

Les 12 phrases de mise en contexte phonétique et linguistique ont été enregistrées. Seules les 6 affirmatives ont été numérisées pour les analyses ultérieures. Les voix ont été évaluées par analyse perceptive classique par un jury composé de 3 experts (un phoniatre et deux orthophoniste) selon la méthode du consensus. La tâche du jury d'écoute était de déterminer les paramètres GRB (de l'échelle GRBAS) de chaque patient, sur chaque phrase, sur une échelle ordinaire de 0 à 3. Ces évaluations se sont déroulées sur 6 sessions (une session par phrase) à une semaine d'intervalle (en 6 semaines). Les sessions ont été réalisées dans l'ordre suivant :

- Semaine 1 : session 1, contexte raucité / position accentuée.
- Semaine 2 : session 2, contexte souffle / position inaccentuée
- Semaine 3 : session 3, contexte dévoisement / position accentuée
- Semaine 4 : session 4, contexte raucité / position inaccentuée
- Semaine 5 : session 5, contexte souffle / position accentuée
- Semaine 6 : session 6, contexte dévoisement / position inaccentuée

Lors de chaque session, les voix des 7 sujets étaient présentées dans un ordre aléatoire différent, afin d'éviter les effets d'apprentissage et de diminuer l'effet de contexte de l'ordre de passation. Selon le même déroulement, une orthophoniste expérimentée a réalisé l'écoute des 6 phrases prononcées par les 7 sujets, selon la méthode d'analyse par phonetic labelling (critères R, B, D).

Une modification de la classification des sujets en grade global de dysphonie est intervenue suite à l'expertise par jury d'écoute (tableau VI.2.). Le sujet 7, évalué en grade 2 par le jury de pré-sélection a été reclassé en grade 3. Le sujet 3, qui est un sujet sain, a été perçu comme une dysphonie légère et classé en grade 1. Il est possible qu'une légère laryngite ait affecté ce sujet lors de l'enregistrement de sa voix.

	Jury de pré-selection (grade global)	Jury d'expertise (grade global)
Sujet 1	1	1
Sujet 2	3	3
Sujet 3	0	1
Sujet 4	1	1
Sujet 5	2	2
Sujet 6	0	0
Sujet 7	2	3

Tableau VI. 2. : classification de recrutement en grade global par le jury de pré-sélection, et classification définitive par le jury d'expertise

Résultats :

- Analyse perceptive classique : chaque sujet a été évalué 6 fois sur la production de 6 phrases différentes. Pour chacune des phrases prononcées par un même sujet, nous avons obtenu le grade d'expertise en GRB. Le tableau VI.3. montre le classement définitif de chaque sujet sur l'ensemble du corpus de phrases. Le chiffre entre parenthèse correspond à l'indice de confiance de l'évaluation finale : par exemple, sur les 6 sessions d'écoute, un sujet noté 1 fois grade 0, 4 fois grade 1, et 1 fois grade 2, intègre le groupe dysphonie légère (grade 1), avec un indice de confiance de 66%.

	G (grade global)	R (raucité)	B (souffle)
Sujet 1	1 (66%)	1 (50%)	1 (83%)
Sujet 2	3 (100%)	3 (100%)	3 (66%)
Sujet 3	1 (66%)	1 (50%)	1 (66%)
Sujet 4	1 (50%)	1 (33%)	1 (66%)
Sujet 5	2 (100%)	2 (66%)	2 (100%)
Sujet 6	0 (100%)	0 (100%)	0 (100%)
Sujet 7	3 (100%)	2 (100%)	3 (100%)

Tableau VI. 3. : résultat de l'expertise par analyse perceptive classique sur GRB sur l'ensemble des phrases. Entre parenthèses l'indice de confiance (pourcentage de réponses identiques).

Les résultats montrent que 2 voix (sujets 6 et 7) obtiennent des résultats identiques sur l'ensemble des paramètres quelle que soit la phrase prononcée. Les sujets 2 et 5 obtiennent des résultats identiques sur deux des trois critères étudiés. Les voix 1, 3 et 4 semblent davantage varier en fonction des phrases prononcées. Les tableaux VI.4., montrent le détail de l'évolution de l'expertise en fonction des contextes pour les critères G (4a), R (4b) et B(4c).

Grade Global	Raucité accentué	Raucité inaccentué	Souffle accentué	Souffle inaccentué	Dévoisement accentué	Dévoisement inaccentué
Sujet 1	1	0	2	1	1	1
Sujet 2	3	3	3	3	3	3
Sujet 3	0	1	2	1	1	1
Sujet 4	0	1	1	0	1	2
Sujet 5	2	2	2	2	2	2
Sujet 6	0	0	0	0	0	0
Sujet 7	3	3	3	3	3	3

Tableau VI. 4a. : évolution de l'expertise en fonction des contextes phonétiques pour l'évaluation du grade global

Raucité	Raucité accentué	Raucité inaccentué	Souffle accentué	Souffle inaccentué	Dévoisement accentué	Dévoisement Inaccentué
Sujet 1	1	0	2	1	2	1
Sujet 2	3	3	3	3	3	3
Sujet 3	0	1	2	0	1	1
Sujet 4	1	0	2	0	1	2
Sujet 5	1	2	1	2	2	2
Sujet 6	0	0	0	0	0	0
Sujet 7	2	2	2	2	2	2

Tableau VI.4b. : évolution de l'expertise en fonction des contextes phonétiques pour l'évaluation de la raucité

Souffle	Raucité accentué	Raucité inaccentué	Souffle accentué	Souffle inaccentué	Dévoisement accentué	Dévoisement Inaccentué
Sujet 1	1	1	1	0	1	1
Sujet 2	1	3	3	1	3	3
Sujet 3	0	1	1	2	1	1
Sujet 4	0	1	1	0	1	1
Sujet 5	2	2	2	2	2	2
Sujet 6	0	0	0	0	0	0
Sujet 7	3	3	3	3	3	3

Tableau VI.4c. : évolution de l'expertise en fonction des contextes phonétiques pour l'évaluation du souffle

Les résultats montrent que le contexte proposé ne semble pas influencer l'évaluation du grade global des voix normales et des dysphonies sévères qui sont toujours notées de la même façon par le jury d'experts (tableau VI.4a.). Le jury est plus sensible au contexte pour la discrimination des dysphonies légères (G1) et moyennes (G2). La phrase dans laquelle s'insère le contexte linguistique de raucité en position accentuée semble minorer la perception du jury. Les voix 3 et 4 (G1, tableau VI.3.) ont été classées en voix normales (tableau VI.4.a) en contexte raucité accentué. En revanche, le contexte de souffle accentué semble majorer la perception du jury d'experts : les voix 1 et 3 (G1, tableau VI.3.) sont notées grade 2 (tableau VI.4a.). Par ailleurs, le contexte de raucité accentué semble minorer la perception du souffle (voix VI.3. et 4, tableau VI.4c.). La voix peut donc être perçue différemment altérée selon le matériau proposé.

- Analyse par phonetic labelling : les résultats de l'analyse par phonetic labelling des contextes spécifiques est présentée dans le tableau 5. Les occurrences pathologiques sont mises en évidence selon un code couleurs décrit dans la légende. Ces couleurs ont été choisies de façon

arbitraire afin de faciliter la lecture du tableau, mais ne correspondent en aucun cas à un critère de « gravité » du paramètre concerné. Le gris correspond aux phonèmes qui n'ont pas été prononcés par le sujet pour des raisons de débit de parole ou d'accent régional. Notons que l'intégralité des phrases affirmatives a fait l'objet de l'analyse par phonetic labelling, mais nous ne présentons ici que les résultats obtenus sur les contextes spécifiques par souci de concision.

	Raucité accentué			Raucité inaccentué			Souffle accentué			Souffle inaccentué			Dévoismt accentué			Dévoismt inaccentué		
Voix	b	ɔ	/	b	ɔ	/	/	a	s	/	a	s	t	i	p	t	i	p
Sujet 1																		
Sujet 2																		
Sujet 3																		
Sujet 4																		
Sujet 5																		
Sujet 6																		
Sujet 7																		
Normal																		
Raucité																		
Souffle																		
Dévoisement																		
Omission																		

Tableau VI.5. : résultats de l'analyse par phonetic labelling portée sur les contextes spécifiques : raucité accentué / raucité inaccentué ; souffle accentué / souffle inaccentué ; dévoisement accentué / dévoisement inaccentué.

A la lecture de ce tableau, on note que le contexte de raucité inaccentué semble plus favorable que le contexte de raucité accentué pour l'apparition des occurrences pathologiques de raucité sur la voyelle (2 en accentué contre 4 en inaccentué). A l'inverse, le souffle semble favorisé par le contexte accentué (4 occurrences de souffle en accentué contre 2 en inaccentué). Enfin, le contexte phonétique /tip/ semble tout à fait favorable à l'émergence du dévoisement avec un léger avantage du contexte inaccentué (3 occurrences de dévoisement en contexte accentué contre 4 en contexte inaccentué).

Le tableau suivant permet de faire la synthèse entre l'expertise par le jury d'écoute et l'analyse par phonetic labelling. Le chiffre en pourcentage correspond au rapport « phonèmes atteints / nombre total de phonèmes voisés prononcés ». Le chiffre entre parenthèses correspond au grade attribué par le jury d'experts.

	Occurrences pathologiques	Occurrences de raucité	Occurrences de souffle	Occurrences de dévoisement
Sujet 1	33% (G 1)	11% (R 1)	23% (B 1)	5%
Sujet 2	68% (G 3)	48% (R 3)	31% (B 3)	12%
Sujet 3	37% (G 1)	6% (R 1)	28% (B 1)	5%
Sujet 4	37% (G 1)	12% (R 1)	18% (B 1)	6%
Sujet 5	47% (G2)	24% (R 2)	29% (B 2)	5%
Sujet 6	11% (G 0)	1% (R 0)	7% (B 0)	3%
Sujet 7	65% (G 3)	34% (R 2)	44% (B 3)	7%

Tableau VI.6. : rapport entre le nombre d'occurrences pathologiques et la totalité des phonèmes voisés prononcés par les sujets. Entre parenthèse le grade attribué par l'expertise lors de l'analyse perceptive classique.

Discussion : L'objectif de cette étude préliminaire était d'élaborer des contextes linguistiques favorisant les occurrences pathologiques de chaque critère de dysphonie afin de mieux caractériser la voix des patients de voix. La première partie de ce travail consistait en la conception de mots dont les contextes phonétiques favorisent l'apparition de la raucité, du souffle ou du dévoisement, puis de placer ces mots dans des phrases dont les contextes sémantiques permettent de retrouver ces mots en position d'accentuation linguistique ou non. Nous avons ensuite procédé à une double analyse de la voix des sujets recrutés pour l'étude : une évaluation perceptive classique et une analyse par phonetic labelling. Avant d'entamer la discussion, nous souhaitons étudier le cas de la voix du sujet 6. Il s'agit d'une voix de contrôle prononcée par un sujet sain. Sur cette voix, on ne retrouve aucune occurrence pathologique sur les contextes sélectionnés (tableau VI.5.). Elle a été notée en grade 0 par notre jury d'écoute avec un indice de confiance de 100%. Cette voix présente un nombre très faible d'occurrences pathologiques lors de l'analyse par phonetic labelling : le ratio « phonèmes atteints / phonèmes voisés » est très faible, de l'ordre de 11% (12 phonèmes atteints sur 110 phonèmes voisés). L'information principale relevée dans l'étude de cette voix est que nos contextes favorisant l'apparition d'occurrences pathologiques n'imposent l'apparition pas d'occurrences « artificielles » sur les voix normales, ce qui est en parfaite adéquation avec notre démarche. La suite de la discussion sera axée sur les réflexions imposées par l'étude des 6 autres voix, toutes jugées dysphoniques.

1- Influence de l'accentuation :

L'étude du graphique VI.5. permet d'observer que contextes phonétiques de raucité et de dévoisement (/bol/ et /tip/) présentent un nombre d'occurrences pathologiques plus importants lorsqu'ils sont en position inaccentuée. Le contexte de souffle (/las/) présente des occurrences différentes dans les deux modes d'accentuation, mais leur nombre est identique. Au total, le nombre d'occurrences

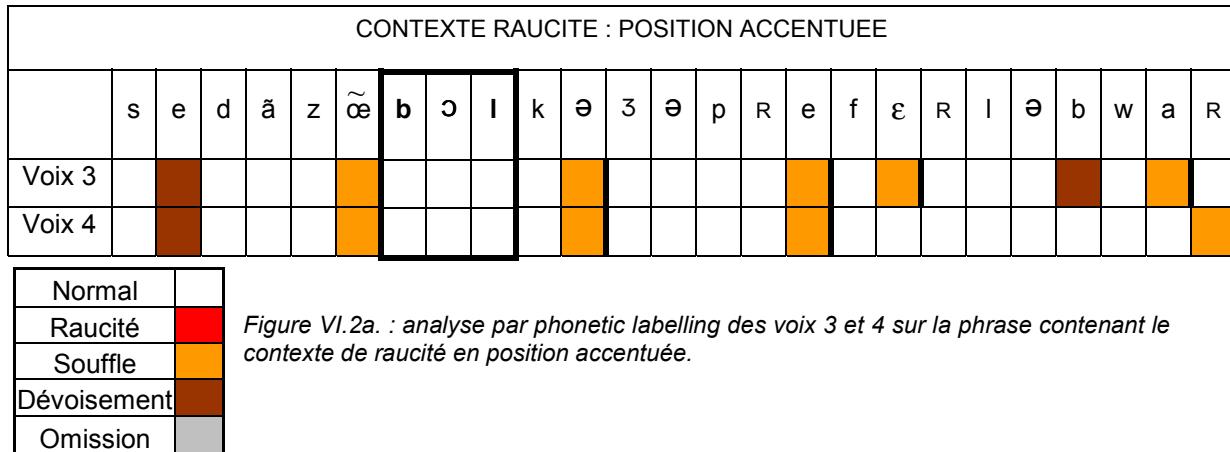
pathologiques sur la voyelle du contexte phonétique en position inaccentuée est au moins égal (contexte souffle) et souvent supérieur (contextes raucité et dévoisement) au nombre d'occurrences en position accentuée. Cette remarque va dans le sens de l'hypothèse d'« hyper-articulation » proposée par De Jong [De Jong, 1993]. En effet, lorsqu'une syllabe est accentuée, la voyelle doit ressortir de façon émergente. L'articulation de cette voyelle est donc particulièrement soignée. Il y a un hyper-ciblage de la voyelle. Ce mode d'articulation renforce les caractéristiques propres de la voyelle [De Jong, 1993]. Ainsi, l'accentuation a pour effet de diminuer la coarticulation entre les phonèmes, et par conséquent, la voyelle est moins soumise à l'influence des consonnes qui l'entourent. L'accentuation de la syllabe neutralise en quelque sorte la mise en contexte phonétique. La poursuite de ce travail devra donc tenir compte de cet élément et proposer les contextes phonétiques en position inaccentuée de manière préférentielle.

2- Le contexte de raucité :

Pour les deux voix de grade 3 (sujets 2 et 7) on retrouve une occurrence pathologique sur notre contexte spécifique, quelle que soit l'accentuation. Cependant, il faut signaler que les dysphonies sévères présentent un grand nombre d'occurrences pathologiques. Dans nos études précédentes, il nous est même fréquemment arrivé de retrouver un ratio « phonèmes atteints / phonèmes voisés » égal à 100%. Les bons résultats obtenus avec les contextes sur les dysphonies sévères ne prouvent pas l'efficacité des contextes mais plutôt l'importance de l'atteinte pathologique.

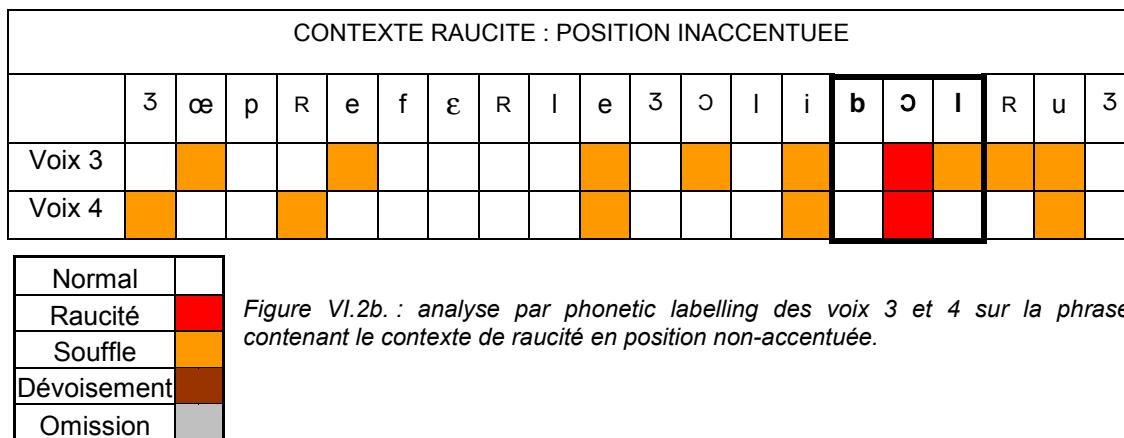
L'étude de cas des voix 3 et 4 (grades 1) fournit une base de réflexion particulièrement intéressante pour l'étude de la raucité. Lors de l'évaluation perceptive classique, ces deux voix obtiennent une cotation de 1 en raucité avec un indice de confiance faible (respectivement 50% et 33%) (tableau VI.3.). Il semble que la perception de la raucité est fortement soumise au matériau linguistique (tableau VI.4b.) : selon la phrase prononcée par le locuteur, la perception globale de raucité varie de 0 (absence totale du paramètre concerné) à 2 (raucité moyenne). Au total, ces deux voix sont classées dysphonies légères (G1), raucité légère (R1), souffle léger (B1). En ce qui concerne la voix 4 qui obtient un indice de confiance de seulement 33% sur le paramètre de raucité, son classement définitif est relativement arbitraire puisqu'il résulte d'une moyenne : en effet, sur 6 phrases prononcées, cette voix est classée 2 fois R0, 2 fois R1, et 2 fois R2. Les résultats par phonetic labelling montrent que 55% des phonèmes pathologiques de la voix 4 sont soufflés contre 36% pour la raucité, quel que soit le matériau phonétique. De même en ce qui concerne la voix 3, 78% des phonèmes atteints d'une occurrence pathologique présentent une occurrence de souffle, contre 17% pour la raucité. Il semble donc exister une nette prévalence du souffle pour ces deux voix.

Si l'on regarde les résultats obtenus par phonetic labelling sur la phrase « raucité accentuée », on constate l'absence totale d'occurrences de raucité (figure VI.2a.).



L'analyse par phonetic labelling (tableau VI.6.) montre que la caractéristique pathologique dominante de ces deux voix est le souffle.

Sur la phrase « raucité inaccentuée » montre une large majorité d'occurrences pathologiques de souffle (plus de 80% des phonèmes atteints). On n'observe aucune occurrence de raucité, sauf sur la mise en contexte spécifique, position inaccentuée (figure VI.2b.).



Au total, il semble que notre contexte phonétique /bɔl/, placé en position inaccentuée soit un facteur favorisant de l'émergence de la raucité pour des voix présentant une raucité légère (R1).

Néanmoins, il semble que notre choix de contexte « raucité » soit encore perfectible. Si l'on regarde attentivement l'ensemble des résultats de l'analyse par phonetic labelling (voir annexes), on constate que, conformément aux bases théoriques présentées dans l'état des lieux, la raucité est plus présente

dans un environnement phonétique voisé et elle apparaît de façon plus manifeste sur les voyelles ouvertes [Ohala, 1993 ; Rees in Laver, 1980]. En ce qui concerne le choix spécifique de la voyelle, nous avions deux possibilités : le /a/ ou le /ɔ/, que nous avons sélectionné, basés sur les travaux de Lorch [Lorch, np] et de Sherman [Sherman, 1954]. Cependant, l'étude de Lorch était basée sur l'étude exclusive de la dysphonie spasmodique dont le modèle physiopathologique est marginal dans l'étude des dysphonies, et peut-être pas généralisable à l'ensemble des dysphonies dysfonctionnelles. L'étude de Sherman portait spécifiquement sur la raucité, chez des sujets recrutés précisément pour le caractère rauque de leur voix. Or, dans notre travail, les caractéristiques vocales de nos patients sont plus variées et n'admettent pas exclusivement la raucité comme composante principale. Pour vérifier quelle est la voyelle notée le plus souvent rauque, nous avons recensé le nombre d'occurrences pathologiques de raucité sur toutes les voyelles de notre corpus de phrases. Il semble que le /a/ soit la voyelle présentant le plus d'occurrences pathologiques de raucité : dans notre corpus, la voyelle /a/ est produite 56 fois ; elle présente 25 occurrences pathologiques de raucité, ce qui donne un ratio « /a/ rauque / /a/ prononcés » de 45%. En ce qui concerne le /ɔ/, il n'est perçu rauque que dans 38% des cas. Il semble donc que la voyelle /a/ favorise davantage l'apparition de la raucité. Le contexte phonétique /bal/ pourrait donc être proposé en remplacement du contexte /bɔl/, utilisé dans cette étude.

3- Le contexte de souffle :

Notre hypothèse était que le contexte phonétique /las/ serait favorable à l'apparition d'occurrences pathologiques de souffle. Nos résultats montrent que les dysphonies moyennes (G2) et sévères (G3) sont plus sensibles à l'influence des contextes phonétiques. Pour les 3 voix de grade global 1, le phonème /a/ est produit une seule fois soufflé (voix 1, contexte souffle accentué) sur les 6 possibilités d'occurrences de souffle (voir annexes). En revanche, les voix 2, 5 et 7 (respectivement G3, G2 et G3), le phonème /a/ est produit 4 fois soufflé sur les 6 contextes exposés, dont 3 fois en position accentuée. Le contexte /las/, en position accentuée permet de révéler l'apparition d'occurrences de souffle. Néanmoins, il faut préciser que la voyelle est également atteinte d'occurrences de raucité, de façon même plus importante : sur l'ensemble des 7 voix, on constate que la voyelle /a/ de /las/ est perçue soufflée 6 fois et rauque 7 fois (dont 4 fois rauque et soufflée). Il semble donc que le contexte phonétique /las/ favorise effectivement les occurrences pathologiques (la voyelle /a/ est atteinte dans 65% des cas), mais pas spécifiquement de souffle.

La contre-performance de ce contexte est très vraisemblablement liée au choix de la voyelle /a/, qui, comme nous l'avons vu dans le paragraphe précédent, semble très favorable à la raucité. Nous avons confronté les occurrences de souffle relevées pour chaque voyelle présente dans notre corpus de phrases. Les résultats ont montré des fréquences similaires d'occurrences de souffle et aucune voyelle

de notre corpus n'est sortie du lot. En effet, l'apparition du souffle semble être davantage dépendant des consonnes qui entourent la voyelle que de la voyelle elle-même. Golb [Golb, 1993] et Ohala [Ohala, 1993] indiquent que la préparation d'une constrictive non-voisée entraîne un effet de souffle sur la voyelle précédente, en particuliers lorsqu'il s'agit d'un /f/. Nos résultats vont dans ce sens puisque la voyelle /e/ du mot /prefER/, présent dans 3 de nos phrases, produit une occurrence pathologique de souffle dans 88% des voix dysphoniques (cas n°6, voix normale, exclu). Un autre contexte consonantique semble très favorable à l'apparition du souffle : une voyelle entourée de deux constrictives non-voisées. Dans la phrase « souffle accentué » (« Elle se sent lasse... »), le /ə/ de /sə/ (/elsəsəlas/) est perçu soufflé dans 83% des voix dysphoniques (5 fois sur 6).

Au total, il semble qu'un contexte intégrant une voyelle pas trop ouverte entre deux constrictives non-voisées serait davantage favorable à l'apparition d'une occurrence pathologique de souffle.

4- Le contexte de dévoisement :

Le contexte phonétique /tip/ favorise les occurrences de dévoisement pour 3 voix pathologiques sur 6 en position accentuée et pour 4 voix sur 6 en position inaccentuée. Dans la phrase intégrant le contexte linguistique accentué (« Le type de voitures... »), les résultats de l'analyse par phonetic labelling montrent que le phonème /p/ de /tip/ n'est pas systématiquement prononcé : il est omis par trois sujets (voix 1, 4 et 5) dont le débit de parole est si rapide que le phonème est escamoté. Dans ce cas, les sujets ne prononcent plus /lətipdəvvwatyr.../, mais /lətidəvvwatyr.../. Le contexte /tip/ est donc transformé en /tid/. La voyelle /i/ n'est donc plus placée entre deux occlusives non voisées (/t/ et /p/) mais entre une occlusive non-voisée /t/ et une occlusive voisée /d/, et n'est donc plus conforme aux bases théoriques sur lesquelles nous nous étions appuyés lors de la conception du contexte. Il n'y a pas de raison particulière pour que /tid/ favorise le dévoisement, puisque la vibration est maintenue sur le /i/ (qui se trouve rallongé) et sur les phonèmes suivants. En revanche, lorsque les trois phonèmes /t/ /i/ /p/ sont respectés, le contexte /tip/ est très favorable à l'apparition d'occurrences pathologiques de dévoisement. En position accentuée, il fonctionne systématiquement, en position inaccentuée, il fonctionne dans 4 cas sur 6. Sur la voix 3, le /i/ de /tip/ présente une occurrence de souffle qui peut correspondre à une forme atténuée de dévoisement. Notons également que la voix 5 ne présente pas d'occurrence de dévoisement sur cette phrase, uniquement des occurrences rauques. Les voix rauques présentent rarement des occurrences de dévoisement, au contraire des voix soufflées. Le fait que la voix 5 ne montre pas de dévoisement montre que les contextes phonétiques incitent l'apparition d'occurrences pour les voix qui ont tendance à être atteintes par ce critère, mais pas pour les autres. Le fonctionnement des contextes respecte les caractéristiques intrinsèques des dysphonies, sans imposer l'apparition d'un critère qui n'est pas constitutif de la voix. On peut donc considérer que les contextes

phonétiques sont incitatifs, mais pas prescriptifs. Cette remarque est importante, car un contexte produisant une occurrence de manière systématique, sans tenir compte du type de dysphonie présenté par le sujet (tendance rauque, ou soufflée,...) n'apporterait aucune information. A l'inverse, l'absence d'occurrences sur la voix normale ou une dysphonie très essentiellement rauque est représentative de l'efficacité du contexte.

Conclusion : L'objectif de ce travail était de proposer une étude préliminaire de l'influence des contextes phonétiques et linguistiques sur l'apparition d'occurrences pathologiques chez les patients dysphoniques. A partir de l'analyse par phonetic labelling, proposée dans nos travaux précédents, nous souhaitions élaborer un corpus de phrases permettant de mettre en évidence les caractéristiques pathologiques de la voix dysphonique. Notre hypothèse était que les matériaux phonétiques et linguistiques influencent la production d'occurrences dysphoniques. Nous avons élaborés 3 contextes phonétiques, placés chacun dans deux contextes prosodiques dans le but de favoriser l'émergence d'occurrences de raucité, de souffle et de dévoisement. Les résultats que nous avons obtenus montrent que la perception du jury est influencée par le matériau phonétique, puisqu'une même voix peut être classée de 0 à 2 en fonction de la phrase prononcée.

L'analyse par phonetic labelling montre que l'influence des contextes est largement soumise à l'accentuation. Les contextes en position inaccentuée semblent davantage favorables à l'apparition d'occurrences spécifiques. Il semble par ailleurs que l'intérêt des mises en contexte soit particulièrement porté sur l'évaluation des dysphonies légères et moyennes (G1 et 2). Dans ce travail, sur les dysphonies légères, l'analyse perceptive classique montre des indices de confiances faibles et même très faibles. L'analyse par phonetic labelling des contextes spécifiques permet de mieux caractériser ces dysphonies « limites », puisqu'ils sont représentatifs des timbres particuliers des voix (plutôt rauque ou plutôt soufflé) sans imposer l'apparition d'une occurrences pathologique non-constitutive de ce timbre. Néanmoins, à l'heure actuelle, l'écoute des contextes seuls ne permet pas un classement suffisamment précis des dysphonies. En effet, après analyse de nos résultats, certains choix théoriques établis à partir de la littérature ne sont peut être pas les plus judicieux. Certains contextes doivent donc être modifiés et testés à nouveau avant que notre matériau phonétique soit validé. Néanmoins les résultats observés dans ce travail ouvrent de nombreuses pistes quant au remaniement des contextes et des perspectives d'application de la méthode. Les prochaines études sur ce thème devraient s'employer à définir de nouveaux contextes et à vérifier la reproductibilité de l'apparition des occurrences intra-locuteur. A terme, il est possible d'imaginer que l'écoute spécifiques de certains mots soit nécessaire et suffisante à la caractérisation de la dysphonie, permettant de simplifier la tâche d'évaluation perceptive et de réduire le temps de passation. De même, l'évolution de la répartition des occurrences pathologiques en cours de traitement pourrait être un indice de l'efficacité de la rééducation, et faire de cette méthode un outil clinique à part entière. A plus long terme, l'analyse

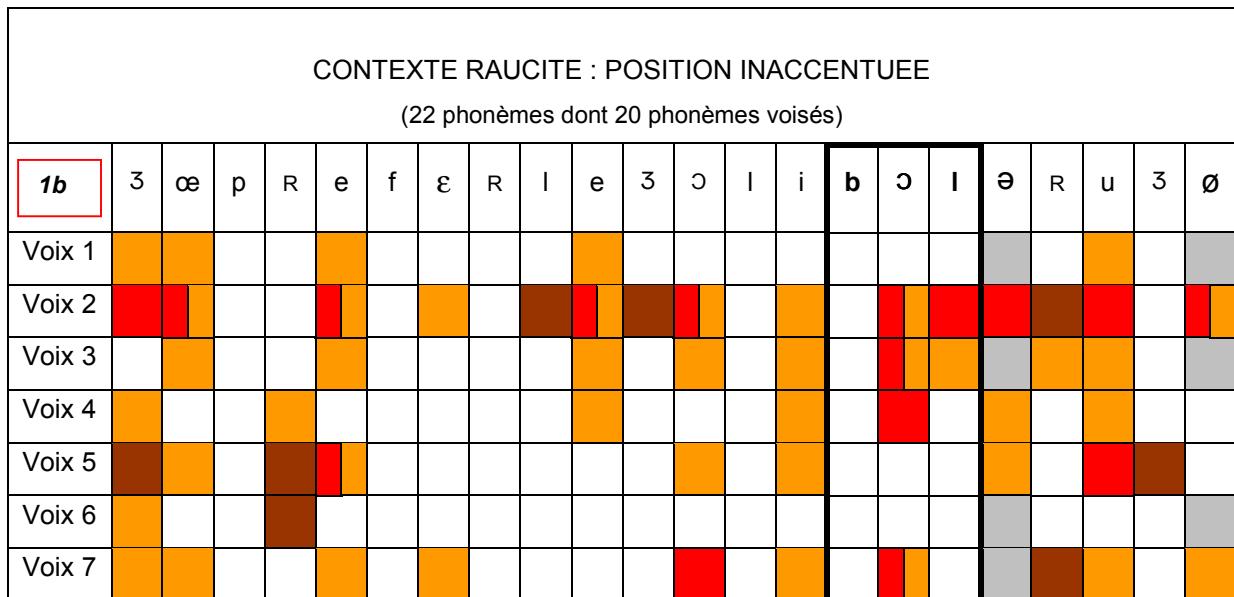
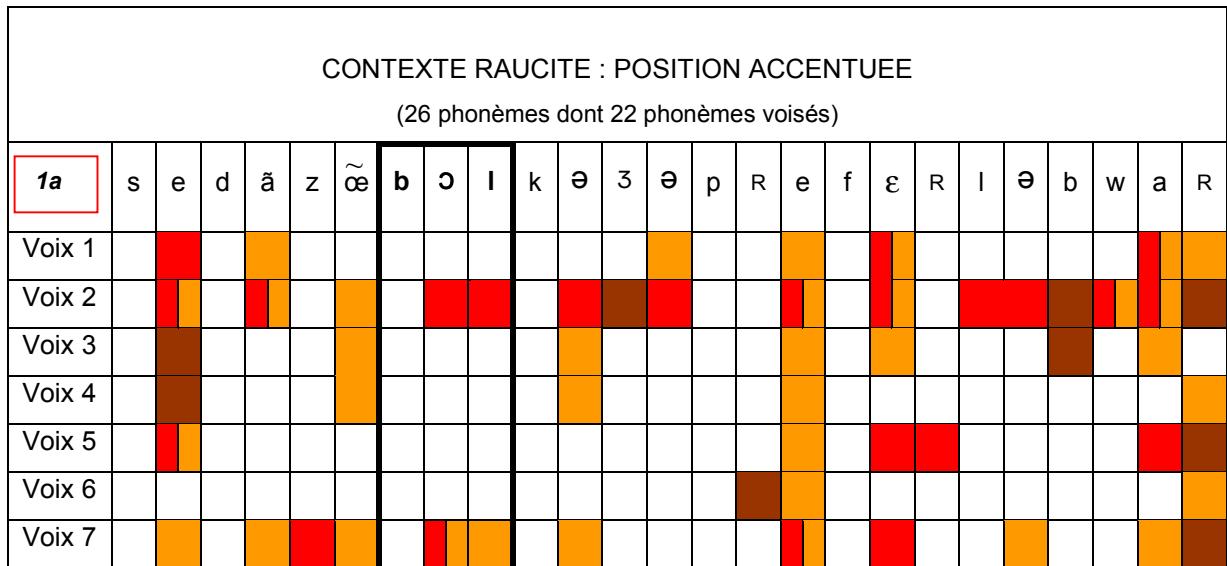
phonétique et linguistique des évènements dysphoniques au cours de la parole pourrait constituer une base de travail unique pour le développement d'outils informatiques de « reconnaissance automatique ». En effet, la recherche consacrée à la reconnaissance automatique du locuteur est en demande d'une banque de données de sujets dysphoniques permettant de tester leurs algorithmes sur la voix pathologique. En effet, les applications diverses de la reconnaissance automatique du locuteur, comme par exemple la sécurisation d'accès par empreinte vocale, nécessitent que l'algorithme fonctionne même si le locuteur habituel présente une voix altérée temporairement par une laryngite. De notre côté, il est possible d'imaginer que les mêmes techniques puissent permettre un regroupement automatique des voix d'après leurs caractéristiques principales et leur niveau global, et même de pointer les phonèmes déviants que l'on pourrait confronter à l'analyse par phonetic labelling.

Annexes :

Annexes 1 : analyse par phonetic labelling des 7 sujets de notre étude sur la mise en contexte de raucité

- Annexe 1a : en position accentuée : « C'est dans un bol que je préfère le boire. »
- Annexe 1b : en position inaccentuée : « Je préfère les jolis bols rouges. »

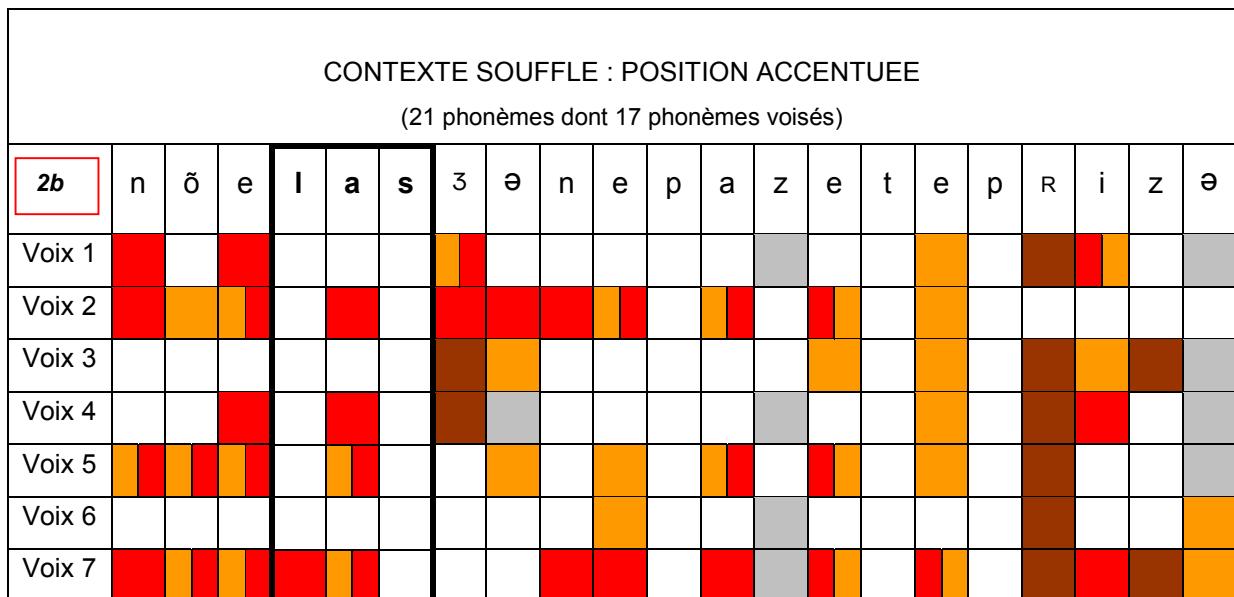
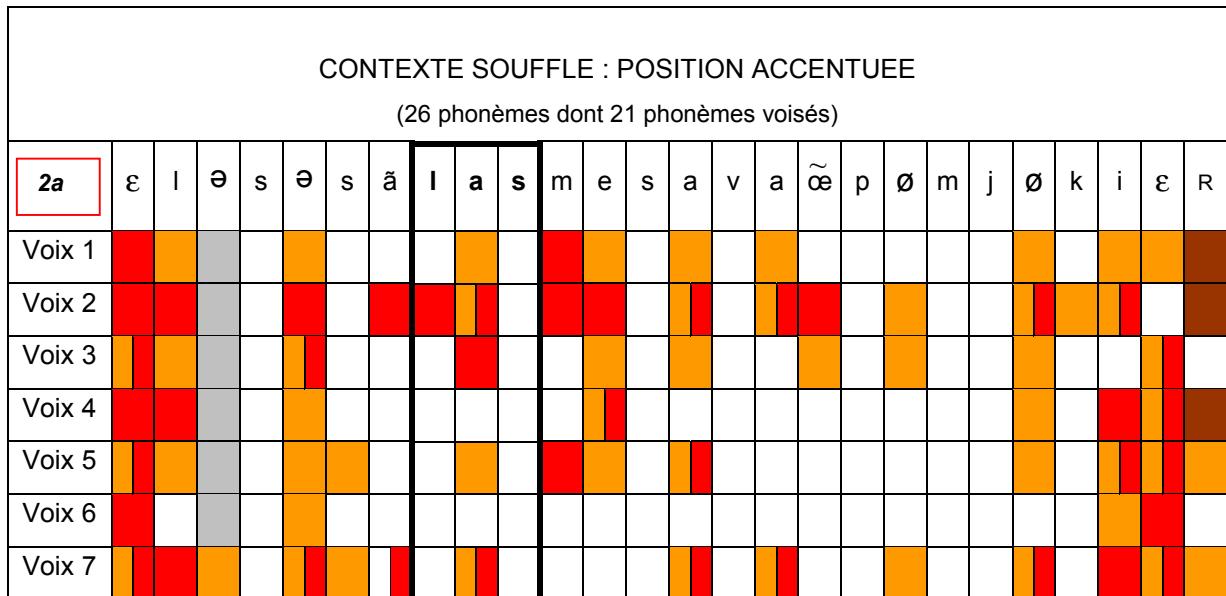
Normal	
Raucité	Red
Souffle	Orange
Dévoisement	Brown
Omission	Grey



Annexes 2 : analyse par phonetic labelling des 7 sujets de notre étude sur la mise en contexte de souffle

- Annexe 2a : en position accentuée : « Elle se sent lasse, mais ça va un peu mieux qu'hier. »
- Annexe 2b : en position inaccentuée : « Non hélas, je n'ai pas été prise. »

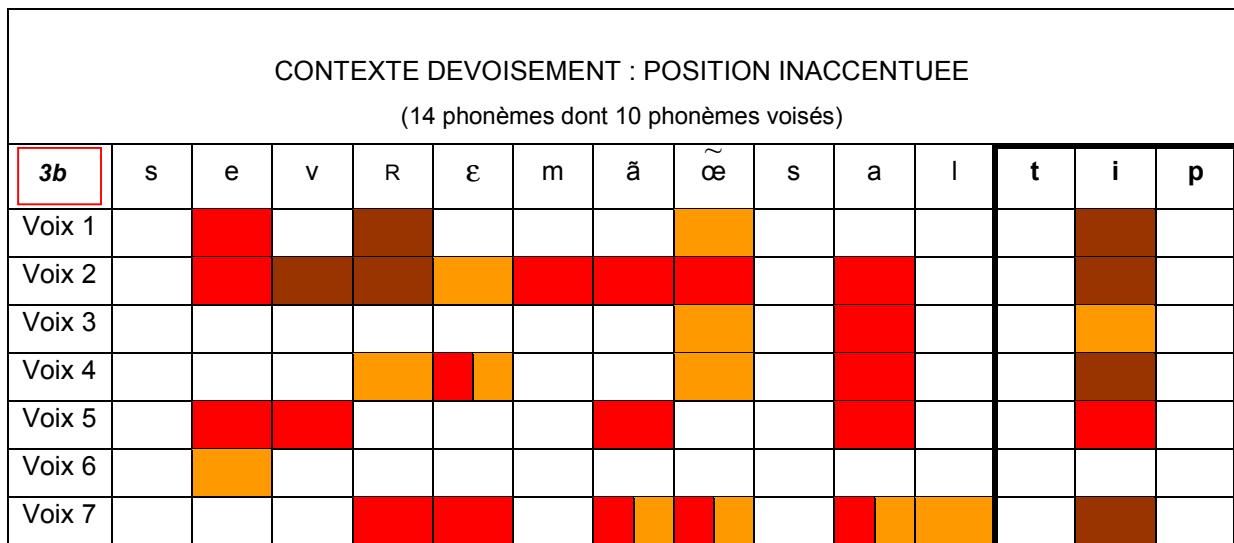
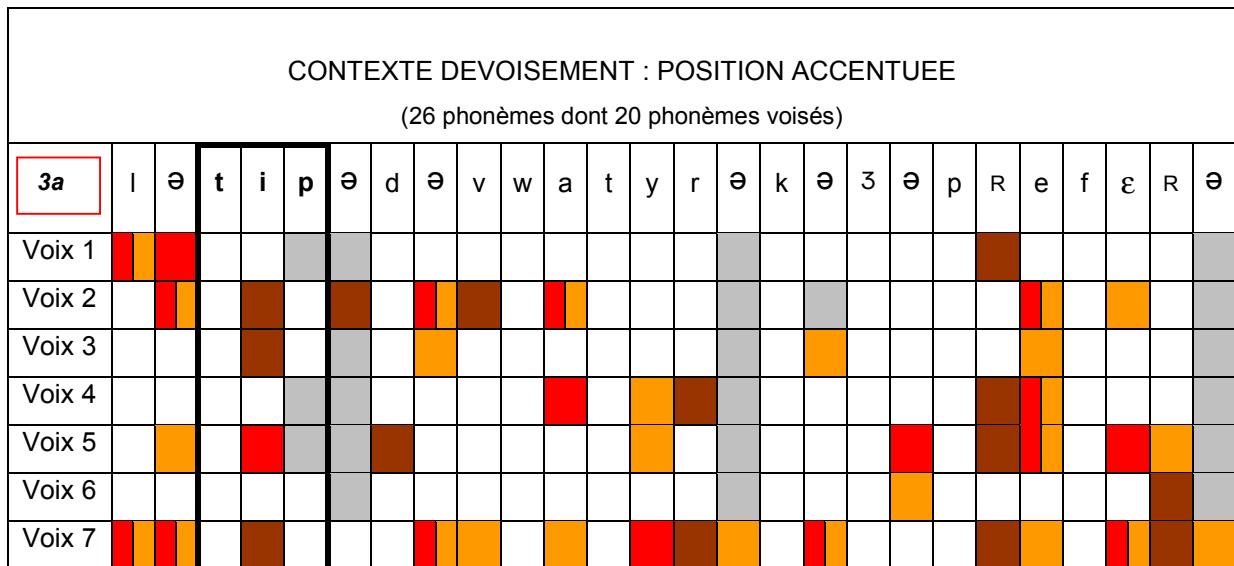
Normal	
Raucité	Red
Souffle	Orange
Dévoisement	Brown
Omission	Grey



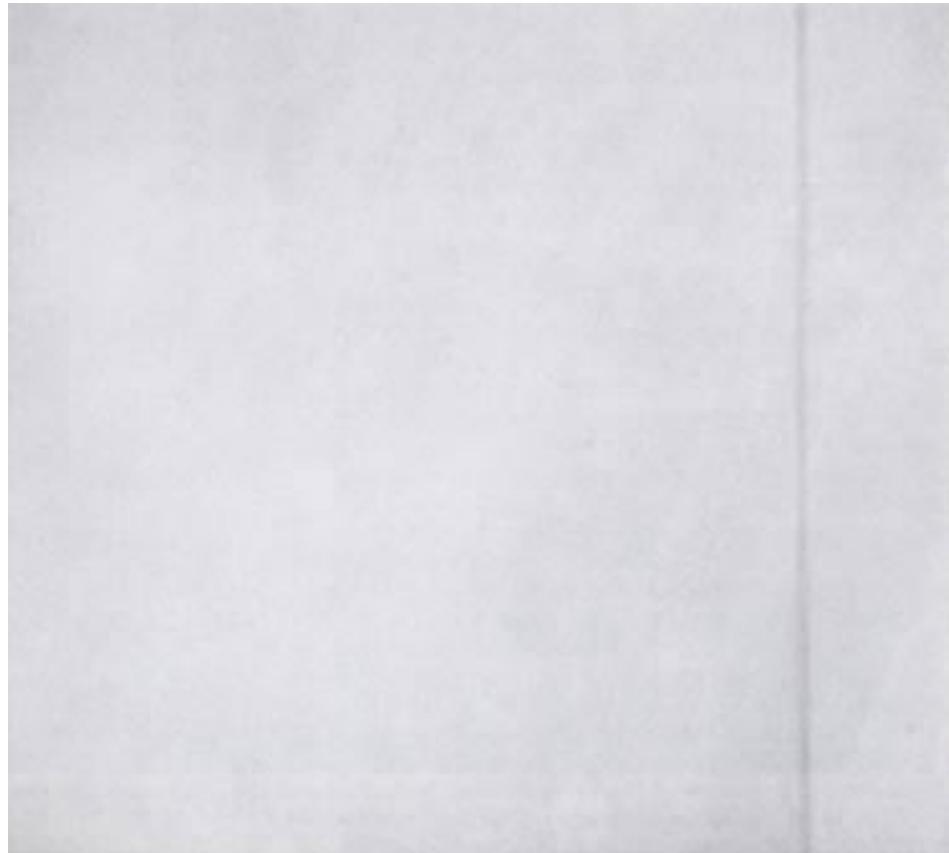
Annexes 3 : analyse par phonetic labelling des 7 sujets de notre étude sur la mise en contexte de dévoisement

- Annexe 3a : en position accentuée : « Le type de voitures que je préfère, ... »
- Annexe 3b : en position inaccentuée : « C'est vraiment un sale type ! »

Normal	
Raucité	Red
Souffle	Yellow
Dévoisement	Brown
Omission	Grey



« *The Voice* » Barnett Newman -1950. New York, Museum of Modern Art



Conclusion générale

Conclusion générale

L'axe de recherche dédié à l'analyse perceptive des dysphonies a été développé dans notre laboratoire dès 1996 avec pour objectif premier d'optimiser la fiabilité des jugements effectués par le jury d'écoute. En effet, l'analyse perceptive étant le standard de validation des mesures instrumentales, il était nécessaire de mettre au point une méthodologie permettant d'améliorer l'adéquation entre ces deux méthodes de mesures.

Notre premier travail, réalisé en 1996, avait pour but de vérifier les capacités cliniques de l'appareillage EVA® par la confrontation des mesures obtenues chez 27 patients avant et après microchirurgie laryngée, avec l'évaluation du changement de qualité vocale par un jury d'experts¹. Les résultats obtenus par l'analyse perceptive montraient des variations si importantes, que la comparaison avec les données instrumentales n'a pu être établie. Néanmoins, ce travail nous a permis d'alimenter notre réflexion concernant l'analyse perceptive des dysphonies et les résultats ambigus que nous avons observés ont servi de base à toute notre démarche scientifique ultérieure. L'analyse des cas litigieux nous a orientés vers trois biais expliquant la variabilité du jury :

- le recrutement du jury d'écoute : le niveau d'expérience des auditeurs était hétérogène
- la méthodologie : l'échelle de mesure quantitative était comprise et utilisée de manière différente par les auditeurs
- le matériau phonétique : les mesures instrumentales étaient effectuées sur un /a/ tenu alors que les auditeurs évaluaient les voix sur un texte lu

La méthodologie de ce travail a été remaniée tenant compte de ces réflexions et l'étude a été publiée en 1999².

Mais dès lors, nos travaux se sont orientés vers l'exploration systématique de trois grands axes :

- Axe Perception/Jury : comment améliorer la fiabilité du jury d'écoute ?

Dans un travail actuellement soumis à publication dans la Revue de Laryngologie, nous proposons une méthodologie de formation des auditeurs à l'analyse perceptive grâce à un paradigme

¹ Giovanni A, Revis J, Triglia JM. « Etude préliminaire de la pertinence des mesures acoustiques et aérodynamiques pour l'évaluation de l'évolution de la voix après microchirurgie laryngée : à propos de 27 cas. » Non Publié.

² Giovanni A, Revis J, Triglia JM. « Objective aerodynamic and acoustic measurement of voice improvement after phonosurgery. » Laryngoscope. 1999 Apr;109(4):656-60.

³ Revis J, Wipf AL, Wuyts FL, Giovanni A. « Elaboration d'un protocole d'entraînement à l'analyse perceptive pour un jury inexpérimenté. » Soumis à Rev Otol Rhinol Laryngol.

« stimulus/réponse/feedback/stimulus »³. Les résultats montrent que la reproductibilité des évaluations des auditeurs novices est améliorée par notre méthode pour se rapprocher des performances du jury expérimenté.

Dans un objectif plus clinique, un autre travail s'est attaché à la conception et la validation d'un logiciel d'aide à l'analyse perceptive (CLAP) destiné aux différents praticiens intervenant dans la prise en charge des patients dysphoniques⁴. Ce logiciel permet de standardiser les évaluations et la méthodologie proposée tient compte des avancées récentes dans ce domaine.

- Axe Perception/Mesures : comment favoriser l'adéquation entre les mesures instrumentales et le jury d'écoute ?

En 2001, le Dr Yu proposait de comparer un protocole de 8 mesures avec l'analyse perceptive réalisée par un auditeur chez 308 femmes. Les résultats montraient 58% de variance commune. Ce mauvais résultat était susceptible de provenir de la méthodologie de l'analyse perceptive, puisque celle-ci était réalisée par un seul auditeur, pas systématiquement le même, au cours de l'examen. Dans le travail suivant⁵, nous avons conçu et appliqué un protocole perceptif allant dans le sens d'une expertise selon la méthode du consensus entre 3 auditores expérimentés. Les résultats ont montré 86% de concordance.

Notre collaboration avec le Dr Yu s'est poursuivie et notre réflexion commune nous a conduits à proposer une nouvelle échelle de mesure quantitative de la perception de la dysphonie⁶. Les résultats ont montré que l'utilisation de l'échelle visuelle analogique modifiée permettait d'atteindre 88% d'adéquation avec les mesures instrumentales.

- Axe Perception/Phonétique : quel est l'impact du matériau phonétique sur la perception de la dysphonie ?

Cet axe constitue le cœur de notre réflexion, basé sur les interactions entre ce qui est produit et ce qui est perçu. L'analyse de la dysphonie est couramment effectuée sur un /a/ tenu, et même la partie stable du /a/ tenu (mesures instrumentales) ou sur de la parole (perception). La quantité d'informations phonétiques est très différente selon ces matériaux : le /a/ tenu contient les transitoires d'attaque et de

⁴ Revis J, Nazarian B, Bingenheimer S, Giovanni A. « Logiciel de CLassement de l'Analyse Perceptive (CLAP) : Validité clinique et influence sur la variabilité du jury . » Soumis à Glossa.

⁵ 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;15(4):529-542.

⁶ Yu P, Revis J, Wuyts FL, Zanaret M, Giovanni A. « Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale. » Folia Phoniatr Logop. 2002;54(6):274-281.

sortie ainsi qu'une partie de vibration stable ; la partie stable du /a/ tenu ne contient que cette partie de vibration stable, alors que la parole contient toutes les informations phonétiques (transitoires, vibration, prosodie, rythme, coarticulation, transitions formantiques, etc.) Notre premier travail publié sur ce thème⁷ a montré que le jury d'écoute était influencé par le matériau phonétique puisque les évaluations portées sur la partie stable du /a/ tenu tendaient à sous-évaluer la dysphonie du fait de l'absence des informations contenues dans l'attaque.

L'étude suivante⁸ nous a permis de caractériser l'influence de l'attaque qui semble jouer un rôle essentiel dans la discrimination des dysphonies légères et moyennes.

Ces travaux nous ont conduits à proposer une mesure temporelle de l'attaque⁹ afin d'en délimiter les contours pour appliquer, à terme, les mesures non-linéaires. Les résultats ont montré que la méthode des « 10%-90% » permettait d'isoler l'attaque, avec un niveau de performance optimal pour les dysphonies légères et moyennes.

De la même façon que la manifestation dysphonique est différente en fonction du matériau étudié, il faut considérer qu'elle n'est pas constante au cours de la parole. La dysphonie n'est pas un bruit constant dans le signal vocal, mais intervient au cours de la parole sous forme d'occurrences pathologiques, en fonction des contextes phonétiques et linguistiques. Dans notre travail suivant¹⁰, nous avons proposé une nouvelle technique d'écoute tenant compte de cette réflexion. Le principe du phonetic labelling est d'écouter et de caractériser chaque phonème d'une phrase en normal/rauque/soufflé/dévoisé. Les résultats ont montré que le nombre d'occurrences pathologiques est fortement corrélé à l'impression dysphonique globale ($r^2=0.87$) et que leur localisation dans la phrase dépend du contexte phonétique (voyelle fermée entre 2 occlusives non-voisées) et linguistique (intonation et accentuation dans la phrase). Une étude menée avec la participation de 20 auditeurs nous a permis de confirmer la validité de la méthode de phonetic labelling¹¹.

La conception de cette méthode a permis d'ouvrir de nouvelles voies dans l'étude de la dysphonie et notamment d'évaluer l'influence de l'attaque au cours de la parole par l'analyse des occurrences pathologiques intervenant dans une phrase ne contenant qu'une seule attaque (Amélie a mis l'eau au moulin) comparée à une phrase contenant de nombreuses attaques (Ton thé t'a-t-il oté ta toux)¹². Ce

⁷ Revis J, Giovanni A, Wuyts FL, Triglia JM. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop*, 1999;51(3):108-116

⁸ Revis J, Giovanni A, Triglia JM. Influence de l'attaque sur l'analyse perceptive des dysphonies. *Folia Phoniatr Logop*. 2002;54:19-25

⁹ Revis J, Barberis S, Giovanni A. Definition of a new temporal voice onset measurement. *Rev Laryngol Otol Rhinol (Bord)*. 2000;121(5):291-6.

¹⁰ Revis J, Nguyen N, Giovanni A. Phonetic labelling of dysphonia : a new perspective in perceptual analysis. Soumis à *J Voice*.

¹¹ Revis J, Gannat M, Giovanni A. Reproductibilité et validation du phonetic labelling. Soumis à *Folia Phoniatr Logop*.

¹² Revis J, Ollivier R, Giovanni A. Influence de l'attaque au cours de la parole. Soumis à *Folia phoniatr Logop*.

travail a montré que l'impression globale de dysphonie était plus importante sur la phrase « attaques » que sur la phrase « stable ». Le phonetic labelling a permis de confirmer que le nombre d'occurrences pathologiques était plus élevé sur la phrase « attaques » et qu'elles étaient effectivement situées sur les voyelles précédées d'un arrêt vibratoire.

Enfin, nous avons également proposé une étude spécifique des contextes phonétiques et linguistiques d'apparition des occurrences pathologiques¹³. Les résultats ont montré qu'il était possible de prédire l'apparition d'occurrences spécifiques (raucité ou souffle ou dévoisement) par la conception de phrases standardisées et calibrées selon les contextes favorisant.

Les travaux basés sur le phonetic labelling sont poursuivis et devraient permettre de proposer une nouvelle méthodologie d'évaluation vocale. Sur le modèle du « test d'effort », l'élaboration de phrases-type mettant la vibration des cordes vocales en difficulté, pourrait permettre aux praticien d'évaluer et de caractériser plus précisément la dysphonie des patients. A terme, ces travaux pourraient faire l'objet d'une publication sous la forme d'un bilan d'investigation destiné spécifiquement à la pratique clinique.

L'avancée de nos connaissances de l'influence phonétique sur la manifestation dysphonique a permis d'ouvrir de nouveaux axes de recherche. Une nouvelle thématique est en cours de développement en collaboration avec le Laboratoire Informatique d'Avignon (LIA) et le Laboratoire Parole et Langage (LPL) d'Aix en Provence¹⁴. Notre objectif commun est d'adapter les techniques de reconnaissance du locuteur à la pathologie vocale dans le but final de proposer un programme de reconnaissance automatique de la dysphonie. A l'heure actuelle, des pré-tests sont en cours à partir d'un corpus d'apprentissage de 60 voix dysphoniques (20 dysphonies légères, 20 moyennes et 20 sévères) et 20 voix normales expertisées par un jury d'écoute professionnel. A l'issue de cette session d'apprentissage, un corpus d'évaluation composé de voix dysphoniques et saines permettra de tester les capacités de discrimination du programme. Les voix devraient être classées en 4 groupes selon leurs ressemblances avec les échantillons du corpus d'apprentissage. Les résultats seront confrontés à une expertise par notre jury d'écoute.

Plusieurs perspectives sont envisagées afin de poursuivre ces pré-tests :

- le repérage automatique des phonèmes déviants sera confronté à une analyse par phonetic labelling des occurrences pathologique.

¹³ Revis J, Gachet F, Giovanni A. Influence du contexte phonétique sur la perception de la dysphonie. Soumis à Phonetica.

¹⁴ Revis J, Briffa C, Bonastre JF, Giovanni A. Caractérisation automatique de la dysphonie : application de méthodes issues de la reconnaissance automatique du locuteur. En cours.

- la caractérisation automatique des dimensions qualitatives de la dysphonie permettra de créer des sous groupes en fonction du type d'altération du timbre des patients (raucité, souffle, asthénie, forçage...)
- une analyse rétrospective des résultats obtenus par ces méthodes innovantes pourra permettre de reconsiderer les méthodologies classiques d'analyse de la dysphonie : en effet, notre hypothèse est que les dimensions qualitatives et quantitatives utilisées actuellement ne sont peut être pas les plus pertinentes et que les résultats obtenus en reconnaissance automatique permettront de mieux caractériser la dysphonie.

« Le véritable esprit scientifique c'est de ne jamais admettre pour vrai quelque chose dont on n'a pas pu démontrer qu'il l'était, mais de ne jamais rejeter quelque chose dont on n'a pas pu démontrer qu'il était faux. »

Paul-Emile Victor

Bibliographie

Bibliographie

1. Alfonsi-Guarella MD. Paramètres acoustiques et aérodynamiques de la voix normale et pathologique chez la femme. DEA Langage et Parole, Université de Provence, 1998
2. Anders L, Hollien H, Hurme P, Sonninen A, Wendler J. Perceptual evaluation of hoarseness by several classes of listeners. *Folia Phoniatr*, 1988;40:91-100
3. Anderson V. Training the speaking voice. New York : Oxford University Press, 1981
4. Askenfeld A, Hammarberg B. Speech waveform perturbation analysis : a perceptual acoustical comparison of seven measures. *J Speech Hear Res*, 1986;29:50-64
5. Bassich CJ, Ludlow CL. The use of perceptual methods by new clinicians for assessing voice quality. *J Speech Hear Dis*, 1986;51:125-133
6. Beckham ME, De Jong K, Sun-Ah J, Sook Hyang L. The interaction of coarticulation and prosody in sound change. *Lang Speech*, 1992;35:45-58
7. Beckman ME, Jung TP. Variability in the production of quantal vowels revisited. *J Acoust Soc Am*, 1995;97:471-490
8. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice selective areas in human auditory cortex. *Nature*, 2000;403:309-311
9. Baken RJ. Clinical measurement of speech and voice. London : Taylor and Francis, 1987.
10. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res*, 1996;39:126-134
11. Black JW. A study of voice merit. *Quarterly Journal of Speech*, 1942;28:67-74
12. Boisacq-Schepens N, Crommelinck M. Neurosciences. Paris : Dunod, 2000
13. Bouchayer M, Cornut G. Microsurgical treatment of benign vocal fold lesions : indications, technique, results. *Folia Phoniatr*, 1992;44:155-184
14. Chan K, Yiu E. The effect of anchor and training on the reliability of perceptual voice evaluation. *J Speech Hear Lang Res*, 2002;45:116-126
15. Chardin D. Que sais-je ? Les surdités. Paris : Presses Universitaires de France, 1982
16. Chardin D. Que sais-je ? L'oto-rhino-laryngologie. Paris : Presses Universitaires de France, 1986
17. Childers D, Lee C. Vocal quality factors : analysis, synthesis, and perception. *J Acoust Soc Am*, 1991;37:2394-2410
18. Cooke A, Ludlow CL, Hallett N, Scott-Selbie W. Characteristics of vocal fold adduction related to voice onset. *J Voice*, 1997;1:12-22
19. Cordes A. The reliability of observational data : I. theories and methods for speech-language pathology. *J Speech Hear Res*, 1994;37:264-278
20. Cowan M. Pitch and intensity characteristics of stage speech. *Arch Speech*, 1936:1-92
21. Crevier-Buchman L, Laccourreyre O. Speech evaluation after suprarecricoid partial laryngectomy : preliminary results. *Actes du Congrès Larynx 97*, Marseille France, 1997;1:19-20
22. Crevier-buchman L, Vaissière J, Maeda S. Etude de l'intelligibilité des consonnes du français après laryngectomie partielle supra-cricoïdienne. *Rev Otol Rhinol Laryngol*, 2002 ;123(5) :307-310

23. De Bodt M, Van de Heining, Wuyts FL, Lambrechts L. The perceptual evaluation of voice disorders. *Acta Otolaryngol Belg*, 1996;50:283-291
24. De Bodt M, Wuyts F, Van de Heyning P, Croux C. Test-retest study of the GRBAS scale : influence of experience and professional background on perceptual ratings of voice quality. *J Voice*, 1997;1:74-80
25. Decoster W, De Bruyne F. Changes in spectral measures and voice onset time with age : a cross sectionnal and longitudinal study. *Folia Phoniatr Logop*, 1997;49:269-280
26. Dejonckere P. Bruit de turbulence et apéridodicité dans la voix pathologique. Une approche multifactorielle. *Rev Laryngol*, 1990;111:353-357
27. Dejonckere P, Obbens C, De Moor GM, Wieneke GH. Perceptual evaluation of dysphonia : reliability and relevance. *Folia Phoniatr*, 1993;45:76-83
28. Dejonckere P, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-buchman L, Millet B. Differentiated perceptual evaluation of pathological voice quality : reliability and correlations with acoustic measurements. *Rev Laryngol Otol Rhinol*, 1996;117(3):219-224
29. Dejonckere P, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-Buchman L, Millet B. Reliability and clinical relevance of perceptual evaluation of pathological voices. *Rev Laryngol Otol Rhinol*, 1998;4:247-248
30. Dejonckere P, Bradiey P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van de Heyning, 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. Guideline elaborated by the Committee on Phoniatics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*, 2001;258:77-82
31. De Jong K, Beckham ME, Edwards J. The interplay between the prosodic structure and coarticulation. *Lang Speech*, 1993;36:197-212
32. De Krom G. Consistency and reliability of voice quality ratings for different types of speech fragments. *J Speech Hear Res*, 1994;37:985-1000
33. De Krom G. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Hear Res*, 1995;38:794-811
34. Feldmeyer JJ. Cerveau et pensée : la conquête des neurosciences. Genève : Georg Editeur, 2002
35. Fujimoto M, Murano E, Niimi S, Kiritani S. Differences in glottal opening patern between Tokyo and Osaka dialect speakers : factors contributing to vowel devoicing. *Folia Phoniatr Logop*, 2002;54:133-143
36. Gazzaniga MS, Ivry RB, Mangun GR. Neurosciences cognitives : la biologie de l'esprit. Bruxelles : De Boeck Université, 2001
37. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. *J Speech Hear Res*, 1993;36:14-20
38. Giovanni A, Molines V, Nguyen N, Teston B, Robert D, Cannoni M, Pech A. Une méthode multiparamétrique d'évaluation vocale objective assistée par ordinateur. *Ann Otolaryng Chir Cervicofac*, 1992;109:200-206
39. Giovanni A. Analyse objective des dysphonies avec l'appareillage EVA : études préliminaires à la définition d'un protocole normalisé. DEA Langage et Parole, Université de Provence, 1995
40. Giovanni A, Robert D, Teston B, Guarella MD, Zanaret M. Etude préliminaire des paramètres acoustiques et aérodynamiques après laryngectomie frontale antérieure de Tucker. *Ann Otolaryngol Chir Cervicofac*, 1996;113:277-284

41. 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 measurements. *Folia Phoniatri Logop*, 1996;48:175-185
42. Giovanni A. Contribution à l'élaboration d'un protocole multiparamétrique d'évaluation objective des dysphonies. Thèse de Science. Université de la Méditerranée, 1998.
43. Giovanni A, Revis J, Triglia JM. Objective aerodynamic and acoustic measurement of voice improvement after phonomicrosurgery. *Laryngoscope*, 1999;109(4):656-60
44. Giovanni A, Vallicioni JM, Gras R, Zanaret M. Clinical experience with Gore-Tex for vocal fold medialization. *Laryngoscope*, 1999;109-2:284-288
45. Giovanni A, Ouaknine M, Triglia JM. Determination of the largest Lyapunov exponents of vocal signal. Application to unilateral laryngeal paralysis. *J Voice*, 1999;13:341-354
46. Giovanni A, Ouaknine M, Guelfucci B, Yu P, Triglia JM, Zanaret M. Non-linear behavior of vocal fold vibration : the role of coupling between the vocal folds. *J Voice*, 1999;13:456-457
47. Giovanni A, Heim C, Demolin D, Triglia JM. Estimated subglottic pressure in normal and dysphonic subjects. *Ann Otol Rhinol Laryngol*, 2000;109:500-504
48. Golb C, Chasaide AN. Acoustic characteristics of voice quality. *Speech Com*, 1992;481-490
49. Golb C, Chasaide AN. Contextual variation of the vowel voice source as function as adjacent consonants. *Lang Speech*, 1993;36:303-330
50. Goldstone RL. Perceptual learning. *Ann Rev Psychol*, 1998 ;49:585-612
51. Hammarberg B, Tritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*, 1980;90:441-451
52. Hammarberg B, Fritzell B, Gauffin J, Sundberg J. Acoustic and perceptual analysis of vocal dysfunction. *Journal of phonetics*, 1986;14:533-547
53. Hammarberg B. Perceptual evaluation of dysphonic voice. *Actes du Congrès Larynx 97*, Marseille France, 1997;1:11-18
54. Herzel H, Berry D, Titze I, Saleh M. Analysis of vocal disorders with methods from nonlinear dynamics. *J Speech Hear Res*, 1994;37:1008-1019
55. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*, 1994;37:769-778
56. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality : dysphonic voices and continuous speech. *J Speech Hear Res*, 1996;39:311-321
57. Hirano M. Psycho-acoustic evaluation of voice : GRBAS Scale for evaluating the hoarse voice. *Clinical examination of voice*. Wien : Springer Verlag, 1981
58. Hirano M, Hibi S, Yoshida T, Hirade Y, Kasuya H, Kikuchi Y. Acoustic analysis of pathological voice. Some results of clinical application. *Acta Otolaryngol*, 1988;105:432-438
59. Hoffman WG. Public speaking today. New York : Mac Graw-Hill Co, 1940
60. Hollien H. That golden voice : talent or training ? *J Voice*, 1993;7:195-205
61. Isshiki N, Okamura H, Tanabe M, Morimoto M. Differential diagnosis of hoarseness. *Folia Phoniatri*, 1969;21:9-12
62. Kant E. (1790). Critique de la faculté de juger. Paris : Librairie Philosophique J Vrin, 2000
63. Kempster G, Kistler DJ, Hillenbrand J. Multidimensional scaling analysis of dysphonia in two speaker groups. *J Speech Hear Res*, 1991;34:534-543
64. Kent RD, Kent JF, Rosenbek JC. Maximum performance tests of speech production. *J Speech Hear Disord*, 1987;52:367-387

65. Koike Y. Application of acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonol*, 1973;7:17-23
66. Köster O, Marx B, Gemmar P, Hess MM, Künzel HJ. Qualitative and quantitative analysis of voice onset by means of a multidimensional voice analysis system (MVAS) using high-speed imaging. *J Voice*, 1999;3:355-374
67. Kreiman J, Papcun G. Comparing discrimination and recognition of unfamiliar voices. *J Speech Hear Res*, 1991;10:265-275
68. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Hear Res*, 1992;35:512-520
69. Kreiman J, Gerratt 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;36:21-40
70. Kreiman J, Gerratt BR. The multidimensional nature of pathologic vocal quality. *J Acoust Soc Am*, 1994;96:1291-1302
71. Kreiman J, Gerratt B. Measuring vocal quality : validity of rating scale measures of voice quality. *J Acoust Soc Am*, 1998;104(3):1598-1608
72. Landercy A, Renard R. *Eléments de phonétique*. Bruxelles : Didier, 1977
73. Laver J. *The phonetic description of voice quality*. Cambridge : Cambridge University Press, 1980
74. Laver J, Wirz S, Mackenzie J, Hiller SM. A perceptual protocol for the analysis of vocal profiles. Edinburgh University Department of Linguistics Work in Progress, 1981;14:139-155
75. Laver J, Hanson RJ. Describing the normal voice. In Darby J (eds). *Speech evaluation in psychiatry*. New York : Grune and Stratton, 1981
76. Laver J, Hanson RJ. Describing the normal voice. In Laver J (eds). *The gift of speech : readings in the analysis of speech and voice*. Edinburgh : Edinburgh University Press, 1991
77. Lucero, GC. Optimal glottal configuration for ease phonation. *J Voice*, 1998;12:151-158
78. Lynch GE. Phonophotographic study of trained and untrained voices reading factual and dramatical material. *Arch Speech*, 1934;1:9-25
79. Mac Adams S, Bigand E. Introduction à la cognition auditive. In Mac Adams S, Bigand E (eds). *Penser les sons : psychologie cognitive de l'audition*. Paris : Presses Universitaires de France, 1994
80. Malmberg B. *Phonétique française*. Berlin : Liber Läromedel, 1976
81. Martin D, Fitch J, Wolfe V. Pathologic voice type and the acoustic prediction of severity. *J Speech Hear Res*, 1995;38:765-771
82. Martin D, Wolfe V. Effects of perceptual training based upon synthesized voice signals. *Perceptual and Motor Skills*, 1996;83:1291-1298
83. Merleau-Ponty M (1945). *Phénoménologie de la perception*. Paris : Gallimard (tel), 2001
84. Millet B, Dejonckere PH. What determines the differences in perceptual rating of dysphonia between experienced raters? *Folia Phoniatr Logop*, 1998;50:305-310
85. Munoz J, Mendoza E, Fresneda MD, Carballo G, Lopez P. Acoustic and perceptual indicators of normal and pathological voice. *Folia Phoniatr Logop*, 2003;55(2):102-114
86. Murray E, Tiffen J. An analysis of some basic aspects of effective speech. *Arch Speech*, 1934;1:61-83
87. Nicolas S. *Les origines de la psychologie cognitive*. In Roulin JL (eds). Rosny : Bréal, 1998

88. Nicolas S, Doré-Mazars K. Eléments d'histoire sur la perception et les liens perception-action. In Coello Y, Honoré J (eds). *Percevoir, s'orienter et agir dans l'espace : approche pluridisciplinaire des relations perception-actio*. Marseille : Solal, 2002
89. Oates J, Russel A. Learning voice analysis using an interactive multi-media package : development and preliminary evaluation. *J Voice*, 1998;12:500-512
90. Ohala J. Coarticulation and phonology. Edda (eds) *Coarticulation*, 1993
91. Papcun G, Kreiman J, Davis A. Long-term memory for unfamiliar voices. *J Acoust Soc Am*, 1988;85:913-925
92. Parret H. *La voix et son temps*. Bruxelles : De Boeck Université, 2002
93. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice : sustained vowels versus continuous speech. *J Speech Lang Hear Res*, 2001;44:327-339
94. Pronovost W. Experimental study of habitual and natural speech levels of superior speakers. *Speech Monographs*, 1943;9:111-123
95. Rabinov R, Kreiman J, Gerratt BR, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J Speech Hear Res*, 1995;38:26-32
96. Revis J, Giovanni A, Wuyts F, Triglia J. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop*, 1999;51(3):108-16
97. Revis J, Barberis S, Giovanni A. Definition of a new temporal voice onset measurement. *Rev Laryngol Otol Rhinol (Bord)*, 2000;121(5):291-6
98. Revis J, Giovanni A, Triglia JM. Influence de l'attaque sur l'analyse perceptive des dysphonies. *Folia Phoniatr Logop*, 2002;54:19-25
99. Russ J. *Panorama des idées philosophiques : de Platon aux contemporains*. Paris : Armand Colin, 2000
100. Schoentgen J. Jitter in sustained vowels and isolated sentences produced by dysphonic speakers. *Speech Com*, 1989;8:61-79
101. Sherman D, Linke E. The influence of certain vowel types on degree of harsh voice quality. *J Speech Hear Dis*, 1954:401-408
102. Snidecor JS. The pitch and durational characteristics of superior female speakers, during oral reading. *J Speech Hear Dis*, 1952;16:44-52
103. Stevens KN. *Acoustic phonetics*. Cambridge, London : The MIT Press, 1998
104. Talley CH. A comparison of conversational and audience speech. *Arch Speech*, 1937;2:28-40
105. Titze I. *Principles of voice production*. Iowa City : Prentice-Hall Inc, 1994
106. Titze IR, Wong D, Milder MA, Hensley SR, Ramig LO. Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. *J Speech Hear Res*, 1995;38:526-535
107. Vilkman E, Aaltonen O, Laine U, Raimo I. Intrinsec pitch of vowels : a complicated problem with an obvious solution ? In Gauffin and Hammarberg (eds). *Vocal fold physiology*, 1991
108. Wolfe V, Cornell R, Fitch J. Acoustic correlates of pathologic voice type. *J Speech Hear Res*, 1991;34:509-516
109. Wolfe V, Fitch J, Cornell R. Acoustic prediction of severity in commonly occurring voice problems. *J Speech Hear Res*, 1995;38:273-279
110. Wolfe V, Martin D, Pamer C. Perception of dysphonic voice quality by naive listeners. *J Speech Lang Hear Res*, 2000;43:697-705

111. Wuyts FL, De Bodt M, Molenberghs G, Bruckers L. Research work of the belgian study group on voice disorders : results. *Acta oto-rhino-laryngologica belg*, 1996;50:331-342
112. Wuyts F, Van de Heyning P, De Bodt M, Heylen L, Remacle M. Dysphonia severity index : a multiparametric approach for the assessment of voice quality. *Actes du Congrès Larynx 97*, Marseille France, 1997;1:73-76
113. 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;13:508-517
114. 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
115. Yamaguchi H, Shrivastav R, Andrews ML, Niimi S. A comparison of voice quality ratings made by Japanese and American listeners using GRBAS scale. *Folia Phoniatr Logop*, 2003;55:147-157
116. Yu P, Ouaknine M, Giovanni A. Intérêt clinique du calcul des coefficients de Lyapunov pour l'analyse objective des dysphonies. *Rev Laryngol Otol Rhinol*, 2000;5:301-305
117. 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;15(4):529-542
118. Yu P, Revis J, Wuyts FL, Zanaret M, Giovanni A. Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale. *Folia Phoniatr Logop*, 2002;54(6):274-281
