

Influence du contexte d'écoute dans la perception des dysphonies : effets de la connaissance de l'état thérapeutique du patient sur le jugement de sa qualité vocale ¹

Context influence on the perception of dysphonia:
When the knowledge of the patient's clinical state can
modify the results of perceptual voice quality assessment

Ghio A. ²
Merienne S. ³
Giovanni A. ^{2/3}
(Marseille)

Résumé

L'objectif de l'étude était d'évaluer dans quelle mesure les connaissances a priori qu'un thérapeute possède sur un locuteur dysphonique, notamment son état pré ou post traitement, peut influencer le jugement de la voix du patient. 53 patients dysphoniques ont été inclus dans l'étude et pour chaque locuteur, nous avons sélectionné une paire de voix enregistrées dans des circonstances différentes. Sept auditeurs orthophonistes, ORL ou phoniatre ont été soumis en aveugle à l'écoute de ces paires de voix et devait fournir un jugement comparatif. Quelques semaines plus tard, les mêmes auditeurs ont subi le test à l'identique sauf que dans cette deuxième session, une information sur le statut des voix du locuteur était donnée : pré ou post, chirurgie ou rééducation. Nous avons équilibré cette information de façon soit à renforcer le jugement porté au préalable (ex : voix préférée en aveugle indiquée comme post thérapeutique en contexte), soit à contrarier le jugement (ex : voix préférée en aveugle indiquée comme pré thérapeutique en contexte). Dans l'écoute influencée renforcée, la préférence est amplifiée de façon significative. Dans l'écoute influencée contrariée, le changement de décision est important. Nous observons des inversions de préférences et la note avec information contradictoire est presque indépendante de la note obtenue en aveugle. Les résultats convergent vers l'extrême nécessité d'utiliser un protocole d'évaluation en aveugle pour réaliser un bilan perceptif d'une dysphonie.

Mots-clés : Dysphonie, évaluation perceptive de la voix, perception, processus descendant montant.

Summary

Two experiments were conducted to examine how knowledge of the patient's clinical state affects the results of perceptual voice quality assessment performed by specialists in voice therapy. This study involved 53 patients with a dysphonia. For each speaker, we selected a pair of recordings made in different circumstances. These pairs of voices were presented to seven listeners (ENT surgeons or speech therapists). The task was to perceptually compare the severity of the dysphonia between the 2 recordings of the pair. Stimuli were presented first in a blind test, then several weeks later with accompanying information about the patient; in particular, whether the voice was pre- or post- treatment was explicitly specified. We balanced this artificial contextual information in order to (α) reinforce the blind judgment (for example, voices perceived as better in the blind test were indicated as post treatment); (β) be inconsistent (in a clinical point of view) compared to the blind test (for example, voices perceived as more disordered during the blind test were indicated as post treatment). Results revealed that in the clinical-consistent context α , the preference was amplified in a significant way. In clinical-inconsistent condition β , we observed an inhibition effect or a change of decision. In this condition, the judgment was more dependant on the contextual information (pre/post treatment) than on the auditory sensation obtained in blind condition. These findings are discussed with reference to results in the literature on visual, olfactory or auditory perception in context. In the frame of perceptual voice assessment, results revealed that only blind tests can provide reliable results.

Key-words: Dysphonia, perceptual voice assessment, constructive perception, top-down bottom-up processes.

1. Communication présentée au LXVI^{ème} Congrès, Société Française de Phoniatrie et Pathologies de la Communication, Paris, 18/10/10.
2. Laboratoire "Parole et Langage" (UMR6057), Université de Provence & CNRS, 29, Avenue Robert Schuman, 13621 Aix-en-Provence cedex 1, France. E-mail: alain.ghio@lpl-aix.fr
3. CHU La Timone, Fédération d'ORL et de Chirurgie Cervico-Faciale, 264 rue St Pierre, 13385 Marseille cedex 05, France.

Article reçu : 04/04/11

accepté : 04/05/11

INTRODUCTION

Evaluer perceptivement la dysphonie : état des lieux

Dans le cadre de la prise en charge de la dysphonie, l'évaluation perceptive de la qualité de la voix est un

outil de dépistage, de diagnostic et d'évaluation des thérapies, qu'elles soient chirurgicales ou rééducatives. Elle vient en complément d'autres examens considérés comme un «tronc commun» pour l'évaluation de la plupart des dysphonies [1] : vidéo-strobo-endoscopie du larynx, analyses acoustiques, examens aérodynamiques, auto-évaluation... Classiquement, l'évaluation perceptive est menée à l'aide d'une échelle standardisée contenant plusieurs paramètres que l'auditeur doit juger à l'oreille. L'échelle GRBAS d'Hirano [2] est l'échelle la plus couramment utilisée. L'énoncé produit par le locuteur peut être une voyelle tenue ou de la lecture d'un texte standardisé, plus pertinente car plus représentative de la voix conversationnelle [3]. Ainsi, l'évaluation perceptive est considérée comme le «gold standard» par les spécialistes de la voix. Cependant, même si elle est très largement utilisée en pratique clinique et si une écoute attentive contribue de façon indéniable à dresser le tableau clinique complet du patient, son importance réelle dans un procédé fiable d'évaluation peut être discutée. En effet, de nombreuses études ont mis en évidence une variabilité certaine, observée dans le jugement porté par différents auditeurs sur une même voix ou pour un même auditeur entre diverses séances d'écoute [4-7]. Ce manque de fiabilité peut être expliqué par l'instabilité des référents internes des auditeurs [4], mais aussi par la sensibilité au contexte des mécanismes de perception de la parole.

L'influence du contexte d'écoute dans la perception de la dysphonie

Dans le cadre de la perception de la qualité vocale, Gerratt et al [8] ont montré que le contexte de présentation des voix peut entraîner une modification des jugements et constatent que le fait de présenter un petit nombre de stimuli moyennement rauques au milieu d'un grand nombre de stimuli «normaux» fait paraître les premiers très rauques. De même, Kreiman et al [5] ont constaté que l'ordre de présentation des stimuli influence l'auditeur. Ainsi, une voix moyennement dysphonique pourra être surévaluée comme sévèrement dysphonique si elle est présentée après un stimulus normal, ou sous-évaluée en légèrement perturbée si elle est écoutée après un stimulus très dégradé. Le moyen d'éviter ce biais est de présenter les échantillons de voix dans un ordre aléatoire différent à chaque session d'écoute.

Un autre biais lié au contexte est le classement des stimuli par groupes pathologiques, par âge ou par sexe, chaque catégorie présentant des caractéristiques particulières pouvant influencer l'auditeur [3].

Enfin, l'étude de Martens et al [9] est une bonne illustration de la sensibilité au contexte des mécanismes de perception de la voix. Dans ce travail, les auteurs montrent que la lecture du spectrogramme de la voix simultanément à son écoute augmente la fiabilité inter-auditeurs pour le jugement perceptif de la qualité vocale. Ces résultats illustrent le fait que la décision issue de la perception est en fait une décision complexe faisant appel non seulement au système perceptif «pur» mais aussi,

pour ce dernier travail, aux informations visuelles du spectrogramme et aux connaissances possédées par les auditeurs sur l'interprétation d'une telle représentation spectro-temporelle.

Ces phénomènes de variabilité de jugement, reflet d'un manque de fiabilité du procédé, sont ainsi déjà observés en conditions expérimentales contrôlées. Qu'en est-il en pratique clinique quotidienne ? Comment peut-on expliquer ces phénomènes ?

LES MÉCANISMES COMPLEXES DE LA PERCEPTION ET L'APPORT DE LA PSYCHOLOGIE COGNITIVE

La perception n'est pas un enregistrement direct de la réalité

La perception de la parole est un processus complexe dans lequel interviennent différentes sources d'information et de décision qui interagissent en réseaux. Il n'existe pas une correspondance parfaite entre la réalité sonore et ce que l'on en perçoit. Si cette équivalence existait, on parlerait de biunivocité, c'est-à-dire de relation bijective entre le monde physique sonore et la représentation perceptive que l'on en fait. Cela signifierait qu'à un objet sonore correspondrait systématiquement une sensation auditive, qu'à une sensation auditive correspondrait systématiquement un objet sonore, qu'à deux objets sonores distincts correspondraient deux sensations auditives distinctes et enfin qu'à deux sensations auditives différentes correspondraient deux sons aux caractéristiques distinctes. De nombreux résultats expérimentaux démontrent que ces propriétés ne sont pas respectées [10].

Pour Gaillard et al [11], «la perception de la réalité sonore n'est pas un enregistrement direct de la réalité. C'est une construction mentale opérée à la suite d'un traitement de l'information disponible, contrainte par nos sens ainsi que nos habitudes sélectives».

Ainsi, évaluer une voix à l'oreille consiste à interpréter à un moment précis le signal sonore qui nous est donné à entendre, avec le risque que le résultat de cette interprétation diffère avec celle d'un autre auditeur sous l'influence d'habitudes sélectives différentes ou diffère lors d'une évaluation ultérieure sous l'influence d'un changement de l'information disponible.

Processus top-down vs. bottom-up

Juger la qualité vocale d'un locuteur est au premier abord un processus de perception ascendant (bottom-up), c'est-à-dire qu'à partir de l'échantillon vocal, l'auditeur va catégoriser la voix par interprétation des indices acoustiques détectés perceptivement. Mais, comme dans tout autre processus de perception de la parole, il ne se réduit pas à ce simple trajet ascendant de l'acoustique vers le cognitif. Des processus descendants (top-down) interviennent et influencent la perception. En effet, lorsque nous entendons un énoncé dégradé, bruité ou phonétiquement appauvri, ces processus top-down

entrent en jeu pour restaurer ce qui est dégradé et optimiser l'intelligibilité du message. Ces mécanismes, qui s'appuient sur le contexte lexical, syntaxique, sémantique, pragmatique et autre, jouent un rôle important dans le décodage de la parole. Ainsi, cet énoncé émis par une personne enrhumée : «ce batin j'ai pris ba boiture pour aller au travail» sera restauré et compris par l'auditeur grâce au contexte lexical, sémantique et pragmatique. L'attention portée au message viendra maximiser ou minimiser les effets de ces processus de restauration. Nous abordons là des effets de restauration phonémique décrits, notamment, par Warren et al [12].

Dans le cadre de la perception de la qualité vocale, nous émettons l'hypothèse que de tels effets existent y compris sur d'autres aspects que le décodage phonémique et que le résultat du processus de perception peut être profondément orienté ou modifié par le contexte de l'écoute.

Le rôle de l'attention sélective dans la perception

Dans le domaine de la perception visuelle, divers travaux ont mis en évidence un certain nombre de caractéristiques pouvant être considérées comme des défauts dans le cas où la perception est utilisée comme un outil d'évaluation ou de décision. Simons et al [13] ont montré, par exemple, que nous pouvons être aveugles à certains éléments saillants et inattendus d'une scène visuelle lorsque notre attention est focalisée sur une autre tâche ou un autre objet de cette scène. Il définit cela par le terme de cécité inattentionnelle (inattentional blindness). Il en fait la démonstration grâce à un film d'une trentaine de secondes¹ dans lequel deux équipes de basketball aux maillots de couleurs différentes se passent le ballon. La consigne insistante donnée aux sujets qui visionnent ce film est de compter le nombre de passes effectuées par les joueurs en blanc. Ce qui intéresse les auteurs de l'expérience n'est pas la capacité des personnes à compter les passes, mais leur aptitude à remarquer un homme déguisé en gorille de couleur noire, qui traverse la scène en dansant. Plus de 50 % des participants, concentrés sur la consigne de comptage, focalisant leur attention sur les joueurs en blanc, ne le remarquent pas. L'attention sélective joue bien un rôle important dans la perception visuelle. La probabilité de détecter un objet inattendu ne dépendrait ni de sa proximité spatiale avec ce qui focalise l'attention, ni de l'habileté, plus ou moins grande du sujet à accomplir la tâche demandée au départ. Elle dépendrait de sa ressemblance avec les autres éléments de la scène, et du niveau de difficulté de la tâche à accomplir. Focaliser son attention sur une tâche précise aurait donc pour effet de rendre aveugle à ce qui peut nous détourner de notre objectif.

Un phénomène analogue défini sous le terme de cécité au changement (change blindness) a fait aussi

l'objet de nombreux travaux. Tout comme nous pouvons être aveugles à certains éléments inattendus d'une scène visuelle, nous pouvons être aussi aveugles aux changements intervenant dans cette scène, qu'elle soit dynamique ou figée [14]. La démonstration la plus connue de ce phénomène est une vidéo² dans laquelle un homme et une femme proposent un tour de cartes. Le sujet qui visionne ce film porte son attention sur les cartes et les explications des acteurs, avec la volonté de comprendre le tour de magie. Dans la majorité des cas, le sujet ne remarque pas les changements de couleurs introduits dans le décor au fur et à mesure de la présentation du tour de magie. Une fois dévoilés, les changements paraissent pourtant parfaitement détectables : ils concernent la couleur des tee-shirts de l'homme et de la femme, la couleur de la nappe recouvrant la table sur laquelle sont posées les cartes, et la couleur de la toile de fond.

Ces résultats ont été répliqués au niveau de la perception auditive sous le terme de surdité au changement (change deafness). Dans l'expérience de Vitevitch [15], les participants avaient pour consigne principale de répéter des mots dont la complexité lexicale variait. Au milieu de la liste, la voix utilisée pour produire les mots à répéter pouvait changer. Au moins 40 % des participants ne détectaient pas ce changement de locuteur.

Même si les phénomènes observés dans la perception visuelle n'avaient pas été répliqués dans la modalité auditive, centre de notre travail, il ne serait pas illégitime d'aller y puiser des résultats expérimentaux. Ainsi, Delorme et al [16] expliquent que, contrairement aux conceptions anciennes qui accordaient aux modalités sensorielles la composante déterminante des perceptions, l'approche actuelle est plutôt celle qui attribue le rôle dominant à l'information contenue dans l'environnement, à laquelle on accède par nos sens et notre cerveau. Il affirme ainsi que «la perception transcende les modalités sensorielles et s'avère donc amodale». Certes, nous ne touchons pas les odeurs et nous ne goûtons pas les images, mais il est parfois possible de visualiser les sons comme dans l'étude de Martens et al [9], et cette association de sens semblent renforcer le résultat perceptif. Dans tous les cas, l'analogie de sens n'est pas abusive, ce qui nous a amené à consulter certains résultats expérimentaux relatifs à l'olfaction.

Le contexte verbal dans la perception des odeurs

L'odorat n'est pas épargné par les effets de distorsion ou d'illusion perceptive avec notamment, une interférence du langage dans cette modalité de perception. Herz et Von Clef se sont intéressées à l'influence du contexte verbal dans la perception des odeurs et ont émis l'hypothèse que le simple fait d'associer un label à une odeur pouvait provoquer une illusion olfactive. Les auteurs ont constaté en effet qu'une même odeur pouvait être jugée différemment selon le nom qu'on lui donne.

Dans une première étude [17], sont présentées à des sujets naïfs des odeurs accompagnées pour chacune d'elle

1. www.simonslab.com/videos.html

2. <http://richardwiseman.wordpress.com/2009/01/07/colour-changing-card-trick-outtakes>

d'un nom à connotation positive et d'un nom à connotation négative lors de 2 sessions d'olfaction distinctes. Par exemple, l'odeur de violette est présentée comme «concombre frais» (connotation positive) lors de la première session, et comme «moisissure» (connotation négative) à la deuxième session. A chaque stimulus, les sujets doivent juger les odeurs sur leur caractère plaisant, familier et intense et remplir un questionnaire écrit d'où est ensuite tirée une «interprétation perceptive». Les auteurs constatent que pour certaines odeurs, et sous l'unique effet du contexte verbal, jusqu'à 88 % des sujets ont une interprétation perceptive complètement différente entre les deux sessions. Plus les labels proposés sont connus et ancrés dans la culture des sujets, plus ils sont susceptibles d'engendrer des illusions perceptives. Pour les divers caractères cités plus haut, elles constatent également une influence significative de la manipulation des labels.

Dans une autre étude [18], Herz montre que les grades donnés aux caractères plaisant, familier et intense de certaines odeurs diffèrent selon que le nom de l'odeur accompagne le stimulus ou pas. Quand un nom accompagne une odeur, ces caractères hédoniques évalués par le sujet sont en accord avec la connotation du nom proposé, même si dans la réalité, le nom proposé ne correspond pas à l'odeur présentée. Plus généralement, concernant des odeurs communes, les grades attribués aux caractères plaisant, familier et intense sont plus élevés lorsque le nom de l'odeur est donné que lorsqu'il ne l'est pas.

Ces résultats mettent en évidence l'interférence produite sur la perception par une connaissance a priori sur le stimulus à percevoir : une connotation positive vs négative engendre un changement de décision perceptive. Des effets de préférence peuvent être amplifiés par l'association au stimulus d'information verbale.

Synthèse

Certains spécialistes de la perception tels Gibson [dans 16] se refusent à voir dans ces manipulations et illusions perceptives une erreur attribuable au sujet. Ils attribuent les résultats au caractère anormal de la stimulation. Cette objection est légitime dans le cas où il s'agit d'illusions d'optique sur des figures géométriques conçues ad hoc, dans lesquelles il y a une ambiguïté ou un appauvrissement de la stimulation. Ce n'est pas le cas des circonstances de perception citées dans les paragraphes précédents ou de la situation prévue dans notre travail. Ce sont des situations contrôlées mais assez proches de situations naturelles.

Inversement, la synthèse des travaux décrits précédemment pourrait nous amener à nous poser la question formulée par Delorme et al [16] :

«La perception est-elle trop incertaine pour assurer une connaissance appropriée de la réalité ? Ou pour reprendre la question dans les termes de la philosophie traditionnelle : si les sens peuvent ainsi être trompés par des illusions, comment peut-on être certain qu'ils ne nous trompent pas constamment ?».

Nous ne nous inscrivons pas dans cette prise de position extrême, position d'ailleurs non défendue par l'auteur. Notre trame de travail est d'apporter des résultats sur les processus de perception de la voix et de mettre en évidence des aspects méthodologiques importants permettant, à terme, de mieux comprendre ces processus et de proposer des conditions expérimentales optimales. Cela nous amène à poser diverses questions.

Dans le cadre de la perception de la qualité vocale, si les processus top-down décrits précédemment viennent restaurer ce qui est dégradé, l'auditeur doit-il les inhiber pour entendre les altérations de la voix et les caractéristiques pathologiques d'une voix dysphonique ? Comment être sûr que cette capacité naturelle à restaurer une parole dégradée n'est pas un handicap pour le thérapeute de la voix ? Nous avons vu aussi que la consigne donnée au sujet oriente profondément sa perception ou qu'un simple label attribué aux stimuli a un effet perturbateur. Une information donnée à l'auditeur sur le patient, pouvant induire une attention focalisée, orientera-t-elle la perception ? L'auditeur portera-t-il d'avantage son attention sur l'information donnée (ex : situation pré/post chirurgie ou rééducation) plutôt que sur la vraie qualité de la voix ? Le fait d'attendre un certain type de voix (une voix «améliorée» après une chirurgie ou une rééducation par exemple) peut-il occulter des éléments non attendus (une «non amélioration» par exemple) ? Une information donnée à l'auditeur sur la situation du patient (et donc une modification du contexte de l'écoute) empêche-t-elle une perception fiable ?

Hypothèse

Nous émettons l'hypothèse que les phénomènes décrits précédemment interviennent dans le processus de décision pour le jugement de la dysphonie. Des mécanismes top-down peuvent masquer des événements acoustiques saillants ou restaurer des phénomènes absents. L'attention sélective de l'auditeur peut le rendre «sourd» à certaines évidences acoustiques : par exemple, dans le cas de paralysies laryngées, l'auditeur peut se focaliser sur des aspects liés à la fuite glottique (voix soufflée) et n'entendra pas d'autres aspects dysfonctionnels ou normaux. De même, il peut aussi être insensible au changement. Enfin, la suggestion verbale de certaines informations simultanément à l'écoute de la voix peut sérieusement modifier le jugement de la qualité vocale.

De façon plus précise, nous postulons que la connaissance, par l'auditeur, de l'état thérapeutique du patient (en pré ou post traitement), peut modifier le jugement perceptif de la qualité vocale.

MATÉRIEL ET MÉTHODES

Design expérimental

Notre protocole expérimental consistait en l'évaluation, par un jury expérimenté, de voix dysphoniques, présentées en aveugle dans une première expérience, puis accompagnées d'informations sur le parcours médical du

patient dans une deuxième séance. Les différences entre les résultats des deux expérimentations pourraient être attribuées à l'apport d'information sous réserve de rigueur méthodologique.

Les voix ont été présentées par paire, chaque paire étant constituée de deux enregistrements d'un même locuteur. Ces enregistrements ayant été effectués à des dates différentes, la qualité vocale des deux éléments de la paire était la plupart du temps différente. Les auditeurs devaient évaluer ces paires de voix par comparaison: après écoute de chacune des deux voix, et ce, plusieurs fois s'il le souhaitait, l'auditeur devait juger leur degré de dysphonie en utilisant une échelle comparative à 7 points :

La voix A est	nettement moins dysphonique	que la voix B
	moins dysphonique	
	légèrement moins dysphonique	
	de même qualité	
	légèrement plus dysphonique	
	plus dysphonique	
	nettement plus dysphonique	

Nous avons opté pour cette échelle de façon (1) à placer l'auditeur dans des conditions proches des usages en pratique clinique où l'intérêt principal réside souvent dans la perception de la quantité de changement (amélioration ou dégradation) lors de la prise en charge thérapeutique (2) pour avoir une sensibilité de mesure suffisante. En effet, nous aurions pu effectuer un test classique de type GRBAS [2] mais cette échelle de notation absolue sur 4 niveaux nous semblait insuffisamment sensible à de légères différences de qualité, différences qui pouvaient être exprimées dans un test comparatif sur 7 niveaux.

Auditeurs

Les participants à l'expérimentation étaient des auditeurs régulièrement confrontés à l'écoute de voix dysphoniques: 3 chirurgiens ORL, 3 orthophonistes et 1 phoniatre. Bien évidemment, afin de ne pas biaiser les résultats de l'expérience, ils ignoraient l'objectif réel de l'étude qui était présentée comme une mise au point d'un protocole informatisé de jugement de la dysphonie en conditions hospitalières.

Corpus

Les enregistrements proposés aux auditeurs ont été sélectionnés dans la base de données MTO de locuteurs dysphoniques enregistrés dans le service ORL du CHU de la Timone à Marseille [19]. Les 53 patients retenus étaient des adultes, porteurs de nodules ou de polypes (44 femmes et 9 hommes). La restriction à ces deux pathologies a été retenue afin de limiter l'hétérogénéité des formes d'expression de la dysphonie mais aussi car elles peuvent être prises en charge à la fois par des traitements chirurgicaux et orthophoniques, conditions nécessaires à la deuxième partie de l'expérience. Les patients sélectionnés étaient ceux pour lesquels nous disposions d'au moins deux enregistrements réalisés à

des dates différentes afin de disposer de paires de voix pour un même locuteur en accord avec le design expérimental décrit ci-avant.

Le style de parole choisi était de la lecture de texte effectuée sur le premier chapitre de «*La chèvre de Monsieur Seguin*» d'Alphonse Daudet. La durée moyenne des énoncés était de 20 secondes.

Protocole

Déroulement des expériences et précautions méthodologiques. Pour le déroulement des expériences, nous avons utilisé le logiciel PERCEVAL avec son extension LANCELOT [20], développé par le Laboratoire Parole et Langage d'Aix en Provence. Les sessions d'écoute ont été effectuées dans un local fermé, sur le même ordinateur, avec la même carte son et le même casque audiophonique. L'expérience se déroulait en quatre phases : deux sessions d'écoute en aveugle (test-retest) suivies de deux sessions d'écoute influencée (test-retest). Dans la condition aveugle, l'auditeur n'avait aucune information sur les locuteurs qu'il écoutait. Dans les sessions d'écoute influencée, une information était affichée à l'écran indiquant la nature du traitement suivi par le patient (chirurgie ou rééducation) et pour chacune des voix de la paire le statut pré ou post traitement.

Pour chacune des sessions, la consigne était présentée par écrit sur l'écran de l'ordinateur. Avant de démarrer le test proprement dit, 3 items d'entraînement étaient proposés à l'auditeur, lui permettant de s'approprier la tâche et l'échelle. Chaque session se déroulait en deux blocs de 16 à 25 paires permettant au participant de faire une pause. Pour minimiser les effets de fatigue et de perte d'attention, la moitié des auditeurs commençait d'abord par le premier bloc du test, l'autre moitié des participants étant confrontée à ce bloc en fin de test. Dans chaque bloc, les paires étaient présentées dans un ordre aléatoire dans le but de minimiser les effets de liste, bien que cet effet soit déjà réduit par la tâche de comparaison intra trial. Enfin, pour chaque auditeur, les sessions d'écoute étaient séparées les unes des autres d'au moins une semaine pour s'affranchir d'éventuels phénomènes de mémorisation.

Le test-retest. Chaque modalité de présentation des stimuli (aveugle ou avec information) était constituée d'un test et d'un retest pour lesquels l'ordre de présentation des stimuli variait d'une part entre les auditeurs, d'autre part entre le test et le retest pour le même auditeur. La répétition du test en retest a permis, pour les deux types d'écoute, de moyenner les résultats et donc de diminuer une part des effets d'erreurs aléatoires (distraction de l'auditeur, mauvaise manipulation dans le choix de la réponse...) au sens de Bele [6].

La construction des sessions avec écoute influencée. A l'issue de l'écoute en aveugle, chaque paire de voix a été jugée 14 fois (7 auditeurs x 2 sessions). Chaque jugement a été converti en note : [nettement moins dysphonique] ⇔ 3, [moins dysphonique] ⇔ 2,

[légèrement moins dysphonique] $\Leftrightarrow$ 1, [équivalent] $\Leftrightarrow$ 0, [légèrement plus dysphonique] $\Leftrightarrow$ -1, [plus dysphonique] $\Leftrightarrow$ -2, [nettement plus dysphonique] $\Leftrightarrow$ -3

La moyenne des 14 notes permettait d'établir un classement décroissant des paires. Plus la moyenne obtenue était proche de 3 en valeur absolue, plus la quantité de changement entre la voix A et la voix B était importante (+3 renseignant une nette préférence pour A, -3 une nette préférence pour B). Plus la moyenne était proche de 0, plus les auditeurs avaient considéré que les stimuli A et B étaient équivalents en terme de qualité vocale.

A partir des résultats de cette première expérience, nous avons scindé le corpus en deux parties équivalentes en terme de distribution de notes (jeu de données α et β) sur un principe simple. Le jeu de données α était constitué des paires positionnées en 1, 3, 5, 7, ... 45, 47, 49 dans le classement ordonné des notes. Le jeu de données β était constitué des paires 2, 4, 6, 8, ... 46, 48, 50 dans le même classement.

Sur le jeu de données α , l'information fournie à l'auditeur pour la deuxième expérience allait être cohérente au sens clinique du terme dans la mesure où les voix jugées moins dysphoniques en aveugle allaient être déclarées comme le résultat post thérapeutique. Sur le jeu de données β , l'information fournie à l'auditeur pour la deuxième expérience allait être incohérente au sens clinique du terme dans la mesure où les voix jugées moins dysphoniques en aveugle allaient être déclarées comme l'état en pré traitement, impliquant ainsi un résultat thérapeutique défavorable. Nous précisons que les informations qui allaient accompagner les voix étaient construites artificiellement pour équilibrer parfaitement les conditions de test et rendre le design expérimental symétrique. Elles ne tenaient pas compte de la réelle situation pré- post traitement.

Pour éviter que les informations incohérentes du jeu de données β apparaissent trop invraisemblables et sèment le doute dans l'esprit des auditeurs, nous avons exclus les paires de voix dont la note moyenne en aveugle était proche des extrêmes, c'est à dire avec une différence de qualité importante entre la voix A et la voix B. En effet, dans ces cas là, la voix évaluée comme nettement plus dysphonique aurait été déclarée dans le jeu de données β comme post traitement, hypothèse peu vraisemblable et difficilement acceptable par des thérapeutes de la voix. Finalement, nous avons sélectionné les paires dont la note moyenne en aveugle se situait entre +1,5 et -1,5, ce qui représentait un échantillon de 32 paires, la même restriction étant appliquée aux deux jeux de données α et β .

Signalons enfin une dernière précaution méthodologique. Afin d'éviter un effet d'ordre d'écoute dans la paire, la voix déclarée comme pré-thérapeutique correspondait parfois à la voix «A» écoutée en premier et parfois à la voix «B», ceci pour limiter l'effet de récence

(meilleure trace en mémoire de la dernière voix écoutée, laquelle aurait été favorisée).

Par la suite, nous appellerons la condition α comme «information cohérente» ou «jugement renforcé». De même, la condition β sera dénommée «information incohérente», «jugement contraire» ou «jugement contrarié».

RÉSULTATS

L'analyse statistique a été effectuée avec le logiciel 'R'³ version 2.12.0. Que ce soit en situation aveugle ou en écoute contextuelle, la note retenue par paire de voix est la note moyenne obtenue sur les 14 jugements (7 auditeurs * 2 écoutes pour chaque condition). Au total, les expériences ont porté sur 700 écoutes de paires de voix (14 jugements * 50 paires) pour l'expérience en aveugle et 448 (14 * 32) pour l'expérience en contexte, soit au minimum 2296 écoutes de voix.

Ecoute aveugle

La répartition des notes obtenues en écoute aveugle sur les 50 paires de voix est fournie en figure 1. Comme précisé précédemment, les notes extrêmes ont été écartées pour éviter que les informations incohérentes fournies dans la deuxième expérience apparaissent trop invraisemblables. Le reste du corpus (32 paires de voix) a été scindé en deux de façon homogène. Aucune différence significative n'a été mesurée entre ces deux jeux de données [F (1,30) = 0.0011 ; p = 0.97].

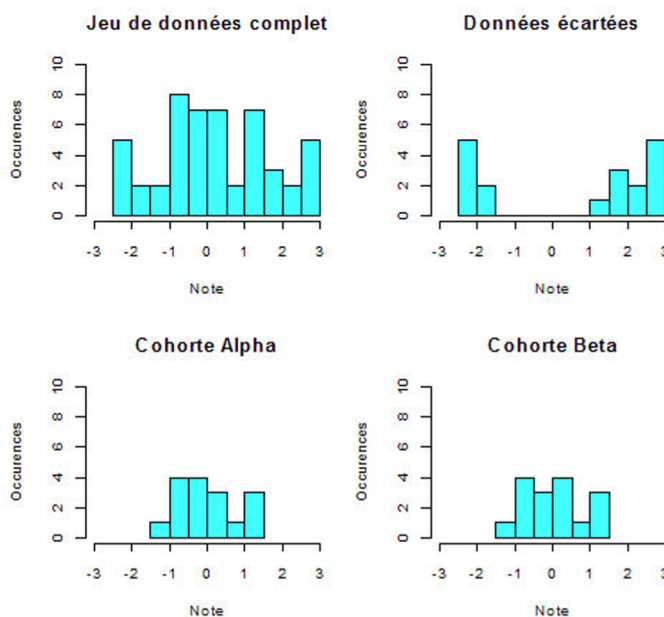


Fig. 1 : Résultats de l'écoute aveugle (en haut à gauche), données écartées pour la suite de l'expérience correspondant aux notes extrêmes (en haut à droite), répartition du corpus restant en cohorte alpha et beta (en bas) correspondant aux deux jeux de données utilisés en situation cohérente (alpha) ou incohérentes (bêta).

3. <http://www.r-project.org>

Ecoute contextuelle

Nous rappelons que l'écoute contextuelle consiste à :

(α) fournir à l'auditeur une information écrite cohérente au sens clinique dans la mesure où les voix jugées moins dysphoniques en aveugle étaient indiquées comme post thérapeutiques.

ou

(β) fournir à l'auditeur une information écrite incohérente au sens clinique dans la mesure où les voix jugées moins dysphoniques en aveugle étaient indiquées comme pré thérapeutiques.

Bien évidemment, lors de la passation du test en condition contextuelle, les stimuli de la cohorte α ou β étaient mélangés et présentés de façon aléatoire. La répartition des notes obtenues sur les 32 paires de voix est fournie en figure 2.

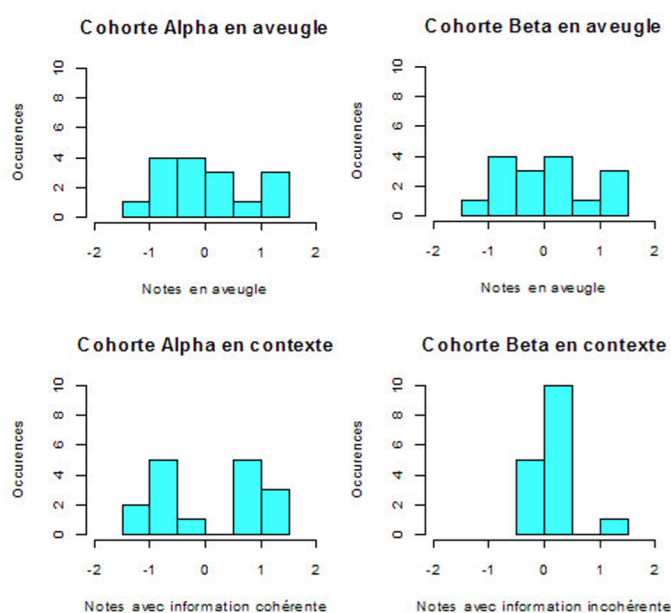


Fig. 2 : Résultats de l'écoute aveugle (en haut) et de l'écoute contextuelle (en bas). La cohorte alpha (resp. beta) a été utilisée en fournissant une information cohérente (resp. incohérente) en situation contextuelle.

Nous observons clairement sur les distributions de la figure 2 un effet net de l'apport d'information lors de l'écoute des voix. Dans la condition où l'information est cohérente, nous observons une distribution quasi bimodale (figure 2, en bas à gauche) correspondant à une décision plus tranchée. Inversement, dans la condition où l'information est incohérente, nous observons une distribution où la majorité des appréciations est centrée autour de zéro (équivalence de qualité vocale des 2 voix de la paire) correspondant à une indécision (figure 2, en bas à droite).

Nous émettons l'hypothèse que dans le cas de la condition cohérente, l'effet de l'apport d'information est

amplificateur : les voix jugées comme légèrement moins dysphoniques en écoute aveugle sont jugées clairement moins dysphoniques car elles sont annoncées comme post thérapeutiques et les voix jugées comme légèrement plus dysphoniques en écoute aveugle sont jugées clairement plus dysphoniques car elles sont annoncées comme pré thérapeutiques. Cette hypothèse rendrait compte de la bimodalité de la distribution de la figure 2. Inversement, nous émettons l'hypothèse que dans le cas de la condition incohérente, l'effet de l'apport d'information est inhibiteur : les voix jugées comme légèrement moins dysphoniques en écoute aveugle sont jugées de qualité vocale équivalente à l'autre voix car elles sont annoncées comme pré thérapeutiques et les voix jugées comme légèrement plus dysphoniques en écoute aveugle sont jugées de qualité vocale équivalente à l'autre voix car elles sont annoncées comme post thérapeutiques.

Pour vérifier cette hypothèse, nous avons réalisé une régression linéaire entre les notes obtenues en contexte en fonction des notes fournies en aveugle pour les catégories α ou β . Les observations sont fournies en figure 3.

La droite d'ordonnée à l'origine 0 et de pente +1 traduit l'absence d'effet de contexte. En effet, un point sur cette droite correspond à deux notes égales en écoute contextuelle et en aveugle. La répartition des notes observées par rapport à cette droite est donc importante.

Dans la condition α où l'apport d'information est cohérent, l'analyse statistique montre que la pente de la droite de régression est de 1.60 ± 0.19 . Cette pente a été obtenue par la méthode des moindres carrés pondérés en tenant compte simultanément des incertitudes sur les données aveugle et contextuelle, où chaque note (aveugle ou contextuelle) est affectée d'un poids qui est inversement proportionnel à la variabilité (erreur standard) inter-juges. Cette pente importante de coefficient directeur 1.6 valide l'hypothèse de l'effet amplificateur du contexte cohérent : la note est accentuée de 60 % par rapport à celle obtenue en aveugle.

Dans la condition β où l'apport d'information est incohérent, l'analyse statistique indique que la pente de la régression linéaire est de 0.17 ± 0.14 , valeur de pente faible validant l'hypothèse inhibitrice de la condition. Rappelons qu'une pente nulle indiquerait la totale indépendance de la note fournie en contexte par rapport à celle obtenue en aveugle. La valeur faible de 0.17 indique que les auditeurs confrontés à une information en contradiction avec leur perception ont eu tendance à se fier à l'information contextuelle et à atténuer fortement les effets perçus en aveugle. On peut même observer des inversions de préférences visibles dans les points situés dans les quadrants supérieurs gauches ou inférieurs droits de la figure 3b. Ces inversions de préférence représentent 50 % des cas de situation incohérente. 100 % des cas d'inversion de préférence sont contraints par l'information contextuelle.

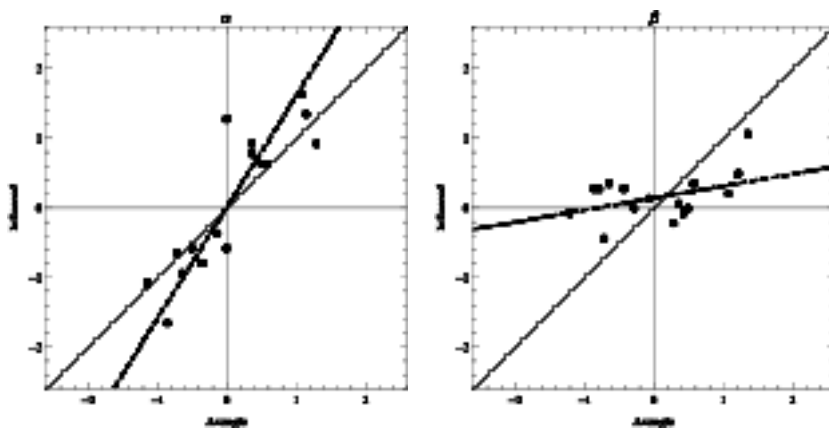


Fig. 3 : Résultats de l'écoute contextuelle influencée (verticalement) en fonction de l'écoute aveugle (horizontalement). A gauche : apport d'information cohérente (cohorte alpha). A droite : apport d'information incohérente (cohorte beta). La droite bissectrice en trait plein est celle de l'absence d'effet (notes identiques entre la condition aveugle et la condition contextuelle). La droite en pointillé est la régression linéaire obtenue à partir des observations.

DISCUSSION

Dans la condition cohérente, l'apport d'information est amplificateur : nous retrouvons les effets observés par Herz dans la perception des odeurs [18] où les préférences sont généralement amplifiées par l'association d'information verbale au stimulus olfactif. Dans notre expérience, les auditeurs sont confortés dans leur jugement par la cohérence des informations fournies en contexte : la qualité vocale est meilleure après qu'avant traitement.

Dans la condition incohérente, nous observons là aussi des analogies avec les résultats de Herz. Les auteurs constatent que pour certaines odeurs, et sous l'unique effet du contexte verbal, jusqu'à 88 % des sujets ont une interprétation perceptive complètement différente entre deux sessions avec connotation positive vs négative. Dans notre expérience, nous observons une inversion de polarité du jugement dans 50 % des cas incohérents.

Nos résultats confirment que la perception «est une construction mentale opérée à la suite d'un traitement de l'information disponible» [11]. Dans le cadre de notre étude, les stimuli auditifs étaient identiques entre les deux conditions d'écoute. Seule variait l'information fournie aux auditeurs sur la nature du locuteur et cette donnée a fait varier de façon importante le résultat. La perception du thérapeute est ainsi influencée par des processus cognitifs top-down («une voix post thérapeutique est meilleure qu'une voix pré thérapeutique») et que cette information le rend «sourd» à des phénomènes qu'il a perçus lors de la phase aveugle. Nous pouvons aussi interpréter ce phénomène comme une capture attentionnelle («on m'indique que c'est post thérapeutique, je n'entends que ce que je m'attends à entendre : l'amélioration»).

Nos auditeurs étaient des professionnels de la prise en charge de la voix. Ils étaient donc, de façon légitime,

en situation de forte implication par rapport aux aspects liés à la réussite thérapeutique. Il serait intéressant d'effectuer ces expériences sur des auditeurs détachés de cette problématique et de vérifier si les résultats restent présents ou disparaissent. De plus, les auditeurs étaient composés de chirurgiens ORL et d'orthophonistes. Ce choix avait été guidé par le fait que lors de l'écoute contextuelle, nous manipulions non seulement la situation pré- post traitement mais aussi la nature du traitement : chirurgie ou rééducation. Nous souhaitions mesurer si les phénomènes d'influence contextuelle pouvaient varier selon l'origine professionnelle : par exemple, nous émettions l'hypothèse que les orthophonistes seraient plus sensibles dans le cas de rééducation que de chirurgie. Le faible effectif de chaque groupe (3 auditeurs par groupe) ne permet pas de mesurer de tels éventuels effets de groupe. Cela nécessite une augmentation du nombre d'auditeurs.

CONCLUSION

L'évaluation d'un résultat thérapeutique lié à une dysphonie est une préoccupation essentielle du phoniatre et de l'orthophoniste. La chirurgie ou la rééducation a-t-elle eu un effet positif, négatif ou négligeable ? L'écoute attentive de la voix avant et après peut être un moyen d'obtenir cette réponse. Mais peut-on estimer qu'il s'agit réellement d'une évaluation dans la mesure où le thérapeute est parfois lui-même juge de son travail thérapeutique et qu'il dispose de nombreuses informations sur le parcours médical de son patient ? Les résultats de cette étude semblent converger vers l'extrême nécessité d'utiliser uniquement des évaluations perceptives en aveugle pour réaliser un bilan perceptif d'une dysphonie.

Remerciements : nous remercions l'Agence Nationale de la Recherche (ANR) pour le financement qu'elle a apporté dans le cadre du projet DESPHO-APADY ANR-08-BLAN-0125 ayant permis la structuration et l'exploitation du corpus de parole pathologique qui a été utilisé dans cette étude.

Références

1. DEJONCKERE PH, BRADLEY P, CLEMENTE P, CORNUT G, CREVIER-BUCHMAN L, FRIEDRICH G, VAN DE HEYNING P, REMACLE M, WOISARD V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). EUR ARCH OTORHINOLARYNGOL. 2001; 258(2), 77-82.
2. HIRANO M. Clinical examination of voice. Springer Verlag, Wien. 1981.
3. REVIS J. L'analyse perceptive des dysphonies. Thèse d'Université Aix-Marseille II. 2004.
4. KREIMAN J, GERRATT BR, PRECODA K, BERKE GS. Individual differences in voice quality perception. J SPEECH HEAR RES. 1992;35(3):512-520.

REV LARYNGOL OTOL RHINOL. 2011;132,1:1-9.

5. 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(1):21-40.
6. BELE I. Reliability in perceptual analysis of voice quality. *JOURNAL OF VOICE.* 2005;19(4):555-573.
7. DE BODT MS, WUYTS FL, VAN DE HEYNING PH, CROUX C. Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *J VOICE.* 1997;11(1):74-80.
8. GERRATT BR, KREIMAN J, ANTONANZAS-BARROSO N, BERKE GS. Comparing internal and external standards in voice quality judgments. *J SPEECH HEAR RES.* 1993;36(1):14-20.
9. MARTENS JWMAF, VERSNEL H, DEJONCKERE PH. The effect of visible speech in the perceptual rating of pathological voices. *ARCH. OTOLARYNGOL. HEAD NECK SURG.* 2007;133(2):178-185.
10. GIOVANNI A, RÉVIS J, GHIO A. Le phoniatre face à l'analyse instrumentale des dysphonies - Use of instrumental vocal analysis in phoniatrics. *REV LARYNGOL OTOL RHINOL (BORD).* 2009;130(1): 3-10.
11. GAILLARD P, BILLIÈRES M, MAGNEN C. La surdit  phonologique illustr e par une  tude de cat gorisation des voyelles fran aises per ues par les hispanophones. *In: Proc. Percepci n y Realidad., Valladolid, Spain.* 2007:187-196.
12. WARREN RM, WARREN RP. Auditory illusions and confusions. *SCI. AM.* 1970;223(6):30-36.
13. SIMONS DJ, CHABRIS CF. Gorillas in our midst: Sustained inattention blindness for dynamic events. *PERCEPTION.* 1999;28(9):1059-1074.
14. SIMONS D, LEVIN D. Change blindness. *TRENDS IN COGNITIVE SCIENCES.* 1997;1(7):261-267.
15. VITEVITCH MS. Change deafness: The inability to detect changes between two voices. *J EXP PSYCHOL HUM PERCEPT PERFORM.* 2003; 29(2):333-342.
16. DELORME A, FLUCKIGER M. Perception et r alit  : une introduction   la psychologie des perceptions. *De Boeck Universit , Bruxelles (Belgique).* 2003.
17. HERZ RS, CLEF JV. The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions? *PERCEPTION.* 2001;30(3):381-391.
18. HERZ RS. The effect of verbal context on olfactory perception. *J EXP PSYCHOL GEN.* 2003;132(4):595-606.
19. GHIO A, POUCHOULIN G, TESTON B, PINTO S, FREDOUILLE C, DE LOOZE C, ROBERT D, VIALLET F, GIOVANNI A. How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? *SPEECH COMMUNICATION.* Accepted in Special Issue "Advanced Voice Assessment". 2011.
20. GHIO A, ANDR  C, TESTON B, CAV  C. PERCEVAL : une station automatis e de tests de PERCEption et d'EVALuation auditive et visuelle. *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA).* 2003;22:115-133.