



**HAL**  
open science

## Voix de patientes avec suspicion d'atteinte du nerf laryngé supérieur : rythme et perception

Hélène Massis, Marie-Hélène Degombert

### ► To cite this version:

Hélène Massis, Marie-Hélène Degombert. Voix de patientes avec suspicion d'atteinte du nerf laryngé supérieur : rythme et perception. Sciences cognitives. 2020. dumas-02951533

**HAL Id: dumas-02951533**

**<https://dumas.ccsd.cnrs.fr/dumas-02951533v1>**

Submitted on 28 Sep 2020

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

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

ACADÉMIE DE PARIS  
SORBONNE UNIVERSITÉ  
MÉMOIRE POUR LE CERTIFICAT DE  
CAPACITÉ D'ORTHOPHONISTE

---

**Voix de patientes avec suspicion  
d'atteinte du nerf laryngé  
supérieur : rythme et perception**

---

*Directeur de mémoire:*

**Mme PILLOT-LOISEAU**

Orthophoniste et maître de Conférences en Phonétique, membre de la section 07 du  
Conseil National des Universités, Laboratoire de Phonétique et Phonologie UMR 7018  
CNRS, Université Paris 3 Sorbonne Nouvelle

ANNÉE UNIVERSITAIRE : 2019-2020

**MASSIS HELENE**

**DEGOMBERT MARIE-HELENE**

## **Remerciements**

Nous tenons à remercier notre maître de mémoire Madame PILLOT-LOISEAU, pour son aide précieuse et ses encouragements ainsi que Monsieur ARDISSON, notre rapporteur.

Nous remercions également tous les professionnels du service d'endocrinologie de l'hôpital Pitié-Salpêtrière où ont été recrutées les patientes, et plus particulièrement le Professeur Tresallet et le Docteur Lazard pour leur disponibilité et leur confiance.

Nous remercions Mademoiselle Le Pape pour avoir accepté de partager le fruit de ses recherches et pour ses éclairages.

De même, nous remercions tous les jurés, experts et naïfs, qui ont accepté de consacrer du temps à l'évaluation perceptive et bien évidemment, toutes les patientes pour leur investissement et leur participation à notre étude.

Enfin, nous tenons à remercier nos professeurs pour leur aide et conseils.

**Attestation de non plagiat :**


Je soussignée, Hélène MASSIS, déclare être pleinement consciente que le plagiat de documents ou d'une partie d'un document publiés sur toutes formes de support, y compris l'Internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce mémoire.

Signature :

**Attestation de non plagiat :**

Je soussignée, Marie-Hélène DEGOMBERT, déclare être pleinement consciente que le plagiat de documents ou d'une partie d'un document publiés sur toutes formes de support, y compris l'Internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce mémoire.

Signature :



## Liste des tableaux et figures

Tableau 1. Résultat du test Shapiro-Wilk.	7
Tableau 2. Corrélation des jugements experts vs naïfs.	11
Figure 1. Distribution du grade G obtenu à partir de la perception de la parole continue par chaque juge des jurys experts 1 et des jurys experts 2 et pour chaque population.	10
Figure 2. Distribution du score VHI30 par population.	12

## Résumé

Après thyroïdectomie totale, la plainte vocale des patients, hors paralysie récurrentielle, est attribuée à l'atteinte de la branche crico-thyroïdienne du nerf laryngé supérieur. Cependant elle reste encore difficilement objectivable. Cette étude cherche à déterminer si des paramètres rythmiques altérés peuvent servir au diagnostic. Elle est complétée par une évaluation perceptive et un questionnaire de ressenti chez 28 femmes avec suspicion d'atteinte de la branche crico-thyroïdienne du NLS (CT-), comparées à 27 autres sans dommage (CT+) après thyroïdectomie totale, au regard d'une population témoin (T). A partir d'épreuves de parole, des paramètres rythmiques ont été mesurés, et une évaluation perceptive réalisée par un jury expert et naïf. Parmi les paramètres rythmiques étudiés, la proportion des intervalles vocaliques sur tout l'énoncé (%V) est le seul à être altéré chez les CT- comparé aux CT+ et T, témoin d'une réduction vocalique. La mobilisation des plis vocaux serait plus complexe pour les CT-. Cette conclusion complète les constatations déjà établies, à savoir la réduction de la plage vocale, et la fatigue vocale pour cette population. L'évaluation perceptive du timbre vocal à partir de l'échelle GRBAS-I ne permet pas de discriminer totalement la voix des 3 populations, mais la voix est perçue comme moins modulée chez les CT- et CT+ comparée à celle des T. L'auto-évaluation des patientes CT- confirme le ressenti d'un handicap vocal parfois sévère. Cette étude innovante sur le rythme et l'évaluation perceptive des CT-, a tout son intérêt pour répondre aux attentes des patients et des professionnels de la voix.

**Mots-clés :** *Nerf laryngé supérieur ; thyroïdectomie ; rythme ; évaluation perceptive ; qualité vocale*

## Abstract

After total thyroidectomy, the patient's voice complaint, excluding recurrent paralysis, is attributed to damage on the cricothyroid branch of the superior laryngeal nerve (SLN). However, it is still difficult to objectify. This study seeks to determine whether altered rhythmic parameters can be used as a diagnosis. It is supplemented by a perceptual evaluation and a questionnaire on the patient's feelings. 28 women with suspected injury on the cricothyroid branch of the SLN (CT-) were compared to 27 others without damage (CT+) after total thyroidectomy, compared to a control population (T). From speech tests, rhythmic parameters were measured, and a perceptual evaluation was carried out by an expert and a naive jury. Among the rhythmic parameters, the proportion of vocal intervals over the whole utterance (%V) is the only one to be altered in CT- compared to CT+ and T, witnessing a vocalic reduction. The mobilization of the vocal folds would be more complex for the CT-. This conclusion complements the findings already established, namely the reduction of the vocal range and vocal fatigue. Perceptual evaluation of the voice timbre from the GRBAS-I scale does not allow the voices of the 3 populations to be totally discriminated. However, the voice is perceived to be less modulated in CT- and CT+ compared to T. Self-assessment of patients CT- confirms the feeling of a severe vocal handicap for a few. Meeting the expectations of patients and voice professionals, this innovative study on the rhythm and perceptual evaluation of CT- has its whole interest.

**Keywords:** *Superior laryngeal nerve; thyroidectomy; rhythm; perceptual evaluation; voice quality*

## **Introduction (H. Massis)**

Suite à une thyroïdectomie totale, les patientes se plaignent de troubles vocaux. Elle reste une chirurgie délicate de par l'atteinte possible de la branche externe du nerf laryngé supérieur (EBNLS) qui fournit les seules fibres motrices au muscle crico-thyroïdien (tenseur des plis vocaux, Barczyński et al., 2013). Sa trajectoire anatomique variable ne facilite pas son identification en peropératoire, rendant le geste chirurgical complexe et entraînant possiblement des lésions (Orestes et Chhetri, 2014). Sans outils diagnostiques fiables, cette lésion et la plainte étaient jusqu'à récemment encore, peu étudiées et sous-estimées. Le terme « suspicion d'atteinte » est d'ailleurs préférablement employé.

La plainte porte sur une fatigabilité vocale, une intensité vocale plus réduite, une aggravation de la voix, une difficulté dans la production des aigus et parfois une impossibilité à terminer une phrase (Orestes et Chhetri, 2014). Ceci laisse envisager une gestion pneumo-phonatoire plus compliquée et donc possiblement une répercussion sur le rythme, au-delà des altérations de hauteur et d'intensité de voix (Orestes et Chhetri, 2014). Les recherches portent davantage sur les altérations du rythme pour des troubles de la parole (Liss et al., 2009 ; Lowit, 2014). Aucune étude à notre connaissance n'a porté sur l'étude du rythme en voix pathologique. Sur la base des capacités de distinction des langues grâce aux variables rythmiques quantitatives (Ramus et al. 1999, cité par Pillot-Loiseau et Xie, 2018), nous pensons que ces mêmes variables sont à même de distinguer des voix pathologiques de voix non pathologiques. A partir d'épreuves de parole, et d'une comparaison de sujets avec suspicion d'atteinte de l'EBNLS (CT-), sans atteinte (CT+) suite à une thyroïdectomie totale et des témoins (T), nous chercherons à mettre en lumière les variables rythmiques permettant de discriminer la voix pathologique de la voix sans dommage. Des paramètres temporels (issus de Ramus et al. 1999 et Pillot-Loiseau et Xie, 2018) dont %V (mesure de la proportion des intervalles vocaliques), paramètre utilisé de préférence par Ramus (1999), devraient être altérés chez les CT-. Pour compléter cette étude, une évaluation perceptive des voix et parole sera menée par deux jurys expert et naïf. Considérée comme le « gold standard » pour caractériser un désordre vocal et mesurer son atteinte (Oates, 2009), elle sera mise en lien avec les mesures rythmiques dans le but de distinguer nos populations. L'auto-évaluation du patient étant primordiale, le VHI servira à évaluer la sévérité de son trouble ou handicap (Jacobson et al., 1997, cités par Oates 2009). Notre étude cherche à identifier un marqueur d'altération rythmique de la parole chez ces patientes, ce qui confirmera notre hypothèse du retentissement de l'atteinte sur le rythme de la parole, et pourra aider à l'objectivation de la plainte et au diagnostic en pratique clinique.



## Méthode (H. Massis)

### *Population*

#### *Patients*

Notre recherche cible des femmes, ayant subi une thyroïdectomie totale avec suspicion d'atteinte de la branche crico-thyroïdienne de l'EBNLS (population CT-), à comparer avec des femmes ayant subi la même chirurgie sans aucun dommage suspecté (population CT+). Notre étude s'est déroulée en collaboration avec un chirurgien expert, ayant opéré toutes les patientes de notre population, et un chirurgien ORL expert du même service posant le diagnostic de suspicion d'atteinte de l'EBNLS. Ces patientes, âgées entre 24 ans et 85 ans, ont toutes été opérées dans ce service entre 2016 et 2019. Elles ont été enregistrées entre 1 à 20 mois de leur date d'opération. Suite à l'intervention, l'altération (parésie ou paralysie) du muscle crico-thyroïdien a été suspectée lors de l'opération, témoignant d'une atteinte de la branche externe du nerf laryngé supérieur. Les examens cliniques et laryngoscopiques, en post-opératoire, ont éliminé toute atteinte de la mobilité des plis vocaux, toute lésion du nerf récurrent, myasthénie ou dystonie. Les populations CT- et CT+ ont été appariées en âge à une population témoin T. Les différences significatives présentes uniquement dans la population CT- seront donc imputables à l'atteinte de l'EBNLS.

Dans chaque groupe, nous n'avons pas inclus toute personne non-francophone, toute personne atteinte de troubles neurologiques et/ou neuromusculaires, de maladie auto-immune, de troubles auditifs et cognitifs, toute personne avec antécédents de troubles vocaux. Les facteurs âge, traitement hormonal de substitution, et les conduites d'atteinte à la voix et à son utilisation ont aussi été pris en compte. De même, n'ont pas été incluses les personnes bilingues avec un accent non francophone, influençant le rythme ainsi que les personnes avec un faible niveau de lecture, et les patientes avec une prise en charge orthophonique. Cette étude faisant suite au précédent mémoire de Gaétane Le Pape (2018) « voix et prosodie de patients avec suspicion d'atteinte du nerf laryngé supérieur », nous avons repris certains sujets de sa cohorte. De nouveaux témoins et de nouvelles patientes ont été recrutées afin notamment de réduire le délai entre la chirurgie et l'enregistrement, point de discussion du précédent mémoire, soit 33 nouveaux enregistrements. Au total, nous avons pu constituer une cohorte de 82 participants (une cohorte moins large que celle de Le Pape, du fait de nos critères d'inclusion et non inclusion plus restrictifs), soit 28 patientes pour la population CT- avec un âge moyen de 56,6 ans (26-78 ans), 27 patientes pour la population CT+ avec un âge moyen de 57 ans (24-85 ans) et 27 sujets pour la population T avec un âge moyen de 55,7 ans (26-81 ans). Pour la population CT-, le délai entre la chirurgie et

l'enregistrement est de 1 à 6 mois pour 11 patientes, de 7 à 12 mois pour 4 patientes et de 13 à 20 mois pour les 13 dernières. Ce même délai est compris entre 1 à 6 mois pour la population CT+.

### *Jury de perception*

20 jurés de langue maternelle française ont participé au test d'évaluation perceptive. Ils n'ont déclaré aucune perte auditive. Parmi eux se trouve un panel d'experts. Il se compose de 9 individus, 7 femmes et 2 hommes, âgés entre 41 et 61 ans (moyenne d'âge (M) = 46,2 ans, et écart-type (ET) = 6,7 ans). Ces experts sont tous spécialistes de la voix (8 orthophonistes et 1 phoniatre) avec de nombreuses années d'expérience dans ce domaine (M = 16,3 années, ET = 6,2 années). Les 11 autres jurés, des femmes uniquement, n'ont aucune formation ni expérience dans la voix et sont désignés comme le jury naïf. Elles sont âgées entre 30 ans et 66 ans (M = 47,8 ans, ET = 12,7 ans). L'ensemble des jurés a participé à l'évaluation perceptive sur la base du volontariat, sans percevoir aucune rémunération.

### *Matériel*

Le matériel utilisé pour cette étude reprend de manière fidèle le dispositif mis en place lors des passations de la précédente étude qui a suivi les recommandations du guide du Committee on Phoniatics (Dejonckere et al., 2001). Les épreuves sont présentées sous format PowerPoint. Les enregistrements sont effectués individuellement dans une pièce calme et isolée du bruit extérieur. Les captations vocales sont enregistrées grâce à un micro casque AKG C520 situé à une distance constante de la bouche (deux doigts de la commissure des lèvres, à un angle de 45°). La voix des sujets est recueillies avec une carte son Audient ID-4 sur le logiciel Audacity® via un ordinateur avec le système d'exploitation macOS Mojave version 10.14.3. La passation du protocole a été faite avec un examinateur formé, dans une pièce isolée à l'Hôpital de la Pitié-Salpêtrière, ou chez le sujet.

### *Corpus*

Avant la passation des épreuves, la signature du formulaire de consentement a été requise et la note d'informations (résumé de la recherche, déroulement du protocole, modalités de confidentialité et de rétractation) a été donnée à chaque sujet. Les épreuves présentées reprennent la totalité du corpus de l'étude précédente afin de recréer les conditions de passation à l'identique. Un test de paires minimales et un test de l'horloge ont été proposés au préalable afin de répondre aux critères d'inclusion et de non inclusion préétablis. La durée totale de la passation est de 20 minutes et comprend différentes épreuves classiques

d'évaluation vocale avec des épreuves spécifiques rendant compte de la modulation vocale (Voir Annexe A). Le corpus sur lequel se fonde notre étude porte sur la lecture d'un extrait du texte de « La poupée rouge » de Pierre Gripari (1994) (Annexe B).

En plus des épreuves du corpus, deux questionnaires ont été adressés aux trois populations avec le *Voice Handicap Index* (VHI) (Jacobson et al., 1997) et un questionnaire de plaintes spécifiques (QPS, questionnaire d'inclusion créé spécialement pour l'étude précédente avec l'équipe hospitalière et pédagogique, portant sur les facteurs protecteurs, les facteurs de risque, les habitudes de vie et les ressentis vocaux).

Tous les enregistrements ont été anonymisés et ne portent aucune mention de la population à laquelle ils appartiennent.

### *Données rythmiques*

Les données rythmiques ont été extraites du dialogue de « La poupée rouge » (Gripari, 1994) permettant de recueillir un échantillon de parole comprenant des modulations variées. Faisant suite aux cinq phrases d'énoncé introductives, le dialogue permet de témoigner de la capacité vocale et de la qualité vocale en fin de lecture, et où la gestion pneumophonique pourra être la plus difficile chez les patientes avec suspicion d'atteinte de la branche externe de l'EBNLS. Ces enregistrements ont été annotés à l'aide du logiciel PRAAT (Boersma et Van Heuven, 2001) par deux examinateurs formés dans le cadre de ce protocole. Toutes les annotations ont été reprises par un seul examinateur afin de respecter la conformité des découpages. Ils comportent un découpage par niveau : phrases, mots, syllabes et phonèmes. Les frontières des voyelles ont été positionnées à la disparition des formants qui les constituent. Le logiciel Correlatore (Mairano et Romano, 2010) a été utilisé afin d'extraire des fichiers TextGrid les variables rythmiques suivantes : %V (ou Vperc), Vmean, Cmean,  $\Delta V$  (ou Vdev),  $\Delta C$  (ou Cdev) (Ramus et al. 1999 ; Pillot-Loiseau et Xie, 2018) :

- %V est une mesure de la proportion des intervalles vocaliques dans l'énoncé (Pillot-Loiseau et Xie, 2018). Elle permet de mettre en avant la complexité des structures syllabiques et également de donner une première impression de la réduction vocalique (Ramus et al. 1999, cité par Pillot-Loiseau et Xie, 2018).
- Vmean représente la moyenne de la durée des voyelles.
- Cmean représente la moyenne de la durée des consonnes.
- $\Delta V$  est une mesure de l'écart-type de la durée des intervalles vocaliques (Pillot-Loiseau et Xie, 2018). Elle permet de caractériser la réduction vocalique. Il est attendu en français que cette mesure ne témoigne pas d'une grande réduction vocalique (Ramus et al. 1999, cité par Pillot-Loiseau et Xie, 2018).

- $\Delta C$  est une mesure de l'écart-type des intervalles consonantiques (« toute séquence ininterrompue de consonnes » ou « intervalles inter-vocaliques », (Ramus, 1999)). Elle permet de mettre en valeur la complexité des structures syllabiques. En français, la valeur n'est pas élevée puisque la structure syllabique tend à être plus régulière (Ramus et al. 1999, cité par Pillot-Loiseau et Xie, 2018).

Un relevé sous Excel a permis de recueillir les dernières variables, à savoir le nombre total de syllabes, le nombre total de pauses et le débit articulatoire (soit le nombre de syllabes divisé par la durée de parole hors pauses). La durée de la parole sera également étudiée afin de constater si les variables subissent des modifications dues à un énoncé plus long ou à la réalisation même de la parole.

### *Tâche de perception*

Deux types de stimuli (voix et parole continue) ont été extraits des 82 enregistrements obtenus pour les trois populations. Le stimulus de la voix se compose du /a/ tenu sur 3 secondes (2 productions consécutives). Celui de la parole continue est la dernière phrase du conte pour enfants « La poupée rouge » (Gripari, 1994) : « Et menteur avec ça, enchaîne le petit diable enchanté, tout ce que j'aime en vérité. Eh bien, puisqu'il en est ainsi, petit paresseux, moi, je vais te rendre heureux. ». Cet échantillon de parole continue, d'une durée moyenne de 19,96 secondes, comprend un nombre moyen de 77,2 syllabes et de 168,7 phonèmes.

Le type de tâches a été différent en fonction du type de jury. En effet, le jury expert a évalué la voyelle tenue puis la parole continue, en utilisant l'échelle GRBAS-I (Hirano, 1981), connue de tous les jurés. Seules les dimensions G (grade de sévérité), R (raucité), B (souffle), et I (instabilité) ont été évaluées, sur 4 niveaux de sévérité (0= normal ; 1= légèrement altéré ; 2= moyennement altéré ; 3= sévèrement altéré). Les dimensions A (asthénie) et S (forçage), recueillant le moins de fiabilité d'après la littérature (Dejonckere et al., 2001), n'ont pas été proposées.

Sur la parole continue, les experts comme les naïfs ont évalué de la même façon 3 paramètres. D'une part, la modulation de la voix avec la question « selon vous la voix est ... ? » avec 4 réponses possibles : non modulée (sans intonation, sans mélodie), un peu modulée (avec un peu d'intonation, de mélodie), normalement modulée (avec une intonation, une mélodie normale), trop modulée (avec trop d'intonation, de mélodie).

Puis la hauteur de la voix a été évaluée avec la question « selon vous la voix est ... ? » avec cinq réponses possibles sur une échelle de 1 (= très grave) à 5 (= très aigue).

Enfin le débit de la phrase a été apprécié à travers la question « selon vous, la phrase est prononcée ... ? » avec trois réponses possibles : de façon lente, de façon normale et de façon accélérée.

La présentation des stimuli (disponibles uniquement en ligne) et le recueil des réponses des jurés ont été effectués par l'intermédiaire d'un formulaire Google. Chaque juré a reçu un lien envoyé par email, avec les prérequis suivants : disposer d'un ordinateur, d'un casque et d'une connexion à internet. L'écoute a donc été individuelle.

L'ordre de présentation des stimuli était aléatoire. Ils pouvaient être écoutés autant de fois que le juré le jugeait nécessaire pour affiner son évaluation perceptive. Cependant, nous avons recommandé aux jurés, en préambule au questionnaire, de faire confiance dans la majorité des cas à leur première écoute. Les jurés pouvaient passer au stimulus suivant une fois que toutes les dimensions avaient reçu une cotation. Après avoir évalué l'ensemble des dimensions pour tous les stimuli, il était demandé aux jurés de donner par écrit une caractéristique perceptive saillante qu'ils auraient identifiée pendant le test d'écoute, de renseigner l'heure de début et de fin du test, et de noter si une pause s'était avérée nécessaire et sa durée.

Face au nombre de stimuli à écouter (2 fois 82 voyelles tenues et 82 phrases, notamment pour les experts) et à évaluer, il a été décidé, pour limiter le temps d'écoute et le coût cognitif, de séparer tous ces stimuli en deux blocs, appariés en classe d'âge et population : un bloc 1 de 42 productions et un bloc 2 de 40 productions. Dans chaque bloc, plus de 25% des stimuli (voyelles et phrases) ont été dupliqués pour permettre d'évaluer l'accord intra-juges. Les jurys experts (E) et naïfs (N) ont donc été divisés en deux : E1, E2, N1 et N2. Le bloc 1 a été évalué par les jurys E1 et N1, le bloc 2 par les jurys E2 et N2.

### ***Statistiques***

Les données ont été rassemblées sur un tableur Excel. Elles ont ensuite été traitées sous le logiciel JMP®. Concernant le rythme, une analyse de la distribution a été effectuée avec le test de Shapiro-Wilk afin de déterminer la normalité de la distribution des données. Les résultats du test ont révélé qu'une partie des données suivait une loi de distribution normale. Pour celles-ci, un test ANOVA a pu être fait afin de voir si les moyennes étaient significativement différentes pour ensuite pouvoir faire un T-Student afin de comparer les populations deux à deux. Pour les données ne suivant pas une distribution normale, le test de Kruskal-Wallis a été effectué afin de comparer les données.

Dans l'évaluation perceptive, l'accord inter et intra-juges chez les experts a été exploré sur le grade G de l'échelle GRB-I (Hirano, 1981). Tous les tests statistiques utilisés sont non paramétriques puisque les variables de l'évaluation perceptive sont des variables qualitatives ordinales. Premièrement le coefficient de Kappa Cohen (k) a été calculé. Ce résultat statistique peut être défini comme une mesure de l'unanimité des évaluations conduites par deux juges sur un même stimulus. Ensuite, pour évaluer l'accord inter-juges à l'échelle d'un jury, nous avons calculé le Fleiss kappa (Fk). Nous avons appliqué les recommandations de Landis et Koch (1977) avec  $k < 0,00$ , désaccord ;  $k = 0,00-0,20$ , accord très faible ;  $k = 0,21-0,40$ , accord faible ;  $k = 0,41-0,60$ , accord modéré ;  $k = 0,61-0,80$ , accord fort ;  $k = 0,81-1,00$ , accord presque parfait. La significativité des moyennes a été estimée à partir du test non paramétrique de Kruskal-Wallis et les comparaisons entre populations à partir du test non paramétrique de Wilcoxon. Enfin le coefficient de Spearman a été calculé pour déterminer les corrélations possibles entre variables.

Tous les tests statistiques seront comparés au nouveau seuil  $p < 0,005$ , le seuil  $p < 0,05$  devenant un seuil de significativité suggestif (Laccourreye et al., 2019).

### Résultats (M-H. Degombert)

#### Résultats des données issues du rythme

Compte tenu du nombre de sujets par groupe (28 sujets dans la population CT-, 27 sujets dans la population CT+ et 27 sujets dans la population T), nous avons voulu vérifier si la distribution des données suivait une loi normale. Un test de Shapiro-Wilk a permis d'analyser cette distribution au sein de nos variables quantitatives : %V, Vmean, Cmean,  $\Delta V$ ,  $\Delta C$ , nombre de syllabes, nombre de pauses et débit articulatoire (tableau 1).

Tableau 1. Résultat du test Shapiro-Wilk.

Variables	Population CT-		Population CT+		Population T		Loi normale
	W	p	W	p	W	p	
%V	0,944	0,1426	0,968	0,5550	0,933	0,0831	Oui
Vmean	0,939	0,1042	0,988	0,9852	0,965	0,4753	Oui
Cmean	0,984	0,9275	0,970	0,5982	0,979	0,5469	Oui
$\Delta V$	0,964	0,4367	0,876	0,0039*	0,950	0,2134	Oui
$\Delta C$	0,970	0,5762	0,931	0,0726	0,955	0,2855	Oui
nombre de pauses	0,956	0,2891	0,963	0,4295	0,942	0,1343	Oui
débit articulatoire	0,972	0,6349	0,958	0,3336	0,958	0,3388	Oui
nombre de syllabes	0,830	0,0004*	0,830	0,0005*	0,602	<0,0001*	Non
durée de la parole	0,941	0,1163	0,984	0,9440	0,976	0,7708	Oui

\* $p < 0,005$

Ce test a ainsi pu démontrer que la majorité de nos variables suivent une distribution de loi normale. Les résultats ont permis de voir que la distribution de la variable nombre de syllabes ne suit pas une loi normale dans les populations de l'étude.

Du fait de la distribution normale de la plupart de nos variables quantitatives, un test ANOVA a été opéré sur celles-ci afin de voir si les moyennes sont significativement différentes. Le résultat du test ANOVA sur %V montre des différences suggestives de significativité des moyennes de cette mesure ( $F_{(2,79)}=5,39$  ;  $p=0,006$ ). La valeur p trouvée étant proche du seuil suggestif de significativité, il a donc été intéressant de poursuivre la comparaison des moyennes de %V. Le test de Student permettant une comparaison par paire, démontre que la moyenne de %V de la population CT- ( $M=52,1$  ;  $ET=0,7$ ) est suggestivement inférieure à celle de la population CT+ ( $M=55,1$  ;  $ET=0,7$ ) avec un t calculé de  $t_{79}=2,820$  ( $p=0,0061$ ), et avec celle de la population T ( $M=55,1$  ;  $ET=0,7$ ) avec un t calculé de  $t_{79}=2,848$  ( $p=0,0056$ ). Cependant, les moyennes de la population CT+ et T ne se distinguent pas ( $t_{79}=0,027$  ;  $p=0,9785$ ). Ces premiers résultats nous permettent de conclure que la moyenne de %V de la population CT- se différencie significativement de celle des populations CT+ et T.

Le test ANOVA a également été effectué sur Vmean. Il démontre que les moyennes ne sont pas distinguables par population ( $F_{(2,79)}=1,24$  ;  $p=0,29$ ), alors qu'elles le sont pour Cmean ( $F_{(2,79)}=6,75$  ;  $p=0,002$ ). Le test de Student sur cette dernière variable révèle que la moyenne Cmean de la population CT- ( $M=99,8$  ;  $ET=1,97$ ) est significativement supérieure à Cmean de la population T ( $M=89,5$  ;  $ET=2,01$ ) avec un t calculé de  $t_{79}=-3,65$  ( $p=0,0005$ ), ce qui n'est pas le cas entre CT- et Cmean de CT+ ( $M=93,7$  ;  $ET=2,01$ ) avec un t calculé de  $t_{79}=-2,17$  ( $p=0,033$ ), et entre CT+ et T ( $t_{79}=-1,47$  ;  $p=0,146$ ). Cmean est donc une variable dont la moyenne permet de distinguer seulement la population CT- de T.

Les tests ANOVA montrent que les moyennes des populations ne sont pas significativement différentes pour  $\Delta V$  ( $F_{(2,79)}=0,98$  ;  $p=0,38$ ),  $\Delta C$  ( $F_{(2,79)}=0,09$  ;  $p=0,91$ ), le débit articulatoire ( $F_{(2,79)}=0,95$  ;  $p=0,39$ ), le nombre de pauses ( $F_{(2,79)}=0,99$  ;  $p=0,38$ ) et la durée de la parole ( $F_{(2,79)}=0,12$  ;  $p=0,89$ ). Ces résultats démontrent que seules les variables %V et Cmean sont des mesures dont les moyennes permettent de distinguer certaines populations entre elles.

La dernière mesure rythmique, ne suivant pas une loi de distribution normale dans l'étude, est le nombre de syllabes. La médiane des syllabes produites de chaque population est de 77. La variabilité des syllabes s'explique dans la réalisation de Schwas supplémentaires (« enchaîne le petit diable enchanté » [ãʃɛnləpətɪdʒabləãʃãte]), la plupart des sujets produisant plutôt : « enchaîne le petit diabl'enchanté » [ãʃɛnləpətɪdʒablãʃãte]. A l'inverse, sur certaines productions le /ə/ n'était pas prononcé : « enchaîne le **p'tit** diable enchanté »

[ãʃɛnləptidjablãʃãte] qui est normalement produit « enchaîne le **petit** diable enchanté » [ãʃɛnləpətidjablãʃãte]. Les moyennes du nombre de syllabes par population sont de 76,7 (ET=0,76) pour la population CT-, 76,8 (ET=0,79) pour la population CT+ et 76,9 (ET=0,44) pour la population T. Le test non paramétrique de Kruskal-Wallis a permis de mettre en avant que les moyennes des populations sur cette variable n'étaient pas significativement différentes ( $\chi^2=1,95$  ;  $p=0,38$ ).

### *Résultats des données issues de la perception*

#### Accord inter juges

Le maximum de concordance des jugements est atteint quand les jurés cotent le grade G sur la parole continue. L'accord inter-juges pour les 4 juges du jury E1 sur le grade G donne un coefficient de Kappa Cohen  $k$  échelonné entre 0,08 et 0,33 ce qui est interprété comme un accord très faible à faible. Pour les 5 juges du jury E2, l'accord inter-juges sur le grade G donne un  $k$  échelonné entre -0,10 et 0,53, interprété comme un désaccord à un accord modéré.

Les valeurs négatives de  $k$  sont systématiquement liées au juré E24. De plus, en analysant le grade G par stimulus des jurys E1 et E2, il apparaît que 18 stimuli semblent avoir posé des difficultés aux juges étant donné la dispersion des réponses enregistrées. Ces valeurs aberrantes sont également réparties entre les blocs de stimuli 1 et 2, avec une certaine disparité par population (8 dans la population CT-, 3 dans la population CT+, et 7 dans la population T) et par classe d'âge : 14 stimuli proviennent d'individus âgés de 50 ans et plus. Après avoir écarté la cotation extrême du jury E24 et les 18 stimuli difficiles à évaluer pour les jurys, le coefficient de Kappa Cohen a été recalculé et donne les valeurs suivantes : pour les juges de E1,  $k$  est échelonné entre 0,15 et 0,43 et l'accord devient très faible à modéré et pour les juges de E2,  $k$  est échelonné entre 0,45 et 0,74 l'accord devient modéré à fort. Pour le jury E1, le Fleiss Kappa ( $F_k$ ) passe de 0,17 à 0,24 quand les calculs sont retraités après élimination des valeurs aberrantes, interprétable comme un accord faible. Pour le jury E2,  $F_k$  passe de 0,19 à 0,57 quand les calculs sont retraités des valeurs aberrantes, soit un accord modéré entre les 4 juges.

L'accord inter-juges pour les juges naïfs donne un  $k$  échelonné sur la question de la modulation entre 0,04 et 0,44 pour le jury N1 ce qui est interprété comme un accord très faible à modéré, entre 0,21 et 0,49 pour le jury N2, soit un accord faible à modéré ; sur la question hauteur entre 0,07 et 0,35 pour le jury N1 ce qui est interprété comme un accord très faible à faible, entre 0,02 et 0,33 pour le jury N2, soit un accord très faible à faible ; sur



la question débit entre 0,11 et 0,57 pour le jury N1 ce qui est interprété comme un accord très faible à modéré, entre 0,11 et 0,41 pour le jury N2, soit un accord faible à modéré.

### Accord intra juges

De la même manière, l'accord intra-juges a été mesuré avec le coefficient de Kappa Cohen  $k$  et retraité après élimination des valeurs aberrantes. Il a été calculé à partir des cotations du grade G et des réponses aux questions portant sur la modulation, la hauteur et le débit pour les stimuli dupliqués dans le bloc 1 et 2 (plus de 25% des stimuli ont été dupliqués).

Pour E1, le coefficient  $k$  est échelonné entre 0,25 et 0,73 pour le grade G, entre 0 et 0,71 pour la modulation, entre 0,18 et 0,57 pour la hauteur et entre 0,25 et 0,71 pour le débit. Pour E2, le coefficient  $k$  est échelonné entre 0,60 et 1 pour le grade G, entre 0,43 et 0,74 pour la modulation, entre 0,52 et 0,74 pour la hauteur et entre 0 et 0,57 pour le débit. Pour N1, le coefficient  $k$  est échelonné entre 0 et 0,31 pour la modulation, entre 0 et 0,67 pour la hauteur et entre 0 et 0,74 pour le débit. Pour N2, le coefficient  $k$  est échelonné entre 0,08 et 1 pour la modulation, entre 0,14 et 0,79 pour la hauteur et entre 0,25 et 0,48 pour le débit.

### Scores de perception

La figure 1 montre la distribution du grade G obtenu à partir de la perception de la parole continue par chaque juge des jurys experts 1 (E11, E12, E13, E14) et jurys experts 2 (E21, E22, E23, E24, E25) et pour chaque population.

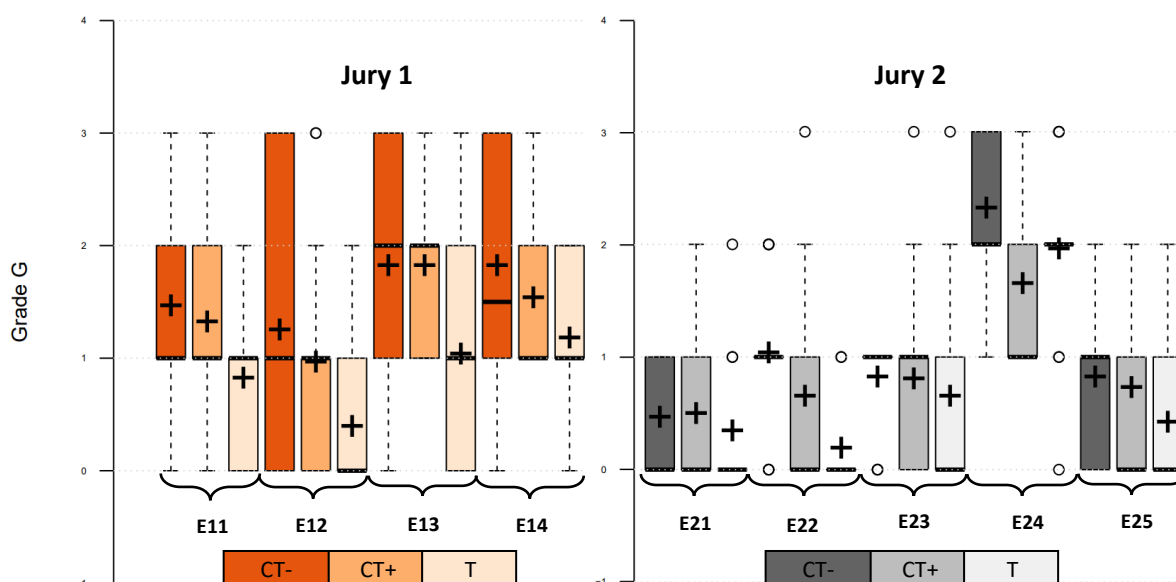


Figure 1. Distribution du grade G obtenu à partir de la perception de la parole continue par chaque juge des jurys experts 1 et des jurys experts 2 et pour chaque population.

La figure 1 indique que les médianes du score G pour les populations CT- et CT+ sont dans la majorité des cas, inférieures aux moyennes respectives, ce qui indique une grande dispersion vers les scores les plus élevés de G. Les écarts-types à la moyenne sont très importants, ce qui ne permet pas de conclure sur la différence significative des moyennes entre ces trois populations (test de Kruskal-Wallis :  $\text{Khi-2} = 6,79$  ;  $p < 0,033$ ).

Cependant la Figure 1 montre que les moyennes du grade de sévérité G sont systématiquement décroissantes : la plus forte étant pour la population CT-, puis la population CT+ et la population T recueille le score G le plus faible, pour tous les juges, sauf E13, E21 et E24.

#### Corrélation entre jurys experts et naïfs

Pour les experts, la durée de l'évaluation perceptive (GRB-I sur la voyelle tenue et la parole continue et trois questions sur la parole continue) est en moyenne de 54 minutes avec un écart-type de 13 minutes et 47 secondes. Pour les naïfs, l'évaluation perceptive (trois questions sur la parole continue) a duré en moyenne 25 minutes, avec un écart-type de 6 minutes. La corrélation entre les jugements des experts et des naïfs pour les questions de modulation, hauteur et débit, calculée avec le coefficient de Spearman, est positive, élevée et significative, sauf pour la population T sur la question concernant le débit, comme illustré dans le Tableau 2.

Tableau 2. Corrélation des jugements experts vs naïfs.

<i>Population</i>	<i>Variables</i>	<i>Coefficient de Spearman</i>	<i>Prob &gt; p</i>
CT-	modulation	0,8182	<0,0001*
CT-	hauteur	0,9062	<0,0001*
CT-	débit	0,8529	<0,0001*
CT+	modulation	0,6147	0,0014*
CT+	hauteur	0,6606	0,0004*
CT+	débit	0,8009	<0,0001*
T	modulation	0,8281	<0,0001*
T	hauteur	0,7534	0,0001*
T	débit	0,5707	0,0086

\* $p < 0,005$

Pour la question modulation, le jugement des experts et naïfs donne une moyenne pour la population CT- de 1,2 (ET = 0,7), pour la population CT+ de 1,5 (ET = 0,6) et pour la population T de 2,1 (ET = 0,5). Le test de Kruskal-Wallis ( $\text{Khi-2} = 18,42$ ,  $p < 0,0001$ ) permet de conclure à la différence significative des moyennes entre ces trois populations. La comparaison des moyennes par la méthode de Wilcoxon donne une différence significative

entre les populations CT- et T ( $Z= 3,8$ ,  $p < 0,0001$ ) et entre CT+ et T ( $Z= 3,28$ ,  $p < 0,001$ ), mais pas entre CT- et CT+ ( $Z= 1,67$ ,  $p = 0,09$ ). La voix est perçue moins modulée chez les patientes avec suspicion d'atteinte et sans suspicion d'atteinte comparée à celle de la population témoin. Pour les questions portant sur la hauteur et le débit, les mêmes tests statistiques ne relèvent aucune différence significative entre les trois populations : les juges ne détectent pas de hauteur de voix significativement différente ni de débit de phrases significativement différent entre les 3 populations.

## VHI

### Scores VHI par population

La Figure 2 montre la distribution du score du VHI30 par population. Le score VHI n'excède pas 32 points sur 120 pour les populations CT+ et T. Le score total VHI pour la population CT- s'étend jusqu'à 67 points sur 120. Pour rappel, un score au-delà de 60 pointe un handicap ressenti comme sévère.

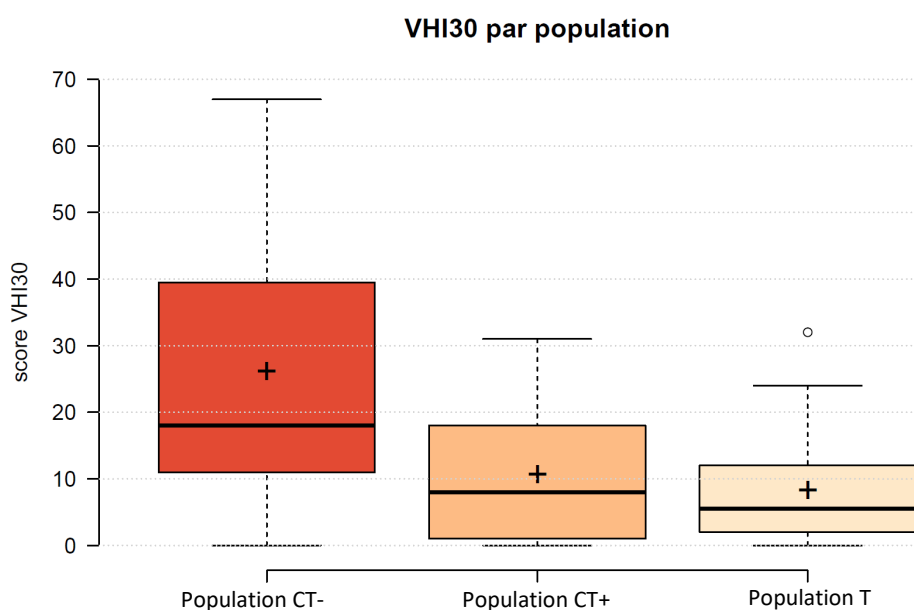


Figure 2. Distribution du score VHI30 par population.

### Significativité des moyennes

Le test non paramétrique de Kruskal-Wallis montre des moyennes de Total VHI significativement différentes ( $Khi-2=13,8$  ;  $p<0,001$ ). La méthode de Wilcoxon de comparaison pour chaque paire montre une différence significative ( $Z= -3,09$  ;  $p<0,002$ ) entre la population CT- ( $M=28,2$  ;  $ET=19,8$ ) et CT+ ( $M=8,5$  ;  $ET=8,2$ ) et CT- et T ( $M=6,7$  ;  $ET=6,7$ ) avec  $Z = -3,29$  ( $p<0,001$ ) mais pas entre CT+ et T ( $Z = -0,41$  ;  $p<0,68$ ).

### Corrélation du Total VHI avec le grade G par population

Le test de corrélation ne révèle aucune corrélation entre le total VHI et le grade G pour les trois populations. Le coefficient de Spearman est de 0,2055 ( $p=0,44$ ) pour la population CT- et de -0,44 ( $p=0,13$ ) pour la population T. Pour la population CT+, le coefficient de Spearman est de 0,48 ( $p=0,023$ ) ce qui indique une corrélation avec une significativité suggestive.

### **Discussion** (M-H. Degombert)

#### *Rythme*

Les résultats montrent que des marqueurs d'altérations rythmiques, tel que %V, permettent de distinguer les patientes avec suspicion d'atteinte de l'EBNLS, des patientes sans atteinte et des témoins. Nous avons notamment identifié %V comme marqueur d'altération du rythme dans la parole pathologique. Selon Ramus (1999), %V est une variable à utiliser préférentiellement du fait de sa « variance moindre » comme il le démontre dans son étude. La différence suggestive de significativité des moyennes de %V est à la limite de  $p<0,005$  entre les moyennes de la population CT- et les populations CT+ ( $t_{79}=2,82$  ;  $p=0,006$ ) et T ( $t_{79}=2,85$  ;  $p=0,0056$ ). Il n'est pas attendu en français qu'il y ait une grande variabilité de la durée des voyelles du fait de l'isochronie syllabique de cette langue (Pillot et Xie, 2018). De ce fait, la différence des moyennes à la limite du seuil interroge tout de même sur les modifications de parole de notre population. Lorsque l'on regarde les moyennes des populations de notre étude, la moyenne de %V des CT- est inférieure à celle des CT+ et T. Selon Ramus (1999), à propos de la distinction des langues, un %V abaissé témoigne de syllabes plus complexes à condition que cela soit suivi par l'augmentation de  $\Delta C$ . Or dans notre étude, nous ne retrouvons pas cette corrélation. En effet, si  $C_{mean}$  témoigne bien d'une modification, ce n'est pas le cas de  $\Delta C$ . Il semble bien que la population CT- a une moyenne de  $C_{mean}$  nettement supérieure à la population T. En moyenne, les patientes CT- réalisent des consonnes plus longues. Cependant,  $\Delta C$  ne montre aucune différence significative entre nos populations ce qui ne permet pas de conclure sur la réalisation des consonnes chez les CT-. Du fait que la durée de parole et le nombre de syllabes réalisées par les sujets ne permettent pas de les différencier, il apparaît donc que notre hypothèse se voit confirmée : %V est la seule mesure montrant une modification de la parole chez les CT-, probable compensation de leurs difficultés vocales. En français, il n'est pas attendu que %V soit réduit, ce qui est pourtant le cas pour les CT- comparé aux CT+ et aux T. Il semble que la réalisation des voyelles, voisées par définition, est ce qui met le plus en difficulté les CT-.

La mise en vibration des plis vocaux est donc bien ce qui est le plus complexe pour la population des CT-. Cette conclusion est en lien avec les constatations déjà faites sur les difficultés rencontrées par les patients avec suspicion d'atteinte de l'EBNLS, autour de la réduction de la plage de variation de la fréquence fondamentale, et la fatigue vocale (Orestes et Chhetri, 2014 ; Neri et al., 2011).

Les résultats de  $\Delta V$  ne vont pas dans le sens d'une réduction vocalique comme %V. Cependant, selon Ramus (1999), cette mesure est très influençable, contrairement à %V et  $\Delta C$ . Beaucoup de précautions sont à prendre quant à son interprétation. En effet,  $\Delta V$  est très influencée par « la réduction vocalique, l'allongement contrastif des voyelles, l'allongement de voyelles dans certains contextes [et] l'existence de voyelles intrinsèquement longues » (Ramus, 1999) comme les voyelles nasales « garçon, menteur, bien » (/õ,ã,ẽ/) et les semi-consonnes en phonétique comme dans « oui, crois, travaille » (/y,w,j/), présentes dans le dialogue de notre corpus. Ces dernières sont décomptées comme des consonnes dans le logiciel Correlatore (Mairano et Romano, 2010). Il est d'ailleurs commun que celles-ci soient classées en consonnes comme le fait l'annotation SAMPA, précédant ou suivant systématiquement une voyelle. Elles ne sont donc pas incluses dans le calcul de  $\Delta V$ . Afin de pouvoir utiliser la variable  $\Delta V$  dans la distinction de nos populations, un autre corpus pourrait être nécessaire, en enlevant les voyelles nasales et les semi-consonnes pouvant entrer en conflit avec la réalisation de la voyelle précédente et suivante. En effet, les annotations sous PRAAT (Boersma et Van Heuven, 2001) sont complexifiées et les frontières entre les semi-consonnes et les voyelles prêtent à discussion.

En outre, selon Arvaniti (2012), les marqueurs du rythme ne sont pas toujours fiables pour distinguer les langues entre elles, du fait des variations intrinsèques des langues. Nous avons pu mettre en lumière cette limitation, car hormis %V, il semble qu'aucune autre variable ne puisse distinguer les paroles de sujets entre eux. Notre étude propose d'étudier ces variations au sein du français entre une population pathologique et des populations témoins. Peu de littérature peut venir appuyer notre démarche innovante. Des études comme celles de Liss et al. (2009) et Lowit (2014) utilisent les variables rythmiques dans un but de distinction de populations d'une même langue. Cependant, aucune des études à notre connaissance ne porte sur l'étude du rythme de la parole dans le contexte particulier de troubles *vocaux*. De plus, Liss (2009) ne fait qu'une distinction intra-pathologie (pour des troubles *de parole*) et non population pathologique versus population témoin. Notre étude est donc une première dans l'utilisation des valeurs rythmiques dans l'objectif de diagnostiquer une pathologie vocale.

## *Perception*

Les jurys experts et naïfs peuvent identifier perceptivement les voix et parole pathologiques de patientes avec suspicion d'atteinte de l'EBNLS et les voix et parole de patientes sans suspicion d'atteinte et de la population témoin. Après avoir conduit une évaluation perceptuelle sur la voyelle tenue /a/ et sur un extrait de parole continue grâce à l'échelle GRB-I (Hirano, 1981), les experts portent un jugement plus sévère sur les stimuli issus de la population avec suspicion d'atteinte versus les autres populations. Pour rappel, cette évaluation est en aveugle et mixe les stimuli de voix pathologiques de différents degrés avec des voix saines, appariées en classe d'âge. La figure 1 montre en effet que les moyennes du grade de sévérité G sont systématiquement décroissantes, la plus forte étant enregistrée pour la population CT-, puis la population CT+ et la population T recueillant le score G le plus faible, pour tous les juges sauf E13, E21 et E24.

Les jurys experts et naïfs ont évalué selon les mêmes modalités les stimuli de parole continue sur les questions de modulation, hauteur et débit. La corrélation entre les jugements des experts et des naïfs pour ces mêmes questions, est positive, élevée et significative, sauf pour la population T sur la question concernant le débit, comme illustré dans le Tableau 2. Ces résultats sont satisfaisants car nous savons que des différences peuvent être observées entre experts et naïfs, expliquées par les ressources additionnelles dont disposent les experts avec une large expérience, pour analyser un signal acoustique et catégoriser une qualité vocale. Le choix de limiter cette évaluation pour le jury naïf à trois questions, formulées de manière simple et concise, sur des notions familières au plus grand nombre sans aucune expérience dans la voix et la parole, a permis d'obtenir des résultats positivement corrélés avec les évaluations des experts.

Bien que les moyennes ne soient pas significativement différentes entre populations sauf pour la question concernant la modulation, ces moyennes sont systématiquement croissantes, la plus faible représentant la population CT- et la plus élevée la population T. Il était attendu que la modulation et la hauteur soient plus altérées chez les patientes CT+ et CT- versus la population T. En effet, au-delà de l'atteinte fréquentielle, les patients atteints d'une paralysie de l'EBNLS témoignent communément d'une difficulté à projeter la voix et l'impossibilité d'atteindre les sons aigus (Potenza, et al., 2017). Dans la parole conversationnelle, la voix est décrite comme affaiblie, soufflée et monotone (Roy et al., 2009, cités par Potenza, et al., 2017). Ces résultats sont concordants.

Cependant les résultats perceptifs ne discriminent pas suffisamment les populations étudiées. En effet, seule la modulation de la voix sur l'extrait de la parole continue, évaluée par les experts et les naïfs, recueille une significativité dans la différence des moyennes entre les

trois populations. Comment expliquer ces résultats peu discriminants pour le grade G de l'échelle GRB-I et les questions portant sur la hauteur et débit, alors que les experts et les naïfs sont capables d'identifier perceptivement les voix et parole pathologiques ? D'une manière générale, il est à noter que ces derniers résultats évaluent avant tout la voix. En effet, l'évaluation GRB-I, même sur des items de parole, concerne bel et bien le timbre de la voix alors que la modulation est davantage liée à la parole, dans sa dimension prosodique, qui fait d'ailleurs l'objet de davantage de plaintes chez les patientes CT-. Pour compléter cette réponse, on peut aussi observer d'une part, que les écarts-types à la moyenne sont très importants ce qui indique une grande dispersion des jugements pour un même stimulus, à l'intérieur d'un même jury. Après élimination des valeurs extrêmes, l'accord inter-juges est faible pour E1 et modéré pour E2. Pour les jurys naïfs, l'accord inter-juges oscille entre très faible à modéré en fonction de la question et du bloc de stimuli. Il est courant de trouver un faible accord inter-juges dans les études d'évaluation perceptives et ces résultats étaient donc attendus. En effet, comme la notion de prototype en psychologie cognitive (Kreiman et al., 1993, cités par Oates, 2009) parle de « standard interne », propre à l'auditeur, sur ce que doit être une voix normale. Ce prototype fait intervenir des paramètres tout à fait personnels et donc subjectifs tels que ses goûts, sa culture, son approche du genre, l'époque... Il n'existe donc pas de norme absolue mais des normes personnelles.

D'autre part, nous observons des scores du grade G plus élevés dans le jury E1, ce qui indique une dispersion des jugements entre jurys. Outre les paramètres personnels et subjectifs propres à chaque juge, se pose alors la question de la composition des jurys E1 et E2. Les deux jurys d'experts sont composés de juges avec une moyenne d'âge équivalente (M de E1 = 46,2 ans (ET=4,7) ; M de E2 = 46,2 (ET=8,6)). Tous ces jurés ont été recrutés sur la base de leurs nombreuses années d'expérience dans l'écoute de patients dysphoniques mais avec des disparités entre E1 et E2 (M de E1 = 20,5 ans (ET=3,3) ; M de E2=13 ans (ET=6,2)). La moyenne des années d'expérience pour le jury E1 étant supérieure à celle du jury E2, le facteur « années d'expérience » dans cette évaluation perceptive pourrait être discriminant, et peut-être expliquer les moyennes plus pertinentes dans le jury E1. Nous savons aussi que le *background* professionnel peut être un facteur important à prendre en compte dans l'évaluation. Or, tous ces juges sont orthophonistes à l'exception d'une phoniatre. Leur pratique clinique s'exerce en cabinet libéral et /ou en milieu hospitalier. L'échelle GRBAS-I leur est familière. La composition des jurys, calibrée en termes d'âge et de *background* professionnel, n'est cependant pas tout à fait équilibrée en termes d'années d'expérience.

La distribution du grade G par juge, des jurys E1 et E2 (figure 1), pose aussi la question du calibrage des blocs de stimuli. Les stimuli du bloc 1 et 2 sont-ils perceptivement équivalents bien que ces blocs aient été construits avec une répartition similaire en termes de population et de classe d'âge ? A l'intérieur de ces blocs de stimuli, nous avons observé une grande dispersion dans les scores de G pour 18 locutrices. Ces évaluations qui ont posé des difficultés, sont également réparties entre les blocs de stimuli 1 et 2, avec une certaine disparité par population (8 dans la population CT-, 3 dans la population CT+, et 7 dans la population T) et par classe d'âge : 14 stimuli proviennent de locutrices âgées de 50 ans et plus. Ces enregistrements sont pourtant de qualité équivalente aux autres stimuli. Dans la littérature, il a été montré qu'un locuteur plus âgé que l'auditeur pourra recueillir un jugement plus sévère (Goy et al., 2016, cités par Delvaux et Pillot-Loiseau, 2019).

Enfin, pour pouvoir évaluer perceptivement une voyelle tenue et la parole continue de 82 locutrices, sur la base d'une échelle connue des professionnels de la voix et sur des questions plus ouvertes afin d'intégrer un jury naïf au jury expert, notre méthodologie initiale a été retravaillée. Cette révision porte sur le nombre de stimuli à évaluer et donc le nombre de jurys pour limiter le temps d'écoute et la charge cognitive. Le jury initial de neuf experts est passé à deux jurys de quatre et cinq experts, et le jury de onze naïfs est passé à deux jurys de six et cinq naïfs. De même, le nombre de stimuli a été divisé par deux : 42 stimuli dans le bloc 1 et 40 stimuli dans le bloc 2. La représentativité de chaque population dans chaque bloc étant moindre, cette méthodologie pourrait expliquer l'absence de significativité pour les variables déjà mentionnées.

Malgré ces dispositions, pour les experts, la durée de l'évaluation perceptive (GRB-I sur la voyelle tenue et la parole continue et trois questions sur la parole continue) est en moyenne de 54 minutes avec un écart-type de 13 minutes et 47 secondes. Or nous savons que des facteurs internes tels que la fatigue, l'attention, la mémoire, générés entre autres, par l'effet de la tâche et la taille de l'échantillon, peuvent influencer les jugements. (Kreiman et al., 1993; Gerratt et al., 1993, cités par Barsties, 2015). Décrite comme un « gold standard » dans la littérature pour décrire les désordres vocaux, l'évaluation perceptive prête à débat du fait de sa nature subjective. Nos résultats ne peuvent prouver la fiabilité et la validité de notre recherche.

La distribution du score du VHI (Figure 2) illustre le handicap vocal ressenti par les patientes et la population témoin. Pour rappel, un score au-delà de 60 pointe un handicap ressenti comme sévère. Les scores peuvent atteindre 67 points sur 120 pour la population CT- alors que pour la population sans dommage et la population témoin, le score maximum est de 32 points. Ces scores témoignent bien de l'altération de la qualité vocale chez ces patientes. Les



moyennes du VHI sont significativement différentes entre les populations CT- et T, et CT- et CT+ (Khi-2=13,81 ;  $p < 0,001$ ). En revanche l'étude n'a pas permis de mettre en lumière une corrélation entre le total VHI et le grade G pour ces 3 populations, hormis une corrélation pour la population CT+ avec  $p=0,0291$  soit une significativité dite suggestive (Laccourreye et al., 2019).

#### *Validité de la recherche*

L'une des difficultés majeures de cette recherche se trouve dans le diagnostic initial de la pathologie étudiée. Aucun examen à ce jour ne permet de confirmer objectivement la paralysie de l'EBNLS (Orestes et Chhetri, 2014). Cette observation est ce qui motive d'une part cette étude et la questionne. Cependant, la population a pu être recrutée suite au geste opératoire et l'observation experte peropératoire d'un seul chirurgien. Le diagnostic de suspicion a ensuite été posé par un médecin ORL expert dans cette pathologie. La réalité clinique de la démarche diagnostique d'atteinte de l'EBNLS impose une rigueur scientifique dans le suivi des patientes prises en charge par un seul et même chirurgien expert puis diagnostiquées par un seul et même médecin ORL expert des paralysies de l'EBNLS.

En vue d'améliorer la validité interne de cette étude, et pour de futures recherches, nous préconisons d'améliorer ou revoir les points suivants. Malgré de nombreux nouveaux enregistrements, les populations n'ont pas pu être appariées de manière satisfaisante en termes de délai d'enregistrement. Nous avons inclus uniquement les patientes avec un délai (entre l'opération et l'enregistrement) de 20 mois maximum. En effet, chez certains patients, une régression spontanée de la paralysie de l'EBNLS suspectée, peut être observée (Orestes et Chhetri, 2014). Barczyński et al. (2013) évoquent des récupérations spontanées mais aucune étude à notre connaissance ne permet de connaître avec précision ce délai de récupération. Il serait encore plus rigoureux de recruter une population de patientes à maximum 6 mois du geste opératoire afin de s'assurer que la récupération spontanée n'opère pas trop tôt et comparer ainsi des voix de patientes au plus proche de la chirurgie. Cette conclusion sur la récupération possible de certaines patientes a pu influencer nos résultats mais notre large cohorte a permis de limiter cet écueil comme l'atteste l'observation de nos résultats.

Les conditions d'enregistrement n'ont pas été strictement identiques pour tous, certains individus ont été enregistrés à leur domicile et non à l'hôpital comme la plupart de notre cohorte. De fait, les enregistrements ont pu être impactés de ces différentes conditions. De plus, les gains réglés sur la carte son ont été conservés à l'identique mais nous n'avons pas

procédé à un calibrage de l'intensité par sonomètre. Une normalisation de l'intensité a dû être faite pour que l'épreuve perceptive se passe dans les meilleures conditions. Enfin le niveau de maîtrise du langage écrit peut avoir un impact sur notre étude. En effet, si l'épreuve de lecture répond aux exigences expérimentales (Arvaniti, 2012), elle a ses limites en imposant des contraintes vocales inhérentes à la lecture à haute voix.

### **Conclusion**

Notre étude met en lumière de manière inédite un paramètre rythmique altéré : %V. En effet, on observe une réduction vocalique témoignant de la difficulté de la mise en vibration des plis vocaux chez les CT-. Les autres paramètres rythmiques étudiés ne permettent pas de discriminer nos populations. Notre étude confirme l'hypothèse selon laquelle certaines composantes du rythme sont altérées chez les voix de patientes avec suspicion d'atteinte de l'EBNLS. L'évaluation perceptive experte de la voix ne permet pas de distinguer nos populations au même titre que les variables rythmiques. Cependant, nos jurys s'accordent sur la question de la modulation vocale concernant la parole, identifiée comme plus faible chez les CT-. Le VHI pointe le handicap vocal ressenti par les patientes. Ces données matérialisent la plainte des patientes et démontrent combien cette plainte ne peut plus être sous-estimée. Le monitoring en peropératoire et un suivi régulier et approfondi en post opératoire sont plus que nécessaires pour prendre en compte une atteinte encore difficilement objectivable. Le résultat visé ne doit plus être la seule réussite de la chirurgie (exérèse d'une tumeur par exemple) mais doit aussi prendre en compte la préservation de l'EBLNS pour minimiser tout risque de handicap vocal. En cas de suspicion d'atteinte, une prise en charge rapide, globale et collaborative entre chirurgien, ORL et orthophoniste doit être préconisée. Cette étude offre de nouvelles pistes pour approfondir la recherche sur le rythme des voix pathologiques, ainsi que pour compléter la connaissance de chaque professionnel et adapter la prise en charge des patientes, dont la plainte est encore peu écoutée. Un diagnostic d'atteinte de l'EBNLS ne peut être envisagé à partir d'un paramètre isolé. Il est donc recommandé de compléter le bilan orthophonique avec ces nouveaux éléments. En plus de l'évaluation de l'étendue vocale à partir d'un texte lu, une analyse du rythme pour mettre en lumière une réduction vocalique serait tout à fait appropriée, de même qu'une évaluation perceptive portant sur le timbre de la voix et la modulation, corrélées à une auto-évaluation du handicap vocal. Tous ces paramètres permettront d'affiner les objectifs thérapeutiques en proposant des matériels de parole spécifiques, intégrant la prosodie, à l'image de certains éléments de notre corpus. Et ceci dans le but d'une prise en charge orthophonique toujours plus efficace et « sur-mesure ».

### Bibliographie Apa 7<sup>ème</sup> édition

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351-373. <https://doi.org/10.1016/j.wocn.2012.02.003>

Barczyński, M., Randolph, G.W., Cernea, C.R., Dralle, H., Dionigi, G., Alesina, P.F., ... International Neural Monitoring Study Group. (2013). External branch of the superior laryngeal nerve monitoring during thyroid and parathyroid surgery : International Neural Monitoring Study Group standards guideline statement. *The Laryngoscope*, 123, S1-S14. <https://doi.org/10.1002/lary.24301>

Barsties, B., et De Bodt, M. (2015). Assessment of voice quality : Current state-of-the-art. *Auris Nasus Larynx*, 42(3), 183-188. <https://doi.org/10.1016/j.anl.2014.11.001>

Boersma, P., et Van Heuven, V. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5(9/10): 341-345.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., et Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology*, 258(2), 77-82. <https://doi.org/10.1007/s004050000299>

Delattre, P. (1966). Les dix intonations de base du français. *French review*, 40(1), 1-14.

Delvaux, V., et Pillot-Loiseau, C. (2019). Perceptual Judgment of Voice Quality in Nondysphonic French Speakers : Effect of Task-, Speaker- and Listener-Related Variables. *Journal of Voice*, 1-12. <https://doi.org/10.1016/j.jvoice.2019.02.013>

Fougeron C., Delvaux V., Pernon M., Lévêque N., Borel S., Pellet P. Bagou O., Trouville R. Ménard L., Catalano S., Lopez U., Kocjancic-Antolik T., et Laganaro M. (2016) Chapitre 14. *MonPaGe : un protocole informatisé d'évaluation de la parole pathologique en langue française*. In Joyeux N. & Topouzkhianian S. (eds) 'Orthophonie et technologies innovantes'. OrthoEdition.

Gripari, A. (1994). *Nouveaux contes* [CD audio]. Paris: Compacts Radio France.

Hirano, M. (1981). GRBAS scale for evaluating the hoarse voice & frequency range of phonation. *Clinical examination of voice*, 5, 83-84.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., et Newman, C. W. (1997). The voice handicap index (VHI) development and validation. *American Journal of Speech-Language Pathology*, 6(3), 66-70. <https://doi.org/10.1044/1058-0360.0603.66>

Laccourreye, O., Lisan, Q., Bonfils, P., Garrel, R., Jankowski, R., Karkas, A., Leboulanger, N., Makeieff, M., Righini, C., Vincent, C., et Martin, C. (2019). Use of P-values and the terms “significant”, “non-significant” and “suggestive” in Abstracts in the European Annals of Otorhinolaryngology, Head & Neck Diseases. *European annals of*

*otorhinolaryngology, head and neck diseases*, 136(6), 469-473.  
<https://doi.org/10.1016/j.anorl.2019.10.008>

Landis, J.R., et Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174. <https://doi.org/10.2307/2529310>

Liss, J.M., White, L., Mattys, S.L., Lansford, K., Lotto, A.J., Spitzer, S.M., et Caviness, J.N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of speech, language, and hearing research*, 52, 1334-1352. [https://doi.org/10.1044/1092-4388\(2009/08-0208\)](https://doi.org/10.1044/1092-4388(2009/08-0208))

Lowit, A. (2014). Quantification of rhythm problems in disordered speech: A re-evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130404. <https://doi.org/10.1098/rstb.2013.0404>

Mairano, P., et Romano, A. (2010) Un confronto tra diverse metriche ritmiche usando Correlatore. In: Schmid, S., Schwarzenbach, M. & Studer, D. (eds.) La dimensione temporale del parlato, (Proc. of the V National AISV Congress, University of Zurich, Collegiengebaude, 4-6 February 2009), Torriana (RN): EDK, 79-100.

Neri, G., Castiello, F., Vitullo, F., De Rosa, M., Ciammetti, G., et Croce, A. (2011). Post-thyroidectomy dysphonia in patients with bilateral resection of the superior laryngeal nerve: a comparative spectrographic study. *Acta Otorhinolaryngologica Italica*, 31(4), 228-234.

Oates, J. (2009). Auditory-Perceptual Evaluation of Disordered Voice Quality. *Folia Phoniatica et Logopaedica*, 61(1), 49-56. <https://doi.org/10.1159/000200768>

Orestes, M. I., et Chhetri, D. K. (2014). Superior laryngeal nerve injury : Effects, clinical findings, prognosis, and management options. *Current opinion in otolaryngology & head and neck surgery*, 22(6), 439-443. <https://doi.org/10.1097/MOO.0000000000000097>

Pillot-Loiseau, C., et Xie, H. (2018). Transfert rythmique du chinois mandarin au français dans l'apprentissage du Français Langue Étrangère : acoustique et perception. In *SHS Web of Conferences* (Vol. 46, p. 09001). EDP Sciences, Congrès Mondial de Linguistique Française CMLF 2018, 1-15. <https://doi.org/10.1051/shsconf/20184609001>

Potenza, A. S., Araujo Filho, V. J. F., et Cernea, C. R. (2017). Injury of the external branch of the superior laryngeal nerve in thyroid surgery. *Gland Surgery*, 6(5), 552-562.

Ramus, F. (1999). La discrimination des langues par la prosodie: Modélisation linguistique et études comportementales. 1<sup>ère</sup> journée d'étude sur l'identification automatique des langues, *De la caractérisation à l'identification des langues*, Lyon, 19 janvier 1999, 131-140.

## Annexe A : corpus

- Lecture du dialogue des dix intonations (1 assertion, 1 question partielle, 1 question totale et 2 exclamations) du français de Delattre (1966) avec une modification justifiée dans le mémoire.
- Lecture de phrases (« Allons donc, prouve le moi! ») énoncées à différentes intensités (voix conversationnelle, voix forte et voix projetée ou implicatrice).
- Tenue de la voyelle /a/ sur 4 secondes deux fois et deux autres fois pour recueillir le temps maximum de phonation (TMP).
- Répétition du dialogue des dix intonations du français de Delattre (1966).
- Sirène et glissando sur un /a/.
- Lecture de six phrases : 3 assertives: « Anne-Marie et moi allons à la mer. / Anne, Marie et moi allons à la mer.» et « Mélanie vend du lilas. ». Puis les mêmes phrases sous la forme d'une question et d'une exclamation (ordre). Ces phrases sont extraites du protocole d'évaluation de la parole pathologique en langue française (paires minimales intonatives) (Monpage, Fougeron et al., 2016).
- Lecture de l'histoire de « La poupée rouge » (Gripari, 1994).
- Répétition des six phrases issues de Monpage.
- Tenue de /s/ et de /z/.
- Lecture de phrases issues du protocole Montréal d'Évaluation de la Communication (MEC), le sujet doit produire une même phrase, avec trois émotions différentes :
  - P1 : « mon mari a brisé mon vase » joyeux (J1), en colère (C1), et triste (T1)
  - P2 : « j'ai vu mon médecin aujourd'hui » joyeux (J2), en colère (C2), et triste (T2)
  - P3 : « J'ai reçu une lettre de mon père » joyeux (J3), en colère (C3) et triste (T3)
- Chant : « Joyeux anniversaire ».
- Répétition du dialogue de l'histoire de « La poupée rouge » (Pierre Gripari).
- Epreuve perceptive sur quatre phrases : le sujet entend-il une phrase montante, descendante, ou les deux ?
- Epreuve perceptive de douze phrases : le sujet doit désigner l'émotion correspondante parmi la joie, la colère ou la tristesse.
- /e/ aigu tenu sur 3 secondes.

## Annexe B

Il était une fois, un petit garçon paresseux. Ça l'ennuyait d'aller à l'école, ça l'ennuyait d'apprendre, ça l'ennuyait de lire, d'écrire, de compter. Ce qu'il voulait, c'était ne rien faire, jamais, et cependant, devenir riche quand même, ce qui n'est pas facile. Un jour, vint à passer le petit diable verdissant, qui s'occupe des enfants désobéissants. Sitôt qu'il aperçoit ce petit fainéant, il frétille de joie et murmure :

« Ça, c'est un bon client pour moi ! » Puis il appelle à haute voix :

- Petit garçon ! Petit garçon !
- Oui, Monsieur le démon ?
- Tu es fort paresseux, je crois ?
- Mais pas du tout ! Je travaille, je travaille beaucoup !
- Et menteur avec ça, enchaîne le petit diable enchanté, tout ce que j'aime en vérité.  
Eh bien, puisqu'il en est ainsi, petit paresseux, moi, je vais te rendre heureux. »

Extrait de l'histoire de « La poupée rouge » de Pierre Gripari (1994).

## **Voix de patientes avec suspicion d'atteinte du nerf laryngé supérieur : rythme et perception – 24 pages – 22 références bibliographiques**

### **Résumé**

Après thyroïdectomie totale, la plainte vocale des patients, hors paralysie récurrentielle, est attribuée à l'atteinte du nerf laryngé supérieur, encore difficilement objectivable. Des paramètres rythmiques altérés pourraient servir au diagnostic, complétés par une évaluation perceptive et un questionnaire de ressenti du patient. 28 femmes avec suspicion d'atteinte du NLS (CT-) ont été comparées à 27 autres sans dommage (CT+) après thyroïdectomie totale, et une population témoin (T). Parmi les paramètres rythmiques étudiés, %V est le seul à être altéré chez les CT- comparé aux CT+ et T, témoin d'une mobilisation des plis vocaux plus complexe. Cette conclusion complète les constatations déjà établies : la réduction de la plage vocale et la fatigue vocale. L'évaluation perceptive à partir de l'échelle GRBAS-I ne permet pas de discriminer les voix des 3 populations. La voix est cependant perçue comme moins modulée chez les CT- et CT+. L'auto-évaluation des patientes confirme le ressenti d'un handicap vocal. Cette étude innovante sur le rythme et l'évaluation perceptive des CT-, a tout son intérêt pour répondre aux attentes des patients et des professionnels de la voix.

***Mots-clés :** Nerf laryngé supérieur ; thyroïdectomie ; rythme ; évaluation perceptive ; qualité vocale*

### **Abstract**

After total thyroidectomy, the vocal complaint of patients, excluding recurrent paralysis, is attributed to damage on the superior laryngeal nerve, which is still difficult to objectify. Altered rhythmic parameters could be used as a diagnosis, supplemented by a perceptual evaluation and a questionnaire on the patient's feelings. 28 women with suspected SLN injury (CT-) were compared to 27 others without damage (CT+) after total thyroidectomy, and a control population (T). Among the rhythmic parameters studied on vocal productions, % V is the only one to be altered in CT- compared to CT+ and T, witnessing a more complex mobilization of the vocal folds. This conclusion complements the findings already established: reduction of the vocal range and vocal fatigue. Perceptual assessment from the GRBAS-I scale does not allow the voices of the 3 populations to be discriminated. The voice is however perceived to be less modulated in CT- and CT+. Patient self-assessment confirms the feeling of a severe vocal handicap for a few. Meeting the expectations of patients and voice professionals, this innovative study on the rhythm and the perceptual evaluation of CT- gets its whole interest.

***Keywords:** Superior laryngeal nerve; thyroidectomy; rhythm; perceptual evaluation; voice quality*