

## **L'AVQI 03.01 appliqué à la langue française : validité et précision diagnostique dans la population adulte. Cette évaluation objective corrèle-t-elle avec la plainte du patient ?**

**Auteur :** Pommée, Timothy

**Promoteur(s) :** MORSOMME, Dominique; 4706

**Faculté :** Faculté de Psychologie, Logopédie et Sciences de l'Éducation

**Diplôme :** Master en logopédie, à finalité spécialisée en voix

**Année académique :** 2017-2018

**URI/URL :** <http://hdl.handle.net/2268.2/4565>

---

*Avertissement à l'attention des usagers :*

*Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.*

*Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.*

---



L'AVQI 03.01 appliqué à la langue française : validité  
et précision diagnostique dans la population adulte.  
Cette évaluation objective corrèle-t-elle avec la plainte  
du patient ?

Mémoire rédigé par POMMÉE Timothy  
Encadré par Mme MORSOMME Dominique  
et M. MARYN Youri (co-promoteur)

Année académique : 2017-2018  
Master en logopédie, à finalité spécialisée en Voix

*Differing statements about the nature of things, differing medical philosophies, different diagnoses and treatments—all of these are to be collected eagerly and be made a part of the individual's collection of puzzle pieces. Ultimately, after many lifetimes, the pieces will fit together and the individual will attain clear and certain knowledge.*

*BEALS, 1976*

*Dans le cadre d'un phénomène complexe, multi-dimensionnel comme la voix, toute mesure isolée est réductrice ; il n'existera aucun paramètre dont la mesure rende compte de toutes les dimensions de la dysphonie.*

*GIOVANNI ET AL., 1996*

*Never trust a computer completely.*

*BAKEN & ORLIKOFF, 2000*

## Table des matières

### PARTIE I : CONTEXTE THÉORIQUE

LISTE DES TABLEAUX.....	V
LISTE DES FIGURES .....	VIII
REMERCIEMENTS .....	X
INTRODUCTION .....	1
LA PHONATION, LA VOIX ET LA QUALITÉ VOCALE.....	4
I. LA PHONATION ET LA VOIX .....	4
II. LA MODULATION DU SON VOCAL.....	5
A. <i>La régulation de la fréquence</i> .....	5
B. <i>La régulation du niveau de pression sonore</i> .....	9
C. <i>Les cavités de résonance</i> .....	11
III. LA QUALITÉ VOCALE .....	13
LES DYSPHONIES.....	14
L'ÉVALUATION VOCALE — ÉTAT DE L'ART .....	16
I. L'ANAMNÈSE .....	17
II. L'IMAGERIE LARYNGÉE : LA VIDÉOSTROSCOPIE .....	18
III. LES MESURES SUBJECTIVES .....	19
A. <i>L'évaluation perceptive par le thérapeute</i> .....	19
a. <i>GRBAS (Hirano, 1981)</i> .....	19
b. <i>Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) (Kempster, Gerratt, Abbott, Barkmeier-Kraemer, &amp; Hillman, 2009)</i> .....	20
c. <i>Les limites de l'évaluation perceptive</i> .....	22
B. <i>L'évaluation par le patient : impact des troubles au quotidien (approche écologique)</i> .....	23
a. <i>VHI (Jacobson et al., 1997)</i> .....	23
b. <i>Voice Symptom Scale (VoiSS) (Deary, Wilson, Carding, &amp; MacKenzie, 2003)</i> .....	25
IV. LES MESURES OBJECTIVES .....	26
A. <i>Les mesures aérodynamiques</i> .....	26
B. <i>Les mesures acoustiques</i> .....	27
a. <i>Dysphonia Severity Index (Wuyts et al., 2000)</i> .....	28
b. <i>Le modèle d'Awan et Roy : Cepstral Spectral Index of Dysphonia (CSID)</i> .....	29
c. <i>Les limites des mesures acoustiques</i> .....	31
V. LES MESURES ACOUSTIQUES ET L'ÉVALUATION SUBJECTIVE PAR LE PATIENT SONT-ELLES CORRÉLÉES ? .....	31
L'AVQI – ACOUSTIC VOICE QUALITY INDEX .....	34
I. L'ÉVOLUTION DE L'AVQI ET LES ÉTUDES DE VALIDATION .....	39
II. L'AVQI, UN OUTIL MULTILINGUE.....	42

### PARTIE II : L'AVQI 03.01 APPLIQUÉ AU FRANÇAIS

L'ÉCHANTILLON EN FRANÇAIS : ANALYSES PRÉLIMINAIRES.....	46
I. L'ÉQUILIBRAGE PHONÉTIQUE DE L'EXTRAIT DU TEXTE « LA BISE ET LE SOLEIL » .....	47
II. L'UTILISATION D'UN EXTRAIT DU TEXTE DE HARMEGNIES.....	49
L'ÉCHANTILLON EN FRANÇAIS : APPLICATION SUR UNE POPULATION PATHOLOGIQUE .....	56
I. LA POPULATION ET LES ENREGISTREMENTS .....	56
II. NOS QUESTIONS INITIALES .....	59
III. LA DURÉE VOISÉE EN SECONDES DE 27 VS 32 SYLLABES DE PAROLE CONTINUE.....	61
IV. LES EFFETS PRINCIPAUX ET LES EFFETS SIMPLES .....	62



A. Les analyses sur scores bruts .....	63
B. Le classement « normophonie-dysphonie » .....	65
<b>V. SYNTHÈSE DES RÉSULTATS .....</b>	<b>67</b>
<b>VI. DISCUSSION .....</b>	<b>68</b>
A. Concernant les scores bruts .....	68
B. Concernant le classement « normophonie – dysphonie » .....	69
C. Concernant les durées syllabiques des échantillons .....	70
<b>LA VALIDITÉ CONCORANTE ET LA PRÉCISION DIAGNOSTIQUE DE L’AVQI 03.01 APPLIQUÉ AU FRANÇAIS .....</b>	<b>71</b>
<b>I. LA POPULATION .....</b>	<b>72</b>
<b>II. LES ENREGISTREMENTS .....</b>	<b>75</b>
<b>III. L’INTERFACE D’ÉVALUATION .....</b>	<b>76</b>
<b>IV. LES SESSIONS D’ÉVALUATION PERCEPTIVE .....</b>	<b>77</b>
A. Les juges .....	77
B. La tâche .....	78
<b>V. LA FIDÉLITÉ INTER-JUGES ET INTRA-JUGE .....</b>	<b>81</b>
<b>VI. LA VALIDITÉ CONCORANTE .....</b>	<b>83</b>
<b>VII. LE SCORE SEUIL OPTIMAL ET LA PRÉCISION DIAGNOSTIQUE .....</b>	<b>84</b>
<b>VIII. LES PARTICIPANTS NORMOPHONIQUES... DYSPHONIQUES ? .....</b>	<b>87</b>
<b>IX. RÉITÉRATION DES ANALYSES SUR LA PAROLE CONTINUE UNIQUEMENT .....</b>	<b>88</b>
A. La fidélité inter-juges et intra-juge .....	88
B. La validité concourante .....	88
C. Le score seuil optimal et la précision diagnostique .....	89
<b>X. SYNTHÈSE DES RÉSULTATS .....</b>	<b>93</b>
<b>L’AVQI ET L’ÉVALUATION DE LA QUALITÉ DE VIE PAR LE PATIENT .....</b>	<b>94</b>
<b>I. L’EBP ET LA CIF .....</b>	<b>94</b>
<b>II. L’AVQI ET LE VHI .....</b>	<b>96</b>
A. Notre questionnaire .....	97
B. La population .....	97
C. Les corrélations AVQI – VHI .....	99
<b>III. SYNTHÈSE DES RÉSULTATS .....</b>	<b>102</b>
<b>DISCUSSION GÉNÉRALE .....</b>	<b>103</b>
<b>RÉFLEXIONS .....</b>	<b>111</b>
<b>PERSPECTIVES DE RECHERCHE .....</b>	<b>115</b>
<b>CONCLUSION GÉNÉRALE .....</b>	<b>118</b>
<b>BIBLIOGRAPHIE .....</b>	<b>120</b>

## ANNEXES

ANNEXE 1 – LE CEPSTRE, KÉZAKO?! .....	150
ANNEXE 2 – ÉTAPES DE L’AVQI .....	160
ANNEXE 3 – EXEMPLE D’APPLICATION DU PROTOCOLE D’ÉVALUATION DE L’ELS .....	161
ANNEXE 4 – CAPE-V .....	162
ANNEXE 5 – PHRASES DE COMBESURE .....	163
ANNEXE 6 – SCRIPT DE L’AVQI 02.03 .....	165
ANNEXE 7 – SCRIPT DE L’AVQI 03.01 .....	171
ANNEXE 8 – PROCESSUS DE SÉLECTION DE LA LONGUEUR SYLLABIQUE OPTIMALE POUR L’ÉTUDE DE VALIDATION DE L’AVQI 03.01 EN FRANÇAIS .....	177
ANNEXE 9 – SCRIPT COMPOSÉ POUR CRÉER LES FICHIERS DE L’ÉVALUATION PERCEPTIVE .....	178
ANNEXE 10 – SCRIPT ADAPTÉ POUR L’ÉVALUATION PERCEPTIVE DU CRITÈRE G DE LA GRBAS ....	181

## Liste des tableaux

TABLEAU 1 CORRESPONDANCE ENTRE LE SCORE G DE LA GRBAS ET LE SCORE CALCULÉ PAR LE DSI .....	28
TABLEAU 2 MESURES PSYCHOMÉTRIQUES DES ÉTUDES DE VALIDATION DE L'AVQI DANS DIFFÉRENTES LANGUES	42
TABLEAU 3 FRÉQUENCES D'OCCURRENCE DES PHONÈMES DANS LA PREMIÈRE PHRASE DU TEXTE « LA BISE ET LE SOLEIL » .....	47
TABLEAU 4 FRÉQUENCES D'OCCURRENCE DES PHONÈMES DANS L'EXTRAIT DU TEXTE « LA BISE ET LE SOLEIL » ET DANS LA LANGUE FRANÇAISE.....	48
TABLEAU 5 TESTS DE NORMALITÉ POUR LES OCCURRENCES DANS L'EXTRAIT DU TEXTE « LA BISE ET LE SOLEIL » ET DANS LA LANGUE FRANÇAISE .....	48
TABLEAU 6 CORRÉLATION DE SPEARMAN ENTRE LES FRÉQUENCES D'OCCURRENCE DANS L'EXTRAIT DU TEXTE « LA BISE ET LE SOLEIL » ET DANS LA LANGUE FRANÇAISE .....	48
TABLEAU 7 TEST DE WILCOXON COMPARANT LES MÉDIANES DES FRÉQUENCES D'OCCURRENCE DE L'EXTRAIT DU TEXTE « LA BISE ET LE SOLEIL » ET DE LANGUE FRANÇAISE .....	49
TABLEAU 8 COMPARAISON DES FRÉQUENCES D'OCCURRENCE DES PHONÈMES DANS LE PREMIER EXTRAIT DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	51
TABLEAU 9 TESTS DE NORMALITÉ POUR LES OCCURRENCES DANS LE PREMIER EXTRAIT DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	51
TABLEAU 10 CORRÉLATION DE SPEARMAN ENTRE LES FRÉQUENCES D'OCCURRENCE DES PHONÈMES VOISÉS DANS LE PREMIER EXTRAIT DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE.....	52
TABLEAU 11 COMPARAISON DES FRÉQUENCES D'OCCURRENCE DES PHONÈMES DANS LE SECOND EXTRAIT DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	52
TABLEAU 12 TESTS DE NORMALITÉ POUR LES OCCURRENCES DANS LE SECOND EXTRAIT DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	52
TABLEAU 13 CORRÉLATION DE SPEARMAN ENTRE LES FRÉQUENCES D'OCCURRENCE DES PHONÈMES VOISÉS DANS LE SECOND EXTRAIT DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	53
TABLEAU 14 COMPARAISON DES FRÉQUENCES D'OCCURRENCE DES PHONÈMES DANS LES 32 PREMIÈRES SYLLABES DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE.....	54
TABLEAU 15 CORRÉLATION DE SPEARMAN ENTRE LES FRÉQUENCES D'OCCURRENCE DES PHONÈMES VOISÉS DANS LES 32 PREMIÈRES SYLLABES DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	55
TABLEAU 16 COMPARAISON DES OCCURRENCES DES PHONÈMES DANS LES 27 PREMIÈRES SYLLABES DU TEXTE ET DANS LA LANGUE FRANÇAISE.....	55
TABLEAU 17 CORRÉLATION DE SPEARMAN ENTRE LES FRÉQUENCES D'OCCURRENCE DES PHONÈMES VOISÉS DANS LES 27 PREMIÈRES SYLLABES DU TEXTE DE HARMEGNIES ET DANS LA LANGUE FRANÇAISE .....	56
TABLEAU 18 CARACTÉRISTIQUES PRINCIPALES DE L'ÉCHANTILLON DE 162 PATIENTS DU CHU DE LIÈGE .....	58
TABLEAU 19 STATISTIQUES DESCRIPTIVES DES DURÉES DES ÉCHANTILLONS DE 27 ET DE 32 SYLLABES .....	61
TABLEAU 20 TESTS DE NORMALITÉ POUR LES ÉCHANTILLONS DE 27 (GAUCHE) ET 32 SYLLABES (DROITE).....	62

TABLEAU 21 TEST DE WILCOXON SUR LES MÉDIANES DES ÉCHANTILLONS DE 27 ET DE 32 SYLLABES.....	62
TABLEAU 22 TESTS DE NORMALITÉ POUR LES ÉCHANTILLONS DE 32 SYLLABES, ANALYSÉS PAR L'AVQI 02.03 .....	63
TABLEAU 23 TESTS DE NORMALITÉ POUR LES ÉCHANTILLONS DE 32 SYLLABES, ANALYSÉS PAR L'AVQI 03.01 .....	63
TABLEAU 24 TESTS DE NORMALITÉ POUR LES ÉCHANTILLONS DE 27 SYLLABES, ANALYSÉS PAR L'AVQI 02.03 .....	63
TABLEAU 25 TESTS DE NORMALITÉ POUR LES ÉCHANTILLONS DE 27 SYLLABES, ANALYSÉS PAR L'AVQI 03.01 .....	63
TABLEAU 26 TEST DE FRIEDMAN SUR LES SCORES BRUTS DES DEUX VERSIONS DE L'AVQI POUR LES DEUX DURÉES SYLLABIQUES .....	63
TABLEAU 27 TEST DE WILCOXON SUR LES MÉDIANES DES SCORES BRUTS CALCULÉS SUR LES ÉCHANTILLONS DE 27 ET DE 32 SYLLABES (LES DEUX VERSIONS DE L'AVQI CONFONDUES) .....	64
TABLEAU 28 TEST DE WILCOXON SUR LES MÉDIANES DES SCORES BRUTS CALCULÉS PAR LES DEUX VERSIONS DE L'AVQI (DURÉES SYLLABIQUES CONFONDUES).....	64
TABLEAU 29 TEST DE WILCOXON SUR LES MÉDIANES DES SCORES BRUTS CALCULÉS PAR LES DEUX VERSIONS DE L'AVQI SUR LES ÉCHANTILLONS DE 32 SYLLABES .....	65
TABLEAU 30 TEST DE WILCOXON SUR LES MÉDIANES DES SCORES BRUTS CALCULÉS PAR LES DEUX VERSIONS DE L'AVQI SUR LES ÉCHANTILLONS DE 27 SYLLABES .....	65
TABLEAU 31 TEST DE WILCOXON SUR LES MÉDIANES DES SCORES BRUTS CALCULÉS SUR LES ÉCHANTILLONS DE 27 ET DE 32 SYLLABES PAR LA VERSION 02.03 DE L'AVQI .....	65
TABLEAU 32 TEST DE WILCOXON SUR LES MÉDIANES DES SCORES BRUTS CALCULÉS SUR LES ÉCHANTILLONS DE 27 ET DE 32 SYLLABES PAR LA VERSION 03.01 DE L'AVQI .....	65
TABLEAU 33 EFFET PRINCIPAL « LONGUEUR SYLLABIQUE » : MODES ET MÉDIANES DES SCORES BRUTS PAR LONGUEUR D'ÉCHANTILLON (INDÉPENDAMMENT DES VERSIONS AVQI).....	65
TABLEAU 34 EFFET PRINCIPAL « VERSION AVQI » : MODES ET MÉDIANES DES SCORES BRUTS PAR VERSION (INDÉPENDAMMENT DES LONGUEURS D'ÉCHANTILLONS) .....	65
TABLEAU 35 EFFETS SIMPLES : MODES ET MÉDIANES DES SCORES BRUTS POUR LES 4 CROISEMENTS .....	65
TABLEAU 36 TEST DE WILCOXON SUR LES JUGEMENTS « NORMOPHONIQUE/DYSPHONIQUE » POUR LES ÉCHANTILLONS DE 27 ET DE 32 SYLLABES .....	66
TABLEAU 37 TEST DE WILCOXON SUR LES JUGEMENTS « NORMOPHONIQUE/DYSPHONIQUE » POUR LES DEUX VERSIONS DE L'AVQI .....	66
TABLEAU 38 POURCENTAGES DES CLASSEMENTS SELON LA LONGUEUR SYLLABIQUE DE L'ÉCHANTILLON .....	66
TABLEAU 39 POURCENTAGES DES CLASSEMENTS SELON LA VERSION DE L'AVQI .....	66
TABLEAU 40 EFFET PRINCIPAL « VERSION AVQI » : MODES ET MÉDIANES DES SCORES BRUTS PAR VERSION (INDÉPENDAMMENT DES LONGUEURS D'ÉCHANTILLONS) .....	68
TABLEAU 41 EFFET PRINCIPAL « LONGUEUR SYLLABIQUE » : MODES ET MÉDIANES DES SCORES BRUTS PAR LONGUEUR D'ÉCHANTILLON (INDÉPENDAMMENT DES VERSIONS AVQI).....	68
TABLEAU 42 EFFETS SIMPLES : MODES ET MÉDIANES DES SCORES BRUTS POUR LES 4 CROISEMENTS .....	68
TABLEAU 43 CRITÈRES D'ATTRIBUTION CATÉGORIELLE DES ÉCHANTILLONS SONORES EN FONCTION DU SCORE BRUT DE L'AVQI .....	72

TABLEAU 44 CARACTÉRISTIQUES PRINCIPALES DE L'ÉCHANTILLON DE 30 PARTICIPANTS DU GROUPE CONTRÔLE	73
TABLEAU 45 CARACTÉRISTIQUES PRINCIPALES DE L'ÉCHANTILLON TOTAL DE 120 PATIENTS (90 PATIENTS DE LA BASE DE DONNÉES ET 30 PARTICIPANTS CONTRÔLES)	73
TABLEAU 46 INTERPRÉTATION DU COEFFICIENT KAPPA SELON LANDIS ET KOCH (1977)	82
TABLEAU 47 COEFFICIENTS KAPPA PONDÉRÉS POUR L'ÉVALUATION DE LA FIDÉLITÉ INTRA-JUGE DES JUGEMENTS PERCEPTIFS SUR LA PAROLE CONTINUE ET LA VOYELLE TENUE	82
TABLEAU 48 COEFFICIENT DE CONCORDANCE DE KENDALL POUR L'ÉVALUATION DE LA FIDÉLITÉ INTER-JUGES SUR LA PAROLE CONTINUE ET LA VOYELLE TENUE	82
TABLEAU 49 INTERPRÉTATION DU COEFFICIENT DE CONCORDANCE DE KENDALL SELON SCHMIDT 1997 (CITÉ PAR REMØY, 2010)	82
TABLEAU 50 CORRÉLATION DE SPEARMAN ENTRE LES SCORES AVQI ET LES ÉVALUATIONS MOYENNES DU DEGRÉ DE SÉVÉRITÉ (CRITÈRE G) ÉVALUÉ PAR LES QUATRE VOCOLOGISTES SUR BASE DE LA PAROLE CONTINUE ET DE LA VOYELLE TENUE	83
TABLEAU 51 TEST DE SIGNIFICATIVITÉ DE L' Aire sous la courbe ROC SELON LES ÉVALUATIONS PERCEPTIVES SUR BASE DE LA PAROLE CONTINUE ET DE LA VOYELLE TENUE	85
TABLEAU 52 COEFFICIENTS KAPPA PONDÉRÉS POUR L'ÉVALUATION DE LA FIDÉLITÉ INTRA-JUGE DES JUGEMENTS PERCEPTIFS SUR LA PAROLE CONTINUE	88
TABLEAU 53 COEFFICIENT DE CONCORDANCE DE KENDALL POUR L'ÉVALUATION DE LA FIDÉLITÉ INTER-JUGES DES JUGEMENTS PERCEPTIFS SUR LA PAROLE CONTINUE	88
TABLEAU 54 CORRÉLATION DE SPEARMAN ENTRE LES SCORES AVQI ET LES ÉVALUATIONS MOYENNES DU DEGRÉ DE SÉVÉRITÉ (CRITÈRE G) ÉVALUÉ PAR LES QUATRE VOCOLOGISTES SUR BASE DE LA PAROLE CONTINUE	89
TABLEAU 55 TEST DE SIGNIFICATIVITÉ DE L' Aire sous la courbe ROC SELON LES ÉVALUATIONS PERCEPTIVES SUR BASE DE LA PAROLE CONTINUE	90
TABLEAU 56 CALCUL DE LA SENSIBILITÉ ET DE LA SPÉCIFICITÉ DE L'AVQI 03.01 EN FONCTION DU SCORE SEUIL	92
TABLEAU 57 CARACTÉRISTIQUES PRINCIPALES DE L'ÉCHANTILLON DE 78 PATIENTS	98
TABLEAU 58 ATTRIBUTION DES VALEURS 0 OU 1 EN FONCTION DU SCORE AVQI 03.01 ET DU SCORE VHI	101
TABLEAU 59 TABLE DE CONTINGENCE POUR L'ANALYSE DE L'ASSOCIATION ENTRE VERDICTAVQI ET VERDICT VHI	101
TABLEAU 60 LIGNES DIRECTRICES POUR L'INTERPRÉTATION DU COEFFICIENT PHI DE CRAMER SELON COHEN (1988) ET REA & PARKER (1992) (KOTRLIK, WILLIAMS & KATHA JABOR, 2011)	101
TABLEAU 61 POURCENTAGES DES ENREGISTREMENTS OBTENANT DES SCORES SÉVÈRES OU NON, SELON L'AVQI 03.01 ET SELON L'ÉVALUATION PERCEPTIVE MOYENNE	108

## Liste des figures

FIGURE 1. VUE DE FACE DES TROIS ÉTAGES LARYNGÉS (ADAPTÉ DE CORNUT, 2014).....	5
FIGURE 2. LES CINQ COUCHES DES PLIS VOCAUX (ADAPTÉ DE CORNUT, 2014). ....	6
FIGURE 3. CONFIGURATION DES PLIS VOCAUX EN FONCTION DE LA CONTRACTION DES MUSCLES ANTAGONISTES THYRO-ARYTÉNOÏDIEN ET CRICO-THYROÏDIEN (STORY & TITZE, 1995).....	7
FIGURE 4. REPRÉSENTATION DE LA VARIATION DE LA PRESSION SOUS-GLOTTIQUE EN FONCTION DE LA FRÉQUENCE FONDAMENTALE (SUNDBERG, 1992). ....	8
FIGURE 5. REPRÉSENTATION DU « SINGING FORMANT » (SUNDBERG, 1972).....	10
FIGURE 6. REPRÉSENTATION DE LA VARIATION DE LA PRESSION SOUS-GLOTTIQUE EN FONCTION DU NIVEAU DE PRESSION SONORE (SUNDBERG, 1992).....	11
FIGURE 7. REPRÉSENTATION SCHÉMATIQUE DE L'INFLUENCE DES CAVITÉS DE RÉSONANCE SUR LE SIGNAL ACOUSTIQUE LARYNGÉ (RUBIN & VATIKIOTIS-BATESON, 1998). ....	11
FIGURE 8. REPRÉSENTATION SPECTRALE D'UN SON VOCAL (CORNUT, 2014).....	12
FIGURE 9. SPECTROGRAMME DU MOT « EFFECTIVEMENT ».....	12
FIGURE 10. TRIANGLE VOCALIQUE DU FRANÇAIS (MÉCHAIN, 2014). ....	13
FIGURE 11. ILLUSTRATION DES VARIATIONS PÉRIODIQUES D'AMPLITUDE (VERSTRAETE, 2017). ....	36
FIGURE 12. REPRÉSENTATION DU « SPECTRAL TILT » (DELGADO, LEÓN, JIMÉNEZ, & IZQUIERDO, 2017).....	37
FIGURE 13. REPRÉSENTATION DU PIC CEPSTRAL (MARYN, DE BODT, BARSTIES, & ROY, 2014). ....	38
FIGURE 14. RÉSULTATS DE L'ENQUÊTE EN LIGNE QUANT À LA QUESTION DE L'UTILISATION DE L'AVQI. ....	45
FIGURE 15. CROISEMENTS DE DONNÉES DISPONIBLES POUR LES ANALYSES STATISTIQUES.....	60
FIGURE 16. VISUALISATION DE LA CONCATÉNATION DES SEGMENTS VOISÉS EXTRAITS DE LA PAROLE CONTINUE ET DES TROIS SECONDES DE LA VOYELLE [A] TENUE PAR UN SUJET SAIN. ....	61
FIGURE 17. EXEMPLE DE FICHIER AUDIO CRÉÉ GRÂCE AU SCRIPT COMPOSÉ. ....	76
FIGURE 18. ILLUSTRATION DE L'INTERFACE D'ÉVALUATION PERCEPTIVE DES CRITÈRES DE LA GRBAS, SELON LE SCRIPT DE VALÉRIE DELVAUX (2016). ....	77
FIGURE 19. ILLUSTRATION DE LA NOUVELLE INTERFACE POUR L'ÉVALUATION DU CRITÈRE G DE LA GRBAS. ....	77
FIGURE 20. FACTEURS INFLUENÇANT L'ÉVALUATION PERCEPTIVE DE LA QUALITÉ VOCALE SELON KREIMAN ET AL. (1993). ....	80
FIGURE 21. COURBE ROC SELON LES ÉVALUATIONS PERCEPTIVES SUR BASE DE LA PAROLE CONTINUE ET DE LA VOYELLE TENUE. ....	84
FIGURE 22. INDEX DE YOUNDEN ET DIFFÉRENCE MINIMALE SENSIBILITÉ-SPÉCIFICITÉ SUR UNE COURBE ROC (HABIBZADEH, HABIBZADEH, & YADOLLAHIE, 2016). ....	85
FIGURE 23. LES SCORES SEUILS OPTIMAUX SELON TROIS MÉTHODES SUR LA COURBE ROC SELON LES ÉVALUATIONS PERCEPTIVES SUR BASE DE LA PAROLE CONTINUE ET DE LA VOYELLE TENUE. ....	86
FIGURE 24. COURBE ROC SELON LES ÉVALUATIONS PERCEPTIVES SUR BASE DE LA PAROLE CONTINUE.....	89

FIGURE 25. LES SCORES SEUILS OPTIMAUX SELON TROIS MÉTHODES SUR LA COURBE ROC SELON LES ÉVALUATIONS PERCEPTIVES SUR BASE DE LA PAROLE CONTINUE.....	91
FIGURE 26. LES TROIS COMPOSANTES PRINCIPALES DE L'EBP (TRADUIT DE AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION, 2017B). .....	94
FIGURE 27. CODAGE DE LA COMPOSANTE « STRUCTURES ET FONCTIONS ANATOMIQUES » DE LA CIF (ORGANISATION MONDIALE DE LA SANTÉ, 2001). .....	95
FIGURE 28. CODAGE DE LA COMPOSANTE « ACTIVITÉS ET PARTICIPATION » DE LA CIF (ORGANISATION MONDIALE DE LA SANTÉ, 2001). .....	95
FIGURE 29. REPRÉSENTATION CEPSTRALE DE LA VOYELLE [A] (MARYN, DICK, VANDENBRUAENE, VAUTERIN, & JACOBS, 2009). .....	150
FIGURE 30. EXEMPLES DE CEPSTRES ET CEPSTROGRAMMES D'UN [A] TENU (DELGADO-HERNÁNDEZ, LEÓN-GÓMEZ, IZQUIERDO-ARTEAGA, & LLANOS-FUMERO, 2018). .....	151
FIGURE 31. DÉCOMPOSITION D'UNE ONDE COMPLEXE EN SES DIFFÉRENTS COMPOSANTS (LADEFOGED, 1995). ...	152
FIGURE 32. EXEMPLE D'UN SPECTRE (HEMAN-ACKAH, MICHAEL, & GODING, 2002). .....	152
FIGURE 33. REPRÉSENTATION TEMPORELLE, SPECTRALE ET CEPSTRALE DU SON (RANDALL, 1973). .....	153
FIGURE 34. REPRÉSENTATION SCHÉMATIQUE DE L'ENVELOPPE SPECTRALE ET DES DÉTAILS SPECTRAUX (NOLL, 1967). .....	154
FIGURE 35. DÉCOMPOSITION DU SPECTRE EN L'ENVELOPPE SPECTRALE ET LES DÉTAILS SPECTRAUX (PRAHALLAD, 2008). .....	155
FIGURE 36. REPRÉSENTATIONS SCHÉMATIQUES SPECTRALE ET CEPSTRALE D'UNE ONDE SONORE PÉRIODIQUE (NOLL, 1967). .....	155
FIGURE 37. REPRÉSENTATION SCHÉMATIQUE DE LA TRANSFORMATION INVERSE DE FOURRIER EFFECTUÉE SUR LE SPECTRE (PRAHALLAD, 2008). .....	156
FIGURE 38. REPRÉSENTATION SPECTRALE ET CEPSTRALE D'UN SON DE 80 HZ (OPPENHEIM & SCHAFFER, 2004). ...	157
FIGURE 39. EXEMPLE DE REPRÉSENTATION CEPSTRALE AVEC LA DROITE DE RÉGRESSION ET LA PROÉMINENCE DU PIC CEPSTRAL (HEMAN-ACKAH, MICHAEL, & GODING, 2002). .....	158
FIGURE 40. REPRÉSENTATION DE LA DÉTERMINATION DE LA PROÉMINENCE DU PIC CEPSTRAL (MARYN, DE BODT, BARSTIES, & ROY, 2014). .....	159

## Remerciements

J'adresse mes remerciements à tous ceux qui ont rendu possible la réalisation de ce travail.

En premier lieu, je remercie Madame Morsomme, ma promotrice pour ce Travail de Fin d'Études, pour le temps qu'elle m'a accordé, pour ses conseils avisés et surtout pour sa patience. Je lui suis également très reconnaissant de parvenir, malgré la charge de travail qui lui incombe, à toujours garder le sourire.

Je remercie également Monsieur Maryn, co-promoteur de ce travail.

Mes remerciements sont aussi adressés à Mesdames Delacroix, Dubru, Simon et Winters pour avoir accepté sans hésiter de participer à ce travail, pour le sérieux avec lequel elles se sont attelées à la tâche d'évaluation perceptive, ainsi que pour leur compréhension face aux imprévus qui ont augmenté leur charge de travail.

Un grand merci aussi à tous les membres de l'équipe du service ORL du CHU de Liège, qui ont été très compréhensifs et m'ont permis d'utiliser le laboratoire vocal à de nombreuses reprises, malgré leurs horaires chargés.

Je tiens aussi à exprimer ma gratitude envers les membres d'autres services du CHU qui ont accepté de me consacrer de leur temps pour la constitution du groupe contrôle.

Enfin, je tiens également à exprimer mes remerciements à tous mes proches, mes amis et collègues qui m'ont aidé de près ou de loin à mener à bien ce travail, et qui m'ont épaulé dans les bons comme dans les mauvais instants.

## Introduction

La voix est un phénomène à la fois fascinant et complexe, ses multiples déterminants rendant son ensemble difficile à saisir, à quantifier. En logopédie de la voix, la qualité vocale est évaluée sur un plan subjectif via des échelles perceptives (par le thérapeute) et des questionnaires d'auto-évaluation (par le patient), et sur un plan objectif via des mesures acoustiques et aérodynamiques. Ces dernières années, avec l'essor de la technologie et de l'informatique, sont apparus des index de mesure de la qualité vocale. Ces index incluent différents paramètres objectifs dans leur calcul et permettent ainsi de quantifier le degré de sévérité de la dysphonie du patient.

Depuis 2013, la législation belge prévoit un remboursement, par l'I.N.A.M.I., de quatre-vingts séances de rééducation vocale logopédique, sur base d'une évaluation complète de la voix du patient. Cette évaluation comprend un certain nombre de tests repris et décrits dans la liste limitative de l'I.N.A.M.I., et doit notamment démontrer au moins une mesure objective pathologique. Pour ce faire, deux index de sévérité de la dysphonie sont cités dans la liste : le Dysphonia Severity Index (DSI) et l'Acoustic Voice Quality Index (AVQI). L'AVQI, développé par Youri Maryn, est un script utilisable sous le logiciel *Praat* qui inclut dans son calcul six paramètres acoustiques. Il permet, sur base de l'enregistrement d'une voyelle tenue et d'un extrait de texte, de calculer un score unique allant de 0 à 10, sachant qu'un score de 0 jusqu'aux alentours de 3 signifie que la voix est non pathologique (le score exact dépend de la langue utilisée), tandis qu'elle devient pathologique au-delà de ce seuil. Les caractéristiques psychométriques de cet outil ont fait l'objet de nombreuses publications, confirmant la validité, la fidélité et la précision diagnostique de l'outil, et ce dans de nombreuses langues. Cependant, sa validation en français repose sur un nombre réduit de patients, qui plus est, non francophones de naissance. De ce fait, la plupart des vocologues en Belgique francophone préfèrent utiliser le DSI pour leurs bilans vocaux. Or, comme nous allons le décrire dans la première partie de ce travail, l'AVQI présente certains avantages par rapport à ce dernier.

Aussi, notre travail a consisté en quatre étapes :

Premièrement, il s'agissait de vérifier que le texte couramment utilisé en clinique francophone serait utilisable pour l'analyse de l'AVQI, et de déterminer l'extrait de ce texte à utiliser pour les analyses.

Deuxièmement, nous avons testé la validité d'utilisation de l'AVQI 03.01 en français sur une population euphonique et dysphonique. Pour cela, quatre vocologues ont effectué un jugement



perceptif du degré général de la dysphonie à l'aide du critère G de la GRBAS, sur 120 enregistrements vocaux de la base de données du CHU de Liège.

Troisièmement, une fois la validité concourante confirmée, nous avons déterminé le score seuil idéal pour l'échantillon en langue française. Ce score correspond au seuil permettant de détecter au mieux les voix pathologiques et de rejeter un maximum de voix normales. La précision diagnostique de l'AVQI 03.01 avec ce score seuil a ensuite été décrite, en calculant le pourcentage de sujets dysphoniques correctement identifiés par l'outil (sensibilité) et le pourcentage de sujets sains correctement rejetés (spécificité de l'outil).

Finalement, une fois les critères psychométriques évalués, nous avons tenté d'apporter des réponses à la question clé de notre travail, à savoir : le score AVQI 03.01 est-il corrélé à l'idée que le patient se fait de sa voix et à l'impact de sa dysphonie dans son quotidien ? Nous avons pour cela mesuré la corrélation entre le score calculé par l'AVQI 03.01 et le handicap que cause réellement la dysphonie dans le quotidien du patient, tel qu'évalué à l'aide du questionnaire Voice Handicap Index (VHI), également disponible dans la base de données du CHU.

Ceci nous a permis de nuancer la pertinence de l'utilisation de l'évaluation vocale objective dans le cadre d'un bilan logopédique. En effet, lorsque Youri Maryn a présenté l'AVQI lors d'un symposium organisé le 21 octobre 2016 par l'Université de Liège en l'honneur du Professeur Johan Sundberg, nous nous sommes demandé : les mesures acoustiques permettent-elles vraiment de rendre compte de la sévérité du trouble vocal, produit de l'interaction entre la qualité vocale et le contexte de vie du patient dysphonique ?

Notre étude nous a permis de confirmer que l'AVQI 03.01 est un outil valide pour l'évaluation vocale en langue française. Il s'est également avéré que le score AVQI corrèle modérément avec le handicap causé par la dysphonie tel que perçu par le patient. L'implication de ce résultat sera nuancée.

## PARTIE I : CONTEXTE THÉORIQUE

Dans ce travail, nous traiterons de l'Acoustic Voice Quality Index (AVQI), outil d'évaluation objective de la voix. Cependant, il nous semble important, avant d'aborder l'évaluation vocale, de définir brièvement ce qu'est la voix, phénomène complexe et multidimensionnel.

## La phonation, la voix et la qualité vocale

### I. La phonation et la voix

Tout d'abord, les termes « phonation » et « voix » doivent être distingués. Il s'agit pour le premier du phénomène physique et physiologique de la mise en vibration des plis vocaux (Giovanni & Lagier, 2009), responsable de la production d'un signal acoustique qui sera modelé par les cavités supra-glottiques : la « voix ». La phonation est donc définie comme le geste vocal, tandis que le second terme « voix » correspond au son qui en résulte. Déjà en 1820, la phonation était définie dans le dictionnaire des sciences médicales comme comprenant « la série de phénomènes qui concourent dans l'homme et les animaux à la production de la voix et de la parole... elle appartient à la vie de relation et elle est un des premiers moyens d'expression de l'instinct, des sentiments et des idées. » (Rullier, 1820)

La voix, « métaphore de notre identité profonde » (Castarède & Konopczynski, 2015), est produite grâce à l'« appareil vocal » (représenté à la figure 1), qui peut être subdivisé en trois étages :

1. **L'étage sous-glottique** : c'est ici que va s'amasser l'énergie aérodynamique en provenance des poumons, qui permettra la mise en vibration des plis vocaux.
2. **L'étage glottique** : ce sont les plis vocaux<sup>1</sup> qui jouent le rôle de vibreur dans la phonation. C'est à cet étage que s'opère la conversion de l'énergie aérodynamique en énergie acoustique/sonore (c'est-à-dire la phonation). Cette transformation se fait grâce à la vibration passive (autoentretenu) des plis vocaux, souvent décrite en référence à la théorie myo-élastique de Ewald (1898, cité par Cornut, 2014 et Giovanni & Lagier, 2009)<sup>2</sup>, ancienne mais toujours partiellement valable. La création du son laryngé implique la participation de nombreux muscles, ligaments et cartilages laryngés.
3. **L'étage sus-glottique** : il comprend le vestibule laryngé et les cavités de résonance, soit le pharynx, la cavité buccale et pour certains sons les fosses nasales. Le son laryngé est ici modulé dans sa hauteur, dans son intensité<sup>3</sup> et dans son timbre

---

<sup>1</sup> Précisons que le terme « corde vocale », autrefois communément utilisé, est peu à peu remplacé par le terme « pli vocal », davantage représentatif de la structure anatomique décrite. (Autesserre et al., 2006)

<sup>2</sup> Selon la théorie myo-élastique d'Ewald (1898, cité par Cornut, 2014), la vibration des plis vocaux est un processus passif (elle se fait grâce à l'air expiré) et les caractéristiques acoustiques (hauteur, niveau de pression sonore et timbre) sont fonction de la pression sous-glottique et de la tension des plis vocaux.

<sup>3</sup> Nous privilégierons dans ce travail le terme « niveau de pression sonore » (sound pressure level).

par les résonateurs (grâce, notamment, à l'amplification de certaines bandes fréquentielles).

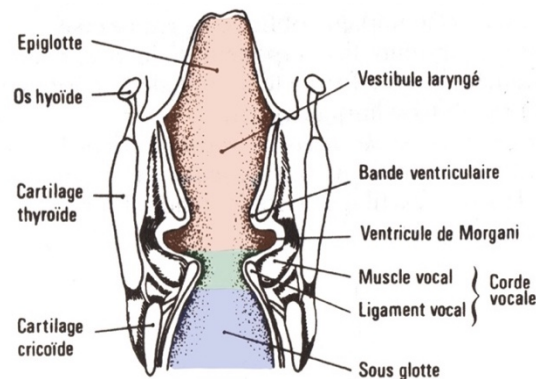


Figure 1. Vue de face des trois étages laryngés (adapté de Cornut, 2014).

En bleu, l'étage sous-glottique ; en vert, l'étage glottique ; en rouge, l'étage sus-glottique.

Nous ne décrivons pas ici l'anatomie du larynx ni les mécanismes physiologiques permettant la création du son laryngé (vibration cordale, théorie myo-élastique d'Ewald, théorie « body-cover » de Hirano...), décrits par de nombreux auteurs (dont Giovanni & Lagier, 2009 et Cornut, 2014). Nous nous concentrerons davantage sur les mécanismes permettant de réguler la hauteur, le niveau de pression sonore et le timbre du son laryngé, et ainsi de donner vie à la voix, produit final d'un processus complexe.

## II. La modulation du son vocal

Le son produit au niveau des plis vocaux, puis modelé par les résonateurs, comporte comme tout autre son quatre paramètres principaux : la hauteur, le niveau de pression sonore, le timbre et la durée (Vaissière, 2015). Titze, en 1994 (cité par Maryn, 2010), décrit quant à lui les quatre caractéristiques suivantes pour le son vocal : la hauteur, le niveau de pression sonore, l'identification phonétique — correspondant au timbre qui permet l'identification de phonèmes, comme les voyelles par exemple — et la qualité vocale. Nous décrivons ci-après les mécanismes de régulation de la hauteur (fréquence), du niveau de pression sonore (amplitude) et du timbre (résonance). La qualité vocale, en relation étroite avec le timbre, sera décrite dans le sous-chapitre suivant.

### A. La régulation de la fréquence

Les plis vocaux vibrent en moyenne jusqu'à 120 fois par seconde chez l'homme, 240 chez la femme, 350 chez l'enfant et jusqu'à plus de 400 fois chez le nouveau-né (Vaissière, 2015). Le son sera perçu comme plus ou moins aigu en fonction de la vitesse d'ondulation de la couverture cordale. La hauteur (ou « tonie », lorsque l'on considère l'impression subjective de

la hauteur perçue) dépend donc de la fréquence de vibration des plis vocaux — la fréquence fondamentale :

$$F0 = \frac{1}{2L} \times \sqrt{\frac{\sigma}{\rho}}$$

Équation de la fréquence dans le cas d'une corde vibrante (Titze, 2011)

La fréquence fondamentale ( $F0$ ) est inversement proportionnelle à la longueur des plis vocaux ( $L$ ) et à leur densité ( $\rho$ ), et directement proportionnelle à leur tension ( $\sigma$ ).

Nous avons préféré utiliser cette équation de Titze (2011) à l'équation traditionnelle suivante :

$$F0 = \frac{1}{2\pi} \times \sqrt{\frac{k}{m}}$$

Équation de la fréquence traditionnellement proposée — pendule simple

Cette dernière prend en compte la tension ( $k$ ) et la masse ( $m$ ) des plis vocaux. Or, Titze explique que la masse n'est pas un paramètre valide pour la détermination de la fréquence fondamentale. L'équation traditionnelle s'applique en effet sur un corps qui se déplace, c'est-à-dire un pendule — une masse («  $m$  ») attachée à un « fil » inextensible et sans masse, de raideur «  $k$  ». Or, les plis vocaux ne se déplacent pas, ils se déforment. Titze souligne que les différentes couches des plis vocaux (représentées à la figure 2) vibrent à une amplitude différente et doivent donc être considérées séparément.

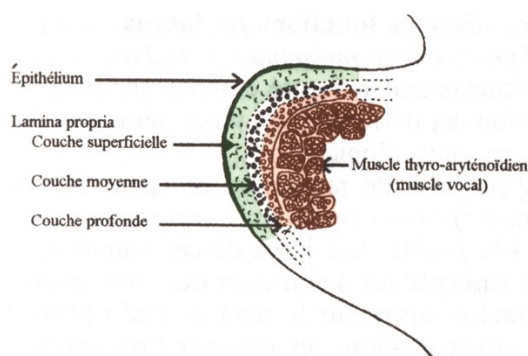


Figure 2. Les cinq couches des plis vocaux (adapté de Cornut, 2014).

Les cinq couches sont : l'épithélium, les trois couches de la lamina propria (espace de Reinke, couche intermédiaire et profonde) et le muscle vocal ; en vert, la couverture cordale ; en orange, le corps cordal (Finck, 2005).

Giovanni et Lagier (2009) évoquent également les propriétés biomécaniques propres à chaque couche. Ceci entraîne un comportement relativement indépendant de ces couches en ce qui concerne la tension résultante d'un changement de longueur. Les auteurs précisent, notamment, que les fibres de collagène (très présentes dans les couches profondes de la lamina propria) « sont plus résistantes à l'élongation que les fibres d'élastine ».

Titze fournit lui aussi des explications exhaustives dans son article de 2011, ainsi que dans son livre « Principles of voice production » (1994), œuvre de référence pour de nombreux

vocologues. En résumé, Titze (2011) cite les trois variables suivantes comme les variables clés de la variation fréquentielle : la *longueur*, la *tension* et l'*activation musculaires*. La densité ( $\rho$ ) reste constante dans les tissus mous et est de  $1.04 \text{ g/cm}^2$ . Il souligne que dans le cas d'une altération tissulaire (polype, nodule...), il ne faut pas uniquement considérer la masse ajoutée par celle-ci, mais aussi et surtout, la raideur qui y est associée.

Les mécanismes qui contribuent à modifier ces paramètres en vue de définir la fréquence fondamentale de la voix sont les suivants :

- 1) La contraction des muscles antagonistes thyro-aryténoïdien et crico-thyroïdien (Story & Titze, 1995) augmente la rigidité des plis vocaux (rigidité active pour le TA, tension longitudinale passive pour le CT), avec laquelle croît la fréquence vibratoire. Par exemple, lorsque la contraction du crico-thyroïdien est prépondérante (mécanisme laryngé dit « léger »), la raideur passive (le muscle thyro-aryténoïdien étant relâché) est alors accrue autant pour le corps que pour la couverture cordale (figure 3, schéma d). La vibration, limitée à deux tiers des plis vocaux, est de faible amplitude horizontale et de fréquence élevée, l'ondulation muqueuse quasiment absente ; la phase de fermeture est courte (dans les émissions les plus aiguës, il n'y a plus de contact entre les plis vocaux).

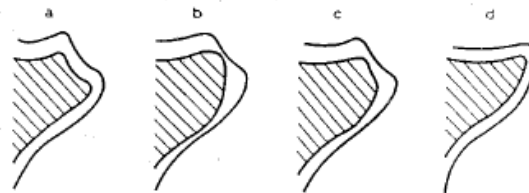


Figure 3. Configuration des plis vocaux en fonction de la contraction des muscles antagonistes thyro-aryténoïdien et crico-thyroïdien (Story & Titze, 1995).

TA = thyro-aryténoïdien ; CT = crico-thyroïdien. Rapports de contraction : (a) TA = CT (b) TA > CT (c) TA légèrement > CT (d) CT > TA ; Rapports de tension induite : (a) corps = couverture (légère) (b) corps >> couverture (c) corps > couverture (d) corps < couverture.

- 2) La contraction du muscle crico-aryténoïdien latéral rapproche les apophyses vocales des cartilages aryténoïdes, ce qui provoque ce qu'on appelle une « compression médiane » (Cornut, 2014). Plus cette dernière est importante, moins grande sera la portion vibrante du pli vocal, et donc plus élevée sera la fréquence fondamentale.
- 3) La pression sous-glottique : selon des études récentes menées sur des larynx humains excisés (Garrel, 2007 cité par Henrich, 2012), un lien direct existe entre la fréquence fondamentale et la pression sous-glottique (représenté à la figure 4). En effet, sans modifier les autres paramètres laryngés (musculaires ou mécaniques), la fréquence fondamentale augmente de façon concomitante à la pression et au débit d'air sous-glottiques.

Une augmentation de la pression subglottique de 1 cm H<sub>2</sub>O amène une élévation de la fréquence fondamentale d'environ 3 à 7 Hz (Baer, 1979). Titze mesure une augmentation de 0.5 à 6 Hz/cm H<sub>2</sub>O (Titze, 1989). Il explique celle-ci par la tension des plis vocaux induite par l'amplitude vibratoire croissante. Cette tension serait d'après lui une « tension intratissulaire d'origine purement dynamique » (Finck, 2008). Sundberg (1992) évoque ainsi la fraction « tension = A/L » : la tension augmente de manière proportionnelle à l'amplitude vibratoire (A) et inversement proportionnelle à la longueur (L) des plis vocaux.

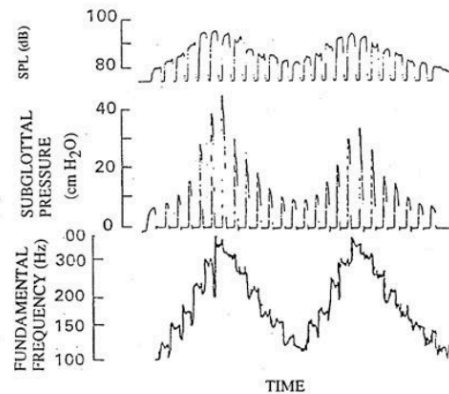


Figure 4. Représentation de la variation de la pression sous-glottique en fonction de la fréquence fondamentale (Sundberg, 1992). La première courbe montre le niveau de pression sonore, la seconde les variations de pression sous-glottique, la troisième la fréquence fondamentale.

Cornut (2014), quant à lui, explique ce phénomène par l'effet Bernoulli : le passage rapide de l'air au niveau du plan glottique provoque une sous-pression qui amène les plis vocaux à se rapprocher (phénomène de rétro-aspiration). Lorsque la pression sous-glottique augmente, la rétro-aspiration est amplifiée, ce qui entraîne un raccourcissement de la phase de fermeture des plis vocaux, et par conséquent une augmentation de la fréquence fondamentale. L'augmentation inhérente de la pression sous-glottique avec la hauteur semble non seulement inévitable, mais également nécessaire (Herbst, Hess, Müller, Svec, & Sundberg, 2015). En effet, les hautes fréquences sont produites grâce à des plis vocaux étirés ; la mise en vibration des plis vocaux ainsi tendus nécessite une plus grande pression sous-glottique, la résistance du plan glottique étant augmentée (Giovanni & Lagier, 2009 ; Sundberg, 1992).

Le seuil minimal de pression nécessaire à la mise en vibration des plis vocaux en voix parlée est appelé *phonation threshold pressure* en anglais (PTP) (Peiyun et al., 2009 ; Titze, Schmidt, & Titze, 1995). Pillot-Loiseau (2011) évoque pour ce seuil une valeur située entre 7 et 10 cm H<sub>2</sub>O à fréquence et à niveau de pression sonore « confortables ». Notons qu'il peut varier en fonction de la raideur de la partie vibrante (Giovanni & Lagier, 2009; Mendelsohn & Zhang, 2011), de la viscosité (c'est-à-dire le degré de résistance aux forces « qui déplacent les unes

par rapport aux autres les particules » constituantes [Centre national de la recherche scientifique, 2016]), de l'épaisseur du bord libre (Giovanni & Lagier, 2009) et de la largeur de la fente glottique pré-phonatoire<sup>4</sup> (Chan & Titze, 2006 ; Giovanni & Lagier, 2009). Ce seuil est donc souvent augmenté dans les conditions pathologiques comme le forçage vocal (dysphonie hypertonique par exemple), mais aussi dans les dysphonies organiques.

Les différents mécanismes décrits ci-dessus permettent donc à l'être humain de varier sa fréquence fondamentale, et ce sur une étendue considérable. Tim Storms a ainsi battu le record mondial de l'étendue vocale en 2008, faisant varier sa fréquence fondamentale de dix octaves, de la note la plus grave à la plus aiguë ! (Guinness World Records, 2017) En moyenne, une personne non chanteuse a une étendue vocale d'une octave et demie (Sicard & Sicard, 2016).

Finalement, il est important de préciser que la régulation fréquentielle est pour ainsi dire indissociable de la régulation du niveau de pression sonore, que nous allons maintenant décrire.

## B. La régulation du niveau de pression sonore

La régulation du niveau de pression sonore se fait aux trois étages de l'appareil vocal (National Center for Voice and Speech, 2016) :

- Au niveau du tractus vocal : la taille, la forme et la position des résonateurs (selon le type morphologique de l'individu mais aussi des mouvements effectués) vont permettre d'amplifier certaines zones fréquentielles, de leur donner davantage d'énergie acoustique. Évoquons ici le « singing formant » : il s'agit du renforcement de la zone fréquentielle autour de 3000 Hz recherché dans le chant, qui permet aux chanteurs d'opéra d'être audibles par-dessus le volume sonore de l'orchestre (qui, lui, montre un pic d'énergie sonore maximale autour de 450 Hz) (Sundberg, 1972). Ce formant, représenté à la figure 5, serait généré en abaissant le larynx et en élargissant le résonateur pharyngé.

---

<sup>4</sup> Notons qu'une étude de Titze (1995) sur une glotte modélisée a permis de mettre en évidence que la relation entre l'ouverture glottique pré-phonatoire et la pression sous-glottique nécessaire à la mise en vibration des plis vocaux n'est pas linéaire : le minimum de pression nécessaire est atteint avec une ouverture glottique de 1.0mm. Pour des ouvertures plus grandes, la pression nécessaire augmente de façon quasi linéaire ; pour des ouvertures plus petites, cette pression nécessaire augmente également. (Titze, Schmidt, & Titze, 1995) Ce phénomène est expliqué par Lucero (1996) comme étant le résultat de la résistance visqueuse des plis vocaux fortement rapprochés.



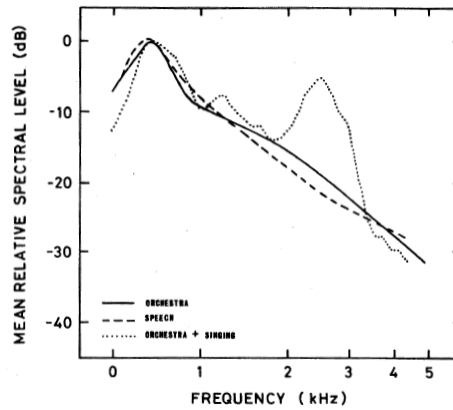


Figure 5. Représentation du « singing formant » (Sundberg, 1972).

Sont représentés les spectres moyens ; en pointillés longs = voix parlée ; en ligne continue = orchestre ; en pointillés fins = chanteur d'opéra accompagné d'un orchestre (pointillés fins) ; sur ce dernier tracé, l'on peut observer le « singing formant » autour de 3000 Hz.

- Au niveau du larynx : la compression médiane des plis vocaux joue un rôle important dans le réglage du niveau de pression sonore. Plus cette compression est importante, plus l'énergie nécessaire afin de séparer les plis vocaux sera grande. De cette manière, la glotte se referme avec plus de vigueur et à une plus grande vitesse. Les « pufs d'air » passent ainsi à l'étage sus-glottique avec un plus grand niveau de pression sonore (Jiang, 2015). Le temps d'adduction est alors allongé, le temps d'abduction (« open quotient ») raccourci (Giovanni & Lagier, 2009). Cette augmentation du temps d'adduction des plis vocaux concomitante à l'augmentation de la pression sous-glottique permet de contrecarrer l'élévation cooccurrence de la fréquence (Giovanni & Lagier, 2009).

Une plus grande intensité vocale implique donc une force de fermeture plus importante. L'on comprendra facilement que si cette augmentation de niveau de pression sonore s'opère de manière inadéquate, un forçage vocal peut rapidement s'installer, exposant les plis vocaux à des stress mécaniques importants.

- Au niveau de la soufflerie/des poumons : la pression sous-glottique est également un élément primordial déterminant le niveau de pression sonore. Elle dépend elle-même de la pression pulmonaire, liée aux forces de l'appareil respiratoire (Giovanni & Lagier, 2009). Plus la pression atteignant les plis vocaux est grande, plus l'amplitude de vibration des plis vocaux sera importante. Notons qu'il existe une relation linéaire entre l'augmentation de la pression sous-glottique et celle du niveau de pression sonore : lorsque la pression est doublée, le niveau de pression sonore augmente de neuf décibels (Fant, 1982). Donc, plus la pression sous-glottique est élevée, plus l'amplitude vibratoire et l'intensité vocale perçue seront également importantes (voir figure 6) (Herbst, Hess, Müller, Svec, & Sundberg, 2015).

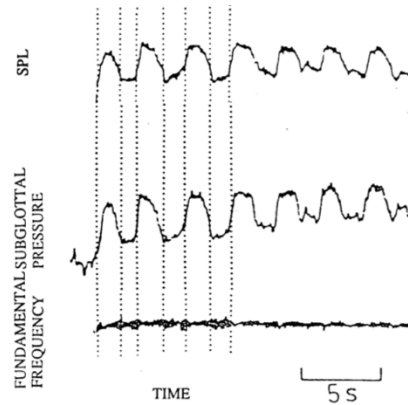


Figure 6. Représentation de la variation de la pression sous-glottique en fonction du niveau de pression sonore (Sundberg, 1992). La première courbe représente les variations du niveau de pression sonore, la seconde la pression sous-glottique, et la troisième la fréquence fondamentale.

### C. Les cavités de résonance

Le signal acoustique provenant des plis vocaux correspond à un son périodique complexe dont la perception pourrait se définir comme un cri d’oie. En effet, lorsque nous observons les spectrogrammes de gauche sur la figure 7, nous remarquons que toutes les fréquences du signal sont émises avec une amplitude qui diminue progressivement vers les fréquences aiguës.

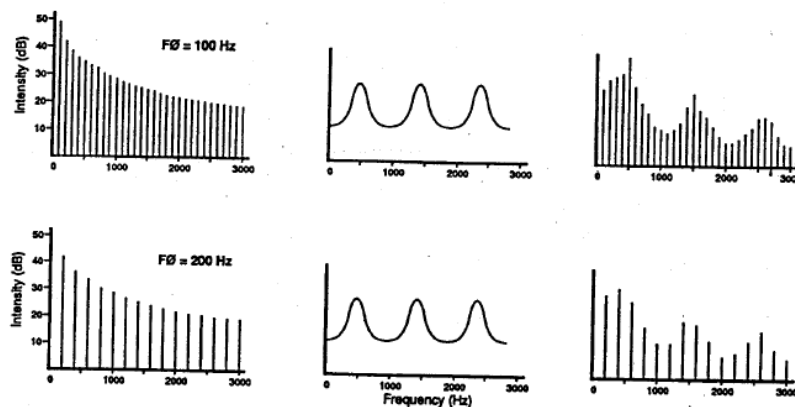


Figure 7. Représentation schématique de l’influence des cavités de résonance sur le signal acoustique laryngé (Rubin & Vatikiotis-Bateson, 1998).

À gauche = spectre du son laryngé ; au milieu = représentation des zones de filtration par les résonateurs ; à droite = spectre résultant ; en haut = un son d’une fréquence fondamentale de 100 Hz ; en bas = un son d’une fréquence fondamentale de 200 Hz.

C’est lors du passage dans l’étage sus-glottique que la voix sera modulée (figure 8) par le vestibule laryngé, l’oropharynx, la cavité buccale et pour certains sons le cavum (rhinopharynx) et les fosses nasales. De par leur configuration, ceux-ci vont renforcer certaines bandes de fréquences (les formants) — proches de leur fréquence propre — et « amortir » les autres.

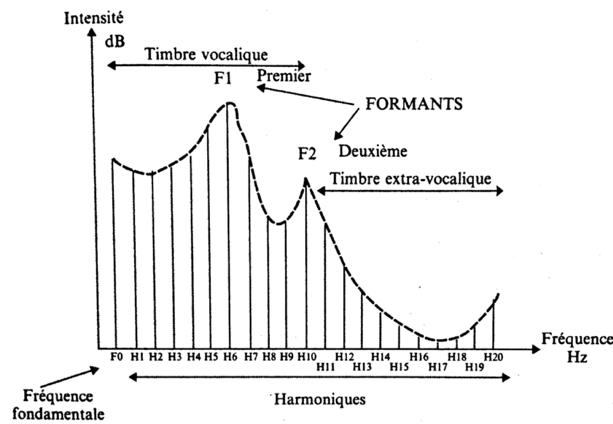


Figure 8. Représentation spectrale d'un son vocal (Cornut, 2014).

L'on peut notamment observer la fréquence fondamentale (F0) et les deux premiers formants (F1 et F2).

De cette manière, le tractus vocal humain de chaque individu modifiera le son laryngé selon la configuration propre (forme, volume) de ses cavités de résonance. C'est ainsi que se caractérise le timbre vocal, notamment, mais c'est aussi par ce mécanisme de résonance que nous pouvons distinguer les voyelles : chaque voyelle est caractérisée principalement par des écarts relativement stables entre les formants 1 (bande fréquentielle renforcée par la cavité pharyngée) et 2 (bande fréquentielle renforcée par la cavité buccale).

Précisons que les fréquences des formants caractérisant les voyelles varient d'un individu à l'autre, mais aussi en fonction du contexte lexical chez un même locuteur. Cette dernière variabilité est principalement due à la coarticulation, qui rend difficile de moyennner la fréquence exacte des formants. En effet, les voyelles transitoires entre deux consonnes sont parfois très courtes et ainsi fortement influencées par le contexte phonémique. Prenons comme exemple concret la voyelle [i] dans le mot « effectivement » à la figure 9. Le second formant présente sur le spectrogramme une pente ascendante assez raide dans sa transition vers le [v].

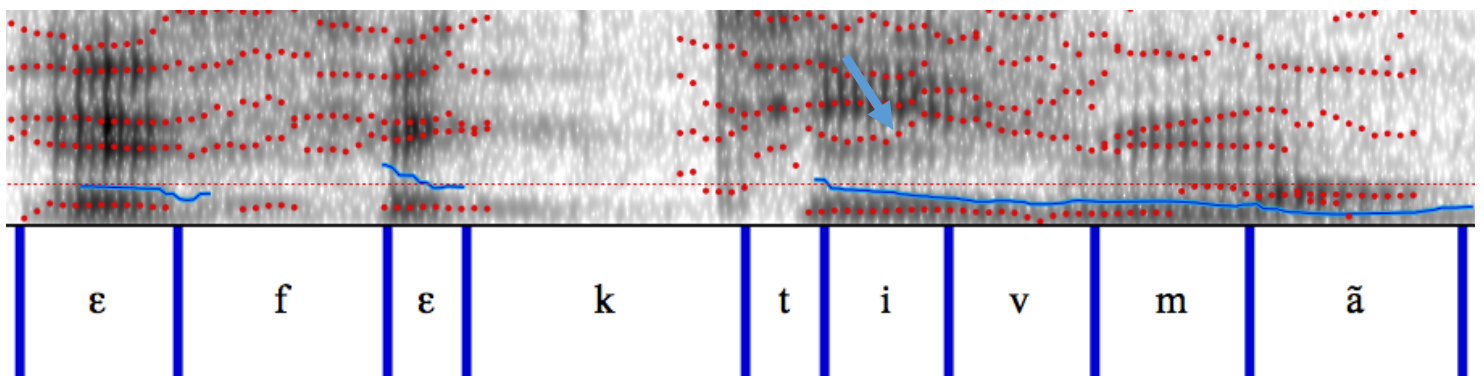


Figure 9. Spectrogramme du mot « effectivement ».

Mesurer une fréquence moyenne serait ici délicat. La variabilité qui apparaît en conséquence n'a cependant pas d'impact sur l'identification des voyelles. En effet, ce ne sont pas les valeurs absolues des formants qui sont primordiales, mais davantage les écarts entre eux.

C'est ainsi qu'a pu être créé le triangle vocalique (figure 10), sur lequel nous pouvons observer les fréquences moyennes des deux premiers formants (F2 en abscisse, F1 en ordonnée) des voyelles de la langue française.

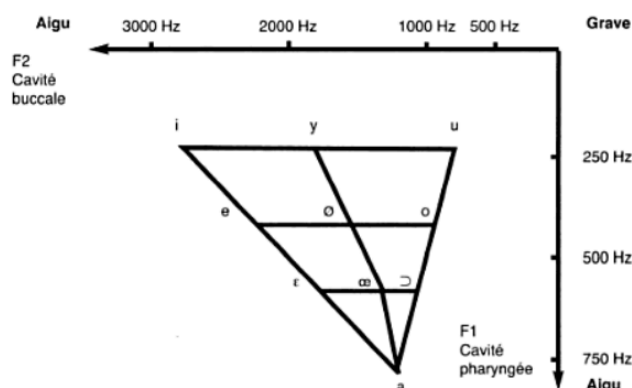


Figure 10. Triangle vocalique du français (Méchain, 2014).

Notons que le 3<sup>e</sup> formant, ainsi que les formants suivants, contribuent moins à l'identification des voyelles. Cependant, ils n'en sont pas moins importants, étant donné qu'ils permettent de caractériser le timbre vocal d'un individu.

### III. La qualité vocale

La qualité vocale, plus communément appelée « timbre » vocal, est décrite par l'Institut de Normalisation Américaine (American National Standards Institute, 1960) comme « l'attribut de la sensation auditive grâce auquel un auditeur peut juger que deux sons présentés de manière similaire et ayant la même intensité et la même hauteur sont différents ».

Ainsi, la qualité vocale englobe tout ce qui ne relève pas de la hauteur ou du niveau de pression sonore (Behrman, cité dans Boone, McFarlane, Von Berg, & Zraick, 2010). Elle constitue de ce fait un aspect perceptif multidimensionnel, c'est-à-dire qui ne peut être défini par un corrélat acoustique unique, comme la hauteur peut être définie par la fréquence, et le niveau de pression sonore par l'amplitude vibratoire (Barsties & De Bodt, 2015). La qualité vocale est ainsi, selon Kreiman et Gerratt (1998), le résultat de l'interaction du signal vocal et de la perception par l'auditeur ; elle n'est donc pas une caractéristique du signal acoustique à proprement parler. C'est pourquoi, soulignent les auteurs, il est primordial de prendre en compte l'évaluation perceptive dans l'évaluation vocale. Au niveau physiologique, Mathieson (2013) précise que la qualité vocale dépend de la qualité d'adduction des plis vocaux, ainsi que de la qualité de la vibration cordale, entendu sa symétrie et sa périodicité. Boone, McFarlane, Von Berg et Zraick (2010) ajoutent que la qualité vocale peut être modifiée non seulement par des modifications de la source laryngée, mais aussi par des modifications dans le tractus vocal (résonateurs).

Pour résumer, le National Center for Voice and Speech (NCVS) définit la qualité vocale selon l'équation suivante, soit la somme de la configuration du tractus vocal et des caractéristiques du larynx, ainsi que de l'apprentissage de l'utilisation de l'appareil phonatoire :

*Qualité Vocale = configuration du tractus vocal + anatomie laryngée + composante apprise*

La qualité vocale est donc définie par les prédispositions génétiques, mais aussi par l'apprentissage. Ainsi, un chanteur professionnel a appris à modifier la configuration de ses résonateurs afin de modifier sa qualité vocale (p.ex grâce au « singing formant ») (National Center for Voice and Speech, 2016).

Une liste exhaustive de termes utilisés pour définir la qualité vocale est également présentée sur la page web du National Center for Voice and Speech, accessible via le lien suivant :

<http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/quality.html>.

Dans le cas d'une dysphonie, la qualité vocale peut être altérée par de nombreux processus. Au niveau perceptif, le terme le plus adéquat pour définir une qualité vocale altérée est l'« enrrouement », selon Barsties et Maryn (2015). Les sous-termes les plus utilisés sont la voix « soufflée », « forcée » ou « rauque ».

## Les dysphonies

De nombreux facteurs de risque peuvent contribuer à la création d'un terrain propice au développement de pathologies vocales. Ces facteurs sont multiples et variés, incluant à la fois des facteurs physiques, sociaux et psychologiques. Ils peuvent apparaître de manière isolée ou combinée. La multitude des facteurs potentiellement à l'origine d'une pathologie vocale souligne l'importance de réaliser une anamnèse complète et une observation clinique avisée. Celles-ci, conjuguées aux résultats obtenus à l'aide de différents outils de mesure décrits plus loin, apportent une vision du patient dysphonique dans son ensemble. Citons quelques facteurs de risque, cette liste n'étant pas exhaustive (Gabriel, 2012) :

- l'utilisation de la voix dans le cadre socioprofessionnel ou dans les hobbies ;
- les caractéristiques psychologiques (grande sensibilité émotionnelle et/ou nervosité) ;
- les substances nocives : l'alcool, le tabac ou toute autre substance pouvant induire une irritation de la muqueuse laryngée (poussières, vapeurs caustiques...) ;
- les facteurs iatrogènes : interventions médicales (chirurgicales, médicamenteuses, hormonales, chimiothérapiques...) ;
- les infections chroniques de la sphère ORL ;
- le reflux gastro-œsophagien (RGO) ;
- la baisse de l'acuité auditive, induisant une perte du contrôle audio-phonatoire ;

– la présence d’une personne malentendante ou dysphonique dans l’entourage du patient.

Les dysphonies, ou pathologies vocales, sont classiquement subdivisées en pathologies d’origine *organique* (appelées dysphonies organiques – kystes intracordaux muqueux, polypes, sulcus...), d’origine *fonctionnelle* (dysphonies fonctionnelles hyper- ou hypokinétiques) ou d’origine *neurologique* (troubles de la mobilité des plis vocaux, dysphonies neurologiques – paralysie récurrentielle, dysarthries...). Notons que ces subdivisions, bien que communément utilisées en pratique clinique, semblent à ce jour désuètes. En effet, les dysphonies causées par une (ou des) lésion(s) structurelle(s) acquise(s) au niveau de l’organe phonateur peuvent entraîner, par compensation, une utilisation inadéquate des supports vocaux. Inversement, un geste vocal inadapté peut, à long terme, conduire à des lésions organiques.

À ces trois grandes catégories s’ajoutaient autrefois également les troubles de la mue ainsi que les troubles de la voix d’origine psychogène. Cependant, l’on sait aujourd’hui que certaines dysphonies fonctionnelles, de même que les troubles de la mue, ont une origine psychologique. Cette catégorisation des dysphonies chroniques<sup>5</sup> est donc aujourd’hui devenue obsolète.

En réponse au manque d’accord quant aux définitions souvent trop floues des pathologies cordales bénignes, Camille Finck a récemment proposé une nouvelle classification originale centrée sur l’aspect du tissu cordal. Elle décloisonne ainsi les « étiquettes » pathologiques classiquement définies (nodules, polypes...). Insistant sur l’importance de l’état de l’espace de Reinke (souplesse et épaisseur) dans la vibration cordale, elle classe les lésions selon le type d’altération de celui-ci : lésions occupantes, lésions destructrices, lésions occupantes et destructrices, lésions déformantes, lésions hémorragiques et pathologies inflammatoires (Finck, 2017).

Depuis une vingtaine d’années, le concept de charge vocale s’est développé dans la littérature. Ce concept permet lui aussi de considérer la dysphonie sous un nouvel angle et de s’écarter de la dichotomie « dysphonie fonctionnelle/organique ».

La charge vocale est décrite comme « la quantité d’utilisation vocale par unité de temps » (Morsomme & Remacle, 2013), définie par trois paramètres :

- la fréquence fondamentale ;
- l’intensité/le niveau de pression sonore ;
- la durée de phonation.

---

<sup>5</sup> Des dysphonies aiguës peuvent également être décrites, d’étiologie traumatique (« coup de fouet ») ou infectieuse (laryngites).

Une augmentation trop importante de l'un de ces trois paramètres (la fréquence et le niveau de pression sonore étant par ailleurs liés) par rapport à la résistance propre des plis vocaux de tout un chacun va amener à ce qui est appelé un *surmenage vocal*. Pour objectivement évaluer la charge vocale dans des situations écologiques et sur une longue période (journées entières), les dosimètres ou accumulateurs vocaux ont vu le jour dans le domaine de la recherche. Malheureusement, le monitoring ambulatoire de la voix est encore trop peu abordable pour une application en clinique. Outre le surmenage, qui concerne donc la quantité d'utilisation vocale par unité de temps, le *malmenage* désigne quant à lui l'utilisation d'un geste vocal inadapté (aspect qualitatif). Celui-ci peut soit faire suite au surmenage, soit se développer de manière indépendante. À long terme, le malmenage et le surmenage vocaux peuvent, suite aux stress mécaniques et aux déformations subis par les plis vocaux (Titze, 2001), entraîner des conséquences néfastes sur le plan anatomophysiologique et ainsi déclencher le cercle vicieux du forçage vocal.

## L'évaluation vocale — état de l'art

Les outils d'évaluation de la voix sont très nombreux et variés. Ceci pose un problème méthodologique notamment pour les méta-analyses et les comparaisons entre traitements. Afin de permettre des comparaisons plus évidentes, une standardisation de la procédure d'évaluation est nécessaire. Ainsi, De Bodt (cité dans Wuyts et al., 2000) proposa en 1997 un modèle pour l'évaluation vocale, comprenant quatre composantes :

- 1) l'inspection laryngée ;
- 2) l'évaluation perceptive par le thérapeute/clinicien ;
- 3) l'évaluation subjective par le patient ;
- 4) l'évaluation objective à l'aide du DSI.

Un modèle plus complet vit le jour en 2001, lorsque la Société européenne de laryngologie (ELS) mit en place un protocole pour l'évaluation fonctionnelle de la voix (Dejonckere et al., 2001), dont un exemple est disponible à l'annexe 3 (page 161). Ce protocole propose cinq étapes d'évaluation :

- 1) l'évaluation perceptive par le clinicien : à l'aide des critères G (degré général de dysphonie), R (raucité) et B (souffle) de l'échelle GRBAS (décrite ci-après, page 19) ;
- 2) la vidéostroboscopie : observation de la fermeture glottique, de la régularité/périodicité de la vibration ainsi que de sa symétrie, et de l'onde muqueuse ;
- 3) les mesures acoustiques : mesures de perturbation (jitter ou « guige vocale » [Autesserre et al., 2006], shimmer), étendue fréquentielle et niveau de pression sonore minimal ;
- 4) une mesure aérodynamique : quotient phonatoire ;

- 5) l'évaluation subjective de sa voix par le patient sur une double échelle visuelle analogique (qualité vocale et répercussions au quotidien).

Des informations quant aux conditions d'enregistrement et aux stimuli à analyser sont également fournies.

Une étude de validation a montré que ce protocole est efficace et utile pour les dysphonies non « extrêmes », mais laisse à désirer pour les cas plus sérieux, telle la dysphonie spasmodique, l'aphonie et les voix de substitution (non produites à l'aide des plis vocaux, comme la voix des bandes ventriculaires) (Friedrich & Dejonckere, 2005). Comme cela avait déjà été mentionné dans l'article original, des indications spécifiques doivent être suivies pour l'évaluation de la voix dans ces cas (Dejonckere et al., 2001).

Nous allons ci-après brièvement décrire les principales étapes de l'évaluation vocale. Nous reprenons pour cela la trame du protocole de l'ELS (2001), principalement adopté en Belgique et en France, en nous attardant seulement sur les éléments pertinents pour ce travail : l'évaluation subjective par le thérapeute et par le patient, et les mesures acoustiques. Nous nous permettons tout d'abord d'évoquer quelques informations quant à l'anamnèse (non reprise dans le protocole de l'ELS), qui nous semble être un élément capital dans l'évaluation vocale.

## **I. L'anamnèse**

Comme toute évaluation logopédique, l'évaluation vocale doit être précédée d'une anamnèse ciblée et complète. En effet, avant d'évaluer la voix, il est primordial de connaître l'histoire du patient, son contexte de vie, l'usage qu'il fait de sa voix (donc, entre autres, sa charge vocale) et les éventuels facteurs de risque liés aux pathologies vocales (décrits ci-avant). La nature exacte de la plainte vocale doit également être définie.

La Japan Society of Logopedics and Phoniatrics conseille de recueillir les informations suivantes (Omori, 2011) :

- la plainte principale : raucité, hauteur ou intensité anormale... ;
- la pathologie actuelle : aiguë/chronique, début des symptômes, historique de traitement... ;
- les facteurs de risque : usage vocal abusif ou inadapté, interventions chirurgicales, rhumes ou affections des voies aériennes supérieures, thérapie hormonale, stress... ;
- les complications : brûlures d'estomac, reflux gastrique, douleur laryngopharyngée, difficultés respiratoires, dysphagie... ;
- l'histoire psychomédicale : troubles neurologiques, psychologiques, endocriniens... ;
- le métier et/ou les hobbies ;



- la médication (par voie orale) : psychotropes, hormones... ;
- les habitudes de vie : tabagisme, consommation d'alcool... ;
- les allergies médicamenteuses.

Une fois l'anamnèse effectuée, l'évaluation vocale en tant que telle peut avoir lieu, dont les résultats devront être considérés en lien avec l'histoire et le contexte de vie du patient, et toute autre donnée récoltée lors de l'anamnèse.

### L'observation et la palpation

En complément de l'anamnèse, l'observation à l'œil nu de la cavité buccale et de l'oropharynx peut déjà apporter des informations quant à l'état des tissus mous des résonateurs. L'on examinera par exemple le voile du palais, la langue et les amygdales, la présence d'une éventuelle inflammation (rougeur des parois pharyngées) ou de sécrétions. D'autres observations telles les contractions lors de la phonation, les mouvements de la pomme d'Adam ou une restriction de la mobilité mandibulaire doivent également être notées (Boone, McFarlane, Von Berg, & Zraick, 2010). La palpation du cou et de la nuque informe généralement sur le degré de crispation de la musculature périlaryngée (Omori, 2011). Elle peut se faire en respiration de repos ainsi qu'en phonation, permettant ainsi une comparaison des deux conditions (Gabriel, 2014).

## **II. L'imagerie laryngée : la vidéostroboscopie**

La première étape, condition sine qua non pour la suite de l'évaluation vocale, est la visualisation du plan glottique par un médecin spécialiste de la voix (ORL ou phoniatre). La vidéostroboscopie permet de considérer l'état non seulement des plis vocaux et du larynx en général, mais aussi de l'étage supra-glottique, en respiration et en phonation.

Nous n'entrerons pas en détail dans la description des différents examens/outils laryngoscopiques existants, dont les plus fréquents sont : la laryngoscopie indirecte (Boone, McFarlane, Von Berg, & Zraick, 2010 ; Yanagisawa, Driscoll, & Sims, 2005), la laryngoscopie directe flexible (Dejonckere, 2010 ; Remacle & Lawson, 2009), la vidéokymographie (Heuer, Hawkshaw, & Sataloff, 2006 ; Svec & Schutte, 1996), la laryngoscopie à haute vitesse (Mehta & Hillman, 2008) et la laryngoscopie directe rigide. Notons qu'outre l'imagerie, d'autres outils permettent l'analyse du comportement laryngé, dont l'électroglottographie (EGG) (Heuer, Hawkshaw, & Sataloff, 2006 ; Remacle, Lawson, Giovanni, & Woisard, 2005) et l'électromyographie (EMG) (Dejonckere, 2010).

### III. Les mesures subjectives

#### A. L'évaluation perceptive par le thérapeute

Bien qu'actuellement l'évaluation objective à l'aide de mesures acoustiques est de plus en plus investiguée (comme nous allons le développer ci-après), l'évaluation à l'écoute reste toujours le « gold standard »<sup>6</sup> par rapport auquel ces nouveaux outils sont comparés et ainsi validés. Après tout, « la parole et la voix sont produites pour être perçues », et les patients dysphoniques consultent souvent lorsqu'eux-mêmes ou des proches perçoivent une altération vocale (Ghio et al., 2007). De plus, comme mentionné plus tôt, la qualité vocale est un aspect perceptif multidimensionnel, résultat de l'interaction du signal vocal et de la perception par l'auditeur ; elle n'est donc pas une caractéristique du signal acoustique à proprement parler (Kreiman & Gerratt, 1998).

Différents outils d'évaluation sont ainsi communément employés par les spécialistes de la voix, souvent en fonction du pays. Nous ne décrivons que les outils pertinents pour ce mémoire (GRBAS et CAPE-V).

##### a. *GRBAS (Hirano, 1981)*

Cette échelle, encore aujourd'hui utilisée internationalement et faisant partie du protocole d'évaluation de l'ELS, a été développée par la Société japonaise de logopédie et de phoniatry (Hirano) en 1981. Le thérapeute juge cinq caractéristiques vocales sur une échelle de 0 (absence) à 3 (sévérité) points :

- G (*Grade of hoarseness*) : degré général de dysphonie ;
- R (*Roughness*) : raucité, résultant de variations irrégulières de la fréquence fondamentale ou du niveau de pression sonore ;
- B (*Breathiness*) : souffle, lié à une fuite aérienne ;
- A (*Asthenia*) : fatigue, manque de puissance dû à un faible niveau de pression sonore ou un manque en harmoniques aigus ;
- S (*Strain*) : forçage, conséquence d'une hypertension – fonction liée à une fréquence fondamentale élevée ou du bruit dans les harmoniques aigus.

Un sixième critère a été ajouté par Dejonckere et al. en 1995 (cité par Morsomme, 2001) :

- I (*Instability*) : instabilité au cours de la phonation de l'une des caractéristiques décrites ci-dessus.

---

<sup>6</sup> Ce terme provient de temps anciens où un standard monétaire avait été établi pour le commerce international, afin de pouvoir définir la valeur d'unités monétaires de pays différents par rapport à une quantité d'or de référence (Claassen, 2005).

Cet outil, quoique largement employé, est soumis à des critiques concernant des critères méthodologiques :

- 1) Les juges : les variations intra- et interévaluateurs sont importantes — surtout dans les dysphonies légères à modérées, un peu moins pour les extrêmes du continuum. Ces variations sont dues aux référents internes que tout un chacun se construit quant aux différentes qualités vocales<sup>7</sup> (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). C'est par rapport à ces standards internes que seront jugées les voix dysphoniques.
- 2) Les paramètres de l'échelle : les cinq éléments de la GRBAS ne présentent pas une fidélité égale. En effet, le critère G semble être le plus robuste, suivi du critère R, puis B. Les critères A et S sont quant à eux les moins fiables. Ceci est notamment expliqué par la difficulté de reproduire le niveau de pression sonore réel de la voix sur base des enregistrements, ce qui amène une moindre utilisation du critère A (Gabriel, 2014). Quant au forçage (S), il est plus facilement évalué avec l'observation du patient, le forçage étant parfois plus « visible » qu'audible (Gabriel, 2014).

L'utilisation des trois critères GRB est donc privilégiée. En Allemagne, une échelle nommée « RBH » pour « Rauigkeit » (raucité), « Behauchtheit » (souffle), « Heiserkeit » (degré général de la dysphonie) est largement utilisée depuis les études de Wendler et al. en 1986 (Dejonckere, 2010 ; Lemm, 2012).

- 3) Le stimulus : l'évaluation dépend également du stimulus ; il peut s'agir soit d'une voyelle tenue ([a]), soit d'un échantillon de parole continue. Une combinaison de ces deux stimuli semble appropriée (Barsties & Maryn, 2017).

***b. Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) (Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009)***

Cet outil d'évaluation a été généré par Kempster et al. à la suite de la conférence de consensus sur l'évaluation perceptive, tenue en 2002 à Pittsburgh. De nombreux spécialistes de la voix et de la perception ainsi que des logopèdes ont donc contribué à sa création, mettant ainsi en commun les avis de chercheurs et de cliniciens. L'objectif était l'obtention d'un outil standardisé (concernant l'administration et la cotation – non pas les normes) afin de permettre une meilleure communication entre les intervenants ainsi qu'une certaine cohérence dans l'utilisation de l'outil d'évaluation.

---

<sup>7</sup> Kreiman et al. (1993) précisent ainsi que nous semblons tous avoir des référents internes comparables et stables concernant les voix normales, car nous y sommes exposés au quotidien de manière relativement équivalente.

Trois tâches sont effectuées par le patient (un exemplaire du CAPE-V se trouve à l'annexe 4, p. 162) :

1. voyelles tenues [a] et [i], chacune trois fois pendant trois à cinq secondes ;
2. lecture de six phrases à contenu phonétique spécifique (effets de coarticulation des voyelles [a], [i], [o], transitions sourde-sonore et démarrages vocaux en douceur, arrêts vocaux, démarrages vocaux brusques, hypo- et hypernasalité) ;
3. conversation naturelle suite à la question « Parlez-moi de votre problème de voix » (évaluation vocale dans un contexte de parole continue).

Six paramètres sont observés : la sévérité globale, la raucité, le souffle, la tension, la hauteur tonale et le niveau de pression sonore. Pour chacun de ces paramètres, l'évaluateur indique son jugement sur une échelle visuelle analogique graduée (sans unités). L'extrême gauche de celle-ci représente une voix « normale » (pour la sévérité globale, la hauteur tonale et le niveau de pression sonore) ou l'absence du paramètre (pour la raucité, le souffle et la tension). Le jugement est indiqué pour chaque tâche, avec comme légende un « 1 » pour les voyelles tenues, un « 2 » pour les phrases lues et un « 3 » pour la parole continue en conversation. L'évaluateur mentionne également si le paramètre observé est stable ou intermittent. De plus, les termes ordinaires « léger », « modéré » et « sévère » ont été ajoutés à cette échelle (à des intervalles non égaux) afin de permettre une description qualitative plus compréhensible. Des descripteurs supplémentaires permettent finalement de qualifier la voix de la personne évaluée de manière plus détaillée.

L'outil a été validé dans plusieurs études, confirmant notamment que le CAPE-V corrèle fortement avec l'échelle GRBAS (les paramètres évalués étant similaires, mis à part l'asthénie de l'échelle GRBAS). Trois différences sont cependant décrites par Kempster et al. (2009) : premièrement, aucun protocole standardisé de l'échelle GRBAS n'a été publié. Aucune information précise n'est donnée quant au stimulus à évaluer ni quant à la cotation. Deuxièmement, l'échelle GRBAS n'inclut que des échelles ordinales (contrairement à l'échelle CAPE-V qui inclut une échelle en intervalles/métrique), rendant moins efficace son utilisation pour les analyses statistiques. Troisièmement, l'outil CAPE-V intègre différents contextes de production vocale (phonation soutenue, lecture à voix haute, parole continue) ainsi que des descripteurs variés afin de permettre une évaluation complète et écologique.

Une étude de Karnell et al. (2007), parmi d'autres, montre que les paramètres du CAPE-V sont fortement corrélés avec ceux de l'échelle GRBAS, mais que le CAPE-V est plus sensible. Nemr et al. (2012) confirment la meilleure sensibilité de l'outil CAPE-V, particulièrement pour les légères altérations vocales.

De nombreux autres outils sont disponibles, mais une description de chacun d'entre eux n'est malheureusement pas possible dans le cadre de ce travail : Hammarberg Scheme/Stockholm Voice Evaluation Approach (Carding, Carlson, Epstein, Mathieson, & Shewell, 2000), Buffalo Voice Profile (Carding, Carlson, Epstein, Mathieson, & Shewell, 2000 ; Webb et al., 2004), Vocal Profile Analysis Scheme (VPAS) (Laver, 1981, cité dans Shewell, 1998 ; San Segundo & Mompean, 2017).

### *c. Les limites de l'évaluation perceptive*

En plus des critiques caractéristiques à chaque outil, de nombreux auteurs pointent les limites de l'évaluation perceptive en général. Nous avons déjà évoqué lors de la description de la GRBAS la limite principale, qui concerne le manque de fidélité intra- et interévaluateurs, c'est-à-dire la stabilité du jugement entre différents juges et la stabilité dans le temps du juge lui-même (test-retest). En effet, sans entraînement préalable ni référence, le pourcentage de variation entre les auditeurs peut atteindre 50 % (Teston, 2004).

D'après Kreiman et al. (1993), ces variations de jugement sont dues à la dissemblance des référents internes par rapport auxquels sont évaluées les voix dysphoniques. Ces référents peuvent différer selon le degré d'expertise dans le domaine de la voix, mais aussi selon des critères personnels (comme les valeurs esthétiques ou la langue maternelle [Révis, Giovanni, Wuyts, & Triglia, 1999]). La variabilité des jugements est également influencée par la définition plus ou moins précise du/des paramètre(s) à évaluer (Kreiman et al., 1993), ainsi que par des facteurs dépendants du moment d'évaluation (fatigue, capacités attentionnelles [Kreiman et al., 1993], stress, état émotionnel...). Notons par exemple qu'une voix modérément altérée sera jugée plus sévèrement si elle est présentée après une voix normale que si elle est entendue après une voix fortement altérée (Goldstone, 1998, cité par Barkat-Defradas et al., 2012).

Barsties et Maryn (2016) mentionnent trois sources principales de variations : les caractéristiques propres de l'évaluateur (connaissance de l'histoire du patient [Wuyts et al., 2000], niveau d'expérience...), les caractéristiques de l'échelle de mesure (ordinaire ou visuelle/métrique) et les caractéristiques du stimulus évalué (phonation soutenue d'une voyelle, parole continue...).

Pour pallier cette critique, de nombreuses études évaluent la fidélité<sup>8</sup> inter- et intra-juges avant d'effectuer les analyses prévues. De plus, un entraînement à l'évaluation perceptive de la voix est

---

<sup>8</sup> Notons qu'il est important de différencier la fidélité inter-juges (jugement parallèle par les différents juges) et l'accord inter-juges (jugement identique pour chaque item par les différents juges) (Kreiman et al., 1993). En effet, si un juge attribue par exemple un score X à un item et Y à un autre, et qu'un autre juge attribue les scores X+1 et Y+1, ils ne sont pas en accord, mais la fidélité inter-juges est « bonne ». Si, cependant, le second juge attribue les scores X et Y+1, la fidélité inter-juges de l'évaluation est compromise.

souvent proposé. Puis, des échantillons de voix pathologiques de différents types et degrés sont présentés au préalable pour garantir l'ancrage d'un référent interne plus stable et ainsi améliorer la standardisation. Nous décrivons dans la seconde partie de ce travail, lorsque nous développerons la méthodologie de notre étude, comment nous avons tenté de pallier cette problématique.

Malgré le point faible de la variabilité des jugements, il est aujourd'hui largement admis que l'évaluation subjective, autant par le thérapeute que par le patient, doit être prise en compte *parallèlement* aux autres données récoltées. Ceci est notamment primordial parce que la qualité vocale n'est pas une caractéristique du signal vocal à proprement parler (Kreiman & Gerratt, 1998).

### B. L'évaluation par le patient : impact des troubles au quotidien (approche écologique)

Les outils suivants ont comme objectif principal d'évaluer la *qualité de vie* du patient et le *handicap* causé par la dysphonie. La qualité de vie est définie par l'OMS comme « la perception qu'a un individu de sa place dans l'existence, dans le contexte de la culture et du système de valeurs dans lequel il vit, en relation avec ses objectifs, ses attentes, ses normes et ses inquiétudes. Il s'agit d'un large champ conceptuel, englobant de manière complexe la santé physique de la personne, son état psychologique, son niveau d'indépendance, ses relations sociales, ses croyances personnelles et sa relation avec les spécificités de son environnement. » (Organisation mondiale de la Santé, 1993) Le handicap quant à lui est défini par l'OMS dans la Classification internationale du fonctionnement, du handicap et de la santé (CIF) comme « un terme générique désignant les déficiences, les limitations d'activité et les restrictions de participation. Il désigne les aspects négatifs de l'interaction entre un individu (ayant un problème de santé) et les facteurs contextuels face auxquels il évolue (facteurs personnels et environnementaux). » (Organisation mondiale de la Santé, 2001)

Dans le cadre de ce travail, nous ne décrivons que les outils en lien avec notre recherche (VHI et VoiSS).

#### a. VHI (Jacobson et al., 1997)

Le Voice Handicap Index a été initialement créé et validé par Jacobson et al. (1997). Il s'agit d'une liste de 30 affirmations étudiant l'impact de la qualité vocale sur le quotidien du patient, et ce à trois niveaux : fonctionnel (impact sur les activités quotidiennes), émotionnel (ressenti, perception de la pathologie vocale) et physique (caractéristiques vocales et inconfort). Le patient évalue chaque item sur une échelle de 0 (jamais ; pas d'impact) à 4 (toujours ; handicap important). Un score total sur 120 points est calculé, ainsi qu'un score pour chaque catégorie. Le VHI est intéressant non seulement pour la mise en place du plan thérapeutique, mais aussi pour les mesures de suivi. En

effet, il permet de rendre compte du progrès perçu par le patient et de l'amélioration de la qualité de vie, qui au bout du compte doit être l'objectif primaire de tout clinicien. Un changement de l'impact de la qualité vocale peut être considéré comme significatif à partir d'une variation supérieure à 18 points entre deux mesures (Jacobson et al., 1997). Notons aussi qu'en répondant à ces affirmations, le patient prend conscience de son trouble vocal, de sa sévérité et de son impact, ce qui peut amener une plus grande motivation pour la prise en charge (Giovanni, 2004).

De nombreuses études ont été menées afin de traduire le questionnaire dans différentes langues : allemand, espagnol, arabe, grec, mandarin, portugais, hindi, norvégien, croate, perse, lithuanien, polonais, français, taiwanais, hébreu, suédois, italien, néerlandais et flamand (Seifpanahi, Jalaie, Nikoo, & Sobhani-Rad, 2015).

Une étude destinée à évaluer les scores seuils de quatre échelles d'auto-évaluation (VHI-30, VHI-10, VoiSS et VPQ) sur un large échantillon (975 participants) a identifié un score seuil optimal de 19 (Behlau et al., 2016). Gräbel, Hoppe et Rosanowski (2009) ont attribué quatre grades à quatre intervalles de scores, sur base d'analyses statistiques effectuées par rapport aux scores du SF-36 (outil de mesure de la qualité de vie liée à la santé en général) :

1. VHI 0-11 ( $< 1\sigma$ ) : aucune souffrance ;
2. VHI 12-28 : souffrance peu apparente ;
3. VHI 29-56 : souffrance observable ;
4. VHI 57-120 : souffrance claire.

Selon ces auteurs, toute personne ayant obtenu un score supérieur à 56 ( $> 1\sigma$ ) devrait donc nécessairement être prise en charge.

D'autres versions ont été élaborées à partir du VHI-30, dont :

- une version pédiatrique (Zur et al., 2007), à remplir par les parents ;
- une version adaptée aux chanteurs (Barsties, Kropp, Dicks, Grzondziel, & Morsomme, 2014 ; Morsomme, Gaspar, Jamart, Remacle, & Verduyck, 2007) ;
- une version abrégée (VHI-10) : les dix items les plus pertinents ont été sélectionnés par des professionnels de la voix sur base d'analyses des données de 100 participants dysphoniques et 159 participants normophoniques. Le VHI-10 semble être plus robuste que la version initiale ; il serait également davantage adapté pour les diagnostics différentiels (Clark et al., 2004).

Ricci-Maccarini, De Maio, Murry et Schindler (2013) ont aussi adapté une version du VHI-10 à l'évaluation de l'impact de la dysphonie sur la qualité de vie chez l'enfant : le CVHI-10. Cet outil, à la différence de la version pédiatrique du VHI-30, évalue le ressenti de l'enfant (et non de ses parents).

Finalement, le CVHI-10-P permet d'inclure, en plus, l'avis des parents (Ricci-Maccarini et al., 2016) ; une utilisation conjointe avec le CVHI-10 est conseillée afin de prendre en compte la perspective de tous les intervenants.

***b. Voice Symptom Scale (VoiSS) (Deary, Wilson, Carding, & MacKenzie, 2003)***

Le VoiSS a été créé afin de proposer aux cliniciens un moyen d'évaluation dont les critères psychométriques ont été vérifiés sur un échantillon conséquent. De plus, ils ont souhaité présenter un contenu justifié, basé sur les symptômes identifiés par des individus dysphoniques présentant une grande diversité de pathologies<sup>9</sup>. La version finale du VoiSS (ayant impliqué en tout et pour tout les réponses de 800 participants) (Wilson et al., 2004) comprend trois catégories : incapacité (quinze items), réponse émotionnelle (huit items) et symptômes physiques (sept items). Le patient répond à ces 30 items sur une échelle de fréquence allant de 0 (jamais) à 4 (toujours). Le score total est calculé (sur un maximum de 120 points), ainsi que les sous-scores « incapacité » (60 points), « émotionnel » (32 points) et « physique » (28 points). Le score seuil permettant un équilibre optimal entre sensibilité et spécificité, identifié dans l'étude de Behlau et al. (2016), est de 16.

Le VoiSS a été adapté au brésilien (langue portugaise) en 2014 (« Escala de Sintomas Vocais » - ESV) (Moreti, Zambon, Oliveira, & Behlau, 2014).

Remarquons les avantages du VoiSS par rapport au VHI, explicités par les auteurs (Wilson et al., 2004) :

1. premièrement, les items du VoiSS proviennent de la description des symptômes vocaux par des patients dysphoniques ;
2. deuxièmement, les trois facteurs du VoiSS ont été déterminés statistiquement à l'aide d'analyses factorielles, au contraire des trois dimensions du VHI. Wilson et al. (2004) n'ont en effet identifié que deux facteurs pour le VHI à l'aide de l'analyse factorielle ;
3. troisièmement, un sujet sur quatre ayant contribué à la création du VHI avait subi une ablation du larynx, présentant de ce fait un profil spécifique non généralisable à l'ensemble de la population des dysphoniques.

Lopes et al. (2017) mentionnent également la supériorité du VoiSS par rapport au VHI, le VoiSS étant considéré comme « le protocole le plus fidèle et le plus robuste

---

<sup>9</sup> En effet, l'une des critiques du VHI est que les items ont été sélectionnés sur base de dossiers médicaux de patients dysphoniques (Wilson et al., 2004). Il y a donc eu un double biais de sélection, d'abord par le médecin spécialiste qui juge ou non pertinent de noter certaines plaintes, puis par les auteurs qui ont rassemblé ces plaintes.



psychométriquement parlant pour l'auto-évaluation vocale ». En effet, les auteurs du VoiSS semblent avoir été les plus rigoureux dans la création de leur outil (Branski et al., 2010).

De nombreux autres questionnaires existent encore : Voice Outcome Survey (VOS) (Gliklich, Glovsky, & Montgomery, 1999) qui est un questionnaire pour les immobilités laryngées (Poulain & Sauvignet-Poulain, 2017), Pediatric Voice Outcome Survey (PVOS) (Hartnick, Volk, & Cunningham, 2003), Voice-Related Quality of Life (VRQOL) (Hogikyan & Sethuraman, 1999)/Pediatric Voice-Related Quality of Life (PVRQOL) (Boseley, Cunningham, Volk, & Hartnick, 2006), Pediatric Voice Symptom Questionnaire (PVSQ)—Questionnaire des Symptômes Vocaux chez l'enfant (QSV-enfant) (Verduyckt, Remacle, Jamart, Benderitter, & Morsomme, 2011), Vocal Performance Questionnaire (Carding, Horsley, & Docherty, 1999), Voice Activity and Participation Profile (Ma & Yiu, 2001), Voice Impact Profile (Martin & Lockhart, 2005).

#### **IV. Les mesures objectives**

##### **A. Les mesures aérodynamiques**

Avant de passer aux mesures aérodynamiques quantitatives, il est important d'observer le comportement respiratoire du patient, au repos et en phonation. En effet, un éventuel manque de coordination pneumo-phonatoire pourrait, par exemple, constituer un élément causal de la pathologie vocale. Il convient donc de noter le mode respiratoire préférentiel du patient, ainsi que la présence d'éventuels symptômes telles les prises d'air inappropriées dans le discours, un mécanisme respiratoire inversé (expansion de l'abdomen et du thorax à l'expiration, rétrécissement à l'inspiration), des sifflements lors de la respiration de repos, ou une toux/des raclements de gorge fréquents (Boone, McFarlane, Von Berg, & Zraick, 2010).

Ensuite, différentes mesures aérodynamiques peuvent être effectuées, qui servent à quantifier le rendement vocal. Nous ne citerons que les trois nécessaires au bilan fonctionnel de la voix selon le protocole de l'ELS :

- la mesure à l'aide d'un spiromètre des différents volumes et capacités pulmonaires, dont la capacité vitale (CV) (Boone, McFarlane, Von Berg, & Zraick, 2010) ;
- le temps maximum phonatoire (TMP) : il s'agit de la durée maximale de la voyelle tenue [a] suivant une inspiration profonde. Le TMP moyen se situe entre quinze et vingt secondes selon Woodson et Cannito (1998, cités dans Autesserre et al., 2006) ;
- le quotient phonatoire (QP) : il s'agit du rapport entre la capacité vitale et le TMP, permettant de quantifier le rendement laryngé (Teston, 2004). Il est calculé pour pallier les différences de volume pulmonaire qui se répercutent sur le TMP, notamment en fonction du genre (volume

pulmonaire en moyenne plus grand chez l'homme), mais aussi de l'âge (volume plus réduit chez l'enfant), et d'éventuelles pathologies ou chirurgies pulmonaires (Autesserre et al., 2006). Notons que la mesure de la pression sous-glottique, mesurée indirectement via la pression intraorale, est également une mesure fortement appréciée par les vocologistes, notamment parce qu'une pression sous-glottique anormalement élevée est un indicateur de forçage vocal (Boone, McFarlane, Von Berg, & Zraick, 2010 ; Mehta & Hillman, 2008).

## B. Les mesures acoustiques

Nous avons vu que de nombreux auteurs défendent que la qualité vocale est une réalité perceptive multidéterminée, ne pouvant pas être définie par un corrélat acoustique unique. Cependant, l'évaluation perceptive — bien que communément utilisée par les professionnels de la voix — a maintes fois été décriée. Aussi, Jensen (cité par Kreiman, Gerratt, Kempster, Erman, & Berke, 1993) s'exprima déjà en 1965 : « ... the assumption that the perceptual characteristics of a person's voice provide sufficient information for a reliable description of vocal deviation and its severity appears to be hazardous ».

Ainsi, un intérêt croissant s'est développé pour les mesures objectives, tentant de chiffrer la qualité vocale. Un avantage important de l'évaluation objective/acoustique métrique est une appréciation plus fiable de l'évolution de la prise en charge ; il est alors primordial d'effectuer les enregistrements dans les mêmes conditions d'une passation à l'autre : au même moment de la journée, en donnant les mêmes consignes... (Jo Verstraete, communication personnelle, 4 mai 2017). Les données métriques peuvent également être intéressantes dans le cadre d'études comparant l'efficacité de différents types de prise en charge, par exemple.

Les paramètres analysés par les outils d'analyse acoustique peuvent se regrouper en trois domaines (Barsties & Maryn, 2013) : temps (représentation temporelle du son, oscillogramme), fréquence (représentation spectrale, spectrogramme) et quéfrencence (représentation quéfrentielle, cepstre – voir explication à l'annexe 1, page 150).

Outre les mesures acoustiques « isolées » (tels les phonétogrammes, représentations en deux dimensions des capacités vocales d'un individu : la dynamique d'intensité en ordonnée et l'étendue fréquentielle en abscisse), de véritables index de la qualité vocale ont vu le jour. Prenant en compte l'idée que la qualité vocale est par définition multidéterminée, ces index mesurent et combinent différents paramètres acoustiques choisis selon des bases théoriques distinctes, afin de quantifier la qualité vocale.

Voici certains de ces index, dont l'AVQI que nous décrirons plus en profondeur.

### *a. Dysphonia Severity Index (Wuyts et al., 2000)*

Le DSI est un outil multiparamétrique d'évaluation objective de la qualité vocale. Il combine le jitter (jitt%) — un indice d'irrégularité de la vibration effectué sur la partie centrale de la voyelle [a] tenue —, le temps maximum phonatoire (TMP)<sup>10</sup> — obtenu sur un [a] produit à fréquence confortable et à niveau de pression sonore conversationnelle —, le niveau de pression phonatoire le plus bas (IMin) et la fréquence la plus haute (Fmax).

Ces quatre paramètres ont été sélectionnés parmi treize à l'aide d'une analyse de régression, prédisant au mieux l'évaluation à l'aide du critère G de la GRBAS. L'équation du DSI correspond donc à la formule suivante :

$$DSI = 0.13 \times TMP + 0.0053 \times FMax - 0.26 \times IMin - 1.18 \times Jitter + 12.4$$

Notons qu'une seule version du DSI peut être utilisée pour les femmes ainsi que pour les hommes, car les facteurs TMP et FMax s'égalisent : alors qu'en moyenne, le TMP est plus élevé chez l'homme, la fréquence maximale est plus basse que chez la femme, et vice-versa (Wuyts et al., 2000). L'âge semble cependant avoir un effet significatif sur le score DSI, celui-ci diminuant progressivement (baisse de FMax, du TMP et augmentation du Jitter) (Hakkesteeft, Brocaar, Wieringa, & Feenstra, 2006). Ces observations ont très récemment été confirmées dans une étude investiguant l'impact de l'âge et du genre sur les scores du DSI et de l'AVQI (Barsties, Ulozaite-Staniene, Maryn, Petrauskas, & Uloza, 2017).

Awan et Ensslen (2010) ont mesuré un DSI fortement supérieur pour les individus ayant pratiqué une activité vocale avec un enseignant/coach (jitter et Imin significativement baissés, Fmax augmenté, TMP similaire) et ont conclu qu'il serait nécessaire d'employer des normes différentes pour ceux-ci. Ceci corrobore les résultats obtenus par Timmermans, De Bodt, Wuyts et Van de Heyning (2004), qui ont mesuré une amélioration significative du DSI après un entraînement vocal de neuf mois.

La corrélation mesurée entre le DSI et le critère G est très forte ( $r_s = .996$ ), ainsi que la corrélation entre le score au DSI et l'évaluation de l'impact de la dysphonie à l'aide du VHI ( $r_s = .79$ ).

Un score au DSI se rapprochant de -5 (G = 3) correspond à une voix pathologique ; un score qui se rapproche de +5 (G = 0) correspond à une bonne qualité vocale (tableau 1).

Tableau 1 *Correspondance entre le score G de la GRBAS et le score calculé par le DSI*

G	DSI	écart-type
0	5	0.23
1	1.02	0.25
2	-1.4	0.3
3	-5	0.8

<sup>10</sup> Comme décrit plus haut, le quotient phonatoire est souvent considéré comme plus pertinent car il permet notamment de pallier les différences de volume pulmonaire. Cependant, l'analyse de régression a montré ici que le TMP contribuait à une meilleure prédiction du G que le quotient phonatoire.

D'après Hakkesteegt et al. (2008), qui confirment la validité concourante externe du DSI, le score seuil permettant le meilleur équilibre entre sensibilité et spécificité est de 3.0. En clinique — et notamment pour l'obtention de l'accord INAMI pour le remboursement des séances de bilan et de prise en charge —, le score seuil utilisé pour définir une voix pathologique est de 1.6.

Le DSI a également été validé comme outil de suivi thérapeutique, permettant de détecter les modifications de la qualité vocale suite à une intervention thérapeutique ou chirurgicale (Hakkesteegt, Brocaar, & Wieringa, 2010).

Initialement, les mesures devaient être prises à l'aide du Computerized Speech Lab de KayPentax (Voice Range Profile et Multi-Dimensional Voice Program). Une étude récente montre que le DSI peut aussi être mesuré à l'aide du gratuiciel Praat, en utilisant un script adapté qui permet d'obtenir des mesures fortement corrélées (mais non strictement similaires) aux mesures de la version originale du DSI (Maryn, Morsomme, & De Bodt, 2017). Des études supplémentaires sont cependant nécessaires afin de valider cette nouvelle version du DSI.

Le DSI est largement utilisé en clinique et en recherche, et fait partie de la liste limitative des outils d'évaluation approuvés par l'INAMI. Il comprend un paramètre aérodynamique, un paramètre acoustique du domaine spectral, ainsi que deux mesures appartenant au profil vocal. D'autres index ont vu le jour, investiguant également le domaine cepstral. En effet, comme nous allons le décrire plus loin, les données du domaine cepstral semblent plus fiables que les mesures de perturbation du domaine spectral, notamment. L'un de ces outils prenant en compte le domaine cepstral est le Cepstral Spectral Index of Dysphonia (CSID) d'Awan et Roy (2009).

#### ***b. Le modèle d'Awan et Roy : Cepstral Spectral Index of Dysphonia (CSID)***

Tout d'abord, les auteurs ont identifié les meilleures mesures acoustiques temporelles et spectrales pour la *catégorisation vocale* (c.-à-d. voix normale, rauque, soufflée ou enrouée...) (Awan & Roy, 2005). Cette catégorisation est importante, car elle peut informer le clinicien quant au processus pathologique sous-jacent et ainsi donner des indications pour la prise en charge. Cinq mesures (effectuées sur une voyelle tenue) ont été sélectionnées : le shimmer, le pic de prééminence cepstrale (*cepstral peak prominence*, CPP)<sup>11</sup>, le rapport de la transformation de Fourier discrète (*discrete Fourier transform ratio*, DFTR)<sup>12</sup>, la fréquence fondamentale moyenne et son écart-type converti en demi-tons. Ces cinq mesures permettaient de classer correctement 79.9 % des enregistrements.

---

<sup>11</sup> Voir l'annexe 1 : « le cepstre, kézako ?! », page 150.

<sup>12</sup> DFTR, Discrete Fourier Transform Ratio : rapport de l'énergie à < 4000 Hz et à > 4000Hz ; cette mesure est comparable au « spectral tilt » du modèle de Maryn et al. (2009) décrit ci-après.

Leur deuxième étude avait pour objectif d'évaluer l'utilisation des marqueurs acoustiques pour la détermination de la *sévérité de la dysphonie* (Awan & Roy, 2006). Les mesures acoustiques prédisant au mieux la sévérité de la dysphonie (en rapport avec le jugement perceptif sur une échelle à sept points) étaient : CPP, shimmer, écart-type de F0 converti en demi-tons (sigmaF0) et DFTR. L'équation ainsi générée par les auteurs pour déterminer la sévérité de la dysphonie à partir des mesures acoustiques effectuées sur une voyelle tenue est :

$$\text{Sévérité} = 6.077 + 1.033 \times \text{shimmer} - 0.895 \times \text{sigmaF0} - 0.017 \times \text{DFTR} - 0.089 \times \text{CPP}$$

Une troisième étude d'Awan et Roy (2009) a permis de confirmer que l'utilisation de leur modèle à quatre facteurs est appropriée pour le suivi thérapeutique. En effet, il présente une sensibilité satisfaisante aux changements vocaux dus à une prise en charge vocale et perçus subjectivement (évaluation perceptive sur une échelle à sept points).

Les auteurs ont alors voulu investiguer l'utilisation de leur outil avec des enregistrements de parole continue (Awan, Roy, & Dromey, 2009), à l'aide de phrases du texte anglais phonétiquement équilibré « The Rainbow Passage ». Le degré général de sévérité de la dysphonie de 208 enregistrements a été jugé à l'aide d'une échelle visuelle analogique à 100 points (0 : voix normale, 100 : voix « profondément anormale »). Le modèle s'est révélé concordant avec la sévérité perçue de la dysphonie ; l'élément prédominant était le CPP, expliquant 66 % de la variance de la sévérité perçue. L'équation générée est la suivante :

$$\text{Sévérité} = 154.59 - (\text{CPP} \times 10.39) - (\text{DFTR}_{ET} \times 3.71) - (\text{DFTR} \times 1.08)$$

Finalement, les auteurs ont souhaité valider leur outil en référence à un outil d'évaluation perceptive lui aussi validé : le CAPE-V (Awan, Roy, Jetté, Meltzner, & Hillman, 2010). Le modèle corrélait fortement avec la sévérité perçue par les juges sur base de la parole continue ainsi que sur base de la phonation soutenue et présentait aussi une bonne précision diagnostique. D'autres études de validation ont été menées, confirmant que le CSID (équation de l'étude de Awan, Roy, & Dromey, 2009) est un outil fiable et valide pour les mesures de sévérité et de suivi (Awan, Roy, Zhang, & Cohen, 2016 ; Peterson et al., 2013). Un score cut-off de 24.3 permet d'atteindre une sensibilité de 79.4 % et une spécificité de 79 % (Awan, Roy, Zhang, & Cohen, 2016).

Bien qu'Awan et Roy aient investigué l'évaluation vocale en phonation soutenue et en parole continue, ils n'ont pas créé de modèle permettant d'analyser ces deux conditions simultanément. C'est ici le plus grand avantage d'un autre index, l'Acoustic Voice Quality index (AVQI) de Maryn et al. (2009), qui combine la phonation soutenue et la parole continue dans une même analyse, pour une meilleure validité du modèle.

### ***c. Les limites des mesures acoustiques***

Une critique pouvant être faite envers les outils d'évaluation objective acoustique est que, très fréquemment, les patients dysphoniques se plaignent de variations importantes de leur qualité vocale dans le temps ; nous comprendrons aisément la problématique, dans ce cas, d'une mesure prise à un instant donné. Les mesures acoustiques effectuées sur deux [a] tenus par une même personne, par exemple, sont différentes, même si les deux [a] ont été enregistrés avec un intervalle de quelques secondes. Plus étonnamment encore, les paramètres mesurés sur un même [a] diffèrent selon l'endroit où est placée la fenêtre d'analyse (Révis, 2013).

De plus, il y a actuellement toujours un manque d'accord quant aux corrélats acoustiques de la qualité vocale (Reynolds et al., 2012). Ainsi, les index de qualité vocale incluent dans leurs calculs des paramètres acoustiques différents et n'évaluent donc pas exactement la même réalité multidimensionnelle. Remarquons aussi que certains parmi ces outils sont basés sur des tâches artificielles (phonation soutenue d'une voyelle, par exemple), et ne reflètent ainsi pas suffisamment l'utilisation vocale dans un contexte écologique. En outre, certains de ces outils d'évaluation incluent des paramètres acoustiques qui ne sont pas optimaux pour l'évaluation des dysphonies sévères, dont le signal est fortement apériodique (Heman-Ackah et al., 2003). Il s'agit par exemple des mesures de perturbation « jitter » et « shimmer », qui ne sont interprétables que lorsqu'elles sont inférieures à 5 % (Leong et al., 2012).

## **V. Les mesures acoustiques et l'évaluation subjective par le patient sont-elles corrélées ?**

De nombreuses études ont été menées afin d'observer le lien entre les mesures acoustiques et l'évaluation perceptive par le clinicien, mais aussi entre ces mesures et l'évaluation par le patient de l'impact de la pathologie vocale sur sa qualité de vie. Ces études montrent principalement des résultats contradictoires, en fonction des outils et de la méthodologie employés. Concernant l'évaluation perceptive par le clinicien, Eadie et Doyle (2005) soulignent que la précision diagnostique n'atteint 100 % que lorsque les évaluations perceptive et acoustique sont combinées. L'on observe effectivement une tendance générale à conseiller la combinaison des mesures subjectives et objectives. Ceci est vrai aussi pour l'évaluation par le patient : la plupart des auteurs considèrent aujourd'hui l'évaluation objective et l'évaluation subjective par le patient comme *complémentaires* et toutes deux indispensables à la réalisation d'un bilan vocal complet. Voici quelques études démontrant cette idée :

Da Costa de Ceballos et al. ont mené une étude sur 476 enseignants, afin de mesurer la validité du VHI-10. Les critères externes choisis étaient les mesures acoustiques de jitter et de

shimmer, ainsi que l'évaluation perceptive par le clinicien via l'échelle GRBAS. Les auteurs ont observé que les indicateurs de validité intrinsèque (sensibilité, spécificité,  $LR^+$ ,  $LR^-$ ) du VHI-10 étaient très faibles lorsque les mesures acoustiques étaient utilisées comme critère externe ; ils étaient quelque peu meilleurs, mais toujours faibles, lorsque le critère externe était l'évaluation subjective par le clinicien (critère G). Cette étude indique que le VHI-10 est davantage en lien avec l'évaluation subjective par le clinicien qu'avec les mesures acoustiques (Da Costa de Ceballos, Carvalho, De Araujo, & Farias Borges dos Reis, 2010).

Une autre étude menée par Gillespie et al. avait pour objectif de mesurer la corrélation entre les changements pré- et post-thérapeutiques mesurés à l'aide du VHI-10 et neuf autres mesures (acoustiques, aérodynamiques et évaluation subjective par le clinicien<sup>13</sup>), et ce dans cinq pathologies vocales différentes. Seule l'évaluation subjective par le clinicien (CAPE-V) était modérément corrélée au VHI-10 (scores pré- et post-thérapie moyennés). Les mesures aérodynamiques avaient évolué après la prise en charge, mais seul le changement du débit d'air moyen en parole continue était corrélé avec le changement du VHI-10, et ce uniquement pour les individus atteints d'une paralysie unilatérale des plis vocaux. Les mesures acoustiques s'avérèrent inchangées au fil du temps. Les mesures acoustiques et le score au VHI semblent donc également indépendants dans cette étude. Les auteurs expliquent ce constat par le fait que les mesures acoustiques objectivent un phénomène physique/physiologique, alors que le VHI-10 en mesure la conséquence (le handicap causé) (Gillespie, Gooding, Rosen, & Gartner-Schmidt, 2014).

Awan, Roy et Cohen (2014) ont investigué la corrélation entre le VHI-30 et leur outil d'évaluation objective, le CSID. Des corrélations faibles à modérées ont été observées entre les différentes mesures spectrales et cepstrales (parole continue et phonation soutenue) et le score au VHI. Aussi, ces auteurs rejoignent l'avis de nombreux autres (dont Lopes et al., 2017), concluant que l'évaluation objective et l'évaluation par le patient doivent toutes deux être considérées comme des outils complémentaires.

Schindler et al. ont eux aussi mesuré la corrélation entre le VHI et des mesures objectives (TMP, jitter, shimmer, HNR, F0 moyenne), dans des groupes dysphoniques de différentes étiologies. Les corrélations significatives mesurées variaient selon l'étiologie de la dysphonie : une corrélation significative existait dans la dysphonie fonctionnelle entre le TMP et les dimensions fonctionnelle et physique du VHI ; dans le cas de nodules, une corrélation entre le jitter, le shimmer, le HNR et la dimension fonctionnelle du VHI. Les auteurs en

---

<sup>13</sup> Les neuf mesures étaient : pression sous-glottique, niveau de pression sonore et débit d'air moyen sur une voyelle tenue à hauteur et intensité confortables ; niveau de pression sonore, fréquence et débit d'air moyen en parole continue ; résistance laryngée, HNR et sévérité globale (CAPE-V).

concluent que les patients ne sont pas attentifs aux mêmes aspects de leur voix en fonction de l'étiologie de leur dysphonie, ce qui donne de bonnes indications pour la prise en charge (Schindler, Mozzanica, Vedrody, Maruzzi, & Ottaviani, 2009).

Dans l'étude de Woisard et al. (2007) concernant le VHI-30, aucune corrélation n'a été trouvée entre les scores au VHI (total et sous-scores) et les mesures acoustiques et aérodynamiques<sup>14</sup>. Seules les mesures du profil vocal (fréquence minimale et étendue vocale) étaient faiblement corrélées avec les scores au VHI (Woisard, Bodin, Yardeni, & Puech, 2007).

Lopes et al. (2017) ont mesuré les corrélations entre différentes mesures acoustiques (déviations standard de F0, jitter, shimmer, glottal to noise excitation ratio) et les outils VHI et VoiSS. Cette étude a montré que toutes les mesures acoustiques étaient faiblement positivement corrélées avec le VoiSS, mais aucune n'était corrélée avec le VHI. Les auteurs expliquent cela par le fait que le VHI, comparé au VoiSS, évalue davantage l'impact psychosocial que les aspects physiques ou physiologiques. Or, l'impact psychosocial de la dysphonie dépend moins de la sévérité et des paramètres de la dysphonie, que de facteurs externes (Awan, Roy, & Cohen, 2014), tels le milieu de vie du patient, son attitude par rapport à la dysphonie, l'attitude de ses partenaires de communication et son état émotionnel.

Ainsi, il est nécessaire de nuancer la présence ou l'absence de corrélations mesurées entre les paramètres acoustiques et les différentes échelles d'évaluation subjective par le patient, qui ne mettent pas toutes l'accent sur le même aspect. Mentionnons aussi les plaintes de patients concernant une voix qui a changé, qui n'est « plus la voix habituelle », même si elle n'est pas de prime abord une voix pathologique !

Toutes ces données soulignent l'importance d'une évaluation complète à l'aide de différents outils complémentaires : anamnèse, mesures objectives acoustiques et aérodynamiques et évaluation subjective par le patient et par le thérapeute. L'intégration de l'ensemble de ces données permet de considérer les différents aspects des conséquences d'une pathologie vocale.

---

<sup>14</sup> Les mesures acoustiques et aérodynamiques étaient : jitter, shimmer, HNR, Fmin et Fmax, étendue vocale, Imin, TMP, débit d'air moyen et pression sous-glottique.



## L'AVQI – Acoustic Voice Quality Index

Cet outil d'évaluation quantitative de la qualité vocale a été développé par Maryn, Corthals, Van Cauwenberge, Roy, et De Bodt (2009) en réponse à la demande grandissante d'outils d'évaluation vocale objective.

Comme mentionné ci-haut, l'AVQI a le grand avantage de combiner la phonation soutenue et la parole continue. De plus, cet index a bénéficié de nombreuses études de validation (par rapport à des outils eux-mêmes fortement documentés et validés) et de réplification, qui en font un modèle robuste et fiable.

Le calcul de l'AVQI (dès la version 02.02) se fait grâce au logiciel gratuit « Praat ». Il se base sur une analyse de régression linéaire prenant en compte six facteurs acoustiques afin de mesurer la qualité vocale globale (correspondant, en évaluation perceptive, au critère G de la GRBAS).

L'originalité de cet outil est l'utilisation conjointe d'un extrait de trois secondes de la voyelle [a] tenue (sans la phase de mise en vibration des plis vocaux — VOT — et la phase finale de la voyelle<sup>15</sup>) et de l'extrait d'un texte phonétiquement équilibré (dépendant de la langue utilisée). En tant que premier outil combinant la voyelle tenue et l'extrait de parole, l'AVQI a voulu garantir une validité écologique maximale.

En effet, Maryn et al. (2009) citent six avantages de la voyelle tenue :

1. elle présente une phonation relativement stable, avec une variabilité fréquentielle et temporelle limitée qui en facilite l'analyse acoustique ;
2. la phonation soutenue ne contient pas de phonèmes non voisés et n'est pas influencée par les délais d'établissement et d'arrêt du voisement (Voice Onset/Offset Time) ni par les modifications prosodiques (sur la fréquence et sur le niveau de pression sonore) ;
3. elle est relativement indépendante des facteurs individuels, telle la vitesse de parole, et des facteurs langagiers tels le contexte phonétique et l'accent tonique ;
4. les mesures traditionnelles de perturbation de la fréquence (jitter) et du niveau de pression sonore (shimmer) se basent sur la détection de la hauteur vocale, qui est mesurée plus efficacement dans la phonation soutenue (sans influence de la prosodie ou du VOT, par exemple) ;

---

<sup>15</sup> Notons que ce découpage de la partie centrale “stable” de la voyelle peut être critiqué par le fait que les phases initiale et finale – respectivement de mise en vibration et d'arrêt du voisement – sont des indices non négligeables en pratique clinique. En effet, la mise en vibration des plis vocaux est souvent problématique lors des conditions vocales pathologiques (Pouchoulin, 2008).

5. sa production demande moins d'effort et permet une mesure standardisée ;
6. elle n'a pas de charge linguistique, et n'est donc pas encline à varier selon les dialectes régionaux, la langue, le niveau cognitif...

Cependant, la phonation soutenue n'est pas représentative de l'utilisation de la voix au quotidien. Les mesures seront plus fiables grâce à l'ajout d'un échantillon de parole, car de nombreux phénomènes glottiques ou supraglottiques sont négligés lors de l'analyse d'une simple voyelle tenue : la prosodie, les intonations, les pauses, les démarrages et arrêts vocaux rapides... De plus, l'analyse de perturbations fréquentielles ou périodiques nécessite également des sources plus variées qu'un simple son (Hauptmann, 2013).

C'est pour cela que l'intégration de la parole continue dans l'analyse permet d'augmenter la validité écologique de l'outil. De plus, les plaintes vocales sont principalement émises par rapport à la parole continue, la dysphonie pouvant même parfois n'apparaître que dans ce contexte. Néanmoins, la parole continue est soumise à de nombreuses variations inter- et intra-individuelles, en fonction notamment de la langue/du dialecte.

Il s'agit donc, pour l'AVQI, d'utiliser les segments voisés de la parole continue (extraits à l'aide du script proposé par Maryn et al., à intégrer au logiciel Praat) et des trois secondes centrales de la voyelle [a] tenue. Les segments voisés et la voyelle sont concaténés et séparés d'un intervalle de deux secondes de silence par le script (sur Praat) pour que puisse y être calculé le score AVQI compris entre 0 et 10.

Les analyses effectuées sur ces deux enregistrements sont au nombre de six. En effet, Maryn et al. (2009) expliquent que la valeur prédictive d'un seul marqueur acoustique est trop faible et n'est pas suffisante pour analyser la voix dont la nature est multidimensionnelle.

Dans la littérature scientifique, de nombreux paramètres acoustiques sont utilisés pour l'appréciation de la voix. Maryn et al. menèrent une méta-analyse reprenant les résultats de 25 études. Sur un total de 87 paramètres acoustiques étudiés, seuls six se révélèrent prometteurs et seule la proéminence du pic cepstral lissé (smoothed cepstral peak prominence) répondait à tous les critères de la méta-analyse. C'est donc cette mesure qui paraît être la plus prometteuse pour la détermination de la sévérité des dysphonies (Maryn, Roy, De Bodt, Van Cauwenberge, & Corthals, 2009).

Les six paramètres utilisés par l'AVQI se décomposent en une analyse cepstrale et cinq analyses spectrales (toutes effectuées avec le logiciel Praat depuis la version 02.02) :

- Les paramètres du domaine spectral :
- o *le rapport harmoniques/bruit* (« harmonics to noise ratio » - HNR) : il s'agit d'un paramètre mesurant le rapport entre les parties harmoniques et non harmoniques du son. C'est le

logarithme en base 10 du rapport entre l'énergie périodique et l'énergie du bruit, multiplié par 10 (Barsties & Maryn, 2015) :

$$10 \times \log_{10} \frac{\text{énergie périodique}}{\text{énergie du bruit}}$$

Un HNR diminué est souvent associé à une voix soufflée (Mathieson, 2013) ;

- *le shimmer local* : c'est une mesure des « perturbations » sonores, c'est à dire des déviations de la périodicité de l'onde sonore. Le shimmer concerne les déviations au niveau de l'amplitude (niveau de pression sonore), comme représenté à la figure 11.

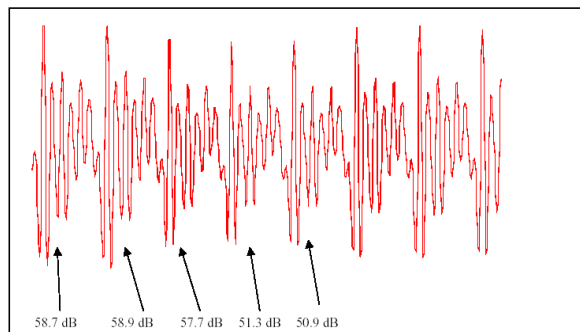


Figure 11. Illustration des variations périodiques d'amplitude (Verstraete, 2017).

Le shimmer local mesure l'écart moyen absolu des amplitudes de différentes périodes consécutives, divisé par l'amplitude moyenne (Boersma, 2017) :

$$\frac{\sum | \text{amplitude}_{\text{période } k} - \text{amplitude}_{\text{période } k+1} |}{n \times \text{amplitude}_{\text{moy}}}$$

Où « n » est le nombre de périodes extraites pour l'analyse.

Les perturbations de l'amplitude sont généralement associées à une voix rauque, parfois soufflée (Ferrer, De Bodt, Maryn, Van de Heyning, & Hernandez-Diaz, 2007 ; Heman-Ackah, Michael, & Goding, 2002).

- *le shimmer local dB* : il s'agit ici de la moyenne absolue du logarithme en base 10 multiplié par 20, de la différence entre les amplitudes de périodes consécutives (Boersma, 2017 ; Teixeira, Oliveira, & Lopes, 2013) :

$$\frac{20 \times \sum \left| \log_{10} \frac{\text{amplitude}_{\text{période } k+1}}{\text{amplitude}_{\text{période } k}} \right|}{n - 1}$$

Avec « n » étant le nombre de périodes analysées.

- *la pente du spectre moyen à long terme* (« *general slope* ») : le spectre moyen à long terme (SMLT ou *long-term average spectrum* — *LTAS* en anglais ; en noir sur la figure 12) est un spectre obtenu en moyennant un grand nombre de spectres à intervalles réguliers (p.ex. toutes les 200 ms, donc cinq spectres/seconde) (Master, De Biase, Pedrosa, & Chiari, 2006). Les mesures du spectre moyen à long terme (tout comme les mesures cepstrales) ont gagné de

l'intérêt dans l'évaluation vocale acoustique, car elles permettent l'évaluation de la sévérité de la dysphonie sur des échantillons de parole continue (Lowell, Colton, Kelley, & Hahn, 2011). L'intensité des différents harmoniques d'un son décroît avec l'augmentation des fréquences ; ainsi, la « pente spectrale » (exprimée en dB), reliant les points d'amplitude maximale entre les harmoniques, chute. On parle généralement d'une chute de 12 dB par octave en voix « normale ». Un son à niveau de pression sonore élevé montre une pente spectrale moindre qu'un son à un faible niveau de pression sonore (Alipour, Scherer, & Finnegan, 2012 ; Baken & Orlikoff, 2000). La pente du spectre moyen à long terme se mesure en divisant l'énergie des fréquences basses (0-1000Hz) par celle des fréquences hautes (1000-10000Hz) à travers le spectre moyen à long terme (Volín & Zimmermann, 2011).

- *l'inclinaison de la courbe de tendance à travers le spectre (« tilt »)* : il s'agit de la différence d'énergie de la droite de régression (prédiction de l'énergie en fonction de la fréquence, en rouge à la figure 12) entre 0-1000Hz et 1000-10000Hz à travers le spectre moyen à long terme (Barsties & Maryn, 2015 ; Smith & Goberman, 2014).

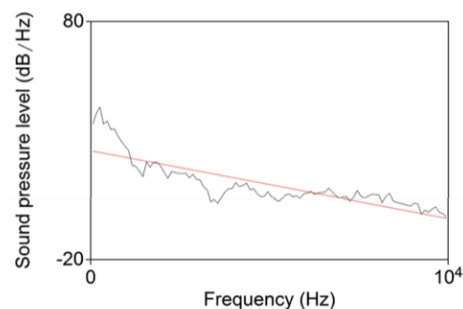


Figure 12. Représentation du « spectral tilt » (Delgado, León, Jiménez, & Izquierdo, 2017).

En noir = spectre moyen à long terme ; en rouge = droite de régression à travers le spectre moyen à long terme.

#### - Le paramètre du domaine cepstral :

Heman-Ackah et al. (2003) conduisirent une étude suite au constat que les analyses spectrales classiques (shimmer, jitter, HNR) étaient peu fiables. En effet, celles-ci sont trop dépendantes de la sévérité de la dysphonie, mais aussi des modulations du tractus vocal dans la parole spontanée et d'éventuels tremblements physiologiques (Schoentgen, 2003). Ces indices vocaux de perturbation sonore ne devraient pour cette raison être utilisés qu'en cas de dysphonie légère, voire modérée, et sur base de sons continus (Autesserre et al., 2006).

Heman-Ackah et al. (2003) analysèrent ainsi l'utilité de l'analyse cepstrale, plus précisément de la proéminence du pic cepstral lissé (smoothed cepstral peak prominence) (représentée à la figure 13), calculée sur l'analyse de Fourier du spectre vocal. C'est cette même mesure qui est employée dans l'AVQI : il s'agit de la distance entre l'amplitude du pic cepstral et l'amplitude

du point ayant la même quéfrence sur la droite de régression à travers le cepstre lissé<sup>16</sup> (donc l'amplitude prédite à ladite quéfrence par la droite de régression) (Hillenbrand, Cleveland, & Erickson, 1994) :

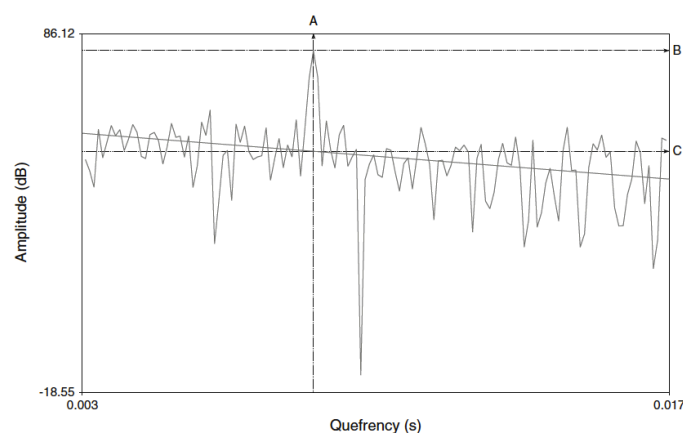


Figure 13. Représentation du pic cepstral (Maryn, De Bodt, Barsties, & Roy, 2014).

La proéminence du pic cepstral (A) est calculée en mesurant la différence entre le sommet de celui-ci (B) et le point de même quéfrence sur la droite de régression (C).

L'utilisation de données du domaine quéfrentiel est intéressante, car, comme le soulignent Heman-Ackah et al. en 2003, les dysphonies plus sévères, caractérisées par un signal peu périodique<sup>17</sup>, se prêtent plus difficilement aux analyses du domaine fréquentiel tels le shimmer, le jitter et le rapport signal/bruit. Ces mesures impliquent un suivi des variations cycle par cycle du signal vocal. Elles sont également plus sensibles à des variations minimales de la vibration — variations non nécessairement associées à une qualité vocale diminuée (Ferrer, De Bodt, Maryn, Van de Heyning, & Hernandez-Diaz, 2007 ; Hillenbrand, 1987).

De plus, les mesures cepstrales permettent (comme les mesures du spectre moyen à long terme) l'évaluation du degré de la dysphonie sur un échantillon de parole continue (Delgado-Hernández, León-Gómez, Izquierdo-Arteaga, & Llanos-Fumero, 2018 ; Lowell et al., 2011 ; Sauder, Bretl, & Eadie, 2017).

Une explication plus détaillée du cepstre et des analyses cepstrales est, au besoin, disponible à l'annexe 1 (page 150).

Les étapes de la procédure d'analyse de l'AVQI sont représentées à l'annexe 2 (page 160).

<sup>16</sup> Le PPC est mesuré après un « lissage », c'est à dire un moyennage d'un certain nombre de fenêtres cepstrales avant et après la fenêtre d'intérêt (Heman-Ackah et al., 2003 ; Hillenbrand & Houde, 1996).

<sup>17</sup> Le National Center for Voice & Speech a publié en 1994 des recommandations quant à l'analyse acoustique de la voix. Dans ce communiqué, Titze décrit trois types de signaux vocaux, selon leur périodicité. Seuls les signaux de type I, pour lesquels l'énergie sonore de la fréquence fondamentale est assez saillante, seraient appropriés pour les analyses de perturbation classiques. Pour les signaux de type 2, il serait préférable d'utiliser des analyses visuelles (par exemple de spectrogramme) ; pour les signaux de type 3, c'est l'évaluation perceptive de l'apériodicité (raucité, souffle...) qui est conseillée (Titze, 1994).

## **I. L'évolution de l'AVQI et les études de validation**

De nombreuses études ont été menées concernant la validité (concourante et interne), la précision diagnostique, ainsi que la sensibilité de cet outil aux changements vocaux dus à une prise en charge vocale. L'AVQI s'en sort avec de très bons résultats pour chaque critère psychométrique, et ce même dans différentes langues, ainsi que chez l'enfant. Voici un aperçu de l'évolution de l'AVQI depuis l'étude initiale de Maryn et al. en 2009 :

- 1) Dans leur étude initiale, Maryn et al. ont voulu évaluer le bien-fondé de l'utilisation conjointe de la phonation soutenue et de la parole continue (Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2009). Un algorithme de détection de vocalisation a été intégré à Praat afin d'extraire les segments voisés des enregistrements de parole continue et de concaténer ceux-ci avec la voyelle tenue. Treize paramètres acoustiques ont alors été mesurés, et, à l'aide d'une régression multiple pas à pas<sup>18</sup>, six d'entre eux ont été retenus. Ces six paramètres ont permis aux auteurs de créer un modèle acoustique pour évaluer la qualité vocale : l'AVQI. L'équation résultante de cette étude est la suivante :

$$AVQI = (3.295 - 0.111 \times CPPs - 0.073 \times HNR - 0.213 \times shimmer\ local + 2.789 \times shimmer\ local\ dB - 0.032 \times slope + 0.077 \times tilt) \times 2.571$$

Le score seuil optimal calculé est de 2.95. Notons que l'évaluation de la qualité vocale à l'aide de l'AVQI est indépendante de l'âge et du genre de l'individu évalué (Barsties & Maryn, 2013). Ceci a été très récemment confirmé par Barsties, Ulozaitė-Staniene, Maryn, Petrauskas et Uloza (2017, article sous presse).

Une relation positive entre l'AVQI et le critère G a également été mesurée, confirmant sa validité externe (généralisabilité à la population) : au plus le score AVQI est grand, au plus la qualité vocale est perturbée. Remarquons que dans sa version initiale, le calcul de l'AVQI se faisait grâce aux deux logiciels gratuits « Praat » et « SpeechTool » (James Hillenbrand, Western Michigan University, Kalamazoo, MI, USA).

- 2) La sensibilité de l'AVQI à des changements dus à un traitement (chirurgical ou comportemental) a ensuite été investiguée (Maryn, De Bodt, & Roy, 2010). Les modifications perçues (critère G de la GRBAS) étaient fortement corrélées avec les

---

<sup>18</sup> Il s'agit d'une méthode couramment utilisée pour déterminer, parmi plusieurs variables indépendantes d'un modèle de régression multiple, lesquelles prédisent au mieux la variable dépendante. On introduit la variable avec le plus grand coefficient de corrélation dans un nouveau modèle et on vérifie la significativité de ce coefficient. On introduit alors pas à pas dans ce modèle les autres variables dans l'ordre décroissant de leur coefficient de corrélation. Si la significativité du coefficient d'une variable est rejetée, on ne la retient pas (Facultés universitaires Notre-Dame de la Paix, 2009).

modifications proportionnellement similaires de l'AVQI. Les auteurs ont donc pu confirmer qu'il s'agissait là d'un bon outil de suivi de traitement, permettant au clinicien ou au chirurgien d'évaluer l'efficacité de leur prise en charge.

- 3) Reynolds et al. ont validé l'utilisation de l'AVQI dans une population pédiatrique anglophone (2012). Le score seuil 2.95 initialement calculé par Maryn et al. (2009) ne semblant pas adapté à l'évaluation chez l'enfant, un nouveau score seuil a été calculé : 3.46.
- 4) En 2013, les auteurs ont confirmé la fidélité test-retest de l'AVQI (Barsties & Maryn, 2013). De plus, ils ont évalué la part de la variance du score AVQI due à chacune des deux mesures — phonation soutenue et parole continue. Ils ont conclu que la phonation soutenue jouait un plus grand rôle que la parole continue dans l'obtention du score total. Selon les auteurs, ce phénomène pouvait être expliqué par les différences de durée de ces deux conditions : en effet, l'échantillon de phonation soutenue analysé a toujours une durée de trois secondes, alors que la durée des fragments voisés extraits de la parole continue varie en fonction des individus. Ainsi, les auteurs émirent comme perspective la tentative d'une égalisation des durées des deux conditions pour le calcul de l'AVQI.
- 5) La composante principale parmi les mesures acoustiques de l'AVQI, la proéminence du pic cepstral lissé, était initialement effectuée à l'aide du logiciel SpeechTool, contrairement aux autres mesures effectuées à l'aide du programme Praat. L'inclusion de toutes les mesures en une seule formule exigeait la réalisation d'un grand nombre d'étapes. Aussi, en 2015, les auteurs ont généré un nouveau script pour le calcul de l'AVQI à partir de l'unique logiciel Praat, rendant le processus d'évaluation bien moins laborieux (Maryn & Weenink, 2015) :

$$AVQI_{02.02} = 9.072 - 0.245 \times CPPS - 0.161 \times HNR - 0.470 \times \text{shimmer local} + 6.158 \times \text{shimmer local} \\ dB - 0.071 \times \text{slope} - 0.170 \times \text{tilt}$$

Le CPPS calculé à l'aide du logiciel SpeechTool et le CPPS calculé par Praat se sont révélés très similaires et l'équivalence entre les deux formules de l'AVQI a été confirmée par les auteurs.

- 6) Suite à l'étude de 2013, ayant montré une supériorité de la phonation soutenue dans l'explication de la variance totale du score AVQI par rapport à la parole continue, les auteurs ont généré une nouvelle version de l'AVQI (03.01). Celle-ci permet d'équilibrer les parts de variance expliquée (Barsties & Maryn, 2015) :

$$AVQI_{03.01} = (4.152 - 0.117 \times CPPS - 0.006 \times HNR - 0.037 \times \text{shimmer local} + 0.941 \times \text{shimmer local} \\ dB + 0.01 \times \text{slope} + 0.093 \times \text{tilt}) \times 2.8902$$

Pour l'utilisation de cette version, la durée moyenne nécessaire pour la parole continue inclut 34 syllabes (du texte néerlandais « Papa en Marloes »).

- 7) Parmi les nombreux facteurs affectant la fidélité des évaluations subjectives, les auteurs ont précisé que l'influence de la longueur de l'échantillon vocal n'avait pas encore été investiguée. Ils ont donc évalué la durée de parole continue nécessaire afin d'obtenir des mesures stables (Barsties & Maryn, 2017). Cette nouvelle étude a permis de conclure qu'une longueur de deux phrases du texte néerlandais ne paraît pas suffisante, et que l'utilisation du texte entier n'amène aucune amélioration significative. Ainsi, afin de permettre une évaluation perceptive correcte tout en économisant le temps de l'évaluation, il est préférable d'utiliser une durée de parole continue équivalente aux trois secondes de la phonation soutenue (une fois les segments voisés extraits), ce qui correspond en moyenne à 35.5 syllabes pour le texte néerlandais.
- 8) Une étude de réplication a ensuite été menée par Barsties et Maryn (2016) afin de confirmer la validité de la troisième version de l'AVQI générée en 2015. Les critères psychométriques étant satisfaisants (validité externe et précision diagnostique), les auteurs ont conclu que la troisième version de l'AVQI semblait être la plus appropriée pour la pratique clinique. En effet, elle combine la facilité d'utilisation apportée par la version 2 (implémentation de la mesure du CPPS dans Praat) et l'équilibrage des proportions de voisement analysées pour les deux conditions vocales (phonation soutenue et parole continue).
- 9) L'AVQI version 03.01 a été validé en japonais (Hosokawa et al., sous presse) et en espagnol (Delgado et al., 2018), avec respectivement une longueur optimale de parole continue de 30 et de 33 syllabes, et un score seuil optimal de 2.06 et de 2.28.
- 10) Une étude récente confirme que l'AVQI (version 02.02) permet, tout comme le DSI, de différencier les voix normophoniques des voix dysphoniques. Il semblerait même que l'AVQI présente une plus grande précision diagnostique que le DSI (Uloza, Barsties, Ulozaite-Staniene, Petrauskas, & Maryn, 2018).
- 11) La validité interne de l'AVQI version 03.01 a été étudiée en langue allemande dans une étude de Barsties et Lehnert (2018). Sa sensibilité aux changements concernant la raucité et le caractère soufflé de la voix permettent de détecter une amélioration de la qualité vocale suite à une prise en charge conservatrice (Barsties & Lehnert, 2018).



## II. L'AVQI, un outil multilingue

Différentes études ont été menées afin d'examiner l'utilisation de l'AVQI sur des extraits de parole continue dans d'autres langues que le néerlandais. Maryn, De Bodet, Barsties et Roy (2014) précisent en effet que l'évaluation vocale peut dépendre de la langue parlée par le sujet. Un trait jugé pathologique dans une langue peut être nécessaire pour la discrimination phonémique dans une autre. Ladefoegd (cité par Maryn et al., 2014) exprime cette réalité ainsi : « la dysphonie d'une personne peut être le phonème d'une autre personne ». Les auteurs citent comme exemple la raucité, qui est nécessaire à la discrimination phonémique dans le danois, alors qu'elle est jugée comme un trait pathologique en néerlandais (et en français). De plus, les contextes phonémiques des voyelles dans la parole continue ainsi que les courbes intonatives varient en fonction des langues (Maryn et al., 2014). Les variations du niveau de pression sonore dues à la prosodie, entre autres, ont un impact important sur les indices de perturbation (Schoentgen, 2016).

Voici dans le tableau 2 le résumé des résultats des études dans différentes langues :

Tableau 2 Mesures psychométriques des études de validation de l'AVQI dans différentes langues

Langue	population	n	ASC	score-seuil	sensibilité	spécificité	LR <sup>+</sup>	LR <sup>-</sup>	validité concourante <sup>19</sup>
<b>Néerlandais</b> (Maryn et al., 2009)	adultes	251	.895	2.95	74 %	96 %	19.98	0.27	.780
<b>Néerlandais</b> (Maryn et al., 2010)	adultes	39	.920	2.95	85 %	100 %	>2000	0.16	.796
<b>Néerlandais</b> (Maryn et al., 2014)	adultes	50	.893	3.19	92 %	73 %	3.41	0.11	.808
<b>Néerlandais</b> (Maryn et al., 2014)	adultes	50	.894	3.66	85 %	80 %	4.25	0.19	.809
<b>Néerlandais : version 03.01</b> (Barsties & Maryn, 2016)	adultes	1058	.923	2.43	78.5 %	93.2 %	11.54	0.23	.815
<b>Anglais</b> (Reynolds et al, 2012)	enfants	107	.876	3.46	82 %	92 %	10.25	0.20	.794
<b>Anglais</b> (Maryn et al., 2014)	adultes	50	.953	3.29	90 %	90 %	9.00	0.11	.868
<b>Anglais</b> (Maryn et al., 2014)	adultes	50	.956	3.25	95 %	82 %	5.28	0.06	.849
<b>Allemand</b> (Barsties & Maryn, 2012)	enfants et adultes	61	.888	2.70	79 %	92 %	/	/	.790
<b>Allemand</b> (Maryn et al., 2014)	adultes	50	.958	3.05	98 %	75 %	3.92	0.03	.858
<b>Français</b> (Maryn et al., 2014)	adultes	50	.869	3.07	97 %	70 %	3.23	0.04	.781

<sup>19</sup> Coefficient de corrélation entre AVQI et G (GRBAS)

<b>Japonais</b> (Hosokawa et al., 2017)	adultes	267	.905	3.15	73 %	95 %	15.1	0.29	.828
<b>Japonais</b> (Hosokawa et al., sous presse) — version 03.01	adultes	455	.915	2.06	72.1 %	93.8 %	11.7	0.298	.873
<b>Finlandais</b> (Kankare et al., 2015)	adultes	50	.898	2.35	/	/	/	/	.810
<b>Coréen</b> (Maryn, Kim, & Kim, 2016)	adultes	56	/	/	/	/	/	/	G : .911
									OS <sup>20</sup> : .924
<b>Lithuanien</b> (Uloza et al., 2017)	adultes	184	G .940	2.97	84 %	94 %	13.24	0.17	.852
			VAS <sup>21</sup> .962	3.48	84 %	92 %	10.81	0.17	.876
<b>Espagnol : version 03.01</b> (Delgado, Gómez, Jiménez, Izquierdo, & Barsties, 2018)	adultes	183	.905	2.28	74.8 %	94.6 %	13.85	0.26	.835

Note. ASC = aire sous la courbe ROC (area under ROC curve) ; LR+ = rapport de vraisemblance positif (positive likelihood ratio) ; LR- = rapport de vraisemblance négatif (negative likelihood ratio).

De nombreux autres outils de mesure acoustique existent, dont : le Göttinger Heiserkeitsdiagramm (Michaelis, 1999), l'Artificial Neutral Network-based Classification (Linder, Albers, Hess, Pöpl, & Schönweiler, 2008) et l'Acoustic Breathiness Index (Barsties, Maryn, Gerrits, & De Bodt, 2017). Nous ne pouvons cependant pas tous les décrire dans le cadre de ce travail.

Notons qu'il existe même un outil (néerlandophone) utilisant les mesures acoustiques pour évaluer la voix trachéo-œsophagienne (Maryn, Dick, Vandenbrouaene, Vauterin, & Jacobs, 2009).

<sup>20</sup> critère "Overall Severity" de l'échelle CAPE-V

<sup>21</sup> critère "Visual Analogic Scale" de l'échelle CAPE-V

**PARTIE II : L'AVQI 03.01 APPLIQUÉ AU FRANÇAIS**

Nous avons vu que l'AVQI de Maryn et al. n'a été validé en français que dans leur étude de 2014 sur des participants quadrilingues, dont la langue maternelle était le néerlandais. De plus, les indices psychométriques (validité concourante et précision diagnostique) obtenus pour la phrase du texte « La bise et le soleil » en français étaient les moins concluants (voir tableau 2, page 42). Les auteurs expliquent cette faiblesse par le fait qu'il s'agit de la seule langue romane parmi celles étudiées, ce qui en fait la langue la plus distante de la langue néerlandaise. Enfin, aucune analyse psychométrique n'a été effectuée concernant la version 03.01 de l'AVQI sur un échantillon en français.

Le premier objectif de notre étude était donc de valider l'AVQI sur un nouvel échantillon en langue française. En effet, une enquête en ligne remplie par 25 logopèdes wallon(ne)s spécialisé(e)s dans les prises en charges vocales révèle que seuls 40 % (10/25) d'entre eux utilisent actuellement l'AVQI. 4 % (1/25) disent en avoir tenté l'utilisation, mais ne pas avoir été convaincus. 56 % (14/25) n'utilisent pas l'AVQI et n'en ont jamais tenté l'utilisation. Les résultats de cette enquête sont représentés sur le diagramme à la figure 14. Le nombre d'années de pratique des logopèdes ayant répondu au sondage varie entre 1 et 28, avec une médiane de 12.5 ans (intervalle interquartile : 15.6 ans). La part de patientèle dysphonique parmi l'entièreté des patients pris en charge varie entre 6 % et 80 %, avec une médiane de 37.5 % (intervalle interquartile : 47.5 %).

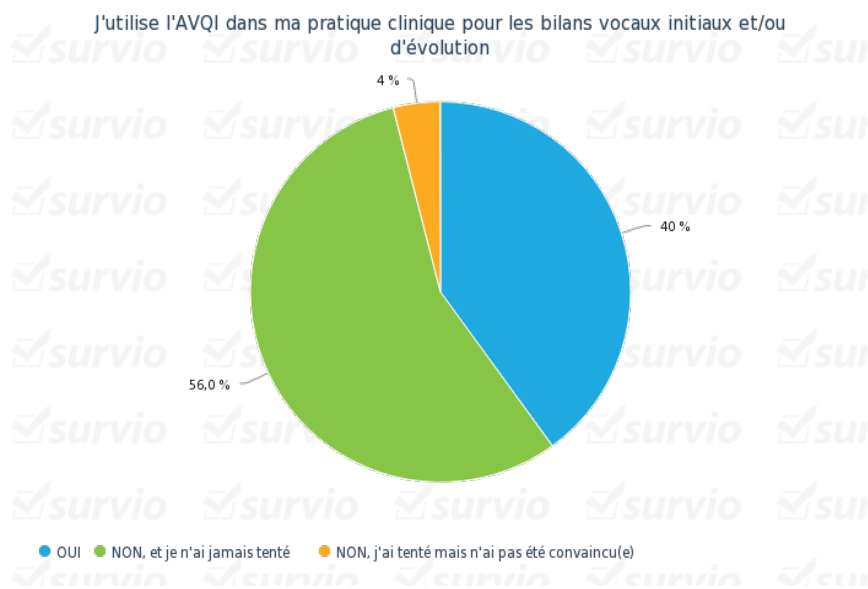


Figure 14. Résultats de l'enquête en ligne quant à la question de l'utilisation de l'AVQI.

Les autres données de cette enquête révèlent que 92 % (23/25) de ces logopèdes affirment que le local dans lequel sont effectués les bilans vocaux n'est pas insonorisé. Concernant le matériel utilisé, la plupart d'entre eux/elles semblent utiliser un microphone sur pied (p.ex. Shure PG27, Shure SM58, Yeti, Sennheiser E935, Vivanco DM 39) ; certains utilisent un micro-casque (p.ex. Logitech H390), d'autres le dictaphone de leur smartphone (p.ex. Honor 5s).

Malheureusement, tou(te)s n'ont pas spécifié le modèle du microphone employé (la question posée dans le sondage ne demandait pas précisément cette information). Les logiciels utilisés varient également ; certains utilisent l'interface CSL de KayPENTAX, d'autres les logiciels payants Vocalab ou DIANA, d'autres encore uniquement le gratuiciel Praat.

Rappelons cependant que ces données ne peuvent pas être généralisées à la pratique de la totalité des logopèdes en Belgique francophone. En effet, selon les chiffres fournis par Madame Coets, secrétaire générale de l'Union Professionnelle des Logopèdes Francophones (UPLF), le nombre de logopèdes affiliés indiquant traiter les dysphonies s'élève à 751 sur un nombre total de logopèdes affiliés de 4763 (en considérant les données depuis l'année 1989). Notre sondage ne représenterait donc que 3.3 % des vocologues belges francophones. Il nous donne néanmoins une indication de l'utilisation de l'AVQI, et s'ajoute aux ouï-dire circulant parmi les vocologues, concernant le manque de guidelines et la réticence à utiliser cet outil par manque d'assurance. En effet, face à l'insécurité, la plupart des vocologues choisissent, par défaut, l'utilisation du DSI, lui aussi repris dans la liste limitative des tests de l'INAMI.

Pour pouvoir valider l'AVQI, il nous a tout d'abord fallu déterminer l'échantillon sur lequel effectuer nos analyses. Nous avons également souhaité observer si une différence significative existait entre la version 02.03 de l'AVQI et la version 03.01, plus récente.

## L'échantillon en français : analyses préliminaires<sup>22</sup>

La première étape de notre étude était de vérifier quel matériel phonétique serait le plus adapté pour l'échantillon de parole continue à analyser par l'AVQI. Nos analyses préliminaires, décrites dans ce sous-chapitre, ont été effectuées entre septembre 2016 et juin 2017, dans le cadre du cours « Travail d'initiation à la recherche » (TIR).

Dans leur étude de 2014, Maryn et al. ont utilisé la phrase suivante : « La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort. » Elle est composée de 21 syllabes et tirée du texte phonétiquement équilibré « La bise et le soleil ». Deux questions sont apparues :

- choisir la première phrase d'un texte phonétiquement équilibré signifie-t-il que cette phrase est elle-même équilibrée ?
- cette phrase est-elle suffisamment représentative de la phonologie de la langue investiguée ?

---

<sup>22</sup> Les analyses statistiques ont été effectuées à l'aide du logiciel « IBM SPSS Statistics Version 22 » (Statistical Package for the Social Sciences, par la compagnie IBM).

## I. L'équilibre phonétique de l'extrait du texte « La bise et le soleil »

Nous avons donc remis en question le choix de la phrase extraite du texte « La bise et le soleil ». Généralement, l'équilibrage phonétique d'un texte se réalise sur sa totalité. Il nous semblait donc de prime abord préférable d'utiliser des phrases phonétiquement équilibrées<sup>23</sup>, dont l'équilibrage serait a priori plus adapté. Cependant, nous avons tout de même souhaité vérifier le bien-fondé de l'utilisation de l'extrait employé par Maryn et al. (2014), afin de confirmer ou d'infirmer notre intuition. Les auteurs ont calculé les fréquences d'occurrence des phonèmes sur les 21 syllabes de la phrase. Ces occurrences sont représentées dans le tableau 3.

« La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort. »

Tableau 3 Fréquences d'occurrence des phonèmes dans la première phrase du texte « La bise et le soleil »

phonème	p	a	ɛ	r	l	s	t	ɔ	ə	i	z	d	ɛ <sup>i</sup>	e
occurrence	2	3	2	2	7	4	2	1	2	2	1	1	1	2
	k	b	y	ʃ	f	œ	ɔ <sup>•</sup>	ɔ <sup>ʷ</sup>	i :					
	2	1	3	1	1	1	1	1	1					

Nous apportons une première correction : contrairement à leur décompte, nous obtenons six phonèmes [l] et non sept.

Nous observons que les voyelles orales [u], [ɑ], [o] et [ø], les voyelles nasales [ɛ̃] et [ɔ̃], les semi-voyelles [w] et [ɥ] ne sont pas présentes dans la phrase, de même que les consonnes [g], [m], [ŋ], [v], [ʒ], et [n]. La représentativité de la phrase sur le plan phonétique n'est donc pas vérifiée. En 1990, Tubach et Boe ont effectué des analyses sur de nombreuses transcriptions phonétiques de la langue française parlée (possédant une meilleure validité écologique par rapport aux transcriptions phonétiques de textes). Cette étude a notamment permis le calcul de la fréquence d'occurrence des phonèmes, des diphtongues, triphongues et des quadruphones dans la langue française parlée. Nous avons donc vérifié l'équilibre phonétique de la phrase utilisée par Maryn et al. (2014) en regard des données de Tubach et Boe. Le nombre total de phonèmes dans cette phrase est dès lors de 45 :

[la bɪzɛlɛsɔləjsədɪspɪtɛ.ʃakœasɪrɑ̃kɪlɛtɛləpɪlyfɔʀ]. Ensuite, nous avons calculé les pourcentages d'occurrence des phonèmes dans la phrase. Nous avons également comparé ces occurrences à celles mesurées dans la langue française par Tubach et Boe (1990). Le tableau 4 reprend ces occurrences, les phonèmes voisés (qui nous intéressent

<sup>23</sup> Combescura présentait par exemple, en 1981, 20 listes de dix phrases phonétiquement équilibrées (voir l'annexe 5, page 163) qui peuvent être utilisées pour des analyses acoustiques (Goujon, 2012).

particulièrement dans ce travail, car ce sont eux qui sont pris en compte dans le calcul de l'AVQI) sont représentés en gras :

Tableau 4 Fréquences d'occurrence des phonèmes dans l'extrait du texte « La bise et le soleil » et dans la langue française

phonème	<i>p</i>	<b>a</b>	<b>ε</b>	R	l	s	t	ɔ	ə	i	z	d
occurrence	2	<b>3</b>	<b>3</b>	2	<b>6</b>	4	2	2	<b>3</b>	<b>3</b>	<b>1</b>	<b>1</b>
% phrase	4.4 %	<b>6.7</b> %	<b>6.7</b> %	4.4 %	<b>13.3</b> %	8.8 %	4.4 %	4.4 %	<b>6.7</b> %	<b>6.7</b> %	<b>2.2</b> %	<b>2.2</b> %
% langue fr	3.74 %	<b>7.98</b> %	<b>3.81</b> %	<b>7.22</b> %	<b>5.74</b> %	6.1 %	5.22 %	<b>2.06</b> %	<b>5.59</b> %	<b>5.5</b> %	<b>1.85</b> %	<b>4.13</b> %
	<b>e</b>	<i>k</i>	<b>b</b>	<b>y</b>	<i>f</i>	<i>f</i>	<b>œ</b>	<b>j</b>	<b>ã</b>			
	<b>2</b>	2	<b>1</b>	<b>3</b>	<i>l</i>	<i>l</i>	<b>1</b>	<b>1</b>	<b>1</b>			
	<b>4.4</b> %	4.4 %	<b>2.2</b> %	<b>6.7</b> %	2.2 %	2.2 %	<b>2.2</b> %	<b>2.2</b> %	<b>2.2</b> %			
	<b>7.19</b> %	4.22 %	<b>1.05</b> %	<b>1.9</b> %	0.47 %	1.5 %	<b>0.54</b> %	<b>1.77</b> %	<b>3.55</b> %			

Note. % phrase = pourcentage relatif dans la phrase ; % langue fr = pourcentage relatif dans la langue française selon Tubach et Boe (1990) ; en gras, les phonèmes voisés, pris en compte pour le calcul de l'AVQI ; en italique, les phonèmes non voisés.

La plupart des occurrences dans la phrase ne semblent effectivement pas équivalentes aux scores moyens établis par les auteurs. Nous avons vérifié si les différences observées étaient statistiquement significatives.

Les résultats obtenus au test de normalité (tableau 5) montrent que les données de l'échantillon ne sont pas normalement distribuées.

Tableau 5 Tests de normalité pour les occurrences dans l'extrait du texte « La bise et le soleil » et dans la langue française

groupe	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistique s	ddl	Sig.	Statistique s	ddl	Sig.
fréqu échant	.209	15	.076	.789	15	<b>.003</b>
LF	.186	15	.170	.930	15	.277

a. Correction de signification de Lilliefors

Note. échant = extrait du texte « La bise et le soleil » ; LF = langue française.

Nous avons donc poursuivi l'analyse avec la corrélation de Spearman, dont la matrice est représentée dans le tableau 6 :

Tableau 6 Corrélation de Spearman entre les fréquences d'occurrence dans l'extrait du texte « La bise et le soleil » et dans la langue française

			échant	LF
Rho de Spearman	échant	Coefficient de corrélation	1.000	<b>.615*</b>
		Sig. (bilatéral)	.	<b>.015</b>
		N	15	15
	LF	Coefficient de corrélation	.615*	1.000
		Sig. (bilatéral)	.015	.
		N	15	15

\*. La corrélation est significative au niveau 0.05 (bilatéral).

Note. échant = extrait du texte « La bise et le soleil » ; LF = langue française.

La corrélation observée est significative avec  $r_s = .615$  et  $p = .015 (< .05)$ . Selon les critères de Cohen (1988)<sup>24</sup>, cette corrélation est forte ( $> .50$ ) ; selon les critères de Hinkle, Wiersma et Jurs (1994), elle est modérée ( $.51-.70$ )<sup>25</sup>.

Pour consolider cette observation, nous avons comparé les médianes des fréquences d'occurrence des deux groupes (phrase du texte « La bise et le soleil » et langue française).

Les résultats obtenus au test non paramétrique de Wilcoxon-Mann-Whitney (tableau 7) nous indiquent que nous n'avons pas assez de preuves (à  $\alpha = .05$ ) pour rejeter l'hypothèse d'égalité de la distribution des fréquences d'occurrence pour la phrase et pour la langue française. La distribution semble ainsi être statistiquement similaire dans les deux groupes.

Tableau 7 Test de Wilcoxon comparant les médianes des fréquences d'occurrence de l'extrait du texte « La bise et le soleil » et de langue française

Récapitulatif du test d'hypothèse				
	Hypothèse nulle	Test	Sig.	Décision
1	La distribution de fréquences est identique sur les catégories de groupe.	Test U de Mann-Whitney d'échantillons indépendants	.305 <sub>1</sub>	Retenir l'hypothèse nulle.

Les significations asymptotiques sont affichées. Le niveau d'importance est .05.

<sub>1</sub> La signification exacte est affichée pour ce test.

La phrase utilisée par Maryn et al. (2014) est donc équilibrée, lorsque nous considérons les fréquences d'occurrence des phonèmes qui y sont présents. Cependant, nous gardons à l'esprit l'absence des phonèmes précités.

## II. L'utilisation d'un extrait du texte de Harmegnies

Pour notre étude, nous avons choisi d'utiliser des enregistrements de la base de données du CHU de Liège. En effet, de nombreux bilans vocaux sont effectués annuellement dans le service ORL de cet hôpital. Depuis 2009, les données médicales des patients sont enregistrées dans une base de données informatisée, qui comprend déjà plus de 1100 dossiers de patients âgés de 4 à 93 ans, employés, étudiants, retraités ou sans-emploi, avec des diagnostics variés (p.ex. nodules, immobilités laryngées, dysphonie fonctionnelle, œdème, RGO, sulcus vocalis et presbyphonie) (Remacle, Petitfils, Finck, & Morsomme, 2017). De plus, le texte phonétiquement équilibré de Harmegnies (1988), utilisé pour l'évaluation vocale au CHU, est également largement employé par les vocologistes francophones en Belgique.

<sup>24</sup> Coefficient de corrélation de .10 : effet de petite taille, corrélation faible ; .30 : effet de taille moyenne, corrélation moyenne ; .50 : effet de grande taille, corrélation forte.

<sup>25</sup> Coefficient de corrélation de .00-.30 : corrélation très faible, voire nulle ; .31-.50 : corrélation faible ; .51-.70 : corrélation modérée ; .71-.90 : corrélation élevée ; .91-1.00 : corrélation très élevée.



Les enregistrements des patients dysphoniques vus en clinique journalière sont essentiellement constitués d'un [a] tenu et de la lecture à deux reprises du texte que voici :

« *Quand René périt, un chat esseulé grogna fort. À cet instant, Vic sortit contempler le jour naissant. À midi, nerveux, il pensa aux blés mondés de son maître et à tous ceux du pur Roi David. Et puis, les larmes aux yeux, il lut tout Kafka.* »

Nous nous sommes demandé si ce texte ou une portion de celui-ci pouvait être considéré comme équivalent à celui utilisé par Maryn et al. En effet, modifier le texte utilisé pour l'évaluation vocale au CHU et constituer une nouvelle base de données serait chronophage et peu rationnel. Notamment, les enregistrements effectués lors des bilans d'évolution à venir, pour les patients ayant déjà consulté, ne pourraient pas être utilisés pour une comparaison fiable des mesures. Nous avons donc décidé de conserver le texte de Harmegnies et d'identifier quelle partie de celui-ci serait à employer afin de garantir une bonne validité conceptuelle. En effet, afin de fournir des mesures valides, l'extrait de la parole continue doit être phonétiquement représentatif de la langue cible, et la durée totale (en secondes) des segments voisés doit s'équilibrer avec les trois secondes de voyelle tenue.

#### ***A. Première sélection : l'équilibre phonétique***

Pour rappel, dans leur étude de 2015, Barsties et Maryn ont observé une durée moyenne de parole continue nécessaire de 34 syllabes (pour le texte « Papa en Marloes »), qui équivaut à trois secondes de segments voisés concaténés. Nous avons donc souhaité identifier un extrait phonétiquement équilibré du texte de Harmegnies permettant également d'obtenir trois secondes de segments voisés. Pour cela, nous avons pris comme point de référence la durée de 34 syllabes mesurée par Barsties et Maryn. Le texte de Harmegnies comprend 62 syllabes. Pour nous rapprocher des 34 syllabes mesurées par Barsties et Maryn (2015), nous avons établi les deux phrases suivantes.

La première : « *Quand René périt, un chat esseulé grogna fort. À cet instant, Vic sortit contempler le jour naissant. À midi, nerveux, il pensa aux blés* » inclut 37 syllabes et 78 phonèmes ; la seconde : « *À midi, nerveux, il pensa aux blés mondés de son maître et à tous ceux du pur Roi David. Et puis, les larmes aux yeux, il lut tout Kafka* » inclut 35 syllabes et 74 phonèmes.

Pour sélectionner l'extrait à employer, nous avons d'abord mesuré les fréquences d'occurrence des phonèmes dans les deux extraits (tableaux 8 et 11). Afin de comparer ces fréquences d'occurrence avec celles des mêmes phonèmes dans la langue française (selon Tubach & Boe, 1990), nous avons ensuite calculé des corrélations.

Les analyses de l'AVQI se font, comme nous l'avons déjà expliqué, sur les segments voisés de l'enregistrement. Nous avons donc effectué nos calculs sur les phonèmes voisés, uniquement. Les consonnes [R] et [l] ont été considérées comme voisées, car précédant ou suivant un autre phonème voisé, elles le deviennent également. Par exemple, dans le mot « barbe », le [R] est voisé : [b a R b ə] ; dans le mot « poutre » cependant, le [R] est dévoisé (Révis, 2013) et se note alors [χ] : [putχ].

### Extrait 1

Tableau 8 Comparaison des fréquences d'occurrence des phonèmes dans le premier extrait du texte de Harmegnies et dans la langue française

[k̄ãRœneperi.œ̄faesœlegrɔnafɔR.asetēstā.viksɔrtikɔtāplelœzurnesā.amidi.nERVø.ilpāsaoble]															
phonème	<i>k</i>	<i>ã</i>	<i>R</i>	<i>œ</i>	<i>n</i>	<i>e</i>	<i>p</i>	<i>i</i>	<i>œ̄</i>	<i>f</i>	<i>a</i>	<i>s</i>	<i>l</i>	<i>g</i>	
occurrence	3	5	7	3	3	6	3	6	1	1	5	6	5	1	
% phrase	3.85 %	6.41 %	8.97 %	3.85 %	3.85 %	7.69 %	3.85 %	7.69 %	1.28 %	1.28 %	6.41 %	7.69 %	6.41 %	1.28 %	
% langue fr	4.22 %	3.55 %	7.22 %	5.59 %	2.79 %	7.19 %	3.74 %	5.5 %	0.54 %	0.47 %	7.98 %	6.10 %	5.74 %	0.55 %	
	<i>ɔ</i>	<i>ɲ</i>	<i>f</i>	<i>ɛ</i>	<i>t</i>	<i>ẽ</i>	<i>v</i>	<i>õ</i>	<i>ʒ</i>	<i>u</i>	<i>m</i>	<i>ø</i>	<i>d</i>	<i>o</i>	<i>b</i>
	3	1	1	3	4	1	2	1	1	1	1	1	1	1	1
	3.85 %	1.28 %	1.28 %	3.85 %	5.13 %	1.28 %	2.56 %	1.28 %	1.28 %	1.28 %	1.28 %	1.28 %	1.28 %	1.28 %	1.28 %
	2.06 %	0.08 %	1.50 %	3.81 %	5.22 %	1.26 %	2.27 %	1.95 %	1.59 %	2.02 %	3.30 %	0.45 %	4.13 %	1.07 %	1.05 %

Note : % phrase = pourcentage relatif dans l'extrait ; % langue fr = pourcentage relatif dans la langue française selon Tubach et Boe (1990) ; en gras, les phonèmes voisés, pris en compte pour le calcul de l'AVQI ; en italique, les phonèmes non voisés.

Nous observons que les phonèmes de la langue française manquant dans l'extrait n° 1 sont : les semi-voyelles [ɥ], [w] et [j], la voyelle [y] et la consonne [z].

Les données de l'extrait n° 1 n'étant pas normalement distribuées (tableau 9), nous avons calculé la corrélation de Spearman (tableau 10) entre les fréquences d'occurrence des phonèmes voisés dans l'extrait n° 1 et dans la langue française (selon Tubach & Boe, 1990). Cette corrélation s'avère significative et élevée selon les critères de Cohen (1988) ainsi que selon ceux de Hinkle, Wiersma et Jurs (1994).

Tableau 9 Tests de normalité pour les occurrences dans le premier extrait du texte de Harmegnies et dans la langue française

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
echant1son	.307	23	.000	.774	23	.000
francais1son	.160	23	.131	.914	23	.051

a. Correction de signification de Lilliefors

Note : echant1son = phonèmes voisés du premier extrait du texte de Harmegnies ; francais1son = ces mêmes phonèmes dans la langue française.

Tableau 10 *Corrélation de Spearman entre les fréquences d'occurrence des phonèmes voisés dans le premier extrait du texte de Harmegnies et dans la langue française*

		echant1son	francais1son
Rho de Spearman	echant1son	1.000	.818**
	Coefficient de corrélation		
	Sig. (bilatéral)	.	.000
	N	23	23
		francais1son	1.000
	Coefficient de corrélation	.818**	
	Sig. (bilatéral)	.000	.
	N	23	23

\*\* . La corrélation est significative au niveau 0.01 (bilatéral).

Note : echant1son = phonèmes voisés du premier extrait du texte de Harmegnies ; francais1son = ces mêmes phonèmes dans langue française (selon Tubach & Boe, 1990).

## Extrait 2

Tableau 11 *Comparaison des fréquences d'occurrence des phonèmes dans le second extrait du texte de Harmegnies et dans la langue française*

[amidi.nERVø.ilpãsaoble mōdedœsōmεtreatusødypyr:wadavid.ep̄i.lelar mozjœ.il:y tuka fka]														
phonème	a	m	i	d	n	ε	ʀ	v	ø	l	p	ã	s	
occurrence	8	4	6	6	1	2	4	2	3	5	3	1	3	
% phrase	10.81 %	5.41 %	8.11 %	8.11 %	1.35 %	2.70 %	5.41 %	2.70 %	4.05 %	6.76 %	4.05 %	1.35 %	4.05 %	
% langue fr	7.98 %	3.30 %	5.5 %	4.13 %	2.79 %	3.81 %	7.22 %	2.27 %	0.45 %	5.74 %	3.74 %	3.55 %	6.10 %	
	o	b	e	õ	œ	t	u	y	w	ɥ	z	j	k	f
	2	1	5	2	1	3	2	3	1	1	1	1	2	1
	2.70 %	1.35 %	6.76 %	2.70 %	1.35 %	4.05 %	2.70 %	4.05 %	1.35 %	1.35 %	1.35 %	1.35 %	2.70 %	1.35 %
	1.07 %	1.05 %	7.19 %	1.95 %	5.59 %	5.22 %	2.02 %	1.9 %	1.08 %	0.46 %	1.85 %	1.77 %	4.22 %	1.5 %

Note : % phrase = pourcentage relatif dans l'extrait ; % langue fr = pourcentage relatif dans la langue française selon Tubach et Boe (1990) ; en gras, les phonèmes voisés, pris en compte pour le calcul de l'AVQI ; en italique, les phonèmes non voisés.

Dans l'extrait n° 2, les phonèmes de la langue française manquants sont : les voyelles [ɔ], [ē] et [œ] et les consonnes [ɲ], [ʃ], [ʒ] et [g].

Le test de normalité (tableau 12) indique une distribution non gaussienne des données.

Le calcul de la corrélation de Spearman (tableau 13) nous indique une corrélation significative entre les fréquences d'occurrence des phonèmes voisés dans l'extrait n° 2 et celles dans la langue française (selon Tubach & Boe, 1990). La corrélation est forte (> .50) selon les critères de Cohen (1988), et modérée (.51-.70) selon les critères de Hinkle, Wiersma et Jurs (1994).

Tableau 12 *Tests de normalité pour les occurrences dans le second extrait du texte de Harmegnies et dans la langue française*

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
echant2son	.245	22	.001	.837	22	.002
francais2son	.173	22	.086	.908	22	.043

a. Correction de signification de Lilliefors

Note : echant2son = phonèmes voisés du second extrait du texte de Harmegnies ; francais2son = ces mêmes phonèmes dans langue française (selon Tubach & Boe, 1990).

Tableau 13 *Corrélation de Spearman entre les fréquences d'occurrence des phonèmes voisés dans le second extrait du texte de Harmegnies et dans la langue française*

		echant2son	francais2son
Rho de Spearman	echant2son	1.000	.588**
	Coefficient de corrélation		
	Sig. (bilatéral)	.	.004
	N	22	22
francais2son	francais2son	.588**	1.000
	Coefficient de corrélation		
	Sig. (bilatéral)	.004	.
	N	22	22

\*\* La corrélation est significative au niveau 0.01 (bilatéral).

Note : echant2son = phonèmes voisés du second extrait du texte de Harmegnies ; francais2son = ces mêmes phonèmes dans langue française (selon Tubach & Boe, 1990).

Nous avons ensuite vérifié si les coefficients de corrélation pour les deux extraits diffèrent significativement. Pour ce faire, nous avons transformé les coefficients  $r$  en scores  $Z$  (qui sont alors normalement distribués) à l'aide de la transformation de Fisher, dont la formule est la suivante :

$$z' = 0.5[\ln(1+r) - \ln(1-r)] \text{ (Andale, 2016)}$$

Soit, pour le coefficient du premier extrait :

$$z' = 0.5 \times [\ln(1 + 0.818) - \ln(1 - 0.818)]$$

$$\Leftrightarrow z' = \mathbf{1.151}$$

Pour le coefficient du second extrait :

$$z' = 0.5 \times [\ln(1 + 0.588) - \ln(1 - 0.588)]$$

$$\Leftrightarrow z' = \mathbf{0.675}$$

Les scores  $Z$  de Fischer ont été comparés à l'aide de la formule suivante (Kenny, 1987) :

$$Z = \frac{z1 - z2}{\sqrt{\frac{1}{n1-3} + \frac{1}{n2-3}}} = 1.486$$

Pour  $\alpha = .05$ , si  $Z > 1.96$  ou  $Z < -1.96$ , nous pouvons rejeter l'hypothèse nulle selon laquelle les deux coefficients de corrélation sont statistiquement équivalents (Warner, 2013).

$Z$  étant dans notre cas inférieur à 1.96, nous n'avons pas assez de preuves pour rejeter l'hypothèse d'égalité des coefficients de corrélation dans les deux extraits, au seuil de significativité  $\alpha = .05$ . Autrement dit, nous devons tolérer l'hypothèse selon laquelle les deux extraits se valent : les corrélations sont significatives dans les deux cas, et la différence entre elles non significative.

Le premier extrait nous intéresse davantage, car la corrélation entre les fréquences d'occurrence des phonèmes voisés dans celui-ci et dans la langue française parlée est quelque peu plus forte ( $r_s = .818$ ,  $p = .000$  pour l'extrait n° 1 contre  $r_s = .588$ ,  $p = .004$  pour l'extrait n° 2). De plus, seule une voyelle est manquante, contre trois dans le second extrait ; sa représentativité est donc meilleure. Aussi, la représentativité semble meilleure en comparaison avec la phrase utilisée par Maryn en 2014, dont la corrélation avec les fréquences d'occurrence des phonèmes dans la langue française selon Tubach

et Boe avait montré un  $r_s = .615$ ,  $p = .015$ . Il y manquait également de nombreux phonèmes (quatre voyelles orales, deux voyelles nasales, deux semi-voyelles et six consonnes).

Nous avons donc opté pour les 37 premières syllabes du texte de Harmegnies (extrait n° 1).

### B. Seconde sélection : trois secondes de voisement

Le calcul de l'AVQI intègre une portion d'un [a] tenu d'une durée de trois secondes et une phrase phonétiquement équilibrée qui doit également respecter une durée de trois secondes de segments voisés. Des analyses préliminaires effectuées sur de multiples lectures de l'extrait 1 par deux sujets sains ont révélé que la durée syllabique nous permettant d'obtenir un échantillon de trois secondes de segments voisés — avec une déviation n'excédant pas 0.1 seconde (Barsties & Maryn, 2015) — est de **32 syllabes** : « Quand René périt, un chat esseulé grogna fort. À cet instant, Vic sortit contempler le jour naissant. À midi, nerveux ».

Les fréquences d'occurrence des phonèmes voisés de cet extrait (représentées dans le tableau 14) corrélaient également avec les fréquences dans la langue française selon Tubach et Boe (1990) : le coefficient de corrélation de Spearman de  $r_s = .802$  (tableau 15) indique une corrélation significative et forte, et ce coefficient n'est pas significativement différent de celui calculé sur les 37 premières syllabes ( $r_s = .818$ ), avec un Z de Fisher de 0.14 (compris dans l'intervalle [-1.96 ; 1.96]).

Tableau 14 Comparaison des fréquences d'occurrence des phonèmes dans les 32 premières syllabes du texte de Harmegnies et dans la langue française

[kãrœnepəri.œʃaesœlegrɔnafɔrasɛtɛstã.viksɔrtikɔtãplelœʒurnesã.amidi.nɛrvø]														
phonème	<i>k</i>	<i>ã</i>	<i>r</i>	<i>œ</i>	<i>n</i>	<i>e</i>	<i>p</i>	<i>i</i>	<i>œ̃</i>	<i>ʃ</i>	<i>a</i>	<i>s</i>	<i>l</i>	<i>g</i>
occurrence	3	4	7	3	3	5	2	5	1	1	4	5	3	1
% phrase	4.41 %	<b>5.88</b> %	<b>10.29</b> %	4.41 %	4.41 %	7.35 %	2.94 %	7.35 %	1.47 %	1.47 %	<b>5.88</b> %	7.35 %	4.41 %	1.47 %
% langue fr	4.22 %	<b>3.55</b> %	7.22 %	<b>5.59</b> %	2.79 %	7.19 %	3.74 %	5.5 %	<b>0.54</b> %	0.47 %	<b>7.98</b> %	6.10 %	5.74 %	<b>0.55</b> %
	<i>ɔ</i>	<i>ɲ</i>	<i>f</i>	<i>ɛ</i>	<i>t</i>	<i>ɛ̃</i>	<i>v</i>	<i>õ</i>	<i>ʒ</i>	<i>u</i>	<i>m</i>	<i>ø</i>	<i>d</i>	
	3	1	1	3	4	1	2	1	1	1	1	1	1	
	4.41 %	1.47 %	1.47 %	4.41 %	5.88 %	1.47 %	2.94 %	1.47 %	1.47 %	1.47 %	1.47 %	1.47 %	1.47 %	
	<b>2.06</b> %	<b>0.08</b> %	1.50 %	<b>3.81</b> %	5.22 %	1.26 %	2.27 %	1.95 %	1.59 %	2.02 %	3.30 %	0.45 %	4.13 %	

Note : % phrase = pourcentage relatif dans l'extrait ; % langue fr = pourcentage relatif dans la langue française selon Tubach et Boe (1990) ; en gras, les phonèmes voisés, pris en compte pour le calcul de l'AVQI ; en italique, les phonèmes non voisés.

Nous observons que les phonèmes de la langue française manquant dans cet extrait sont : les semi-voyelles [ɥ], [w] et [j], les voyelles [y] et [o] et les consonnes [z] et [b].

Tableau 15 *Corrélation de Spearman entre les fréquences d'occurrence des phonèmes voisés dans les 32 premières syllabes du texte de Harmegnies et dans la langue française*

			echant32son	echantFr32son
Rho de Spearman	echant32son	Coefficient de corrélation	1.000	.802**
		Sig. (bilatéral) N	.	.000 21
	echantFr32son	Coefficient de corrélation	.802**	1.000
		Sig. (bilatéral) N	.000 21	.

\*\* . La corrélation est significative au niveau 0.01 (bilatéral).

Note : echant32son = phonèmes voisés de l'extrait de 32 syllabes du texte de Harmegnies ; echantFr32son = ces mêmes phonèmes dans langue française (selon Tubach & Boe, 1990).

Nous nous sommes tout de même interrogés quant à l'utilité d'inclure les cinq dernières syllabes (« A midi, nerveux ») dans l'analyse. En effet, il s'agit là d'un début de phrase inachevée, dénuée de sens et grammaticalement incorrecte à l'état isolé, qui pourrait notamment influencer le jugement perceptif par le jury de vocologistes. L'extrait de 27 syllabes, obtenu après suppression de ces cinq syllabes, montre également une corrélation forte avec les fréquences d'occurrence des phonèmes voisés dans la langue française selon Tubach et Boe (tableaux 16 et 17). En transformant les coefficients de corrélation pour les extraits de 32 et de 27 syllabes en scores  $z'$  de Fisher et en les comparant, nous obtenons une statistique  $Z$  de -0.43, comprise dans l'intervalle [-1.96 ; 1.96]. Nous ne pouvons donc pas rejeter l'hypothèse d'égalité des deux coefficients de corrélation à  $\alpha = .05$ . Sur le plan de l'équilibre phonétique des phonèmes présents dans les phrases, ces deux extraits se valent. Notons cependant l'absence plus importante de phonèmes voisés dans l'extrait de 27 syllabes : les semi-voyelles [ɥ], [w] et [j], les voyelles [y], [ø] et [o] ainsi que les consonnes [z], [m], [d] et [b] sont manquantes.

Ainsi, nous choisissons d'effectuer les analyses à venir en deux temps : une première fois sur l'échantillon de parole continue de **32 syllabes**, ainsi qu'une seconde fois sur l'échantillon de parole continue de **27 syllabes** (ne comprenant donc que les deux premières phrases entières).

Tableau 16 *Comparaison des occurrences des phonèmes dans les 27 premières syllabes du texte et dans la langue française*

[k̄ãrœnep̄erĩ.œ̄jaesœlegrɔ̄nãfɔ̄raset̄estã.viksɔ̄rtikȭtãplelœ̄ʒurnesã]														
phonème	<i>k</i>	<i>ã</i>	<i>r</i>	<i>œ</i>	<i>n</i>	<i>e</i>	<i>p</i>	<i>i</i>	<i>œ̄</i>	<i>ʃ</i>	<i>a</i>	<i>s</i>	<i>l</i>	<i>g</i>
occurrence	3	4	6	3	2	5	2	3	1	1	3	5	3	1
% phrase	5.17 %	<b>6.90</b> %	<b>10.34</b> %	5.17 %	3.45 %	<b>8.62</b> %	3.45 %	5.17 %	1.72 %	1.72 %	5.17 %	8.62 %	5.17 %	1.72 %
% langue fr	4.22 %	<b>3.55</b> %	7.22 %	<b>5.59</b> %	2.79 %	<b>7.19</b> %	3.74 %	5.5 %	<b>0.54</b> %	0.47 %	<b>7.98</b> %	6.10 %	5.74 %	<b>0.55</b> %
	<i>ɔ</i>	<i>ɲ</i>	<i>f</i>	<i>ɛ</i>	<i>t</i>	<i>ẽ</i>	<i>v</i>	<i>õ</i>	<i>ʒ</i>	<i>u</i>				
	3	1	1	2	4	1	1	1	1	1				
	5.17 %	1.72 %	1.72 %	3.45 %	6.90 %	1.72 %	1.72 %	1.72 %	1.72 %	1.72 %				
	<b>2.06</b> %	<b>0.08</b> %	1.50 %	<b>3.81</b> %	5.22 %	1.26 %	2.27 %	1.95 %	1.59 %	2.02 %				

Note : % phrase = pourcentage relatif dans l'extrait ; % langue fr = pourcentage relatif dans la langue française selon Tubach et Boe (1990) ; en gras, les phonèmes voisés, pris en compte pour le calcul de l'AVQI ; en italique, les phonèmes non voisés.

Tableau 17 *Corrélation de Spearman entre les fréquences d'occurrence des phonèmes voisés dans les 27 premières syllabes du texte de Harmegnies et dans la langue française*

		echant27son	echantFr27son
Rho de Spearman	echant27son	Coefficient de corrélation	1.000
		Sig. (bilatéral)	.
		N	18
	echantFr27son	Coefficient de corrélation	.850**
		Sig. (bilatéral)	.000
		N	18

\*\* La corrélation est significative au niveau 0.01 (bilatéral).

Note : echant27son = phonèmes voisés de l'extrait de 27 syllabes ; echantFr27son = ces mêmes phonèmes dans langue française.

## L'échantillon en français : application sur une population pathologique<sup>26</sup>

Dès septembre 2017, nous avons souhaité vérifier l'adéquation des extraits choisis (32 et 27 syllabes) sur un échantillon aléatoire de productions de patients dysphoniques provenant de la banque de données du CHU de Liège.

### I. La population et les enregistrements

Pour pouvoir effectuer cette vérification, nous avons sélectionné 268 enregistrements de 167 patients<sup>27</sup>. Ceux-ci ont été effectués dans le laboratoire vocal du service ORL, en utilisant un microphone serre-tête cardioïde à condensateur AKG C520 maintenu à 3-6cm de la bouche, via l'interface « CSL Main Program » d'un poste de travail informatique « Computerized Speech Lab » de KayPENTAX, modèle 4500, en enregistrement « mono », avec un taux d'échantillonnage de 11025 Hz (pour rappel, Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2009 préconisent un taux d'échantillonnage de 44100 Hz) et une résolution de 16 bits (recommandé par Chial, 2003). Notons que le niveau de pression sonore d'entrée n'est pas maintenu constant pour chaque sujet. En effet, il est diminué lorsque la voix paraît trop puissante (pour éviter les phénomènes de saturation), et augmenté en cas de voix de très faible niveau de pression sonore. Nous sommes conscients de cette faiblesse qui a un impact sur la précision des mesures acoustiques. Idéalement, à des buts de recherche scientifique, un même niveau d'entrée est systématiquement utilisé. En clinique, il n'est pas toujours possible de respecter cet idéal, sachant par exemple que les chanteurs d'opéra font également partie des patients qui consultent le centre. Leur positionnement résonantiel, même en voix parlée, a un impact sur leur niveau de pression sonore qui s'avère toujours plus élevé que la moyenne des locuteurs.

<sup>26</sup> Les analyses statistiques ont été effectuées à l'aide du logiciel « SAS 9.4 » (Statistical Analysis Software, par le SAS Institute Inc.).

<sup>27</sup> Après suppression préalable des enregistrements pour lesquels (a) il nous manquait l'une des deux conditions de phonation (parole continue ou phonation soutenue) (b) les deux conditions n'avaient pas été recueillies le même jour (c) le [a] tenu ne durait pas plus de trois secondes (d) une partie de la lecture à voix haute était coupée.

Toutefois, il est important d'utiliser le même niveau d'entrée pour un même sujet, sachant que plusieurs enregistrements sont effectués à différents moments du suivi du patient (p.ex. pour un bilan d'évolution), et ce de manière à pouvoir comparer les données d'un instant à l'autre de manière fiable, en évitant les variations liées aux paramètres d'enregistrement.

Les enregistrements ont été effectués dans un local calme (non insonorisé) ; le microphone serre-tête unidirectionnel permet de limiter l'impact du bruit environnant. En effet, le bruit ambiant augmente de manière importante les mesures de perturbation telles que le jitter et le shimmer (Ingrisano, Perry, & Jepson, 1998 ; Perry, Ingrisano, Palmer, & McDonald, 2000). Un rapport signal-bruit supérieur à 42 dB est recommandé afin de pouvoir mener à bien des analyses acoustiques fiables (avec une précision de 99 %) (Deliyski, Shaw, & Evans, 2005). Dans certaines études, les auteurs mesurent ainsi en post-hoc le rapport signal-bruit sur chaque enregistrement, en visant des valeurs situées entre 30 et 42 dB (Barsties & De Bodt, 2015). Le rapport signal-bruit (signal-to-noise ratio, SNR) est considéré comme identique au rapport harmoniques-bruit (harmonics-to-noise ratio, HNR) dans le cas d'un signal supposé périodique (Boersma & Weenink, 2003). Dans notre cas, comme notre échantillon comprend des enregistrements de voix fortement perturbées, nous ne pouvons pas supposer que les signaux soient périodiques. La mesure du rapport harmoniques-bruit ne nous permettrait donc pas de tirer des conclusions quant au rapport signal-bruit (et donc au bruit ambiant). Nous avons pour cela mesuré le bruit ambiant dans le local à l'aide d'un décibel-mètre PCE-353 (PCE Instruments, Strasbourg, France), à quatre reprises, et avons obtenu une moyenne de 50.33 dB. Considérant que le rapport signal-bruit recommandé doit être supérieur à 42 dB (Deliyski, Shaw, & Evans, 2005), le niveau de bruit ambiant dans le laboratoire vocal semble insatisfaisant. En effet, si le bruit ambiant est de 50 dB, le niveau de pression sonore vocal devrait être supérieur à 92 dB pour permettre un rapport signal-bruit satisfaisant. Or, ce niveau de pression sonore est bien supérieur à la voix conversationnelle, et même à la voix projetée. Nous avons également effectué des mesures thermo-hygrométriques à l'aide d'un hygromètre Hygromaster BLD7700 (Protimeter, Pennsylvanie, États-Unis), qui ont indiqué un pourcentage d'humidité relative de 26.6 % (avec une température moyenne de 22.65 °C). Les pourcentages situés entre 0 et 30 % représentent un air sec. L'air sec peut avoir un effet sur la biomécanique vocale (augmentation de la viscosité et de la raideur des plis vocaux) (Hemler, Wieneke, Lebacqz, & Dejonckere, 2001), et ainsi modifier les paramètres acoustiques tels le jitter et le shimmer (Hemler, Wieneke, & Dejonckere, 1997). Cependant, le temps d'un bilan vocal n'est probablement pas assez long que pour voir apparaître des altérations significatives des paramètres acoustiques.



Nous avons effectué un second tri des 268 enregistrements, en supprimant :

- les enregistrements de mauvaise qualité : bruits de fond, saturation du microphone, bruits de souffle dans le microphone, écho ;
- les enregistrements des participants dont la langue maternelle ne semblait pas être le français. Notons que cette information n’était pas toujours disponible dans les dossiers des patients. Nous avons donc dû effectuer un jugement subjectif, basé principalement sur la présence d’un accent ;
- les enregistrements qui comprenaient des erreurs de lecture concernant les voyelles (p.ex. « enseulé », au lieu de « esseulé ») et les consonnes sonores. Nous avons toléré les erreurs de lecture concernant les consonnes sourdes — non prises en compte dans le calcul de l’AVQI — à l’exception des sonorisations de ces dernières (p.ex. « ézeulé » au lieu de « esseulé ») ;
- les enregistrements de participants présentant un accent régional fort prononcé, amenant des modifications phonémiques importantes (ici encore, le jugement est clairement subjectif) ;
- les enregistrements comprenant de nombreux raclements de gorge ou autres bruits parasites ;
- les enregistrements d’enfants et d’adolescents.

De plus, nous avons souhaité ne garder qu’un seul enregistrement par patient (la plupart d’entre eux ayant été enregistrés à plusieurs reprises, notamment pour des bilans d’évolution), afin d’éviter les erreurs statistiques liées aux échantillons non indépendants.

Finalement, notre cohorte inclut 162 enregistrements provenant de la production de 162 patients, dont les caractéristiques sont reprises dans le tableau 18 :

Tableau 18 *Caractéristiques principales de l’échantillon de 162 patients du CHU de Liège*

Nombre total	162		
Genre	Femme : 92 (56.79 %) Homme : 70 (43.21 %)		
Age <sup>a</sup>	Médiane <sup>b</sup> = 49.99 Intervalle interquartile = 21.62 Min = 19.74 Max = 88.34		
Pathologies (n et %) <sup>c</sup>			
	– Immobilité laryngée	35	21.60 %
	– Dysphonie fonctionnelle <sup>d</sup>	16	9.88 %
	– pathologie non renseignée	14	8.64 %
	– Post-phonochirurgie	13	8.02 %
	– Œdème cordal	12	7.41 %
	– Consultation sans pathologie <sup>e</sup>	12	7.41 %
	– Lésion nodulaire	10	6.17 %
	– Presbyphonie	9	5.56 %
	– Polype	6	3.70 %
	– Sulcus/vergeture	6	3.70 %
	– Lésion tumorale	5	3.09 %
	– Kyste	4	2.47 %
	– Cicatrice cordale/fibrose	4	2.47 %
	– Dysphonie spasmodique	3	1.85 %
	– Cordectomie	3	1.85 %
	– Affection neurologique	2	1.23 %

	– Laryngite chronique	2	1.23 %
	– Granulome	1	0.62 %
	– Atrophie cordale	1	0.62 %
	– Aryténoïdectomie	1	0.62 %
	– Reflux gastro-œsophagien	1	0.62 %
	– Dysphonie sur insuffisance respiratoire	1	0.62 %
	– Mue faussée	1	0.62 %
Professions (n et %) <sup>f</sup>	– Pensionnés/retraités	31	19.14 %
	– Profession non renseignée	27	16.67 %
	– Professionnels de la formation initiale et de la formation continue (enseignants)	22	13.58 %
	– Professionnels de la santé (professions médicales et paramédicales)	11	6.79 %
	– Personnel des services administratifs et commerciaux (employé de bureau, comptable, secrétaire médical, agent administratif...)	11	6.79 %
	– Personnel de la distribution et de la vente (commerçant, délégué commercial...)	9	5.56 %
	– Cadres administratifs et professionnels de l'information et de la communication (PDG, juriste, avocat, assistant parlementaire...)	8	4.94 %
	– Personnel, techniciens et cadres du transport et de la logistique (chauffeur routier, chauffeur de taxi, magasinier, agent distributeur...)	8	4.94 %
	– Personnel des services aux personnes et à la collectivité (technicienne de surface, nettoyage industriel, militaire, coiffeur...)	6	3.70 %
	– Chômage, invalidité	6	3.70 %
	– Étudiants	4	2.47 %
	– Professionnels des arts et du spectacle (chanteur, photographe, harpiste...)	4	2.47 %
	– Personnel de type artisanal (ébéniste, boulanger...)	4	2.47 %
	– Personnel du bâtiment, des travaux publics et de l'extraction (ouvrier en bâtiment)	2	1.23 %
	– Agents de maîtrise, techniciens et ingénieurs du bâtiment, des travaux publics et de l'extraction (architecte, métreur-deviseur)	2	1.23 %
	– Personnel des industries graphiques (copiste)	1	0.62 %
	– Cadres commerciaux (courtier en assurance)	1	0.62 %
	– Personnel des industries de process (ouvrier de production)	1	0.62 %
	– Techniciens industriels (technicien en mécanique)	1	0.62 %
	– Personnel de l'industrie hôtelière (collaboratrice dans un fast food)	1	0.62 %
– Professionnels de l'intervention sociale, du développement local et de l'emploi (assistant social)	1	0.62 %	
– Personnel de la production agricole (jardinier)	1	0.62 %	

Note. <sup>a</sup> Les âges de cinq participants n'ont pas pu être intégrés dans le calcul de la médiane, les données disponibles dans leurs dossiers de la base de données du CHU étant incomplètes. Ce calcul est donc effectué sur 157 participants. <sup>b</sup> Le test de Shapiro-Wilk montre une statistique W de .981, avec  $p = .032$ . Nous devons donc rejeter l'hypothèse de normalité de la distribution des âges dans notre échantillon. Ainsi, nous préférons la médiane et l'intervalle interquartile à la moyenne et à l'écart-type. <sup>c</sup> Les diagnostics laryngologiques ont été effectués à l'aide d'un vidéo-laryngo-stroboscope EndoSTROB (Xion GmbH, Berlin). Nous tenons à souligner que dans notre base de données, les dysphonies liées au cancer n'incluent que les patients de stade T1N0M0. <sup>d</sup> Les dysphonies dites « psychogènes » sont incluses dans cette catégorie. <sup>e</sup> Sont ici compris entre autres les bilans préopératoires dans le cadre d'interventions chirurgicales touchant la glande thyroïde, d'interventions chirurgicales ou toute autre intervention au cours de laquelle peut avoir lieu une lésion des nerfs laryngés. <sup>f</sup> Nous avons choisi de regrouper les professions selon l'arborescence du service public wallon de l'emploi et de la formation (Forem), accessible via le lien suivant : <https://www.leforem.be/former/horizonemploi/metier-profession-index.html> (Service public wallon de l'emploi et de la formation, 2016)

## II. Nos questions initiales

Comme nous l'avons décrit, nous avons effectué les 162 analyses en deux temps : une première fois sur l'échantillon de parole continue de 32 syllabes, ainsi qu'une seconde fois sur l'échantillon de parole continue de 27 syllabes (ne comprenant donc que les deux premières phrases entières).

De plus, les analyses ont été effectuées à l'aide des deux versions les plus récentes de l'AVQI : l'AVQI 02.03 et l'AVQI 03.01, dont les scripts sont disponibles aux annexes 6 (page 165) et 7 (page 171), respectivement.

Ainsi, nous avons obtenu les croisements suivants (figure 15), avec quatre conditions finales :

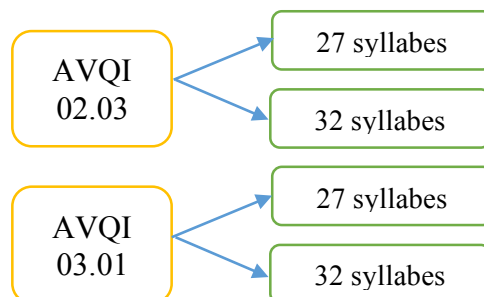


Figure 15. Croisements de données disponibles pour les analyses statistiques.

Nous pouvons donc effectuer de nombreuses analyses statistiques afin de répondre aux questions suivantes :

- *Effet principal « version » de l'AVQI* : y a-t-il une différence statistiquement significative entre les scores des deux versions de l'AVQI en français, toutes durées syllabiques confondues ?
- *Effet principal « durée syllabique »* : y a-t-il une différence statistiquement significative entre les scores selon que l'analyse est effectuée sur 27 ou sur 32 syllabes, les deux versions de l'AVQI confondues ?
- *Effets simples* : y a-t-il des différences entre les scores des deux versions de l'AVQI lorsque l'analyse est effectuée sur 27 syllabes ? ... lorsque l'analyse est effectuée sur 32 syllabes ? Y a-t-il des différences entre les scores selon que l'analyse est effectuée sur 27 ou sur 32 syllabes, lorsque l'on utilise la version 02.03 ? ... lorsque l'on utilise la version 03.01 ?
- Pouvons-nous utiliser un extrait d'uniquement 27 syllabes de la parole continue afin d'éviter un début de phrase inachevée ? C'est-à-dire, en termes statistiques :
  - Quels sont la durée moyenne et l'écart-type des segments voisés concaténés, pour 27 et pour 32 syllabes<sup>28</sup> ? L'une des deux durées s'équilibre-t-elle davantage avec les trois secondes de la voyelle tenue ?
  - *Effet simple « durée syllabique en version 03.01 »* : lorsque l'on utilise la version 03.01 de l'AVQI, y a-t-il une différence significative entre les scores selon que l'on effectue l'analyse sur 27 ou sur 32 syllabes ?

<sup>28</sup> Notons que les durées extraites sont exactement similaires dans les deux versions de l'AVQI.

### III. La durée voisée en secondes de 27 vs 32 syllabes de parole continue

Tout d'abord, nous avons voulu savoir s'il existait une différence significative entre les durées des segments voisés extraits de 27 et de 32 syllabes de parole continue. Via Praat, nous avons mesuré ces durées pour chaque enregistrement. L'AVQI, après avoir effectué l'analyse, fournit un fichier audio comprenant la concaténation des segments voisés de la parole continue et du [a] tenu. Nous avons ainsi pu aisément localiser la limite entre les deux parties et mesurer la durée de chacune d'elles, comme illustré à la figure 16.

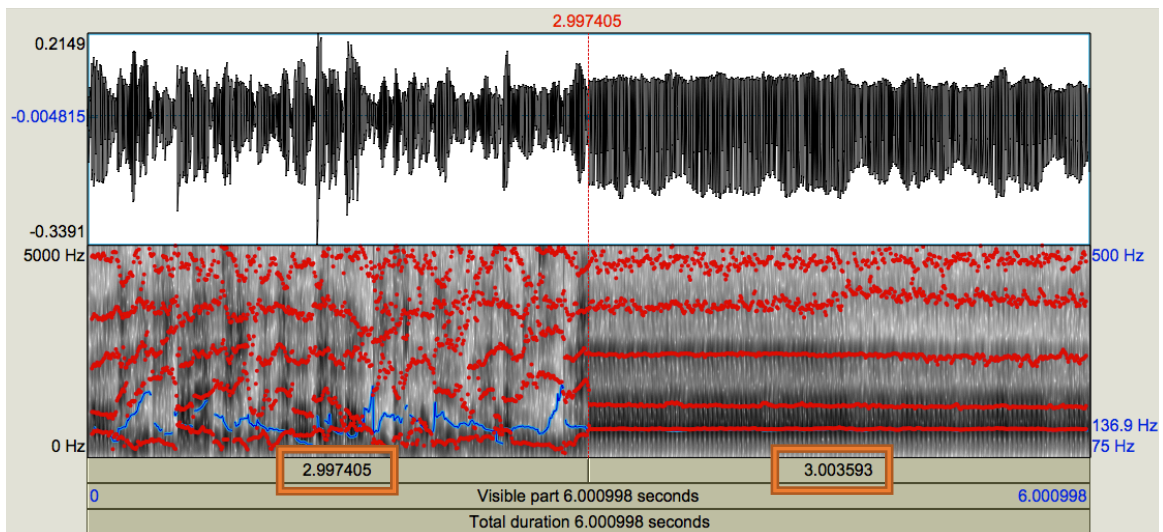


Figure 16. Visualisation de la concaténation des segments voisés extraits de la parole continue et des trois secondes de la voyelle [a] tenue par un sujet sain.

Moitié gauche = segments voisés de la parole continue (32 syllabes) ; moitié droite = voyelle tenue ; les deux parties sont parfaitement équilibrées au niveau de leur durée.

Le tableau 19 illustre les statistiques descriptives concernant les durées des segments voisés extraits des échantillons de 27 et de 32 syllabes. Les données n'étant pas normalement distribuées, nous ne présentons pas les moyennes et écarts-types, moins fiables dans le cas d'une distribution non gaussienne<sup>29</sup>. Nous optons pour la médiane et l'écart interquartile, qui sont dans ce cas plus adaptés.

Tableau 19 *Statistiques descriptives des durées des échantillons de 27 et de 32 syllabes*

Variable	N	Médiane	Intervalle interquartile	Mode	Minimum	Maximum
Syll32	162	3.63	0.8125	3.57	1.14	6.90
Syll27	162	3	0.755	3.75	1.01	5.69

Note. Syll32 = échantillons de 32 syllabes ; Syll27 = échantillons de 27 syllabes.

<sup>29</sup> À titre informatif, les moyennes et écarts-types, malgré une distribution non gaussienne, sont quasiment similaires aux médianes et aux intervalles interquartiles. Pour l'échantillon de 32 syllabes, la moyenne est de 3.66 ( $\sigma = 0.84$ ), pour l'échantillon de 27 syllabes, la moyenne est de 2.99 ( $\sigma = 0.73$ ).

Nous constatons déjà que la médiane de la durée de 27 syllabes se rapproche davantage des trois secondes recherchées, et que l'indice de dispersion est quelque peu moins élevé. La médiane de la durée de 32 syllabes diffère de plus que la marge de tolérance de 0.1 seconde (Barsties & Maryn, 2015).

Nous avons alors souhaité vérifier si la différence entre les deux médianes était statistiquement significative. Le test de normalité Shapiro-Wilk (tableau 20) — conseillé pour les effectifs modestes — nous montre que les deux échantillons ne sont pas normalement distribués, puisque la probabilité de dépassement est inférieure à .05 :

Tableau 20 Tests de normalité pour les échantillons de 27 (gauche) et 32 syllabes (droite)

Tests for Normality				Tests for Normality					
Test	Statistic		p Value	Test	Statistic		p Value		
Shapiro-Wilk	W	0.981033	Pr < W	0.0255	Shapiro-Wilk	W	0.977642	Pr < W	0.0100
Kolmogorov-Smirnov	D	0.065974	Pr > D	0.0839	Kolmogorov-Smirnov	D	0.074192	Pr > D	0.0277
Cramer-von Mises	W-Sq	0.139141	Pr > W-Sq	0.0346	Cramer-von Mises	W-Sq	0.180454	Pr > W-Sq	0.0093
Anderson-Darling	A-Sq	0.798958	Pr > A-Sq	0.0395	Anderson-Darling	A-Sq	1.015449	Pr > A-Sq	0.0114

Nous avons donc employé le test non paramétrique de Wilcoxon pour échantillons appariés (tableau 21) afin de comparer les médianes des deux durées :

Tableau 21 Test de Wilcoxon sur les médianes des échantillons de 27 et de 32 syllabes

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	40.56415	Pr >  t	<.0001
Sign	M	80.5	Pr >=  M	<.0001
Signed Rank	S	6520.5	Pr >=  S	<.0001

Nous observons que la probabilité de dépassement est largement inférieure au seuil .05, ce qui nous permet de rejeter l'hypothèse d'égalité des médianes. Il y a donc une différence statistiquement significative entre les durées totales des segments voisés extraits de 27 et de 32 syllabes.

#### **IV. Les effets principaux et les effets simples**

Nous avons ensuite souhaité savoir si les scores AVQI différaient selon la version utilisée et/ou selon que l'on effectue l'analyse sur 27 ou sur 32 syllabes.

La condition de distribution normale des scores pour les quatre croisements a tout d'abord été vérifiée. Les résultats du test de normalité Shapiro-Wilk (tableaux 22 à 25) montrent une distribution non gaussienne pour les quatre conditions :

Tableau 22 Tests de normalité pour les échantillons de 32 syllabes, analysés par l'AVQI 02.03

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.929164	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.111406	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.582854	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.810801	Pr > A-Sq	<0.0050

Tableau 23 Tests de normalité pour les échantillons de 32 syllabes, analysés par l'AVQI 03.01

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.920821	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.13302	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.782841	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.62184	Pr > A-Sq	<0.0050

Tableau 24 Tests de normalité pour les échantillons de 27 syllabes, analysés par l'AVQI 02.03

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.924067	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.120965	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.638915	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.054602	Pr > A-Sq	<0.0050

Tableau 25 Tests de normalité pour les échantillons de 27 syllabes, analysés par l'AVQI 03.01

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.924034	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.13056	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.771999	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.494095	Pr > A-Sq	<0.0050

Nous avons donc utilisé une ANOVA non paramétrique à mesures répétées (test de Friedman), en considérant les quatre croisements comme quatre niveaux séparés d'une mesure répétée.

#### A. Les analyses sur scores bruts

Le test de Friedman (tableau 26) effectué sur les scores bruts calculés par les deux versions AVQI sur les échantillons de 27 et de 32 syllabes nous donne les résultats suivants : avec une statistique Q de 161.8228 et une probabilité de dépassement < .001, nous devons rejeter l'hypothèse d'égalité des médianes.

Tableau 26 Test de Friedman sur les scores bruts des deux versions de l'AVQI pour les deux durées syllabiques

Cochran-Mantel-Haenszel Statistics (Based on Rank Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	61.4694	<.0001
2	Row Mean Scores Differ	3	161.8228	<.0001

Total Sample Size = 648

Nous avons alors effectué des comparaisons des médianes deux à deux (post-hoc) afin de pouvoir tirer des conclusions plus précises. Pour cela, nous avons utilisé des tests de Wilcoxon pour échantillons appariés (tableaux 27 à 32). Notons qu'avec la répétition des tests, le risque d'erreur de première espèce augmente à au moins 10 % au lieu de 5 (Chippaux, 2004). Afin de pallier ce risque, nous avons utilisé la correction de Bonferroni, classiquement appliquée afin de contrebalancer l'effet préjudiciable des comparaisons multiples. En divisant le niveau  $\alpha$  (.05) par le nombre de comparaisons effectuées (six), nous obtenons un seuil de .0083.

Ainsi, nous avons pu investiguer les effets principaux et simples suivants, qui se sont tous révélés statistiquement significatifs au seuil .0083. Les tableaux 33 à 35 reprennent les statistiques descriptives permettant de faciliter la compréhension des implications des différences mesurées :



1) *Les effets principaux :*

- a. L'effet de la longueur syllabique (32 vs 27 syllabes) de l'échantillon analysé sur les scores bruts calculés (tableau 27) : les scores bruts sont globalement plus élevés lorsque l'analyse AVQI est effectuée sur 32 syllabes (tableau 33).
- b. L'effet de la version de l'AVQI utilisée (version 02.03 vs 03.01) sur les scores bruts calculés (tableau 28) : les scores bruts sont globalement plus élevés lorsque l'analyse est effectuée à l'aide de la version 02.03 de l'AVQI (tableau 34).

2) *Les effets simples (tableau 35) :*

- a. L'effet de la version AVQI utilisée (version 02.03 vs 03.01) sur les scores bruts calculés sur 32 syllabes (tableau 29) : les scores bruts sont plus élevés avec la version 02.03 de l'AVQI.
- b. L'effet de la version AVQI utilisée (version 02.03 vs 03.01) sur les scores bruts calculés sur 27 syllabes (tableau 30) : les scores bruts sont plus élevés avec la version 02.03 de l'AVQI.
- c. L'effet de la longueur syllabique (32 vs 27 syllabes) de l'échantillon analysé sur les scores bruts calculés avec la version 02.03 de l'AVQI (tableau 31) : les scores bruts sont plus élevés lorsque l'analyse est effectuée sur 32 syllabes.
- d. L'effet de la longueur syllabique (32 vs 27 syllabes) de l'échantillon analysé sur les scores bruts calculés avec la version 03.01 de l'AVQI (tableau 32) : les scores bruts sont plus élevés lorsque l'analyse est effectuée sur 32 syllabes.

Nous pouvons ainsi conclure que les scores bruts diffèrent significativement selon la version AVQI utilisée (scores plus élevés pour la version 02.03), autant pour les échantillons de 32 que pour ceux de 27 syllabes ; de plus, les scores diffèrent significativement selon la longueur d'échantillon analysée (scores plus élevés sur 32 syllabes), pour les deux versions de l'AVQI. Rappelons que nous avons précédemment identifié que la longueur de 27 semblait davantage adaptée à l'analyse par l'AVQI, car elle permettait d'obtenir trois secondes de segments voisés.

Tableau 27 *Test de Wilcoxon sur les médianes des scores bruts calculés sur les échantillons de 27 et de 32 syllabes (les deux versions de l'AVQI confondues)*

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	5.456893	Pr >  t	<.0001
Sign	M	39	Pr >=  M	<.0001
Signed Rank	S	8536	Pr >=  S	<.0001

Tableau 28 *Test de Wilcoxon sur les médianes des scores bruts calculés par les deux versions de l'AVQI (durées syllabiques confondues)*

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	16.22048	Pr >  t	<.0001
Sign	M	102.5	Pr >=  M	<.0001
Signed Rank	S	20560	Pr >=  S	<.0001

Tableau 29 Test de Wilcoxon sur les médianes des scores bruts calculés par les deux versions de l'AVQI sur les échantillons de 32 syllabes

Tests for Location: Mu0=0			
Test	Statistic		p Value
Student's t	t	11.86719	Pr >  t  <.0001
Sign	M	50	Pr >=  M  <.0001
Signed Rank	S	5252	Pr >=  S  <.0001

Tableau 30 Test de Wilcoxon sur les médianes des scores bruts calculés par les deux versions de l'AVQI sur les échantillons de 27 syllabes

Tests for Location: Mu0=0			
Test	Statistic		p Value
Student's t	t	11.05903	Pr >  t  <.0001
Sign	M	52.5	Pr >=  M  <.0001
Signed Rank	S	5051.5	Pr >=  S  <.0001

Tableau 31 Test de Wilcoxon sur les médianes des scores bruts calculés sur les échantillons de 27 et de 32 syllabes par la version 02.03 de l'AVQI

Tests for Location: Mu0=0			
Test	Statistic		p Value
Student's t	t	4.013228	Pr >  t  <.0001
Sign	M	23.5	Pr >=  M  0.0002
Signed Rank	S	2283.5	Pr >=  S  <.0001

Tableau 32 Test de Wilcoxon sur les médianes des scores bruts calculés sur les échantillons de 27 et de 32 syllabes par la version 03.01 de l'AVQI

Tests for Location: Mu0=0			
Test	Statistic		p Value
Student's t	t	3.712318	Pr >  t  0.0003
Sign	M	15.5	Pr >=  M  0.0157
Signed Rank	S	1969.5	Pr >=  S  0.0004

Tableau 33 Effet principal « longueur syllabique » : modes et médianes des scores bruts par longueur d'échantillon (indépendamment des versions AVQI)

Variable	Mode	Median
syll32	1.1900000	3.4400000
syll27	2.7900000	3.2450000

Tableau 34 Effet principal « version AVQI » : modes et médianes des scores bruts par version (indépendamment des longueurs d'échantillons)

Variable	Mode	Median
version2	3.1000000	3.6650000
version3	2.7900000	2.8350000

Tableau 35 Effets simples : modes et médianes des scores bruts pour les 4 croisements

Variable	Mode	Median
syll32vers2	3.1300000	3.7300000
syll32vers3	1.1900000	2.9050000
syll27vers2	1.2200000	3.5900000
syll27vers3	2.7900000	2.7900000

## B. Le classement « normophonie-dysphonie »

Ce qui nous intéresse principalement dans notre étude (notamment pour l'octroi d'un remboursement INAMI partiel pour la prise en charge vocale), c'est la répartition des participants dans les catégories « normophonie » et « dysphonie ». Nous avons donc classé les enregistrements en deux groupes : « normophonie » (AVQI 02.03 < 2.95 ; AVQI 03.01 < 2.43) et « dysphonie » (AVQI 02.03 ≥ 2.95 ; AVQI 03.01 ≥ 2.43). Ensuite, nous avons effectué des tests de Wilcoxon sur échantillons appariés afin de voir si le jugement « normophonique »/« dysphonique » différait selon la longueur de l'échantillon analysé et/ou la version de l'AVQI employée. À nouveau, nous avons employé le seuil  $\alpha = .0083$ . Les tableaux 38 et 39 reprennent les statistiques descriptives permettant de faciliter la compréhension des implications des différences mesurées<sup>30</sup> :

<sup>30</sup> En considérant les nombreux ex æquo dans nos données ordinales, montrer les médianes ne serait pas informatif. En effet, le test de Friedman considère la distribution des rangs par colonne. Si cette distribution est aléatoire (donc que les petits ou les grands rangs ne se concentrent pas dans une seule colonne spécifique), nous devons tolérer l'hypothèse d'égalité des médianes. Ce ne sont donc pas les médianes en tant que telles qui sont directement comparées.



a. *La longueur de l'échantillon* (tableau 36) :

Les deux versions de l'AVQI confondues, avec  $W = 34$  et  $p = .08$  ( $> .0083$ ), nous ne pouvons pas rejeter l'hypothèse d'égalité des médianes et pouvons ainsi conclure que le classement « normophonique »/« dysphonique » ne diffère pas significativement selon la longueur de l'échantillon analysé. Effectivement, celui-ci ne diffère que de 2.47 % (tableau 38). Les effets simples sont également non significatifs (effet de la longueur syllabique en version 02.03 :  $W = 13.5$ ,  $p = 0.07$  ; en version 03.01 :  $W = 4.5$ ,  $p = 0.73$ ).

b. *La version de l'AVQI* (tableau 37) :

Les deux longueurs syllabiques confondues, avec  $W = 367.5$  et  $p < .0001$  ( $< .0083$ ), nous rejetons l'hypothèse d'égalité des médianes et pouvons ainsi conclure que le classement « normophonique »/« dysphonique » diffère selon la version de l'AVQI employée. Les effets simples sont également significatifs (effet de la version sur 32 syllabes :  $W = 104.5$ ,  $p < .0001$  ; effet de la version sur 27 syllabes :  $W = 75$ ,  $p = .0007$ ). La version 02.03 est plus « sévère », assignant davantage d'échantillons à la catégorie « dysphonique » (tableau 39). Nous notons une différence de classement de 10.81 % entre les deux versions de l'AVQI.

Tableau 36 *Test de Wilcoxon sur les jugements « normophonique/dysphonique » pour les échantillons de 27 et de 32 syllabes*

Tests for Location: $\mu_0=0$				
Test		Statistic		p Value
Student's t	t	2.009353	Pr >  t	0.0453
Sign	M	4	Pr >=  M	0.0768
Signed Rank	S	34	Pr >=  S	0.0768

Tableau 37 *Test de Wilcoxon sur les jugements « normophonique/dysphonique » pour les deux versions de l'AVQI*

Tests for Location: $\mu_0=0$				
Test		Statistic		p Value
Student's t	t	5.72814	Pr >  t	<.0001
Sign	M	17.5	Pr >=  M	<.0001
Signed Rank	S	367.5	Pr >=  S	<.0001

Tableau 38 *Pourcentages des classements selon la longueur syllabique de l'échantillon*

	Jugement « normophonie »	Jugement « dysphonie »
<b>32 syllabes</b>	37.34 %	62.66 %
<b>27 syllabes</b>	39.81 %	60.19 %
différence	2.47 %	2.47 %

Tableau 39 *Pourcentages des classements selon la version de l'AVQI*

	Jugement « normophonie »	Jugement « dysphonie »
<b>Version 02.03</b>	33.33 %	66.67 %
<b>Version 03.01</b>	44.14 %	55.86 %
différence	10.81 %	10.81 %

## V. SYNTHÈSE DES RÉSULTATS

### CONCERNANT LES SCORES BRUTS

- **Q1 : Effet principal « version de l'AVQI » et effets simples** : y a-t-il une différence statistiquement significative entre les scores calculés par les deux versions de l'AVQI (02.03 vs 03.01) en français ?

→ **OUI**. Les scores calculés par la version 02.03 sont globalement plus élevés.

- **Q2 : Effet principal « durée syllabique » et effets simples** : y a-t-il une différence statistiquement significative entre les scores selon que l'analyse est effectuée sur 27 ou sur 32 syllabes ?

→ **OUI**. Les scores calculés sur 32 syllabes sont globalement plus élevés.

### CONCERNANT LE CLASSEMENT « NORMOPHONIE » - « DYSPHONIE »

- **Q3 : Effet principal « version de l'AVQI » et effets simples** : y a-t-il une différence statistiquement significative entre les classements selon la version de l'AVQI (02.03 vs 03.01) utilisée ?

→ **OUI**. La version 02.03 classe davantage de voix dans la catégorie des dysphonies.

- **Q4 : Effet principal « durée syllabique » et effets simples** : y a-t-il une différence statistiquement significative entre les classements selon que l'analyse est effectuée sur 27 ou sur 32 syllabes ?

→ **NON**. Bien que les scores bruts diffèrent significativement selon la longueur syllabique analysée, les classements « normophonie » - « dysphonie » ne diffèrent pas significativement.

### CONCERNANT LA DURÉE SYLLABIQUE DE L'ÉCHANTILLON

- **Q5 : Pouvons-nous utiliser un extrait d'uniquement 27 syllabes afin d'éviter un début de phrase inachevée ?**

→ **OUI**. La durée des segments voisés concaténés des extraits de 27 syllabes s'équilibre davantage avec les trois secondes de la voyelle tenue.

## VI. Discussion

### A. Concernant les scores bruts

Les deux effets principaux (« durée syllabique de l'échantillon analysé » et « version de l'AVQI ») sont significatifs, de même que les effets simples.

Voici les tableaux reprenant les médianes et les modes des scores bruts pour ces différents effets (tableaux 40 à 42). Pour rappel, les scores n'étant pas normalement distribués, nous ne pouvons pas comparer les moyennes.

Tableau 40 Effet principal « version AVQI » : modes et médianes des scores bruts par version (indépendamment des longueurs d'échantillons)

Variable	Mode	Median
version2	3.1000000	3.6650000
version3	2.7900000	2.8350000

Tableau 41 Effet principal « longueur syllabique » : modes et médianes des scores bruts par longueur d'échantillon (indépendamment des versions AVQI)

Variable	Mode	Median
syll32	1.1900000	3.4400000
syll27	2.7900000	3.2450000

Tableau 42 Effets simples : modes et médianes des scores bruts pour les 4 croisements

Variable	Mode	Median
syll32vers2	3.1300000	3.7300000
syll32vers3	1.1900000	2.9050000
syll27vers2	1.2200000	3.5900000
syll27vers3	2.7900000	2.7900000

#### ***L'effet de la version utilisée (tableau 40) :***

Les scores bruts sont généralement plus élevés pour la version 02.03 que pour la version 03.01. La version 02.03 peut donc être considérée comme « plus sévère ». Ceci peut être expliqué par le fait que chaque version repose sur une équation propre. En effet, la version 02.03 utilise l'équation suivante :

$$AVQI_{02.03} = \{[3.295 - (0.111 \times CPPS) - (0.073 \times HNR) - (0.213 \times Shim) + (2.789 \times ShdB) - (0.032 \times Slope) + (0.077 \times Tilt)] \times 2.208\} + 1.797$$

L'équation de la version 03.01 est la suivante :

$$AVQI_{03.01} = [4.152 - (0.177 \times CPPS) - (0.006 \times HNR) - (0.037 \times Shim) + (0.941 \times ShdB) + (0.01 \times Slope) + (0.093 \times Tilt)] \times 2.8902$$

Comme nous pouvons le constater, les pondérations des mesures sont différentes. L'ajustement de l'équation de l'AVQI a été effectué suite au constat que la voyelle tenue expliquait une plus grande part de la variance du score AVQI que la parole continue (Barsties & Maryn, 2015). En d'autres termes, dans la version 02.03, la voyelle tenue a une influence plus importante sur le score AVQI que l'échantillon de parole continue. La version 03.01 est une version adaptée

permettant d'égaliser l'influence des deux types de phonation. En conséquence, il est cohérent que les scores calculés par les deux équations diffèrent.

***L'effet de la durée syllabique analysée (tableau 41) :***

Les scores bruts diffèrent significativement en fonction de la durée de l'échantillon de parole continue analysée. Les scores sur 32 syllabes sont quelque peu plus élevés que ceux sur 27 syllabes (même si, comme nous le voyons dans le tableau 41, les médianes semblent proches : 3.44 vs 3.25). Une explication possible pour cette différence est que les 5 dernières syllabes prises en compte dans le calcul de l'AVQI sur 32 syllabes contiennent uniquement des phonèmes voisés. Ainsi, les médianes des durées totales des segments voisés concaténés (3.63 secondes pour 32 syllabes, 3 secondes pour 27 syllabes) diffèrent de 0.63 seconde. Ceci représente presque 20 % de ces durées totales, ce qui est proportionnellement non négligeable. Ce pourrait être une portion suffisamment longue pour amplifier certaines des mesures de perturbation. De plus, les cinq syllabes supplémentaires constituent le début d'une nouvelle phrase (« A midi, nerveux, ... »). Pour chaque participant, nous observons donc une initiation de la vibration cordale sur la voyelle [a] (« A midi »). Aussi, certains participants marquent un silence après la virgule séparant le second et le troisième mot (« A midi, nerveux, »), et ont donc à effectuer un arrêt de la vibration cordale, suivi d'une seconde initiation phonatoire. Comme nous le savons, la mise en vibration des plis vocaux est dans de nombreux cas problématique pour les individus souffrant d'une dysphonie, comme dans le cas d'une lésion cordale induisant des modifications de masse et de tension (Ikuma, Kunduk, Fink, & McWhorter, 2016 ; Schaeffler, Beck, & Jannetts, 2015). De ce fait, l'inclusion de ce début de phrase et donc de deux initiations de voisement supplémentaires dans les analyses de l'AVQI contribue très probablement à l'augmentation de certaines mesures de perturbation.

**B. Concernant le classement « normophonie – dysphonie »**

***L'effet de la version utilisée :***

Nous avons observé que la version 02.03 de l'AVQI semblait plus « sévère » que la version 03.01, résultant en des scores bruts globalement plus élevés. Néanmoins, rappelons la différence des scores seuils des deux versions : en effet, bien que la version 02.03 de l'AVQI semble calculer des scores plus sévères, le score seuil est plus élevé (2.95). Au contraire, le calcul de la version 03.01 résulte en des scores moins élevés, mais le score seuil est lui aussi plus bas (2.43). Nous avons ainsi émis l'hypothèse que la différence de classement « normophonie »/« dysphonie » pourrait être amoindrie par la différence des scores seuils.

Cependant, nos résultats montrent que le classement diffère effectivement selon la version de l'AVQI employée, en prenant en compte les scores seuils différents pour les deux versions. Nous pouvons ainsi observer que la fréquence d'attribution aux deux catégories diffère de 10.81 %. Davantage d'échantillons sont jugés dysphoniques par la version 02.03. La version 03.01 est moins « sévère », assignant davantage d'échantillons à la catégorie « normophonie ».

#### ***L'effet de la durée syllabique analysée :***

Nous avons observé que les scores bruts différaient significativement selon la longueur syllabique analysée par l'AVQI. La longueur de l'échantillon analysé n'a cependant pas d'effet significatif sur le classement « normophonie » - « dysphonie ». En reprenant notre raisonnement explicité ci-avant, nous émettons l'hypothèse que les différences significatives pour les scores bruts concernent principalement les dysphonies modérées à sévères. En effet, nous avons expliqué que l'extrait de 32 syllabes comprend deux initiations phonatoires supplémentaires. Or, la mise en vibration des plis vocaux est d'autant plus difficile que la dysphonie est sévère. Ainsi, les scores situés aux alentours du score seuil de l'AVQI (aux environs de 3) ne seraient pas affectés de la même manière par l'ajout de deux initiations phonatoires que les scores situés plus haut dans le continuum de sévérité. Aussi, imaginons que nous effectuions une analyse comparant non seulement le classement « normophonie » - « dysphonie » selon la longueur syllabique, mais également le classement par l'AVQI dans des catégories plus précises (normophonie, dysphonie légère, dysphonie modérée et dysphonie sévère, par exemple). Nous émettons alors l'hypothèse que les classements différeraient davantage au niveau des dysphonies modérées à sévères, selon la longueur syllabique analysée.

#### **C. Concernant les durées syllabiques des échantillons**

Nous observons une différence statistiquement significative entre les durées des segments voisés concaténés de 27 et de 32 syllabes. La durée de 27 syllabes s'équilibre avec les trois secondes de voyelle tenue, tandis que la durée médiane pour les 32 syllabes s'en écarte de 0.63 seconde, ce qui excède la marge de tolérance de 0.1 seconde (Barsties & Maryn, 2015). De plus, nous avons constaté des différences significatives entre les scores bruts selon la longueur syllabique analysée, qui n'était cependant pas observée dans le classement « normophonie - dysphonie ». À choisir entre les deux extraits, nous pensons donc que, de par son équilibre temporel avec les trois secondes de voyelle tenue, l'extrait de 27 syllabes est le plus adéquat pour l'analyse de l'AVQI 03.01 en français.

## VII. Conclusion

Les *27 premières syllabes* du texte phonétiquement équilibré de Harmegnies seront utilisées pour l'étude de validation de l'AVQI 03.01 en langue française. Cet extrait semble plus représentatif de la langue cible que les 21 premières syllabes employées dans l'étude de validation de Maryn et al. en 2014. En effet, il y manque moins de phonèmes voisés, et l'équilibre phonétique est meilleur. Cet extrait permet aussi un équilibre temporel adéquat entre l'extrait lu et la voyelle tenue, puisqu'il respecte les trois secondes préconisées. Un schéma résumant le fil de pensée ayant conduit au choix final de cette longueur syllabique est disponible en annexe 8 (page 177). Les scores AVQI bruts diffèrent significativement selon la version employée. La *version 03.01* présente une meilleure cohérence interne, car elle prend en compte l'équilibre des durées de la parole continue et de la phonation tenue. Nous choisissons cette version, plus récente et davantage « perfectionnée » et aboutie. De plus, la version 03.01 de l'AVQI n'a jamais été validée sur un échantillon en français, ce qui souligne la nécessité de valider cette version dans notre étude.

### La validité concourante et la précision diagnostique de l'AVQI 03.01 appliqué au français

La prochaine étape de notre étude était d'évaluer la validité concourante de l'AVQI 03.01 sur l'extrait de 27 syllabes du texte en français phonétiquement équilibré de Harmegnies. L'objectif était donc ici de mesurer la corrélation entre les scores calculés par l'AVQI 03.01 et le jugement à l'aide du critère G de la GRBAS par des vocologistes expérimentés. En effet, rappelons que l'évaluation à l'écoute reste toujours le « gold standard » par rapport auquel les nouveaux outils d'évaluation vocale sont comparés et validés. La qualité vocale est un aspect perceptif multidimensionnel, résultat de l'interaction du signal vocal et de la perception par l'auditeur ; elle n'est donc pas une caractéristique du signal acoustique à proprement parler (Kreiman & Gerratt, 1998).

Initialement, nous souhaitions vérifier le lien entre le score AVQI et l'évaluation perceptive par le clinicien à l'aide de l'outil CAPE-V (décrit en page 20). En effet, Kempster et al. (2009) ont montré que cet outil présente certains avantages par rapport à l'échelle GRBAS. Il comporte un protocole standardisé, et les échelles sont métriques, facilitant ainsi les analyses statistiques. Karnell et al. (2007) observent que l'outil CAPE-V est plus sensible que l'échelle GRBAS. Cette observation est confirmée par Nemr et al. (2012), qui mesurèrent une forte corrélation entre les deux outils, mais une sensibilité plus importante pour l'outil CAPE-V concernant la détection de légères altérations vocales. Nous avons néanmoins utilisé le critère G de la GRBAS, étant donné

que Maryn et al. l'ont employé dans leurs études, mais également parce que son utilisation est largement répandue dans le monde de par sa validation dans de nombreuses langues. L'utilisation de cet outil permettra donc d'éventuelles comparaisons avec d'autres études, notamment dans le cadre d'une revue systématique.

## I. La population

Pour étudier la validité de l'AVQI, nous avons choisi d'utiliser 90 enregistrements parmi les 162, ceci afin de créer des groupes à effectifs égaux pour trois degrés de dysphonie : normophonie, dysphonie légère à modérée, et dysphonie sévère. Pour cela, nous avons classé les scores bruts en trois catégories en utilisant les critères définis dans le tableau 43.

Une répartition équitable hommes/femmes a également été respectée dans chacun des trois groupes, soit 3x2x15. La sélection s'est aussi opérée en fonction de l'accord obtenu entre les deux versions de l'AVQI (02.03 et 03.01) selon les degrés de sévérité attribués (1 : normophonie, 2 : dysphonie légère à modérée, 3 : dysphonie sévère). Nous avons donc choisi de n'utiliser que des enregistrements pour lesquels la même conclusion est tirée en fonction du score calculé par les deux versions de l'AVQI.

Tableau 43 *Critères d'attribution catégorielle des échantillons sonores en fonction du score brut de l'AVQI*

Degré de sévérité	Scores AVQI 02.03	Scores AVQI 03.01
Normophonique (attribution à la catégorie 1)	score < 2.95	score < 2.43
Dysphonie légère à modérée (attribution à la catégorie 2)	$2.95 \leq \text{score} < 6^{31}$	$2.43 \leq \text{score} < 6$
Dysphonie sévère (attribution à la catégorie 3)	$6 \leq \text{score}$	$6 \leq \text{score}$

De plus, il nous a semblé nécessaire d'ajouter un groupe contrôle, apparié en genre et en âge aux participants dysphoniques du groupe expérimental. Bien que 30 individus parmi les 90 ont été diagnostiqués normophoniques, nous estimons qu'ils appartiennent bien au groupe expérimental, dans la mesure où la plupart d'entre eux avaient décidé de consulter pour un trouble de la voix, avec une plainte vocale formulée. De ce fait, nous ne pouvons pas les considérer comme participants contrôle. Nous avons donc recruté 30 personnes supplémentaires, quinze hommes et quinze femmes, afin d'enregistrer les deux premières phrases du texte de Harmegnies (extrait à 27 syllabes), ainsi que la voyelle [a] tenue. La moyenne d'âge des 90 participants du groupe

<sup>31</sup> Le seuil de 6 a été choisi subjectivement en mettant en lien les scores AVQI et les scores G (échelle GRBAS) attribués au moment du bilan vocal par les logopèdes vocologistes du CHU. En effet, nous avons constaté qu'un score AVQI supérieur à 6 correspondait généralement à un G de 3. Ce seuil correspond également aux données récoltées par Maryn et al. (2012) et par Reynolds et al. (2012).

expérimental est de 52.85 ans<sup>32</sup> (min : 22.72, max : 88.34), la moyenne d'âge des participants du groupe contrôle est de 46.27 (min : 36.15 max : 60.87)<sup>33</sup>. Ces individus n'avaient aucun antécédent de trouble de la voix ni aucune plainte vocale actuelle. Nous avons volontairement refusé d'employer un grand nombre de critères d'exclusion supplémentaires, car nous souhaitons que le groupe contrôle soit représentatif de la population générale sans trouble vocal. Les seuls critères d'inclusion étaient donc l'absence d'un trouble vocal passé avéré, l'absence d'une plainte vocale subjective actuelle, l'absence d'une dysphonie telle que perceptivement évaluée par le testeur, ainsi que la présence d'un score AVQI inférieur à 2.43.

Les caractéristiques des participants contrôles et de l'échantillon total sont représentées dans les tableaux 44 et 45, respectivement.

Tableau 44 *Caractéristiques principales de l'échantillon de 30 participants du groupe contrôle*

Nombre total	30		
Genre	Femme : 15 (50 %) Homme : 15 (50 %)		
Âge	$\mu^a = 46.27$ $\sigma = 7.91$ Min = 36.15 Max = 60.87		
Professions (n et %) <sup>b</sup>	Professionnels de la santé (professions médicales et paramédicales, professionnels médico-techniques)	15	50 %
	Personnel des services aux personnes et à la collectivité (agents de sécurité, brancardiers...)	3	10 %
	Personnel des services administratifs et commerciaux (employés administratifs, secrétaire...)	10	33.33 %
	Professionnels de l'informatique (informaticiens)	2	6.67 %

Note. <sup>a</sup> La distribution des âges suit la loi normale selon le test Shapiro Wilk ( $W = .975, p = .09$ ). Nous pouvons donc ici renseigner la moyenne et l'écart-type. <sup>b</sup> Nous avons choisi de regrouper les professions selon l'arborescence du service public wallon de l'emploi et de la formation (Forem), accessible via le lien suivant : <https://www.leforem.be/former/horizonemploi/metier-profession-index.html> (Service public wallon de l'emploi et de la formation, 2016)

Tableau 45 *Caractéristiques principales de l'échantillon total de 120 patients (90 patients de la base de données et 30 participants contrôles)*

Nombre total	120		
Genre	Femme : 60 (50 %) Homme : 60 (50 %)		
Age <sup>a</sup>	Médiane <sup>b</sup> = 49.04 Intervalle interquartile = 20.06 Min = 22.72 Max = 88.34		

<sup>32</sup> Les âges de deux participants n'ont pas pu être intégrés dans le calcul de la moyenne, les données disponibles dans leurs dossiers de la base de données du CHU étant incomplètes. Les calculs sont donc effectués sur 88 participants.

<sup>33</sup> Le groupe contrôle n'est donc pas tout à fait apparié au niveau de l'âge avec le groupe expérimental. Nous avons recruté les participants du groupe contrôle en nous basant sur l'écart-type de l'âge du groupe expérimental : avec une moyenne de 52 ans et un écart-type de 16, nous avons choisi de recruter des personnes âgées entre 36 et 68 ans (donc  $52 \pm 16$ ). Il aurait été préférable d'apparier les deux groupes en nombre (donc de recruter 90 participants contrôles), pour ensuite pouvoir correctement apparier les deux groupes au niveau de l'âge. Une autre solution aurait été d'apparier les 30 participants contrôles avec les 30 participants du groupe expérimental diagnostiqués comme « normophoniques ».



Pathologies (n et %) <sup>c</sup>	– Consultation sans pathologie <sup>d</sup>	38	31.67 %
	– Immobilité laryngée	22	18.33 %
	– Dysphonie fonctionnelle <sup>e</sup>	9	7.5 %
	– Post-phonochirurgie	8	6.67 %
	– Presbyphonie	8	6.67 %
	– Œdème cordal	7	5.83 %
	– pathologie non renseignée	7	5.83 %
	– Polype	4	3.33 %
	– Lésion nodulaire	3	2.5 %
	– Sulcus/vergeture	3	2.5 %
	– Lésion tumorale	3	2.5 %
	– Kyste	3	2.5 %
	– Cicatrice cordale/fibrose	2	1.67 %
	– Dysphonie spasmodique	1	0.83 %
	– Corpectomie	1	0.83 %
– Aryténoïdectomie	1	0.83 %	
Professions (n et %) <sup>f</sup>	– Pensionnés/retraités	24	20 %
	– Professionnels de la santé (professions médicales et paramédicales)	24	20 %
	– Personnel des services administratifs et commerciaux (employé de bureau, comptable, secrétaire médical, agent administratif...)	14	11.67 %
	– Professionnels de la formation initiale et de la formation continue (enseignants)	13	10.83 %
	– Profession non renseignée	9	7.5 %
	– Personnel, techniciens et cadres du transport et de la logistique (chauffeur routier, chauffeur de taxi, magasinier, agent distributeur...)	7	5.83 %
	– Personnel des services aux personnes et à la collectivité (technicienne de surface, nettoyage industriel, militaire, coiffeur...)	6	5 %
	– Cadres administratifs et professionnels de l'information et de la communication (PDG, juriste, avocat, assistant parlementaire...)	5	4.17 %
	– Chômage, invalidité	3	2.5 %
	– Personnel de la distribution et de la vente (commerçant, délégué commercial...)	2	1.67 %
	– Étudiants	2	1.67 %
	– Professionnels des arts et du spectacle (chanteur, photographe...)	2	1.67 %
	– Personnel de type artisanal (ébéniste, boulanger...)	2	1.67 %
	– Professionnels de l'informatique	2	1.67 %
	– Agents de maîtrise, techniciens et ingénieurs du bâtiment, des travaux publics et de l'extraction (architecte, métreur-deviseur)	1	0.83 %
	– Personnel des industries graphiques (copiste)	1	0.83 %
	– Techniciens industriels (technicien en mécanique)	1	0.83 %
	– Professionnels de l'intervention sociale, du développement local et de l'emploi (assistant social)	1	0.83 %
	– Personnel du bâtiment, des travaux publics et de l'extraction (ouvrier en bâtiment)	1	0.83 %

Note. <sup>a</sup> Les âges de deux participants n'ont pas pu être intégrés dans le calcul de la médiane, les données disponibles dans leurs dossiers de la base de données du CHU étant incomplètes. Ce calcul est donc effectué sur 118 participants. <sup>b</sup> Le test de Shapiro-Wilk montre une statistique W de .975, avec  $p = .027$ . Nous devons donc rejeter l'hypothèse de normalité de la distribution des âges dans notre échantillon. Ainsi, nous préférons la médiane et l'intervalle interquartile à la moyenne et à l'écart-type. <sup>c</sup> Les diagnostics laryngologiques ont été effectués à l'aide d'un vidéo-laryngo-stroboscope EndoSTROB (Xion GmbH, Berlin). Nous tenons à souligner que dans notre base de données, les dysphonies liées au cancer n'incluent que les patients de stade T1N0M0. <sup>d</sup> Sont ici compris (a) les participants contrôles, (b) les participants ayant consulté, mais ne présentant pas de pathologie avérée, y compris entre autres les bilans préopératoires dans le cadre d'interventions chirurgicales touchant la glande thyroïde, d'interventions chirurgicales ou toute autre intervention au cours de laquelle peut avoir lieu une lésion des nerfs laryngés. <sup>e</sup> Les dysphonies dites « psychogènes » sont incluses dans cette catégorie. <sup>f</sup> Nous avons choisi de regrouper les professions selon l'arborescence du service public wallon de l'emploi et de la formation (Forem), accessible via le lien suivant : <https://www.leforem.be/former/horizonemploi/metier-profession-index.html> (Service public wallon de l'emploi et de la formation, 2016).

## **II. Les enregistrements**

Les enregistrements des participants contrôles ont été effectués dans les mêmes conditions que ceux effectués lors des bilans vocaux au CHU de Liège. En effet, ils ont été réalisés dans le même local, avec le même support écrit pour le texte de Harmegnies (texte écrit en grande police, lecture du texte en entier), en utilisant le microphone serre-tête cardioïde à condensateur AKG C520 maintenu à 3-6cm de la bouche, via l'interface « CSL Main Program » d'un poste de travail informatique « Computerized Speech Lab » de KayPENTAX, modèle 4500, et ce avec les mêmes réglages d'enregistrement que lors des bilans (enregistrement « mono », taux d'échantillonnage de 11025 Hz, 16 bits). De plus, nous avons mesuré le rapport harmoniques-bruit sur ces enregistrements en post-hoc, afin de vérifier si le bruit ambiant était suffisamment faible que pour pouvoir prétendre à des analyses acoustiques de qualité. En effet, le rapport harmoniques-bruit est considéré comme correspondant au rapport signal-bruit dans le cadre d'un signal supposé périodique (Boersma & Weenink, 2003). Nous avons donc mesuré le rapport harmoniques-bruit sur les enregistrements de voyelle tenue — dont le signal est supposé périodique. Les valeurs sont considérées comme acceptables lorsqu'elles se situent au-delà de 30 dB ; des valeurs supérieures à 42 dB sont recommandées (Deliyski, Shaw, & Evans, 2005). D'autres auteurs mentionnent un SNR acceptable au-delà de 25 dB (Carson, Ingrisano, & Eggleston, 2002). La moyenne des rapports harmoniques-bruit pour nos enregistrements est de 25.22 dB, avec un écart-type de 3.28 dB. Même en considérant les données moins strictes de Carson et al. (2002), ces valeurs sont insatisfaisantes. Les conditions de bruit ambiant ne semblent donc, selon ces mesures, pas optimales. L'utilisation d'un décibel-mètre PCE-353 (PCE, Melrose, États-Unis) nous avait déjà permis de confirmer que le bruit environnant dans le laboratoire vocal était en moyenne de 50.33 dB (moyenné sur 4 mesures à différents moments d'une journée) et dépassait ainsi les seuils conseillés.

Avant de pouvoir faire évaluer perceptivement ces 120 enregistrements par les vocologistes, nous avons créé des fichiers sons adéquats. Aussi, nous avons concaténé les 27 syllabes du texte lu et les trois secondes centrales de voyelle tenue, afin d'obtenir un fichier son identique à celui analysé par l'AVQI. Pour extraire les trois secondes de voyelle identiques à celles extraites par l'AVQI, nous avons employé la partie du script responsable de cette action. Pour la parole continue, nous avons déjà sauvegardé les extraits de 27 syllabes utilisés pour les analyses AVQI. Ensuite, nous avons ajouté à notre nouveau script le contenu d'un autre script (du site web « Praatpfanne »), permettant de concaténer les deux fichiers sons (parole continue et voyelle tenue) avec un silence de deux secondes qui les sépare (Mayer, 2013).

Nous pouvons ainsi garantir pour l'évaluation perceptive l'emploi d'extraits strictement analogues à ceux utilisés pour l'évaluation acoustique<sup>34</sup>. Nous avons effectué une dizaine d'analyses AVQI en utilisant les nouveaux fichiers ainsi créés, et avons ainsi pu vérifier que les scores AVQI étaient bien identiques à ceux calculés auparavant.

Pour résumer : le script que nous avons assemblé (disponible à l'annexe 9, page 178) nous a permis de concaténer les extraits identiques à ceux utilisés par l'AVQI, et de les séparer d'un intervalle de deux secondes comme préconisé par Maryn et al. (2009). Voici à la figure 17 un exemple de fichier assemblé :

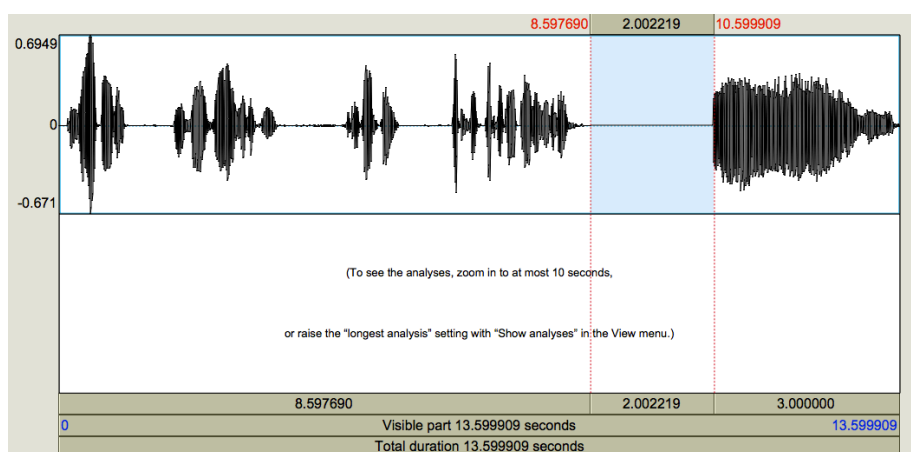


Figure 17. Exemple de fichier audio créé grâce au script composé.

À gauche = l'extrait de parole continue ; à droite = les trois secondes centrales de la voyelle tenue ; en bleu = intervalle de deux secondes.

### III. L'interface d'évaluation

Une fois les fichiers sons créés, nous avons composé un script définissant l'interface pour l'évaluation perceptive. Nous nous sommes inspirés du script de Véronique Delvaux (2016), lui-même adapté d'un script disponible sur le site web « Praatpfanne » (Mayer, 2011). Ce script permet de présenter dans un ordre aléatoire tous les fichiers « .wav » présents dans un dossier. Lors de la lecture du fichier son, le juge attribue aux différents critères de la GRBAS un score entre 0 et 3, comme illustré à la figure 18. Le juge a également la possibilité de réécouter l'enregistrement autant de fois que nécessaire. Une fois toutes les évaluations complétées, le script génère un fichier « .txt » reprenant l'identification de l'évaluateur, les noms des fichiers dans leur ordre de présentation, ainsi que le score attribué pour chacun d'eux.

<sup>34</sup> Notons à ce propos que chacun des quatre juges a rapporté avoir eu des difficultés pour l'évaluation perceptive dans la mesure où leur jugement différait parfois fortement selon qu'il était effectué sur base de la voyelle tenue ou sur l'extrait de parole continue ; les deux jugements contradictoires devaient alors être « moyennés ». Ce point sera discuté plus loin.

	Replay			
Grade	0	1	2	3
Roughness	0	1	2	3
Breathiness	0	1	2	3
Asthenia	0	1	2	3
Strain	0	1	2	3
Instability	0	1	2	3
	Next			

Figure 18. Illustration de l'interface d'évaluation perceptive des critères de la GRBAS, selon le script de Valérie Delvaux (2016).

Ne souhaitant utiliser que le critère G de la GRBAS, nous avons supprimé les lignes du script permettant aux juges d'évaluer les autres critères, afin d'épurer l'interface et d'éviter toute distraction inutile. Ensuite, nous avons quelque peu adapté le script afin de réharmoniser la présentation visuelle. Le résultat final est illustré à la figure 19 ; le script est disponible à l'annexe 10 (page 181).

	Replay			
Grade	0	1	2	3
	Next			

Figure 19. Illustration de la nouvelle interface pour l'évaluation du critère G de la GRBAS.

## IV. Les sessions d'évaluation perceptive

### A. Les juges

Nous pouvons enfin faire évaluer perceptivement les 120 fichiers sons. Quatre logopèdes spécialisées dans le domaine de la voix (quatre vocologistes), régulièrement en contact avec des patients dysphoniques et ayant eu au moins 5 ans d'expérience dans l'évaluation vocale perceptive (Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2009) ont accepté de participer à cette étude. Le choix d'évaluateurs ayant au moins 5 ans d'expérience dans l'écoute de voix pathologiques est soutenu par les données des études de Kreiman et al. (1993), qui soulignent que l'établissement de représentations stables pour les différents paramètres vocaux (raucité, souffle, degré général de dysphonie...) prend plusieurs années. En effet, certaines

études montrent qu'un entraînement intensif de courte durée à l'évaluation perceptive n'améliore pas significativement la fidélité des jugements.

Nous avons effectué un screening audiométrique pour chacune des quatre logopèdes, à l'aide d'un audiomètre MAICO MA 50 (MAICO Diagnostic GmbH, Berlin), à 250, 500, 1000, 2000, 4000 et 8000 Hz. Les seuils d'audition étaient tous inférieurs ou égaux à 20 dB entre 500 et 8000 Hz pour les quatre juges. Seule l'une d'entre elles présentait un seuil de 30 dB à 250 Hz à l'oreille droite. Notons cependant la présence de bruit ambiant important dans le local où a été effectué le testing ; il s'agissait d'un bruit d'une fréquence grave, ce qui pourrait expliquer un effet masquant plus important à ladite fréquence, et donc un seuil d'audition légèrement augmenté. De plus, la procédure de screening classique, ainsi que le calcul du déficit audiométrique moyen, n'incluent pas les deux valeurs extrêmes de 250 et de 8000 Hz. Celles-ci ne font effectivement pas partie de la zone des fréquences conversationnelles (Maltby & Knight, 2015). Notons aussi que de nombreux auteurs considèrent le seuil d'audition « normale » à 25 dB (Fausti, Wilmington, Helt, Helt, & Konrad-Martin, 2005) ; ce seuil est également conseillé par l'ASHA pour le screening chez l'adulte (American Speech-Language-Hearing Association, 2017a). Les quatre logopèdes ont ainsi pu être intégrées à notre étude. Un dernier critère d'inclusion était l'obtention d'une fidélité test-retest au moins « moyenne », c'est-à-dire un accord moyen entre les évaluations répétées d'un même juge, signé par un coefficient Kappa de Cohen pondéré supérieur à .41 (comme explicité ci-après, sous le titre « La fidélité inter-juges et intra-juge » en page 81).

## B. La tâche

La tâche d'évaluation a été effectuée par les quatre logopèdes à leur domicile. Des consignes précises ont été données quant aux conditions de passation. Il s'agissait de faire écouter les enregistrements dans les conditions optimales mentionnées par différents auteurs (Awan & Roy, 2006 ; Barsties & Maryn, 2015 ; Barsties & Maryn, 2016 ; Maryn et al., 2009) :

- les stimuli ont été présentés individuellement dans un casque, à intensité confortable ;
- l'évaluation a été effectuée dans un endroit calme, sans distracteur visuel ou auditif ;
- les juges ne connaissaient pas l'identité ni la pathologie des différents participants ;
- les enregistrements ont été présentés dans un ordre aléatoire<sup>35</sup> ;

---

<sup>35</sup> Notons que les enregistrements n'ont donc pas été présentés dans le même ordre aux quatre évaluateurs. Nous sommes conscients que ceci peut représenter une limite à notre étude, dans la mesure où l'ordre de présentation peut influencer les jugements de sévérité (par exemple, une voix présentée juste après une voix sévèrement dysphonique n'est pas évaluée de la même manière que lorsqu'elle est présentée après une voix normophonique). Ceci peut donc avoir un impact sur la fidélité inter-juges.

- la tâche a été effectuée en une seule « session », c'est-à-dire qu'il a été demandé aux logopèdes de ne pas fragmenter l'évaluation sur plusieurs journées, voire sur différents moments éloignés d'une même journée (p.ex. pendant les pauses de travail) ;
  - des pauses ont été proposées au participant dès qu'il en ressentait le besoin. En effet, certains auteurs proposent une pause systématique après un certain nombre d'enregistrements. Cependant, nous sommes de l'avis que chacun a des capacités d'attention soutenue propres. Imposer un même rythme à chaque juge pourrait donc diminuer l'effet bénéfique des pauses. Ainsi, chaque juge a reçu l'instruction de faire une pause de minimum dix minutes et de maximum vingt minutes. Selon des données récentes, la durée idéale pour une pause après cinquante-deux minutes de travail serait de dix-sept minutes (Draugiem Group, 2014). Cette durée permettrait au cerveau de « se désengager » afin de se sentir à nouveau d'attaque, sans toutefois induire une perte totale du focus attentionnel et du dynamisme (White, 2014).
  - les juges avaient la possibilité de répéter chaque enregistrement autant de fois que nécessaire.
- Afin d'augmenter la fidélité inter-juges, une définition écrite du critère G a été présentée au préalable, ainsi que huit voix de référence (deux par degré de dysphonie 0-3, dont une pour chaque genre). Ces repères externes ont été présentés à nouveau après chaque pause. Comme nous l'avons déjà décrit dans la partie théorique, la critique principale de l'évaluation perceptive concerne la stabilité du jugement entre différents évaluateurs et la stabilité dans le temps du juge lui-même (test-retest). Cette variabilité est plus accentuée pour les voix dysphoniques que pour les voix normales (Kreiman, Gerratt, Precoda, & Berke, 1992), et ce d'autant plus dans le cas de dysphonies légères à modérées que dans les extrêmes du continuum (Kreiman et al., 1993). Rappelons que l'importance de cette variabilité est communément expliquée par les référents idiosyncrasiques dissemblables, par rapport auxquels sont évaluées les voix dysphoniques. Chacun se construirait ses propres référents en fonction de ses expériences avec les voix dysphoniques, ce qui résulterait en des représentations internes qui diffèrent notamment en fonction des détails constituants (Kreiman et al., 1993). Ainsi, selon ces mêmes auteurs, chacun aurait en mémoire un nombre propre de degrés de sévérité pour les différents paramètres à évaluer. Représenté à la figure 20, voici un modèle proposé par Kreiman et al. (1993) pour expliquer les facteurs influençant l'évaluation perceptive de la qualité vocale. Ces facteurs sont :
- des facteurs propres à chaque évaluateur, dont l'expérience/l'exposition aux différents types de voix, les habitudes/stratégies perceptives et la sensibilité globale au facteur à

évaluer. Ces facteurs sont ancrés et difficiles à modifier, et ont un impact principalement sur la fidélité inter-juges ;

- des facteurs d'« erreur » affectant autant la fidélité inter-juges qu'intra-juge, tels la fatigue, les décrochages attentionnels, les maladresses ;
- des facteurs relatifs à la tâche d'évaluation, comme la définition plus ou moins claire du paramètre à évaluer, la réalité perceptive de celui-ci, ainsi que les caractéristiques de l'échelle d'évaluation (ordinaire, métrique...) et des stimuli à évaluer.

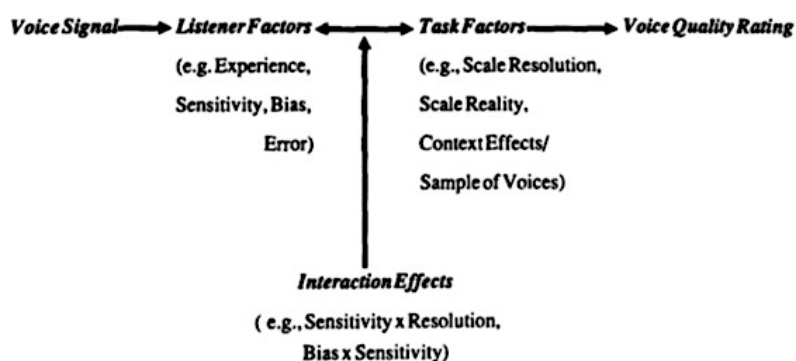


Figure 20. Facteurs influençant l'évaluation perceptive de la qualité vocale selon Kreiman et al. (1993).

C'est donc dans le but de garantir l'ancrage d'un référent interne plus stable et commun à tous les évaluateurs — comme l'ont également fait les auteurs de l'AVQI — que nous avons présenté des voix de référence. Ainsi, nous espérons une diminution de la variabilité inter-juges et de la variabilité test-retest. Notons toutefois que selon certains auteurs, la présentation de référents auditifs amène un jugement moins sévère (Eadie & Kapsner-Smith, 2011).

La sélection des voix de référence a été effectuée sur base de l'accord unanime de deux juges sur le critère G de la GRBAS. Ces enregistrements contenaient le même extrait de texte de 27 syllabes que les enregistrements qui ont été, par la suite, évalués par les quatre vocologistes. Les voix de référence n'ont bien évidemment pas été reprises dans les enregistrements à évaluer par le groupe de vocologistes.

En plus des voix de référence, nous avons aussi présenté la définition écrite du critère G de la GRBAS, comme suit : « *Le degré général de dysphonie est une mesure globale exprimant à quel point la voix est jugée comme étant "bonne" ou "mauvaise" par le juge* » (traduit de Eadie & Doyle, 2005). En effet, Awan et Lawson (2009) ont montré que la combinaison de référents auditifs (voix de référence) et écrits (définitions écrites des paramètres) semble être la meilleure solution pour améliorer la fidélité inter-juges. Ceci serait particulièrement important afin d'améliorer la distinction perceptive entre les dysphonies légères et les voix normophoniques. Brinca et al. ont eux aussi plus récemment démontré une amélioration de la fidélité inter-juges grâce à l'utilisation de référents auditifs et textuels, et ce principalement

pour les critères G et B sur de la parole continue. La fidélité pour le critère R reste, d'après ces auteurs, plus difficile à évaluer (Brinca, Batista, Tavares, Pinto, & Araujo, 2015).

Notons que sur tous les enregistrements à évaluer, les deux conditions phonatoires étaient présentées dans le même ordre, comme suit : parole continue – pause de deux secondes – voyelle tenue. Lors de l'interprétation des résultats de notre étude, nous garderons en tête que ceci peut constituer un biais. Effectivement, nous savons que l'évaluation perceptive est généralement plus positive quand elle est effectuée sur base de la parole continue, par rapport au jugement effectué sur une voyelle tenue (Maryn & Nelson, 2012). Nous aurions donc pu choisir de présenter certains enregistrements dans l'ordre inverse (voyelle tenue, puis parole continue), de manière aléatoire. Cependant, nous avons préféré garder le même ordre pour l'évaluation perceptive que celui utilisé par l'AVQI, afin d'effectuer des analyses de corrélation sur des données analogues. Notons aussi que le fait de ne pas prendre en compte les attaques et les finales des voyelles tenues entraîne probablement une diminution de la différence de jugement entre les deux conditions. La voyelle tenue sera moins sévèrement évaluée lorsque les délais de mise en vibration des plis vocaux et de l'arrêt du voisement sont retirés. Les jugements sur les deux conditions sont ainsi quelque peu mieux équilibrés, et le possible biais décliné.

## **V. La fidélité inter-juges et intra-juge**

Pour pallier la critique de manque de stabilité des jugements, de nombreux auteurs évaluent la fidélité inter- et intra-juges avant d'effectuer les analyses prévues. C'est aussi ce que nous avons choisi de faire afin de nous assurer que les données récoltées soient exploitables. Pour évaluer la fidélité intra-juges, nous avons répété, à l'insu des juges, 20 % de l'échantillon, c'est-à-dire 24 fichiers sons. En tout et pour tout, les vocologistes ont donc évalué 144 échantillons sonores.

La fidélité intra-juge (test-retest) a été évaluée pour chaque juge en mesurant le coefficient Kappa de Cohen ( $\kappa$ ) pondéré, sur les 24 enregistrements évalués à deux reprises à l'insu de l'évaluateur. Le coefficient Kappa de Cohen classique mesure les désaccords entre les jugements ordinaux de deux évaluateurs, ou entre deux jugements d'un même évaluateur. Cependant, il ne prend pas en compte le degré des désaccords. Cette faiblesse est compensée par l'utilisation du coefficient pondéré (qui attribue un « poids » à chaque désaccord) (Jakobsson & Westergren, 2005), qui ne peut être utilisé que dans le cas de valeurs ordinales (Santos, 2018). La valeur du coefficient de Kappa se situe entre -1 et 1 ; une valeur  $\kappa$  de 1 équivaut à un accord parfait, alors qu'un  $\kappa$  de 0 ou moins signe un désaccord total entre les deux jugements (McHugh, 2012).



Les valeurs obtenues peuvent être interprétées selon les recommandations de Landis et Koch décrites dans le tableau 46 (1977, cité par Shweta, Ram, & Himanshu, 2015) :

Tableau 46 *Interprétation du coefficient Kappa selon Landis et Koch (1977)*

$\kappa$	Interprétation de l'accord
< 0	Pauvre
.00-.20	Léger
.21-.40	Raisonnable
.41-.60	Modéré
.61-.80	Substantiel
.81-1.00	Presque excellent

Comme nous pouvons le constater dans le tableau 47, les quatre vocologistes présentent une fidélité test-retest importante, avec un coefficient Kappa pondéré moyen de .727. Les jugements effectués à deux reprises sont donc globalement concordants.

Tableau 47 *Coefficients Kappa pondérés pour l'évaluation de la fidélité intra-juge des jugements perceptifs sur la parole continue et la voyelle tenue*

Évaluateur	Coefficient de Kappa pondéré	p	Interprétation
1	.6717	< .0001	Accord substantiel, significatif
2	.7303	< .0001	Accord substantiel, significatif
3	.7778	< .0001	Accord substantiel, significatif
4	.7280	< .0001	Accord substantiel, significatif

Nous avons alors mesuré la fidélité inter-juges à l'aide du coefficient de concordance de Kendall — adapté pour comparer les jugements ordinaux de plus de deux juges (Perinetti, 2017) — dont les résultats sont représentés dans le tableau 48.

Avec un coefficient de concordance de Kendall de .88467 et une probabilité de dépassement de  $p < .0001$ , nous pouvons rejeter l'hypothèse nulle selon laquelle l'accord entre les juges est dû au facteur chance. Selon les recommandations d'interprétation de Schmidt (1997, cité par Remøy, 2010) décrites dans le tableau 49, l'accord inter-juges est élevé ( $> .70$ ).

Tableau 48 *Coefficient de concordance de Kendall pour l'évaluation de la fidélité inter-juges sur la parole continue et la voyelle tenue*

Coeff of Concordance	F	Num DF	Denom DF	Prob>F
0.88467	23.01	118.5	355.5	<.0001

Tableau 49 *Interprétation du coefficient de concordance de Kendall selon Schmidt 1997 (cité par Remøy, 2010)*

Kendall's W	Interpretation	Confidence in Ranks
.10	Very weak agreement	None
.30	Weak agreement	Low
.50	Moderate agreement	Fair
.70	Strong agreement	High
.90	Unusually strong agreement	Very high
W = 1	Complete agreement	Very high

Les jugements perceptifs semblent donc fiables, autant en intra-juge (test-retest) qu'entre les quatre évaluateurs, ce qui nous permet de poursuivre nos analyses.

## VI. La validité concurrente

Pour investiguer la validité concurrente de l'AVQI 03.01 appliqué à l'extrait en français du texte de Harmegnies (27 premières syllabes), nous avons mesuré la corrélation de Spearman entre les scores bruts calculés par l'AVQI et les degrés de sévérité moyens<sup>36</sup> (critère G de l'échelle GRBAS) tels qu'évalués par les quatre vocologistes. Les résultats sont représentés dans le tableau 50.

Avec un coefficient de corrélation de Spearman<sup>37</sup> de .85 et  $p < .0001$ , nous pouvons rejeter l'hypothèse nulle selon laquelle il n'y a pas de relation entre les scores de l'AVQI 03.01 et les jugements perceptifs du degré de sévérité de la dysphonie (G de la GRBAS).

La corrélation entre les scores AVQI 03.01 et les jugements perceptifs par les quatre vocologistes est forte ( $> .50$ ) selon les critères d'interprétation de Cohen (1988)<sup>38</sup> ; elle est également élevée (.71-.90) selon les critères plus sévères de Hinkle, Wiersma et Jurs (1994)<sup>39</sup>.

Tableau 50 *Corrélation de Spearman entre les scores AVQI et les évaluations moyennes du degré de sévérité (critère G) évalué par les quatre vocologistes sur base de la parole continue et de la voyelle tenue*

2 Variables: avqi gmean						
Simple Statistics						
Variable	N	Mean	Std Dev	Median	Minimum	Maximum
avqi	120	3.46933	2.91240	2.69500	-0.48000	10.02000
gmean	120	1.76667	0.98930	1.75000	0	3.00000

Spearman Correlation Coefficients, N = 120 Prob >  r  under H0: Rho=0		
	avqi	gmean
avqi	1.00000	0.85185 <.0001
gmean	0.85185 <.0001	1.00000

Nous pouvons donc conclure à une bonne validité concurrente de l'AVQI 03.01 appliqué sur un échantillon en langue française, lorsque les scores sont mis en lien avec l'évaluation perceptive, gold standard de l'évaluation vocale.

<sup>36</sup> Notons que de nombreux statisticiens déconseillent le calcul des moyennes sur des données catégorielles ordonnées (tel le critère G de la GRBAS). Malgré ce fait, d'après nos recherches dans la littérature scientifique, c'est pourtant bien ainsi que procèdent souvent les auteurs pour valider un outil en le comparant à un standard dont les résultats sont catégoriels.

<sup>37</sup> Certains auteurs emploient aussi le coefficient de détermination avec le coefficient de corrélation de Spearman. Cependant, son utilisation dans ce contexte est fortement critiquée. En effet, la corrélation de Spearman est utilisée sur les données transformées en rangs ; c'est un outil non paramétrique. L'utilisation du coefficient de détermination présume une relation linéaire, alors que la relation présumée lors de l'utilisation de la corrélation de Spearman est monotone, c'est-à-dire qu'elle évolue toujours dans le même sens, mais pas forcément de façon linéaire (GraphPad Software Inc., 2016).

<sup>38</sup> Coefficient de corrélation de .10 : corrélation faible ; .30 : corrélation moyenne ; .50 : corrélation forte.

<sup>39</sup> Coefficient de corrélation de .00-.30 : corrélation très faible, voire nulle ; .31-.50 : corrélation faible ; .51-.70 : corrélation modérée ; .71-.90 : corrélation élevée ; .91-1.00 : corrélation très élevée.

## VII. Le score seuil optimal et la précision diagnostique

Afin d'investiguer la précision diagnostique de l'AVQI 03.01 appliqué à un échantillon en français, nous avons reproduit graphiquement le compromis entre sensibilité et spécificité<sup>40</sup> du test en fonction de différents scores seuils, en créant une courbe ROC (Receiver operating characteristic curve) (Delacour, Servonnet, Perrot, Vigezzi, & Ramirez, 2005). Pour cela, nous avons employé la macro mise à disposition par la société SAS (2014). La courbe résultante est illustrée à la figure 21. Sur la courbe ROC, le taux de vrais positifs (c.-à-d. sensibilité) est représenté en ordonnée, et le taux de faux positifs (c.-à-d. 1- spécificité) en abscisse. La droite diagonale représente la courbe « un test d'apport nul » (Delacour et al., 2005).

Comme précisé par Barsties et Maryn (2016), les voix ont été considérées comme normophoniques lorsque le critère G moyenné sur l'évaluation des quatre vocologues était < 0.5. Une voix était considérée comme dysphonique lorsque le G moyen était ≥ 0.5.

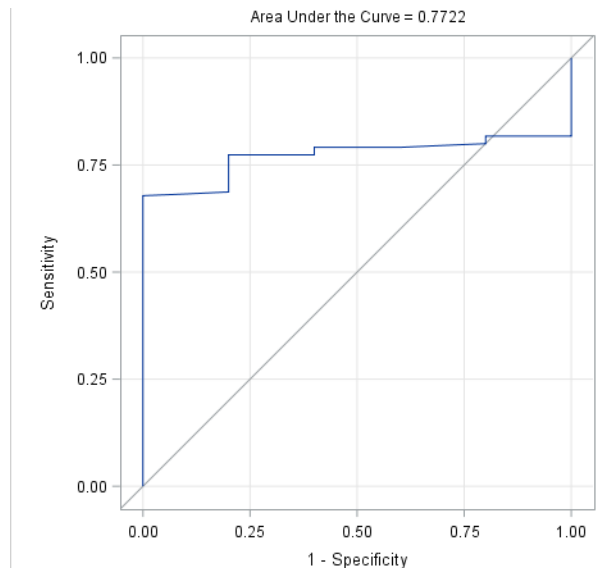


Figure 21. Courbe ROC selon les évaluations perceptives sur base de la parole continue et de la voyelle tenue.

Afin d'estimer la capacité de l'AVQI 03.01 à distinguer les voix normophoniques et dysphoniques, nous avons calculé l'aire sous la courbe ( $A_{ROC}$  : area under the ROC curve). Celle-ci était de .7722, indiquant un pouvoir discriminant suffisant (Simundić, 2009)<sup>41</sup>. Une aire de .7722 signifie que l'AVQI classe correctement des paires de participants pathologiques/non pathologiques aléatoires dans 77.22 % des cas (Hart, 2016). Comme l'intervalle de confiance de 95 % [0.6870 ; 0.8573] ne contient pas 0.5000 (test d'apport nul) (tableau 51), nous pouvons conclure que l'aire sous la

<sup>40</sup>  $sensibilité = \left( \frac{vrais\ positifs}{vrais\ positifs + faux\ négatifs} \right) \times 100$  ;  $spécificité = \left( \frac{vrais\ négatifs}{vrais\ négatifs + faux\ positifs} \right) \times 100$  (Meulemans, 2016)

<sup>41</sup> Un outil avec une ASC = 0.50 n'a aucune valeur discriminatoire ; une ASC situé entre 0.50 et 0.70 correspond à une précision faible ; une ASC situé entre 0.70 et 0.90 correspond à un test avec une précision « suffisante pour certains objectifs » ; un outil avec une ASC situé entre 0.90 et 1 a une bonne précision discriminative ; une ASC de 1 correspond à un test de précision parfaite (Swets, 1988).

courbe est significativement plus élevée que le facteur chance, et donc que le test (AVQI 03.01) présente une précision diagnostique acceptable. Le test de Chi-carré, évaluant si l'aire sous la courbe est significativement différente de 0.5000, est également significatif (Hart, 2016).

Tableau 51 Test de significativité de l'aire sous la courbe ROC selon les évaluations perceptives sur base de la parole continue et de la voyelle tenue

ROC Association Statistics							
ROC Model	Mann-Whitney				Somers' D	Gamma	Tau-a
	Area	Standard Error	95% Wald Confidence Limits				
Model	0.7722	0.0434	0.6870	0.8573	0.5443	0.5462	0.0438
ROC1	0.5000	0	0.5000	0.5000	0		0

ROC Contrast Test Results			
Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Model	1	39.2665	<.0001

Afin d'identifier le score seuil optimal, nous avons employé 3 indices (Pandey & Jain, 2016), illustrés sur la courbe ROC à la figure 23 :

- 1) **L'index de Youden** : il identifie sur la courbe ROC le résultat maximal de la formule  $sensibilité + spécificité - 1$ , et permet de sélectionner le score seuil optimal en donnant un poids maximal équivalant pour la sensibilité et pour la spécificité (Ruopp, Perkins, Whitcomb, & Schisterman, 2008). En d'autres mots, il s'agit aussi du point le plus éloigné verticalement de la droite diagonale de 45 ° (droite « chance ») sur la courbe ROC (figure 22).

Le score seuil optimal ainsi identifié dans notre étude, permettant la meilleure délimitation entre les voix normophoniques et dysphoniques selon l'index de Youden, est de 1.41. À ce score, la sensibilité, de 67.8 %, est acceptable, et la spécificité, de 100 %, est excellente. Cela signifie que 100 % des voix normophoniques seront détectées comme telles, mais que seulement 67.8 % des voix dysphoniques seront détectées comme pathologiques.

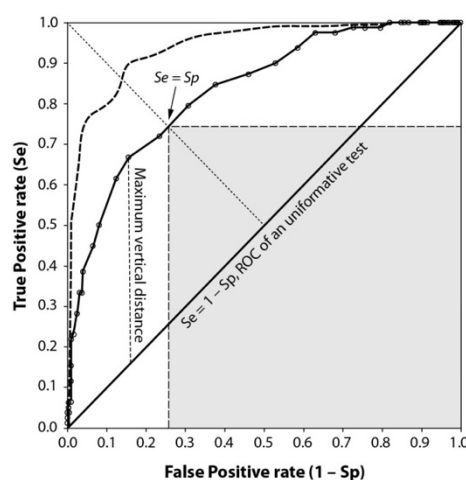


Figure 22. Index de Youden et différence minimale sensibilité-spécificité sur une courbe ROC (Habibzadeh, Habibzadeh, & Yadollahie, 2016).

« Maximum vertical distance » = index de Youden ; « Se = Sp » = différence minimale entre sensibilité et spécificité.

- 2) **La différence minimale sensibilité-spécificité** : ici, nous avons recherché le score qui permet une différence absolue minimale entre la spécificité et la sensibilité. Il s'agit du point où la courbe ROC coupe la diagonale partant du point supérieur gauche (figure 22). Ainsi, nous avons identifié le score seuil de 0.98. Celui-ci, par rapport au score seuil de 1.44, permet d'atteindre une spécificité moindre mais toujours satisfaisante, de 80 %, ainsi qu'une meilleure sensibilité, de 77.4 %.
- 3) **La distance minimale du point 0,1** : il s'agit ici du point le plus proche du point idéal sur la courbe ROC, qui est le point de sensibilité maximale (100 %) sur l'axe des ordonnées, et le point minimal sur l'axe des abscisses (1-spécificité = 0). C'est donc, sur la courbe ROC, le point situé en haut à gauche. Dans notre cas, le score se situant le plus près du point optimal est le même que celui permettant une différence absolue minimale entre la sensibilité et la spécificité, c'est-à-dire 0.98 (sensibilité = 77.4 %, spécificité = 80 %).

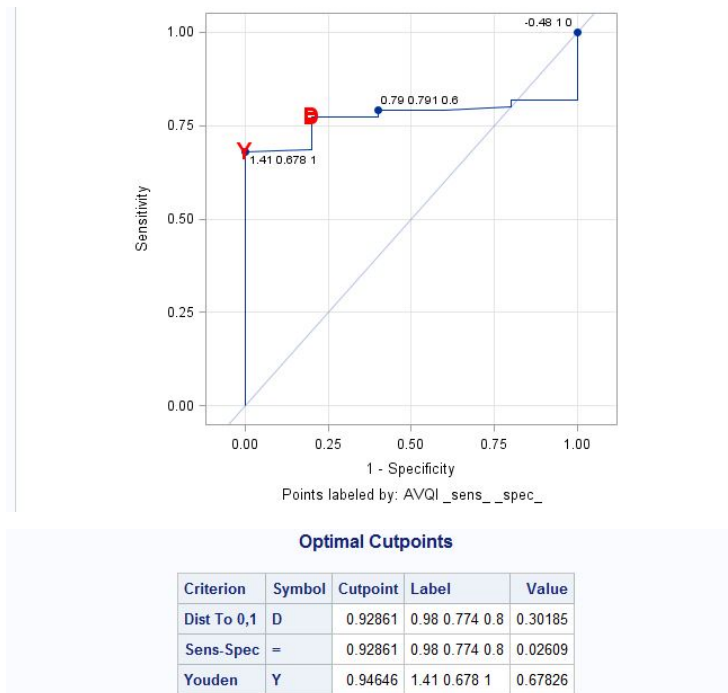


Figure 23. Les scores seuils optimaux selon trois méthodes sur la courbe ROC selon les évaluations perceptives sur base de la parole continue et de la voyelle tenue.

Dist To 0,1 = distance minimale du point 0,1 (D rouge sur le graphique); Sens-Spec = distance minimale sensibilité-spécificité (symbole « = » rouge sur le graphique, caché par le D); Youden : index de Youden (Y rouge sur le graphique; les trois nombres dans la colonne Label correspondent, dans l'ordre, au score AVQI, à la sensibilité et à la spécificité de l'outil à ce score seuil.

Nous avons remarqué que les effectifs des deux groupes « normophonique » et « dysphonique » étaient très différents. En effet, nous avons dans notre échantillon de 120 enregistrements 115 dont le critère G moyen est  $\geq 0.5$  (donc 96 %), et cinq dont le G moyen est  $< 0.5$  (donc 4 %). Dans ce cas, il n'est pas conseillé de baser le choix du score seuil uniquement sur la sensibilité et la spécificité, car ces deux indicateurs sont fortement dépendants des tailles des échantillons

et de la prévalence de patients dans l'échantillon (Dollaghan, 2007, cité par Barsties & Maryn, 2015). Ainsi, nous avons calculé les rapports de vraisemblance (« likelihood ratios »), qui prennent en compte à la fois la sensibilité et la spécificité, et dont l'utilisation est conseillée dans le cas d'effectifs fortement différents entre le groupe pathologique et non pathologique (Barsties & Maryn, 2015). Le « rapport de vraisemblance pour un résultat positif » (likelihood ratio for a positive result, LR+) permet d'estimer la probabilité qu'un individu soit dysphonique si le test est positif (Barsties & Maryn, 2015). Le LR+ correspond à  $\frac{\text{sensibilité}}{1-\text{spécificité}}$ . Le « rapport de vraisemblance pour un résultat négatif » (LR-) permet au contraire d'estimer la probabilité qu'un individu soit normophonique lorsque le test est négatif. Le LR- correspond à  $\frac{1-\text{sensibilité}}{\text{spécificité}}$ . La précision d'un outil peut être considérée comme élevée lorsque le LR+ est  $\geq 10$ , et que le LR- est  $\leq 0.1$  (Akobeng, 2006 ; Barsties & Maryn, 2015).

Si nous reprenons nos deux scores seuils identifiés ci-avant :

- 1) Pour le score seuil de 0.98, nous calculons un LR+ de 3.87 et un LR- de 0.28. Dans ces conditions, le clinicien n'est donc pas assuré qu'un patient obtenant un score pathologique ( $\geq 0.98$ ) soit effectivement dysphonique ; de même, si le score est non pathologique ( $< 0.98$ ), le clinicien ne peut pas garantir que le patient soit normophonique.
- 2) Pour le score seuil de 1.41 (identifié à l'aide de l'index de Youden), le LR+ tend vers l'infini, et le LR- est de 0.32. Dans ce cas, le clinicien peut être assuré qu'un patient obtenant un score  $\geq 1.41$  est dysphonique. Cependant, lorsque le score est  $< 1.41$ , il ne peut garantir que le patient soit normophonique.

### **VIII. Les participants normophoniques... dysphoniques ?**

Les résultats des jugements perceptifs du groupe contrôle nous ont mis la puce à l'oreille. En effet, sur nos 30 participants normophoniques, seuls trois ont obtenu un G moyen  $< 0.5$ . Cela signifie que pour les 27 autres, au moins deux des vocologistes ont estimé le G  $\geq 1$ .

Rappelons que chacune des quatre vocologistes avait rapporté des difficultés pour évaluer les voix en fonction des deux conditions phonatoires (parole continue ET voyelle soutenue), les jugements différant souvent fortement entre les deux. En effet, comme Maryn et Roy (2012) l'ont souligné, les jugements sont significativement plus sévères sur les voyelles tenues que sur la parole continue. La remarque selon laquelle cette évaluation était ainsi peu écologique ne pouvait donc pas être ignorée : en pratique clinique, le jugement subjectif à l'aide de la GRBAS n'est pas effectué sur une voyelle tenue, mais bien sur la parole continue, plus représentative de la voix du patient au quotidien.

Ainsi, nous avons demandé aux quatre vocologistes de réévaluer les 120 échantillons, cette fois-ci uniquement sur base des 27 syllabes de parole continue. Nous avons alors réitéré nos analyses de validation de l'AVQI en utilisant ces nouveaux jugements perceptifs, qui tout de suite — à vue d'œil — nous semblaient bien plus concordants.

## **IX. Réitération des analyses sur la parole continue uniquement**

### **A. La fidélité inter-juges et intra-juge**

Lors de ces nouvelles évaluations perceptives, la fidélité test-retest s'est avérée quelque peu plus élevée que lorsque l'évaluation était effectuée sur la concaténation des deux conditions phonatoires. En effet, le coefficient Kappa pondéré moyen est de .7783 (contre .727 précédemment), indiquant un accord significatif substantiel pour chacun des quatre juges (tableau 52).

Tableau 52 Coefficients Kappa pondérés pour l'évaluation de la fidélité intra-juge des jugements perceptifs sur la parole continue

Évaluateur	Coefficient de Kappa pondéré	p	Interprétation
1	.8089	< .0001	Accord presque excellent, significatif
2	.7220	< .0001	Accord substantiel, significatif
3	.7973	< .0001	Accord substantiel, significatif
4	.7851	< .0001	Accord substantiel, significatif

La fidélité inter-juges, évaluée à l'aide du coefficient de concordance de Kendall (Perinetti, 2017) est elle aussi très satisfaisante (tableau 53). Avec un coefficient de concordance de Kendall de .89517 et une probabilité de dépassement de  $p < .0001$ , nous pouvons rejeter l'hypothèse nulle selon laquelle l'accord entre les juges est dû au facteur chance. Selon les recommandations d'interprétation de Schmidt (1997, cité par Remøy, 2010) décrites dans le tableau 49 (page 82), l'accord inter-juges est « exceptionnellement élevé » (.90).

Tableau 53 Coefficient de concordance de Kendall pour l'évaluation de la fidélité inter-juges des jugements perceptifs sur la parole continue

Coeff of Concordance	F	Num DF	Denom DF	Prob>F
0.89517	25.62	118.5	355.5	<.0001

### **B. La validité concurrente**

Nous avons à nouveau mesuré la corrélation de Spearman entre les scores bruts calculés par l'AVQI et les degrés de sévérité moyens (critère G de l'échelle GRBAS) tels qu'évalués par les quatre vocologistes sur base de la parole continue. Les résultats sont représentés dans le tableau 54. Avec un coefficient de corrélation de Spearman de .84 et  $p < .0001$ , nous pouvons rejeter l'hypothèse nulle selon laquelle il n'y a pas de relation entre les scores de l'AVQI et les jugements perceptifs du degré de sévérité de la dysphonie (G de la GRBAS) sur base de la parole continue.

La corrélation entre les scores AVQI et les jugements perceptifs par les quatre vocologues est forte ( $> .50$ ) selon les critères d'interprétation de Cohen (1988) ; elle est également élevée (.71-.90) selon les critères plus sévères de Hinkle, Wiersma et Jurs (1994).

Tableau 54 *Corrélation de Spearman entre les scores AVQI et les évaluations moyennes du degré de sévérité (critère G) évalué par les quatre vocologues sur base de la parole continue*

2 Variables: avqi gmean						
Simple Statistics						
Variable	N	Mean	Std Dev	Median	Minimum	Maximum
avqi	120	3.46933	2.91240	2.69500	-0.48000	10.02000
gmean	120	1.52500	1.01170	1.25000	0	3.00000

Spearman Correlation Coefficients, N = 120 Prob >  r  under H0: Rho=0		
	avqi	gmean
avqi	1.00000	0.83546 <.0001
gmean	0.83546 <.0001	1.00000

Nous pouvons donc conclure à une bonne validité concourante de l'AVQI 03.01 appliqué sur un échantillon en langue française, lorsque mis en lien avec l'évaluation perceptuelle sur base de la parole continue.

### C. Le score seuil optimal et la précision diagnostique

Nous avons également reproduit graphiquement le compromis entre sensibilité et spécificité du test en fonction de différents scores seuils, en créant une courbe ROC (Receiver operating characteristic curve). Comme précédemment, les voix ont été considérées comme normophoniques lorsque le critère G moyenné sur l'évaluation des quatre vocologues était  $< 0.5$ . Une voix était considérée comme dysphonique lorsque le G moyen était  $\geq 0.5$  (Barsties & Maryn, 2016).

La courbe résultante est illustrée à la figure 24.

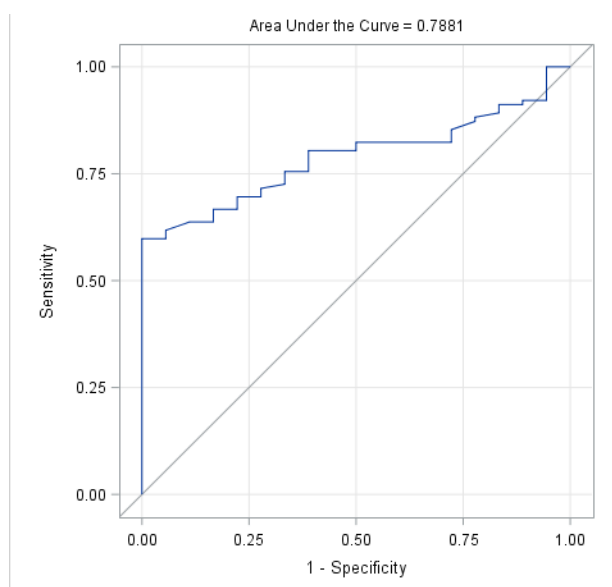


Figure 24. Courbe ROC selon les évaluations perceptuelles sur base de la parole continue.



Afin d'estimer la capacité de l'AVQI 03.01 à distinguer les voix normophoniques et dysphoniques, nous avons calculé l'aire sous la courbe ( $A_{ROC}$  : area under the ROC curve). Celle-ci était de .7881, indiquant un pouvoir discriminant suffisant (Simundić, 2009). Une aire de 0.7881 signifie que l'AVQI classe correctement des paires de participants pathologiques/non pathologiques dans 78.81 % des cas (Hart, 2016). Comme l'intervalle de confiance de 95 % [0.7054 ; 0.8699] ne contient pas 0.5000 (tableau 55), nous pouvons conclure que l'aire sous la courbe est significativement plus élevée que le facteur chance, et que l'AVQI 03.01 présente une précision diagnostique satisfaisante. Le test de Chi-carré, évaluant si l'aire sous la courbe est significativement différente de 0.5000, est également significatif (Hart, 2016).

Tableau 55 Test de significativité de l'aire sous la courbe ROC selon les évaluations perceptives sur base de la parole continue

ROC Association Statistics							
ROC Model	Mann-Whitney			Somers' D	Gamma	Tau-a	
	Area	Standard Error	95% Wald Confidence Limits				
Model	0.7881	0.0417	0.7064	0.8699	0.5763	0.5781	0.1482
ROC1	0.5000	0	0.5000	0.5000	0	.	0

ROC Contrast Test Results			
Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Model	1	47.7192	<.0001

Afin d'identifier le score seuil optimal, nous avons employé les mêmes indices que précédemment (figure 25) :

- 1) **L'index de Youden** : le score seuil optimal identifié, permettant la meilleure délimitation entre les voix normophoniques et dysphoniques selon l'index de Youden, est de 2.33. À ce score, la sensibilité, de 59.8 %, est acceptable, et la spécificité, de 100 %, est excellente. Cela signifie que 100 % des voix normophoniques seront détectées comme telles, mais que seulement 59.8 % des voix dysphoniques seront détectées comme pathologiques.
- 2) **La différence minimale sensibilité-spécificité** : le score seuil identifié à l'aide de cette méthode est de 1.41. Celui-ci, par rapport au score seuil de 2.33, permet d'atteindre une spécificité moindre mais toujours satisfaisante, de 72.2 %, ainsi qu'une meilleure sensibilité, de 71.6 %.
- 3) **La distance minimale du point 0,1** : le score se situant le plus près du point optimal est de 1.79, avec une sensibilité de 66.7 % et une spécificité de 83.3 %.

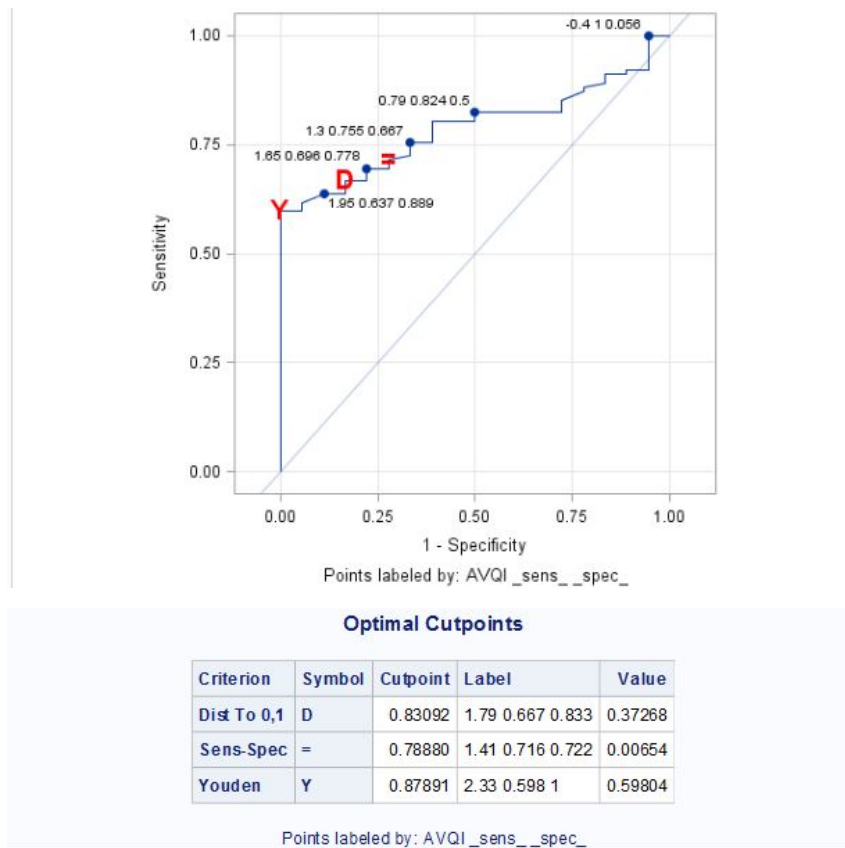


Figure 25. Les scores seuils optimaux selon trois méthodes sur la courbe ROC selon les évaluations perceptives sur base de la parole continue.

Dist To 0,1 = distance minimale du point 0,1 (D rouge sur le graphique); Sens-Spec = distance minimale sensibilité-spécificité (symbole « = » rouge sur le graphique); Youden : index de Youden (Y rouge sur le graphique); les trois nombres dans la colonne Label correspondent, dans l'ordre, au score AVQI, à la sensibilité et à la spécificité de l'outil à ce score seuil.

Dans notre échantillon de 120 enregistrements, 102 ont obtenu un critère G moyen  $\geq 0.5$  (donc 85 %), et 18 un G moyen de  $< 0.5$  (donc 15 %). À nouveau, il n'est donc pas conseillé de baser le choix du score seuil sur la sensibilité et la spécificité, de par leur dépendance des tailles des échantillons et de la prévalence de patients dans l'échantillon (Dollaghan, 2007, cité par Barsties & Maryn, 2015). Ainsi, nous avons calculé les rapports de vraisemblance (« likelihood ratios »). Pour rappel, la précision d'un outil peut être considérée comme élevée lorsque le LR+ est  $\geq 10$ , et que le LR- est  $\leq 0.1$  (Akobeng, 2006 ; Barsties & Maryn, 2015).

Si nous reprenons nos trois scores seuils identifiés ci-avant :

- 1) Pour le score seuil de 2.33 (identifié à l'aide de l'index de Youden), nous calculons un LR+ dont la valeur tend vers l'infini, et un LR- de 0.4. Dans ces conditions, le clinicien peut donc avec assurance conclure qu'un patient obtenant un score pathologique ( $\geq 2.33$ ) est effectivement dysphonique ; cependant, si le score est non pathologique ( $< 2.33$ ), le clinicien ne peut pas garantir que le patient soit normophonique.

- 2) Pour le score seuil de 1.41, le LR+ est de 2.58, et le LR- de 0.39. Dans ce cas, le clinicien ne peut pas assurer qu'un patient obtenant un score  $\geq 1.41$  soit dysphonique. De même, lorsque le score est  $< 1.41$ , il ne peut pas garantir que le patient soit normophonique.
- 3) Pour le score seuil de 1.79, le LR+ est de 3.99, et le LR- de 0.4. Encore une fois, le clinicien ne peut pas assurer qu'un patient obtenant un score  $\geq 1.79$  soit dysphonique. De même, lorsque le score est  $< 1.79$ , il ne peut pas garantir que le patient soit normophonique.

Ainsi, l'index de Youden nous permet d'obtenir le score seuil optimal, de **2.33**, en tenant compte des différences des effectifs dans les deux groupes « normophonique » et « dysphonique ». Ce score permet cependant uniquement au clinicien de garantir qu'un patient ayant un score pathologique ( $\geq 2.33$ ) est effectivement dysphonique.

Notons que dans leur étude de validation de l'AVQI 03.01 (Barsties & Maryn, 2016), les auteurs avaient eux aussi obtenu un LR- supérieur à 0.1 (LR- = 0.23) pour le score-seuil optimal en néerlandais, de 2.43. La même prudence doit donc être observée quant à la conclusion tirée en présence d'un score non pathologique. Appliqué à nos données, ce score seuil de 2.43 (très proche de notre score seuil optimal) permet d'obtenir une sensibilité de 58.82 % et une spécificité de 100 % (les calculs sont représentés dans le tableau 56), avec un LR+ tendant vers l'infini, et un LR- de 0.41.

Tableau 56 Calcul de la sensibilité et de la spécificité de l'AVQI 03.01 en fonction du score seuil

score seuil 2.33		G <sub>moyen</sub>	
		patho	non patho
AVQI <sub>03.01</sub>	patho	61	0
	non patho	41	18
Sensibilité = vrais positifs / vrais positifs + faux négatifs = 61/102 = <b>59.8 %</b>			
Spécificité = vrais négatifs / vrais négatifs + faux positifs = 18/18 = 100 %			
score seuil 2.43		G <sub>moyen</sub>	
		patho	non patho
AVQI <sub>03.01</sub>	patho	60	0
	non patho	42	18
Sensibilité = vrais positifs / vrais positifs + faux négatifs = 60/102 = <b>58.82 %</b>			
Spécificité = vrais négatifs / vrais négatifs + faux positifs = 18/18 = 100 %			

## **X. SYNTHÈSE DES RÉSULTATS**

### **A. Jugement sur la parole continue et la voyelle tenue**

#### **La fidélité inter-juges et intra-juge :**

- La fidélité test-retest est satisfaisante pour les quatre évaluateurs ( $\kappa = .67, .73, .78$  et  $.73$  ;  $p < .0001$ ).
- La fidélité inter-juges est également élevée ( $W = .88, p < .0001$ ).

#### **La validité concourante :**

Avec une corrélation de Spearman élevée ( $r_s = .85, p < .0001$ ) entre les scores AVQI 03.01 et les jugements perceptifs du degré de sévérité de la dysphonie (G de la GRBAS), nous pouvons conclure que l'AVQI 03.01 est un outil valide pour l'évaluation de la qualité vocale en français.

#### **Le score seuil optimal et la précision diagnostique :**

L'aire sous la courbe est de  $.7722$  ; le score-seuil optimal (identifié à l'aide de l'index de Youden) est de  $1.41$ , avec une sensibilité de  $67.8\%$ , une spécificité de  $100\%$ , un  $LR+ = \infty$  et un  $LR- = 0.32$ .

### **B. Jugement sur la parole continue uniquement**

#### **La fidélité inter-juges et intra-juge :**

- La fidélité test-retest est satisfaisante pour les quatre évaluateurs ( $\kappa = .81, .72, .80$  et  $.79$  ;  $p < .0001$ ).
- La fidélité inter-juges est « exceptionnellement élevée » ( $W = .90, p < .0001$ ).

#### **La validité concourante :**

Avec une corrélation de Spearman élevée ( $r_s = .84, p < .0001$ ) entre les scores AVQI 03.01 et les jugements perceptifs du degré de sévérité de la dysphonie (G de la GRBAS), nous pouvons conclure que l'AVQI 03.01 est un outil valide pour l'évaluation de la qualité vocale en français.

#### **Le score seuil optimal et la précision diagnostique :**

L'aire sous la courbe est de  $.7881$  ; le score-seuil optimal (identifié à l'aide de l'index de Youden) est de **2.33**, avec une sensibilité de  $59.8\%$ , une spécificité de  $100\%$ , un  $LR+ = \infty$  et un  $LR- = 0.4$ .

## L'AVQI et l'évaluation de la qualité de vie par le patient

Pour la dernière partie de ce travail, nous nous sommes interrogés quant à l'existence d'une corrélation entre le score à l'AVQI 03.01 calculé sur un échantillon en français et l'évaluation par le sujet dysphonique de sa qualité de vie liée à la pathologie vocale. En effet, comme nous l'avons vu, de nombreuses études d'investigent le lien entre les mesures objectives et les mesures de qualité de vie. L'AVQI s'est avéré concordant par rapport à l'évaluation subjective par le clinicien (ainsi le démontre la corrélation entre le score AVQI et le critère G de l'échelle GRBAS). Cependant, aucune étude concernant la question de la corrélation du score à l'AVQI avec l'évaluation subjective par le patient n'a encore été menée. Des études répondant à la question du lien entre l'évaluation objective acoustique et l'évaluation subjective par le patient ont pourtant été menées entre autres pour le DSI (Hauptmann, 2013). L'absence d'études répondant à cette interrogation pour le cas de l'AVQI a été confirmée par Youri Maryn lors du symposium organisé par l'Université de Liège le 21 octobre 2016, en l'honneur du Professeur Johan Sundberg.

Pourtant, la tendance actuelle — autant dans le domaine médical que paramédical — tend à privilégier des soins centrés sur le patient, avec une implication de plus en plus importante de ce dernier notamment dans les prises de décisions cliniques (DiLollo & Favreau, 2010).

### **I. L'EBP et la CIF**

L'inclusion du point de vue du patient dans l'évaluation est primordiale dans la démarche de l'Evidence Based Practice (EBP, figure 26). L'EBP est un concept qui sert actuellement de ligne conductrice dans beaucoup de pays, autant pour l'évaluation que pour la prise en charge de nombreuses pathologies. Cette démarche prend en compte à la fois les données scientifiques (recherche, documentation), mais aussi l'expérience clinique du thérapeute et les besoins, souhaits et préférences du patient. À ce modèle, Satterfield a ajouté le contexte organisationnel et l'environnement du patient, qui influent également sur la prise de décision clinique (Satterfield, 2009).

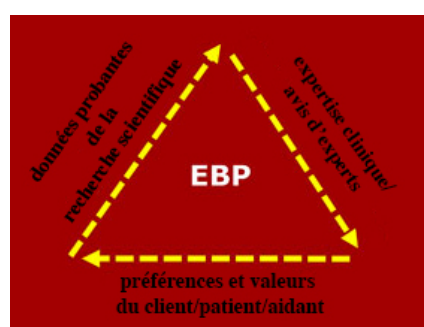


Figure 26. Les trois composantes principales de l'EBP (traduit de American Speech-Language-Hearing Association, 2017b).

La Classification internationale du fonctionnement, du handicap et de la santé (CIF) de l'Organisation mondiale de la Santé (OMS, 2001) rejoint cette perspective EBP. La CIF a été créée afin de permettre une communication plus efficace entre les différents intervenants, notamment dans les équipes pluridisciplinaires. Il s'agit d'une classification qui peut s'appliquer à toute pathologie, en utilisant un codage précis pour chaque composante. La CIF se compose de quatre éléments :

- 1) Les structures et les fonctions anatomiques (figure 27) : dans le cas de la voix, par exemple, il peut s'agir de l'état des structures laryngées, des plis vocaux...

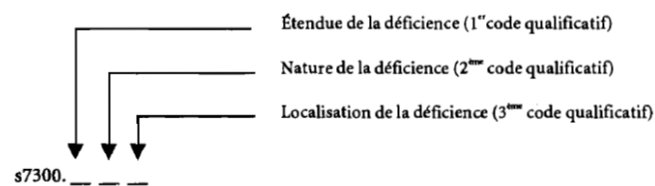


Figure 27. Codage de la composante « structures et fonctions anatomiques » de la CIF (Organisation mondiale de la Santé, 2001).

- 2) Les activités et la participation de l'individu dans les situations de vie quotidienne (figure 28) : par exemple le métier de l'individu (enseignant, animateur radio...) ou ses passe-temps (chant, mouvements de jeunesse...).

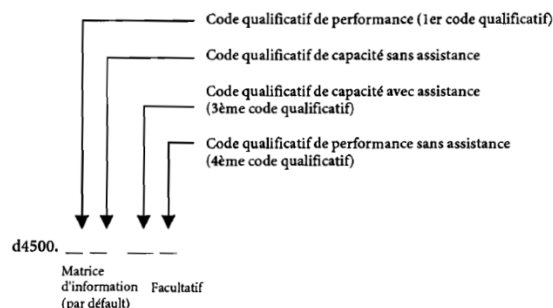


Figure 28. Codage de la composante « activités et participation » de la CIF (Organisation mondiale de la Santé, 2001).

- 3) Les facteurs environnementaux : appliqués à la voix, ces facteurs peuvent inclure des facteurs du milieu (telle la climatisation, un milieu familial bruyant...), mais aussi et surtout les réactions des personnes de l'entourage de l'individu par rapport à sa dysphonie, les croyances, l'influence culturelle... ;
- 4) Les facteurs personnels : les caractéristiques physiques et sociales de l'individu, son âge, son genre... mais aussi son attitude par rapport à sa dysphonie, ses émotions.

Imaginons une personne dysphonique, de genre masculin, présentant des lésions nodulaires des plis vocaux. Cette personne se plaint d'une qualité vocale fortement altérée et d'une gêne sociale, notamment dans les conversations avec ses proches. Elle a d'ailleurs été licenciée de son poste dans un centre d'appels. Heureusement, elle bénéficie d'un suivi intensif par un logopède.

Des exemples de codes pourraient être les suivants :

- s3400.274 : changement qualitatif (7), modérée (2), à la partie antérieure (4) des plis vocaux (s3400)
- b3101.4 : altération sévère (4) de la qualité vocale (b3101)
- d3503.3\_ : difficulté grave dans la capacité (3) à faire la conversation avec une autre personne (d3503)
- d8451.4\_ : difficulté absolue (4) dans la capacité à garder un emploi (d8451)
- e2600.4 : très mauvaise (4) qualité de l'air à l'intérieur (e2600)
- e430.4 : l'attitude de la personne en position d'autorité (e430) représente un obstacle absolu (4)
- e355+3 : le professionnel de la santé (e355) représente un élément facilitateur substantiel (3)

La CIF, en intégrant dans son modèle les facteurs personnels de l'individu, souligne l'importance du point de vue de la personne dysphonique dans l'évaluation et la prise en charge.

## **II. L'AVQI et le VHI**

Comme nous l'avons décrit ci-avant, les études investiguant la corrélation entre les mesures acoustiques et les outils d'évaluation subjective, autant par le patient que par le clinicien, montrent des résultats contradictoires. Il est donc vivement conseillé de prendre en compte les données objectives et subjectives de façon concomitante.

Dans leur étude, Behrman, Sulica, et He (2004) n'ont pas trouvé de corrélation significative entre le VHI et l'évaluation perceptive de la voix par le thérapeute à l'aide de l'échelle CAPE-V, ni entre le VHI et la qualité de la vibration cordale, la fermeture glottique, l'âge et le genre. Selon l'INSERM, il n'y aurait donc « pas de lésion bénigne plus gênante qu'une autre ; les critères médicaux ne reflètent pas le degré de handicap de vie. » (Autesserre et al., 2006)

Ainsi, ils concluent que la perception subjective de la sévérité de la dysphonie par le patient est une variable relativement indépendante, qui doit être prise en compte en tant que telle dans l'évaluation vocale et dans la planification de la prise en charge (Behrman, Sulica, & He, 2004).

Ces données indiquent donc que la perception du handicap vocal n'est pas directement en lien avec les autres mesures (subjectives et objectives) de la voix. Citons encore l'INSERM, qui exprime très bien la réalité complexe et les conséquences variées de la dysphonie : « L'impact d'un trouble de la voix est plus qu'une pathologie du larynx ou un trouble acoustique, car il retentit sur la vie professionnelle et sociale d'un individu, retentit sur le mental, le physique, l'émotionnel et la communication, et il a des conséquences sur l'emploi, le social et la santé » (Autesserre et al., 2006).

L'objectif final de cette étude était donc d'investiguer une éventuelle corrélation entre le score calculé par l'AVQI 03.01 et les scores au VHI complété par le patient. Suite aux nombreuses lectures effectuées, nous avons conclu que le VHI et le VoiSS semblaient être les outils les plus adaptés pour l'évaluation de l'impact de la dysphonie (sensibilité et spécificité maximales) (Behlau et al., 2016). Parmi ces deux outils, c'est le VoiSS qui paraît être le plus satisfaisant. En effet, Lopes et al. (2017) décrivent ses avantages, déjà mentionnés plus haut : premièrement, les items du VoiSS proviennent de la description des symptômes vocaux par les patients dysphoniques, et deuxièmement les trois facteurs du VoiSS ont été déterminés statistiquement à l'aide d'analyses factorielles. Ces auteurs décrivent le VoiSS comme « le protocole le plus fidèle et le plus robuste psychométriquement parlant pour l'auto-évaluation vocale » (Lopes et al., 2017). Le VoiSS, généré avec beaucoup de rigueur (Branski et al., 2010), nous semblait donc optimal pour notre étude. Malheureusement, il s'est avéré que celui-ci n'est pas disponible dans la banque de données du CHU, contrairement au VHI. C'est donc ce dernier qui sera utilisé dans notre travail.

#### A. Notre questionnaire

Pour rappel, le VHI (Jacobson et al., 1997) est une liste de 30 affirmations étudiant l'impact de la qualité vocale sur le quotidien du patient, et ce à trois niveaux : fonctionnel (impact sur les activités quotidiennes), émotionnel (ressenti, perception de la pathologie vocale) et physique (caractéristiques vocales et inconfort). Le patient évalue chaque item sur une échelle de 0 (jamais ; pas d'impact) à 4 (toujours ; handicap important). L'on obtient donc quatre scores : un score global du degré de handicap perçu (sur un total de 120), et les trois sous-scores physique, émotionnel et fonctionnel (chacun sur un total de 40).

Nous avons souhaité tout d'abord mesurer la corrélation entre le score calculé par l'AVQI 03.01 sur l'échantillon de 27 syllabes du texte de Harmegnies, et le score global au VHI. Ensuite, nous souhaitions connaître, plus précisément, les liens entre le score AVQI 03.01 et les scores aux trois sous-échelles du VHI.

#### B. La population

L'échantillon analysé comporte les scores de 78 individus, dont treize pour chaque degré de sévérité selon le score AVQI 03.01 (voix non pathologique, dysphonie légère à modérée, dysphonie sévère), pour chaque genre. Cet échantillon est décrit dans le tableau 57.



Tableau 57 Caractéristiques principales de l'échantillon de 78 patients

Nombre total	78		
Genre	Femme : 39 (50 %) Homme : 39 (50 %)		
Âge	Moyenne <sup>a</sup> = 52.85 Écart-type = 16.82 Min = 22.72 Max = 88.34		
Pathologies (n et %) <sup>b</sup>	– Immobilité laryngée	22	28.21 %
	– Consultation sans pathologie <sup>c</sup>	10	12.82 %
	– Dysphonie fonctionnelle <sup>d</sup>	9	11.54 %
	– Presbyphonie	7	8.74 %
	– Post-phonochirurgie	6	7.69 %
	– Œdème cordal	6	7.69 %
	– Lésion nodulaire	5	6.41 %
	– Kyste	3	3.85 %
	– Polype	3	3.85 %
	– Sulcus/vergeture	2	2.56 %
	– Lésion tumorale	2	2.56 %
	– pathologie non renseignée	1	1.28 %
	– Dysphonie spasmodique	1	1.28 %
	– Cordectomie	1	1.28 %
Professions (n et %)	– Pensionnés/retraités	21	26.92 %
	– Professionnels de la formation initiale et de la formation continue (enseignants)	14	17.95 %
	– Professionnels de la santé (professions médicales et paramédicales)	9	11.54 %
	– Personnel des services administratifs et commerciaux (employé de bureau, comptable, secrétaire médical, agent administratif...)	6	7.69 %
	– Personnel, techniciens et cadres du transport et de la logistique (chauffeur routier, chauffeur de taxi, magasinier, agent distributeur...)	6	7.69 %
	– Cadres administratifs et professionnels de l'information et de la communication (PDG, juriste, avocat, assistant parlementaire...)	5	6.41 %
	– Personnel des services aux personnes et à la collectivité (technicienne de surface, nettoyage industriel, militaire, coiffeur...)	3	3.85 %
	– Étudiants	2	2.56 %
	– Chômage, invalidité	2	2.56 %
	– Professionnels des arts et du spectacle (chanteur, photographe...)	2	2.56 %
	– Personnel de type artisanal (ébéniste, boulanger...)	2	2.56 %
	– Agents de maîtrise, techniciens et ingénieurs du bâtiment, des travaux publics et de l'extraction (architecte, métreur-deviseur)	2	2.56 %
	– Personnel de la distribution et de la vente (commerçant, délégué commercial...)	1	1.28 %
	– Personnel des industries graphiques (copiste)	1	1.28 %
	– Techniciens industriels (technicien en mécanique)	1	1.28 %
	– Professionnels de l'intervention sociale, du développement local et de l'emploi (assistant social)	1	1.28 %

Note. <sup>a</sup> La distribution des âges dans notre échantillon suit la loi gaussienne, tel que démontré par le test de Shapiro-Wilk ( $W = .974$ ,  $p = .103$ ). <sup>b</sup> Les diagnostics laryngologiques ont été effectués à l'aide d'un vidéo-laryngo-stroboscope EndoSTROB (Xion GmbH, Berlin). Nous tenons à souligner que dans notre base de données, les dysphonies liées au cancer n'incluent que les patients de stade T1N0M0. <sup>c</sup> Sont ici inclus les participants ayant consulté, mais ne présentant pas de pathologie avérée, y compris entre autres les bilans préopératoires dans le cadre d'interventions chirurgicales touchant la glande thyroïde, d'interventions chirurgicales ou toute autre intervention au cours de laquelle peut avoir lieu une lésion des nerfs laryngés. <sup>d</sup> Les dysphonies dites « psychogènes » sont incluses dans cette catégorie. <sup>e</sup> Nous avons choisi de regrouper les professions selon l'arborescence du service public wallon de l'emploi et de la formation (Forem), accessible via le lien suivant : <https://www.leforem.be/former/horizonemploi/metier-profession-index.html> (Service public wallon de l'emploi et de la formation, 2016).

### C. Les corrélations AVQI – VHI

Comme nous l'avons déjà observé, les scores bruts de l'AVQI 03.01 calculés sur 27 syllabes ne sont pas normalement distribués. La corrélation de Pearson ne peut donc être employée, ce qui nous a conduit à utiliser la corrélation de Spearman ( $r_s$ ).

Nous avons tout d'abord mesuré la corrélation entre le score calculé par l'AVQI 03.01 et le score global au VHI (sur un total de 120). Puis, nous avons calculé les corrélations entre le score AVQI 03.01 et les scores aux trois sous-échelles du VHI, à savoir les scores « physique », « fonctionnel » et « émotionnel » (sur un total de 40, respectivement). Pour toutes ces corrélations, nous nous sommes également interrogés quant à l'influence de l'âge du patient : un patient âgé de 80 ans ne ressentira probablement pas le même handicap lié à sa qualité vocale qu'un patient âgé de 25 ans<sup>42</sup>. Cette potentielle différence de perception du handicap pourrait jouer sur la corrélation entre le score AVQI03.01 et l'auto-évaluation par le patient. Pour cette raison, nous avons également effectué des corrélations partielles de Spearman en contrôlant l'effet de l'âge. Voici donc les résultats des analyses effectuées :

#### 1) VHI global – score AVQI 03.01

Le coefficient de corrélation de Spearman calculé entre le score global au VHI et le score AVQI 03.01 est de .584, indiquant une corrélation positive entre les deux scores ; cette corrélation est de plus significative, avec  $p < .0001$ . Lorsque nous effectuons une corrélation partielle en contrôlant l'effet de l'âge, le coefficient de corrélation est plus élevé :  $r_s = .62$ ,  $p < .0001$ .

#### 2) VHI fonctionnel – score AVQI 03.01

Le coefficient  $r_s$  calculé entre le score aux items « fonctionnels » du VHI et le score AVQI 03.01 est de .618, indiquant une corrélation positive entre les deux scores ; cette corrélation est de plus significative, avec  $p < .0001$ . Lorsque nous effectuons une corrélation partielle en contrôlant l'effet de l'âge, le coefficient de corrélation est plus élevé :  $r_s = .643$ ,  $p < .0001$ .

#### 3) VHI émotionnel – score AVQI 03.01

Le coefficient  $r_s$  calculé entre le score aux items « émotionnels » du VHI et le score AVQI 03.01 est de .520 indiquant une corrélation positive entre les deux scores ; cette corrélation est de plus significative, avec  $p < .0001$ . Lorsque nous effectuons une corrélation partielle en contrôlant l'effet de l'âge, le coefficient de corrélation est plus élevé :  $r_s = .543$ ,  $p < .0001$ .

---

<sup>42</sup> Notons que l'âge ne doit pas être considéré comme un facteur isolé. En effet, il doit être vu dans un cadre plus large, incluant notamment le contexte professionnel et social dans lequel évolue l'individu, mais aussi les aspects psychologiques. Mentionnons à ce sujet le « paradoxe du bien-être », selon lequel le bonheur augmenterait avec l'âge malgré l'accumulation d'évènements stressants et/ou négatifs au courant de la vie.

#### 4) VHI physique – score AVQI 03.01

Le coefficient  $r_s$  calculé entre le score aux items « physiques » du VHI et le score AVQI 03.01 est de .458, indiquant une corrélation positive entre les deux scores ; cette corrélation est de plus significative, avec  $p < .0001$ . Lorsque nous effectuons une corrélation partielle en contrôlant l'effet de l'âge, le coefficient de corrélation est plus élevé :  $r_s = .514$ ,  $p < .0001$ .

#### 5) Comparaison des coefficients de corrélation

Nous avons ensuite comparé les coefficients de corrélation des trois sous-échelles (en contrôlant l'effet de l'âge), afin de mesurer s'ils différaient significativement entre eux. Pour cela, nous les avons transformés en  $z$  « de Fisher à l'aide de la formule suivante :

$$z' = 0.5[\ln(1+r) - \ln(1-r)] \text{ (Andale, 2016)}$$

Nous obtenons ainsi les  $Z$  suivants :

- $Z_{AVQI-VHI\text{fonctionnel}}$  : 0.763
- $Z_{AVQI-VHI\text{émotionnel}}$  : 0.608
- $Z_{AVQI-VHI\text{physique}}$  : 0.568

Afin de comparer ces  $Z$  de Fisher, nous avons utilisé la formule de Kenny (1987) :

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

Voici les  $Z$  ainsi calculés pour les trois comparaisons :

- comparaison entre le  $Z_{AVQI-VHI\text{fonctionnel}}$  et le  $Z_{AVQI-VHI\text{émotionnel}}$  :  $Z = 0.666$
- comparaison entre le  $Z_{AVQI-VHI\text{fonctionnel}}$  et le  $Z_{AVQI-VHI\text{physique}}$  :  $Z = 0.839$
- comparaison entre le  $Z_{AVQI-VHI\text{émotionnel}}$  et le  $Z_{AVQI-VHI\text{physique}}$  :  $Z = 0.173$

Pour  $\alpha = .05$ , si  $Z > 1.96$  ou  $Z < -1.96$ , nous pouvons rejeter l'hypothèse nulle selon laquelle les deux coefficients de corrélation comparés sont statistiquement équivalents (Warner, 2013). Dans notre cas, les  $Z$  calculés pour les trois comparaisons sont tous compris dans l'intervalle  $[-1.96 ; 1.96]$ , ce qui nous conduit à tolérer l'hypothèse nulle d'égalité des coefficients de corrélation. Nous n'avons donc pas suffisamment de preuves pour conclure que les corrélations diffèrent.

Tous ces résultats nous permettent de conclure que le score calculé par l'AVQI 03.01 est en lien avec l'auto-évaluation de l'impact de la dysphonie sur le quotidien du patient. Les corrélations entre le score AVQI 03.01 et les scores aux sous-échelles du VHI sont statistiquement homogènes. Si nous considérons les critères d'interprétation de Cohen (1988), toutes les corrélations mesurées peuvent être considérées comme étant fortes ( $> .50$ ). Cependant, si nous considérons les critères plus sévères de Hinkle, Wiersma et Jurs (1994) elles doivent toutes être considérées comme étant modérées ( $.51 - .70$ ) ; aucun coefficient de corrélation ne signe alors une relation forte entre l'AVQI 03.01 et les scores au VHI.

6) « Normophonie/dysphonie » : AVQI 03.01 – VHI

Nous avons finalement souhaité savoir si le jugement « normophonie-dysphonie » des deux outils concordait. Pour cela, nous avons mesuré le coefficient phi de Cramer, destiné à mesurer le degré de la liaison entre deux variables dichotomiques. Nos deux variables dichotomiques ont été définies comme représenté dans le tableau 58 :

Tableau 58 Attribution des valeurs 0 ou 1 en fonction du score AVQI 03.01 et du score VHI

	AVQI 03.01	VHI
Valeur 0 (« normophonie »)	Si score < 2.33	Si score < 20/120
Valeur 1 (« dysphonie »)	Si score ≥ 2.33	Si score ≥ 20/120 <sup>43</sup>

La table de contingence est représentée dans le tableau 59. Avec un coefficient phi ( $\phi$ ) = .52, et  $p = .000$ , nous pouvons conclure que les verdicts « pathologique » ou « non pathologique » sont concordants entre les deux outils d'évaluation.

Selon les lignes directrices d'interprétation de Cohen (1988) (tableau 60), la relation entre les deux verdicts est forte. Selon les lignes directrices de Rea et Parker (1992) (tableau 60), la relation est « relativement élevée ».

Tableau 59 Table de contingence pour l'analyse de l'association entre verdict AVQI et verdict VHI

Frequency Percent Row Pct Col Pct	Table of verdictAVQI by verdictVHI		
	verdictAVQI	verdictVHI	
		0	1
0	14 17.95 53.85 77.78	12 15.38 46.15 20.00	26 33.33
1	4 5.13 7.69 22.22	48 61.54 92.31 80.00	52 66.67
Total	18 23.08	60 76.92	78 100.00

Note. verdictAVQI = attribution de la valeur 0 ou 1 en fonction du score AVQI 03.01 ; verdictVHI = attribution de la valeur 0 ou 1 en fonction du score total au VHI.

Tableau 60 Lignes directrices pour l'interprétation du coefficient Phi de Cramer selon Cohen (1988) et Rea & Parker (1992) (Kotrlík, Williams & Katha Jabor, 2011)

Reference	Effect size statistic	Values	Interpretation of effect size
Cohen, 1988	Cramer's <i>Phi</i> or Cramer's <i>V</i> for nominal data	.10	Small effect size
		.30	Medium effect size
		.50	Large effect size
Rea & Parker, 1992	Cramer's <i>V</i> for nominal data	.00 and under .10	Negligible association
		.10 and under .20	Weak association
		.20 and under .40	Moderate association
		.40 and under .60	Relatively strong association
		.60 and under .80	Strong association
	.80 and under 1.00	Very strong association	

<sup>43</sup> Jacobson, B., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M., & Newman, C. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology*, 6, 66-70.

### III. SYNTHÈSE DES RÉSULTATS

#### L'AVQI et l'évaluation de la qualité de vie par le patient (VHI) :

En contrôlant l'effet de l'âge des patients, le score calculé par l'AVQI 03.01 concorde avec l'auto-évaluation de l'impact de la dysphonie sur le quotidien du patient à l'aide du VHI :

$$r_{s \text{ AVQI } 03.01 - \text{VHI}_{\text{total}}} = .62, p < .0001$$

De plus, les corrélations entre le score AVQI 03.01 et les scores aux trois sous-échelles du VHI sont toutes statistiquement significatives ( $p < .0001$ ), et ce de façon homogène :

-  $r_{s \text{ AVQI } 03.01 - \text{VHI}_{\text{fonctionnel}}} = .643$

-  $r_{s \text{ AVQI } 03.01 - \text{VHI}_{\text{émotionnel}}} = .543$

-  $r_{s \text{ AVQI } 03.01 - \text{VHI}_{\text{physique}}} = .514$

Ces corrélations peuvent toutes être considérées comme fortes selon les critères de Cohen (1988), et modérées selon les critères de Hinke et al. (1994).

Le jugement « normophonie – dysphonie » des deux outils d'évaluation est également corrélé :

$$\phi = .52, p = .000$$

Cette corrélation est élevée selon les critères de Cohen (1988), et relativement élevée selon les critères de Rea & Parker (1992).

## Discussion générale

Comme dans les nombreuses autres études concernant l'AVQI, cet index se révèle être une fois encore un outil assez robuste. Nous allons ici principalement discuter les résultats obtenus sur base des jugements perceptifs effectués sur la parole continue uniquement. En effet, la fidélité intra-juge et inter-juges s'est avérée plus forte dans cette condition (fidélité intra-juge moyenne = .78, contre .73 pour l'évaluation perceptive sur la parole continue et la voyelle tenue ; fidélité inter-juges = .90, contre .88). De plus, les données obtenues étaient plus cohérentes, notamment concernant l'évaluation des sujets normophoniques.

Notre étude nous a permis de confirmer la validité concourante de l'AVQI 03.01 sur un échantillon de parole continue en français. Nous obtenons un coefficient de corrélation  $r_s = .84$  entre les scores AVQI 03.01 et l'évaluation perceptive par les vocologistes sur base de la parole continue ( $r_s = .85$  pour l'évaluation perceptive sur la parole continue et la voyelle tenue). Pour rappel, dans leur étude en 2014, Maryn et al. ont mesuré un  $r_s$  de .781 pour l'échantillon en français (version 02.03) ; dans leur étude de 2016, Barsties et Maryn ont mesuré un  $r_s$  de .815 pour la version 03.01 (échantillon en néerlandais).

La précision diagnostique s'avère satisfaisante, et le score seuil optimal à employer pour un échantillon en français, tel qu'identifié par l'index de Youden dans notre étude, est de 2.33. Ce score est proche de celui identifié par Barsties et Maryn (2016) pour l'AVQI 03.01 appliqué à un échantillon en néerlandais (2.43).

La précision diagnostique de l'AVQI 03.01 semble, en général, quelque peu moins satisfaisante que celle de l'AVQI 02.03. En effet, la version 02.03 appliquée à un échantillon en néerlandais montre un LR+ > 2000 et un LR- de 0.16 ; la version 03.01 appliquée à un échantillon en néerlandais montre un LR+ de 11.54 et un LR- de 0.23. D'après nos résultats, l'AVQI 03.01 appliqué à un échantillon en français montre une tendance similaire, avec une spécificité très élevée, mais une sensibilité moins satisfaisante (LR+ tendant vers l'infini et LR- de 0.4)<sup>44</sup>.

Dans notre cas, nous notons l'absence totale de faux positifs (AVQI 03.01 pathologique alors que  $G_{moyen}$  indique une voix non pathologique), expliquant la spécificité parfaite (100 %) de l'outil ; Barsties et Maryn (2016) ont calculé une spécificité de 93.2 %, avec un LR+ de 11.54 (supérieur à 10). Dans leur étude, la sensibilité était moindre (78.5 %), avec ainsi un LR- de 0.23, supérieur à 0.1. Ce phénomène est également observé dans notre étude, mais de façon plus accentuée encore (sensibilité 59.8% ; LR- = 0.4). Nous tenons ainsi à rappeler que, bien

---

<sup>44</sup> Rappelons que Maryn et al. (2014) ont expliqué les plus faibles critères psychométriques obtenus par l'AVQI 02.03 sur un échantillon en français par le fait qu'il s'agit d'une langue romane, assez distante de la langue néerlandaise.

que le clinicien puisse garantir qu'un patient ayant un score supérieur ou égal à 2.33 est dysphonique ( $LR+ = \infty$ ), le contraire n'est pas valable : lorsque le patient a un score AVQI 03.01 non pathologique (inférieur à 2.33), le clinicien ne peut pas conclure avec assurance que le patient soit effectivement normophonique ( $LR- = 0.4$ ). Ceci est dû à la présence de nombreux faux négatifs (AVQI 03.01 non pathologique alors que  $G_{moyen}$  indique une voix pathologique), diminuant de façon conséquente la sensibilité de l'outil.

Nous avons analysé de plus près ces faux négatifs. Pour ceux-ci, la moyenne des scores AVQI 03.01 est de 0.94 ( $\sigma = 0.69$ ). La moyenne des  $G_{moyen}$  est de 0.88 ( $\sigma = 0.34$ ). Ainsi, nous pouvons tout de même observer que, bien qu'il y ait présence de faux négatifs, lorsque les enregistrements sont erronément détectés comme non pathologiques par l'AVQI 03.01, le  $G_{moyen}$  associé est généralement situé entre 0.5 et 1.5. Seules deux exceptions sont observées, avec un  $G$  moyen respectivement de 1.75, et de 2. Ainsi, nous soulignons que la probabilité qu'un score AVQI 03.01 non pathologique soit associé à une dysphonie sévère est minime. Les faux négatifs correspondent donc généralement à une dysphonie légère ( $G = 1$ ), rarement modérée ( $G = 2$ ).

Considérons maintenant les caractéristiques des participants faux négatifs. 17 parmi ces 41 participants appartiennent au groupe contrôle. À ce propos, nous tenons à attirer l'attention sur le fait que les vocologistes ayant effectué l'évaluation perceptive sont en quelque sorte entraînés à percevoir des traits pathologiques dans la qualité vocale. Ainsi, alors que les participants du groupe contrôle sont effectivement évalués comme normophoniques selon l'AVQI, les juges, cherchant à évaluer des voix pathologiques, attribuent assez facilement un  $G$  de 1. Parmi les patients « faux négatifs » du groupe expérimental, nous comptons autant d'hommes que de femmes (12, respectivement), âgés de 29 à 84 ans (moyenne : 54 ans), avec des contextes pathologiques variés<sup>45</sup> (immobilités laryngées, presbyphonies, dysphonies fonctionnelles...). À partir de ces observations, nous ne pouvons malheureusement pas tirer de conclusion quant à la raison de la présence de ces faux négatifs. En effet, ni les pathologies, ni l'âge, ni le genre des patients ne permettent d'expliquer les classements erronés par l'AVQI 03.01. Nous devons donc conclure que ces faux négatifs apparaissent de manière aléatoire et soulignons ainsi la prudence quant à l'interprétation d'un score non pathologique de l'AVQI 03.01.

---

<sup>45</sup> Notons que le degré de sévérité de la dysphonie n'est pas forcément corrélé avec le degré de la pathologie qui en est la cause. En effet, l'expression symptomatique de toute pathologie est hétérogène, et bien que l'on puisse – globalement parlant – citer certaines caractéristiques souvent rencontrées (p.ex. le souffle dans le cas des immobilités laryngées en abduction), l'on ne peut ni généraliser celles-ci, ni prédire le degré de l'altération vocale sur base du diagnostic médical. Rappelons aussi la conclusion de Schindler, Mozzanica, Vedrody, Maruzzi, et Ottaviani (2009), expliquant l'hétérogénéité des corrélations entre les sous-échelles du VHI et les mesures objectives dans leur étude, selon les pathologies analysées : les patients ne seraient pas attentifs aux mêmes aspects de leur voix selon l'étiologie de leur dysphonie.

Notons qu'aucun guideline d'interprétation des tailles d'effet de la sensibilité et de la spécificité n'est disponible dans la littérature. En effet, leur interprétation est fortement dépendante de l'objectif du test. Ainsi, dans certains cas, l'on préférera avoir l'assurance de détecter tous les malades (vrais positifs), nonobstant le risque élevé de faux positifs (sensibilité élevée, faible spécificité). Dans d'autres cas, l'on préférera rejeter le plus possible de vrais négatifs quitte à rejeter erronément certains patients atteints de la pathologie (faux négatifs). Notamment, les tests de screening sont employés pour effectuer un « premier tri », et présentent parfois une précision diagnostique « asymétrique ». Dans ce cas (par exemple un test montrant une sensibilité élevée, mais une spécificité moindre), l'idée est de soumettre le patient à un second test avec les caractéristiques inverses, afin de pouvoir tirer une conclusion plus fiable (Lalkhen & McCluskey, 2008). Cependant, dans le cas de l'AVQI 03.01, nous sommes de l'avis qu'une haute sensibilité serait préférable à une spécificité élevée. En effet, notre objectif est de pouvoir détecter toutes les dysphonies afin de pouvoir aider les patients qui en ont besoin. L'objectif contraire, de rejeter un maximum de patients normophoniques, quitte à également rejeter certains patients dysphoniques, signerait une perspective politico-financière visant par exemple à minimiser les remboursements octroyés par le système de santé. Ceci n'est bien évidemment pas un point de vue que nous soutenons. Ainsi, nos observations concernant la sensibilité limitée de l'outil nous permettent encore une fois de souligner l'importance d'une combinaison de l'AVQI 03.01 avec d'autres outils d'évaluation (notamment subjective), dans le cadre d'un bilan vocal complet. Un score AVQI 03.01 non pathologique ne doit pas à lui seul conduire à la conclusion d'une qualité vocale non altérée ne nécessitant pas de prise en charge.

Ensuite, nous tenons aussi à recenser les principales faiblesses de notre travail, afin de nuancer quelque peu nos résultats.

Premièrement, reprenons les limites concernant les enregistrements vocaux utilisés dans notre étude. Les enregistrements effectués au laboratoire vocal ont été effectués avec un taux d'échantillonnage de 11025 Hz (au lieu de 44100 Hz tel que recommandé par Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2009). Le niveau de pression sonore d'entrée n'est pas maintenu constant pour chaque sujet, ce qui peut avoir un impact sur la précision des mesures acoustiques. Enfin, la plupart des enregistrements ont été effectués dans un local calme, mais non insonorisé. Le bruit ambiant dans ce local, mesuré à quatre reprises à l'aide d'un décibelmètre PCE-353 (PCE Instruments, Strasbourg, France), était en moyenne de 50.33 dB. En considérant que le rapport signal-bruit recommandé doit être supérieur à 42 dB (Deliyski, Shaw, & Evans, 2005), le niveau de bruit ambiant dans le laboratoire vocal semble donc insatisfaisant. Les mesures thermo-hygrométriques, effectuées à l'aide d'un hygromètre Hygromaster BLD7700



(Protimeter, Pennsylvanie, États-Unis), indiquaient également un pourcentage d'humidité relative de 26.6 % (avec une température moyenne de 22.65 °C), signant un air sec qui peut avoir un effet sur la biomécanique vocale (Hemler, Wieneke, Lebacq, & Dejonckere, 2001), et sur les paramètres acoustiques (Hemler, Wieneke, & Dejonckere, 1997). Nous pensons cependant que la durée d'un bilan vocal n'est probablement pas assez longue que pour voir apparaître des altérations significatives des paramètres acoustiques.

De plus, lors du tri des enregistrements, nous avons dû effectuer un jugement subjectif, basé principalement sur l'accent, afin de ne garder que les individus dont le français nous semblait être la langue maternelle, ainsi que ceux dont l'accent régional n'était « pas trop prononcé ». Cependant, comme le décrit très bien Joana Révis dans son ouvrage « La voix et soi » (Révis, 2013), chacun d'entre nous a un accent pour quelqu'un issu d'un autre groupe linguistique. Les variations diatopiques (basées sur l'origine géographique) et diastratiques (basées sur l'origine socioéconomique/démographique) du français sont nombreuses, autant en France et en Belgique que dans le monde entier, et nous nous interrogeons quant à la validité de l'AVQI 03.01 sur l'ensemble de ces variations, dont certaines peuvent conséquemment différer.

Deuxièmement, lors de l'évaluation perceptive des échantillons vocaux à l'aide du critère G de la GRBAS, les enregistrements n'ont pas été présentés dans le même ordre aux quatre évaluateurs. Ceci peut représenter une limite à notre étude, dans la mesure où l'ordre de présentation peut influencer les jugements de sévérité. Par exemple, une voix présentée juste après une voix sévèrement dysphonique n'est pas évaluée de la même manière que lorsqu'elle est présentée après une voix normophonique (Goldstone, 1998, cité par Barkat-Defradas et al., 2012). Cette lacune peut donc avoir eu un impact sur la fidélité inter-juges.

Troisièmement, les jugements perceptifs par les quatre vocologistes ont été effectués à domicile, avec des consignes précises. Nous n'avons aucun contrôle sur l'application correcte des instructions (respect des pauses, qualité du casque employé, concentration de l'évaluateur, événements distrayants, ...). Concernant la qualité et le sérieux des évaluations perceptives effectuées, nous pouvons donc uniquement nous fier à nos mesures de fidélité.

Quatrièmement, nous souhaitons souligner que l'évaluation perceptive par les vocologistes est quelque peu biaisée. En effet, ceux-ci, en contact régulier avec des patients dysphoniques, sont en quelque sorte « drillés » à percevoir des traits pathologiques dans la qualité vocale. Ainsi, parmi les 30 participants contrôles, seuls 12 ont obtenu un G moyen < 0.5. En effet, dans les 18 autres cas, au moins deux parmi les quatre vocologistes ont attribué un G de 1, la moyenne étant alors supérieure

ou égale à 0.5. Une solution possible aurait été de considérer une voix comme pathologique non pas lorsque le G moyen est strictement inférieur à 0.5, mais lorsqu'il est inférieur ou égal à ce même seuil.

De plus, nous avons finalement basé nos conclusions concernant la précision diagnostique de l'outil sur les jugements perceptifs effectués uniquement sur la parole continue. Comme nous l'avons développé, la combinaison des deux conditions phonatoires ne nous a pas permis d'obtenir des données interprétables. Aussi, nous ne parvenons pas à expliquer pourquoi, dans d'autres études, la combinaison des deux conditions n'a pas posé problème lors de l'évaluation perceptive. En effet, chacune des quatre vocologues a fait un retour concernant cet aspect, sans pour autant qu'elles n'aient été en contact entre elles. Finalement, une alternative à ce problème aurait été de présenter la parole continue et la voyelle soutenue, mais de faire évaluer les deux séparément (donc, permettre aux évaluateurs d'attribuer un score G de la GRBAS pour chaque condition). C'est effectivement la méthode employée par Hosokawa et al. dans leur étude de validation de l'AVQI 03.01 en japonais (sous presse). Les résultats peuvent alors être analysés soit en considérant les deux conditions phonatoires individuellement, soit en calculant la moyenne des scores pour les deux conditions.

Nous tenons aussi à souligner que notre méthodologie peut présenter un biais test-retest, car les évaluations perceptives par les vocologues ont été effectuées à deux reprises dans un intervalle de 15 à 32 jours. Ce biais pourrait en partie expliquer l'amélioration de la fidélité de leurs jugements.

Pour étudier la validité de l'AVQI 03.01, nous avons créé des groupes à effectifs égaux pour trois degrés de dysphonie : normophonie, dysphonie légère à modérée, dysphonie sévère. Malheureusement, pour cela, nous ne disposons pas pour l'entièreté des individus des grades G (échelle GRBAS) attribués par les vocologues du CHU au moment du bilan. Nous avons donc dû classer les scores bruts en trois catégories en utilisant les critères définis dans le tableau 43 (page 72). Le seuil de 6 (limite entre la dysphonie légère à modérée et la dysphonie sévère) a donc été choisi subjectivement en mettant en lien les scores AVQI et les quelques scores G dont nous disposons. En effet, nous avons constaté qu'un score AVQI supérieur à 6 correspondait généralement à un G de 3. Ce seuil coïncide cependant également avec les données récoltées par Maryn et al. (2012) et par Reynolds et al. (2012). Les données que nous avons finalement obtenues dans le cadre de notre travail vont dans le même sens. En effet, comme illustré dans le tableau 61, parmi les 120 enregistrements analysés, un AVQI 03.01 supérieur ou égal à 6 correspond à un G moyen supérieur à 2.5 (c.-à-d. que la majorité des juges a octroyé un score de 3) dans 88 % des cas. De la même manière, lorsque le G moyen est inférieur ou égal à 2.5 (c.-à-d. que la majorité des juges a octroyé un G de maximum 2), le score AVQI 03.01 est inférieur à 6 dans 92 % des cas.

Tableau 61 Pourcentages des enregistrements obtenant des scores sévères ou non, selon l'AVQI 03.01 et selon l'évaluation perceptive moyenne

	AVQI <sub>03.01</sub> ≥ 6	AVQI <sub>03.01</sub> < 6
G <sub>moy</sub> ≤ 2.5	8.42 %	91.59 %
G <sub>moy</sub> > 2.5	88 %	12 %

Note. Le tableau est à lire en lignes.

Ainsi, l'utilisation du score seuil de 6 pour différencier les dysphonies « légères à modérées » des dysphonies « sévères » semble tout de même justifiée.

Enfin, nous avons dû faire un choix entre (a) l'équilibre temporel entre les deux conditions de phonation et (b) la représentativité phonétique de l'extrait. En effet, l'extrait de 27 syllabes permet d'obtenir trois secondes de voisement. Cependant, rappelons qu'il y manque les phonèmes voisés suivants (entre parenthèses, les fréquences d'occurrences dans la langue française selon Tubach & Boe, 1990) : les semi-voyelles [ɥ] (0.46 %), [w] (1.08 %), [j] (1.77 %), les voyelles [y] (1.9 %), [ø] (0.45 %) et [o] (1.07 %), ainsi que les consonnes [z] (1.85 %), [m] (3.30 %), [d] (4.13 %) et [b] (1.05 %).

L'on peut donc, malgré le bon équilibre phonétique des phonèmes présents dans l'extrait, se poser la question de la représentativité de l'extrait par rapport à la phonologie de la langue française parlée. Rappelons cependant que l'extrait de « La bise et le soleil » présentait une faiblesse similaire, avec davantage de phonèmes manquants, dont la voyelle [a] très fréquente en langue française : les voyelles [u] (2.02 %), [ɑ] (7.98 %), [o] (1.07 %), [ø] (0.45 %), [ɛ̃] (1.26 %) et [ɔ̃] (1.95 %), les semi-voyelles [w] (1.08 %) et [ɥ] (0.46 %), et les consonnes [g] (0.55 %), [m] (3.30 %), [ɲ] (0.08 %), [v] (2.27 %), [ʒ] (1.59 %), et [n] (2.79 %).

Finalement, nous pouvons aussi remettre en question la pertinence des fréquences d'occurrences mesurées par Tubach et Boe en 1990, sur un corpus dont la collection a débuté en 1967. Il serait en effet intéressant de vérifier si ces données sont toujours d'application aujourd'hui, les langues évoluant constamment. De plus, les fréquences d'occurrence ont été mesurées sur la langue française en France (à Paris et à Grenoble, principalement), qui présente de nombreuses différences de prononciation et de vocabulaire par rapport au « français de Belgique » (en considérant bien sûr l'ensemble des multiples variations régionales). Il serait alors nécessaire de disposer d'un recensement récent des fréquences d'occurrence des phonèmes dans la langue française parlée en Belgique, aujourd'hui.

En tenant compte de ces faiblesses concernant notre méthodologie, il nous semble important de souligner qu'une étude de réplification serait nécessaire afin de confirmer nos conclusions.

Pour celle-ci, il serait notamment intéressant d'utiliser la méthodologie d'Hosokawa et al. (sous presse) concernant l'évaluation perceptive des enregistrements vocaux.

Enfinement, en réponse à notre questionnement initial, nos résultats montrent que le score AVQI 03.01 et les scores au VHI sont modérément corrélés. En effet, le score AVQI 03.01 corrèle autant avec le score total au VHI qu'avec les trois sous-échelles, et les verdicts finaux posés selon le score seuil de l'AVQI 03.01 (2.33) et selon le score seuil du VHI (20/120) sont concordants. Nous souhaitons cependant attirer l'attention sur les coefficients de corrélation modérés ; de plus, nous tenons à rappeler, pour la pratique clinique, qu'il s'agit ici de tests de probabilité statistique. Aussi, il est important de garder à l'esprit que nous devons considérer le patient dans sa globalité, et interpréter les scores obtenus en lien avec les autres données du bilan vocal. En effet, un score AVQI 03.01 non pathologique ne signe pas nécessairement une absence de handicap perçu. Parmi nos résultats, nous pouvons par exemple observer un score total au VHI élevé, de 51/120, avec un score AVQI très bas, de 0.07. Le cas contraire peut également être observé dans nos résultats. Effectivement, même si la plupart des scores AVQI 03.01 pathologiques sont associés à un score VHI pathologique, dans quatre cas, nous observons une exception à cette règle. Il s'agit d'hommes âgés de 67, 68, 70 et 71 ans, qui ont un score AVQI 03.01 pathologique (p.ex. 7.09) avec un score VHI total non pathologique (p.ex. 14/120). Donc, un score pathologique à l'AVQI 03.01 n'est pas forcément vécu comme un handicap ; d'après nous, de nombreux facteurs peuvent influencer le degré de handicap perçu, tels l'âge et le contexte professionnel. Notons de plus que les patients inclus dans notre étude ont tous consulté au laboratoire vocal du CHU de Liège, et avaient donc, pour la plupart (sauf pour les consultations préopératoires), une plainte vocale initiale. Il s'agit ainsi ici clairement d'un biais de sélection, et les résultats doivent donc être interprétés avec précaution.

Enfin, rappelons qu'il est aujourd'hui largement admis que l'évaluation subjective, autant par le thérapeute que par le patient, doit être prise en compte *parallèlement* aux autres données récoltées. Bien que l'AVQI 03.01 s'avère être un outil d'évaluation robuste de la qualité vocale, actuellement, il y a toujours un manque d'accord quant aux corrélats acoustiques de cette dernière (Reynolds et al., 2012). Ainsi, les index de qualité vocale incluent dans leurs calculs des paramètres acoustiques différents et n'évaluent donc pas exactement la même réalité multidimensionnelle. Certains auteurs vont même jusqu'à définir la qualité vocale comme n'étant pas une caractéristique du signal vocal à proprement parler (Kreiman & Gerratt, 1998). Il semble

donc que les définitions actuelles de la qualité vocale ne permettent pas de concevoir un outil d'évaluation faisant l'unanimité pour quantifier ce phénomène perceptif multidimensionnel. De plus, la fiabilité des mesures acoustiques en général peut être remise en question dans certains cas. Par exemple, très fréquemment, les patients dysphoniques se plaignent de variations importantes de leur qualité vocale dans le temps. Se fier uniquement à des mesures acoustiques effectuées à un instant *t* est alors fortement déconseillé. Aussi, les mesures acoustiques effectuées sur deux [a] tenus par une même personne, même avec un intervalle de quelques secondes, montreront toujours une certaine inconstance. Plus étonnamment encore, les paramètres mesurés sur un même [a] différeront selon l'endroit où est placée la fenêtre d'analyse (Révis, 2013). Toutes ces données soulignent l'importance d'une évaluation complète à l'aide de différents outils complémentaires : VLS, anamnèse, mesures objectives acoustiques et aérodynamiques et évaluation subjective par le patient et par le thérapeute. Un bilan vocal de haute qualité ne doit faire l'impasse sur aucun de ces facteurs. C'est uniquement grâce à l'intégration de l'ensemble de ces données que le clinicien pourra estimer les implications et conséquences d'une éventuelle pathologie vocale.

## Réflexions

De nombreuses questions sont apparues au courant de la rédaction de ce travail, à propos de l'utilisation de l'index AVQI. Nous en avons recensé les plus pertinentes à nos yeux, afin de susciter la réflexion du lecteur et de pouvoir éventuellement inspirer l'un ou l'autre projet de recherche.

Premièrement, nous nous sommes interrogés quant au champ d'application de l'AVQI au sein de la pratique clinique des vocologistes. En effet, l'AVQI est considéré — selon la liste limitative de l'INAMI — comme un outil d'évaluation globale de la qualité vocale, toutes circonstances confondues. Il nous semble ici nécessaire de souligner que, bien que l'AVQI ait été créé pour l'évaluation de la qualité vocale, il revient au bon sens du clinicien et à la pratique de l'EBP de juger de la pertinence de l'utilisation de cet outil pour chaque patient, individuellement. En effet, nous rencontrons dans notre pratique clinique des personnes présentant des plaintes ou problèmes vocaux dans le cadre desquels l'AVQI sera moins informatif que d'autres moyens d'évaluation vocale. Citons quelques exemples afin d'illustrer ce point :

La fréquence fondamentale usuelle n'est pas prise en compte dans le calcul de l'AVQI. Nous tenons à souligner qu'il ne s'agit pas là d'une limite de l'AVQI, celui-ci n'ayant pas été conçu à cet effet. L'AVQI est un outil d'évaluation de la qualité vocale — la qualité vocale englobant tout ce qui ne relève pas de la hauteur ou du niveau de pression sonore (Behrman, cité dans Boone, McFarlane, Von Berg, & Zraick, 2010). Aussi, le clinicien doit rester dans une démarche de réflexion concernant la pertinence des outils de mesure pour chaque cas spécifique, et adapter le bilan vocal et les mesures utilisées par rapport au patient qu'il a devant lui. Par exemple, dans le cas d'un patient consultant pour un trouble de la mue, la prise en compte de la fréquence fondamentale sera primordiale<sup>46</sup>. En effet, sa voix ne présentera pas nécessairement des paramètres acoustiques pathologiques. La plainte vocale, pourtant bel et bien présente, ne sera probablement pas corroborée par un score AVQI pathologique. Comme deuxième exemple, nous pouvons citer les modifications structurales des plis vocaux, qui induisent quelquefois des changements importants de la hauteur vocale (bien souvent dans le sens d'une aggravation). Certes, ces modifications résultent généralement également en une altération de la qualité vocale en termes notamment de raucité ou de voix soufflée. Cependant,

---

<sup>46</sup> Rappelons que la fréquence fondamentale est un critère présent dans la liste limitative de l'INAMI, et peut donc également contribuer à l'accord d'une prise en charge lorsqu'il s'agit d'un paramètre pathologique saillant.

ces paramètres vocaux ne sont pas dans tous les cas les plus saillants, ce qui pourrait donc dans certains cas conduire à un score AVQI peu élevé, voire non pathologique.

Lorsqu'un patient chuchote, par exemple dans le cas d'une aphonie psychogène, ou que le voisement est quasiment absent de l'échantillon vocal, il convient également d'être prudent quant à l'emploi de l'AVQI. En effet, rappelons que le calcul de ce dernier s'opère uniquement sur les segments voisés de la parole continue.

Le nasonnement n'est a priori pas pris en compte dans l'analyse de l'AVQI. Lorsqu'il s'agit du paramètre essentiel de la plainte vocale, encore une fois, il ne conviendra donc pas d'utiliser cet outil comme mesure principale de quantification de la qualité vocale du patient.

Aussi, nous nous interrogeons quant aux bruits inspiratoires (stridor), présents notamment dans certains cas d'immobilité laryngée. A priori, nous pensons qu'ils auront un impact sur le score AVQI calculé sur la parole continue. En effet, ces inspirations sonores seront probablement détectées par l'AVQI comme des segments voisés, ce qui rendra moins aisée l'interprétation du score obtenu. Encore une fois, il faut donc nuancer la pertinence de l'utilisation de l'AVQI dans ce cas particulier.

De même, l'AVQI n'est pas prévu pour les pathologies affectant la fluence/le débit de parole. Nous pouvons ici notamment citer le bégaiement, mais aussi les dysphonies spasmodiques. La qualité vocale dans ces contextes pathologiques est souvent très instable, et sa quantification sur base d'un court extrait de parole continue n'est pas assez informative. D'autres outils seront dans ce cas plus adaptés, tel que notamment le modèle AMPEX (Auditory Model Based Pitch Extractor) créé par Van Immerseel et Martens en 1992 (Siemons-Lühring et al., 2009), ou l'outil BioVoice2 (Giordano, Dejonckere, & Manfredi, 2011), tous deux adaptés pour l'évaluation acoustique de signaux très irréguliers.

Ainsi, par ce premier point, nous tenons à rappeler aux cliniciens qu'aucun outil n'a été créé pour répondre aux besoins de l'ensemble des pathologies vocales. Il convient donc de se baser sur les sources scientifiques récentes, sur son expérience clinique et sur les spécificités du patient afin de choisir les bons outils pour l'évaluation vocale.

De plus, il nous semble également important de souligner que pour les études de validation d'outils d'évaluation vocale, la population étudiée devrait correspondre à la population d'application de l'outil. Il serait donc nécessaire d'exclure les contextes pathologiques particuliers, pour lesquels l'utilisation de l'outil étudié est mise en question. En effet, nous avons constaté que dans les études concernant l'AVQI, les échantillons de patients dysphoniques comprennent notamment des mues faussées, des patients transgenres, et des

dysphonies spasmodiques. Nous aussi, nous avons inclus dans notre étude certaines pathologies que — suite aux réflexions a posteriori — nous aurions probablement dû exclure.

Deuxièmement, nous nous sommes demandé comment procéder lorsque nous devons effectuer un bilan vocal avec une personne dyslexique, illettrée, ou analphabète. En d'autres termes, comment faire lorsque la fluence de la lecture du texte est compromise ? Cette interrogation concerne bien évidemment tous les outils d'évaluation nécessitant la lecture de phrases. Comme élément de réponse, nous avons pensé qu'il serait envisageable de procéder à une répétition des phrases : le thérapeute prononce les phrases, segment par segment, et le patient répète. Ensuite, il faudrait extraire les productions du patient de l'enregistrement, afin de pouvoir effectuer les analyses. Cependant, cette méthode pose le problème du phénomène d'imitation : le patient aura tendance à imiter la voix et les inflexions du thérapeute, ses répétitions ne correspondant alors pas à la voix qui aurait été employée spontanément.

Notre troisième réflexion porte sur les mesures de perturbation incluses dans le calcul de l'AVQI. Pour rappel, l'AVQI utilise deux mesures de shimmer, ainsi que le HNR — des mesures de perturbation classiques largement controversées dans le cadre de l'évaluation de signaux fortement apériodiques. Comme nous l'avons mentionné dans la partie théorique de ce travail, les mesures de perturbation « jitter » et « shimmer » ne sont interprétables que lorsqu'elles sont inférieures à 5 % (Leong et al., 2012). Or, bien souvent, les calculs de l'AVQI résultent en des seuils bien plus élevés. Dans ce cas, peut-on toujours interpréter raisonnablement le score AVQI ? Si l'on considérait la réponse négative, cela reviendrait à dire que l'AVQI serait surtout utilisable dans le cadre des dysphonies de type 1 selon Titze (1994), pour lesquelles l'énergie sonore de la F0 est assez saillante. Nous sommes cependant de l'avis que tel n'est pas le cas : une mesure de shimmer très élevée, même si elle n'est pas interprétable à l'état isolé, signe tout de même une qualité vocale fortement perturbée. Ainsi, nous pensons pouvoir conclure en soulignant qu'il sera difficile de distinguer perceptivement deux voix dont le score AVQI est respectivement de 8 et de 9. Cependant, si nous considérons un score de 3 et un score de 8, même si les mesures de perturbation seront sans doute ininterprétables en tant que telles dans ce dernier cas, les deux scores AVQI renverront très probablement à deux réalités perceptives qui font sens au clinicien. De plus, la mesure cepstrale, élément principal dans la formule de l'AVQI, justifie également l'utilisation de cet outil dans les dysphonies sévères.



Quatrièmement, nous nous sommes demandé si les mesures effectuées sur les trois secondes centrales de la voyelle tenue sont représentatives de la « performance vocale » du patient lors de cette tâche. En effet, nombreux sont les enregistrements sur lesquels nous percevons une qualité vocale fortement altérée lors de l'attaque (attaque dure, serrage, difficultés de mise en vibration des plis vocaux, attaque soufflée...) ou lors de la finale de la voyelle tenue, les trois secondes centrales étant cependant stables et de bien meilleure qualité. Aujourd'hui, à notre connaissance, tous les index mesurant des paramètres acoustiques sur base d'une voyelle tenue font fi des attaques et des finales. Or, les perturbations de la mise en vibration et de l'arrêt du voisement participent de manière importante à la qualité vocale perçue. Dans le travail thérapeutique des vocologistes, l'initiation de la vibration des plis vocaux est d'ailleurs d'une grande importance. Nous nous trouvons donc là face à une dissonance entre les aspects pris en compte dans l'évaluation vocale objective et les facteurs primordiaux aux yeux des cliniciens dans la prise en charge de la voix. Ainsi, une amélioration de l'initiation vibratoire sera perçue comme un progrès par le clinicien — et probablement par le patient —, alors qu'elle ne sera pas détectée par les analyses acoustiques.

Enfin, nous tenons à signaler que, dans le script de l'AVQI version 02.03, il nous semble qu'une erreur s'est glissée dans les lignes codant l'illustration finale du score calculé (autant en mode simple qu'en mode illustré) :

```
Paint rectangle... green 0 2.91 0 1  
Paint rectangle... red 2.91 10 0 1
```

En effet, le score seuil (pour le néerlandais) est normalement de 2.95, et non de 2.91. Ainsi, un score de 2.92 (normalement normophonique) sera situé dans la zone rouge sur le graphique, ce qui pourrait amener à des erreurs d'interprétation.

Une erreur du même type a été décelée dans le script de l'AVQI version 03.01, cette fois-ci uniquement dans le mode illustré : au lieu du nouveau score seuil de 2.43 y figure celui de 2.91. Un score de 2.60, par exemple (normalement pathologique selon la version 03.01), sera donc ici situé dans la zone verte sur le graphique, ce qui peut à nouveau entraîner une erreur d'interprétation.

## Perspectives de recherche

Ci-après, nous développons quelques pistes de réflexion supplémentaires concernant l'AVQI et les mesures de corrélation entre celui-ci et d'autres outils d'évaluation de la voix. Nous espérons que, peut-être, ces pistes pourront inspirer des projets de recherche ultérieurs.

*Concernant l'AVQI :*

- Revenons sur la problématique de l'utilisation de l'AVQI pour les dysphonies caractérisées par de nombreux épisodes de dévoisement. Le calcul de l'AVQI s'opère sur les segments voisés concaténés de la parole continue. Au contraire, pour la voyelle tenue, l'AVQI n'effectue pas d'extraction des segments voisés, mais prend simplement en compte les trois secondes centrales sans analyse préliminaire. Ne pourrait-on pas imaginer, dans une version ultérieure de l'AVQI, effectuer une extraction des segments voisés de la voyelle soutenue afin de les concaténer, comme c'est le cas pour la parole continue ?
- Le « shimmer local », utilisé dans le calcul de l'AVQI, mesure l'écart moyen absolu des amplitudes des périodes consécutives, divisé par l'amplitude moyenne (Boersma, 2017) :

$$\frac{\sum |amplitude_{période\ k} - amplitude_{période\ k+1}|}{n \cdot amplitude_{moy}}$$

Où « n » est le nombre de périodes extraites pour l'analyse.

Cette mesure prend donc en compte les différences d'amplitude cycle par cycle, de façon très précise. L'on pourrait envisager de prendre en compte un shimmer calculé sur des « fenêtres » plus larges (moyennes courantes). Ainsi, ne diminuerait-on pas les biais liés aux variations minimales d'amplitude liées notamment aux tremblements physiologiques ou aux modulations du tractus vocal pendant la parole spontanée ?

L'on pourrait par exemple imaginer utiliser le shimmer « apq5 » — « five point amplitude perturbation quotient » — (ou l'« apq11 »), pour lequel l'amplitude de chaque période est moyennée sur les quatre (ou dix) périodes voisines :

$$\frac{\sum \left| amplitude_{période\ k} - \left( \frac{ampl_{période\ k-2} + ampl_{période\ k-1} + ampl_{période\ k} + ampl_{période\ k+1} + ampl_{période\ k+2}}{5} \right) \right|}{n \cdot amplitude_{moy}}$$

Où « n » est le nombre de périodes extraites pour l'analyse.

- L'on pourrait s'intéresser au pouvoir de classification de l'AVQI 03.01 : cet outil permet-il de différencier les degrés de sévérité de dysphonie, comme c'est le cas pour le DSI (Hakkesteegt, 2008) ?
- L'on pourrait également évaluer la validité de l'AVQI 03.01 comme mesure de suivi en français, en investiguant si cet outil permet de détecter de manière fiable les changements

de qualité vocale après une intervention chirurgicale et/ou conservatrice. Rappelons que Barsties et Lehnert (2018) ont confirmé la capacité de l'AVQI 03.01 à détecter les modifications du souffle perçu et de la raucité après prise en charge conservatrice.

- Quelles sont finalement les recommandations générales pour l'utilisation de l'AVQI 03.01 en pratique clinique? En effet, il serait intéressant de pouvoir fournir aux cliniciens des recommandations précises, notamment quant aux conditions d'enregistrements (type de micro, niveau maximal de bruit ambiant, ...). Bien sûr, il s'agit là d'une question sensible, dans la mesure où, par exemple, l'on ne peut pas exiger des cliniciens de s'approprier un matériel trop onéreux. En effet, la plupart des supports informatiques et instruments qui seraient optimaux pour un bilan vocal de haute qualité sont coûteux. Cependant, en évitant un discours prescriptif, des conseils pratiques pourraient orienter l'attention des cliniciens sur certains aspects primordiaux pour un bilan vocal de qualité, et ainsi garantir des mesures plus fiables.
- Aussi, il pourrait être intéressant d'ajouter au script une ligne permettant au clinicien de constater plus clairement le « verdict » final. En effet, avec les différentes versions de l'AVQI et sa validation dans de nombreuses langues, de nombreux scores seuils différents existent. Le clinicien « naïf » pourrait se perdre parmi ces chiffres. Ainsi, l'on pourrait concevoir qu'il ait la possibilité de sélectionner la langue de l'échantillon à analyser, sélection qui conditionnera l'affichage du score final en fonction du score seuil de la langue cible. En plus de l'interprétation visuelle à partir du cadre rouge et vert, il pourrait aussi être intéressant d'ajouter un commentaire écrit de type « pathologique : OUI/NON », afin d'éviter toute erreur d'interprétation.

*Concernant le lien entre l'AVQI et l'évaluation de la qualité de vie par le patient :*

- Comme évoqué dans ce travail, le VoiSS semble être un outil plus robuste encore que le VHI. Il serait donc intéressant de mesurer la corrélation entre le score AVQI 03.01 et les scores au VoiSS.
- Selon Barsties et Maryn (2013), l'évaluation de la qualité vocale à l'aide de l'AVQI est indépendante de l'âge et du genre du sujet évalué (Barsties & Maryn, 2013). Concernant le VHI, les études montrent des résultats controversés. Maertens et De Jong (2007), par exemple, observent un effet de l'âge sur les scores au VHI, mais pas du genre ni de la profession. Rodrigues Bastilha, De Moraes Lima, et Aparecida Cielo (2014) observent cependant des effets de l'âge, du genre et de la profession.

D'après notre (courte) expérience clinique, nous formulons l'hypothèse que le handicap perçu diffère selon l'activité professionnelle du patient. Ainsi, il serait intéressant de mesurer la corrélation entre l'AVQI 03.01 et le VHI en fonction de la profession du patient. Nous n'avons malheureusement pas des effectifs suffisamment importants pour chaque

profession afin de répondre à cette question. Dans la même ligne de pensée, il pourrait être intéressant de contribuer aux discussions concernant les effets de l'âge et du genre sur les scores au VHI et sur l'AVQI. Puis, finalement, l'on pourrait également analyser l'influence de ces deux facteurs sur la corrélation entre l'AVQI et le VHI.

- L'on pourrait aussi mesurer les corrélations entre chacune des six mesures acoustiques de l'AVQI 03.01 et les scores au VHI, afin de déterminer si certains paramètres acoustiques corréleraient davantage avec l'une des sous-échelles du VHI.
- Il serait aussi intéressant de mesurer s'il existe une différence des moyennes des paramètres acoustiques de l'AVQI 03.01 entre les deux groupes « VHI pathologique » et « VHI non pathologique », ainsi qu'entre les deux groupes « VoiSS pathologique » et « VoiSS non pathologique ». Effectivement, une étude menée par Lopes et al. en 2017 montre que les patients avec un handicap autoévalué à l'aide du VoiSS présentent une plus grande déviation des paramètres acoustiques investigués (déviations standard de la F0, jitter, shimmer et glottal-to-noise excitation ratio) que ceux qui ne ressentent pas de handicap lié à la qualité vocale. Cependant, aucune différence significative n'a été trouvée pour les moyennes de ces paramètres entre les deux groupes « VHI pathologique » et « VHI non pathologique ».
- Également, l'on pourrait mesurer s'il existe une différence de corrélation AVQI 03.01 – VHI entre les dysphonies fonctionnelles et les dysphonies organiques. Schindler et al., en 2009, ont par exemple investigué les différences des corrélations entre le VHI et différentes mesures objectives (jitter, shimmer, HNR, TMP et F0) selon les quatre groupes suivants : dysphonie fonctionnelle, immobilité laryngée unilatérale, dysphonie structurelle (kystes, polypes et œdèmes de Reinke) et nodules. Nous n'avons pas d'effectifs suffisants pour chaque groupe de pathologie afin d'effectuer des analyses similaires.

*Concernant le lien entre l'AVQI et d'autres outils d'évaluation vocale :*

- Nous avons vu que, selon certains auteurs, l'outil CAPE-V est un outil plus sensible que l'échelle GRBAS. Une prochaine étude pourrait ainsi être consacrée à la mesure de la corrélation entre le score AVQI 03.01 et l'évaluation perceptive à l'aide de l'outil CAPE-V. Ceci permettrait de confirmer la validité concurrente de l'AVQI 03.01.
- Aussi, l'on pourrait investiguer la question d'une corrélation entre le score de l'AVQI 03.01 et le score du DSI. L'on pourrait notamment analyser si le pourcentage de remboursements INAMI diffère significativement selon l'outil d'évaluation objective employé.

## Conclusion générale

Notre étude a permis de conclure que l'Acoustic Voice Quality Index version 03.01 est un outil valide d'évaluation acoustique de la qualité vocale globale en langue française. En effet, nous avons mesuré une validité concourante élevée en comparant les scores AVQI 03.01 aux jugements perceptifs à l'aide du critère G de la GRBAS par quatre vocologistes.

Le score seuil optimal pour l'utilisation de l'AVQI 03.01 en langue française, permettant la meilleure précision diagnostique, est de 2.33. Nos mesures démontrent une précision diagnostique satisfaisante, avec une très bonne spécificité, mais une sensibilité moindre. Une interprétation prudente est donc conseillée en cas de score non pathologique.

Nous avons aussi mesuré une corrélation significative modérée entre les scores AVQI 03.01 et les scores au Voice Handicap Index. L'évaluation objective de la qualité vocale par l'AVQI 03.01 est donc globalement en lien avec l'évaluation de l'impact de cette qualité vocale sur le quotidien du patient, autant sur le plan fonctionnel que physique et émotionnel.

Notons cependant que la corrélation entre l'AVQI 03.01 et les scores au VHI n'est pas parfaite, et que, si nous considérons les critères d'interprétation de Hinkle, aucune n'atteint un degré élevé.

Ainsi, comme mot de la fin, nous citons Baken et Orlikoff (2000), pour souligner l'importance de la prise en compte *concomitante* des données objectives et subjectives : « *measurements can be no better than the knowledge and skills of the clinician who chooses and obtains them.* » L'évolution de la technologie — malgré tous les avantages qu'elle nous offre — ne doit pas amener à ce que le clinicien se retranche derrière celle-ci : l'évaluation objective doit être considérée non pas comme un remplacement de, mais bien comme un outil complémentaire à l'évaluation perceptive, ainsi que — au même niveau d'importance — à l'évaluation par le patient de l'impact de la dysphonie sur sa qualité de vie. Dans le cadre du remboursement des séances de rééducation vocale, nous sommes de l'avis qu'en aucun cas l'évaluation objective à elle seule ne doit conditionner la décision de l'accès à la prise en charge. Tout compte fait, l'objectif primaire du logopède est d'améliorer la qualité de vie de ses patients dysphoniques. Ainsi, et ce dans la philosophie contemporaine des soins centrés sur le patient, il sera toujours indispensable de prendre en compte son avis dans les prises de décisions cliniques. Deux voix

dysphoniques sur lesquelles est calculé un même score objectif ne seront pas vécues de la même manière par les personnes qui les génèrent, dans leurs contextes de vie et de travail spécifique. En outre, l'évaluation perceptive par l'oreille experte du vocologue restera à nos yeux — tout du moins dans l'avenir proche — le gold standard pour l'évaluation vocale. En effet, aucun outil n'est en ce jour capable de saisir, aussi pleinement que l'oreille humaine, l'entière du phénomène multidéterminé qu'est la voix. Citons Le Breton (2011), qui en décrit merveilleusement bien la complexité :

*Aucune science n'épuise la question de la voix même si l'acoustique, la phonétique ou la linguistique essaient de la résorber dans leur savoir, elle fuit de partout, elle ne se laisse pas circonvenir. L'émotion liée à l'audition d'une voix ne tient pas à ses propriétés acoustiques, mais à son impact sur le désir de celui qui écoute. Il en va de même du visage, les deux éléments les plus intimes, les plus singularisés de l'humain et ceux qui se dérobent le plus. En donnant chair au langage la voix le donne à entendre. Pourtant elle est volatile, elle ne connaît en apparence aucune frontière même si, sans moyen technique pour l'amplifier, elle ne porte pas très loin. Elle ne pèse rien, mais elle peut changer le monde, bouleverser pour le meilleur ou pour le pire. Bien entendu une voix est une matière sonore à la fois sociale, culturelle, sexuée, affective, singulière, marquée par des ritualités et des émotions propres à une communauté linguistique à un moment de son histoire.*

Parviendra-t-on un jour à mesurer fidèlement un phénomène perceptif qui, comme l'ont très bien décrit Kreiman et Gerratt (1998), est somme toute le résultat de l'interaction complexe du signal vocal et de la perception par l'auditeur ? De nouveaux modèles d'analyse acoustique, notamment ceux inspirés des systèmes de reconnaissance automatique du locuteur (Fredouille et al., 2009) — eux-mêmes basés sur l'apprentissage automatique — parviendront-ils à combler notre désir de quantification de la qualité vocale ? L'avenir nous en dira plus.

## Bibliographie

- Akobeng, A. K. (2006). Understanding diagnostic tests 2: Likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Pædiatrica*, 96, 487-491. <https://doi.org/10.1111/j.1651-2227.2006.00179.x>
- Alipour, F., Scherer, R. C., & Finnegan, E. (2012). Measures of spectral slope using an excised larynx model. *Journal of Voice*, 26(4), 403-411. <https://doi.org/10.1016/j.jvoice.2011.07.002>
- American Speech-Language-Hearing Association. (2011, 4 mars). *ASHA Cape-V Form*. Récupéré le 22 octobre 2016 du site de l'association : <https://www.asha.org/Form/CAPE-V/>
- American Speech-Language-Hearing Association. (2017a, 26 septembre). *Adult Hearing Screening*. Récupéré le 10 mars 2018 du site Practice Portal : [https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589942721&section=Key\\_Issues](https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589942721&section=Key_Issues)
- American Speech-Language-Hearing Association. (2017b, 21 février). *Evidence-Based Practice (EBP)*. Récupéré le 9 mars 2017 du site de l'association : <http://www.asha.org/Research/EBP/>
- Andale. (2016, 10 novembre). *Fisher Z-Transformation*. Récupéré le 11 mars 2017 du site Statistics How To : <http://www.statisticshowto.com/fisher-z/>
- Autesserre, D., Charpy, N., Crevier-Buchman, L., Dejonckere, P., Eluard, F., Fresnel, E., . . . Timmermans, B. (2006). *La voix : Ses troubles chez les enseignants*. Paris, France : Les éditions INSERM.
- Awan, S. N., & Ensslen, A. J. (2010). A comparison of trained and untrained vocalists on the dysphonia severity index. *Journal of Voice*, 24(6), 661-666. <https://doi.org/10.1016/j.jvoice.2009.04.001>

- Awan, S. N., & Lawson, L. L. (2009). The effect of anchor modality on the reliability of vocal severity ratings. *Journal of Voice*, 23(3), 341-352. <https://doi.org/10.1016/j.jvoice.2007.10.006>
- Awan, S. N., & Roy, N. (2005). Acoustic prediction of voice type in women with functional dysphonia. *Journal of Voice*, 19(2), 268-282. <https://doi.org/10.1016/j.jvoice.2004.03.005>
- Awan, S. N., & Roy, N. (2006). Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. *Clinical Linguistics & Phonetics*, 20(1), 35-49. <https://doi.org/10.1080/02699200400008353>
- Awan, S. N., & Roy, N. (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research*, 52(2), 482-499. [https://doi.org/10.1044/1092-4388\(2009/08-0034\)](https://doi.org/10.1044/1092-4388(2009/08-0034))
- Awan, S. N., Roy, N., & Cohen, S. M. (2014). Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). *Journal of Voice*, 28(4), 430-439. <http://doi.org/10.1016/j.jvoice.2013.12.008>
- Awan, S. N., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical Linguistics & Phonetics*, 23(11), 825-841. <https://doi.org/10.3109/02699200903242988>
- Awan, S. N., Roy, N., Jetté, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics*, 24(9), 742-758. <https://doi.org/10.3109/02699206.2010.492446>
- Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M. (2016). Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a screening tool for voice disorders: Development of clinical cutoff scores. *Journal of Voice*, 30(2), 130-144. <http://doi.org/10.1016/j.jvoice.2015.04.009>



- Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *The Journal of the Acoustical Society of America*, 65(5), 1271-1275. <https://dx.doi.org/10.1121/1.382795>
- Baken, R. J., & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice* (2<sup>e</sup> éd.). Boston, MA : Cengage Learning.
- Barkat-Defradas, M., Busseuil, C., Chauvy, O., Hirsch, F., Fauth, C., Révis, J., & Amy de la Bretèque, B. (2012). Dimension esthétique des voix normales et dysphoniques : Approches perceptive et acoustique. *Travaux interdisciplinaires sur la parole et le langage*, 28, 1-19. <https://doi.org/10.4000/tipa.204>
- Barsties, B., & De Bodt, M. (2015). Assessment of voice quality: Current state-of-the-art. *Auris Nasus Larynx*, 42, 183-188. <https://doi.org/10.1016/j.anl.2014.11.001>
- Barsties, B., & Lehnert, B. (2018). Interne Validität des Acoustic Voice Quality Index Version 03.01 und des Acoustic Breathiness Index. *Laryngo-Rhino-Otologie*. <https://doi.org/10.1055/a-0596-7819>
- Barsties, B., & Maryn, Y. (2012). Der Acoustic Voice Quality Index in Deutsch: Ein Messverfahren zur allgemeinen Stimmqualität. *HNO*, 60, 715-720. <https://doi.org/10.1007/s00106-012-2499-9>
- Barsties, B., & Maryn, Y. (2013). Test-Retest-Variabilität und interne Konsistenz des Acoustic Voice Quality Index. *HNO*, 61(5), 399-403. <https://doi.org/10.1007/s00106-012-2649-0>
- Barsties, B., & Maryn, Y. (2015). The improvement of the internal consistency of the Acoustic Voice Quality Index. *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, 36(5), 647-656. <https://doi.org/10.1016/j.amjoto.2015.04.012>
- Barsties, B., & Maryn, Y. (2016). External validation of the Acoustic Voice Quality Index version 03.01 with extended representativity. *Annals of Otolaryngology, Rhinology & Laryngology*, 125(7), 571-583. <https://doi.org/10.1177/0003489416636131>

- Barsties, B., & Maryn, Y. (2017). The influence of voice sample length in the auditory-perceptual judgment of overall voice quality. *Journal of Voice*, 31(2), 202-210. <https://doi.org/10.1016/j.jvoice.2016.07.006>
- Barsties, B., Kropp, J., Dicks, P., Grzondziel, V., & Morsomme, D. (2014). Reliabilit t und Validit t des « Voice Handicap Index (VHI) adaptiert an die Gesangsstimme ». *Laryngo-Rhino-Otologie*, 94(7), 441-446. <https://doi.org/10.1055/s-0034-1387765>
- Barsties, B., Maryn, Y., Gerrits, E., & De Bodt, M. (2017). The Acoustic Breathiness Index (ABI): A multivariate acoustic model for breathiness. *Journal of Voice*, 31(4), 511.e11-511.e27. <https://doi.org/10.1016/j.jvoice.2016.11.017>
- Barsties, B., Ulozaite-Staniene, N., Maryn, Y., Petrauskas, T., & Uloza, V. (sous presse). The influence of gender and age on the Acoustic Voice Quality Index and Dysphonia Severity Index: A normative study. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2017.11.011>
- Beals, A. R. (1976). Strategies of resort to curers in South India. Dans C. M. Leslie ( d.), *Asian medical systems: A comparative study* (p. 184-200). Berkeley, CA : University of California Press.
- Behlau, M., Madazio, G., Moreti, F., Oliveira, G., Dos Santos, L., Paulinelli, B. R., & Couto Junior, E. (2016). Efficiency and cutoff values of self-assessment instruments on the impact of a voice problem. *Journal of Voice*, 30(4), 506.e9-506.e18. <https://doi.org/10.1016/j.jvoice.2015.05.022>
- Behrman, A., Sulica, L., & He, T. (2004). Factors predicting patient perception of dysphonia caused by benign vocal fold lesions. *The Laryngoscope*, 114(10), 1693-1700. <https://doi.org/10.1097/00005537-200410000-00004>
- Boersma, P. (2017, 21 janvier). *Voice 3. Shimmer*. R cup r  le 3 f vrier 2017 du site Phonetic Sciences, Amsterdam : [http://www.fon.hum.uva.nl/praat/manual/Voice\\_3\\_\\_Shimmer.html](http://www.fon.hum.uva.nl/praat/manual/Voice_3__Shimmer.html)

- Boersma, P., & Weenink, D. (2003, 10 juin). *Harmonicity*. Récupéré le 28 février 2018 du site Phonetic Sciences, Amsterdam : <http://www.fon.hum.uva.nl/praat/manual/Harmonicity.html>
- Boone, D. R., McFarlane, S. C., Von Berg, S. L., & Zraick, R. I. (2010). *The Voice and Voice Therapy* (8<sup>e</sup> éd.). Boston, MA : Allyn & Bacon.
- Boseley, M. E., Cunningham, M. J., Volk, M. S., & Hartnick, C. J. (2006). Validation of the Pediatric Voice-Related Quality-of-Life Survey. *Archives of Otolaryngology - Head & Neck Surgery*, *132*, 717-720. <https://doi.org/10.1001/archotol.132.7.717>
- Branski, R. C., Cukier-Blaj, S., Pusic, A., Cano, S. J., Klassen, A., Mener, D., . . . Kraus, D. H. (2010). Measuring quality of life in dysphonic patients: A systematic review of content development in patient-reported outcomes measures. *Journal of Voice*, *24*(2), 193-198. <https://doi.org/10.1016/j.jvoice.2008.05.006>
- Brinca, L., Batista, A. P., Tavares, A. I., Pinto, P. N., & Araujo, L. (2015). The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *Journal of Voice*, *29*(6), 776.e7-776.e14. <https://doi.org/10.1016/j.jvoice.2015.01.007>
- Carding, P. N., Horsley, I. A., & Docherty, G. J. (1999). A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *Journal of Voice*, *13*(1), 72-104. [https://doi.org/10.1016/S0892-1997\(99\)80063-0](https://doi.org/10.1016/S0892-1997(99)80063-0)
- Carding, P., Carlson, E., Epstein, R., Mathieson, L., & Shewell, C. (2000). Formal perceptual evaluation of voice quality in the United Kingdom. *Logopedics Phoniatrics Vocology Journal*, *20*, 133-138. <https://doi.org/10.1080/14015430050175860>
- Carson, C. P., Ingrisano, D., & Eggleston, D. (2002). The effect of noise on computer-aided measures of voice: A comparison of CSpeechSP and the Multi-Dimensional Voice Program software using the CSL 4300B M-module and Multi-Speech for Windows. *Journal of Voice*, *17*(1), 12-20. [https://doi.org/10.1016/S0892-1997\(03\)00031-6](https://doi.org/10.1016/S0892-1997(03)00031-6)

- Castarède, M.-F., & Konopczynski, G. (2015). *Au commencement était la voix*. Toulouse, France : Erès.
- Centre national de la recherche scientifique. (2016, mai). *Glossaire GéoManips*. Récupéré le 15 novembre 2016 du site de l'organisation : <http://www.cnrs.fr/cw/dossiers/dosgeol/glossaire/develop.html>
- Chan, R. W., & Titze, I. R. (2006). Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics. *The Journal of the Acoustical Society of America*, 119(4), 2351-2362. <https://doi.org/10.1121/1.2173516>
- Chial, M. R. (2003). *Suggestions for computer-based audio recording of speech samples for perceptual and acoustic analyses*. Madison, WI : University of Wisconsin-Madison.
- Chippaux, J.-P. (2004). *Pratique des essais cliniques en Afrique*. Marseille, France : IRD Éditions.
- Claassen, J. (2005). The gold standard: Not a golden standard. *British Medical Journal*, 330, 1121. <https://doi.org/10.1136/bmj.330.7500.1121>
- Clark, A., Rosen, M., Annie, S., Lee, M., Jamie, O., Thomas, Z., & Thomas, M. (2004). Development and validation of the Voice Handicap Index-10. *The Laryngoscope*, 114, 1549-1556. <https://doi.org/10.1097/00005537-200409000-00009>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>e</sup> éd.). Hillsdale, NJ : Erlbaum Associates.
- Combescure, P. (1981). 20 listes de dix phrases phonétiquement équilibrées. *Revue d'acoustique*, 56, 34-38.
- Cornut, G. (2014). *La voix* (8<sup>e</sup> éd.). Paris, France : Presses Universitaires de France.
- Da Costa de Ceballos, A. G., Carvalho, F. M., De Araujo, T. M., & Farias Borges dos Reis, E. J. (2010). Diagnostic validity of Voice Handicap Index-10 (VHI-10) compared with perceptive-auditory and acoustic speech pathology evaluations of the voice. *Journal of Voice*, 24(6), 715-718. <https://doi.org/10.1016/j.jvoice.2009.03.006>

- Deary, I. J., Wilson, J. A., Carding, P. N., & MacKenzie, K. (2003). VoiSS: A patient-derived Voice Symptom Scale. *Journal of Psychosomatic Research*, 54, 483-489. [https://doi.org/10.1016/S0022-3999\(02\)00469-5](https://doi.org/10.1016/S0022-3999(02)00469-5)
- Dejonckere, P. H. (2010). Assessment of voice and respiratory function. Dans M. Remacle, & H. E. Eckel (Éds.), *Surgery of Larynx and Trachea* (p. 11-26). Berlin, Allemagne : Springer-Verlag.
- Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., . . . 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: Guideline elaborated by the Committee on Phoniatics of the European Laryngological Society (ELS). *European Archives of Oto-Rhino-Laryngology*, 258, 77-82. <https://doi.org/10.1007/s004050000299>
- Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J., & Ramirez, J. (2005). La courbe ROC (receiver operating characteristic) : Principes et principales applications en biologie clinique. *Annales de Biologie Clinique*, 63(2), 145-154.
- Delgado, J., Gómez, L., Jiménez, A., Izquierdo, L., & Barsties, B. (2018). Validation of the Acoustic Voice Quality Index version 03.01 and the Acoustic Breathiness Index in the spanish language. *The Annals of Otolaryngology, Rhinology, and Laryngology*, 127(5), 317-326. <https://doi.org/10.1177/0003489418761096>
- Delgado, J., León, N., Jiménez, A., & Izquierdo, L. (2017). Acoustic analysis of the voice: Temporal, spectral and cepstral measurements in the normal voice with the Praat in a sample of Spanish speakers. *Revista de Investigacion en Logopedia*, 7(2), 108-127.
- Delgado, J., León, N., Izquierdo, L., & Llanos, Y. (2018). Cepstral analysis of normal and pathological voice in Spanish adults. Smoothed cepstral peak prominence in sustained

- vowels versus connected speech. *Acta Otorrinolaringológica Española*, 69(3), 134-140. <https://doi.org/10.1016/j.otorri.2017.05.006>
- Deliyski, D., Shaw, H., & Evans, M. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, 19(1), 15-28. <https://doi.org/10.1016/j.jvoice.2004.07.003>
- DiLollo, A., & Favreau, C. (2010). Person-centered care and speech and language therapy. *Seminars in Speech and Language*, 31(2), 90-97. <https://doi.org/10.1055/s-0030-1252110>
- Draugiem Group. (2014, 20 août). *The secret of the 10% most productive people? Breaking!* Récupéré le 28 février 2018 du site DeskTime : <https://deskttime.com/blog/17-52-ratio-most-productive-people/>
- Eadie, T. L., & Doyle, P. C. (2005). Classification of dysphonic voice: Acoustic and auditory-perceptual measures. *Journal of Voice*, 19(1), 1-14. <https://doi.org/10.1016/j.jvoice.2004.02.002>
- Eadie, T. L., & Kapsner-Smith, M. (2011). The effect of listener experience and anchors on judgments of dysphonia. *Journal of Speech, Language, and Hearing Research*, 54, 430-447. [https://doi.org/10.1044/1092-4388\(2010/09-0205\)](https://doi.org/10.1044/1092-4388(2010/09-0205))
- Facultés universitaires Notre-Dame de la Paix. (2009, 7 juin). *La régression pas à pas (stepwise)*. Récupéré le 24 janvier 2018 du site Pratique des biostatistiques : <http://webapps.fundp.ac.be/biostats/biostat/modules/module230/page6.html>
- Fant, G. (1982). Preliminaries to analysis of the human voice source. *STL-QPSR*, 23(4), 1-27.
- Fausti, S. A., Wilmington, D. J., Helt, P. V., Helt, W. J., & Konrad-Martin, D. (2005). Hearing health and care: The need for improved hearing loss prevention and hearing conservation practices. *Journal of Rehabilitation Research & Development*, 42(4), 45-62.
- Ferrer, C., De Bodt, M., Maryn, Y., Van de Heyning, P., & Hernandez-Diaz, M. (2007). Properties of the cepstral peak prominence and its usefulness in vocal quality

- measurements. *5th international workshop on models and analysis of vocal emissions for biomedical applications* (p. 93-96). Florence, Italie : Firenze University Press.
- Finck, C. (2005). Structure cordale et pathologies vocales. *Revue de Laryngologie - Otologie - Rhinologie*, 5, 295-300.
- Finck, C. (2008). *Implantation d'acide hyaluronique estérifié lors de la microchirurgie des lésions cordales bénignes* (thèse de doctorat, Université de Liège, Belgique). Récupéré du site ORBi : <http://hdl.handle.net/2268/142631>
- Finck, C. (2017). Les pathologies cordales bénignes de l'adulte. Dans C. Klein-Dallant (Éd.), *De la voix parlée au chant : bilans, rééducations, pathologies de la voix parlée et chantée* (p. 207-235). Paris, France : Klein-Dallant.
- Fraile, R., & Godino-Llorente, J. I. (2014). Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, 14, 42-54. <https://doi.org/10.1016/j.bspc.2014.07.001>
- Fredouille, C., Pouchoulin, G., Ghio, A., Révis, J., Bonastre, J.-F., & Giovanni, A. (2009). Back-and-forth methodology for objective voice quality assessment: From/to expert knowledge to/from automatic classification of dysphonia. *EURASIP Journal on Advances in Signal Processing*, 2009(1), 1-13. <https://doi.org/10.1155/2009/982102>
- Friedrich, G., & Dejonckere, P. H. (2005). Das Stimm diagnostik-Protokoll der European Laryngological Society (ELS) : Erste Erfahrungen im Rahmen einer Multizenterstudie. *Laryngo-Rhino-Otologie*, 84(10), 744-752. <https://doi.org/10.1055/s-2005-861450>
- Gabriel, F. (2012). *1.1. Étude et perfectionnement de la voix, de l'articulation, de la parole et du langage* (support de cours). Liège, Belgique : Haute École de la Ville de Liège.
- Gabriel, F. (2014). *3.6. Troubles de la voix* (support de cours). Liège, Belgique : Haute École de la Ville de Liège.

- Ghio, A., Pouchoulin, G., Giovanni, A., Fredouille, C., Teston, B., Révis, J., . . . Marchal, A. (2007). Approches complémentaires pour l'évaluation des dysphonies : Bilan méthodologique et perspectives. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence*, 26, 33-74.
- Gillespie, A. I., Gooding, W., Rosen, C., & Gartner-Schmidt, J. (2014). Correlation of VHI-10 to voice laboratory measurements across five common voice disorders. *Journal of Voice*, 28(4), 440-448. <https://doi.org/10.1016/j.jvoice.2013.10.023>
- Giordano, A., Dejonckere, P., & Manfredi, C. (2011). Acoustic assessment of spasmodic dysphonia using a new multipurpose voice analysis tool. *7th international workshop on models and analysis of vocal emissions for biomedical applications* (p. 147-150). Florence, Italie : Firenze University Press.
- Giovanni, A. (2004). *Le bilan d'une dysphonie : État actuel et perspectives*. Marseille, France : Solal.
- Giovanni, A., & Lagier, A. (2009). Anatomie et physiologie de la phonation. Dans P. Dugero, & M. Remacle (Éds.), *Précis d'audiophonologie et de déglutition* (vol. 2, p. 27-42). Marseille, France : Solal.
- Giovanni, A., Robert, D., Estublier, N., Teston, B., Zanaret, M., & Cannoni, M. (1996). Objective evaluation of dysphonia: Preliminary results of a device allowing simultaneous acoustic and aerodynamic measurements. *Folia Phoniatica et Logopaedica*, 48(4), 175-185. <https://doi.org/10.1159/000266406>
- Gliklich, R. E., Glosky, R. M., & Montgomery, W. W. (1999). Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngology - Head and Neck Surgery*, 120, 153-158. [https://doi.org/10.1016/S0194-5998\(99\)70399-2](https://doi.org/10.1016/S0194-5998(99)70399-2)
- Goldberg, R., & Riek, L. (2000). *A practical handbook of speech coders*. Boca Raton, FL : CRC Press.



- Goujon, F. (2012). *Audiométrie vocale : Étude de l'intelligibilité dans le bruit chez le normo-entendant et détermination de courbes vocales de référence* (thèse de maîtrise, Université de Lorraine, Nancy). Récupéré de [http://docnum.univ-lorraine.fr/public/BUPHA\\_MAUDIO\\_2012\\_GOUJON\\_FLORIAN.pdf](http://docnum.univ-lorraine.fr/public/BUPHA_MAUDIO_2012_GOUJON_FLORIAN.pdf)
- Gräbel, E., Hoppe, U., & Rosanowski, F. (2009). Graduierung des Voice-Related-Quality-of-Life-Index. *HNO*, 57, 896-901. <https://doi.org/10.1007/s00106-009-1953-9>
- GraphPad Software Inc. (2016, 4 juin). *GraphPad Statistics Guide - Interpreting results: Correlation*. Récupéré le 14 février 2017 du site GraphPad Software : [http://www.graphpad.com/guides/prism/6/statistics/index.htm?stat\\_interpreting\\_results\\_correlati.htm](http://www.graphpad.com/guides/prism/6/statistics/index.htm?stat_interpreting_results_correlati.htm)
- Guinness World Records. (2017, 5 octobre). *Greatest vocal range, male*. Récupéré le 14 janvier 2018 du site Guinness World Records : <http://www.guinnessworldrecords.com/world-records/greatest-vocal-range-male>
- Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: The case of tests with continuous results. *Biochemia Medica*, 26(3), 297-307. <https://doi.org/10.11613/BM.2016.034>
- Hakkestegt, M. M., Brocaar, M. P., & Wieringa, M. H. (2010). The applicability of the Dysphonia Severity Index and the Voice Handicap Index in evaluating effects of voice therapy and phonosurgery. *Journal of Voice*, 24(2), 199-205. <https://doi.org/10.1016/j.jvoice.2008.06.007>
- Hakkestegt, M. M., Brocaar, M. P., Wieringa, M. H., & Feenstra, L. (2008). The relationship between perceptual evaluation and objective multiparametric evaluation of dysphonia severity. *Journal of Voice*, 22(2), 138-145. <https://doi.org/10.1016/j.jvoice.2006.09.010>

- Hakkesteeft, M., Brocaar, M., Wieringa, M., & Feenstra, L. (2006). Influence of age and gender on the Dysphonia Severity Index: A study of normative values. *Folia Phoniatica et Logopaedica*, 58, 264-273. <https://doi.org/10.1159/000093183>
- Harmegnies, B., & Landercy, A. (1988). Intra-speaker variability of the long-term speech spectrum. *Speech communication*, 7(1), 81-86. [https://doi.org/10.1016/0167-6393\(88\)90023-4](https://doi.org/10.1016/0167-6393(88)90023-4)
- Hart, P. (2016). Receiver Operating Characteristic (ROC) curve analysis: A tutorial using Body Mass Index (BMI) as a measure of obesity. *Journal of Physical Activity Research*, 1(1), 5-8. <http://doi.org/10.12691/jpar-1-1-2>
- Hartnick, C. J., Volk, M., & Cunningham, M. (2003). Establishing normative voice-related quality of life scores within the pediatric otolaryngology population. *Archives of Otolaryngology - Head & Neck Surgery*, 129, 1090-1093. <https://doi.org/10.1001/archotol.129.10.1090>
- Hauptmann, J. H. (2013). *Die Bedeutung des Acoustic Voice Quality Index (AVQI) bei der Beurteilung der Stimmfunktion*. Norderstedt, Allemagne : GRIN Verlag.
- Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., . . . Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otolaryngology, Rhinology & Laryngology*, 112(4), 324-33. <https://doi.org/10.1177/000348940311200406>
- Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1), 20-27. [https://doi.org/10.1016/S0892-1997\(02\)00067-X](https://doi.org/10.1016/S0892-1997(02)00067-X)
- Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., . . . Hillenbrand, J. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice*, 28(6), 783-788. <https://doi.org/10.1016/j.jvoice.2014.05.005>

- Hemler, R., Wieneke, G., & Dejonckere, P. (1997). The effect of relative humidity of inhaled air on acoustic parameters of voice in normal subjects. *Journal of Voice*, 11(3), 295-300. [https://doi.org/10.1016/S0892-1997\(97\)80007-0](https://doi.org/10.1016/S0892-1997(97)80007-0)
- Hemler, R., Wieneke, G., Lebacqz, J., & Dejonckere, P. (2001). Laryngeal mucosa elasticity and viscosity in high and low relative air humidity. *European Archives of Oto-Rhino-Laryngology*, 258, 125-129. <https://doi.org/10.1007/s004050100321>
- Henrich, N. (2012). Physiologie de la voix chantée : Vibrations laryngées et adaptations phonorésonnantes. *40es entretiens de médecine physique et de réadaptation* (p. 17-32). Montpellier, France : Sauramps Médical.
- Herbst, C. T., Hess, M., Müller, F., Svec, J. G., & Sundberg, J. (2015). Glottal adduction and subglottal pressure in singing. *Journal of Voice*, 29(4), 391-402. <https://doi.org/10.1016/j.jvoice.2014.08.009>
- Heuer, R. J., Hawkshaw, M. J., & Sataloff, R. T. (2006). The clinical voice laboratory. Dans R. T. Sataloff (Éd.), *Vocal health and pedagogy : Science and assessment* (2<sup>e</sup> éd., vol. 1, p. 159-198). San Diego, CA : Plural Publishing.
- Hillenbrand, J. M. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research*, 30(4), 448-461. <https://doi.org/10.1044/jshr.3004.448>
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39, 311-321. <https://doi.org/10.1044/jshr.3902.311>
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37, 769-778. <https://doi.org/10.1044/jshr.3704.769>

- Hinkle, D., Wiersma, W., & Jurs, S. (1994). *Applied statistics for the behavioral sciences* (3<sup>e</sup> éd., vol. 663). Boston, MA : Houghton Mifflin.
- Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure Voice-Related Quality of Life (V-RQOL). *Journal of Voice*, *13*(4), 557-569. [https://doi.org/10.1016/S0892-1997\(99\)80010-1](https://doi.org/10.1016/S0892-1997(99)80010-1)
- Hosokawa, K., Barsties, B., Iwahashi, T., Iwahashi, M., Kato, C., Iwaki, S., . . . Maryn, Y. (2017). Validation of the Acoustic Voice Quality Index in the Japanese language. *Journal of Voice*, *31*(2), 260.e1-260.e9. <http://doi.org/10.1016/j.jvoice.2016.05.010>
- Hosokawa, K., Barsties, B., Iwahashi, T., Iwahashi, M., Iwaki, S., Kato, C., . . . Maryn, Y. (sous presse). The Acoustic Voice Quality Index version 03.01 for the Japanese-speaking population. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2017.10.003>
- Ikuma, T., Kunduk, M., Fink, D., & McWhorter, A. J. (2016). A spatiotemporal approach to the objective analysis of initiation and termination of vocal-fold oscillation with high-speed videoendoscopy. *Journal of Voice*, *30*(6), 756.e21-756.e30. <https://doi.org/10.1016/j.jvoice.2015.09.007>
- Ingrisano, D., Perry, C., & Jepsen, K. (1998). Environmental noise: A threat to automatic voice analysis. *American Journal of Speech-Language Pathology*, *7*, 91-96. <https://doi.org/10.1044/1058-0360.0701.91>
- Jacobson, B., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M., & Newman, C. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology*, *6*(3), 66-69. <https://doi.org/10.1044/1058-0360.0603.66>
- Jakobsson, U., & Westergren, A. (2005). Statistical methods for assessing agreement for ordinal data. *Scandinavian Journal of Caring Sciences*, *19*(4), 427-431. <https://doi.org/10.1111/j.1471-6712.2005.00368.x>

- Jiang, J. (2015). Physiology of voice production: How does the voice work? Dans M. S. Benninger, T. Murry, & M. M. Johns (Éds.), *The Performer's Voice* (2<sup>e</sup> éd., p. 39-52). San Diego, CA : Plural Publishing.
- Kankare, E., Barsties, B., Maryn, Y., Ilomäki, I., Laukkanen, A.-M., Tyrmi, J., . . . Vilpas, S. (2015, août). *A preliminary study of the Acoustic Voice Quality Index in finnish speaking population*. Communication présentée à la 11<sup>e</sup> conférence PEVOC - Pan-European Voice Conference, Florence, Italie.
- Karnell, M. P., Melton, S. D., Childes, J. M., Coleman, T. C., Dailey, S. A., & Hoffman, H. T. (2007). Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Journal of Voice*, *21*(5), 576-590. <https://doi.org/10.1016/j.jvoice.2006.05.001>
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, *18*, 124-132. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. Boston, MA : Little, Brown and Company.
- Kotrlik, J. W., Williams, H. A., & Khata Jabor, M. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education*, *52*(1), 132-142. <http://doi.org/10.5032/jae.2011.01132>
- Kreiman, J., & Gerratt, B. R. (1998). Validity of rating scale measures of voice quality. *The Journal of the Acoustical Society of America*, *104*(3), 1598-1608. <https://doi.org/10.1121/1.424372>
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, *36*, 21-40. <https://doi.org/10.1044/jshr.3601.21>

- Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research*, 35(3), 512-520. <https://doi.org/10.1044/jshr.3503.512>
- Ladefoged, P. (1995). *Elements of acoustic phonetics* (2<sup>e</sup> éd.). Chicago, IL : University of Chicago Press.
- Lalkhen, A. G., McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*, 8(6), 221-223. <http://doi.org/10.1093/bjaceaccp/mkn041>
- Le Breton, D. (2011). *Éclats de voix : Une anthropologie des voix*. Paris, France : Éditions Métailié.
- Lemm, L. (2012). *Zur Wertigkeit videostroboskopischer und lupenlaryngoskopischer Tonaufnahmen für die objektive Stimmanalyse* (thèse de doctorat, Georg-August-Universität zu Göttingen, Allemagne). Récupéré du site Wevosys : [http://www.wevosys.de/wissen/\\_data\\_wissen/156.pdf](http://www.wevosys.de/wissen/_data_wissen/156.pdf)
- Leong, K., Hawkshaw, M. J., Dentchev, D., Gupta, R., Lurie, D., & Sataloff, R. T. (2012). Reliability of objective voice measures of normal speaking voices. *Journal of Voice*, 22(2), 170-176. <https://doi.org/10.1016/j.jvoice.2012.07.005>
- Linder, R., Albers, A. E., Hess, M., Pöpl, S. J., & Schönweiler, R. (2008). Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features. *Journal of Voice*, 22(2), 155-163. <https://doi.org/10.1016/j.jvoice.2006.09.003>
- Lopes, L. W., Da Silva, J. D., Simões, L. B., Evangelista, D. d., Silva, P. O., Almeida, A. A., & De Lima-Silva, M. F. (2017). Relationship between acoustic measurements and self-evaluation in patients with voice disorders. *Journal of Voice*, 31(1), 119.e1-119.e10. <https://doi.org/10.1016/j.jvoice.2016.02.021>

- Lowell, S., Colton, R., Kelley, R., & Hahn, Y. (2011). Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25(5), e223-e232. <https://doi.org/10.1016/j.jvoice.2010.06.007>
- Lucero, J. C. (1996). Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis. *The Journal of the Acoustical Society of America*, 100(4), 2551-2554. <https://doi.org/10.1121/1.417365>
- Ma, E. P., & Yiu, E. M. (2001). Voice Activity and Participation Profile : Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language and Hearing Research*, 44(3), 511-524. [https://doi.org/10.1044/1092-4388\(2001/040\)](https://doi.org/10.1044/1092-4388(2001/040))
- Maertens, K., & De Jong, F. (2007). The voice handicap index as a tool for assessment of the biopsychosocial impact of voice problems. *B-ENT*, 3(2), 61-66.
- Maltby, M. A., & Knight, P. (2015). *Audiology: An introduction for teachers & other professionals*. New York, NY : Routledge.
- Martin, S., & Lockhart, M. (2005). *VIP: Voice Impact Profile*. Bicester, Angleterre : Speechmark Publishing.
- Maryn, Y. (2010). *Acoustic measurement of overall voice quality in sustained vowels and continuous speech* (thèse de doctorat, Université de Gand, Belgique). Récupéré de <http://hdl.handle.net/1854/LU-888156>
- Maryn, Y. (2017). *Cepstrography (v.01.01)*. Récupéré le 17 février 2018 du site Phonanium : <https://www.phonanium.com/product/cepstrography-v-01-01/>
- Maryn, Y., & Roy, N. (2012). Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. *Jornal da Sociedade Brasileira de Fonoaudiologia*, 24(2), 107-112. <https://doi.org/10.1590/S2179-64912012000200003>

- Maryn, Y., & Weenink, D. (2015). Objective dysphonia measures in the program Praat: Smoothed cepstral peak prominence and Acoustic Voice Quality Index. *Journal of Voice*, 29(1), 35-43. <https://doi.org/10.1016/j.jvoice.2014.06.015>
- Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2009). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice*, 24(5), 540-555. <https://doi.org/10.1016/j.jvoice.2008.12.014>
- Maryn, Y., De Bodt, M., & Roy, N. (2010). The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders. *Journal of Communication Disorders*, 43(3), 161-174. <https://doi.org/10.1016/j.jcomdis.2009.12.004>
- Maryn, Y., De Bodt, M., Barsties, B., & Roy, N. (2014). The value of the Acoustic Voice Quality Index as a measure of dysphonia severity in subjects speaking different languages. *European Archives of Oto-Rhino-Laryngology*, 271(6), 1609-1619. <https://doi.org/10.1007/s00405-013-2730-7>
- Maryn, Y., Dick, C., Vandenbruaene, C., Vauterin, T., & Jacobs, T. (2009). Spectral, cepstral and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels. *Laryngoscope*, 119(12), 2384-2394. <https://doi.org/10.1002/lary.20620>
- Maryn, Y., Kim, H.-T., & Kim, J. (2016). Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech. *Journal of Voice*, 30(5), 587-594. <https://doi.org/10.1016/j.jvoice.2015.06.011>
- Maryn, Y., Morsomme, D., & De Bodt, M. (2017). Measuring the Dysphonia Severity Index (DSI) in the program Praat. *Journal of Voice*, 31(5), 644.e29-644.e40. <https://doi.org/10.1016/j.jvoice.2017.01.002>



- Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America*, 126(5), 2619-2634. <https://doi.org/10.1121/1.3224706>
- Master, S., De Biase, N., Pedrosa, V., & Chiari, B. M. (2006). The long-term average spectrum in research and in the clinical practice of speech therapists. *Pró-Fono Revista de Atualização Científica*, 18(1), 111-120. <http://doi.org/10.1590/S0104-56872006000100013>
- Mathieson, L. (2013). *Greene and Mathieson's the voice and its disorders* (6<sup>e</sup> éd.). Hoboken, NJ : John Wiley & Sons.
- Mayer, J. (2011, 13 mai). *Praat Skripte : Auditive Stimmanalyse (GRBAS) mit dem Demo-Window*. Récupéré le 2 janvier 2018 du site Die Praatpfanne : [http://praatpfanne.lingphon.net/downloads/demo\\_GRBAS.txt](http://praatpfanne.lingphon.net/downloads/demo_GRBAS.txt)
- Mayer, J. (2013, 15 novembre). *Praat Skripte : Sound-Objekte zusammenfügen*. Récupéré le 3 janvier 2017 du site Die Praatpfanne : <http://praatpfanne.lingphon.net/downloads/concatenateWithPause.txt>
- McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276-282. <https://doi.org/10.11613/BM.2012.031>
- Méchain, C. (2014). *Utilisation du triangle vocalique dans l'apprentissage de la voix œsophagienne* (thèse de maîtrise, Université de Bordeaux, France). Récupéré du site DUMAS : <https://dumas.ccsd.cnrs.fr/dumas-01081035/document>
- Mehta, D. D., & Hillman, R. E. (2008). Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 16(3), 211-215. <https://doi.org/10.1097/MOO.0b013e3282fe96ce>
- Mendelsohn, A. H., & Zhang, Z. (2011). Phonation threshold pressure and onset frequency in a two-layer physical model of the vocal folds. *The Journal of the Acoustical Society of America*, 130(5), 2961-2968. <https://doi.org/10.1121/1.3644913>

- Meulemans, T. (2016). *PSYC5872-1 Neuropsychologie* (support de cours). Liège, Belgique : Presses Universitaires de Liège.
- Michaelis, D. (1999). *Das Göttinger Heiserkeits-Diagramm : Entwicklung und Prüfung eines akustischen Verfahrens zur objektiven Stimmgütebeurteilung pathologischer Stimmen* (thèse de doctorat, Georg-August-Universität zu Göttingen, Allemagne). Récupéré du site eDiss : [https://ediss.uni-goettingen.de/bitstream/handle/11858/00-1735-0000-0022-5D57-6/michaelis\\_re.pdf?sequence=1](https://ediss.uni-goettingen.de/bitstream/handle/11858/00-1735-0000-0022-5D57-6/michaelis_re.pdf?sequence=1)
- Moreti, F., Zambon, F., Oliveira, G., & Behlau, M. (2014). Cross-cultural adaptation, validation, and cutoff values of the brazilian version of the Voice Symptom Scale - VoiSS. *Journal of Voice*, 28(4), 458-468. <https://doi.org/10.1016/j.jvoice.2013.11.009>
- Morsomme, D. (2001). *Contribution à la détermination de paramètres subjectifs et objectifs pour l'étude de la voix* (thèse de doctorat, Université Catholique de Louvain, Belgique). Récupéré du site Orbi : [https://orbi.uliege.be/bitstream/2268/26003/1/Morsomme\\_Dominique\\_THESE2001.pdf](https://orbi.uliege.be/bitstream/2268/26003/1/Morsomme_Dominique_THESE2001.pdf)
- Morsomme, D., & Remacle, A. (2013). La charge vocale. *Rééducation Orthophonique*, 254, 85-102.
- Morsomme, D., Gaspar, M., Jamart, J., Remacle, M., & Verduyck, I. (2007). Adaptation du Voice Handicap Index à la voix chantée. *Revue de Laryngologie - Otologie - Rhinologie*, 128(5), 305-314.
- National Center for Voice and Speech. (2016, 8 septembre). *3 Levels of Intensity Control: Lungs, Glottis & Vocal Tract*. Récupéré le 14 mars 2017 du site de l'organisation : <http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/three.html>
- National Center for Voice and Speech. (2016, 9 octobre). *Voice Qualities*. Récupéré le 16 mars 2017 du site de l'organisation : <http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/quality.html>

- Nemr, K., Simões-Zenari, M., Cordeiro, G., Tsuji, D., Ogawa, A., Ubrig, M., & Menezes, M. (2012). GRBAS and Cape-V Scales: High reliability and consensus when applied at different times. *Journal of Voice*, 26(6), 812.e17-812.e22. <https://doi.org/10.1016/j.jvoice.2012.03.005>
- Noll, A. M. (1964). Short-time spectrum and “cepstrum” techniques for vocal-pitch detection. *The Journal of the Acoustical Society of America*, 36(2), 296-302. <https://doi.org/10.1121/1.1918949>
- Noll, A. M. (1967). Cepstrum pitch determination. *The Journal of the Acoustical Society of America*, 41(2), 293-309. <https://doi.org/10.1121/1.1910339>
- Omori, K. (2011). Diagnosis of voice disorders. *Japan Medical Association Journal*, 54(4), 248-253.
- Oppenheim, A. V., & Schaffer, R. W. (2004). From frequency to quefrency: A history of the cepstrum. *IEEE Signal Processing Magazine*, 21(5), 95-106. <https://doi.org/10.1109/MSP.2004.1328092>
- Organisation mondiale de la Santé. (1993). Study protocol for the World Health Organization project to develop a quality of life assessment instrument (WHOQOL). *Quality of Life Research*, 2, 153-159. <https://doi.org/10.1007/BF00435734>
- Organisation mondiale de la Santé. (2001). *Classification internationale du fonctionnement, du handicap et de la santé*. Genève, Suisse : Organisation Mondiale de la Santé.
- Pandey, M., Jain, A. (2016, mai). ROC Curve: Making way for correct diagnosis. Dans *Proceedings of the 6<sup>th</sup> PharmaSUG Conference*. Denver, CO : PharmaSUG. Récupéré du site PharmaSUG : <https://www.pharmasug.org/2016-proceedings.html>
- Peiyun, Z., Sprecher, A. J., Hoffman, M. R., Zhang, Y., Fourakis, M., Jiang, J. J., & Wei, C. S. (2009). Phonation threshold flow measurements in normal and pathological phonation. *Laryngoscope*, 119(4), 811-815. <https://doi.org/10.1002/lary.20165>

- Perinetti, G. (2017). StaTips Part III: Assessment of the repeatability and rater agreement for nominal and ordinal data. *South European Journal of Orthodontics and Dentofacial Research*, 4(1), 3-4. <https://doi.org/10.5937/sejodr4-15199>
- Perry, C., Ingrisano, D., Palmer, M., & McDonald, E. J. (2000). Effects of environmental noise on computer-derived voice estimates from female speakers. *Journal of Voice*, 14(2), 146-153. [https://doi.org/10.1016/S0892-1997\(00\)80021-1](https://doi.org/10.1016/S0892-1997(00)80021-1)
- Peterson, E. A., Roy, N., Awan, S. N., Merrill, R. M., Banks, R., & Tanner, K. (2013). Toward validation of the Cepstral Spectral Index of Dysphonia (CSID) as an objective treatment outcomes measure. *Journal of Voice*, 27(4), 401-410. <https://doi.org/10.1016/j.jvoice.2013.04.002>
- Pillot-Loiseau, C. (2011). Pression sous-glottique et débit oral d'air expiré comme aides à la pose du diagnostic de dysodie ; implications pour la rééducation vocale. *Entretiens d'orthophonie*, 32-45.
- Pouchoulin, G. (2008). *Approche statistique pour l'analyse objective et la caractérisation de la voix dysphonique* (thèse de doctorat, Université d'Avignon et des Pays de Vaucluse, France). Récupéré du site Archives ouvertes HAL : <https://hal.archives-ouvertes.fr/tel-01472450>
- Poulain, J.-C., & Sauvignet-Poulain, A. (2017). La rééducation vocale des immobilités laryngées. Dans C. Klein-Dallant (Éd.), *De la voix parlée au chant : Bilans, rééducations, pathologies de la voix parlée et chantée* (p. 261-274). Paris, France : Klein-Dallant.
- Prahalad, K. (2008, 8 septembre). *Speech technology - A practical introduction: Spectrogram, cepstrum and mel-frequency analysis*. Récupéré le 10 février 2017 sur le site de la Carnegie Mellon University : [http://www.speech.cs.cmu.edu/15-492/slides/03\\_mfcc.pdf](http://www.speech.cs.cmu.edu/15-492/slides/03_mfcc.pdf)
- Randall, R. (1973). *Cepstrum analysis and gearbox fault diagnosis*. Copenhagen, Danemark : B&K Instruments.

- Remøy, H. T. (2010). *Out of office: A study on the cause of office vacancy and transformation as a means to cope and prevent*. Amsterdam, Pays-Bas: IOS Press.
- Remacle, A., Petitfils, C., Finck, C., & Morsomme, D. (2017). Description of patients consulting the voice clinic regarding gender, age, occupational status, and diagnosis. *European Archives of Oto-Rhino-Laryngology and Head & Neck*, 274(3), 1567-1576. <https://doi.org/10.1007/s00405-016-4332-7>
- Remacle, M., & Lawson, G. (2009). Exploration du larynx. Dans P. Dulguerov, & M. Remacle (Éds.), *Précis d'audiophonologie et de déglutition : Les voies aéro-digestives supérieures* (p. 109-120). Marseille, France : Solal.
- Remacle, M., & Morsomme, D. (2009). Evaluation objective de la voix. Dans P. Dulguerov, & M. Remacle (Éds.), *Précis d'audiophonologie et de déglutition : Les voies aéro-digestives supérieures* (p. 121-134). Marseille, France : Solal.
- Remacle, M., Lawson, G., Giovanni, A., & Woisard, V. (2005). Exploration du larynx. *EMC - Oto-rhino-laryngologie*, 2(4), 401-419. <https://doi.org/10.1016/j.emcorl.2005.10.002>
- Révis, J. (2013). *La voix et soi*. Paris, France : De Boeck-Solal.
- Révis, J., Giovanni, A., Wuyts, F., & Triglia, J.-M. (1999). Comparison of different voice samples for perceptual analysis. *Folia Phoniatica et Logopaedica*, 51, 108-116. <https://doi.org/10.1159/000021485>
- Reynolds, V., Buckland, A., Bailey, J., Lipscombe, J., Nathan, E., Vijayasekaran, S., . . . French, N. (2012). Objective assessment of pediatric voice disorders with the acoustic voice quality index. *Journal of Voice*, 26(5), 672.e1-672.e7. <https://doi.org/10.1016/j.jvoice.2012.02.002>
- Ricci-Maccarini, A., De Maio, V., Murry, T., & Schindler, A. (2013). Development and validation of the Children's Voice Handicap Index-10 (CVHI-10). *Journal of Voice*, 27(2), 258.e23-258.e28. <https://doi.org/10.1016/j.jvoice.2012.10.006>

- Ricci-Maccarini, A., De Maio, V., Murry, T., & Schindler, A. (2016). Development and validation of the Children's Voice Handicap Index-10 for Parents. *Journal of Voice*, 30(1), 120-126. <https://doi.org/10.1016/j.jvoice.2014.10.004>
- Rodrigues Bastilha, G., De Moraes Lima, J. P., & Aparecida Cielo, C. (2014). Influence of gender, age, occupation and phonoaudiological diagnosis in the voice quality of life. *Revista CEFAC*, 16(6), 1900-1907. <http://doi.org/10.1590/1982-0216201415913>
- Rubin, P., & Vatikiotis-Bateson, E. (1998). Measuring and modeling speech production. Dans S. L. Hopp, M. J. Owren, & C. S. Evans (Éds.), *Animal acoustic communication: Sound analysis and research methods* (p. 251-290). Berlin, Allemagne : Springer-Verlag.
- Rullier. (1820). Phonation. Dans C. L. Panckoucke (Éd.), *Dictionnaire des sciences médicales* (Vol. 41, p. 483-484). Paris, France : Panckoucke. Récupéré du site BIU Santé : <http://www.biusante.parisdescartes.fr/histoire/medica/resultats/index.php?do=page&cote=47661x41&p=491>
- Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., & Schisterman, E. F. (2008). Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal*, 50(3), 419-430. <https://doi.org/10.1002/bimj.200710415>
- San Segundo, E., & Mompean, J. A. (2017). A simplified Vocal Profile Analysis protocol for the assessment of voice quality and speaker similarity. *Journal of Voice*, 31(5), 644.e11-644.e27. <https://doi.org/10.1016/j.jvoice.2017.01.005>
- Santos, F. (2018, 4 avril). *Le kappa de Cohen : Un outil de mesure de l'accord interjuges sur des caractères qualitatifs*. Récupéré du site PACEA : [http://www.pacea.u-bordeaux1.fr/IMG/pdf/Kappa\\_Cohen.pdf](http://www.pacea.u-bordeaux1.fr/IMG/pdf/Kappa_Cohen.pdf)
- Satterfield, J. S. (2009). Toward a transdisciplinary model of evidence-based practice. *The Milbank Quarterly*, 87(2), 368-390. <https://doi.org/10.1111/j.1468-0009.2009.00561.x>

- Sauder, C., Bretl, M., & Eadie, T. (2017). Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice*, 31(5), 557-566. <https://doi.org/10.1016/j.jvoice.2017.01.006>
- Schaeffler, F., Beck, J., & Jannetts, S. (2015). Phonation stabilisation time as an indicator of voice disorder. Dans *proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences (ICPhS)*. Glasgow, Écosse : International Phonetic Association. Récupéré du site de l'association : <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0331.pdf>
- Schindler, A., Mozzanica, F., Vedrody, M., Maruzzi, P., & Ottaviani, F. (2009). Correlation between the Voice Handicap Index and voice measurements in four groups of patients with dysphonia. *Otolaryngology - Head and Neck Surgery*, 141, 762-769. <https://doi.org/10.1016/j.otohns.2009.08.021>
- Schoentgen, J. (2003). Spectral models of additive and modulation noise in speech and phonatory excitation signals. *The Journal of the Acoustical Society of America*, 113(1), 553-562. <https://doi.org/10.1121/1.1523384>
- Schoentgen, J. (2016). *Phonétique clinique* (support de cours). Liège, Belgique : Université de Liège.
- Seifpanahi, S., Jalaie, S., Nikoo, M. R., & Sobhani-Rad, D. (2015). Translated versions of Voice Handicap Index (VHI)-30 across languages: A systematic review. *Iranian Journal of Public Health*, 44(4), 458-469.
- Service public wallon de l'emploi et de la formation. (2016, 19 décembre). *Métiers et professions*. Récupéré le 26 janvier 2017 sur le site Le Forem : <https://www.leforem.be/HE/former/horizonemploi/metier/index-metier-categorie.html>
- Shewell, C. (1998). The effect of perceptual training on ability to use the Vocal Profile Analysis Scheme. *International Journal of Language & Communication Disorders*, 33, Suppl:322-326. <https://doi.org/10.3109/13682829809179444>

- Shweta, B., Ram, B., & Himanshu, C. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), 20-27.
- Sicard, A. M., & Sicard, E. (2016). *Évaluation et réhabilitation de la voix : Approche clinique et objective*. Louvain-la-Neuve, Belgique : De Boeck Supérieur.
- Siemons-Lühring, D. I., Moerman, M., Martens, J.-P., Deuster, D., Müller, F., & Dejonckere, P. (2009). Spasmodic dysphonia, perceptual and acoustic analysis: Presenting new diagnostic tools. *European Archives of Oto-Rhino-Laryngology*, 266, 1915-1922. <https://doi.org/10.1007/s00405-009-0995-7>
- Simundić, A.-M. (2009). Measures of diagnostic accuracy: Basic definitions. *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine*, 19(4), 203-211.
- Smith, L. K., & Goberman A. M. (2014). Long-time average spectrum in individuals with Parkinson disease. *NeuroRehabilitation*, 35(1), 77-88. <https://doi.org/10.3233/NRE-141102>
- Statistic Analysis System. (2014, 5 août). *Plot ROC curve with cutpoint labeling and optimal cutpoint analysis*. Récupéré du site de la société : <http://support.sas.com/kb/25/018.html>
- Story, B. H., & Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *The Journal of the Acoustical Society of America*, 97(2), 1249-1260. <https://doi.org/10.1121/1.412234>
- Sundberg, J. (1972). A perceptual function of the 'singing formant'. *STL-QPSR*, 13, 61-63.
- Sundberg, J. (1992). Breathing behavior during singing. *STL-QPSR*, 33(1), 49-64.
- Svec, J. G., & Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10(2), 201-205. [https://doi.org/10.1016/S0892-1997\(96\)80047-6](https://doi.org/10.1016/S0892-1997(96)80047-6)
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293. <http://doi.org/10.1126/science.3287615>



- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis: Jitter, shimmer and HNR parameters. *Procedia Technology*, 9, 1112-1122. <https://doi.org/10.1016/j.protcy.2013.12.124>
- Teston, B. (2004). L'évaluation instrumentale des dysphonies : État actuel et perspectives. Dans A. Giovanni (Éd.), *Le bilan d'une dysphonie : État actuel et perspectives*. Marseille, France : Solal.
- TIBCO Software Inc. (n.d.). *Fisher z' Transformation (Pearson r to z')*. Récupéré le 12 février 2017, sur Statistica Help : <http://documentation.statsoft.com/STATISTICAHelp.aspx?path=glossary/GlossaryTwo/F/FisherzTransformationPearsonrtoz>
- Timmermans, B., De Bodt, M., Wuyts, F., & Van de Heyning, P. (2004). Voice quality change in future professional voice users after 9 months of voice training. *European Archives of Oto-Rhino-Laryngology*, 261, 1-5. <https://doi.org/10.1007/s00405-003-0652-5>
- Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, 85(2), 901-906. <https://doi.org/10.1121/1.397562>
- Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ : Prentice Hall.
- Titze, I. R. (2001). Criteria for occupational risk in vocalization. Dans P. Dejonckere (Éd.), *Occupational voice: Care and cure* (p. 1-10). La Haye, Pays-Bas : Kugler Publications.
- Titze, I. R. (2011). Vocal fold mass is not a useful quantity for describing F0 in vocalization. *Journal of Speech, Language, and Hearing Research*, 54(2), 520-522. [https://doi.org/10.1044/1092-4388\(2010/09-0284\)](https://doi.org/10.1044/1092-4388(2010/09-0284))
- Titze, I. R., Schmidt, S. S., & Titze, M. R. (1995). Phonation threshold pressure in a physical model of the vocal fold mucosa. *The Journal of the Acoustical Society of America*, 97(5), 3080-3084. <https://doi.org/10.1121/1.408054>
- Tubach, J., & Boe, L. (1990). *Un corpus de transcription phonétique (300000 phones) : Constitution et exploitation statistique*. Paris, France : École nationale supérieure des télécommunications.

- Uloza, V., Barsties, B., Ulozaite, N., Petrauskas, T., & Maryn, Y. (2018). A comparison of Dysphonia Severity Index and Acoustic Voice Quality Index measures in differentiating normal and dysphonic voices. *European Archives of Oto-Rhino-Laryngology*, 275(4), 949-958. <https://doi.org/10.1007/s00405-018-4903-x>
- Uloza, V., Petrauskas, T., Padervinskis, E., Ulozaite, N., Barsties, B., & Maryn, Y. (2017). Validation of the Acoustic Voice Quality Index in the lithuanian language. *Journal of Voice*, 31(2), 257.e1-257.e11. <https://doi.org/10.1016/j.jvoice.2016.06.002>
- Vaissière, J. (2015). *La phonétique* (3<sup>e</sup> éd.). Paris, France : Presses Universitaires de France.
- Van Immerseel, L., & Martens, J. (1992). Pitch and voiced/unvoiced determination with an auditory model. *The Journal of the Acoustical Society of America*, 91(6), 3511-3526. <https://doi.org/10.1121/1.402840>
- Verduyckt, I., Remacle, M., Jamart, J., Benderitter, C., & Morsomme, D. (2011). Voice-related complaints in the pediatric population. *Journal of Voice*, 25(3), 373-380. <https://doi.org/10.1016/j.jvoice.2009.11.008>
- Verstraete, J. (2017, 30 avril). *Vocal assessment* (support de cours). L-Imsida, Malte : University of Malta.
- Volín, J., & Zimmermann, J. (2011). Spectral slope parameters and detection of word stress (actes de conférence). Dans *Proceedings of the Technical Computing Prague*. Prague, République tchèque : Humusoft. Récupéré sur le site Humusoft : [http://www2.humusoft.cz/www/papers/tcp11/125\\_volin.pdf](http://www2.humusoft.cz/www/papers/tcp11/125_volin.pdf)
- Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques* (2<sup>e</sup> éd.). Thousand Oaks, CA : SAGE Publications.
- Watts, C. R., Awan, S. N., & Maryn, Y. (2017). A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice*, 31(3), 387.e1-387.e10. <https://doi.org/10.1016/j.jvoice.2016.09.012>

- Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N., & Wilson, J. A. (2004). The reliability of three perceptual evaluation scales for dysphonia. *European Archives of Oto-Rhino-Laryngology*, *261*, 429-434. <https://doi.org/10.1007/s00405-003-0707-7>
- White, M. (2014, 20 octobre). *The exact perfect amount of time to take a break, according to data*. Récupéré le 28 février 2018 du site Time : <http://time.com/3518053/perfect-break/>
- Wilson, J., Webb, A., Carding, P., Steen, I., MacKenzie, K., & Deary, I. (2004). The Voice Symptom Scale (VoiSS) and the Vocal Handicap Index (VHI): A comparison of structure and content. *Clinical Otolaryngology and Allied Sciences*, *29*, 169-174. <https://doi.org/10.1111/j.0307-7772.2004.00775.x>
- Woisard, V., Bodin, S., Yardeni, E., & Puech, M. (2007). The Voice Handicap Index: Correlation between subjective patient response and quantitative assessment of voice. *Journal of Voice*, *21*(5), 623-631. <https://doi.org/10.1016/j.jvoice.2006.04.005>
- Wuyts, F. L., De Bodt, M., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., . . . Van de Heyning, P. H. (2000). The Dysphonia Severity Index: An objective measure of vocal quality based on a multiparameter approach. *Journal of Speech Language and Hearing Research*, *43*(3), 796-809. <https://doi.org/10.1044/jslhr.4303.796>
- Yanagisawa, E., Driscoll, B. P., & Sims, H. S. (2005). Laryngeal photography and videography. Dans R. T. Sataloff (Éd.), *Clinical Voice Assessment* (p. 103-124). San Diego, CA : Plural Publishing.
- Zur, K. B., Cotton, S., Kelchner, L., Baker, S., Weinrich, B., & Lee, L. (2007). Pediatric Voice Handicap Index (pVHI): A new tool for evaluating pediatric dysphonia. *International Journal of Pediatric Otorhinolaryngology*, *71*, 77-82. <https://doi.org/10.1016/j.ijporl.2006.09.004>

## ANNEXES

## Annexe 1 – le cepstre, kézako?!

En très bref, le cepstre est la double transformation de Fourier d'un signal sonore (Delgado, León, Izquierdo, & Llanos, 2018). Initialement employée par Hillenbrand comme corrélat acoustique du caractère soufflé de la voix, il a peu à peu été utilisé plus globalement comme corrélat acoustique de la qualité vocale (Fraile & Godino-Llorente, 2014).

L'objectif de l'analyse cepstrale est la déconvolution du signal afin de séparer l'influence de la source sonore (vibration des plis vocaux) de celle du conduit vocal (modulation par les cavités de résonance supraglottiques). Nous pouvons ainsi identifier sur le cepstre de la voyelle [a] à la figure 29 la contribution de la source sonore grâce au pic marqué vers 10 ms, appelé pic cepstral (Maryn, Dick, Vandenbruaene, Vauterin, & Jacobs, 2009). Les pics de plus petite quéfrencé sur le cepstre, à gauche du pic proémminent, représentent les modulations par les résonateurs.

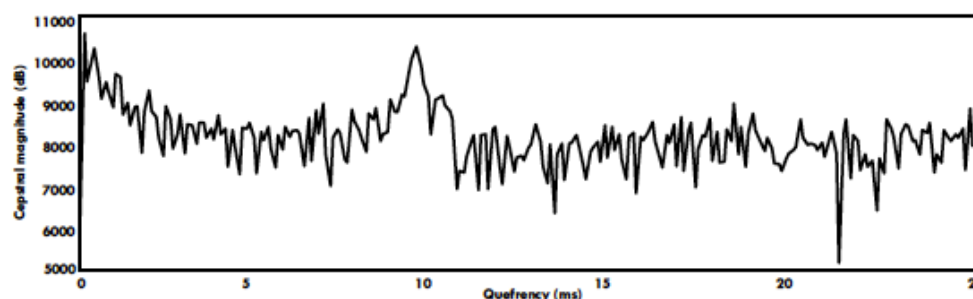


Figure 29. Représentation cepstrale de la voyelle [a] (Maryn, Dick, Vandenbruaene, Vauterin, & Jacobs, 2009).

Le pic cepstral permet de déterminer la fréquence fondamentale du son<sup>47</sup> ; en outre, plus il est prononcé, plus le spectre sonore est périodique. Ainsi, plus une voix est dysphonique, moins la composition spectrale harmonique sera « propre », et moins le pic sera discernable (figure 30). La proéminence du pic cepstral est donc une **mesure de la périodicité du spectre vocal**. Cette mesure ne repose pas sur la détection des cycles du signal sonore, au contraire des mesures de perturbation classiquement employées (Delgado, León, Izquierdo, & Llanos, 2018). Le cepstre<sup>48</sup> représente les structures répétitives (les espacements entre les harmoniques) dans le spectre vocal (Goldberg & Riek, 2000). Ferrer et al. précisent qu'il est important de ne pas confondre la périodicité du signal acoustique avec la périodicité harmonique du spectre. C'est cette différence qui fait de cet outil une mesure vocale plus fiable. En effet, les indices de perturbation sonore classiques (shimmer et jitter, HNR) se mesurent grâce à un algorithme de détection de la fréquence fondamentale sur la représentation temporelle de l'onde sonore.

<sup>47</sup> Noll précise que le pic cepstral n'est présent que dans le cas de sons voisés. Ainsi, l'analyse cepstrale peut aussi permettre la distinction entre les sons voisés/non voisés par sa présence ou son absence (Noll, 1964).

<sup>48</sup> Notons que – tout comme le spectrogramme (ou sonagramme) représente le spectre en fonction du temps – le cepstrogramme est également une représentation du cepstre dans le temps (Maryn, 2017), comme représenté à la figure 30.

Ces mesures semblent être trop sensibles à des variations minimales de la vibration — variations non nécessairement associées à une qualité vocale diminuée (Ferrer, De Bodt, Maryn, Van de Heyning, & Hernandez-Diaz, 2007). De plus, la variation d'un seul paramètre (perturbation d'amplitude, de fréquence, ou présence de bruit) semble généralement affecter les trois paramètres simultanément (Hillenbrand, 1987). La mesure de la proéminence du pic cepstral est sensible aux perturbations d'amplitude et de fréquence (modifiant la structure harmonique du spectre), ainsi qu'à l'ajout de bruit au signal vocal (diminuant les creux entre les harmoniques). Cette mesure semble donc davantage adaptée pour l'évaluation de la qualité vocale. Heman-Ackah et al. (2003) décrivent eux aussi la supériorité de mesures cepstrales, surtout pour les dysphonies plus sévères caractérisées par un signal peu périodique et se prêtant donc plus difficilement aux analyses du domaine fréquentiel.

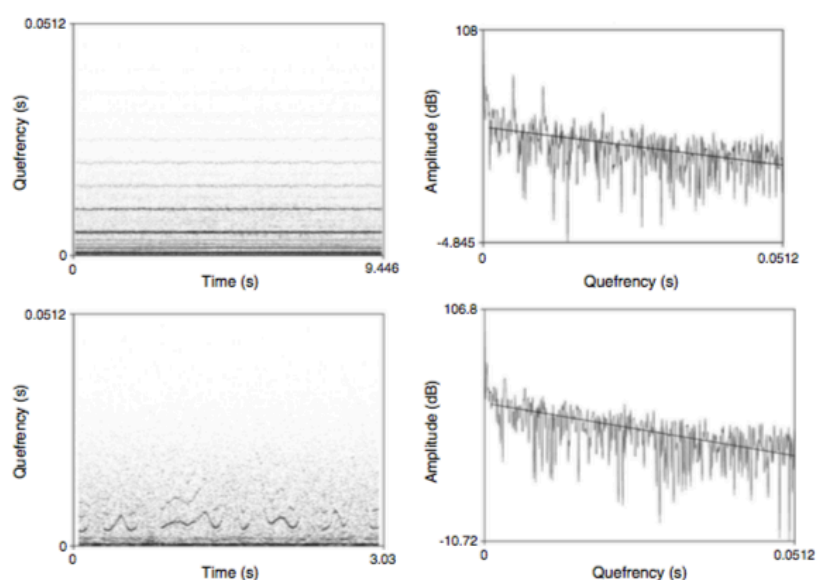


Figure 30. Exemples de cepstres et cepstrogrammes d'un [a] tenu (Delgado-Hernández, León-Gómez, Izquierdo-Arteaga, & Llanos-Fumero, 2018).

À droite = cepstre ; à gauche = cepstrogramme ; en haut = voix normophonique ; en bas = voix dysphonique.

Décrivons brièvement les étapes permettant l'obtention d'une représentation cepstrale.

La source vocale primaire peut être représentée par une onde sinusoïdale, caractérisée par sa fréquence (nombre de cycles par seconde) et par son amplitude (sur l'axe des ordonnées).

La voix humaine étant un phénomène complexe, l'onde vocale est une onde elle aussi dite complexe, c'est-à-dire composée de multiples ondes sinusoïdes d'amplitude et de fréquence différentes : la fréquence fondamentale (F0) et les harmoniques (multiples entiers de la F0). La représentation de cette *onde complexe* sur un oscillogramme est une **représentation « temporelle »** (figure 31) : on observe sur l'axe des abscisses l'évolution dans le temps, et sur l'axe des ordonnées, l'amplitude.

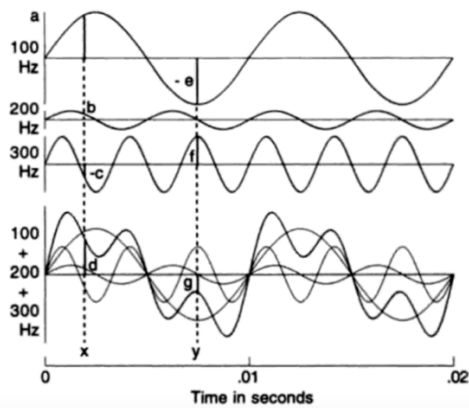


Figure 31. Décomposition d'une onde complexe en ses différents composants (Ladefoged, 1995).

En bas = représentation d'une onde complexe ; en haut = la fréquence fondamentale ; second et troisième tracé = les harmoniques.

C'est en effectuant une analyse de Fourier sur un tel signal que l'on peut visualiser les amplitudes de toutes les ondes constituantes, à un moment donné dans le temps : on obtient ainsi le *spectre* de l'onde complexe (figure 32). Il s'agit dans ce cas d'une **représentation « fréquentielle »** : sur l'axe des abscisses, nous observons alors les différentes fréquences, et sur l'axe des ordonnées, leur amplitude. Notons que les amplitudes représentées sur le spectre sont généralement des logarithmes des amplitudes absolues, permettant ainsi une plus grande précision d'analyse.

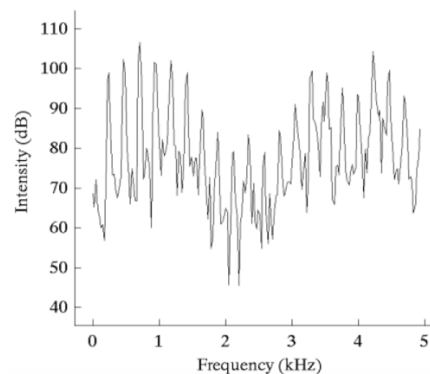


Figure 32. Exemple d'un spectre (Heman-Ackah, Michael, & Goding, 2002).

La fréquence fondamentale correspond à la source sonore, c'est-à-dire à la vibration des plis vocaux. C'est généralement la plus petite fréquence sur la représentation spectrale, et qui a la plus grande amplitude. Les zones de fréquences renforcées par les résonateurs — dites harmoniques — sont des multiples entiers de cette fréquence fondamentale et ont une amplitude plus réduite. Sur le spectre, ces harmoniques sont espacés par des intervalles réguliers réciproques à la période fondamentale de l'onde sonore ( $1/T$ ) (figure 33, graphique b).

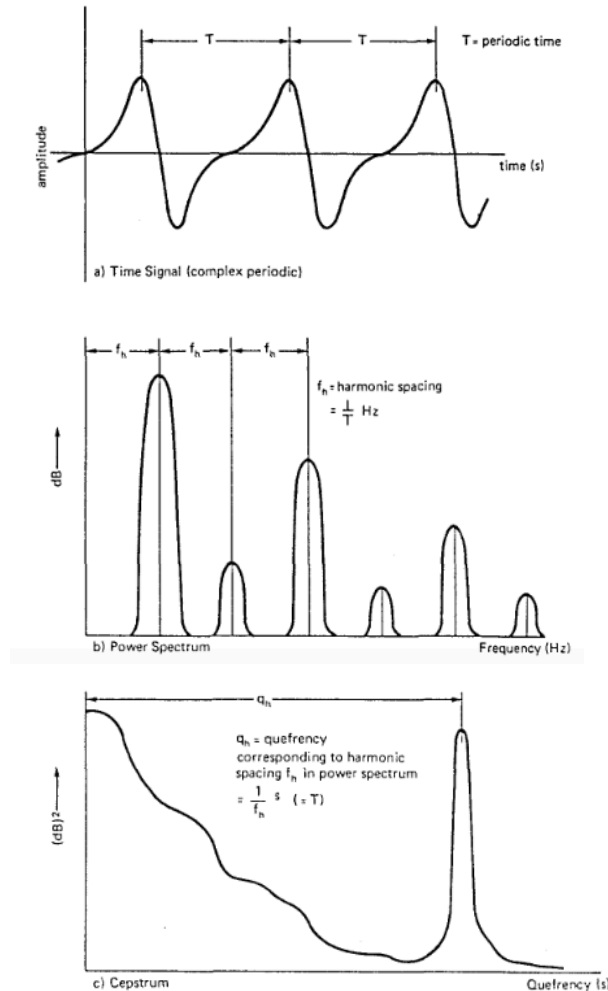


Figure 33. Représentation temporelle, spectrale et cepstrale du son (Randall, 1973).

(a) Représentation temporelle du son ; (b) représentation spectrale ; (c) représentation cepstrale.

Lorsque la voix est quasiment périodique (voix « normale »), les pics d'amplitude de la fréquence fondamentale et des harmoniques se distinguent clairement du bruit de fond. Cependant, lorsque l'onde vocale est peu périodique — dans le cas d'une dysphonie sévère — la distinction devient plus difficile. C'est ici que la représentation cepstrale devient intéressante.

Comme nous l'avons mentionné, la représentation cepstrale est obtenue en appliquant une transformation inverse de Fourier sur le spectre vocal, plus précisément sur son logarithme. Le cepstre peut donc être résumé ainsi :

$$\text{cepstre} = Fi[\log(\text{spectre})]$$

(avec « Fi » : transformation inverse de Fourier)

Aussi, c'est comme si l'on considérait le spectre comme étant l'oscillogramme d'une onde sonore complexe — quasiment périodique et avec des variations d'amplitude — et qu'on souhaitait en faire une représentation spectrale pour en révéler la fréquence fondamentale et les harmoniques. Le cepstre est donc « une représentation spectrale du spectre » (Heman-Ackah



et al., 2003). Afin d'éviter toute confusion, Tukey a proposé de renommer les fréquences dans le cepstre « **quéferences** » (Noll, 1964). Celles-ci sont notées sur l'axe des abscisses, et sont exprimées en unités de temps (ms) (Noll, 1967). Notons que les fréquences élevées du signal, situées du côté droit dans la représentation spectrale, seront au contraire situées du côté gauche dans la représentation cepstrale. L'axe des ordonnées, lui, représente la magnitude ou « **gamnitude** » cepstrale (Maryn, Roy, De Bodt, Van Cauwenberge, & Corthals, 2009). L'analyse cepstrale se distingue donc de l'analyse spectrale par sa nature non plus fréquentielle, mais temporelle.

Considérons le spectre A d'un signal acoustique. Selon la théorie source-filtre, comme représenté aux figures 34 et 35, le spectre A est composé :

- de l'enveloppe spectrale B : c'est la courbe lisse qui relie les harmoniques de la représentation spectrale et qui représente les modulations par le tractus vocal ;
- des détails spectraux C : ils représentent la source sonore, c.-à-d. les impulsions induites par la vibration des plis vocaux.

D'où la formule :  $A = B \times C$  (Noll, 1964, 1967)

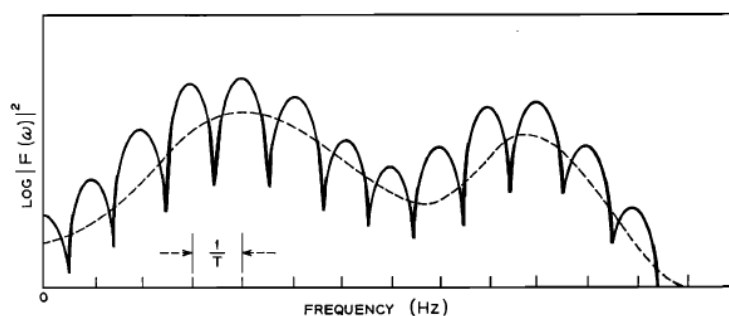


Figure 34. Représentation schématique de l'enveloppe spectrale et des détails spectraux (Noll, 1967).

En pointillés = l'enveloppe spectrale ; en ligne continue = les détails spectraux.

L'enveloppe spectrale (B) — si l'on considère le spectre comme étant une onde complexe — possède une « fréquence basse » ou une « longue longueur d'onde » (Noll, 1964) ; les détails spectraux (C), ondulations périodiques, ont une « fréquence élevée » ou une « courte longueur d'onde » (Noll, 1964) (figures 34 et 35).

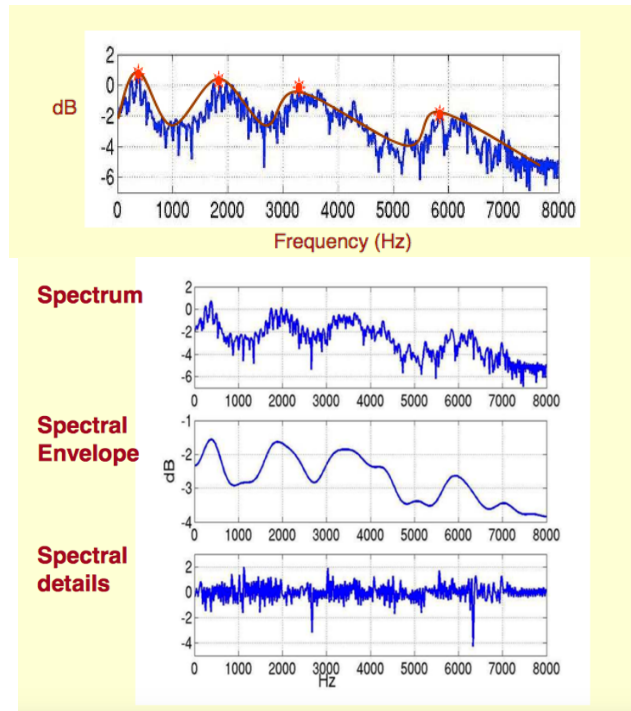


Figure 35. Décomposition du spectre en l'enveloppe spectrale et les détails spectraux (Prahallad, 2008).

En haut = représentation spectrale ; au milieu = l'enveloppe spectrale (modulations par le tractus vocal) ; en bas = les détails spectraux (source sonore).

Afin de pouvoir séparer, filtrer (« lifter ») les deux composantes B et C et obtenir le cepstre (figure 36), on applique une transformation inverse de Fourier sur le logarithme du spectre. Celle-ci nous permet de séparer les effets du conduit vocal et du signal glottique, car le logarithme d'un produit est égal à la somme des logarithmes des multiplicandes :

$$\log A = \log (B \times C) \rightarrow \log A = \log B + \log C$$

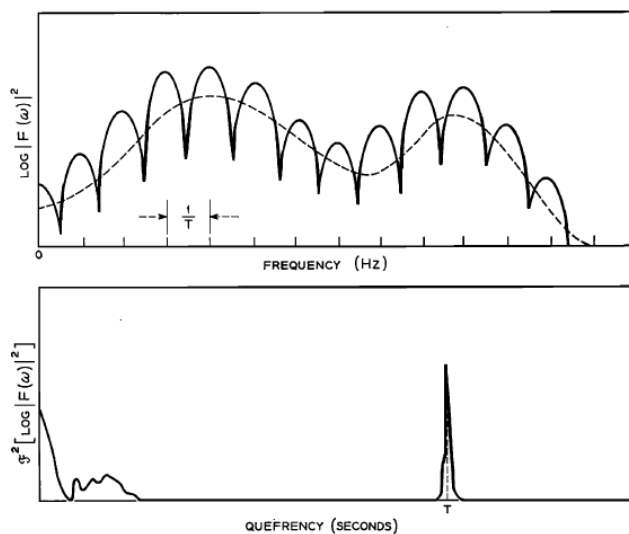


Figure 36. Représentations schématiques spectrale et cepstrale d'une onde sonore périodique (Noll, 1967).

En haut = spectre ; en bas = cepstre, avec le pic proéminent correspondant à la période T de l'onde sonore.

Une transformation de Fourier sur l'« onde sonore spectrale » nous fournira ainsi schématiquement deux composantes principales sur l'axe quéfrentiel (« pseudo-fréquentiel » à la figure 37) :

- un pic large dans les quéfrences basses (Noll, 1967), correspondant à l'*enveloppe spectrale*. Ce pic représente les modulations du son vocal par les *résonateurs*. L'on parlera également de *rhamoniques*. Sur le spectre, il correspond à la composante de longue longueur d'onde.
- un pic étroit mais d'une grande magnitude (« gamnitude »), dans les quéfrences plus élevées (Noll, 1967), correspondant aux *détails spectraux*. Ce pic représente la source sonore (vibration des plis vocaux), et correspond à la composante de courte longueur d'onde sur le spectre.

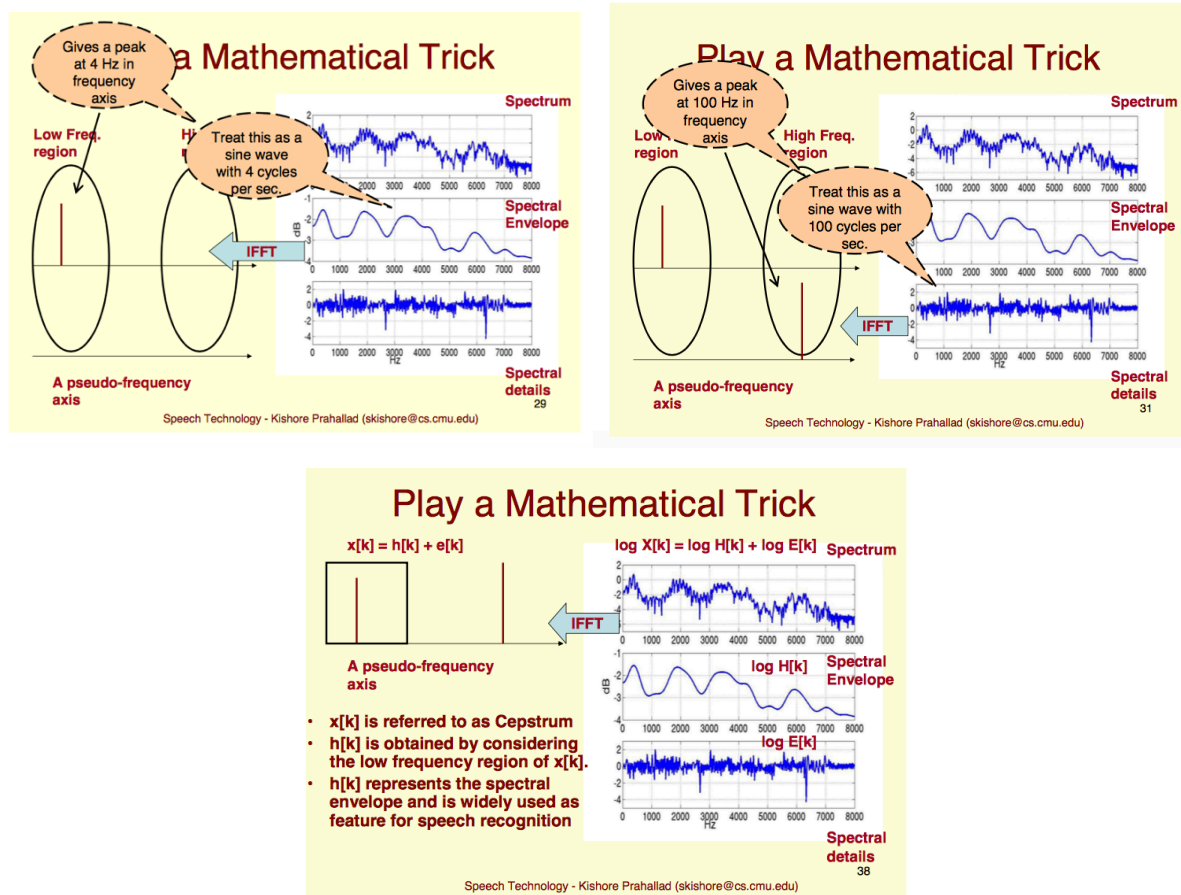


Figure 37. Représentation schématique de la transformation inverse de Fourier effectuée sur le spectre (Prahallad, 2008). La transformation inverse de Fourier sur le spectre convertit l'enveloppe spectrale (en haut à gauche) et les détails spectraux (en haut à droite) en deux éléments bien distincts sur le cepstre (en bas). C'est ainsi que le cepstre permet de déconvoluer la source et le filtre.

La *source sonore* (vibration des plis vocaux) étant périodique, si sa période (sur l'oscillogramme) est de T secondes, les harmoniques du spectre seront séparés de  $T^{-1}$  Hz (figure 33). La représentation spectrale (fréquentielle) est donc « périodique » sur l'axe des fréquences, les harmoniques étant séparés d'un intervalle réciproque à la période de la

représentation temporelle (Noll, 1967). Sur le cepstre, le pic représentant la source sonore aura une quéfrence correspondant à la période  $T$  (en ms). La quéfrence du pic proéminent sur le cepstre (**pic cepstral**) est donc réciproque à la distance séparant les harmoniques du spectre (et correspond à la période fondamentale du son) (Hillenbrand & Houde, 1996).

Aussi, pour un son d'une fréquence fondamentale de 80 Hz (figure 38), la période  $T$  sera de  $1/80 = 0.0125$  seconde, soit 12.5 millisecondes. Sur la représentation spectrale, les harmoniques seront séparés de  $1/T$ , soit 0.08 Hz, et le pic cepstral se situera à  $T$  millisecondes, soit 12.5 ms.

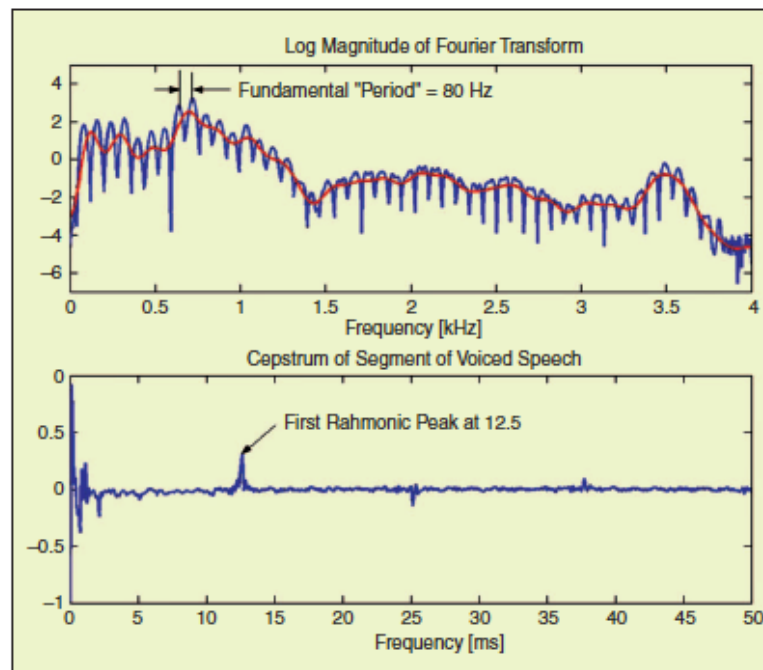


Figure 38. Représentation spectrale et cepstrale d'un son de 80 Hz (Oppenheim & Schaffer, 2004).

En haut = représentation spectrale ; en bas = cepstre correspondant ; les harmoniques sont séparés d'un intervalle de 0.08, et le pic cepstral se situe à  $1/0.08$ , c'est-à-dire 12.5 ms.

Nous pouvons donc identifier, par la quéfrence du pic cepstral, l'intervalle périodique qui sépare les harmoniques du spectre (sans devoir identifier au préalable la fréquence fondamentale !), et ainsi la période et la fréquence fondamentale du son.

En conséquence, dans le cas d'un signal fortement périodique, nous pouvons observer sur le spectre des harmoniques séparés par un intervalle régulier, correspondant à  $1/T$ . Sur la représentation cepstrale, nous observons un pic cepstral saillant (à  $T$  millisecondes). L'**amplitude du pic cepstral** reflète l'organisation harmonique du spectre (plus le son est périodique, plus les intervalles séparant les harmoniques seront réguliers, et plus le pic cepstral sera grand).

Mais cette amplitude est aussi influencée par le niveau de pression sonore global du signal (Hillenbrand, Cleveland, & Erickson, 1994). C'est pour cela que l'on calcule une droite de régression sur le cepstre, par rapport à laquelle l'on mesurera ensuite l'amplitude du

pic (figure 39) : la **proéminence du pic cepstral (PPC)** correspond à la distance entre le sommet du pic de proéminence et cette droite de régression. La droite de régression permet de normaliser les variations d'amplitude liées à la taille de la fenêtre d'analyse et à l'énergie sonore globale du signal (Hillenbrand, Cleveland, & Erickson, 1994). Ainsi, la proéminence du pic cepstral est calculée par rapport à l'amplitude moyenne de phonation, permettant donc des comparaisons objectives.

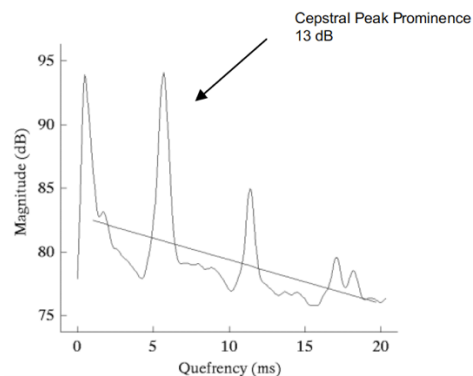


Figure 39. Exemple de représentation cepstrale avec la droite de régression et la proéminence du pic cepstral (Heman-Ackah, Michael, & Goding, 2002).

Résumons l'évolution, jusqu'ici, de l'analyse cepstrale : elle a été introduite en 1964 par Michael Noll. Ensuite, elle s'est développée dès 1994 grâce à des méthodes de calcul plus efficaces par Hillenbrand et al. L'utilisation de la droite de régression a alors été introduite, et l'analyse de la proéminence du pic cepstral a été décrite.

Finalement, l'analyse de la proéminence du pic cepstral a été adaptée en 1996 par l'ajout d'un « lissage » qui se fait en mesurant la PPC sur la moyenne d'un certain nombre de fenêtres cepstrales avant et après la fenêtre d'intérêt (Heman-Ackah et al., 2003 ; Hillenbrand & Houde, 1996).

L'**AVQI** utilise donc l'analyse cepstrale afin de déterminer la **proéminence du pic cepstral lissé** (smoothed cepstral peak prominence) à l'aide du logiciel Praat (initialement, ce calcul se faisait à l'aide du logiciel SpeechTool)<sup>49</sup>. Il s'agit donc de la distance entre l'amplitude (la « gamnitude ») du pic cepstral et l'amplitude du point ayant la même quéfrence sur la droite de régression à travers le cepstre lissé (donc l'amplitude prédite à ladite quéfrence par la droite de régression) (Hillenbrand, Cleveland, & Erickson, 1994), comme illustré à la figure 40.

<sup>49</sup> Notons que la proéminence du pic cepstral est calculable grâce à différents logiciels d'analyse acoustique, dont Analysis of Dysphonia in Speech and Voice (ADSV), SpeechTool et Praat. Les valeurs obtenues, bien que fortement corrélées, ne sont pas strictement équivalentes (Maryn & Weenink, 2015 ; Watts, Awan, & Maryn, 2017).

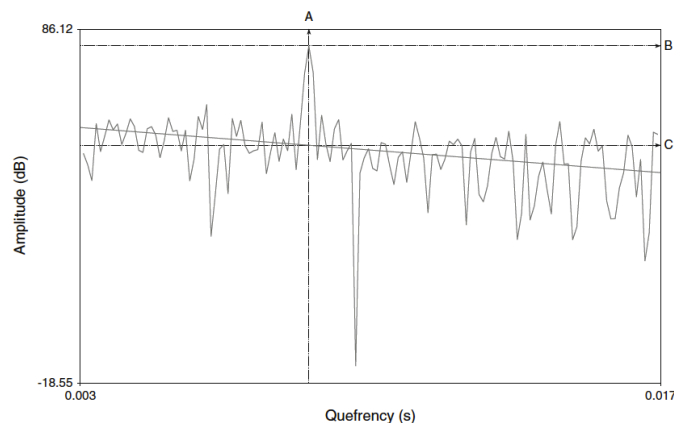


Figure 40. Représentation de la détermination de la proéminence du pic cepstral (Maryn, De Bodt, Barsties, & Roy, 2014). La proéminence du pic cepstral (A) est calculée en mesurant la différence entre le sommet de celui-ci (B) et le point de même quéfreny sur la droite de régression (C).

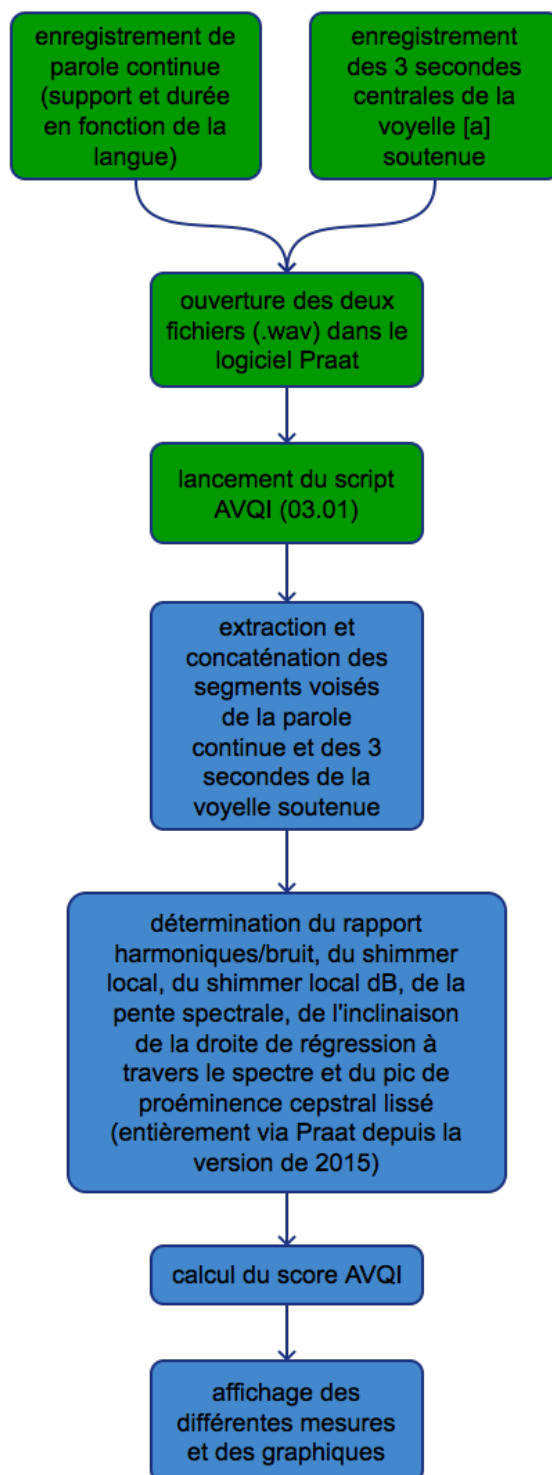
Notons qu'une étude récente de Delgado-Hernández et al. a montré que la proéminence du pic cepstral est un outil d'évaluation de la qualité vocale valide autant lorsqu'il est utilisé sur une voyelle tenue que lorsqu'il est appliqué à un échantillon de parole continue. Cependant, une plus grande corrélation a été mesurée entre le degré général de la dysphonie (G de la GRBAS) et le PPC mesuré sur la parole continue (Delgado, León, Izquierdo, & Llanos, 2018). Les auteurs soulignent que l'analyse cepstrale effectuée sur un échantillon de parole continue est influencée par le contenu de celui-ci, par les accents régionaux, mais aussi par la fluidité de la lecture de l'échantillon. Ainsi, il est fortement recommandé que le sujet effectue une ou deux lectures au préalable, afin de se familiariser avec l'échantillon à lire et ainsi garantir une fluidité de lecture acceptable.

Une dernière remarque au sujet du pic de proéminence cepstral concerne les données normatives. En anglais, Heman-Ackah et al. observent que la proéminence du pic cepstral mesurée à l'aide de SpeechTool sur un échantillon de parole continue peut être considérée comme normale lorsqu'elle est supérieure ou égale à 4 dB (Heman-Ackah et al., 2014).

Aucune donnée normative n'est à ce jour disponible pour le français ni pour le PPC calculé à l'aide du logiciel Praat (et ce, pour aucune langue) — qui tend à être légèrement plus élevé que celui calculé par le logiciel SpeechTool (Maryn & Weenink, 2015). Ceci pourrait donc être, pourquoi pas, le sujet d'une prochaine étude !

## Annexe 2 – étapes de l'AVQI

En vert, les étapes à effectuer par l'examineur ; en bleu, les étapes effectuées par le script AVQI 03.01.



Annexe 3 – exemple d'application du protocole d'évaluation de l'ELS  
(Remacle & Morsomme, 2009)

**Rapport d'un protocole de base :**

Femme, 26, nodules de cordes vocales, avant traitement

PERCEPTION : G34 B52 R18d

STROBOSCOPIE : Ferm 40go Reg10 OM 25 Sym0

AERODYNAMIQUE : QP 285 ml/s (TMP 13 s).

ACOUSTIQUE : Ji 1.2 %, Shi 6.1 %, étendue Fo - : c - g1, intensité faible : 53 dB (A) 30 cm.

EVALUATION SUBJECTIVE : Vo30 Hand50

---

**Commentaires :**

Perception est cotée sur 3 échelles visuelles analogiques de 100 mm. Grade est mesuré 34/100, etc.. "d" signifie que la patiente présente de la diplophonie.

Stroboscopie est mesurée sur 4 échelles analogiques visuelles de 100 mm : fermeture, régularité, qualité de l'ondulation muqueuse et symétrie. Dans ce cas la béance en forme de sablier : « go ». La symétrie est normale chez cette patiente

Aérodynamique : Quotient Phonatoire (ml/s) et Temps Maximal Phonatoire (s). VC était ici de 3 705 ml.

Acoustique : Jitter et Shimmer sur un /a:/ tenu, à une hauteur et une force confortables : c correspond à 131 Hz et g1 à 392 Hz pour la hauteur ; 53 dB (A) correspond à la force. Comme pour le phonétogramme, la distance du microphone doit être de 30 cm.

Evaluation subjective : fournie par le patient lui-même sur une double échelle visuelle analogique de 100 mm. La première échelle (Vo) se rapporte à l'impression à propos de la qualité vocale stricto sensu (30/100 : faible à modéré), pendant que la seconde (Hand) se rapporte à l'impression au sujet de la répercussion de problème vocal sur la vie de tous les jours et, éventuellement sur la vie professionnelle et sociale (50/100 : modéré à sévère).





## Annexe 5 – phrases de Combescure (Combescure, 1981)

### Liste N° 1 :

1. Il se garantira du froid avec ce bon capuchon.
2. Annie s'ennuie loin de mes parents.
3. Les deux camions se sont heurtés de face.
4. Un loup s'est jeté immédiatement sur la petite chèvre.
5. Dès que le tambour bat, les gens accourent.
6. Mon père m'a donné l'autorisation.
7. Vous poussez des cris de colère ?
8. Ce petit canard apprend à nager.
9. La voiture s'est arrêtée au feu rouge.
10. La vaisselle propre est mise sur l'évier.

### Liste N° 3 :

1. Lentement des canes se dirigent vers la mare.
2. Une goélette déploie ses voiles.
3. Le facteur va porter le courrier.
4. Bien sûr, je connais son nom.
5. Maman prend un verre et une assiette.
6. Désormais, je me tournerai quand il partira.
7. Les avions tournent au-dessus de la place.
8. Mettez la faux, ici sous ma tente.
9. Je suis resté sourd à ses cris.
10. Le chameau est loin de son abri.

Nombre total de phonèmes : 213  $\chi^2 = 3,501$

### Liste N° 5 :

1. Le menuisier a scié une planche et l'a rabotée.
2. Maman a préparé une galette pour jeudi.
3. Le football, voilà ce qui l'intéresse.
4. C'est un charmant spectacle, je t'assure.
5. Ils m'ont apporté des friandises à mon anniversaire.
6. Ces élèves prendront l'autocar tout à l'heure.
7. Parfois, mon épicière vend à crédit.
8. Personne n'a applaudi ce beau discours.
9. Je me demande pourquoi on court sans cesse.
10. Il se reprend de ce qu'il vient de faire.

Nombre total de phonèmes : 259  $\chi^2 = 3,936$

### Liste N° 7 :

1. Il s'est réfugié dans ma chambre.
2. Le troupeau s'abreuvait au ruisseau.
3. Le client s'attend à ce que vous fassiez une réduction.
4. Chaque fois que je me lève, ma plaie me tire.
5. Une rançon est exigée par les ravisseurs.
6. Ainsi cette comédie est en un acte.
7. Papa aime mon vin quand il est bon.
8. Le ciel est tout noir, il va tomber des cordes.
9. On dit que l'essor de ce village est important.
10. Ce soir, nous ne nous coucherons pas tard.

Nombre total de phonèmes : 239  $\chi^2 = 4,080$

### Liste N° 9 :

1. L'été, tout le monde se mettait aux fenêtres.
2. Le cocher a fouetté sa jument.
3. Je rends souvent visite à mon oncle.
4. Ma soirée se passera sans incident.
5. La police veut les papiers du chauffeur.
6. Jean, quant à lui, est très grand pour son âge.
7. Le microscope, qui est sur pied, est le mien.
8. Le jardin entoure un petit lac.
9. Il a broyé du noir depuis la perte de son ami.
10. Le forçat s'est évadé du bagne.

Nombre total de phonèmes : 231  $\chi^2 = 4,153$

### Liste N° 2 :

1. Leur chienne a hurlé toute la nuit.
2. Pour se protéger, il s'est couché près de ma porte.
3. Sa voisine est inimitable.
4. Le renard se hâte vers son gîte.
5. Le bouillon fume dans les assiettes.
6. Le caractère de cette femme est moins calme.
7. Le camp d'été s'est passé au bord du fleuve.
8. Un train entre déjà en gare.
9. Souvent, je m'accoude au muret de ce pont.
10. A l'Ouest, mes pommiers donnent peu.

### Liste N° 4 :

1. Il pense être de retour ici, avant la nuit.
2. Des chiens nous montraient leurs crocs pointus.
3. La jeune fille se peigne devant sa glace.
4. Il a été condamné pour un vol de voiture.
5. Je ne veux pas que vous le changiez pour le moment.
6. Nous avons pris froid en jouant au tennis.
7. Il est désormais accablé par son travail.
8. Ce bonbon contenait trop de sucre.
9. A la hâte, le métayer ensilaït ses récoltes avant l'hiver.
10. Une brume épaisse s'est formée sur la mer.

Nombre total de phonèmes : 257  $\chi^2 = 3,846$

### Liste N° 6 :

1. Des gens se sont levés dans les tribunes.
2. Vous éplucherez les légumes du pot-au-feu.
3. Ce chasseur projette encore de partir d'ici, ce matin.
4. La poire est un fruit à pépins.
5. Plus nous le connaissons, plus nous le respectons.
6. Là-haut, monte la voix du pâtre qui ramène ses moutons.
7. Le courrier arrive en retard en ce moment.
8. Cette cage contient mon oiseau.
9. Des lièvres jouent à l'orée du bois.
10. Je te dis que ma bouteille s'abîme à la cave.

Nombre total de phonèmes : 246  $\chi^2 = 4,021$

### Liste N° 8 :

1. Vous avez du plaisir à jouer avec ceux qui ont un bon caractère.
2. Le chevrier a corné pour rassembler ses troupeaux.
3. Mon cordonnier a ressemelé tes souliers.
4. L'oie est dans sa main, son cœur bat et saute.
5. Une rivière dessinait des méandres dans sa prairie.
6. L'alpiniste continuait à grimper le long d'une roche.
7. Effrayé par l'insecte, je rentre précipitamment.
8. Je me suis entretenu avec l'institutrice de ma jeune fille.
9. Quand le soleil se lève, je saute de mon lit.
10. Le fermier est parti pour la foire.

Nombre total de phonèmes : 301  $\chi^2 = 4,129$

### Liste N° 10 :

1. Un fort crédit est consenti par une banque.
2. Le passereau lance une roulade et s'enfuit.
3. Des hannetons voletaient autour de ce prunier.
4. Ces légendes me rappellent les temps anciens.
5. Qu'est ce que vous regardez comme oiseau ?
6. Ce sont mes meilleurs chevaux dont voici les noms.
7. Je parcours les rues des villages avec sa mère.
8. Ma partition est sous ce pupitre.
9. Il arrive demain d'Italie par la route.
10. Le tapis était élimé sur le bord.

Nombre total de phonèmes : 238  $\chi^2 = 4,323$



**Liste N° 11 :**

1. Ma mère et moi faisons de courtes promenades.
2. La poupée fait la joie de cette très jeune fille.
3. Mais le temps lui a manqué.
4. Il aura été retardé par quelqu'importun.
5. Une grenouille verte saute sur les nénuphars.
6. Des violettes emplissent l'air de subtiles senteurs.
7. Au bois, j'ai ramassé de si bons champignons.
8. Fais ce que je veux dès ce midi !
9. Papa coupe l'herbe dans le jardin.
10. Vous porterez ces caisses dans vos voitures.

Nombre total de phonèmes : 237  $\chi^2 = 4,326$

**Liste N° 13 :**

1. La lune se lève maintenant au-dessus des arbres.
2. Des rires montent de la cour de récréation.
3. Ça et là, la prairie se piquait de fleurs.
4. Une grosse poutre maintient la misérable charpente.
5. Tout s'est animé, dès que le soleil s'est levé.
6. Ma voiture est en panne devant ce pont.
7. Ces femmes portent encore une coiffe.
8. Je lui rapporte des fruits très rouges.
9. Je vois ma table en bois vert.
10. Dans le taillis est caché un nid de fauvelles.

Nombre de phonèmes : 240  $\chi^2 = 4,354$

**Liste N° 15 :**

1. Vous achèterez mes moules minuscules !
2. Une jolie bague scintille au doigt de ta fille aînée.
3. A six heures, un voyageur attendait le train.
4. Ses locataires sont rentrés très tard.
5. Ce que j'ai prévu se produira.
6. Le capitaine regarde par le hublot de sa cabine.
7. Virginie a mis le couvert pour sa fête.
8. Votre portrait est exposé au salon.
9. Maman se demande ce qu'il va dire.
10. Des moineaux se sont querellés dans mon champ.

Nombre total de phonèmes : 245  $\chi^2 = 4,486$

**Liste N° 17 :**

1. Elle habite à proximité du champ de foire.
2. Ma concierge veillait sur mon appartement durant les vacances.
3. Ils sont allés travailler bien qu'ils fussent fatigués.
4. Un colonel commandait le régiment.
5. Je vous dis de recoudre ce bouton.
6. Nos parents sont nos tuteurs naturels.
7. Les mésanges y pondaient des œufs tachetés.
8. Vous lui défendez de jouer sur cette route le soir.
9. Ma goélette noire est rentrée au port.
10. Ce passeport n'avait pas de visa.

Nombre total de phonèmes : 253  $\chi^2 = 4,536$

**Liste N° 19 :**

1. Je ménage une surprise à mon ami.
2. Les boulangers façonnent des pains.
3. Vos livres devront être couverts.
4. Mangeras-tu de cette tarte aux prunes.
5. Le chapeau de Monique est sur la table.
6. Il s'est glissé loin des spectateurs.
7. Il s'empresse de réclamer ce qu'on lui a promis.
8. Vous tremblez parce que vous avez froid.
9. C'est le soir qu'il travaille le mieux.
10. Un serpent noir fuit sous une pierre.

Nombre total de phonèmes : 227  $\chi^2 = 4,670$

**Liste N° 12 :**

1. A midi les collégiens vont au réfectoire.
2. Des pommes mûres se détachent de l'arbre.
3. Il tombe lourdement sur un sol plat.
4. Nous partons avant demain vers Paris.
5. Il a été arrêté par des policiers.
6. Cette voyageuse a loué une voiture sans chauffeur.
7. Jean semblait calme tout à coup.
8. Je n'irai sûrement pas danser à son mariage.
9. Elle le lui redit sans cesse.
10. Une guerre nucléaire ferait de nombreux morts.

Nombre de phonèmes : 232  $\chi^2 = 4,342$

**Liste N° 14 :**

1. On entend les gazouillis d'un oiseau dans le jardin.
2. La barque du pêcheur a été emportée par une tempête.
3. Ce livre provient de la bibliothèque.
4. J'en conclus qu'il n'y a personne à voir.
5. Le mal s'envenime, faute de soins.
6. Je suis sûr que vous connaissez ces noms.
7. Il s'arrêtait tout l'été, ici.
8. Voilà toujours deux choux pour le repas de midi.
9. Les manches de son manteau sont décousues.
10. Ce vaisseau parcourt les mers à travers le monde.

Nombre de phonèmes : 248  $\chi^2 = 4,417$

**Liste N° 16 :**

1. Une société de musique va bientôt défiler.
2. Le juge veut prolonger l'interrogatoire.
3. Ici, ma mère a acheté des coupons de tissu.
4. Pierre cogne par derrière comme un sourd.
5. La pluie ne fait pas le beau temps.
6. Sans fleurs, la maison est triste.
7. Elle a vraiment toujours des doigts menus.
8. Ce boucher n'a encore plus de lard à l'étalage.
9. Confie-moi à quoi tu penses.
10. Ce dont nous discutons vous laisse rêveur.

Nombre total de phonèmes : 230  $\chi^2 = 4,505$

**Liste N° 18 :**

1. La pieuvre saisit sa proie avec ses tentacules.
2. La neige couvre la cime des montagnes.
3. Un mouflon se cache dans les anfractuosités.
4. J'ai entendu ce que vous tâchez de jouer au piano.
5. Ce moyeu de roue grince continuellement.
6. Nous voulons tous nous promener à bicyclette.
7. Je ne peux atteindre les bocaux de confiture.
8. Dans cette crèmerie, on vend du fromage fort.
9. La pie se précipite vers ce qui brille.
10. Un petit lièvre est terré dans le buisson.

Nombre total de phonèmes 261  $\chi^2 = 4,575$

**Liste N° 20 :**

1. Je me souviens des beaux jours que j'ai vécus.
2. Mon cousin a été très vexé par ce qu'elle avait dit.
3. Grand-père, sois donc un peu raisonnable !
4. En ce moment, les soirées à l'opéra sont données.
5. Tu as beaucoup changé depuis que tu es parti.
6. Il a souffert pendant des semaines.
7. Vous voyez tout le temps cette femme triste.
8. Ce sentier mène à la route du village.
9. Le docteur a ordonné un médicament.
10. Il faut aussi arriver à temps.

Nombre total de phonèmes : 239  $\chi^2 = 4,787$

## Annexe 6 – script de l’AVQI 02.03 (Maryn, 2016)

*Ce script est disponible en téléchargement libre sur le site « Vlaamse Vereniging voor Logopedisten » (VVL), sous l’adresse <https://www.vvl.be/documenten-en-paginas/praat-script-avqi-v0203>.*

# TITLE OF THE SCRIPT: ACOUSTIC VOICE QUALITY INDEX (AVQI) v.02.03

# Form for introduction and/or parameterization

form Acoustic Voice Quality Index v.02.03

comment >>> It is advocated to estimate someone’s dysphonia severity in both continuous speech (i.e., 'cs') and sustained vowel (i.e., 'sv') (Maryn et al., 2010). This script therefore runs on these two types of recordings, and it is important to name these recordings 'cs' and 'sv', respectively.

comment >>> This script automatically (a) searches, extracts and then concatenates the voiced segments of the continuous speech recording to a new sound; (b) concatenates the sustained vowel recording to the new sound, (c) determines the Smoothed Cepstral Peak Prominence, the Shimmer Local, the Shimmer Local dB, the LTAS-slope, the LTAS-tilt and the Harmonics-to-Noise Ratio of the concatenated sound signal, (d) calculates the AVQI-score mostly based on the method of Maryn et al. (2010), and draws the oscillogram, the narrow-band spectrogram with LTAS and the power-cepstrogram with power-cepstrum of the concatenated sound signal to allow further interpretation.

comment >>> For the AVQI to be reliable, it is imperative that the sound recordings are made in an optimal data acquisition conditions.

comment >>> There are two versions in this script: (1) a simple version (only AVQI with data of acoustic measures), and (2) an illustrated version (AVQI with data of acoustic measures and above-mentioned graphs).

choice version: 2

button simple

button illustrated

comment >>> Additional information (optional):

sentence name\_patient

sentence left\_dates\_(birth\_-\_assessment)

sentence right\_dates\_(birth\_-\_assessment)

comment

comment Script credits: Youri Maryn (PhD) and Paul Corthals (PhD)

endform

Erase all

Select inner viewport... 0.5 7.5 0.5 4.5

Axes... 0 1 0 1

Black

Text special... 0.5 centre 0.6 half Helvetica 12 0 Please wait an instant. Depending on the duration and/or the sample rate of the recorded

Text special... 0.5 centre 0.4 half Helvetica 12 0 sound files, this script takes more or less time to process the sound and search for the AVQI.

# -----

# PART 0:

# HIGH-PASS FILTERING OF THE SOUND FILES.

# -----

select Sound cs

Filter (stop Hann band)... 0 34 0.1

Rename... cs

select Sound sv

Filter (stop Hann band)... 0 34 0.1

Rename... sv

# -----

# PART 1:

# DETECTION, EXTRACTION AND CONCATENATION OF

# THE VOICED SEGMENTS IN THE RECORDING

# OF CONTINUOUS SPEECH.

# -----

```

select Sound cs
Copy... original
samplingRate = Get sampling frequency
intermediateSamples = Get sampling period
Create Sound... onlyVoice 0 0.001 'samplingRate' 0
select Sound original
To TextGrid (silences)... 50 0.003 -25 0.1 0.1 silence sounding
select Sound original
plus TextGrid original
Extract intervals where... 1 no "does not contain" silence
Concatenate
select Sound chain
Rename... onlyLoud
globalPower = Get power in air
select TextGrid original
Remove

select Sound onlyLoud
signalEnd = Get end time
windowBorderLeft = Get start time
windowWidth = 0.03
windowBorderRight = windowBorderLeft + windowWidth
globalPower = Get power in air
voicelessThreshold = globalPower*(30/100)

select Sound onlyLoud
extremeRight = signalEnd - windowWidth
while windowBorderRight < extremeRight
    Extract part... 'windowBorderLeft' 'windowBorderRight' Rectangular 1.0 no
    select Sound onlyLoud_part
    partialPower = Get power in air
    if partialPower > voicelessThreshold
        call checkZeros 0
        if (zeroCrossingRate <> undefined) and (zeroCrossingRate < 3000)
            select Sound onlyVoice
            plus Sound onlyLoud_part
            Concatenate
            Rename... onlyVoiceNew
            select Sound onlyVoice
            Remove
            select Sound onlyVoiceNew
            Rename... onlyVoice
        endif
    endif
    select Sound onlyLoud_part
    Remove
    windowBorderLeft = windowBorderLeft + 0.03
    windowBorderRight = windowBorderLeft + 0.03
    select Sound onlyLoud
endwhile
select Sound onlyVoice

procedure checkZeros zeroCrossingRate

    start = 0.0025
    startZero = Get nearest zero crossing... 'start'
    findStart = startZero
    findStartZeroPlusOne = startZero + intermediateSamples
    startZeroPlusOne = Get nearest zero crossing... 'findStartZeroPlusOne'
    zeroCrossings = 0
    strips = 0

    while (findStart < 0.0275) and (findStart <> undefined)
        while startZeroPlusOne = findStart
            findStartZeroPlusOne = findStartZeroPlusOne + intermediateSamples
            startZeroPlusOne = Get nearest zero crossing... 'findStartZeroPlusOne'

```

```

        endwhile
        afstand = startZeroPlusOne - startZero
        strips = strips + 1
        zeroCrossings = zeroCrossings + 1
        findStart = startZeroPlusOne
    endwhile
    zeroCrossingRate = zeroCrossings/afstand
endproc

# -----
# PART 2:
# DETERMINATION OF THE SIX ACOUSTIC MEASURES
# AND CALCULATION OF THE ACOUSTIC VOICE QUALITY INDEX.
# -----

select Sound sv
durationVowel = Get total duration
durationStart=durationVowel-3
if durationVowel>3
Extract part... durationStart durationVowel rectangular 1 no
Rename... sv2
elsif durationVowel<=3
Copy... sv2
endif

select Sound onlyVoice
durationOnlyVoice = Get total duration
plus Sound sv2
Concatenate
Rename... avqi
durationAll = Get total duration
minimumSPL = Get minimum... 0 0 None
maximumSPL = Get maximum... 0 0 None

# Narrow-band spectrogram and LTAS

To Spectrogram... 0.03 4000 0.002 20 Gaussian
select Sound avqi
To Ltas... 1
minimumSpectrum = Get minimum... 0 4000 None
maximumSpectrum = Get maximum... 0 4000 None

# Power-cepstrogram, Cepstral peak prominence and Smoothed cepstral peak prominence

select Sound avqi
To PowerCepstrogram... 60 0.002 5000 50
cpsps = Get CPPS... no 0.01 0.001 60 330 0.05 Parabolic 0.001 0 Straight Robust
To PowerCepstrum (slice)... 0.1
maximumCepstrum = Get peak... 60 330 None

# Slope of the long-term average spectrum

select Sound avqi
To Ltas... 1
slope = Get slope... 0 1000 1000 10000 energy

# Tilt of trendline through the long-term average spectrum

select Ltas avqi
Compute trend line... 1 10000
tilt = Get slope... 0 1000 1000 10000 energy

# Amplitude perturbation measures

select Sound avqi
To PointProcess (periodic, cc)... 50 400
Rename... avqi1

```

```

select Sound avqi
plus PointProcess avqi1
percentShimmer = Get shimmer (local)... 0 0 0.0001 0.02 1.3 1.6
shim = percentShimmer*100
shdb = Get shimmer (local_dB)... 0 0 0.0001 0.02 1.3 1.6

# Harmonic-to-noise ratio

select Sound avqi
To Pitch (cc)... 0 75 15 no 0.03 0.45 0.01 0.35 0.14 600
select Sound avqi
plus Pitch avqi
To PointProcess (cc)
Rename... avqi2
select Sound avqi
plus Pitch avqi
plus PointProcess avqi2
voiceReport$ = Voice report... 0 0 75 600 1.3 1.6 0.03 0.45
hnr = extractNumber (voiceReport$, "Mean harmonics-to-noise ratio: ")

# Calculation of the AVQI

avqi = ((3.295-(0.111*cpps)-(0.073*hnr)-(0.213*shim)+(2.789*shdb)-(0.032*slope)+(0.077*tilt))*2.208)+1.797

# -----
# PART 3:
# DRAWINGS ALL THE INFORMATION AND THE GRAPHS.
# -----

# Title and patient information

Erase all
Solid line
Line width... 1
Black
Helvetica
Select inner viewport... 0 8 0 0.5
Font size... 1
Select inner viewport... 0.5 7.5 0.1 0.15
Axes... 0 1 0 1
Text... 0 Left 0.5 Half Script: Youri Maryn (PhD) and Paul Corthals (PhD)
Font size... 12
Select inner viewport... 0.5 7.5 0 0.5
Axes... 0 1 0 1
Text... 0 Left 0.5 Half ##ACOUSTIC VOICE QUALITY INDEX (AVQI) v.02.03#
Font size... 8
Select inner viewport... 0.5 7.5 0 0.5
Axes... 0 1 0 3
Text... 1 Right 2.3 Half %%'name_patient$%'
Text... 1 Right 1.5 Half %%'left_dates$%'
Text... 1 Right 0.7 Half %%'right_dates$%'

# Simple version

if version = 1

# Data

Font size... 10
Select inner viewport... 0.5 7.5 0.5 2
Axes... 0 7 6 0
Text... 0.05 Left 0.5 Half Smoothed cepstral peak prominence (CPPS): ##'cpps:2'#
Text... 0.05 Left 1.5 Half Harmonics-to-noise ratio: ##'hnr:2' dB#
Text... 0.05 Left 2.5 Half Shimmer local: ##'shim:2' \% #
Text... 0.05 Left 3.5 Half Shimmer local dB: ##'shdb:2' dB#
Text... 0.05 Left 4.5 Half Slope of LTAS: ##'slope:2' dB#
Text... 0.05 Left 5.5 Half Tilt of trendline through LTAS: ##'tilt:2' dB#

```

```
Select inner viewport... 0.5 3.8 0.5 2
Draw inner box
Font size... 7
Arrow size... 1
Select inner viewport... 4 7.5 1.25 2
Axes... 0 10 1 0
Paint rectangle... green 0 2.91 0 1
Paint rectangle... red 2.91 10 0 1
Draw arrow... avqi 1 avqi 0
Draw inner box
Marks top every... 1 1 yes yes no
Font size... 16
Select inner viewport... 4 7.5 0.5 1.15
Axes... 0 1 0 1
Text... 0.5 Centre 0.5 Half AVQI: ##'avqi:2'#
```

```
# Copy Praat picture
```

```
Select inner viewport... 0.5 7.5 0 2
Copy to clipboard
```

```
# Illustrated version
```

```
elsif version = 2
```

```
# Oscillogram
```

```
Font size... 7
Select inner viewport... 0.5 5 0.5 2.0
select Sound avqi
Draw... 0 0 0 0 no Curve
Draw inner box
One mark left... minimumSPL no yes no 'minimumSPL:2'
One mark left... maximumSPL no yes no 'maximumSPL:2'
Text left... no Sound pressure level (Pa)
One mark bottom... 0 no yes no 0.00
One mark bottom... durationOnlyVoice no no yes
One mark bottom... durationAll no yes no 'durationAll:2'
Text bottom... no Time (s)
```

```
# Narrow-band spectrogram
```

```
Select inner viewport... 0.5 5 2.3 3.8
select Spectrogram avqi
Paint... 0 0 0 4000 100 yes 50 6 0 no
Draw inner box
One mark left... 0 no yes no 0
One mark left... 4000 no yes no 4000
Text left... no Frequency (Hz)
One mark bottom... 0 no yes no 0.00
One mark bottom... durationOnlyVoice no no yes
One mark bottom... durationAll no yes no 'durationAll:2'
Text bottom... no Time (s)
```

```
# LTAS
```

```
Select inner viewport... 5.4 7.5 2.3 3.8
select Ltas avqi
Draw... 0 4000 minimumSpectrum maximumSpectrum no Curve
Draw inner box
One mark left... minimumSpectrum no yes no 'minimumSpectrum:2'
One mark left... maximumSpectrum no yes no 'maximumSpectrum:2'
Text left... no Sound pressure level (dB/Hz)
One mark bottom... 0 no yes no 0
One mark bottom... 4000 no yes no 4000
Text bottom... no Frequency (Hz)
```



# Power-cepstrogram

```
Select inner viewport... 0.5 5 4.1 5.6
select PowerCepstrogram avqi
Paint: 0, 0, 0, 0, 80, "no", 30, 0, "yes"
Draw inner box
One mark left... 0.00303 no yes no 0.003
One mark left... 0.01667 no yes no 0.017
Text left... no Quefrency (s)
One mark bottom... 0 no yes no 0.00
One mark bottom... durationOnlyVoice no no yes
One mark bottom... durationAll no yes no 'durationAll:2'
Text bottom... no Time (s)
```

# Power-cepstrum

```
Select inner viewport... 5.4 7.5 4.1 5.6
select PowerCepstrum avqi_0_100
Draw... 0.00303 0.01667 0 0 no
Draw tilt line... 0.00303 0.01667 0 0 0.00303 0.01667 Straight Robust
Draw inner box
One mark left... maximumCepstrum no yes no 'maximumCepstrum:2'
Text left... no Amplitude (dB)
One mark bottom... 0.00303 no yes no 0.003
One mark bottom... 0.01667 no yes no 0.017
Text bottom... no Quefrency (s)
```

# Data

```
Font size... 10
Select inner viewport... 0.5 7.5 5.9 7.4
Axes... 0 7 6 0
Text... 0.05 Left 0.5 Half Smoothed cepstral peak prominence (CPPS): ##'cpps:2'#
Text... 0.05 Left 1.5 Half Harmonics-to-noise ratio: ##'hnr:2' dB#
Text... 0.05 Left 2.5 Half Shimmer local: ##'shim:2' \% #
Text... 0.05 Left 3.5 Half Shimmer local dB: ##'shdb:2' dB#
Text... 0.05 Left 4.5 Half Slope of LTAS: ##'slope:2' dB#
Text... 0.05 Left 5.5 Half Tilt of trendline through LTAS: ##'tilt:2' dB#
Select inner viewport... 0.5 3.8 5.9 7.4
Draw inner box
Font size... 7
Arrow size... 1
Select inner viewport... 4 7.5 6.75 7.4
Axes... 0 10 1 0
Paint rectangle... green 0 2.91 0 1
Paint rectangle... red 2.91 10 0 1
Draw arrow... avqi 1 avqi 0
Draw inner box
Marks top every... 1 1 yes yes no
Font size... 16
Select inner viewport... 4 7.5 5.9 6.65
Axes... 0 1 0 1
Text... 0.5 Centre 0.5 Half AVQI: ##'avqi:2'#
```

# Copy Praat picture

```
Select inner viewport... 0.5 7.5 0 7.4
Copy to clipboard
```

endif

# Remove intermediate objects

```
select all
minus Sound cs
minus Sound sv
Remove
```

## Annexe 7 – script de l’AVQI 03.01 (Barsties & Maryn, 2016)

# TITLE OF THE SCRIPT: ACOUSTIC VOICE QUALITY INDEX (AVQI) v.03.01

# Form for introduction and/or parameterization

form Acoustic Voice Quality Index v.03.01

comment >>> It is advocated to estimate someone’s dysphonia severity in both continuous speech (i.e., ‘cs’) and sustained vowel (i.e., ‘sv’) (Maryn et al., 2010). This script therefore runs on these two types of recordings, and it is important to name these recordings ‘cs’ and ‘sv’, respectively.

comment >>> This script automatically (a) searches, extracts and then concatenates the voiced segments of the continuous speech recording to a new sound; (b) concatenates the sustained vowel recording to the new sound, (c) determines the Smoothed Cepstral Peak Prominence, the Shimmer Local, the Shimmer Local dB, the LTAS-slope, the LTAS-tilt and the Harmonics-to-Noise Ratio of the concatenated sound signal, (d) calculates the AVQI-score based on the equation of Barsties & Maryn (2015), and draws the oscillogram, the narrow band spectrogram with LTAS and the powercepstrogram with power cepstrum of the concatenated sound signal to allow further interpretation.

comment >>> To be reliable for the AVQI analysis, it is imperative that the sound recordings are made in an optimal data acquisition conditions.

comment >>> There are two versions in this script: (1) a simple version (only AVQI with data of acoustic measures), and (2) an illustrated version (AVQI with data of acoustic measures and above-mentioned graphs).

choice version: 1

button simple

button illustrated

comment >>> Additional information (optional):

sentence name\_patient

sentence left\_dates\_(birth\_-\_assessment)

sentence right\_dates\_(birth\_-\_assessment)

comment

comment Script credits: Youri Maryn (PhD), Paul Corthals (PhD), and Ben Barsties

endform

Erase all

Select inner viewport... 0.5 7.5 0.5 4.5

Axes... 0 1 0 1

Black

Text special... 0.5 centre 0.6 half Helvetica 12 0 Please wait an instant. Depending on the duration and/or the sample rate of the recorded

Text special... 0.5 centre 0.4 half Helvetica 12 0 sound files, this script takes more or less time to process the sound and search for the AVQI.

# -----

# PART 0:

# HIGH-PASS FILTERING OF THE SOUND FILES.

# -----

select Sound cs

Filter (stop Hann band)... 0 34 0.1

Rename... cs2

select Sound sv

Filter (stop Hann band)... 0 34 0.1

Rename... sv2

# -----

# PART 1:

# DETECTION, EXTRACTION AND CONCATENATION OF

# THE VOICED SEGMENTS IN THE RECORDING

# OF CONTINUOUS SPEECH.

# -----

select Sound cs2

Copy... original

samplingRate = Get sampling frequency

```

intermediateSamples = Get sampling period
Create Sound... onlyVoice 0 0.001 'samplingRate' 0
select Sound original
To TextGrid (silences)... 50 0.003 -25 0.1 0.1 silence sounding
select Sound original
plus TextGrid original
Extract intervals where... 1 no "does not contain" silence
Concatenate
select Sound chain
Rename... onlyLoud
globalPower = Get power in air
select TextGrid original
Remove

select Sound onlyLoud
signalEnd = Get end time
windowBorderLeft = Get start time
windowWidth = 0.03
windowBorderRight = windowBorderLeft + windowWidth
globalPower = Get power in air
voicelessThreshold = globalPower*(30/100)

select Sound onlyLoud
extremeRight = signalEnd - windowWidth
while windowBorderRight < extremeRight
    Extract part... 'windowBorderLeft' 'windowBorderRight' Rectangular 1.0 no
    select Sound onlyLoud_part
    partialPower = Get power in air
    if partialPower > voicelessThreshold
        call checkZeros 0
        if (zeroCrossingRate <> undefined) and (zeroCrossingRate < 3000)
            select Sound onlyVoice
            plus Sound onlyLoud_part
            Concatenate
            Rename... onlyVoiceNew
            select Sound onlyVoice
            Remove
            select Sound onlyVoiceNew
            Rename... onlyVoice
        endif
    endif
    select Sound onlyLoud_part
    Remove
    windowBorderLeft = windowBorderLeft + 0.03
    windowBorderRight = windowBorderLeft + 0.03
    select Sound onlyLoud
endwhile
select Sound onlyVoice

procedure checkZeros zeroCrossingRate

    start = 0.0025
    startZero = Get nearest zero crossing... 'start'
    findStart = startZero
    findStartZeroPlusOne = startZero + intermediateSamples
    startZeroPlusOne = Get nearest zero crossing... 'findStartZeroPlusOne'
    zeroCrossings = 0
    strips = 0

    while (findStart < 0.0275) and (findStart <> undefined)
        while startZeroPlusOne = findStart
            findStartZeroPlusOne = findStartZeroPlusOne + intermediateSamples
            startZeroPlusOne = Get nearest zero crossing... 'findStartZeroPlusOne'
        endwhile
        afstand = startZeroPlusOne - startZero
        strips = strips + 1
        zeroCrossings = zeroCrossings + 1
    endwhile
endprocedure

```

```

        findStart = startZeroPlusOne
    endwhile
    zeroCrossingRate = zeroCrossings/afstand
endproc

# -----
# PART 2:
# DETERMINATION OF THE SIX ACOUSTIC MEASURES
# AND CALCULATION OF THE ACOUSTIC VOICE QUALITY INDEX.
# -----

select Sound sv2
durationVowel = Get total duration
durationStart=durationVowel-3
if durationVowel>3
Extract part... durationStart durationVowel rectangular 1 no
Rename... sv3
elsif durationVowel<=3
Copy... sv3
endif

select Sound onlyVoice
durationOnlyVoice = Get total duration
plus Sound sv3
Concatenate
Rename... avqi
durationAll = Get total duration
minimumSPL = Get minimum... 0 0 None
maximumSPL = Get maximum... 0 0 None

# Narrow-band spectrogram and LTAS

To Spectrogram... 0.03 4000 0.002 20 Gaussian
select Sound avqi
To Ltas... 1
minimumSpectrum = Get minimum... 0 4000 None
maximumSpectrum = Get maximum... 0 4000 None

# Power-cepstrogram, Cepstral peak prominence and Smoothed cepstral peak prominence

select Sound avqi
To PowerCepstrogram... 60 0.002 5000 50
cpps = Get CPPS... no 0.01 0.001 60 330 0.05 Parabolic 0.001 0 Straight Robust
To PowerCepstrum (slice)... 0.1
maximumCepstrum = Get peak... 60 330 None

# Slope of the long-term average spectrum

select Sound avqi
To Ltas... 1
slope = Get slope... 0 1000 1000 10000 energy

# Tilt of trendline through the long-term average spectrum
select Ltas avqi
Compute trend line... 1 10000
tilt = Get slope... 0 1000 1000 10000 energy

# Amplitude perturbation measures
select Sound avqi
To PointProcess (periodic, cc)... 50 400
Rename... avqi1
select Sound avqi
plus PointProcess avqi1
percentShimmer = Get shimmer (local)... 0 0 0.0001 0.02 1.3 1.6
shim = percentShimmer*100
shdb = Get shimmer (local_dB)... 0 0 0.0001 0.02 1.3 1.6

```

```

# Harmonic-to-noise ratio

select Sound avqi
To Pitch (cc)... 0 75 15 no 0.03 0.45 0.01 0.35 0.14 600
select Sound avqi
plus Pitch avqi
To PointProcess (cc)
Rename... avqi2
select Sound avqi
plus Pitch avqi
plus PointProcess avqi2
voiceReport$ = Voice report... 0 0 75 600 1.3 1.6 0.03 0.45
hnr = extractNumber (voiceReport$, "Mean harmonics-to-noise ratio: ")

# Calculation of the AVQI

avqi = (4.152-(0.177*cpps)-(0.006*hnr)-(0.037*shim)+(0.941*shdb)+(0.01*slope)+(0.093*tilt))*2.8902

# -----
# PART 3:
# DRAWINGS ALL THE INFORMATION AND THE GRAPHS.
# -----

# Title and patient information

Erase all
Solid line
Line width... 1
Black
Helvetica
Select inner viewport... 0 8 0 0.5
Font size... 1
Select inner viewport... 0.5 7.5 0.1 0.15
Axes... 0 1 0 1
Text... 0 Left 0.5 Half Script: Youri Maryn (PhD) and Paul Corthals (PhD)
Font size... 12
Select inner viewport... 0.5 7.5 0 0.5
Axes... 0 1 0 1
Text... 0 Left 0.5 Half ##ACOUSTIC VOICE QUALITY INDEX (AVQI) v.03.01#
Font size... 8
Select inner viewport... 0.5 7.5 0 0.5
Axes... 0 1 0 3
Text... 1 Right 2.3 Half %%'name_patient$'%
Text... 1 Right 1.5 Half %%'left_dates$'%
Text... 1 Right 0.7 Half %%'right_dates$'%

# Simple version

if version = 1

    # Data

    Font size... 10
    Select inner viewport... 0.5 7.5 0.5 2
    Axes... 0 7 6 0
    Text... 0.05 Left 0.5 Half Smoothed cepstral peak prominence (CPPS): ##'cpps:2'#
    Text... 0.05 Left 1.5 Half Harmonics-to-noise ratio: ##'hnr:2' dB#
    Text... 0.05 Left 2.5 Half Shimmer local: ##'shim:2' \% #
    Text... 0.05 Left 3.5 Half Shimmer local dB: ##'shdb:2' dB#
    Text... 0.05 Left 4.5 Half Slope of LTAS: ##'slope:2' dB#
    Text... 0.05 Left 5.5 Half Tilt of trendline through LTAS: ##'tilt:2' dB#
    Select inner viewport... 0.5 3.8 0.5 2
    Draw inner box
    Font size... 7
    Arrow size... 1
    Select inner viewport... 4 7.5 1.25 2
    Axes... 0 10 1 0

```

```
Paint rectangle... green 0 2.43 0 1
Paint rectangle... red 2.43 10 0 1
Draw arrow... avqi 1 avqi 0
Draw inner box
Marks top every... 1 1 yes yes no
Font size... 16
Select inner viewport... 4 7.5 0.5 1.15
Axes... 0 1 0 1
Text... 0.5 Centre 0.5 Half AVQI: ##'avqi:2'##
```

```
# Copy Praat picture
```

```
Select inner viewport... 0.5 7.5 0 2
Copy to clipboard
```

```
# Illustrated version
```

```
elsif version = 2
```

```
# Oscillogram
```

```
Font size... 7
Select inner viewport... 0.5 5 0.5 2.0
select Sound avqi
Draw... 0 0 0 0 no Curve
Draw inner box
One mark left... minimumSPL no yes no 'minimumSPL:2'
One mark left... maximumSPL no yes no 'maximumSPL:2'
Text left... no Sound pressure level (Pa)
One mark bottom... 0 no yes no 0.00
One mark bottom... durationOnlyVoice no no yes
One mark bottom... durationAll no yes no 'durationAll:2'
Text bottom... no Time (s)
```

```
# Narrow-band spectrogram
```

```
Select inner viewport... 0.5 5 2.3 3.8
select Spectrogram avqi
Paint... 0 0 0 4000 100 yes 50 6 0 no
Draw inner box
One mark left... 0 no yes no 0
One mark left... 4000 no yes no 4000
Text left... no Frequency (Hz)
One mark bottom... 0 no yes no 0.00
One mark bottom... durationOnlyVoice no no yes
One mark bottom... durationAll no yes no 'durationAll:2'
Text bottom... no Time (s)
```

```
# LTAS
```

```
Select inner viewport... 5.4 7.5 2.3 3.8
select Ltas avqi
Draw... 0 4000 minimumSpectrum maximumSpectrum no Curve
Draw inner box
One mark left... minimumSpectrum no yes no 'minimumSpectrum:2'
One mark left... maximumSpectrum no yes no 'maximumSpectrum:2'
Text left... no Sound pressure level (dB/Hz)
One mark bottom... 0 no yes no 0
One mark bottom... 4000 no yes no 4000
Text bottom... no Frequency (Hz)
```

```
# Power-cepstrogram
```

```
Select inner viewport... 0.5 5 4.1 5.6
select PowerCepstrogram avqi
Paint: 0, 0, 0, 0, 80, "no", 30, 0, "yes"
Draw inner box
```

One mark left... 0.00303 no yes no 0.003  
One mark left... 0.01667 no yes no 0.017  
Text left... no Quefreny (s)  
One mark bottom... 0 no yes no 0.00  
One mark bottom... durationOnlyVoice no no yes  
One mark bottom... durationAll no yes no 'durationAll:2'  
Text bottom... no Time (s)

# Power-cepstrum

Select inner viewport... 5.4 7.5 4.1 5.6  
select PowerCepstrum avqi\_0\_100  
Draw... 0.00303 0.01667 0 0 no  
Draw tilt line... 0.00303 0.01667 0 0 0.00303 0.01667 Straight Robust  
Draw inner box  
One mark left... maximumCepstrum no yes no 'maximumCepstrum:2'  
Text left... no Amplitude (dB)  
One mark bottom... 0.00303 no yes no 0.003  
One mark bottom... 0.01667 no yes no 0.017  
Text bottom... no Quefreny (s)

# Data

Font size... 10  
Select inner viewport... 0.5 7.5 5.9 7.4  
Axes... 0 7 6 0  
Text... 0.05 Left 0.5 Half Smoothed cepstral peak prominence (CPPS): ##'cpps:2'#  
Text... 0.05 Left 1.5 Half Harmonics-to-noise ratio: ##'hnr:2' dB#  
Text... 0.05 Left 2.5 Half Shimmer local: ##'shim:2' \% #  
Text... 0.05 Left 3.5 Half Shimmer local dB: ##'shdb:2' dB#  
Text... 0.05 Left 4.5 Half Slope of LTAS: ##'slope:2' dB#  
Text... 0.05 Left 5.5 Half Tilt of trendline through LTAS: ##'tilt:2' dB#  
Select inner viewport... 0.5 3.8 5.9 7.4  
Draw inner box  
Font size... 7  
Arrow size... 1  
Select inner viewport... 4 7.5 6.75 7.4  
Axes... 0 10 1 0  
Paint rectangle... green 0 2.91 0 1  
Paint rectangle... red 2.91 10 0 1  
Draw arrow... avqi 1 avqi 0  
Draw inner box  
Marks top every... 1 1 yes yes no  
Font size... 16  
Select inner viewport... 4 7.5 5.9 6.65  
Axes... 0 1 0 1  
Text... 0.5 Centre 0.5 Half AVQI: ##'avqi:2'#

# Copy Praat picture

Select inner viewport... 0.5 7.5 0 7.4  
Copy to clipboard

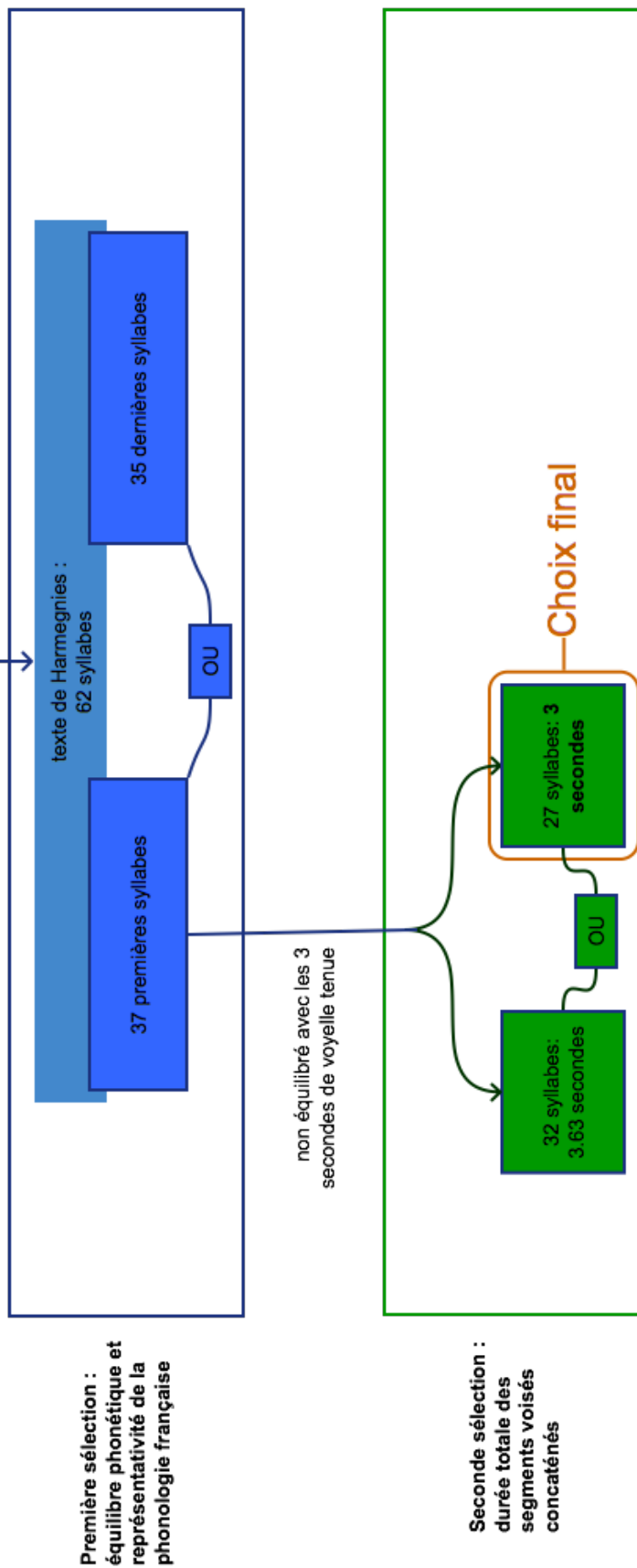
endif

# Remove intermediate objects

select all  
minus Sound sv  
minus Sound cs  
minus Sound avqi  
Remove

## Quelle longueur doit avoir l'échantillon ?

idéalement: 34 syllabes, en néerlandais (Barsties & Maryn, 2015)





## Annexe 9 – script composé pour créer les fichiers de l'évaluation perceptive

```
select Sound sv
Filter (stop Hann band)... 0 34 0.1
Rename... sv2
select Sound sv2
durationVowel = Get total duration
durationStart=durationVowel-3
if durationVowel>3
Extract part... durationStart durationVowel rectangular 1 no
Rename... sv3
elsif durationVowel<=3
Copy... sv3
endif

select all
minus Sound cs
minus Sound sv3
Remove

select Sound sv3
Rename... sv

select all

### concatWithPause.praat
#
# Description:
# Concatenate sounds with a user defined pause (silence) duration between sounds
# Options:
# - resampling of sounds if they have different sampling frequencies
# (only copies are resampled, not the original sounds; choice of 7 (re-) sampling frequencies)
# - custom order of sounds
#
# Usage:
# - select 2 or more sounds
# - start the script
# - if selected sounds have different sampling frequencies you are asked
# to specify a resampling frequency
# - specify pause (silence) duration in seconds
# - check option if you want to rearrange sounds (default: order of appearance in object list)
# - if checked: specify custom order
#
### jm, 2013-11-17

# determine number of selected sounds
no_Snd = numberOfSelected ("Sound")

# check if 2 or more sounds are selected
if no_Snd = 0
    exit No Sound selected.
elsif no_Snd = 1
    exit Only one Sound selected.
endif

# get IDs and names of selected sounds
for i to no_Snd
    soundID [i] = selected ("Sound", i)
    soundName$ [i] = selected$ ("Sound", i)
endfor

# check for varying sampling frequencies (fs)
no_fs = 0
for i to no_Snd
    selectObject (soundID [i])
    fs = do ("Get sampling frequency")
```

```

# get some fs at first pass
if no_fs = 0
    fsOcc [1] = fs
    no_fs = 1
# get other fs if there are any
# and count number of different fs (no_fs)
else
    # assume we found a new fs
    newfs = 1
    for k to no_fs
        if fs = fsOcc [k]
            # if the new fs equals to an already occurring fs we didn't find a new fs'
            newfs = 0
        endif
    endfor
    # if we indeed found a new fs increment counter and add new fs to array of occurring fs
    if newfs = 1
        no_fs = no_fs + 1
        fsOcc [no_fs] = fs
    endif
endif
endifor

# get user input
beginPause ("Concatenate Sounds")
# if varying fs occur ask for resampling frequency
if no_fs > 1
    comment ("The Sounds you selected have different sampling frequencies")
    comment ("and can't be concatenated without resampling.")
    comment ("If you want to concatenate resampled copies of the original Sounds")
    comment ("you must choose a (re-) sampling frequency.")
    optionMenu ("Sampling frequency", 5)
        option ("8000")
        option ("10000")
        option ("11025")
        option ("16000")
        option ("22050")
        option ("44100")
        option ("48000")
    comment ("If you want to abort click Cancel.")
    comment ("")
    comment ("")
endif
# ask for pause/silence duration
comment ("Specify the silence duration between Sounds.")
real ("Silence duration (s)", 0.5)
# shall we rearrange sounds?
comment ("By default, Sounds are concatenated in the order they appear")
comment ("in the list. Check below if you want to change the order.")
boolean ("Change order", 0)
clicked = endPause ("Cancel", "OK", 2, 1)
if clicked = 1
    goto FINISH
endif

# for reordering we need a new array (see below); we'll use that new array for the
# default case (no reordering) too in order to stay consistent
if change_order = 0
    # simply copy old array to new array
    for i to no_Snd
        sound [i] = soundID [i]
    endfor
# in the case of reordering it's less simple...
elseif change_order = 1
    # ask the user for new order
    beginPause ("Reorder Sounds")
        for i to no_Snd

```

```

                optionMenu ("Sound 'i'", i)
                for k to no_Snd
                    option (soundName$ [k])
                endfor
            endfor
        clicked = endPause ("Cancel", "OK", 2, 1)
        if clicked = 1
            goto FINISH
        else
            # use the numeric specification of the chosen option as index
            # for the old array to generate the new array
            for i to no_Snd
                sound [i] = soundID [sound_'i']
            endfor
        endif
    endif

# assign the standard fs to a consistent variable;
# standard fs is either the fs of all sounds or the chosen resampling frequency
if no_fs = 1
    sampling_frequency = fsOcc [1]
else
    sampling_frequency = number (sampling_frequency$)
endif

# generate pause
silence = do ("Create Sound from formula...", "silence", 1, 0, silence_duration, sampling_frequency, "0")

# copy/resample sounds and pause to achieve desired order
for i to no_Snd
    selectObject (sound [i])
    thefs = do ("Get sampling frequency")
    # if resampling is on and sound needs resampling generate a resampled copy
    if no_fs > 1 && thefs <> sampling_frequency
        sound [i] = do ("Resample...", sampling_frequency, 50)
    # if sound doesn't need resampling generate a simple copy'
    else
        sound [i] = do ("Copy...", "tmp")
    endif
    # generate a copy of pause after each sound (except the last)
    if i < no_Snd
        selectObject (silence)
        sil [i] = do ("Copy...", "tmp")
    endif
endfor
removeObject (silence)

# now everything is ready to go (correct order, consistent fs), so let's select it...
selectObject (sound [1])
for i to no_Snd - 1
    plusObject (sil [i])
    plusObject (sound [i+1])
endfor
# ... concatenate it ...
concat = do ("Concatenate")
do ("Rename...", "concatenated")
# ... and delete it
selectObject (sound [1])
for i to no_Snd - 1
    plusObject (sil [i])
    plusObject (sound [i+1])
endfor
do ("Remove")
selectObject (concat)

label FINISH
exit

```

## Annexe 10 – script adapté pour l'évaluation perceptive du critère G de la GRBAS

```
#####
#####
## This script implements the Grade parameter of the GRBASI voice quality assessment.
## Randomized presentation of wav files, evaluation form with 4 levels.
## This script has been adapted from Valerie Delvaux's (2016) adaptation of the script available at
## http://praatpfanne.lingphon.net/praat-ressourcen/resources-english,
## by Timothy Pommée for the validation phase of the AVQI in the french language. January 2018.
#####
#####
form Input
  word Path
  word Subject xx
  word Judge yy
endform

fileListID = Create Strings as file list: "FileList", path$ + "*.wav"
Randomize
numberOfFiles = Get number of strings

resultTableID = Create Table with column names: "Results_" + subject$, numberOfFiles,
... "Subject Judge Stimulus Grade"

demo Erase all
demo Select outer viewport: 0, 100, 0, 100
demo Axes: 0, 100, 0, 100
demo Text: 50, "centre", 50, "half", "Click to start!"
demoWaitForInput ()

for i to numberOfFiles
  selectObject: fileListID
  fileName$ = Get string: i
  stimulusID = Read from file: path$ + fileName$
  demo Erase all
  demo 12
  demo Times
  demo Grey
  demo Text: 0, "left", 99, "half", string$ (i) + "/" + string$ (numberOfFiles)
  demo 18
  demo Black
  demo Text: 25, "right", 50, "half", "Grade"
  @drawButtons: 46, 54, 50
  demo Paint rectangle: "navy", 41, 79, 81, 89
  demo White
  demo Text: 60, "centre", 85, "half", "Replay"
  demo Grey
  demo Paint rectangle: "maroon", 41, 79, 11, 19
  demo Text: 60, "centre", 15, "half", "Next"
  asynchronous Play
  grade = 10
  while demoWaitForInput ()
    if demoClicked ()
      if demoClickedIn (41, 49, 46, 54)
        grade = 0
        @drawButtons: 46, 54, 50
        @selectButton: 41, 49, 46, 54
      elseif demoClickedIn (51, 59, 46, 54)
        grade = 1
        @drawButtons: 46, 54, 50
        @selectButton: 51, 59, 46, 54
      elseif demoClickedIn (61, 69, 46, 54)
        grade = 2
        @drawButtons: 46, 54, 50
        @selectButton: 61, 69, 46, 54
      elseif demoClickedIn (71, 79, 46, 54)
```

```

grade = 3
@drawButtons: 46, 54, 50
@selectButton: 71, 79, 46, 54

elseif demoClickedIn (41, 79, 81, 89)
  asynchronous Play
endif
if grade < 10
  demo White
  demo Text: 60, "centre", 15, "half", "Next"
  if demoClickedIn (41, 79, 11, 19)
    select resultTableID
    Set string value: i, "Subject", subject$
    Set string value: i, "Judge", judge$
    Set string value: i, "Stimulus", fileName$
    Set numeric value: i, "Grade", grade
    goto NEXT
  endif
endif
endif
endwhile
label NEXT
removeObject: stimulusID
selectObject: resultTableID
Save as tab-separated file: "Results_" + subject$ + ".txt"

endifor

demo Erase all
demo Black
demo Text: 50, "centre", 50, "half", "Thank you!"
removeObject: fileListID

procedure drawButtons: .row1, .row2, .row3
  demo Paint rectangle: "blue", 41, 49, .row1, .row2
  demo Paint rectangle: "blue", 51, 59, .row1, .row2
  demo Paint rectangle: "blue", 61, 69, .row1, .row2
  demo Paint rectangle: "blue", 71, 79, .row1, .row2
  demo White
  demo Text: 45, "centre", .row3, "half", "0"
  demo Text: 55, "centre", .row3, "half", "1"
  demo Text: 65, "centre", .row3, "half", "2"
  demo Text: 75, "centre", .row3, "half", "3"
endproc

procedure selectButton: .x1, .x2, .y1, .y2
  demo Magenta
  demo Line width: 3
  demo Draw rectangle: .x1+1, .x2-1, .y1+1, .y2-1
  demo Line width: 1
endproc

```

---

**Résumé.** *L'AVQI 03.01 appliqué à la langue française : validité et précision diagnostique dans la population adulte. Cette évaluation objective corrèle-t-elle avec la plainte du patient ?*

L'Acoustic Voice Quality Index (AVQI) est un outil d'évaluation quantitative du degré général de la dysphonie. Son calcul comprend six paramètres acoustiques collectés via le logiciel gratuit Praat. Il est réalisé sur base des enregistrements d'une voyelle tenue ainsi que d'une partie d'un texte lu à voix haute. Les qualités psychométriques de cet outil ont été confirmées dans de nombreuses publications, et ce en plusieurs langues. Cependant, sa validation en français repose sur une petite cohorte de patients non francophones de naissance. De plus, la version 03.01 de l'AVQI n'a pas encore été validée sur un échantillon en français. Par conséquent, ce travail comporte trois étapes principales. D'abord, une phase préparatoire vise à identifier l'extrait optimal d'un texte français couramment utilisé en clinique journalière. Le second objectif est d'évaluer la validité critérielle ainsi que la précision diagnostique de l'AVQI 03.01 appliqué sur l'extrait identifié dans la phase préparatoire. Enfin, nous étudions la corrélation entre les scores AVQI et le handicap perçu par le patient, évalué à l'aide du VHI.

Ce travail nous permet entre autres d'inviter à la prudence dans l'utilisation de la technologie et de l'informatique pour l'évaluation vocale. À ce jour, l'évaluation à l'écoute par le thérapeute semble rester le seul outil permettant de percevoir dans sa totalité la réalité complexe et multidimensionnelle de la voix.

---

**Abstract.** *The AVQI 03.01 applied to the French language: validity and diagnostic accuracy in the adult population. Does this objective assessment correlate with the patient's complaint?*

The Acoustic Voice Quality Index (AVQI) is a tool for quantitative assessment of the overall severity of dysphonia. Its computation includes six acoustic parameters, which are all carried out by the freeware Praat. It is based on the recordings of a sustained vowel and a part of a text read aloud. The psychometric qualities of this tool have been confirmed by numerous publications in various languages. However, its validation in French relies on a small cohort of patients, who were not native French speakers. Furthermore, the version 03.01 of the AVQI has not yet been validated at all on a French sample. Hence, this work consists of three main stages. First a preparatory phase is intended to identify the optimal part of a commonly used French text to be used for the validation study. The second aim is to assess the criterion-related concurrent validity and the diagnostic accuracy of the AVQI 03.01 applied to this text extract. Eventually, we investigate the correlation between the AVQI scores and the handicap perceived by the patient, as assessed by the VHI.

This work allows us to call for caution in the use of technology and computer science for vocal assessment. To date, the perceptual assessment by the clinician still seems to be the only way to catch the complex and multidimensional vocal reality in its entirety.

---

**Mots-clés.** Acoustic Voice Quality Index – dysphonie – analyse acoustique vocale – évaluation vocale – Français – qualité de vie

**Keywords.** Acoustic Voice Quality Index – dysphonia – acoustic voice analysis – voice assessment – French – quality of life