

## UNIVERSITÉ PARIS III — SORBONNE-NOUVELLE

UFR de Linguistique et Phonétique Générales et Appliquées (ILPGA)  
Ecole Doctorale 268 "Langage et Langues : Description, théorisation, transmission"

### THESE

Pour obtenir le grade de  
DOCTEUR DE L'UNIVERSITÉ PARIS III

**Discipline : Phonétique**

## **Etudes multiparamétriques de la voix et de la parole après cordectomie laser par voie endoscopique de type II-III**

Présentée et soutenue publiquement par  
**Lucille Wallet**

le 17 décembre 2012

Composition du Jury :

Madame Jacqueline Vaissière (Directeur), Professeur  
Madame Lise Crevier-Buchman (Co-directeur), Chargée de Recherche, CNRS  
Monsieur Antoine Giovanni (Rapporteur), Professeur  
Monsieur Rudolph Sock (Rapporteur), Professeur  
Monsieur Didier Demolin (Examinateur), Professeur  
Monsieur Stéphane Hans (Examinateur), Maître de Conférences



Lucille Wallet, 2012

**Etudes multiparamétriques de la voix et de la parole  
après cordectomie laser par voie endoscopique de  
type II-III**  
xx+308 p.

Ce document a été préparé avec LyX, un traitement de texte utilisant L<sup>A</sup>T<sub>E</sub>X 2<sub>&</sub>, et la classe *memoir* de Peter Wilson, téléchargeable sur <http://www.ctan.org/tex-archive/macros/latex/contrib/memoir/>. Les références bibliographiques ont été préparées à l'aide de BIBT<sub>E</sub>X, d'Oren Patashnik et du style *phonology* créé par Jean-Marc Beltzung. Les jeux de caractères utilisés dans ce document sont Time New Roman 11pt, Computer Modern Sans Serif (cmss) gras étendu (bx) 30pt pour les titres de chapitres et Tipa 11pt pour les transcriptions phonétiques.

# Table des matières

<b>Table des matières</b>	<b>iii</b>
<b>Table des figures</b>	<b>xi</b>
<b>Liste des tableaux</b>	<b>xv</b>
<b>Liste des symboles</b>	<b>xx</b>
<b>Remerciements</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<b>1 Anatomie et physiologie du système phonatoire</b>	<b>9</b>
1.1 Introduction . . . . .	10
1.2 Le système respiratoire . . . . .	11
1.2.1 Les mécanismes respiratoires . . . . .	11
1.2.2 Les muscles respiratoires . . . . .	12
1.2.2.1 Muscles de l'inspiration . . . . .	12
1.2.2.2 Muscles de l'expiration . . . . .	13
1.2.3 Respiration et phonation . . . . .	13
1.3 Le système phonatoire . . . . .	14
1.3.1 Configuration interne du larynx . . . . .	14
1.3.2 Les muscles du larynx . . . . .	15
1.3.2.1 Les muscles laryngés intrinsèques . . . . .	15
1.3.2.2 Les muscles laryngés extrinsèques . . . . .	17
1.3.3 Les cartilages du larynx . . . . .	18
1.3.3.1 Les structures cartilagineuses . . . . .	19
1.3.3.2 L'articulation crico-aryténoïdienne . . . . .	19
1.3.4 Les plis vocaux . . . . .	20
1.3.4.1 L'espace glottique . . . . .	20
1.3.4.2 Structure interne des plis vocaux . . . . .	21
1.3.4.3 Modèles mécaniques des plis vocaux . . . . .	22
1.3.5 Physiologie de la phonation . . . . .	25
1.3.5.1 Les théories de la phonation : aperçu historique . . . . .	25
1.3.5.2 La théorie actuelle de la phonation . . . . .	26
1.3.5.3 Typologie d'un cycle de vibration des plis vocaux . . . . .	27
1.3.5.4 Les paramètres acoustiques de la voix . . . . .	28
1.4 Le système articulatoire . . . . .	29
1.4.1 Les organes supraglottiques . . . . .	29

1.4.2	La théorie source-filtre (Fant 1960) . . . . .	30
1.5	Conclusion . . . . .	31
<b>2</b>	<b>Les chirurgies du larynx</b>	<b>33</b>
2.1	Introduction . . . . .	34
2.2	Les cancers du larynx . . . . .	34
2.2.1	Infiltration tumorale . . . . .	34
2.2.2	La classification TNM Internationale . . . . .	35
2.3	Préliminaires . . . . .	36
2.3.1	Objectifs . . . . .	36
2.3.2	Historique des chirurgies . . . . .	38
2.4	La chirurgie partielle verticale du plan glottique . . . . .	38
2.4.1	La cordectomie . . . . .	39
2.4.1.1	Principe chirurgical . . . . .	39
2.4.2	La laryngectomie frontolatérale . . . . .	39
2.4.2.1	Principe chirurgical . . . . .	39
2.4.2.2	Conséquences sur la phonation . . . . .	40
2.4.3	La laryngectomie frontale antérieure avec épiglottoplastie (dite de Tucker et Kambic) . . . . .	41
2.4.3.1	Principe chirurgical . . . . .	41
2.4.3.2	Conséquences sur la phonation . . . . .	41
2.4.4	L'hémilaryngectomie . . . . .	42
2.4.4.1	Principe chirurgical . . . . .	42
2.4.4.2	Conséquences sur la phonation . . . . .	42
2.5	La chirurgie partielle horizontale . . . . .	43
2.5.1	Les laryngectomies partielles supraglottiques . . . . .	43
2.5.1.1	Principe chirurgical . . . . .	43
2.5.1.2	Conséquences sur la phonation . . . . .	44
2.5.2	Les laryngectomies partielles supracricoïdiennes (LPSC) . . . . .	44
2.5.2.1	Principes chirurgicaux . . . . .	44
2.5.2.2	Conséquences sur la phonation . . . . .	46
2.6	Conclusion . . . . .	47
<b>3</b>	<b>La cordectomie laser : approche multiparamétrique</b>	<b>49</b>
3.1	Introduction . . . . .	50
3.2	Principes chirurgicaux . . . . .	50
3.2.1	La cordectomie de type I (ou sous-épithéliale) . . . . .	50
3.2.2	La cordectomie de type II (ou sous-ligamentaire) . . . . .	51
3.2.3	La cordectomie de type III (ou transmusculaire) . . . . .	51
3.2.4	La cordectomie de type IV (ou totale) . . . . .	52
3.2.5	Cordectomie de type V (ou étendue) . . . . .	52
3.2.5.1	Cordectomie de type Va . . . . .	53
3.2.5.2	Cordectomie de type Vb . . . . .	53
3.2.5.3	Cordectomie de type Vc . . . . .	53

3.2.5.4	Cordectomie de type Vd . . . . .	54
3.3	Revue de la littérature . . . . .	54
3.3.1	Les populations étudiées . . . . .	55
3.3.2	Corpus . . . . .	56
3.3.3	Méthodologies . . . . .	56
3.3.3.1	Analyses acoustiques . . . . .	56
3.3.3.2	Analyses perceptives . . . . .	57
3.3.3.3	Analyses aérodynamiques . . . . .	58
3.3.3.4	Analyses morphologiques . . . . .	58
3.4	Les analyses perceptives . . . . .	66
3.4.1	Définition . . . . .	66
3.4.2	Les types d'échelles . . . . .	66
3.4.2.1	L'échelle d'Osgood . . . . .	66
3.4.2.2	L'échelle de Voiers . . . . .	67
3.4.2.3	L'échelle GRBAS . . . . .	67
3.4.2.4	L'échelle Laver's Voice Profil Analysis Sheme (VPAS) . . . . .	68
3.4.2.5	Le Stockholm Voice Evaluation Consensus Model (SVEC) . . . . .	68
3.4.3	Applications en pathologie . . . . .	68
3.5	Les autoévaluations . . . . .	69
3.5.1	Les types d'échelles . . . . .	69
3.5.1.1	Le Medical Outcomes Study Short Form 36 (MOS SF 36) . . . . .	70
3.5.1.2	Le Voice Handicap Index (VHI) . . . . .	70
3.5.1.3	Le Voice Related Quality Of Life (V-RQOL) . . . . .	70
3.5.2	Applications en pathologie . . . . .	70
3.6	Les analyses acoustiques . . . . .	71
3.6.1	Introduction . . . . .	71
3.6.2	Qu'est ce qu'une onde acoustique ? . . . . .	71
3.6.2.1	Le son pur . . . . .	72
3.6.2.2	Le son complexe . . . . .	72
3.6.3	Analyses fréquentielles . . . . .	73
3.6.3.1	Préambules . . . . .	73
3.6.3.2	Un outil : le spectrogramme . . . . .	74
3.6.4	Principaux paramètres utilisés en pathologie . . . . .	76
3.6.4.1	La fréquence fondamentale ( $F_0$ ) . . . . .	76
3.6.4.2	L'intensité sonore . . . . .	77
3.6.4.3	Le jitter . . . . .	78
3.6.4.4	Le shimmer . . . . .	79
3.6.4.5	Les mesures de bruit de turbulence : le HNR et le NNE . . . . .	80
3.7	Les analyses aérodynamiques . . . . .	81
3.7.1	Le temps maximum de phonation (TMP) . . . . .	81
3.7.2	La pression sous-glottique (PSG) . . . . .	81
3.7.3	La pression intra-orale (PIO) . . . . .	82

3.7.4	Le débit d'air buccal (DAB) . . . . .	83
3.8	Les analyses morphologiques . . . . .	84
3.8.1	La fibroscopie . . . . .	85
3.8.2	L'électroglottographe (EGG) . . . . .	85
3.9	Conclusion . . . . .	86
<b>4</b>	<b>Population et Corpus</b>	<b>89</b>
4.1	Introduction . . . . .	90
4.2	Population . . . . .	90
4.2.1	Les patients . . . . .	91
4.2.1.1	Période de recueil de données . . . . .	91
4.2.1.2	Le choix des patients . . . . .	91
4.2.1.3	L'âge . . . . .	92
4.2.1.4	Les lésions . . . . .	92
4.2.1.5	Antécédents et facteurs de risque . . . . .	92
4.2.1.6	TraITEMENT et suites postopératoires . . . . .	92
4.2.2	Les témoins . . . . .	94
4.3	Motivation d'un corpus clinique . . . . .	94
4.3.1	Les voyelles tenues . . . . .	95
4.3.2	Corrélatifs acoustico-articulatoires des voyelles : un argument dans le choix du protocole d'étude . . . . .	95
4.3.3	La parole en contexte . . . . .	96
4.4	Présentation succincte du corpus . . . . .	97
4.4.1	Données acoustiques . . . . .	97
4.4.1.1	Les voyelles tenues . . . . .	97
4.4.1.2	Les logatomes . . . . .	98
4.4.2	Données aérodynamiques . . . . .	98
4.4.3	Données EGG . . . . .	98
4.5	Outils d'analyses . . . . .	98
4.5.1	Présentation de Praat . . . . .	99
4.5.2	Présentation de EVA et Phonetedit . . . . .	99
4.6	Conclusion . . . . .	101
<b>5</b>	<b>Cinq études relatives à la F<sub>0</sub> et à l'intensité de la voix</b>	<b>103</b>
5.1	Introduction . . . . .	105
5.2	Notes préliminaires sur les tests statistiques . . . . .	105
5.2.1	Paramètres communs de statistique descriptive . . . . .	106
5.2.1.1	La moyenne arithmétique . . . . .	106
5.2.1.2	La variance . . . . .	106
5.2.1.3	L'écart-type . . . . .	106
5.2.2	Les Tests statistiques . . . . .	107
5.2.2.1	Hypothèses et seuil de tolérance alpha . . . . .	107
5.2.2.2	Test de normalité des échantillons . . . . .	108
5.2.2.3	Test d'homogénéité des variances . . . . .	109

5.2.2.4	Les tests de Student . . . . .	110
5.2.2.5	L'ANOVA . . . . .	111
5.3	Etude n°1 : la $F_0$ intrinsèque et le jitter des voyelles tenues isolées . . . . .	112
5.3.1	Méthodologie . . . . .	112
5.3.1.1	Recueil des données . . . . .	112
5.3.1.2	Mesures et paramètres étudiés . . . . .	112
5.3.1.3	Tests statistiques préalables . . . . .	112
5.3.2	Présentation des résultats . . . . .	115
5.3.2.1	La fréquence fondamentale . . . . .	115
5.3.2.2	Le jitter . . . . .	117
5.3.3	Discussion des résultats . . . . .	118
5.3.3.1	La fréquence fondamentale . . . . .	118
5.3.3.2	Le jitter . . . . .	123
5.4	Etude n°2 : la plage de variation de $F_0$ . . . . .	126
5.4.1	Méthodologie . . . . .	126
5.4.1.1	Présentation du corpus synthétique . . . . .	126
5.4.1.2	Recueil de données . . . . .	130
5.4.1.3	Elimination des stimuli problématiques . . . . .	130
5.4.1.4	Mesures et paramètres étudiés . . . . .	131
5.4.1.5	Une échelle logarithmique : le demi-ton . . . . .	131
5.4.2	Présentation des résultats . . . . .	131
5.4.2.1	Comparaison moyenne théorique vs observée . . . . .	131
5.4.2.2	Comparaison inter-groupes . . . . .	136
5.4.3	Discussion des résultats . . . . .	140
5.4.3.1	Echelle hertzienne ou échelle logarithmique ? . . . . .	140
5.4.3.2	Comparaison moyenne théorique vs observée . . . . .	140
5.4.3.3	Comparaison inter-groupes . . . . .	142
5.5	Etude n°3 : l'intensité intrinsèque et le shimmer sur des voyelles tenues isolées . . . . .	144
5.5.1	Méthodologie . . . . .	144
5.5.1.1	Recueil des données . . . . .	144
5.5.1.2	Mesures et paramètres étudiés . . . . .	144
5.5.1.3	Tests statistiques préalables . . . . .	145
5.5.2	Présentation des résultats . . . . .	148
5.5.2.1	L'intensité . . . . .	148
5.5.2.2	Le shimmer . . . . .	149
5.5.3	Discussion des résultats . . . . .	150
5.5.3.1	L'intensité . . . . .	150
5.5.3.2	Le shimmer . . . . .	155
5.6	Excursus : intensité intrinsèque et $F_0$ intrinsèque des voyelles tenues . . . . .	158
5.6.1	Etude n°4 : l'intensité intrinsèque . . . . .	158
5.6.1.1	Méthodologie . . . . .	160
5.6.1.2	Résultats . . . . .	161
5.6.1.3	Discussion des résultats . . . . .	165

5.6.2	Etude n°5 : la $F_0$ intrinsèque . . . . .	167
5.6.2.1	Méthodologie . . . . .	168
5.6.2.2	Résultats . . . . .	169
5.6.2.3	Discussion des résultats . . . . .	170
5.7	Conclusion . . . . .	171
<b>6</b>	<b>Quatre études relatives au voisement</b>	<b>175</b>
6.1	Introduction . . . . .	177
6.2	Définitions du voisement . . . . .	178
6.2.1	Définition phonologique . . . . .	178
6.2.1.1	Les traits laryngaux . . . . .	178
6.2.1.2	Typologie des processus phonologiques impliquant le trait de voisement . . . . .	180
6.2.2	Définition aérodynamique . . . . .	184
6.2.3	Définition articulatoire . . . . .	186
6.3	Les corrélats acoustiques du voisement . . . . .	187
6.3.1	Le temps d'établissement de voisement . . . . .	187
6.3.2	Paramètres acoustiques secondaires liés au voisement . . . . .	190
6.4	Les occlusives orales du français . . . . .	191
6.4.1	Caractéristiques articulatoires . . . . .	191
6.4.2	Caractéristiques acoustiques . . . . .	192
6.4.3	Caractéristiques aérodynamiques . . . . .	193
6.5	Etude n°6 : études aérodynamiques . . . . .	193
6.5.1	Les phases aérodynamiques des occlusives . . . . .	193
6.5.2	Etude préliminaire sur les fuites glottiques des voyelles isolées . . . . .	195
6.5.2.1	Tests statistiques préalables . . . . .	195
6.5.2.2	Méthodologie . . . . .	196
6.5.2.3	Présentation des résultats . . . . .	197
6.5.2.4	Discussion . . . . .	201
6.5.3	Etudes qualitatives des courbes de la PIO et du DAB des logatomes . . . . .	204
6.5.3.1	Description des patterns de PIO observés . . . . .	204
6.5.3.2	Observation des courbes générales du DAB des logatomes . . . . .	207
6.5.3.3	Synchronisation des courbes de PIO et de DAB . . . . .	209
6.5.4	Etude du trait de voisement des occlusives en contexte . . . . .	213
6.5.4.1	Méthodologie . . . . .	213
6.5.4.2	Présentation des résultats . . . . .	215
6.5.4.3	Discussion . . . . .	219
6.6	Etude n°7 : études acoustiques . . . . .	229
6.6.1	Méthodologie . . . . .	229
6.6.1.1	Procédure initiale . . . . .	229
6.6.1.2	Segmentation des données . . . . .	230
6.6.1.3	Mesures acoustiques . . . . .	232

6.6.2	Présentation des résultats . . . . .	234
6.6.3	Discussion . . . . .	242
6.7	Etude n°8 : études électroglottographiques (EGG) . . . . .	250
6.7.1	Préliminaires . . . . .	250
6.7.1.1	Analyse du signal EGG . . . . .	250
6.7.1.2	Les paramètres glottiques . . . . .	252
6.7.2	Etude préliminaire sur le quotient ouvert des voyelles isolées . . . . .	254
6.7.2.1	Méthodologie . . . . .	254
6.7.2.2	Difficultés rencontrées . . . . .	255
6.7.2.3	Présentation des résultats . . . . .	260
6.7.2.4	Discussion . . . . .	261
6.7.3	Le quotient ouvert peut-il être considéré comme un paramètre robuste pour l'opposition de voisement ? Le cas des occlusives en contexte . . . . .	263
6.7.3.1	Méthodologie . . . . .	263
6.7.3.2	Présentation des résultats . . . . .	264
6.7.3.3	Discussion . . . . .	265
6.8	Etude n°9 : étude perceptive . . . . .	267
6.8.1	Méthodologie . . . . .	268
6.8.2	Présentation des résultats . . . . .	269
6.8.2.1	La position initiale . . . . .	269
6.8.2.2	La position intervocalique . . . . .	271
6.8.3	Discussion . . . . .	273
6.9	Conclusion . . . . .	275
<b>Conclusion générale</b>		<b>279</b>
<b>Bibliographie</b>		<b>281</b>



# Table des figures

1.1	Les organes participant à la production de la parole . . . . .	10
1.2	Vues antérieure et latérale des muscles intercostaux internes et externes . . . . .	12
1.3	Action des muscles respiratoires durant l'inspiration et l'expiration . . . . .	13
1.4	Vue postérieure du larynx montrant les étages laryngés . . . . .	15
1.5	Vues postérieure et latérale des muscles intrinsèques du larynx . . . . .	16
1.6	Muscles extrinsèques du larynx (sous et sus-hyoïdiens) . . . . .	18
1.7	Vue post-médiale du cartilage arytenoïde . . . . .	19
1.8	Vue supérieure schématique de l'articulation crico-aryténoidienne . . . . .	20
1.9	Schéma d'une coupe frontale d'un pli vocal . . . . .	22
1.10	Deux représentations des plis vocaux dans un modèle à une masse . . . . .	23
1.11	Deux représentations des plis vocaux dans un modèle à deux masses . . . . .	24
1.12	Modèle des plis vocaux à trois masses . . . . .	25
1.13	Schéma d'un cycle complet de vibration des plis vocaux . . . . .	27
1.14	Vue sagittale de l'appareil vocal et des différents organes vocaux . . . . .	29
1.15	Représentations des spectres correspondant aux différentes étapes de la théorie source filtre . . . . .	31
2.1	Les différentes infiltrations tumorales . . . . .	35
2.2	Cordectomie . . . . .	39
2.3	Laryngectomie frontolatérale . . . . .	40
2.4	Laryngectomie frontale antérieure avec épiglottoplastie . . . . .	41
2.5	Hémilaryngectomie . . . . .	42
2.6	Laryngectomie partielle supraglottique . . . . .	43
2.7	Laryngectomie supracricoïdienne avec crico-hyoïdo-épiglottopexie . . . . .	45
2.8	Laryngectomie supracricoïdienne avec crico-hyoïdo-pexie . . . . .	45
3.1	Cordectomie Laser de type I . . . . .	51
3.2	Cordectomie Laser de type II . . . . .	51
3.3	Cordectomie Laser de type III . . . . .	52
3.4	Cordectomie Laser de type IV . . . . .	52
3.5	Cordectomie Laser de type Va . . . . .	53
3.6	Cordectomie Laser de type Vb . . . . .	53
3.7	Cordectomie Laser de type Vc . . . . .	54
3.8	Cordectomie Laser de type Vd . . . . .	54
3.9	Représentation d'une onde périodique simple . . . . .	72
3.10	Représentation d'un son périodique complexe . . . . .	73
3.11	Exemple d'un spectrogramme à bandes larges du mot «femme» issu de Praat. . . . .	76

3.12 Exemple d'un spectrogramme à bandes étroites du mot «femme» issu de Praat. . . . .	76
3.13 Représentation schématique du jitter . . . . .	79
3.14 Représentation schématique du shimmer . . . . .	80
4.1 Dispositif EVA2 . . . . .	100
4.2 Pièce à main du dispositif EVA2 . . . . .	100
5.1 Courbe de Gauss représentant les régions de rejet de l'hypothèse nulle pour un seuil $\alpha = .05$ . . . . .	108
5.2 Distribution des valeurs de $F_0$ pour les groupes des témoins et des patients pour la voyelle [a] ( $n = 20$ ). . . . .	113
5.3 Distribution des valeurs de $F_0$ pour les groupes des témoins et des patients pour la voyelle [i] ( $n = 20$ ). . . . .	114
5.4 Distribution des valeurs de $F_0$ pour les groupes des témoins et des patients pour la voyelle [u] ( $n = 20$ ). . . . .	114
5.5 Valeurs moyennes de la $F_0$ (en Hz) des voyelles en fonction des groupes . . . . .	116
5.6 Valeurs moyennes du jitter (en %) des voyelles en fonction des groupes . . . . .	117
5.7 Image prise par vidéo nasofibroscopie (25 images/sec) des plis vocaux en situation pré-phonatoire du patient n°2. . . . .	122
5.8 Image prise par vidéo nasofibroscopie (25 images/sec) des plis vocaux en situation respiratoire du patient n°9. . . . .	122
5.9 Spectrogramme à bandes larges de la voyelle [a] synthétisée sous Klatt avec une $F_0$ de 90 Hz . . . . .	128
5.10 Spectrogramme à bandes larges de la voyelle [i] synthétisée sous Klatt avec une $F_0$ de 150 Hz . . . . .	129
5.11 Spectrogramme à bandes larges de la voyelle [u] synthétisée sous Klatt avec une $F_0$ de 230 Hz . . . . .	129
5.12 Spectrogramme à bandes larges représentant l'imitation de la voyelle [a] à 70 Hz du patient n°1 (Praat). . . . .	130
5.13 Valeurs moyennes de la fréquence fondamentale des différentes modalités étudiées pour la voyelle [a] produite par les témoins et les patients . . . . .	136
5.14 Valeurs moyennes de la fréquence fondamentale des différentes modalités étudiées pour la voyelle [i] produite par les témoins et les patients . . . . .	137
5.15 Valeurs moyennes de la fréquence fondamentale des différentes modalités étudiées pour la voyelle [u] produite par les témoins et les patients . . . . .	137
5.16 Exemple d'un phonétogramme . . . . .	143
5.17 Distribution des valeurs de l'intensité pour les groupes des témoins et des patients pour la voyelle [a] ( $n = 20$ ). . . . .	146
5.18 Distribution des valeurs de l'intensité pour les groupes des témoins et des patients pour la voyelle [i] ( $n = 20$ ). . . . .	146
5.19 Distribution des valeurs de l'intensité pour les groupes des témoins et des patients pour la voyelle [u] ( $n = 20$ ). . . . .	147
5.20 Valeurs moyennes de l'intensité (en dB) des voyelles en fonction des groupes . . . . .	148

5.21 Valeurs moyennes du shimmer (en %) des voyelles en fonction des groupes . . . . .	149
5.22 Corrélation entre les valeurs de $F_0$ et de l'intensité intrinsèques des voyelles [a], [i], [u] des témoins et des patients ( $n = 60$ ) . . . . .	155
5.23 Valeurs moyennes de l'intensité (en dB) en fonction des sujets, des voyelles et des corpus . . . . .	163
5.24 Différences intra-sujets $I_{V1} - I_{V2}$ (en dB) pour la voyelle [a] ( $n = 40$ ) . . . . .	164
5.25 Différences intra-sujets $I_{V1} - I_{V2}$ (en dB) pour la voyelle [i] ( $n = 40$ ) . . . . .	164
5.26 Différences intra-sujets $I_{V1} - I_{V2}$ (en dB) pour la voyelle [u] ( $n = 40$ ) . . . . .	165
5.27 Valeurs moyennes de $F_0$ (en Hz) en fonction des sujets, des voyelles et des tâches . . . . .	170
 6.1 Mesures en ms du VOT en français et en anglais . . . . .	189
6.2 Courbes de PIO et de DAB pendant la production d'un [p] . . . . .	194
6.3 Courbes obtenues pour les voyelles tenues [a, i, u] sur Phonetit . . . . .	197
6.4 Valeurs moyennes du DAB (en $dm^3/s$ ) des voyelles en fonction des groupes . . . . .	198
6.5 Indices de fuites glottiques en ( $cm^3/dB/s$ ) des voyelles en fonction des groupes . . . . .	200
6.6 Modélisations des patterns de PIO retrouvés dans nos données pour les logatomes contenant les labiales [p] et [b]. . . . .	205
6.7 Courbe du DAB d'une voyelle [i] . . . . .	207
6.8 Exemples de courbes de DAB pour le logatome [papapap] (témoin n°10) . . . . .	208
6.9 Exemple d'un logatome [pipipip] où les événements de PIO et DAB sont synchronisés pour la position 1 (témoin n°2). . . . .	210
6.10 Exemple d'un logatome [papapap] où les événements de PIO et DAB sont désynchronisés, à savoir que la courbe de DAB commence à augmenter avant la diminution de la PIO pour la position 2 (patient n°10). . . . .	211
6.11 Exemple d'un logatome [pupupup] où les événements de PIO et DAB sont désynchronisés, à savoir que la courbe de PIO commence à descendre avant le début de l'augmentation de DAB pour la position 2 (patient n°2). . . . .	212
6.12 Courbes obtenues pour le logatome [pipipip] sur Phonetit (Exemple du témoin n°1) . . . . .	214
6.13 Valeurs moyennes de la PIO (en hPa) en fonction des sujets, du type de voyelle, du voisement et de la position dans le logatome . . . . .	216
6.14 Valeurs moyennes de Umax — pic du DAB — (en $dm^3/s$ ) en fonction des sujets, du type de voyelle, du voisement et de la position dans le logatome . . . . .	217
6.15 Valeurs moyennes du DPM — relevées sur une fenêtre temporelle de 50 ms, représentant la portion stable de la voyelle — (en $dm^3/s$ ) en fonction des sujets, du type de voyelle, du voisement et de la position dans le logatome . . . . .	218
6.16 Coupe sagittale montrant (i) la propagation du flux d'air (U) et (ii) la localisation des différentes résistances (R) dans le conduit vocal et profils sagittaux des voyelles [a, i, u] . . . . .	222
6.17 Exemple de forme du signal pour une segmentation fine . . . . .	231
6.18 Segmentation du logatome [tututut] sous Praat . . . . .	232

6.19 Valeurs moyennes des durées consonantiques (ms) en position intervocalique (V_V) en fonction des sujets, du type de voyelle et du contexte de voisement . . . . .	235
6.20 Durées moyennes de la tenue et du relâchement (ms) des consonnes en position intervocalique (V_V) des témoins en fonction du contexte de voisement et du type de voyelle . . . . .	236
6.21 Durées moyennes de la tenue et du relâchement (ms) des consonnes en position intervocalique (V_V) des patients en fonction du contexte de voisement et du type de voyelle . . . . .	237
6.22 Valeurs moyennes de l'intensité relative du relâchement (en dB) des consonnes en fonction des sujets, du type de voyelle, du contexte de voisement et de la position dans le logatomes . . . . .	238
6.23 Valeurs moyennes des durées vocaliques (ms) en fonction des sujets, du type de voyelle, du contexte de voisement et de la position dans le logatomes	239
6.24 Valeurs moyennes de $F_0$ (en Hz) au début des voyelles en fonction des sujets, du type de voyelle, du contexte de voisement et de la position dans le logatome . . . . .	240
6.25 Valeurs moyennes du VOT des occlusives (ms) en fonction des sujets, du type de voyelle, du lieu d'articulation et de la position dans le logatomes .	241
6.26 Exemple d'une courbe d'intensité (dB) pour le logatome [tititit] du témoin n°2 (Praat). . . . .	247
6.27 Exemple d'un signal EGG et de sa dérivée . . . . .	252
6.28 Signal DEGG présentant des pics d'ouverture et de fermeture, ce qui permet la mesure du fondamental (T0) et du quotient ouvert (Oq) . . . . .	253
6.29 Méthode de mesure du Oq sur le signal EGG par détection des seuils (35, 50%) . . . . .	254
6.30 Exemples typiques de 6 périodes sur des signaux EGG et DEGG pour la voyelle [u] . . . . .	255
6.31 Signaux EGG et DEGG montrant un double pic d'ouverture ou un double pic de fermeture . . . . .	256
6.32 Signaux EGG et DEGG montrant un doublement de période . . . . .	257
6.33 Signaux EGG et DEGG montrant un triplement de période . . . . .	257
6.34 Signaux EGG et DEGG acquis sur 5 cycles pour la voyelle [a] (à gauche) et sur 6 cycles pour la voyelle [i] (à droite) (patient n°4). . . . .	259
6.35 Valeurs moyennes des quotients ouverts selon le type de voyelle pour huit de nos patients . . . . .	261
6.36 Valeurs moyennes des quotients ouverts des voyelles en contexte occlusif pour huit de nos patients en fonction des voyelles, de la position dans le logatome et du voisement . . . . .	265
6.37 Exemple d'un test de perception . . . . .	269

# Liste des tableaux

1.1	Actions des muscles intrinsèques du larynx . . . . .	17
1.2	Actions des muscles extrinsèques sus-hyoïdiens du larynx . . . . .	17
1.3	Actions des muscles extrinsèques sous-hyoïdiens du larynx . . . . .	18
2.1	Classification des tumeurs (T) du plan glottique. . . . .	36
2.2	Classification des atteintes ganglionnaires (N). . . . .	36
2.3	Classification des métastases à distance (M). . . . .	36
3.1	Tableau synthétique des différentes études sur la voix après cordectomie .	65
3.2	L'échelle GRBAS . . . . .	68
3.3	Synthèse des différentes corrélations acoustico-perceptives. . . . .	80
4.2	Tableau récapitulatif des patients . . . . .	93
4.3	Tableau récapitulatif des témoins . . . . .	94
5.1	Extrait de la table de Fisher pour un seuil $\alpha = .05$ . . . . .	110
5.2	Extrait de la table de la Loi de Student (test $t$ ) . . . . .	111
5.3	Test de normalité de Lilliefors appliqué sur la $F_0$ moyenne des voyelles [a], [i] et [u] ( $n = 60$ ). . . . .	113
5.4	Résultats du test F de Fisher-Snedecor sur les valeurs de $F_0$ pour les trois voyelles [a], [i] et [u] ( $n = 60$ ). . . . .	115
5.5	Résultats des tests-t non appariés conduits sur les valeurs de $F_0$ (en Hertz) des voyelles [a], [i] et [u] . . . . .	116
5.6	Résultats des tests-t non appariés conduits sur les valeurs du jitter (en pourcentages) des voyelles [a], [i] et [u] . . . . .	118
5.7	Le synthétiseur de Klatt . . . . .	127
5.8	Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en hertz) de la voyelle [a] selon les témoins et les patients . . . . .	132
5.9	Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en demi-tons) de la voyelle [a] selon les témoins et les patients . . . . .	133
5.10	Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en hertz) de la voyelle [i] selon les témoins et les patients . . . . .	134
5.11	Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en demi-tons) de la voyelle [i] selon les témoins et les patients . . . . .	134

5.12 Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en hertz) de la voyelle [u] selon les témoins et les patients . . . . .	135
5.13 Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en demi-tons) de la voyelle [u] selon les témoins et les patients . . . . .	135
5.14 Résultats des différents tests de Student (tests-t non appariés) conduits sur les différentes modalités de la voyelle [a] . . . . .	138
5.15 Résultats des différents tests de Student (tests-t non appariés) conduits sur les différentes modalités de la voyelle [i] . . . . .	139
5.16 Résultats des différents tests de Student (tests-t non appariés) conduits sur les différentes modalités de la voyelle [u] . . . . .	139
5.17 Tables synthétiques des écarts significatifs (en gris) entre les cibles attendues et les imitations des voyelles [a], [i] et [u] selon les groupes des témoins et des patients (d'après les données exprimées en Hertz). . . . .	142
5.18 Test de normalité de Lilliefors appliqué sur les intensités moyennes des voyelles [a], [i], et [u] ( $n = 60$ ). . . . .	145
5.19 Résultats synthétiques du test F de Fisher-Snedecor sur les valeurs de l'intensité pour les trois voyelles [a], [i] et [u] ( $n = 60$ ). . . . .	147
5.20 Résultats des tests-t non appariés conduits sur les valeurs de l'intensité (en dB) des voyelles [a], [i] et [u] . . . . .	149
5.21 Résultats des tests-t non appariés conduits sur les valeurs du shimmer (en pourcentages) des voyelles [a], [i] et [u] . . . . .	150
5.22 Calculs de l'atténuation/gain (dB) en fonction de la distance du micro à la bouche. . . . .	159
5.23 Résultats des tests-t non appariés conduits sur les valeurs de l'intensité (en dB) des voyelles [a], [i] et [u] en contexte . . . . .	161
5.24 Tableaux récapitulatifs des différences $I_{V1} - I_{V2}$ (en dB) . . . . .	163
5.25 Résultats synthétiques des tests-t non appariés conduits sur les valeurs de la $F_0$ (en Hz) des voyelles [a], [i] et [u] en contexte . . . . .	169
 6.1 Système des occlusives orales du français standard . . . . .	192
6.2 Test de normalité de Lilliefors appliqué sur les valeurs moyennes du DAB des voyelles [a], [i], et [u] ( $n = 60$ ). . . . .	195
6.3 Résultats synthétiques du test F sur les valeurs du DAB pour les trois voyelles [a], [i], [u] ( $n = 60$ ). . . . .	196
6.4 Résultats des tests-t non appariés conduits sur les valeurs de DAB (en $dm^3/s$ ) des voyelles [a], [i] et [u]. . . . .	198
6.5 Résultats des tests-t non appariés conduits sur les valeurs d'intensité (en dB) des voyelles [a], [i] et [u] . . . . .	199
6.6 Résultats des tests-t non appariés conduits sur les valeurs des indices de fuites glottiques (en $cm^3/dB/s$ ) des voyelles [a], [i] et [u] . . . . .	200
6.7 Coefficients de variation (exprimés en pourcentages) des mesures de DAB pour nos deux populations . . . . .	202

6.8 Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction des patterns de PIO observés pour la position d'attaque en initiale absolue (#_) . . . . .	206
6.9 Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction des patterns de PIO observés pour la position d'attaque en intervocalique (V_V) . . . . .	206
6.10 Répartitions (en %) des différentes courbes de DAB des logatomes conte-nant les labiales [p] et [b] . . . . .	209
6.11 Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction de la synchronisation des évènements aérodynamiques pour la position d'attaque en initiale absolue (#_) . . . . .	212
6.12 Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction de la synchronisation des évènements aérodynamiques pour la position d'attaque en intervocalique (V_V) . . . . .	213
6.13 Tableau récapitulatif des différents tests ANOVA pour les paramètres aé-rodynamiques étudiés . . . . .	218
6.14 Valeurs moyennes de la PIO (en hPa) et calculs de la pression transglot-tique estimée (en hPa) selon les deux positions syllabiques (attaque en ini-tiale absolue et attaque en intervocalique) pour les témoins et les patients en contexte vocalique [a] . . . . .	219
6.15 Valeurs moyennes de la PIO (en hPa) et calculs de la pression transglot-tique estimée (en hPa) selon les deux positions syllabiques (attaque en ini-tiale absolue et attaque en intervocalique) pour les témoins et les patients en contexte vocalique [i] . . . . .	219
6.16 Valeurs moyennes de la PIO (en hPa) et calculs de la pression transglot-tique estimée (en hPa) selon les deux positions syllabiques (attaque en ini-tiale absolue et attaque en intervocalique) pour les témoins et les patients en contexte vocalique [u] . . . . .	219
6.17 Résultats des principales études concernant les estimations de la pression transglottique chez des sujets sains . . . . .	227
6.18 Tableau récapitulatif des différents tests ANOVA pour les paramètres acous-tiques étudiés (où « ** » note des différences hautement significatives ( $p < .0001$ ), « * » note des différences significatives ( $p < .05$ ), « ns » signifie « non significatif » et « na » signifie « non applicable »). . . . .	242
6.19 Quotients moyens ouverts des huit patients et de notre témoin de référence pour les voyelles [a], [i], [u] et résultats des tests-t univariés . . . . .	260
6.20 Matrice de confusion globale pour la position initiale absolue . . . . .	270
6.21 Matrice de confusion pour le contexte [a], soit [#Ca] ( $n = 1320$ ). . . . .	270
6.22 Matrice de confusion pour le contexte [i], soit [#Ci] ( $n = 1320$ ). . . . .	271
6.23 Matrice de confusion pour la voyelle [u], soit [#Cu] ( $n = 1320$ ). . . . .	271
6.24 Matrice de confusion globale pour la position intervocalique . . . . .	271
6.25 Matrice de confusion pour le contexte [a], soit [aCa] ( $n = 1320$ ). . . . .	272
6.26 Matrice de confusion pour le contexte [i], soit [iCi] ( $n = 1320$ ). . . . .	272
6.27 Matrice de confusion pour le contexte [u], soit [uCu] ( $n = 1320$ ). . . . .	273



# Liste des symboles

AR	Air résiduel
C	Consonne
CAL	Muscle crico-aryténoïdien latéral
CAP	Muscle crico-aryténoïdien postérieur
Cq	Quotient de fermeture
CRF	Capacité résiduelle fonctionnelle
CT	Muscle crico-thyroïdien
CV	Capacité vitale
D	Durée
DAB	Débit d'air buccal
DPM	Débit phonatoire moyen
EGG	Electroglottographe
F0	Fréquence fondamentale
F1	Premier formant
F2	Second formant
F3	Troisième formant
F4	Quatrième formant
Fg	Fuite glottique
H1	Premier harmonique
H2	Second harmonique
HNR	Harmonic noise ratio
I	Intensité
IA	Muscle inter-aryténoïdien
IO	Muscle aryténoïdien oblique
IT	Muscle aryténoïdien transverse
LPH	Laryngectomie partielle horizontale
LPSC	Laryngectomie partielle supracricoïdienne

- LPV Laryngectomie partielle verticale  
NNE Normalized noise energy  
Oq Quotient d'ouverture  
PA pression atmosphérique  
PIO Pression intraorale  
PO Postopératoire  
PSG Pression sous-glottique  
PSGE Pression sous-glottique estimée  
PSP Pression de seuil phonatoire  
TA Muscle thyro-aryténoïdien  
TMP Temps maximum de phonation  
Umax Pic maximal de la courbe du débit d'air buccal  
V Voyelle  
VC Volume courant  
VHI Voice handicap index  
VOT Voice onset time  
VRE Volume de réserve expiratoire  
VRI Volume de réserve respiratoire

# Remerciements

"Ce n'est pas tant l'aide de nos amis qui nous aide que

notre confiance dans cette aide."

– Epicure

Je tiens à exprimer toutes ma reconnaissance aux personnes qui ont contribué directement ou indirectement à l'aboutissement de cette thèse.

En premier lieu, mes remerciements vont à ma directrice de recherches, Jacqueline Vaissière, qui durant ces cinq dernières années a su me guider et me conseiller. Son énergie débordante, sa capacité d'écoute et sa confiance ont été des moteurs pour l'achèvement de ce travail.

A Lise Crevier Buchman qui a accepté de co-diriger cette thèse. Je lui suis tout particulièrement reconnaissante de m'avoir ouvert les portes de l'hôpital Européen Georges Pompidou et de m'avoir entourée si patiemment dans ma démarche clinique. Nos échanges ainsi que ses pertinents conseils ont contribué au succès de ce travail.

J'adresse ma gratitude à l'ensemble des membres de mon jury de soutenance pour l'honneur qu'ils m'ont accordé en acceptant de juger et de valider ce travail. Je remercie Antoine Giovanni qui, d'un œil expert tant du point de vue clinique que du point de vue phonétique, a accepté d'évaluer ce document et d'en être le premier rapporteur. Je remercie également Rudolph Sock pour sa gentillesse et pour l'attention montrée à mon égard en acceptant d'être le second rapporteur de ce travail. Un grand merci également à Stéphane Hans pour ses conseils et relectures expertes dans le domaine de la pathologie. Enfin, je remercie Didier Demolin pour son engagement dans ce travail.

J'exprime ma profonde gratitude à Jean-Marc Beltzung qui n'a eu de cesse de me motiver durant ces pénibles mais enrichissantes années de thèse. Sans sa patience, ses conseils, ses relectures et ses encouragements, ce travail ne serait peut-être pas ce qu'il est aujourd'hui. Ses compétences en phonologie et en phonétique — et depuis peu en scripts informatiques — font de lui un linguiste accompli et irremplaçable.

J'exprime ma reconnaissance à tous les membres du Laboratoire de Phonétique et Phonologie de Paris III qui m'ont aidé à mener à bien cette thèse. Un grand merci à Shinji Maeda qui, à travers nos discussions, a su éclairer les points les plus obscurs. Merci aux ingénieurs du laboratoire Angélique Amelot et Nicolas Audibert pour leurs performances informatiques. Merci à Rachid Ridouane, Cécile Fougeron, Claire Pillot, Pierre Hallé, Cédric Gendrot, Annie Rialland, Bernard Gautheron, Antonia Collazo pour leurs échanges, leur disponibilité et leurs conseils. Merci également aux doctorants avec lesquels j'ai pu discuter.

Je remercie également tous les sujets qui ont accepté de participer à cette étude, patients et témoins. Les enregistrements parfois complexes, les reconversations ont été abordées avec gentillesse. Sans eux, ce travail n'existerait évidemment pas.

Enfin, je remercie chaleureusement tous les membres de ma famille ainsi que mes proches amis. Tous m'ont toujours encouragée et aidée dans cette délicate entreprise. Je pense particulièrement à mes parents qui ont toujours été réconfortants et aimants, à mon frère et à mes grands-parents qui m'ont toujours soutenue. Vous avez toute ma gratitude et ma tendresse.

Je dédie ce travail à mon pazuzu adoré, avec tout mon amour...

# Introduction

"Ecoute, sinon ta langue te rendra sourd."  
– Siting Bull

## Langage, langue et parole

Tous les êtres vivants sont capables de communiquer avec les individus de leur espèce, y compris les animaux. La communication animale — parfois très complexe — soulève bien des interrogations auprès des scientifiques quant à leur proximité avec le langage humain. Von Frisch (1965), par exemple, a montré que les abeilles pouvaient indiquer à leurs congénères la qualité mais aussi la direction et la distance d'une source de nourriture par rapport à la ruche, par l'exécution de diverses danses sophistiquées. D'autres études, comme le chant des baleines ou les cris des primates, mettent en évidence le besoin de ces êtres vivants à communiquer. Dans son ouvrage, Benveniste (1966), a cherché à mettre en évidence les différences et les ressemblances entre le langage humain et la communication animale à travers une critique des recherches de Von Frisch (1967). Celui-ci note que la communication des abeilles est basée uniquement sur un ensemble de mouvements et ne fait, en aucun cas, intervenir un appareil vocal. Il ajoute que le message transmis par les ouvrières n'appelle aucun dialogue ou interaction, ce qui est le propre du langage humain. Aussi, les messages des abeilles restent motivés puisqu'il existe une correspondance entre la qualité des danses et les renseignements qu'elles donnent, ce qui est totalement contraire au langage humain. De plus, la fonction métalinguistique du langage qui implique que les individus puissent intervenir sur leur propre langage est ignorée dans ce système de communication. Enfin, ce type de communication ne peut être analysé du point de vue du domaine de la combinatoire. Finalement, ces observations permettent de conclure que les abeilles — et plus largement les animaux — ont une capacité à communiquer, en particulier pour assurer leur survie, mais que cette faculté de « langage » est réellement le propre de l'homme.

Dès le XVII<sup>e</sup> siècle, les philosophes assimulent l'idée que la fonction langagière est principalement liée à l'expression de la pensée. Selon Descartes (1637 [1840]), seul le langage parlé ou gestuel permet d'émettre des idées et de les communiquer au sein d'une société. Il y a bien là une relation de cause à effet entre langage et société. Plus particulièrement, le langage parlé permet d'établir des liens sociaux. De plus, la société permet une transmission spontanée et efficace du langage d'une génération à une autre, ce qui renforce ce lien culturel.

La langue est le produit social du langage, lequel est régi par un ensemble de codes linguistiques. De ce point de vue, Saussure (1916) définit la langue comme un ensemble de signes arbitraires qui associe, par convention, une idée ou un concept à son

image acoustique. En outre, comme nous l'avons énoncé précédemment, la productivité est un élément caractéristique du langage humain qui le différencie du langage animal. Martinet (1968) ajoute que les langues sont caractérisées par une double articulation : un nombre limité de phonèmes<sup>1</sup> peut former un grand nombre de morphèmes<sup>2</sup> qui peuvent également se combiner en un nombre infini de mots et de phrases. Cette théorie met en avant la complexité des systèmes et la grande variété des énoncés linguistiques.

Pour parler, c'est-à-dire pour actualiser sa langue, l'individu doit pouvoir contrôler et mettre en relation les principaux articulateurs liés à la fonction phonatoire. La phona-tion nécessite une source d'énergie aérienne provenant des poumons, un transducteur d'énergie, les plis vocaux — qui transforme le flux d'air en impulsions vibratoires — et des cavités supraglottiques qui servent de résonateurs et modifient le signal glottal en sons de la parole. Le larynx joue un rôle majeur dans la communication humaine.

Toutes les langues utilisent le larynx dans diverses configurations pour distinguer les consonnes et les voyelles. D'un point de vue phonologique, le larynx peut être identifié par sa fonction linguistique (en particulier pour l'identification du trait de voisement des consonnes, des tons des voyelles, de l'accentuation et de l'intonation des langues), sa fonction paralinguistique (qui dénote des attitudes contrôlées des locuteurs) et sa fonction extralinguistique (qui traduit l'état, l'identité, les émotions ou la pathologie des locuteurs, cf. Vaissière 1997).

## Les voix pathologiques

Il peut arriver que les fonctions linguistiques soient altérées suite à une maladie ou à un handicap moteur par exemple. Ainsi, les études sur les voix pathologiques sont essentielles dans la mesure où elles permettent une meilleure compréhension de ces modèles particuliers de voix. Elles impliquent la validation de protocoles d'études complets et adaptés à chaque type de pathologie. Ces derniers doivent pouvoir rendre compte des caractéristiques articulatoires, acoustiques et perceptives de ces voix afin de caractériser au mieux les mécanismes de production mis en jeu. Ainsi, les productions pathologiques offrent un potentiel d'étude extrêmement pertinent pour comprendre l'étendue de la variabilité dans la production et des stratégies compensatoires qui peuvent être mises en place pour la production d'un message et la préservation de la communication. Finalement, un modèle complet doit pouvoir rendre compte aussi bien de la parole chez le sujet sain que chez le sujet pathologique. D'autre part, ces recherches en phonétique clinique permettent à la fois aux professionnels une meilleure adaptation de leurs méthodes rééducatives mais aussi une meilleure connaissance des patients quant à leur nouvelle voix et leurs comportements de parole qui contribuent à une réinsertion sociale et professionnelle plus satisfaisante.

En laryngologie, la plus traumatisante des pathologies est le cancer du larynx. Celui-ci peut être limité à l'étage glottique ou s'étendre aux structures adjacentes. Face

1. Le phonème se définit comme la plus petite unité distinctive
2. Le morphème représente la plus petite unité distinctive de sens.

à cette maladie, plusieurs traitements s'offrent aux patients : la radiothérapie, la chimiothérapie, la chirurgie ou une combinaison des trois. La décision thérapeutique reste essentiellement liée au principe de préservation d'organes qui implique une ablation tumorale complète tout en conservant les fonctions principales du larynx, à savoir la respiration, la déglutition et la phonation (Laccourreye 1993). A partir des années 70, les cordectomies laser par voie endoscopique furent introduites pour des tumeurs limitées au pli vocal (Strong 1975). Ces dernières sont classées en fonction de l'importance de l'exérèse. Ainsi, les différentes études sur le sujet montrent des résultats fonctionnels satisfaisants mais des résultats phonatoires aléatoires et dépendants de l'étendue de la chirurgie. De notre point de vue, nous nous sommes intéressés aux cordectomies de type II et III qui sont proposées principalement pour des lésions de type Tis (*in-situ*) ou T1 selon la classification TNM. La cordectomie de type II (ou sous-ligamentaire) consiste en l'exérèse de l'*épithélium*, du ligament vocal et de la *lamina propria* dans sa totalité. La cordectomie de type III (ou transmusculaire), quant à elle, résègue également une partie variable du muscle vocal (Remacle & al. 2000). Ces chirurgies sont de bonnes alternatives, quand cela reste possible, aux chirurgies plus invasives ou totales dans la mesure où elles assurent de meilleures suites postopératoires et de moindres coûts médico-économiques.

## Originalité de notre travail

Notre recherche se justifie par l'originalité du sujet. À notre connaissance, très peu d'études se sont intéressées aux résultats fonctionnels vocaux après cordectomie laser (Zeitels & al. 2002, Peretti & al. 2003, Mirghani 2009, Claeys 2010, entre autres). La plupart des recherches sont strictement médicales et étudient les résultats carcinologiques afin de valider la chirurgie. Les études qui traitent des résultats fonctionnels sur la voix restent assez réductrices d'un point de vue méthodologique avec le recours, pour la plupart du temps, aux simples corpus vocaliques. De plus, la plupart de ces travaux mêlent des populations hétérogènes, ce qui rend l'interprétation des résultats délicate. Cette thèse est innovante dans la mesure où il s'agit de la première étude prospective qui décrit et analyse les caractéristiques phonétiques acoustiques et aérodynamiques, en particulier, de la voix et de la parole chez 10 sujets ayant subi une cordectomie de type II-III. L'intérêt de cette recherche réside dans l'homogénéité de notre population, dans la reflexion apportée sur le trait de voisement et dans l'apport d'une nouvelle méthodologie pour l'étude de la fréquence fondamentale. Ainsi, les plages de variation de  $F_0$  ont été appréhendées d'un point de vue didactique à l'aide de tâches d'imitations de voyelles synthétiques. De plus, l'implémentation du trait de voisement chez nos patients a été analysée de façon multiparamétrique, ce qui confère une certaine solidité à notre étude.

## Objectifs de notre recherche

Comme nous venons de le mentionner, la plupart des études sur la voix après cordectomie portent uniquement sur les segments vocaliques. Selon nous, ce traitement est réducteur dans le sens où il ne met pas en avant les aspects multimodales que recouvrent la voix et la parole. Nous avons dû nous interroger sur certains problèmes méthodologiques pour caractériser au mieux ces modèles de voix particuliers. En outre, les changements anatomico-physiologiques engendrés par la chirurgie soulèvent bien des questions quant aux modifications du vibrateur laryngé, et plus largement sur le maintien de la fonction linguistique du larynx et de l'acte de communication. Ainsi, les questions auxquelles nous tenterons de répondre dans cette thèse peuvent se résumer aux questions suivants :

- Dans quelles mesures les paramètres de la voix seront-ils perturbés après cordectomie de type II-III ?
- La chirurgie a t'elle un impact sur le maintien de l'opposition du trait de voisement en français ? Quels seront les paramètres les plus altérés ? Existe t'il des phénomènes compensatoires ?

## Organisation de la thèse

La présente thèse se divise en six chapitres. Le premier chapitre est consacré à la description des organes et des mécanismes mis en jeu dans la production de la parole. En effet, il nous a semblé important de rappeler quelques notions physiologiques ainsi que certaines théories fondamentales qui expliquent le fonctionnement du système phonatoire, et en particulier le mode vibratoire des plis vocaux. Dans un second chapitre, nous évoquerons les différentes chirurgies mises à disposition dans le traitement des cancers du larynx. Nous tenterons d'aborder les différents traitements spécifiques aux cancers de l'étage glottique d'un point de vue historique afin de déterminer les conséquences d'apparition et de pratique des cordectomies laser. Notre troisième chapitre est entièrement consacré à la notion même de cordectomie laser par voie endoscopique. Cette technique, indiquée principalement pour les états pré-cancéreux, les carcinomes *in-situ*, les tumeurs T1 et quelques T2 sélectionnés, permet une récupération fonctionnelle plus rapide et de moindres coûts financiers, ce qui représente des arguments de choix dans son développement depuis les années 70. Après avoir détaillé les principales cordectomies laser correspondant à l'étendue de la lésion aux différentes structures laryngées, nous pouvons alors mettre en évidence les différences, les similitudes et les problèmes posés par les quelques études sur les résultats fonctionnels sur la voix à travers une revue de la littérature sur le sujet. Toutes ces recherches montrent, néanmoins, qu'une analyse multiparamétrique est nécessaire pour qualifier au mieux ces modèles de voix. Par conséquent, nous proposerons, dans la suite de ce chapitre, un récapitulatif des principales mesures quantitatives et qualitatives utilisées en phonétique clinique. Le quatrième chapitre fait le lien entre la partie théorique et

la partie expérimentale puisqu'il décrit précisément nos populations, nos corpus et les matériels dont nous avons eu besoin pour nos différentes études. Nous mettrons également en évidence l'importance de comparer une population de sujets dysphoniques avec des sujets sains afin d'établir un référentiel. De plus, comme nous le précisions précédemment, nous soutiendrons l'utilité d'un corpus varié ainsi que des mesures pluriparamétriques afin de caractériser au mieux ces voix. Le cinquième chapitre comprend trois études, augmentées d'un excursus, qui tendent à montrer l'impact de la chirurgie sur les paramètres de la fréquence fondamentale, de l'intensité sonore et de leurs indices de perturbation : le jitter et le shimmer. Les plages de variation de  $F_0$  des sujets seront analysées. Plus particulièrement, ce chapitre consiste à montrer l'impact de la chirurgie sur les résultats vocaux et, d'autre part, à mettre en évidence l'importance de coupler des corpus vocaliques avec des corpus de parole — dans notre cas, de la parole de laboratoire car notre corpus est composé de logatomes — afin de valider nos hypothèses. A ce titre, il nous a paru utile d'introduire une nouvelle méthodologie pour l'étude des plages de variation de  $F_0$  des patients. Enfin, le dernier chapitre est entièrement consacré à l'implémentation du trait de voisement ou, plus précisément, au maintien du contraste de voisement des occlusives après ces types de chirurgies. Ainsi, des études acoustiques, aérodynamiques, électroglottographiques et perceptives nous permettront de mettre en évidence, s'il y a lieu, les difficultés engendrées par les changements anatomico-physiologiques liés aux interventions.



# Anatomie et physiologie du système phonatoire

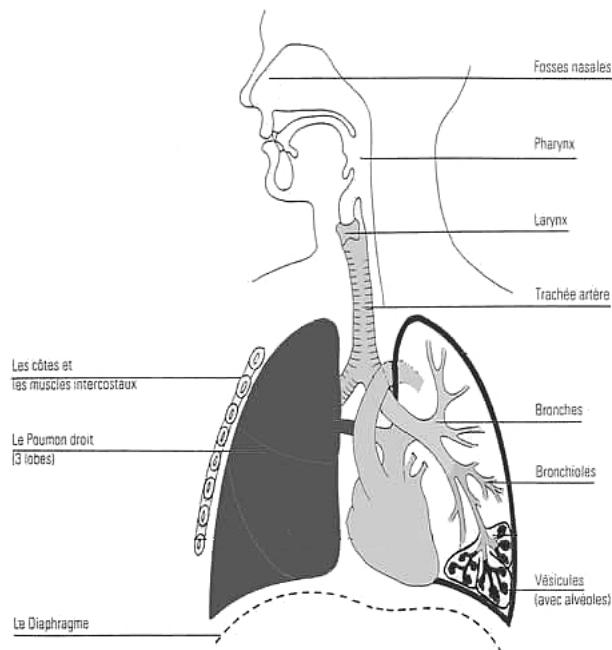
"Beware of the man who does not talk, and the dog that does not bark."  
– Proverbe Cheyenne

## Résumé

Ce chapitre est un rappel anatomique et biomécanique de la phonation. L'individu doit pouvoir contrôler les principaux articulateurs liés à la fonction phonatoire. La phonation nécessite une source d'énergie aérienne provenant des poumons, un transducteur d'énergie, fonction assurée par les plis vocaux — qui transforme le flux d'air en impulsions vibratoires — et des cavités supraglottiques qui servent de résonateurs et modifient le signal glottal en sons de la parole. La connaissance anatomique de ces différents organes et, en particulier, la modélisation de la structure des plis vocaux, ont permis le développement de modèles mécaniques et de théories expliquant la physiologie de la phonation. La théorie actuelle, qui fait appel à un modèle à trois masses (cf. théorie body cover, Hirano 1981), combine un facteur myoélastique qui associe les caractéristiques des plis vocaux à la qualité de l'ondulation, un facteur aérodynamique qui impose un flux d'air transglottique pour initier et maintenir les vibrations et un concept oscillo-impédantiel qui implique que les deux plis vocaux sont des oscillateurs qui transforment la pression de l'air expirée en une pression pulsée dont la fréquence est similaire à celle de l'oscillateur. Dès lors, la modification de la structure du pli vocal, après une chirurgie par exemple, aurait des conséquences plus ou moins importantes sur le mécanisme vibratoire et donc, *in extenso*, sur la voix et ses paramètres acoustiques.

## 1.1 Introduction

La production de la parole nécessite l'intervention de trois ensembles d'organes — que nous nommerons ici « niveaux » — qui tiennent chacun un rôle caractéristique dans ce processus : le système *respiratoire*, composé essentiellement des poumons, le système *phonatoire*, composé du larynx et le système *articulatoire*, composé de l'ensemble des articulateurs supra-laryngés (cf. figure 1.1).



**Figure 1.1** – Les organes participant à la production de la parole (d'après [www.paramed-prepa.com](http://www.paramed-prepa.com)).

La connaissance des différents organes de la parole, tant d'un point de vue fonctionnel qu'anatomique, est un élément essentiel dans la compréhension du mécanisme dont il est question ici, c'est-à-dire la production de la voix et de la parole. Le but de ce chapitre est d'aborder le fonctionnement des trois niveaux évoqués plus haut en essayant d'y apporter, lorsque cela sera nécessaire, une description anatomique et physiologique détaillée des différents organes qui les composent. Dans la première section, nous évoquerons le fonctionnement du système respiratoire. La deuxième section sera consacrée au système phonatoire (anatomie et théories de la phonation principalement) et la troisième section sera dédiée à la description des différents articulateurs de la parole. Dans cette section, nous aborderons notamment la théorie source-filtre (Fant 1960), considérée comme une théorie fondatrice et désormais classique, dans la production des sons de la parole.

## 1.2 Le système respiratoire

D'un point de vue physiologique, le système respiratoire a pour fonction primaire de permettre les échanges gazeux. Chez les êtres vivants, celui-ci fournit l'oxygène et rejette le gaz carbonique. Par ailleurs, le système respiratoire intervient dans le mécanisme phonatoire. Ce dernier, comparé à une « *soufflerie* », reste primordial dans la production de la parole car il fournit l'air nécessaire à la vibration des plis vocaux. La voix représente en effet une expiration sonorisée (Draper & al. 1959, Draper & al. 1960, Cornut 1959, Cornut 1983, Hixon 1973, Hixon 1976, Le Huche & Allali 1991, Ormezzano 2000, Mcfarland 2006).

Le thorax, qui représente la partie supérieure du tronc et qui est principalement constitué d'une armature osseuse, la cage thoracique, contient les organes principaux à cette phase respiratoire : les poumons. De nombreux muscles du thorax et de l'abdomen interviennent dans le processus respiratoire.

### 1.2.1 Les mécanismes respiratoires

Le cycle respiratoire est automatique et vital. Ces cycles restent sensibles à l'activité du corps humain. Ainsi, nous sommes capables d'adapter nos efforts respiratoires aux diverses situations. Par exemple, l'effort physique requiert un apport d'oxygène plus important, ce qui entraîne une accélération du rythme respiratoire. L'acte de respiration s'effectue en deux mouvements : une inspiration pour laquelle la pression à l'intérieur des poumons doit être plus faible que la pression atmosphérique et une expiration pour laquelle la pression dans les poumons doit être supérieure à la pression extérieure.

Le mouvement d'*inspiration* se caractérise principalement par une augmentation majeure de la cage thoracique associée à une dilatation des poumons qui se remplissent d'air. L'air pénètre par le nez et/ou la bouche, descend dans le larynx, rejoint la trachée puis s'engouffre dans les deux bronches qui se subdivisent en bronchioles contenant les alvéoles pulmonaires.

Lors du mouvement d'*expiration*, le parcours est inversé. L'air quitte les poumons par les bronches et la trachée. Il rencontre alors un obstacle, les plis vocaux qui vont entrer en vibration sous l'effet de la pression, dans le cadre d'une phonation. Lors d'une phase expiratoire simple, la glotte reste ouverte et l'air peut continuer de s'échapper à travers le conduit vocal.

De manière générale, les volumes d'air mobilisés pour la ventilation quotidienne représentent, chez un adulte, 20000 litres d'air inspiré et expiré. Aussi, la respiration se décompose en plusieurs phases qui mobilisent des volumes d'air variables selon les besoins vitaux, l'état physiologique ou la situation dans laquelle évolue l'individu. Les différents volumes d'air utilisés se décomposent de la manière suivante :

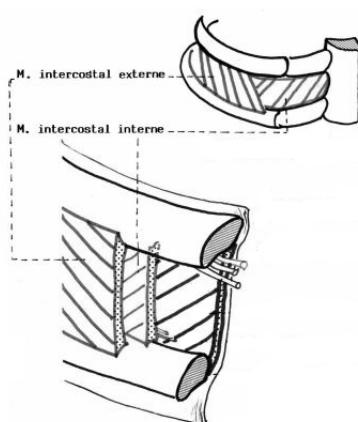
- ▷ La respiration de repos représente 0,4 à 0,7 litres d'air, c'est le *volume courant* (VC)

- ▷ L'inspiration forcée qui, mobilise 1,5 à 2,5 litres d'air, s'effectue lors d'une prolongation d'un mouvement inspiratoire. Les poumons se remplissent totalement, c'est le *volume de réserve inspiratoire* (VRI)
- ▷ L'expiration forcée se traduit par une prolongation d'une expiration normale. En cas de nécessité, nous pouvons vider complètement nos poumons lors d'une expiration complète de 1 à 2 litres environ, c'est le *volume de réserve expiratoire* (VRE).
- ▷ A la fin d'une expiration forcée, nos poumons ne se vident jamais complètement. Ils contiennent toujours une quantité d'air qui ne peut pas être expulsé, soit 1,2 litre environ, c'est l'*air résiduel* (AR)
- ▷ On appelle *capacité résiduelle fonctionnelle* (CRF) la somme du VRE + AR
- ▷ Finalement, la *capacité vitale* (CV) est la somme de l'ensemble des volumes d'air intervenant dans la respiration, soit :  $CV = VC + VRI + VRE$  (Draper & al. 1959, Draper & al. 1960, Cornut 1983, Léothaud 1994, Hans 2003).

## 1.2.2 Les muscles respiratoires

### 1.2.2.1 Muscles de l'inspiration

Le muscle qui représente la principale force inspiratoire est le *diaphragme*. Placé entre le thorax et l'abdomen, le diaphragme est un muscle en forme de dôme. Lors d'une contraction, le diaphragme s'abaisse et va appuyer sur les viscères, ce qui va permettre l'augmentation de la cage thoracique. La dilatation des côtes est également provoquée par l'association des contractions des *muscles intercostaux externes* (cf. figure 1.2).



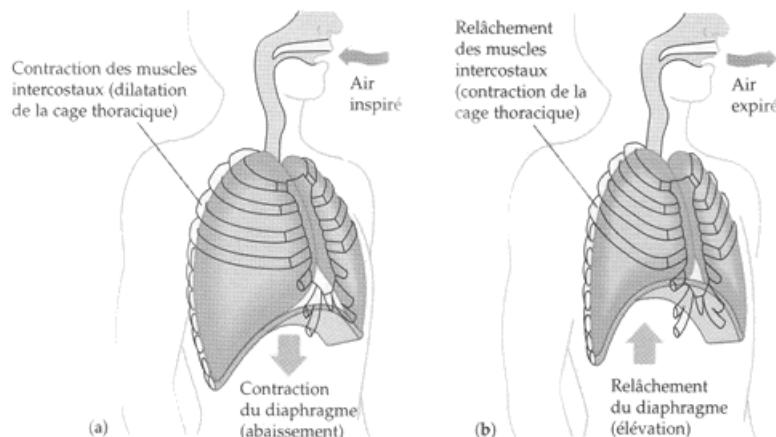
**Figure 1.2** – Vues antérieure et latérale des muscles intercostaux internes et externes (où M= muscles, d'après [www.univ-brest.fr/S\\_Commun/Biblio/ANATOMIE/Web\\_anat/Thorax/Paroi/Orifice\\_Super.htm](http://www.univ-brest.fr/S_Commun/Biblio/ANATOMIE/Web_anat/Thorax/Paroi/Orifice_Super.htm)).

D'un point de vue anatomique, les muscles intercostaux externes sont tendus entre le bord inférieur des côtes supérieures et la face supérieure des côtes inférieures. Ceux-ci occupent tout l'espace intercostal depuis le rachis jusqu'au sternum. Ces muscles

sont fibreux entre les cartilages costaux. Ces fibres sont obliques vers le bas, l'avant et l'intérieur. Ils sont doublés, à leur face profonde, par la membrane intercostale externe.

### 1.2.2.2 Muscles de l'expiration

Outre les *muscles intercostaux internes*<sup>1</sup> (cf. figure 1.2), qui participent à la fermeture costale, les *muscles abdominaux* jouent un rôle important dans cette phase respiratoire. Ces muscles, en se contractant, abaissent les côtes et rétrécissent la ceinture abdominale, provoquant la fermeture thoracique (Draper & al 1959, Lebrun 1966, Le Huche & Allali 1991). Leur action est donc antagoniste à celle du diaphragme. La figure 1.3 résume l'action des muscles respiratoires durant les phases d'inspiration et d'expiration.



**Figure 1.3 – Action des muscles respiratoires durant les phases (a) d'inspiration et (b) d'expiration** (d'après www.paramed-prepa.com).

### 1.2.3 Respiration et phonation

Dans un acte de respiration calme, le rythme des cycles d'inspiration et d'expiration semble régulier. La durée des deux temps respiratoires est relativement comparable, bien que le temps d'expiration soit légèrement allongé.

Lors de la phonation, le rythme respiratoire est volontairement modifié. L'inspiration se raccourcit, devient plus intense et prend le sens « d'un élan du geste phonatoire » tandis que l'expiration s'allonge et devient « souffle phonatoire ». La voix résulte donc d'une expiration sonorisée. Les volumes d'air mobilisés s'accroissent également par rapport au « volume courant » lors d'une phase de respiration calme. Le locuteur va alors mobiliser 60 à 80% de sa capacité vitale.

En outre, l'obstacle de l'air au niveau du larynx et plus spécifiquement des plis vocaux va éléver la pression sous-glottique (PSG). Les muscles respiratoires vont devoir

1. Les muscles intercostaux internes sont tendus entre la lèvre externe des côtes supérieures et la face supérieure des côtes inférieures.

s'adapter pour participer à cette augmentation de la PSG et surtout pour maintenir la vibration. De manière schématique, nous relevons trois types d'activités musculaires durant la phonation. Premièrement, les muscles inspiratoires doivent se contracter plus longtemps pour ralentir le processus de fermeture de la cage thoracique. Deuxièmement, à mesure que l'air pulmonaire décroît, les muscles inspiratoires se relâchent. La pression doit se maintenir par l'action des muscles expiratoires qui prennent le relai. L'action des muscles expiratoires permet d'obtenir des mouvements expiratoires plus rapides, plus amples et prolongés. Finalement, à la fin de l'émission sonore et lorsque la pression de relaxation devient négative, l'intervention des muscles abdominaux devient importante.

## 1.3 Le système phonatoire

Le larynx et, en particulier, les plis vocaux sont à l'origine de la phonation. Il se situe au carrefour entre les voies aériennes et digestives. Néanmoins, cette fonction phonatoire est secondaire chez l'homme. Le larynx reste avant tout un organe sphinctérien (Négus 1949). Il participe à la respiration, à la toux et son action est essentielle dans la protection des voies respiratoires lors de la déglutition. Il est constitué d'un ensemble de cartilages, reliés entre eux par différents ligaments et muscles.

Dans le cadre d'une phonation, le larynx fonctionne comme « *un vibrateur* ». L'air pulmonaire est transformé en vibrations sonores lors de son passage entre les deux plis vocaux. Un bourdonnement laryngé est perceptible. Ainsi, Guerrier & al. (2004 : 3), résument le mécanisme de phonation comme :

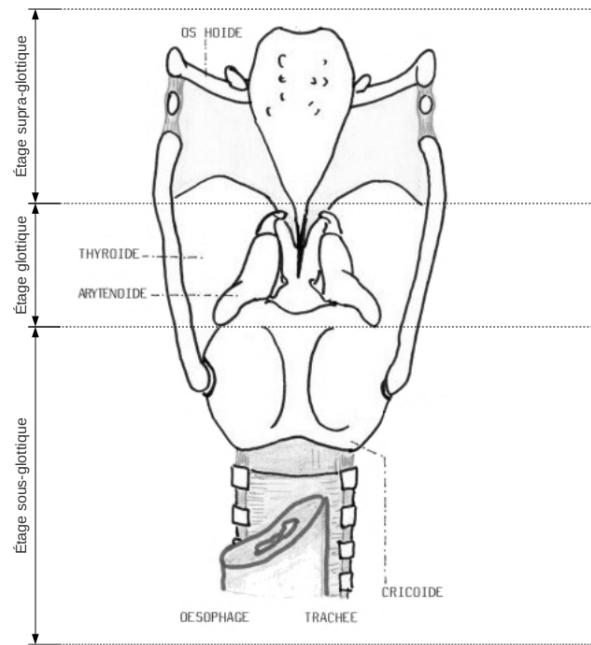
« [...] l'ensemble des phénomènes physiologiques qui permettent la production d'une vibration sonore au niveau du bord libre des cordes vocales. Le larynx est donc l'organe central de la production vocale. La phonation correspond à la vibration de l'air dans certaines circonstances : fermeture contrôlée du larynx pour produire un rétrécissement suffisant, mise en pression de l'air trachéal pour une expiration active. »

Dans cette section, nous aborderons l'anatomie du larynx ainsi que le fonctionnement des éléments qui le constituent. Nous verrons que celui-ci est composé d'un ensemble complexe de muscles et de cartilages qui concourent, directement ou indirectement, à la mise en vibration des plis vocaux. Après avoir examiné la structure interne des plis vocaux et évoqué certains modèles mécaniques simulant leur comportement, nous porterons notre attention sur les différentes théories de la phonation, passées ou actuelles.

### 1.3.1 Configuration interne du larynx

Le larynx est divisé en trois étages distincts : l'étage *sous-glottique*, l'étage *glottique* et l'étage *supraglottique*. L'étage sous-glottique correspond à l'espace situé entre le haut de la trachée et la face inférieure des plis vocaux. L'étage glottique correspond à la glotte, formée par l'espace entre les deux plis vocaux. Cette ouverture triangulaire est limitée par la commissure antérieure à l'avant et les cartilages arytenoïdes à

l’arrière. L’étage supraglottique correspond aux cavités situées au dessus des plis vocaux. Cet étage appelé « vestibule laryngé » est situé entre les bandes ventriculaires et l’ouverture du larynx. Cet étage comprend également l’épiglotte et ouvre sur la partie inférieure du pharynx.



**Figure 1.4** – Vue postérieure du larynx montrant les différents étages laryngés (d’après Huu Nguyen & al. 1999 :74).

### 1.3.2 Les muscles du larynx

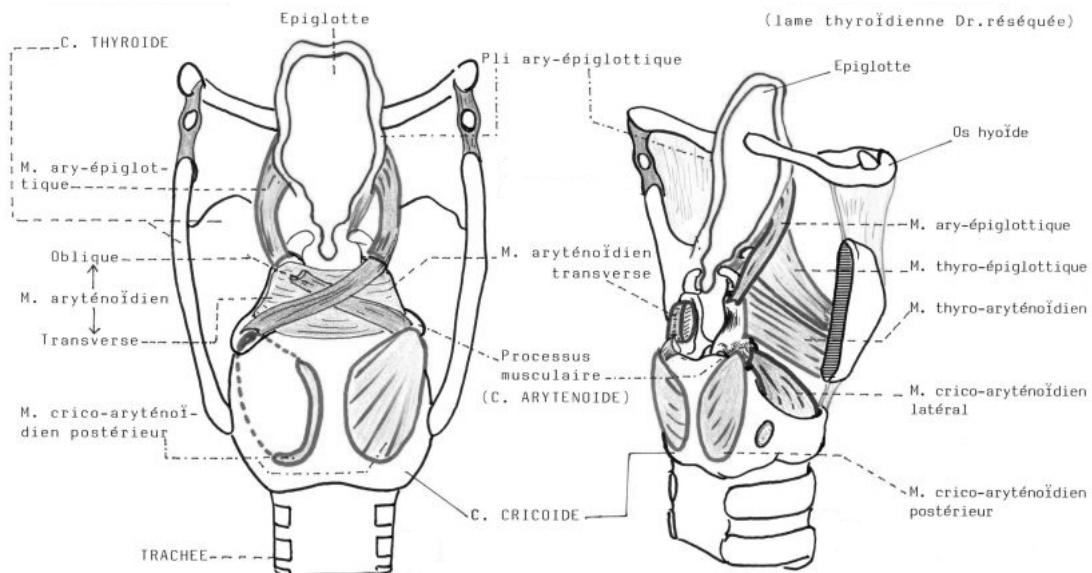
Nous pouvons séparer les muscles laryngés *intrinsèques*, qui permettent le contrôle de l’abduction, de l’adduction et le réglage de la tension des plis vocaux, des muscles *extrinsèques* qui contrôlent le réhaussement et l’abaissement du larynx (Fink & Desmaret 1978).

#### 1.3.2.1 Les muscles laryngés intrinsèques

Parmi les muscles intrinsèques, le muscle sans doute le plus important est le muscle thyro-aryténoidien (TA) qui représente le muscle du pli vocal. Celui-ci permet de raccourcir et d’épaissir le muscle vocal. En conséquence, sa contraction provoque une diminution de la partie vibrante du pli vocal et donc une baisse de la fréquence fondamentale.<sup>2</sup> Le muscle crico-thyroïdien (CT) est le muscle tenseur du pli vocal. Sa contraction est associée à un mouvement de bascule du cartilage cricoïde sur le cartilage thyroïde, ce qui provoque un allongement du pli vocal, élément essentiel dans le

2. La fréquence fondamentale, notée F0, correspond au nombre de vibrations par seconde des plis vocaux. Nous détaillerons plus précisément ce paramètre dans le chapitre 3, section 3.6.4.1.

réglage de la hauteur. Le muscle crico-aryténoïdien postérieur (CAP) est le seul muscle abducteur de la glotte. Sa contraction est responsable de l'écartement des plis vocaux, et donc de l'ouverture de la glotte. Son action serait à rattacher à l'abaissement de  $F_0$  et de la sonie et donc à la production des consonnes non voisées. Le muscle crico-aryténoïdien latéral (CAL), quant à lui, est un muscle adducteur dans le sens où il permet le rapprochement des plis vocaux. Celui-ci joue un rôle dans l'augmentation de la compression des plis vocaux et l'élévation de  $F_0$ . Enfin, les muscles aryténoïdiens transverses et obliques (IT et IO)<sup>3</sup>, vont de pair avec l'action des CAL et sont centraux dans la phonation. Leur contraction provoque également l'adduction des plis vocaux via le rapprochement des deux aryténoïdes (cf. figure 1.5). Ces muscles sont essentiels pour la mise en vibration des muscles vocaux (cf. Le Huche & Allali 1991, Crevier Buchman & al. 2003, Norton 2009).



**Figure 1.5 –**Vues postérieure et latérale des muscles intrinsèques du larynx (M = muscles et C = cartilage, d'après Huu Nguyen & al. 1999 :79).

Le tableau ci-dessous donne un récapitulatif des différentes actions musculaires (muscles intrinsèques) mises en oeuvre dans la phonation :

3. Cet ensemble de muscles constitue anciennement les muscles inter-aryténoïdiens (IA).

Muscles	Abbréviations	Action(s) du muscle	Effets sur la phonation
Thyro-aryténoïdien	TA	Raccourcit et épaisse le muscle vocal/ Diminue la tension des plis vocaux	Abaissement de la $F_0$ . Fonction sphinctérienne de la glotte.
Crico-thyroïdien	CT	Tenseur des plis vocaux	Elévation de $F_0$
Crico-aryténoïdien postérieur	CAP	Abducteur des plis vocaux. Dilatateur de la glotte	Abaisse $F_0$ . Réalisation des sons sourds
Crico-aryténoïdien latéral	CAL	Constricteur de la glotte/ Adduction des plis vocaux	Élévation de $F_0$
Aryténoïdiens transverses et obliques	AT et IO	Adduction des plis vocaux et des aryténoïdes	Mise en vibration des plis vocaux

**Table 1.1 – Actions des muscles intrinsèques du larynx (d'après Norton 2009).**

### 1.3.2.2 Les muscles laryngés extrinsèques

Les muscles extrinsèques constituent le système de suspension du larynx. Ils permettent le réhaussement et l’abaissement de ce dernier lors de la déglutition ou de la phonation. Ces muscles sont divisés en muscles *sous-hyoïdiens* et muscles *sus-hyoïdiens* selon leur insertion à l’os hyoïde.

Les muscles sus-hyoïdiens (digastrique, mylo-hyoïdien, génio-hyoïdien et stylo-hyoïdien) permettent la liaison entre la mandibule, la base du crane et l’os hyoïde principalement. De leur activité dépendent l’ouverture de la mandibule et l’élévation du larynx (cf. figure 1.6). Le tableau suivant montre l’action de ces muscles et, s’il y a lieu, leurs effets sur l’articulation des sons :

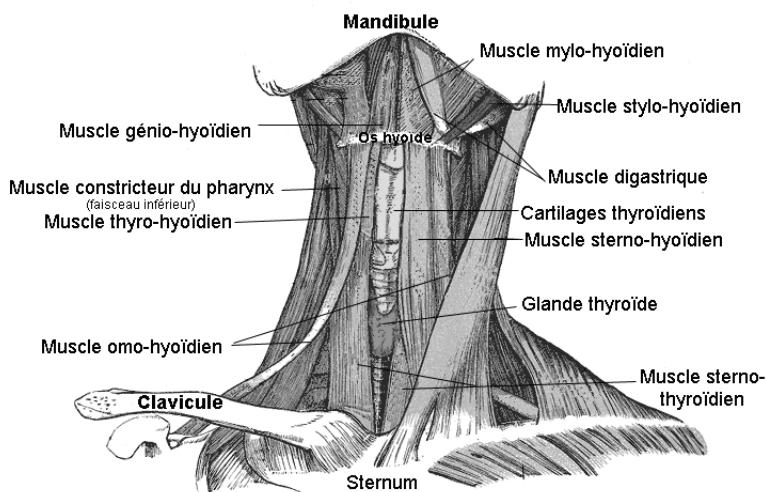
Muscles	Action du muscle	Effet sur l’articulation
Digastrique	Elève l’os hyoïde/ Contribue à l’ouverture de la mandibule	Réalisation articulatoire des sons
Mylo-hyoïdien	Elève l’os hyoïde/ Elève le plancher de la bouche	
Génio-hyoïdien	Aide à la propulsion de l’os hyoïde et de la langue	
Stylo-hyoïdien	Elève l’os hyoïde/ Rétropulsion de l’os hyoïde	

**Table 1.2 – Actions des muscles extrinsèques sus-hyoïdiens du larynx (d’après Norton 2009).**

Les muscles sous-hyoïdiens (omo-hyoïdien, sterno-thyroïdien, thyro-hyoïdien, sterno-hyoïdien) participent quant à eux, à la mobilisation du cartilage thyroïde dans les changements de tension des plis vocaux. Ces muscles ont un rôle dans l'abaissement du larynx et dans la modification des registres (cf. Figure 1.6). Le tableau ci-dessous montre l'action de ces muscles et, s'il y a lieu, leurs effets sur l'articulation des sons :

Muscles	Action du muscle	Effet sur la phonation
Omo-hyoïdien	Abaisse l'os hyoïde et le larynx	
Sterno-hyoïdien	Abaisse l'os hyoïde et épaisse les plis vocaux	Abaisse la $F_0$ . Rôle préparatoire à la phonation
Thyro-hyoïdien	Abaisse l'os hyoïde et le larynx	Abaisse la $F_0$ . Rôle dans l'articulation vélaire
Sterno thyroïdien	Abaisse la larynx	

**Table 1.3 – Actions des muscles extrinsèques sous-hyoïdiens du larynx (d'après Norton 2009).**



**Figure 1.6 – Muscles extrinsèques du larynx (sous et sus-hyoïdiens) (d'après [www.wikipedia.org/wiki/Fichier:Muscle\\_stylo-hyoidien](http://www.wikipedia.org/wiki/Fichier:Muscle_stylo-hyoidien)).**

### 1.3.3 Les cartilages du larynx

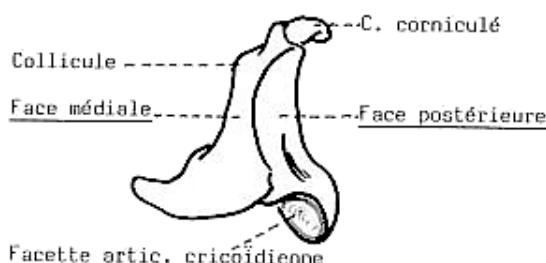
Le larynx compte cinq cartilages principaux : le cartilage cricoïde, les cartilages arytenoïdes, le cartilage thyroïde et l'épiglotte associés à un unique os : l'os hyoïde.

### 1.3.3.1 Les structures cartilagineuses

*Le cartilage cricoïde* (cf. figure 1.5), est décrit comme étant en forme de bague et dont la pierre serait tournée vers l’arrière. Il s’articule avec le cartilage thyroïde formant l’articulation crico-thyroïdienne et le cartilage arytenoïde formant l’articulation crico-aryténoïdienne. De manière générale, il supporte l’ensemble de la structure laryngée. Comme nous le verrons ultérieurement, il reste indispensable dans le processus respiratoire car il permet au larynx de se maintenir en position ouverte. Sa conservation est donc nécessaire dans les chirurgies partielles laryngées (Crevier Buchman 1999, 2003, Brasnu & al. 2005, Norton 2009).

*Le cartilage thyroïde* (cf. figure 1.5), en forme de livre ouvert, forme une saillie en avant qui correspond à la pomme d’Adam. L’angle est plus large chez la femme, ce qui explique que la pomme d’Adam soit plus proéminente chez l’homme. Il est également constitué de deux cornes postérieures qui permettent l’articulation du cartilage thyroïde avec le cartilage cricoïde.

*Les deux cartilages arytenoïdes* (cf. figure 1.7) sont situés en arrière sur le cartilage cricoïde. Ils sont composés d’un petit cartilage incurvé, le cartilage corniculé, qui recouvre le haut de la pyramide aryténoïdienne. Leur particularité est qu’ils peuvent s’écartier et se rapprocher, assurant ainsi la majeure partie de la fonction laryngée.



**Figure 1.7 – Vue post-médiale du cartilage arytenoïde (C= cartilage, d’après Huu Nguyen & al. 1999 :75).**

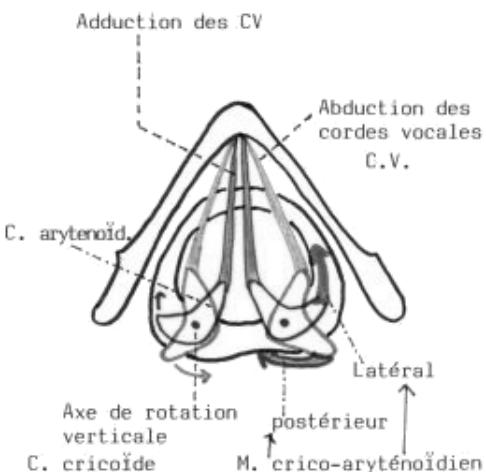
*L’épiglotte*, dressée verticalement, a une fonction de clapet à l’entrée du larynx. Son rôle principal est de participer à la fermeture du larynx lors de la déglutition.

*L’os hyoïde* est l’unique os qui compose le larynx. En forme de U, ouvert vers l’arrière, cet os est un carrefour de rattachement à la majorité des muscles du cou. C’est aussi sur lui que la langue est majoritairement insérée.

### 1.3.3.2 L’articulation crico-aryténoïdienne

Cette unité anatomique et fonctionnelle (Arnoux-Sindt 1992) est constituée du cartilage cricoïde, d’un cartilage arytenoïde et des muscles CAP, CAL, IO et IT. L’articulation crico-aryténoïdienne est à l’origine de deux mouvements distincts (cf. figure 1.8). Premièrement, une rotation sur l’axe vertical du cartilage cricoïde a pour effet de

rapprocher (adduction) ou d'éloigner (abduction) les plis vocaux. Deuxièmement, un glissement antérieur ou postérieur a pour effet de modifier la longueur des plis vocaux.



**Figure 1.8** – Vue supérieure schématique de l'articulation crico-aryténoïdienne (d'après Huu Nguyen & al. 1999 :76).

En anticipant ce qui sera évoqué dans le chapitre suivant, celle-ci permet de conserver une déglutition normale, une respiration et une fonction sphinctérienne intacte. Sa conservation est donc essentielle après laryngectomie partielle pour permettre un fonctionnement normal du carrefour aérodigestif du larynx. Brasnu & al. (2005 :15) ajoutent par exemple que :

« La préservation d'au moins une unité crico-aryténoïdienne fonctionnelle est nécessaire. Elle constitue l'unité de base du larynx, elle permet de préserver les fonctions physiologiques sphinctériennes du larynx. L'unité crico aryténoïdienne permet la déglutition et la phonation. »

Par conséquent, l'articulation crico-aryténoïdienne constitue l'unique vibrateur qui reste dans l'endolarynx dans certaines chirurgies partielles.

### 1.3.4 Les plis vocaux

#### 1.3.4.1 L'espace glottique

L'espace entre les plis vocaux est la glotte. Celle-ci se divise en deux parties, à savoir une glotte postérieure et cartilagineuse délimitant « la glotte respiratoire » et une glotte antérieure ou musculo-ligamentaire délimitant « la glotte phonatoire ». Ces plis vocaux jouent le rôle « *d'oscillateur* » puisque ils transforment l'air en un bourdonnement. Les plis vocaux s'accroient mollement, grâce aux différentes actions des muscles laryngés, pour la production des sons voisés, à savoir les consonnes voisées et les voyelles. Lors de l'articulation des sons non voisés, la glotte reste ouverte. De manière plus précise, Ladefoged & Maddieson (1998) ou Stevens (1977) ont mis en

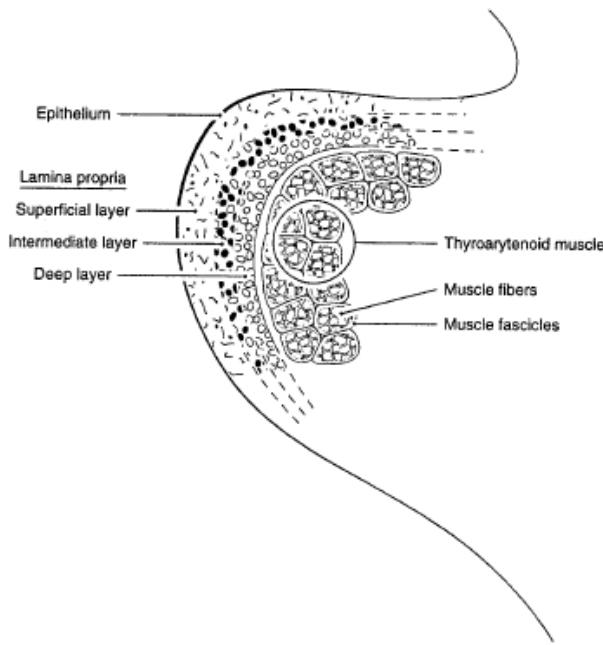
évidence deux modes de production des consonnes et trois modes de production des voyelles. En fonction de la compression ou de l'adduction des plis vocaux, on obtiendra soit une voix soufflée dite « *breathy* » soit une voix rauque dite « *creaky* ». Les consonnes soufflées sont réalisées lorsque les plis vocaux ne s'accroient pas totalement et vibrent avec une fuite d'air postérieure. En phonétique acoustique, le caractère soufflé est reconnaissable à la baisse de l'amplitude du F1, à la diminution de l'amplitude des harmoniques supérieurs, à la diminution de l'énergie dans les hautes fréquences et à l'apparition de bruit. Au contraire, les sons pressés sont obtenus par un rapprochement serré et intense des arytenoïdes et une tension importante des plis vocaux. La vibration des plis vocaux devient incomplète et irrégulière en fréquence et en amplitude (Malmberg 1954, Stevens 1991, Crevier Buchman 1999, entre autres).

En ce qui concerne les voyelles, nous relevons trois modes de production : modale, soufflée (*breathy*) ou serrée (*creaky*). Les voyelles soufflées sont produites avec un espace glottique qui ne ferme pas complètement, c'est-à-dire avec des arytenoïdes écartés. Dans ce modèle, les plis vocaux sont détendus et le larynx en position basse. Pour les voyelles pressées, nous constatons une configuration glottique resserrée où les cartilages arytenoïdes et les plis vocaux sont tendus et extrêmement rapprochés.

#### 1.3.4.2 Structure interne des plis vocaux

Les plis vocaux sont soutenus par les cartilages arytenoïdes. Ils sont essentiellement constitués de trois éléments : un muscle vocal, une couche moyenne (la *lamina propria*) et une muqueuse (*épithélium*) qui recouvre et protège le tout. La figure 1.9 montre le schéma d'une coupe frontale d'un pli vocal. La couche extérieure du pli vocal est fine et correspond à un *épithélium*<sup>4</sup> stratifié (Hirano 1981). Il mesure de 0.05 à 0,10 mm d'épaisseur et agit comme une couverture dont le but est de maintenir la forme du pli vocal. Celui-ci constitue la partie la plus mobile du pli vocal puisqu'il ondule sous l'effet des muscles environnants et surtout de la pression sous-glottique (Hirano & al, 1986). La *lamina propria*, un système composé de fibres de collagène et d'élastine (tissu conjonctif), se trouve entre l'*épithélium* et le muscle thyro-arytenoïdien. Celle-ci peut être divisée en trois couches : superficielle, intermédiaire et profonde. La couche superficielle ou espace de Reinke, qui peut être comparée à une masse de gélatine molle, est constituée de fibres de protéine élastiques entourées par les liquides interstitiels. La couche intermédiaire, qui peut être comparée à un faisceau de bandes de caoutchouc souple, est constituée principalement de fibres élastiques. En ce qui concerne la couche profonde, celle-ci se compose de fibres de collagène qui pourraient être comparées à un paquet de fils de coton. La couche superficielle mesure environ 0,5 mm d'épaisseur au milieu du pli vocal alors que les couches intermédiaires et profondes mesurent environ 1 à 2 mm à elles deux. Le muscle *thyro-arytenoïdien* constitue la majeure partie (le corps) du pli vocal, soit environ 7-8 mm d'épaisseur (Titze 1994).

4. L'*épithélium* est un tissu mince formé de couches de cellules juxtaposées qui recouvre divers organes du corps (cavités internes, glandes, etc.).

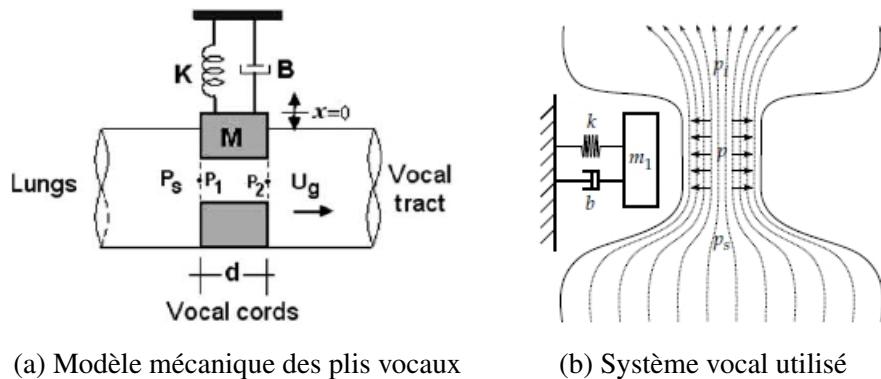


**Figure 1.9 – Schéma d'une coupe frontale d'un pli vocal (d'après Titze 1994).**

La fréquence fondamentale du son — exprimée en Hertz — produite par la vibration des plis vocaux dépend de la masse, de la longueur vibrante de ces derniers, de l'élasticité du conduit vocal et de la pression sous-glottique. Toute modification de l'oscillateur entraînera des modifications du mode et de la régularité de la vibration. Si l'on augmente la masse vibrante, il y aura un abaissement de  $F_0$ . Si les plis vocaux ne s'accroient pas correctement, une fuite glottique sera constatée. Enfin, une asymétrie de masse sera responsable de perturbations de la vibration en fréquence et en amplitude.

#### 1.3.4.3 Modèles mécaniques des plis vocaux

Les modèles masse-ressort constituent une première approche pour simuler le comportement des plis vocaux. Ils sont construits en amalgamant la masse des plis vocaux à quelques éléments de masse discrète, reliés les uns aux autres, et à une frontière rigide avec des éléments de ressort et d'amortissement. La figure 1.10 montre un modèle simple des plis vocaux faisant intervenir une masse.



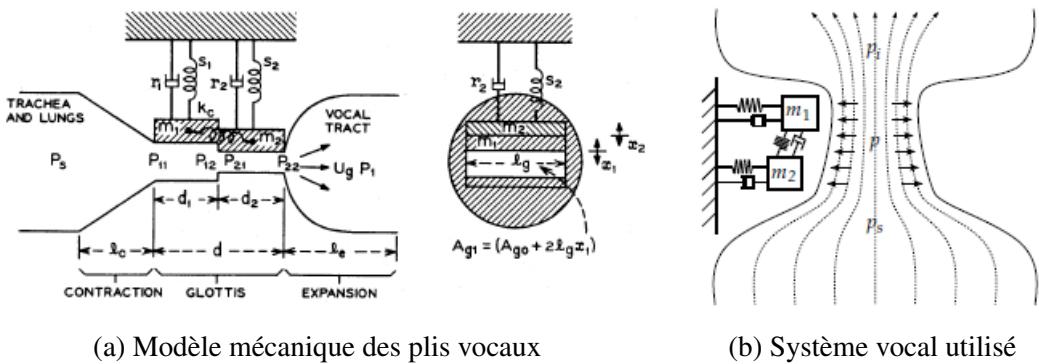
**Figure 1.10 –** Deux représentations des plis vocaux dans un modèle à une masse (d’après Titze 1994).

Dans un modèle des plis vocaux à une masse, ces derniers sont considérés comme un simple système constitué d’une masse, d’un ressort et d’un amortisseur (où  $m$  représente la masse,  $k$  la constante du ressort et  $b$  la constante de l’amortisseur, Flanagan & Landgraf 1968). La masse correspond à la masse totale des plis vocaux humain. La constante du ressort correspond à l’élasticité du tissu alors que l’amortissement rend compte de la perte d’énergie vibratoire causée par la viscosité du tissu. La constante du ressort peut être variée selon la tension des plis vocaux et l’amortissement peut être déterminé de manière expérimentale. Le système est excité par la force  $F(t)$ , résultant du produit de la pression de l’air dans la glotte et de l’aire de la surface intra-glottale. La force, qui agit sur les plis vocaux, est distribuée et sa résultante est appliquée à la masse  $m$ .

Toutefois, le problème que pose un modèle des plis vocaux à une masse est que celui-ci n’est capable de représenter que le déplacement *latéral* des plis vocaux. Puisque l’onde mucosale ne peut pas être modélisée, ce modèle n’est en mesure de maintenir à lui seul une oscillation, qu’en présence d’un conduit vocal et de l’inertance<sup>5</sup> de la colonne d’air à l’intérieur du conduit (cf. Titze 1994). Par conséquent, un modèle à une masse de ce type exagère l’effet du couplage du conduit vocal et des plis vocaux et n’est généralement pas suffisant pour rendre compte des mécanismes mis en jeu. Par exemple, dans l’étude de Titze (1994), le degré d’interaction acoustique produite entre la source et le conduit vocal était plus grande que celle observée dans la parole humaine. Néanmoins, Avanzini & al. (2001) ont récemment proposé un modèle des plis vocaux à une masse amélioré dans lequel la simplicité du modèle original a été retenue mais qui, par ailleurs, est capable de produire des formes d’onde glottique réalistes.

Devant les problèmes soulevés par la modélisation à une masse de Flanagan & Landgraf (1968), l’idée a été d’introduire davantage de propriétés physiologiques et des représentations à masses multiples ont alors été considérées. Ainsi, dans le modèle d’Ishizaka & Flanagan (1972) par exemple, chaque pli vocal est représenté par deux oscillateurs couplés masse-ressort-amortisseur (cf 1.11).

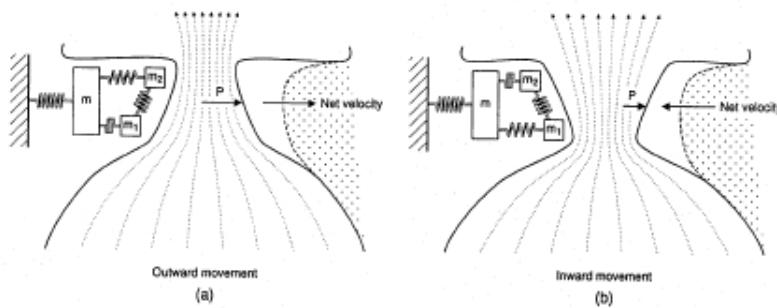
5. En acoustique, l’inertance correspond à l’inertie qu’oppose une masse mécanique à une variation rapide de vitesse.



**Figure 1.11** – Deux représentations des plis vocaux dans un modèle à deux masses (d'après Ishizaka & Flanagan 1972).

Le modèle à deux masses est en mesure de simuler la différence de phase entre les parties hautes et basses des plis vocaux. En plus du déplacement général des tissus, les deux degrés de liberté (les masses) permettent de représenter l'onde de la muqueuse, ce que n'étaient pas en mesure de faire les modèles à une masse. Par conséquent, un modèle à deux masses facilite l'oscillation et ce même en l'absence d'un conduit vocal et de l'inertance. La limite des modèles des plis vocaux à deux masses est que leur discrétisation du tissu — c'est-à-dire le fait de remplacer une variable continue, une partie du pli vocal dans le cas qui nous intéresse ici, par une variable discrète (i.e. un ressort) — ne tient pas compte de la structure en couches des plis vocaux (Story 2002). Malgré le fait que la masse inférieure ( $m_2$ ) dans le modèle de Ishizaka and Flanagan (1972) est plus épaisse et plus lourde afin de simuler l'effet du muscle thyro-aryténoïdien, la configuration ne permet pas une oscillation conjuguée de la couverture (*épithélium* et première couche de la *lamina propria*) et du corps du pli vocal. Par conséquent, le modèle à deux masses peut être interprété comme un modèle qui rend compte uniquement de l'oscillation de la couverture et non de l'oscillation conjuguée de la couverture et du corps du pli vocal. En outre, dans les plis vocaux humains, la tension est contrôlée par la contraction du muscle thyro-aryténoïdien. Dans le modèle à deux masses de Ishizaka and Flanagan (1972), il n'existe aucune corrélation physiologique directe entre la tension des ressorts et l'effet de la contraction musculaire (Story and Titze 1995), ce qui pose un problème non négligeable à ce type de modélisation.

Les problèmes liés à cette modélisation peuvent néanmoins être résolus en ajoutant une masse supplémentaire au modèle (cf. figure 1.12), créant ainsi un modèle à trois masses (Hirano 1977, Story and Titze 1995). En effet, selon Hirano (1977), le pli vocal peut être décrit comme un modèle à trois masses où l'*épithélium* et l'espace de Reinke représentent les parties les plus souples et les plus propices aux forces aérodynamiques, la couche profonde de la *lamina propria* assure quant à elle une stabilité du pli vocal lors de la mise en vibration et où le muscle vocal est stable et ne bouge que très peu lors de la vibration des plis vocaux.



**Figure 1.12 – Modèle des plis vocaux à trois masses (d’après Hirano 1977).**

Dans ce type de représentation, la masse supplémentaire ( $m$ ) simule le corps du pli vocal et est positionnée de manière latérale par rapport aux deux masses qui simulent la couverture du pli vocal ( $m_1$  et  $m_2$ ). La connexion de ces deux masses représente aussi bien la tension propre à la couverture que la tension qui existe entre le corps et la couverture du pli vocal. La masse du corps est quant à elle connectée à une frontière rigide. Cette connexion représente la tension propre au corps du pli vocal, laquelle dépend du degré de contraction du muscle thyro-aryténoïdien. Par conséquent, ce modèle permet de rendre compte de manière réaliste des paramètres physiologiques qui caractérisent les composants du corps et de la couverture du pli vocal. Ainsi, la contraction du muscle thyro-aryténoïdien augmente la tension du corps du pli vocal mais pas nécessairement celle de la couverture.

### 1.3.5 Physiologie de la phonation

La connaissance et la caractérisation de la structure du pli vocal — modélisations y compris — sont essentielles et surtout à l’origine du développement des théories de la phonation. Dans cette partie, nous apporterons un aperçu historique des différentes théories de la phonation avant de nous concentrer sur la théorie adoptée actuellement dans le domaine. Cette théorie permet de rendre compte de manière satisfaisante du cycle vibratoire des plis vocaux. Enfin, nous examinerons brièvement les différents paramètres acoustiques de la voix.

#### 1.3.5.1 Les théories de la phonation : aperçu historique

A notre connaissance, la première théorie de la phonation a été postulée par Ewald (1898). Il s’agit d’une *théorie myoélastique* qui considère principalement la vibration des plis vocaux comme passive. Elle stipule que ces vibrations sont uniquement influencées par l’air expulsé des poumons. Selon lui, les différents paramètres acoustiques dépendent seulement de la pression sous-glottique et de la tension des plis vocaux. Husson (1950) pointera plus tard du doigt que cette théorie ne permet pas d’expliquer les variations indépendantes de hauteur et d’intensité.

Plus d’un demi siècle plus tard, Husson (1950) souligne les insuffisances de la théorie d’Ewald et propose une théorie avec un modèle vibratoire actif (*théorie neuro-chronaxique*). La fréquence de vibration des plis vocaux serait induite par la fréquence

de potentiels d'actions véhiculés par le nerf récurrent. Ainsi, la fréquence du son dépendrait de ces afflux nerveux alors que l'intensité serait liée à la pression sous-glottique. Dans ce cas, nous sommes effectivement en présence de deux mécanismes distincts.

Perello (1962) met, quant à lui, en évidence l'importance de la muqueuse du larynx dans la phonation, ce que Husson rejettait totalement. Dans sa *théorie mucocindulatoire*, Perello note que la muqueuse du pli vocal semble onduler sous la pression de l'air et que son altération ou inflammation modifie la voix. Il fut l'un des précurseurs dans la mise en évidence du *phénomène de Bernouilli* (défini à partir de la Théorie Physique des Fluides) adapté au larynx : lorsque l'on fait circuler un fluide dans un tube dont le diamètre varie, on observe une diminution de la vitesse d'écoulement dans les zones où le diamètre du tube est le plus grand et une diminution de la pression dans les zones où le diamètre du tube est plus petit. Dans ce cas, la vitesse devient plus importante. Par définition, la glotte représente un rétrécissement du conduit aérien. Lorsqu'un flux d'air (suffisamment rapide) circule à travers la glotte entrouverte, une baisse de la pression de l'air est constatée à ce niveau. Cette dépression entraîne l'accroissement de la muqueuse des plis vocaux et donc la fermeture de la glotte. À ce stade, la dépression disparaît, ce qui provoque l'augmentation de la pression sous-glottique (PSG) et la réouverture de la glotte.

Van Den Berg (1958, 1968), qui s'appuie sur les mêmes arguments que Perello, complète la théorie en restreignant l'action de la muqueuse aux sons les plus faibles. De plus, sa *théorie aérodynamique myoélastique* vise à confirmer le phénomène d'adduction, à savoir un rapprochement des arytenoïdes via l'action des muscles interarytenoïdiens afin de permettre aux plis vocaux de vibrer.

Sylvestre & MacLeod (1968) proposent, quant à eux, une *théorie neuro-oscillatoire* basée sur celle de Husson et précisent que la vibration du pli vocal est essentiellement due au muscle vocal. Ces travaux comparent le muscle vocal aux muscles asynchrones des insectes. Ainsi, l'oscillation dépend uniquement de l'inertie et de l'élasticité des structures en mouvement, après afflux nerveux.

Pour finir, Dejonckere (1981) propose une *théorie oscillo-impédantielle*, basée sur la théorie myoélastique de Van Den Berg dans laquelle s'applique le phénomène de Bernouilli. Selon celle-ci, le larynx est en fait un oscillateur à faible amortissement. En position préphonatoire, les plis vocaux pourraient être mis en vibration par la seule force aérodynamique. En phonation, l'énergie acoustique résulterait de la fluctuation de la PSG. Aussi, l'auteur compare la fréquence des oscillations des plis vocaux à la fréquence des battements d'ailes d'un insecte. Ces oscillations sont réglées par des propriétés d'impédance mécanique des parties vibrantes de la muqueuse. Finalement, le timbre dépendrait de la qualité de l'ondulation qui serait elle-même corrélée à la qualité de la muqueuse associée à une vibration régulière des plis vocaux.

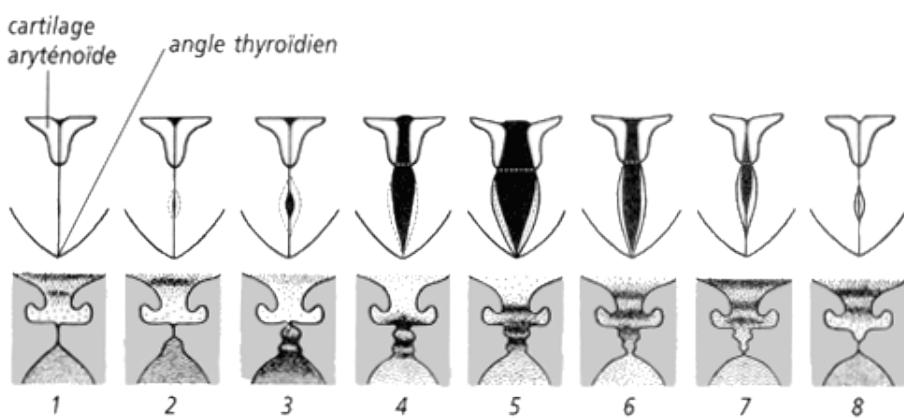
### 1.3.5.2 La théorie actuelle de la phonation

De toutes les théories évoquées précédemment, la théorie adoptée actuellement est celle de Van Den Berg (1958, 1968), laquelle a été complétée par Hirano (1981) et De-

jonckere (1981). Cette théorie prend en compte trois facteurs. Premièrement, le *facteur myoélastique* rend compte du fait que l'ondulation du pli vocal dépend essentiellement de la masse de ce dernier, de son élasticité, de sa longueur, de sa tension, de sa raideur et de son inertie (Hirano 1981). Les caractéristiques vibratoires sont également corrélées à la longueur et à la tension musculaire du larynx. Deuxièmement, le *facteur aérodynamique* précise que l'acte de phonation nécessite un équilibre entre pression sous-glottique (PSG) importante et une basse pression supraglottique afin qu'il se crée un flux transglottique. Enfin, le *concept oscillo impédantiel* implique que les deux plis vocaux seraient des oscillateurs qui transformeraient la pression de l'air expiré en une pression pulsée dont la fréquence serait similaire à celle de l'oscillateur.

### 1.3.5.3 Typologie d'un cycle de vibration des plis vocaux

Cette théorie actuelle de la phonation rend compte de la mise en vibration des plis vocaux. Comme le montre la figure 1.13, en position préphonatoire, les plis vocaux sont totalement fermés (1). Au fur et à mesure, le flux d'air devient plus important dans la partie sous-glottique. La résistance mécanique au niveau des deux plis vocaux serrés est vaincue par cette pression sous-glottique (PSG) qui continue de pousser. Cette poussée contribue à écarter les deux plis vocaux, en commençant par leur bord inférieur (2-4). La poussée de l'air permet l'ouverture complète de la glotte (5), ce qui abaisse progressivement la PSG. Les plis vocaux se referment par la suite, du bord inférieur vers le bord supérieur, grâce à leur élasticité mais surtout par l'effet Bernouilli (6-8). Ce phénomène physique est caractérisé par un effet de succion des bords libres des deux plis vocaux, soit un effet de *rétroaspiration*. La fermeture glottique permet le rétablissement de la résistance glottique et de la PSG. Finalement, tant que la PSG est maintenue, le mouvement vibratoire des plis vocaux est cyclique.



**Figure 1.13 –**Schéma d'un cycle complet de vibration des plis vocaux. Les figures du haut représentent une vue supérieure de la glotte et les figures du bas une vue sagittale de la glotte (Ormezzano 2000 :103) (d'après Hirano 1981).

### 1.3.5.4 Les paramètres acoustiques de la voix

Il existe quatre principaux paramètres qui caractérisent la voix sur le plan acoustique et nous renseigne sur son efficacité. Une voix doit être efficace pour que la communication puisse s'effectuer sans problème particulier. Les évaluations perceptives et les spectrogrammes<sup>6</sup> permettent de décrire les voix, qu'elles soient « normales » ou « pathologiques ». Dans cette partie, nous décrirons brièvement l'ensemble des paramètres acoustiques de la voix. Les paramètres pertinents étudiés dans cette thèse seront approfondis dans les chapitres suivants.

*La fréquence fondamentale* de la voix ( $F_0$ ), mesurée en Hertz, est corrélée à la fréquence de vibration, c'est-à-dire à la vitesse d'exécution des cycles des plis vocaux. Comme nous l'avons déjà mentionné plus haut, elle dépend de la masse, de la longueur vibrante des plis vocaux, de l'élasticité du système vocal et de la pression sous-glottique. Certains individus peuvent développer des voix plutôt graves, caractérisées par une  $F_0$  basse, tandis que d'autres peuvent avoir une voix davantage aigüe, synonyme d'une  $F_0$  élevée.

*L'intensité* d'un son traduit l'amplitude de la variation de pression correspondant à ce son. L'intensité vocale varie avec la pression sous-glottique. L'intensité exprimée en décibels (dB) permet de visualiser l'énergie du sujet pendant son acte phonatoire.

*La durée* est mesurée en millisecondes (ms) et secondes (sec). La durée d'un son et, plus largement, de la phrase est mesurable sur un spectrogramme.

Chaque son a son propre *timbre*. Une voyelle ou une consonne sont qualifiées de « claires » lorsque l'énergie est principalement regroupée dans les hautes fréquences, c'est-à-dire pour les voyelles, quand il existe une distance significative entre les deux premiers formants<sup>7</sup> et pour les consonnes, lorsque les résonances de fréquences supérieures sont excitées. Le [i] et le [s] en sont les exemples principaux. Une voyelle ou consonne « sombre », telles le [u] ou le [p], est caractérisée par un regroupement de l'énergie dans les basses fréquences. Les deux premiers formants de la voyelle sont donc regroupés sous 1000 Hz. Quant aux voyelles « centrales », l'énergie est répartie de façon plus uniforme.

*La qualité de voix*, quant à elle, est un paramètre qualitatif de richesse en harmoniques et d'agrément vocal. Elle est propre à chaque individu. Comme nous l'avons vu précédemment, une voix efficace correspond à des plis vocaux qui se ferment énergiquement et s'accroissent sur la totalité de leur masse vibrante (voix dite « creaky »). Au contraire, des plis vocaux qui s'accroissent plus que mollement provoquent des fuites glottiques, caractéristiques des voix dites « breathy » ou « voilées ».

Ces paramètres sont utiles aux thérapeutes, professionnels de la voix (orthophonistes et médecins), qui peuvent — et doivent — associer ces paramètres acoustiques à d'autres mesures types, telles que le débit d'air par exemple, afin de qualifier au mieux

6. Les spectrogrammes sont des représentations tri-dimensionnelles de la parole. En abscisses, nous distinguons le temps, en ordonnées les fréquences et la noirceur représente l'énergie acoustique. Une description plus approfondie de cet outil sera effectuée dans le chapitre 3, section 3.6.3.2.

7. Les formants sont des zones où les harmoniques sont renforcés. Les harmoniques, quant à eux, sont des multiples entiers de la  $F_0$ .

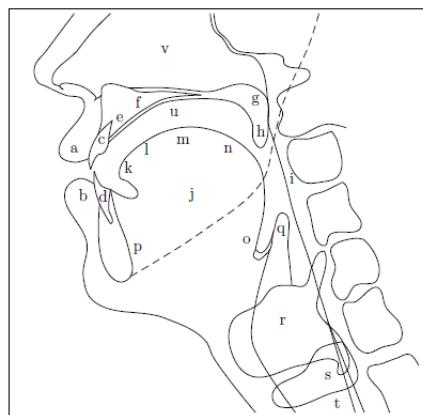
les voix pathologiques, le but étant d'évaluer les aspects qualitatifs, phonétiques et linguistiques. La rééducation se base essentiellement sur ces évaluations de la voix.

## 1.4 Le système articulatoire

Dans cette section, nous aborderons l'étage supraglottique et les organes qui le composent. Nous décrirons les différentes cavités et organes qui jouent le rôle de résonateurs. La théorie source-filtre de Fant (1960) sera également évoquée car elle représente une théorie classique dans la compréhension de la production des sons de la parole.

### 1.4.1 Les organes supraglottiques

L'articulation est l'ultime étape permettant la transformation de l'onde sonore obtenue par le passage de l'air dans le larynx, en sons concrets de la parole. Les résonateurs correspondent aux différentes cavités que le son laryngé traverse avant de ressortir par la bouche sous la forme de sons de la parole (Cornut 1983, Vaissière 2006, entre autres). Du plan glottique à la bouche, le parcours du son laryngé est d'environ 17 cm chez l'homme. La figure 1.14 montre une coupe sagittale de l'appareil vocal humain et des différents articulateurs.



**Figure 1.14 –** Vue sagittale de l'appareil vocal et des différents organes vocaux (d'après Laver 1994 :120).

- |                        |                         |                       |
|------------------------|-------------------------|-----------------------|
| a. Lèvre supérieure    | i. Parois pharyngale    | q. Epiglotte          |
| b. Lèvre inférieure    | j. Corps de la langue   | r. Cartilage thyroïde |
| c. Dents supérieures   | k. Pointe de la langue  | s. Cartilage Cricoïde |
| d. Dents inférieures   | l. Lame de la langue    | t. Trachée            |
| e. Barrière alvéolaire | m. Avant de la langue   | u. Cavité orale       |
| f. Palais dur          | n. Arrière de la langue | v. Cavité nasale      |
| g. Palais mou          | o. Racine de la langue  |                       |
| h. Uvule               | p. Mandibule            |                       |

Le *larynx* peut être un premier lieu d'articulation dans la production des sons. En effet, il participe à l'articulation des glottales et permet la production des consonnes éjectives et implosives dans certaines langues telles que l'arabe par exemple (Ladefoged 1973, Stevens 1977, Ladefoged & Maddieson 1998).

Le *pharynx* est la première cavité que le son laryngé traverse. Celui-ci se divise en trois parties : l'*oropharynx*, situé en arrière de la bouche dont il est séparé par le vélum, le *nasopharynx*, situé en arrière des fosses nasales et par lequel l'air s'achemine pour la production des sons nasals, et l'*hypopharynx*, situé en dessous des deux premiers. Le pharynx est essentiellement constitué de muscles dits constricteurs.

Le *vélum* ou *voile du palais*, sert principalement de séparateur entre la cavité buccale et les fosses nasales. Connu également sous le nom de *palais mou* ou *voile mou*, il prolonge le palais dur dans la bouche et se termine par une extrémité pendante : la luette. Le voile est constitué d'un ensemble complexe de muscles qui s'étend de la langue vers l'os hyoïde et la base du crâne. Le rôle principal de ce dernier est de s'abaisser et de se relever en fonction des sons de la parole produits. Ainsi, lors de la production de sons nasals, le vélum va s'abaisser afin de laisser passer l'air à la fois par la bouche et les fosses nasales. Au contraire, lors de la production de sons oraux, le voile reste relevé, l'air est bloqué au niveau du nasopharynx et ne peut s'écouler que par la bouche.

Comme nous l'évoquions plus haut, les *fosses nasales* sont séparées de la cavité buccale par le voile du palais. Placées horizontalement et d'avant en arrière entre les narines et les choanées, ces cavités tiennent leur rôle dans les sons nasals dans la production de la parole. Lors de l'articulation de voyelles ou de consonnes nasales, le vélum s'abaisse, laissant l'air s'échapper par la bouche et le nez. La cavité nasale, parcourue par le flux d'air, devient à elle seule zone de résonance.

En articulant, la *mandibule*, est largement mobilisée, ce qui induit des changements de forme et de volume de la cavité buccale. Des muscles reliant la mandibule à la base du crâne ou à l'os hyoïde permettent des mouvements d'abaissement et d'élévation de cette dernière, ce qui génère des sons ouverts ou au contraire fermés.

La *langue* est un organe mobile et complexe pourvue de 17 muscles. Celle-ci se divise en plusieurs parties : la pointe, bien visible dans la bouche et appelée communément *apex*, une partie intermédiaire formée de la *lame* et du *dos*, et une partie postérieure nommée *racine*. La langue s'étend de l'apex jusqu'à sa base qui s'implante sur l'os hyoïde, formant ainsi la paroi antérieure de l'oropharynx.

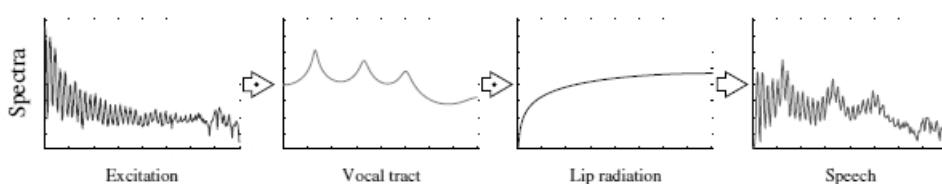
Les *lèvres* prolongent et terminent le conduit vocal. L'anatomie musculaire de celles-ci est très complexe et l'ensemble des muscles de la face doivent être mobilisés pour l'articulation des sons de la parole. Ces muscles permettent aux lèvres de s'ouvrir, se serrer, s'avancer, s'étirer, etc.

#### 1.4.2 La théorie source-filtre (Fant 1960)

Outre les modèles physiologiques évoqués précédemment qui faisaient intervenir des masses, une autre approche consiste à estimer la forme d'onde acoustique de manière directe et ce, sans prendre en compte les détails physiologiques des plis vocaux.

Cette approche, la théorie source-filtre (Fant 1960), consiste à diviser la production d'un son de la parole en deux phases distinctes. Ainsi, le son prendrait sa *source* au niveau glottique (passage de l'air pulmonaire entre les plis vocaux) et les différentes fréquences seraient amplifiées par les organes supraglottiques du conduit vocal qui agirait tel un *filtre*. Pour résumer, la source correspond donc à la vibration des plis vocaux et le filtre dépend des formes et des volumes que peuvent prendre les différentes cavités du tractus vocal (cavité buccale, cavité pharyngale, cavité nasale, cavité labiale). Ainsi, la mise en action des différents articulateurs va modifier le conduit vocal et ses cavités qui vont avoir un rôle d'amplificateurs ou d'amortisseurs de l'onde sonore qui les parcourt. En effet, chaque cavité possède sa propre fréquence de résonance, certaines fréquences du son laryngal vont être amplifiées alors que l'amplitude diminue pour toutes les autres fréquences. Finalement, le son sera modifié à chaque fois que l'un de ces résonateurs sera modifié. La dernière étape à l'élaboration d'un son de la parole se situe au passage entre les lèvres. La pente spectrale au niveau glottique est de  $-12$  dB par octave<sup>8</sup>. Cette pente spectrale est relevée de  $6$  dB par octave lors de cette étape ultime de l'articulation supraglottique, par un phénomène de radiation.

Comme nous venons de le montrer, la production d'un son de la parole met en jeu trois activités distinctes que la figure 1.15 met en évidence. Cette théorie source-filtre peut être traduite sous la forme d'une équation simple où le signal de parole résultant  $S(z)$  est encodé de la manière suivante :  $S(z) = G(z) * F(z) * R(z)$  (où  $G(z)$  représente le bourdonnement laryngal,  $F(z)$  le filtre et  $R(z)$  la radiation au niveau des lèvres) (Airas 2008).



**Figure 1.15** – Représentations des spectres correspondant aux différentes étapes de la théorie source filtre (Airas 2008 :41).

Selon cette théorie, la source et le tractus vocal sont à considérer indépendamment l'un de l'autre. Il existe une interaction minime apparente entre le larynx (source) et les différentes cavités supraglottiques (filtre) (Fant & Lin 1987). Ainsi, les différentes phases nécessaires à l'élaboration d'un son de la parole (source puis filtre et enfin radiation aux lèvres) pourraient être étudiées de manière asynchrone.

## 1.5 Conclusion

Dans ce chapitre, nous avons pu mettre en évidence les différents niveaux nécessaires à la production des sons de la parole (niveau pulmonaire, niveau glottique et

8. D'un point de vue acoustique, une octave peut être définie comme l'intervalle séparant deux sons dont la fréquence fondamentale de l'un vaut le double de la fréquence de l'autre

niveau supraglottique). Après avoir décrit les différents organes qui les composent ainsi que les différents mécanismes mis en jeu, nous avons pu constater la complexité des processus mis en oeuvre. La physiologie du larynx, et en particulier des plis vocaux, reste importante dans l'élaboration du son vocal. Un certain nombre de théories concernant la phonation et la modélisation mécanique des plis vocaux ont été évoquées. Aujourd'hui, la connaissance anatomico-physiologique des plis vocaux plaide en faveur d'un modèle ondulatoire faisant intervenir trois masses (Hirano 1977, entre autres). Ainsi, la phonation dépendrait à la fois de facteurs aérodynamiques, myoélastiques et oscillo-impédantiels. Dès lors, la modification de la structure du pli vocal, par chirurgie par exemple, aurait de lourdes conséquences sur le mécanisme vibratoire et donc *à fortiori* sur la voix et ses paramètres acoustiques.

#### En résumé...

- La phonation nécessite une source d'énergie aérienne provenant des poumons, un transducteur d'énergie — les plis vocaux — qui transforme le flux d'air en impulsions vibratoires, et des cavités supraglottiques, qui servent de résonateurs.
- La connaissance anatomico-physiologique des plis vocaux plaide en faveur d'un modèle à 3 masses combinant un facteur myoélastique, un facteur aérodynamique et un concept oscillo-impédantiel.
- La modification de la structure du pli vocal implique des perturbations des cycles vibratoires.

# Les chirurgies du larynx

"Vox et praeterea nihil."

– Plutarque (Apophthegmata Laconica)

## Résumé

Ce chapitre est une présentation succincte des chirurgies partielles du larynx. Il précise, entre autres, le contexte médical qui a favorisé l'apparition des cordecomies laser par voie endoscopique. Les cancers du larynx peuvent être limités à l'étage glottique ou s'étendre aux structures adjacentes. Face aux cancers, plusieurs traitements sont envisageables, selon la gravité de la situation : la radiothérapie, la chimiothérapie, la chirurgie ou une combinaison des trois. La décision thérapeutique implique une ablation tumorale complète tout en conservant les fonctions principales du larynx, à savoir la respiration, la déglutition et la phonation (Laccourreye 1993). A ce titre, trois grands principes de chirurgies conservatrices laryngées ont émergé : la chirurgie partielle verticale du plan glottique, la chirurgie partielle horizontale supraglottique et la chirurgie partielle horizontale supracricoïdienne. Les premières (cordecomies, laryngectomie frontolatérale, laryngectomie de Tucker et Kambic et hémilaryngectomie) consistent en l'ablation plus ou moins importante d'un hémi-larynx. Les études sur les conséquences sur la phonation rapportent une altération des paramètres phonétiques proportionnelle à l'étendue de l'exérèse. Les laryngectomies supraglottiques ne concernent pas l'étage glottique — les plis vocaux ne sont pas touchés —, ce qui implique une fonction phonatoire intacte. Enfin, les dernières (Crico-hyoïdo-épiglottoplastie CHEP et crico-hyoïdo-pexie CHP) visent la résection d'un étage complet du conduit laryngé par un raccourcissement de celui-ci. De nombreuses structures sont réséquées, y compris les deux plis vocaux, ce qui engendre des voix très altérées dites « de substitution ».

## 2.1 Introduction

Dans le domaine des pathologies vocales, la plus traumatisante des maladies reste le cancer du larynx. Ce dernier, lié principalement au tabagisme et à la consommation d'alcool, est très fréquent en France. Il représente 3500 décès par an, soit 7,5% de la mortalité par cancer chez l'homme entre 40 et 70 ans. Il peut s'agir de cancers limités à l'étage glottique, mais l'envahissement tumoral peut être plus important et s'étendre aux trois étages du larynx puis au pharynx et à la base de la langue. Face à cette pathologie, plusieurs traitements s'offrent aux patients : la radiothérapie, la chimiothérapie et la chirurgie (selon le cas et l'étendue de la tumeur, le chirurgien pourra choisir une laryngectomie totale, une laryngectomie partielle ou une pharyngolaryngectomie), (Le Huche 2001, Brasnu & al. 2005, Brasnu & al. 2008).

Comme nous l'avons évoqué précédemment, le larynx est un organe complexe dont le rôle est primordial dans la respiration, la déglutition et la phonation. La modernité et l'avancement des techniques thérapeutiques permettent actuellement de préserver au mieux les organes ainsi que les fonctions du larynx tout en assurant des résultats oncologiques satisfaisants (Brasnu & al. 2005). La combinaison de ces différentes stratégies, et en particulier le concept de préservation d'organes, se sont rapidement développées afin d'éviter les laryngectomies totales, trop mutilantes tant sur l'aspect communicationnel que sur la qualité de vie des patients.

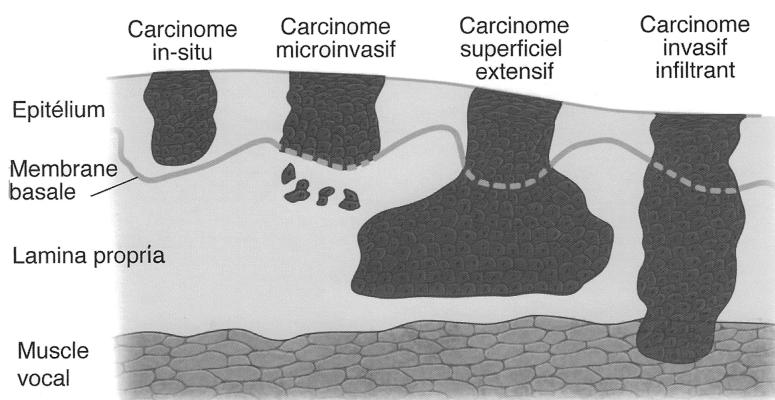
L'objectif de ce chapitre est de présenter les différentes chirurgies partielles du larynx. Ainsi, nous distinguerons deux groupes de chirurgies : les chirurgies partielles verticales et les chirurgies partielles horizontales. Ces dernières, plus mutilantes, consistent en l'exérèse d'un des trois étages (ou plus) du larynx. Dans la première section, nous évoquerons les différents stades d'évolution des cancers du larynx et la classification TNM internationale. Par la suite, nous apporterons une définition du concept de « laryngectomies partielles ». Finalement, nous décrirons très brièvement les principales chirurgies laryngées en les regroupant selon deux groupes distincts : les chirurgies partielles verticales et les chirurgies partielles horizontales.

## 2.2 Les cancers du larynx

### 2.2.1 Infiltration tumorale

Comme nous l'avons abordé dans le chapitre 1, le pli vocal se présente sous une forme stratifiée. De manière générale, nous distinguons différentes couches de la plus superficielle à la plus profonde, le cancer pouvant envahir ces différentes structures (cf. figure 2.1). La couche externe et visible du pli vocal est l'*épithélium*. Le cancer se limitant à cet espace est appelé *carcinome in-situ*. La membrane basale sépare l'*épithélium* de la *lamina propria*. Cette membrane joue le rôle de séparateur entre les carcinomes *in-situ* et les carcinomes les plus invasifs. Lorsque le cancer atteint cette limite, il est nommé *carcinome microinvasif*. Dès lors, la membrane basale est rompue et la couche superficielle de la *lamina* est atteinte. Les carcinomes *in-situ* et *microinvasifs* restent des états précancéreux qui nécessitent une prise en charge précoce.

Comme nous venons de le mentionner, la couche intermédiaire du pli vocal est donc la *lamina propria*. Celle-ci se divise en trois sous-couches : superficielle, moyenne (espace de Reinke) et profonde. L'envahissement tumoral de la *lamina propria* est dénommé *carcinome superficiel extensif*. A ce stade d'envahissement tumoral, le muscle thyro-aryténoïdien est encore sain. Enfin, l'espace le plus profond du pli vocal est constitué principalement du ligament vocal et du muscle vocal, à savoir le muscle thyro-aryténoïdien. Lorsque le cancer touche cette partie du pli vocal, le cancer est dit *carcinome invasif infiltrant*. Ces deux dernières tumeurs représentent des cancers de stade avancé qui doivent alerter les médecins (Brasnu & al. 2005, Hans & al. 2006, Brasnu & al. 2008).



**Figure 2.1** – Les différentes infiltrations tumorales (d'après Brasnu & al. 2005:151).

## 2.2.2 La classification TNM Internationale

Les cancers du plan glottique représentent la moitié des cancers des voies aéro-digestives supérieures et le pronostic d'ensemble semble plutôt positif. Depuis 1987, la classification TNM (où T= Tumeur, N= adénopathies et M= Métastases) permet de classifier les tumeurs selon leur taille, leur localisation (infiltration tumorale) et leur extension. Par conséquent, le chirurgien peut alors prévoir le traitement le plus adapté à la pathologie du patient.

Les tumeurs du plan glottique sont classées ainsi :

<b>T1</b>	Tumeur limitée à un pli vocal (T1a) ou aux deux plis vocaux (T1b) avec une mobilité normale. L'envahissement de la commissure antérieure ou de la commissure postérieure est possible.
<b>T2</b>	Tumeur limitée au larynx avec atteinte des régions sus ou sous-glottique et/ou réduction de la mobilité du pli vocal touché.
<b>T3</b>	Tumeur limitée au larynx avec fixation du pli vocal.
<b>T4</b>	Tumeur étendue aux structures extra-laryngées (cartilage thyroïde ou tissus extralaryngés, à savoir trachée, muscles profonds du cou, muscles extrinsèques de la langue, muscles sous-hyoidiens, glande thyroïde ou œsophage).

**Table 2.1 – Classification des tumeurs (T) du plan glottique.**

Les atteintes ganglionnaires sont, quant à elles, classées de la manière suivante :

<b>N0</b>	Pas d'envahissement ganglionnaire.
<b>N1</b>	Métastase dans un ganglion homolatéral de moins de 3 cm.
<b>N2</b>	Métastase dans un ganglion homolatéral de plus de 3 cm mais ne dépassant pas 6 cm (N2a) ; ou métastases aux ganglions multiples homolatéraux (moins de 6 cm) mais ne dépassant pas 6 cm (N2b) ; ou métastases aux ganglions bilatéraux inférieures à 6 cm (N2c).
<b>N3</b>	Métastase dans un ganglion de plus de 6 cm.

**Table 2.2 – Classification des atteintes ganglionnaires (N).**

Enfin, les métastases à distance sont classées ainsi :

<b>M0</b>	Pas de métastase.
<b>M1</b>	Une ou plusieurs métastases à distance.

**Table 2.3 – Classification des métastases à distance (M).**

## 2.3 Préliminaires

### 2.3.1 Objectifs

Comme nous l'avons mentionné plus haut, les cancers du larynx se traitent par chimiothérapie, radiothérapie mais aussi par chirurgie ; ce que l'on appelle « laryngectomie ». Etymologiquement, le terme de « laryngectomie », pris dans sa globalité,

signifie l’ablation du larynx. Portmann & Guerrier (1987) définissent cet acte chirurgical comme « l’exérèse partielle ou totale du larynx où siège une tumeur bénigne ou maligne ».

L’avancée des techniques chirurgicales et les connaissances en termes de physiologie laryngée et de règles oncologiques ont permis une généralisation des chirurgies partielles du larynx. Quand cela reste possible, ces interventions, moins mutilantes sur les plans communicationnel, physiologique et social que les laryngectomies totales, sont préférées à ces dernières. Son objectif principal est d’assurer l’ablation tumorale totale tout en conservant les fonctions du larynx, à savoir la respiration, la déglutition et la phonation (Laccourreye 1993). A cette fin, ces chirurgies tendent à préserver au mieux les organes. En effet, ces trois fonctions nécessitent la conservation de la dynamique du néolarynx, ce qui passe par la préservation de deux structures indispensables : le cartilage cricoïde et l’articulation crico-aryténoidienne (Arnoux-Sindt 1992). Comme nous l’avons décrit dans le chapitre précédent, cette articulation constitue le seul vibrateur restant de l’endolarynx (cf. section 1.3.3.2). Finalement, Brasnu & al. (2005) qualifient d’ailleurs ces interventions de «chirurgies conservatrices».

Pourtant, historiquement, les traitements des cancers de stades précoces et avancés du larynx restent controversés. Dans les années 1950, les traitements des cancers avancés T3-T4 se limitaient essentiellement à la radiothérapie et aux laryngectomies totales. Dès les années 1980, le développement de la chimiothérapie changea la donne. Aujourd’hui, les indications varient en fonction de l’équipe médicale mobilisée et des zones géographiques. Finalement, quel que soit le protocole de préservation d’organe sélectionné, le but majeur est d’éviter l’échec total. Brasnu & al (2008) expliquent que les cancers T3-T4 du larynx peuvent être traités soit par radiothérapie avec chimiothérapie associée, soit par une chirurgie laryngée en fonction de la réponse à la chimiothérapie d’induction. En fonction de la réponse au traitement par chimiothérapie, il existe trois possibilités : (i) le patient n’est pas un « bon répondeur », alors une laryngectomie totale est décidée (20% des cas), (ii) le patient est « en réponse clinique complète » (réponse tumorale de 100%), alors une radiothérapie avec chimiothérapie associée est proposée et (iii) le patient est un « bon répondeur » (réponse supérieure à 50% mais inférieure à 100%). Dans ce dernier cas, le traitement proposé est soit une chimio-radiothérapie concomitante, soit une chirurgie totale, soit, dans certains cas, une chirurgie partielle laryngée. En ce qui concerne les cancers de stades précoces T1-T2 et certains T3 sélectionnés, le traitement de référence fut pendant de nombreuses années en Europe du Nord et aux Etats-Unis, la radiothérapie avec une solution de rattrapage par laryngectomie totale en cas d’échec. Dès les années 1950, en France et dans les pays de l’Europe du sud, le développement des chirurgies partielles laryngées a été à l’origine de stratégies thérapeutiques différentes. Brasnu & al. (2008) précisent qu’actuellement les cancers précoces du plan glottique peuvent être traités soit par radiothérapie, soit par chirurgie endoscopique au Laser CO<sub>2</sub>, soit par laryngectomie partielle verticale. Pour les cancers supraglottiques classés T1-T2, le traitement proposé est une laryngectomie partielle de type supraglottique.

La décision d’une chirurgie partielle laryngée nécessite une étude du praticien, des examens préopératoires suivis et une technique chirurgicale. Le choix de l’inter-

vention dépend largement de la connaissance de la tumeur et de ses extensions. Des examens préopératoires sont indispensables pour juger de l'importance et de l'étendue du cancer. Ce choix de chirurgie nécessite également une connaissance approfondie de la physiologie laryngée et des techniques chirurgicales puisque le chirurgien doit pouvoir adapter son traitement au diagnostic établi, maîtriser son geste et être capable d'adapter et/ou modifier la chirurgie en fonction des extensions tumorales. Le suivi postopératoire doit évidemment être pris en compte dans le choix du traitement (Hans 2003).

### 2.3.2 Historique des chirurgies

Le concept de laryngectomie partielle a émergé au XIXème siècle. Dès 1833, Bowes réalisa la première cordectomie par thyrotomie. Plusieurs chirurgiens tels que Braouers de Louvain en 1834 ou Solis Cohen en 1860, entre autres, s'essayèrent à cette technique (Laccourreye 2000). Cependant, face aux problèmes postopératoires et au taux de mortalité, la laryngectomie totale se développa. C'est en 1873 que Billroth décrivit la première exérèse totale du larynx.

C'est seulement en 1896 que la laryngectomie partielle retrouva grâce aux yeux des chirurgiens contemporains. A la fin du XIXème siècle, trois chirurgies principales étaient donc pratiquées : la cordectomie par thyrotomie, l'hémilaryngectomie totale et la laryngectomie totale. Néanmoins, les résultats fonctionnels étaient toujours décevants et le taux de mortalité toujours élevé.

Au cours du XXème siècle, le développement de l'anesthésie, la diffusion des antibiotiques et l'avancée des connaissances — que ce soit sur les chirurgies, la réanimation, la physiologie ou sur le matériel — ont permis d'affiner les prises en charge. Ainsi, trois périodes se sont succédées, mettant en lumière trois grands principes de chirurgies conservatrices laryngées : la chirurgie partielle verticale du plan glottique, la chirurgie partielle horizontale supraglottique et la chirurgie partielle horizontale supracricoïdienne.

## 2.4 La chirurgie partielle verticale du plan glottique

La chirurgie partielle verticale (laryngectomie partielle verticale ou LPV) a été développée en Europe et aux Etats-Unis dès le début du XXème siècle. Elle concerne les cancers dont l'étendue est encore limitée à la région du plan glottique ou sous-glottique avec une mobilité normale du pli vocal qui doit être opéré. Elle repose également sur l'indépendance des deux plis vocaux. C'est un traitement défini après un diagnostic T1 ou certains T2 sélectionnés. Elle consiste en l'ablation, plus ou moins importante, d'un hémilarynx. Dans cette section, nous décrirons quelques unes des principales laryngectomies verticales en les classant selon l'importance de l'exérèse (de la moins importante à la plus importante).

## 2.4.1 La cordectomie

Cette chirurgie est l'une des interventions les plus anciennes. Comme nous l'avons brièvement décrit précédemment, elle fut pratiquée pour la première fois par un chirurgien nommé Bowes en 1833. La cordectomie par voie externe est réservée aux patients dont la tumeur du plan glottique est classée T1a et limitée au tiers moyen du pli vocal encore mobile. La commissure antérieure doit être totalement saine pour ce type d'intervention.

### 2.4.1.1 Principe chirurgical

Cette chirurgie consiste en l'exérèse du pli vocal cancéreux, depuis la commissure antérieure jusqu'au processus vocal de l'aryténoïde (figure 2.2). L'ablation du pli vocal tumoral crée une fuite glottique importante, entraînant des troubles vocaux variables. Pour pallier à ce problème et garantir aux patients une voix efficace, les équipes chirurgicales réalisent souvent une reconstruction du pli vocal reséqué par « plastie des bandes ventriculaires ». Cette reconstruction consiste en l'abaissement de la bande ventriculaire homolatérale au niveau de la muqueuse sous-glottique (Burgess 1993). Cette néocorde permet une reprise immédiate de la phonation et une amélioration certaine de cette dernière (Biacabe & al. 1998, 1999, 2001).

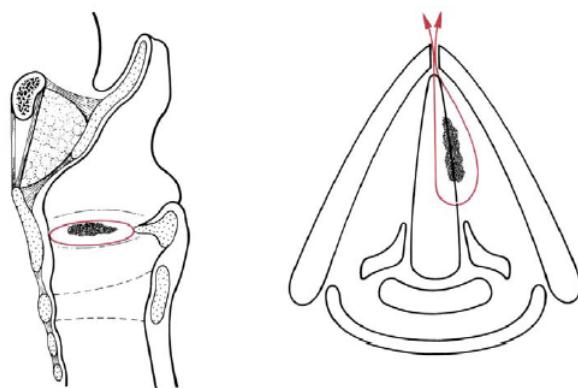


Figure 2.2 – Cordectomie (d'après Lefebvre & Chevalier 2005).

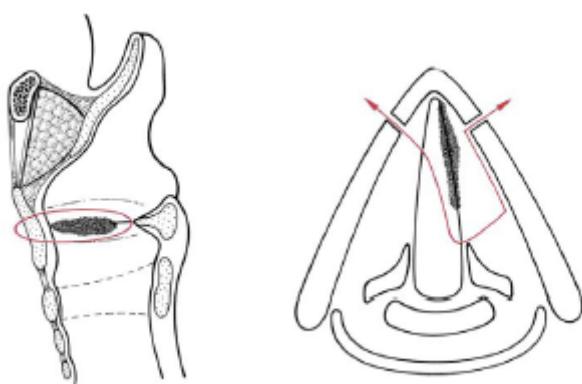
## 2.4.2 La laryngectomie frontolatérale

Cette chirurgie fut proposée par Leroux-Robert en 1956. Elle est indiquée pour les cancers de type T1a ou certaines tumeurs classées T2, sans atteinte de la commissure antérieure avec une mobilité cordale normale.

### 2.4.2.1 Principe chirurgical

L'intervention consiste en l'ablation d'un pli vocal (muqueuse, ligament vocal et muscle TA), de la commissure antérieure, de la partie antérieure du cartilage thyroïde

adjacent et d'une petite portion du pli vocal opposé. L'envahissement sous-glottique doit être néanmoins limité et la bande ventriculaire intacte. Comme pour les cordeectomies par voie externe, les chirurgiens peuvent compléter la chirurgie par une reconstruction du pli vocal reséqué grâce à un abaissement de la bande ventriculaire au niveau de la muqueuse sous-glottique afin d'améliorer la fonction vocale (cf. figure 2.3).



**Figure 2.3 – Laryngectomy frontolatérale (d'après Lefebvre & Chevalier 2005).**

#### 2.4.2.2 Conséquences sur la phonation

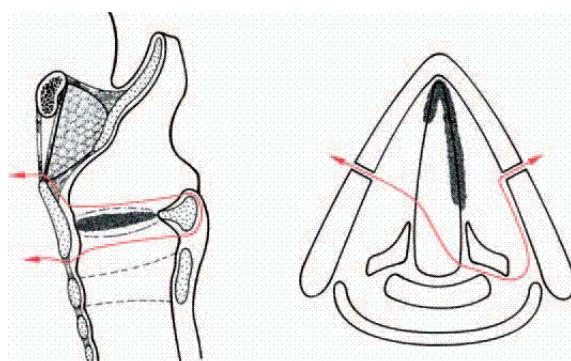
A l'instar de l'ensemble des laryngectomies verticales, peu d'études traitent des résultats fonctionnels après laryngectomy fronto-latérale. L'intervention réalisée modifie le vibrateur et le conduit vocal. Elle réalise également un raccourcissement de la glotte, ce qui est propice à la production de sons aigus. La néoglotte conserve sa forme en V, mais l'intervention crée toutefois une asymétrie entre le pli vocal sain et la partie reséquée, ce qui constitue une explication aux fuites glottiques. Les recherches de Biacabe & al. (1998, 1999, 2001) montrent des paramètres acoustiques (jitter, shimmer, rapport signal/bruit, degré de dévoisement) et de parole (TMP, nombre de mots lus/minute) significativement altérés par rapport à une population témoin. Les auteurs notent cependant une amélioration des paramètres fréquentiels pendant les premiers mois post-opératoires pour l'ensemble des patients ainsi qu'une différence significative pour le degré de dévoisement et le HNR entre le groupe des patients ayant subi une reconstruction glottique et le groupe de patients sans reconstruction. Enfin, les auteurs observent une fermeture glottique statistiquement meilleure pour le groupe des patients ayant subi une reconstruction par rapport aux patients ayant subi uniquement la laryngectomy fronto-latérale. En outre, Wallet (2007) a démontré que d'un point de vue intonatif, ces patients étaient capables de reproduire le contraste entre phrases interrogatives et déclaratives. La réalisation de la focalisation est également comparable au groupe des témoins. Après une analyse perceptive de l'ensemble de ces phrases modales, Wallet (2008, 2009) conclut à une bonne intelligibilité de ce groupe de patients en situation communicationnelle.

### 2.4.3 La laryngectomie frontale antérieure avec épiglottoplastie (dite de Tucker et Kambic)

Décrise dès les années 1960, cette chirurgie fut proposée pour la première fois par Tucker et Kambic en 1979 aux Etats-Unis et par Pech en 1982 en France. Elle est indiquée dans les tumeurs du plan glottique bilatérales avec atteinte des deux plis vocaux (T1b), sans altération de la mobilité du pli vocal ou de l'aryténioïde (Giovanni & al. 2001).

#### 2.4.3.1 Principe chirurgical

L'intervention consiste en l'exérèse des deux plis vocaux, et parfois du cartilage arytenoïde du côté tumoral, des deux bandes ventriculaires et le tiers ou les deux tiers antérieurs du cartilage thyroïde. La reconstruction de la commissure antérieure est réalisée par l'abaissement de l'épiglotte que le praticien rattache au cartilage cricoïde. Ainsi, celle-ci prend la place du cartilage thyroïde réséqué (cf. figure 2.4).



**Figure 2.4 – Laryngectomie frontale antérieure avec épiglottoplastie (d'après Le-febvre & Chevalier 2005).**

#### 2.4.3.2 Conséquences sur la phonation

Selon Pech & al. (1988), le mécanisme phonatoire semblerait se rapprocher de celui des laryngectomies partielles supracricoïdiennes. La phonation résulte de la mise en place de mouvements complexes de substitution combinant les structures conservées. Finalement, ces derniers mettent en évidence trois façons de produire un son. Premièrement, par un affrontement sagittal des deux arytenoïdes ou de l'arytenoïde restant contre un repli muqueux latéral. Deuxièmement, par un affrontement frontal de l'arytenoïde contre la face postérieure de l'épiglotte. Enfin, par un affrontement mixte combinant les deux premiers mécanismes compensatoires.

Les études de Giovanni & al. (1996, 2002) montrent des paramètres aérodynamiques (TMP et débit d'air phonatoire) et acoustiques ( $F_0$ , jitter, HNR) perturbés. Ainsi, en 2002, ces auteurs comparèrent 21 patients traités par laryngectomie de Tucker avec 10 sujets témoins contrôles sur la portion la plus stable d'un [a] tenu. Le

constat est que le jitter et le shimmer restent statistiquement plus élevés par rapport aux témoins. De même le débit d'air oral et la pression sous-glottique estimée (PSGE) sont supérieurs par rapport aux sujets contrôles. Les auteurs expliquent que l'élévation de la PSGE serait une compensation à l'inefficacité glottique et que cette hausse permettrait de mieux initier la vibration.

#### 2.4.4 L'hémilaryngectomie

Pratiquée dans les années 1920, l'hémilaryngectomie est indiquée dans les cancers du pli vocal mobile, descendant vers la région sous-glottique moyenne. Cette intervention permet la résection de la totalité de l'hémilarynx et de la margelle laryngée homolatérale.

##### 2.4.4.1 Principe chirurgical

Cette intervention consiste en l'ablation de la moitié du cartilage thyroïde, d'une bande ventriculaire, d'un pli vocal et d'un arytenoïde. Nous noterons donc la création d'une néoparoï latérale et d'une néoglotte, formée via l'accolement de l'hémilarynx restant à cette nouvelle néoparoï pharyngée. Cette chirurgie ne semble toutefois plus pratiquée de nos jours (cf. figure 2.5).

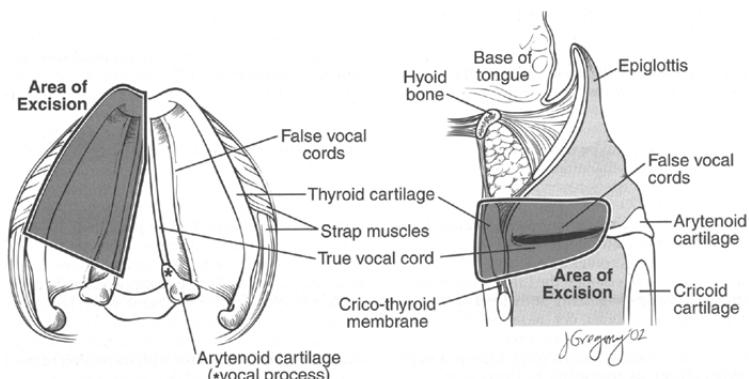


Figure 2.5 – Hémilaryngectomie (d'après [www.msdlatinamerica.com](http://www.msdlatinamerica.com)).

##### 2.4.4.2 Conséquences sur la phonation

La phonation résulte de l'accolement du pli vocal restant contre le bourrelet cicatriciel qui se forme de l'autre côté. Une néoglotte est créée permettant l'établissement d'une voix de substitution. A notre connaissance, aucune étude ne fait état des résultats fonctionnels après ce type d'intervention.

## 2.5 La chirurgie partielle horizontale

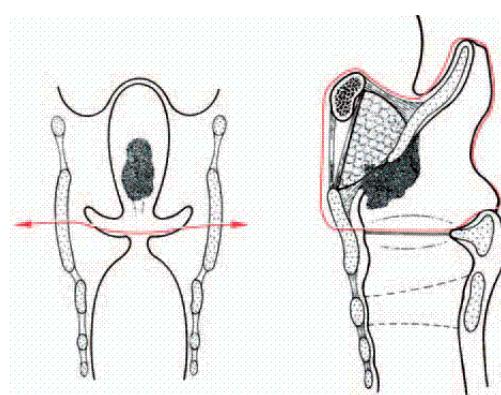
La chirurgie partielle horizontale (laryngectomie partielle horizontale ou LPH) s'adresse aux patients atteints d'un cancer de stade plus avancé. Elle consiste en l'exérèse d'un des trois étages (ou plus) du larynx ; à savoir la région supraglottique (composée des ventricules de Morgani, des bandes ventriculaires, de l'épiglotte, des cartilages arytenoïdes), la région glottique (composée des plis vocaux et de la commissure antérieure) ou la région sous-glottique (qui s'étend de la face inférieure des plis vocaux au bord supérieur du cartilage cricoïde). Dans cette section, nous décrirons très brièvement les techniques chirurgicales des laryngectomies partielles supraglottiques et supracricoïdiennes.

### 2.5.1 Les laryngectomies partielles supraglottiques

Les laryngectomies partielles supraglottiques sont indiquées dans les tumeurs de la région supraglottique. Selon l'étendue du cancer limité au larynx supraglottique ou atteignant le pharynx, le traitement chirurgical sera différent.

#### 2.5.1.1 Principe chirurgical

La laryngectomie partielle supraglottique limitée au larynx, à savoir l'épiglotte, la margelle et la bande ventriculaire, consiste en l'exérèse du tiers supérieur du cartilage thyroïde, des deux tiers antérieurs des bandes ventriculaires et de l'épiglotte. Ainsi, les deux cartilages arytenoïdes et les deux plis vocaux restent intacts. Dans le cas où l'extension tumorale se propagerait à la base de la langue et/ou au pharynx, la laryngectomie supraglottique devrait être élargie. Le chirurgien devrait alors pratiquer une laryngopharyngectomie supraglottique (cf. figure 2.6).



**Figure 2.6 –** Laryngectomie partielle supraglottique (d'après Lefebvre & Chevalier 2005).

### 2.5.1.2 Conséquences sur la phonation

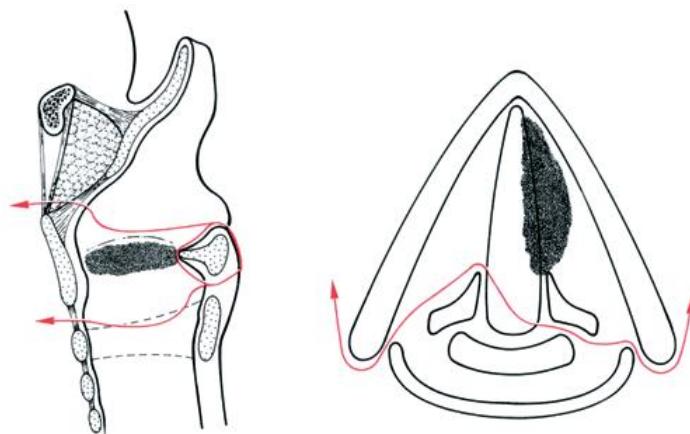
Dans cette chirurgie, le plan glottique est totalement épargné. Ainsi, les deux plis vocaux sont préservés, ce qui permet un maintien de la fonction vocale.

## 2.5.2 Les laryngectomies partielles supracricoïdiennes (LPSC)

La LPSC a été introduite en 1959 par Majer & Reider et, en 1971, par Labayle. Ces chirurgies visent la résection d'un étage complet du conduit laryngé par un raccourcissement de celui-ci. Cette résection doit cependant respecter le cartilage cricoïde et au moins un cartilage arytenoïde. Nous décrirons les deux principales interventions de ce type : la CHEP où l'épiglotte est conservée et la CHP où l'épiglotte est supprimée.

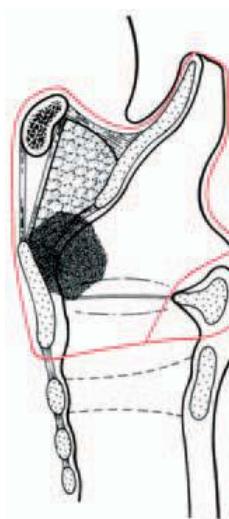
### 2.5.2.1 Principes chirurgicaux

**La laryngectomie supracricoïdienne avec crico-hyoïdo-épiglottopexie (CHEP).** Cette intervention proposée par Majer en 1959 et généralisée en France par Piquet en 1974 est, dans la plupart des cas, réalisée dans les cancers du plan glottique d'indication difficile (T2/T3). Il s'agit de tumeurs du pli vocal avec envahissement de la commissure antérieure et/ou du ventricule, du cartilage thyroïde, dans lesquelles une diminution de la mobilité du pli vocal est constatée voire une fixité totale de ce pli vocal. Le cartilage arytenoïde doit, toutefois, rester mobile. Elle consiste en l'ablation totale du cartilage thyroïde, des deux plis vocaux, des deux plis ventriculaires, avec conservation d'un ou des deux cartilages arytenoïdes. La particularité de cette intervention réside dans la conservation d'une partie de l'épiglotte (cf. figure 2.7). Du point de vue de la suture, la trachée est remontée du fait que le cricoïde est rattaché à l'os hyoïde et à la portion de l'épiglotte restante. Cette reconstruction permet la fixation de l'épiglotte pour éviter sa bascule en arrière. La phonation réside dans la mise en vibration d'une néoglotte. Néanmoins, la qualité de la phonation peut être très altérée à cause de l'affrontement frontal ou latéral de l'arytenoïde conservé, avec la face postérieure de l'épiglotte.



**Figure 2.7 – Laryngectomie supracricoidienne avec crico-hyoïdo-epiglotto-pexie** (d'après Lefebvre & Chevalier 2005).

**La laryngectomie supracricoidienne avec crico-hyoïdo-pexie (CHP).** Cette chirurgie, réalisée par Labayle dès 1970, concerne les cancers du vestibule laryngé trop importants pour bénéficier d'une laryngectomie horizontale supraglottique. Elle est une alternative à la laryngectomie totale dans des cas particuliers de cancers supraglottiques et transglottiques de type T3 avec atteinte du ventricule, du plan glottique et/ou de la mobilité du pli vocal. Cette intervention, proche de la CHEP, requiert l'ablation totale de l'épiglotte (cf. figure 2.8). Selon l'infiltration tumorale, un cartilage arytenoïde peut être réséqué. Au niveau de la reconstruction, nous pouvons constater la remontée de la trachée par rattachement du cricoïde à l'os hyoïde. Celle-ci a pour volonté de permettre l'affrontement de la muqueuse de la langue avec celle du cartilage arytenoïde afin de conserver une fonction phonatoire.



**Figure 2.8 – Laryngectomie supracricoidienne avec crico-hyoïdo-pexie** (d'après Lefebvre & Chevalier 2005).

### 2.5.2.2 Conséquences sur la phonation

Les études acoustiques et perceptives de la voix après LPSC sont peu nombreuses. Sur le plan perceptif, les auteurs font état d'une voix rauque, soufflée et serrée. D'un point de vue acoustique, ces recherches démontrent une forte perturbation de l'ensemble des paramètres, à savoir la  $F_0$ , le jitter, le shimmer ou le HNR, entre autres (Pech & al. 1988, Laccourreye & al. 1995, Crevier Buchman & al. 1995, Crevier Buchman & al. 1998, Crevier Buchman & al. 2001, Crevier Buchman & al. 2002, Crevier Buchman & al. 2003, Crevier Buchman 1999, Makeieff & al. 2005, 2007). Les travaux les plus complets concernant la voix et la parole après LPSC-CHEP sont ceux réalisés par Crevier Buchman. Dans sa thèse de 1999 portant sur une étude longitudinale de 20 patients traités par CHEP, celle-ci met en évidence les points suivants : sur le plan perceptif, l'échelle GRBAS n'est pas adaptée à l'analyse de ce type de voix de substitution. Le voisement, la sonie, la hauteur, la mélodie et le débit seraient des paramètres pertinents à prendre en compte. Sur le plan acoustique, la voix après LPSC-CHEP reste très instable tant en fréquence qu'en amplitude. L'auteur met également l'accent sur les problèmes de détection de la  $F_0$  par les différents logiciels d'analyse vocale. Sur le plan articulatoire, Crevier Buchman précise que ces patients privilient le lieu d'articulation des consonnes au détriment du trait de voisement. Ainsi, un dévoisement des consonnes voisées au profit de leurs correspondantes non voisées (même lieu et mode d'articulation) est réalisé. Les confusions de lieu concernent essentiellement les consonnes palato-vélaires au profit des consonnes alvéolaires (par exemple, la consonne [k] est perçue [t]). L'auteur ne note toutefois aucune confusion sur le mode d'articulation (occlusive vs. fricative) et sur le trait de nasalité. Sur le plan aérodynamique, le seul paramètre pris en compte est le temps maximal de phonation (TMP). Celui ci semble être fortement perturbé en période post-opératoire. En 2003, dans sa thèse portant sur l'étude de quelques paramètres aérodynamiques, Hans a mis en lumière que la pression intra-orale (PIO) de ces patients était statistiquement plus élevée que celle de patients traités par cordectomie ou laryngectomie fronto-latérale. Makeieff & al. (2005) établissent une corrélation entre des mesures acoustiques/aérodynamiques et des notes perceptives chez 61 patients traités par LPSC-CHEP. Ceux-ci concluent que la majorité des patients étaient classés G3 ou G4, indice d'une dysphonie sévère. En outre, les auteurs précisent que les paramètres du jitter, du shimmer, de la PSGE, du HNR et du débit d'air oral sont corrélés aux différents grades perceptifs de dysphonies. Dans une étude de 2007 sur 25 patients traités par LPSC-CHP, Makeieff & al. insistent principalement sur l'intérêt de coupler aux analyses acoustiques des analyses morphologiques et laryngoscopiques. Ce type d'analyse permet de visualiser les mouvements ainsi que les occlusions de la néoglotte. Ainsi, les phénomènes de compensation peuvent être mis en lumière et l'image vidéolaryngoscopique peut servir d'explication aux altérations des divers paramètres acoustiques (Pech & al. 1988, Arnoux Sindt 1992, Crevier Buchman 1999). De plus, d'un point de vue prosodique, Wallet (2007) a démontré que les patients traités par LPSC-CHEP/CHP avaient des modulations aléatoires dans la production de phrases interrogatives et déclaratives du fait des problèmes de détection de  $F_0$  par rapport à un groupe témoin. D'un point de

vue acoustique, la focalisation n'a jamais été réalisée correctement par ce groupe de patients. L'auteur relève également une hausse de l'intensité par rapport à la cohorte des sujets contrôles, signe d'une compensation éventuelle des irrégularités de  $F_0$ .

## 2.6 Conclusion

Dans ce chapitre, nous avons pu mettre en évidence et décrire quelques unes des chirurgies partielles laryngées disponibles dans le traitement des cancers du larynx. La taille, la localisation et l'extension de la tumeur sont des indices dans le choix de la meilleure chirurgie à adopter. Néanmoins, aujourd'hui, le concept de préservation d'organes reste une priorité pour les praticiens afin de conserver intactes les trois principales fonctions laryngées : la respiration, la déglutition et la phonation. En conséquence, nous avons vu que les chirurgies partielles étaient de bonnes alternatives aux laryngectomies totales. Les laryngectomies partielles se divisent en deux groupes distincts : les chirurgies verticales (qui concernent un hémilarynx) et les chirurgies horizontales (qui concernent un étage ou plus du larynx). Chacune de ces interventions doit être choisie selon des principes et indications particulières et requiert une technique chirurgicale complexe. Du point de vue des conséquences sur la voix, nous avons pu noter que plus la résection était importante et plus les caractéristiques de la voix (acoustiques, perceptifs, aérodynamiques et prosodiques) étaient perturbées. Les paramètres de la voix et de la parole après laryngectomie partielle verticale (en particulier la cordectomie par voie externe et la laryngectomie frontolatérale), bien que perturbés par rapport à des populations témoins, restent bien meilleurs que leurs correspondantes après laryngectomie horizontale (en particulier les laryngectomies partielles supracricoïdiennes). Mais qu'en est il réellement pour le type de chirurgie dont il est question ici, la cordectomie par voie endoscopique au laser CO<sub>2</sub> ?

### En résumé...

- L'ablation tumorale complète associée au principe de conservation d'organes sont au cœur des préoccupations des chirurgiens.
- Nous relevons trois types de chirurgies conservatrices laryngées : la chirurgie partielle verticale du plan glottique, la chirurgie partielle horizontale supraglottique et la chirurgie partielle horizontale supracricoïdienne.
- Les études sur les résultats fonctionnels montrent une altération des paramètres phonétiques proportionnelle à l'étendue de l'exérèse.



# La cordectomie laser : approche multiparamétrique

"Ce qui est dans la parole est dans le silence."  
– Proverbe berbère

## Résumé

Ce chapitre est une description des différentes cordectomies laser. Nous présentons une revue de la littérature qui met en évidence les résultats fonctionnels après chirurgie, et surtout qui insiste sur la dimension multiparamétrique des approches scientifiques. C'est à partir des années 70 que la cordectomie laser fut introduite pour les tumeurs limitées au pli vocal (Strong 1975, Brasnu & al. 2005). Celle-ci connut un essor du fait de ses nombreux avantages : diminution de la durée d'hospitalisation, de la morbidité et des coûts médico-économiques par rapport aux chirurgies par voie externe. Nous présentons cinq types de cordectomies laser qui sont classées en fonction de l'étendue de l'exérèse. La plus simple emporte uniquement l'épithélium et la plus complexe va jusqu'à la résection du pli vocal dans son ensemble ainsi que certaines structures adjacentes. Le développement récent de ces chirurgies a vu émerger un certain nombre de recherches portant sur la qualité vocale en postopératoire. Après avoir évoqué ces différents travaux, nous avons souligné que l'hétérogénéité des populations étudiées, la diversité des seuils de normalité adoptés ou la différence entre les logiciels utilisés, ainsi que la divergence des méthodologies (matériels de prises de données et corpus constitués) rendent les comparaisons transversales difficiles. Néanmoins, les différents auteurs s'accordent quant à l'importance d'effectuer des analyses multiparamétriques afin de mieux comprendre le comportement laryngé de ces patients : analyses perceptives, analyses acoustiques, mesures aérodynamiques et morphologiques.

### 3.1 Introduction

Le concept d'exérèse des tumeurs laryngées par voie endoscopique est ancien. Avant même l'apparition du laser, les chirurgiens pratiquaient ce type d'intervention. C'est à partir des années 70 que la cordectomie par voie endoscopique au laser CO2 fut introduite pour les tumeurs limitées au pli vocal (Strong 1975, Brasnu & al. 2005). Actuellement, cette chirurgie connaît un essor du fait de ses nombreux avantages : diminution de la durée d'hospitalisation, de la morbidité et des coûts médico-économiques. En outre, de récents travaux rapportant des résultats oncologiques et vocaux positifs plaident en faveur de l'exérèse par voie endoscopique au laser. Cette chirurgie attestée pour les états précancéreux, les carcinomes *in-situ* et les tumeurs T1 (cf. classification TNM exposée dans le chapitre 2) se développe de plus en plus pour certaines tumeurs classées T2 ou T3. Cette technique permet une récupération fonctionnelle rapide mais les résultats sur la phonation dépendent largement de l'étendue de l'exérèse. Le but principal de ce chapitre est, d'une part, de décrire les principes chirurgicaux correspondant aux différentes cordectomies laser et d'autre part, de faire une revue de la littérature détaillée des divers articles impliquant les résultats fonctionnels sur la voix. Finalement, nous ferons un rappel des principales mesures acoustiques, perceptives, aérodynamiques et morphologiques utilisées dans les études présentées.

### 3.2 Principes chirurgicaux

De manière générale, nous distinguons plusieurs types de cordectomies par voie endoscopique au laser CO2, la plus simple emportant uniquement l'*épithélium* et la plus complexe reséquant l'ensemble du pli vocal. La Société Européenne de Laryngologie propose en 2000 une classification de ces différentes interventions que nous exposons ci-dessous (Remacle & al. 2000) :

#### 3.2.1 La cordectomie de type I (ou sous-épithéliale)

Cette intervention est indiquée dans les états précancéreux et les carcinomes *in-situ* limités à l'*épithélium* du pli vocal. Elle consiste en l'exérèse complète de l'*épithélium* et la partie superficielle de la *lamina propria*. Cette chirurgie permet un traitement correct de la lésion tout en préservant le ligament vocal (cf. figure 3.1).

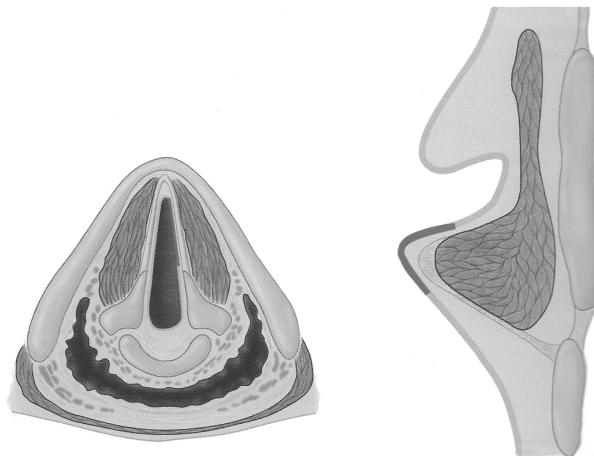


Figure 3.1 – Cordectomie Laser de type I (d'après Brasnu & al. 2005:152).

### 3.2.2 La cordectomie de type II (ou sous-ligamentaire)

Cette intervention est indiquée pour des carcinomes *in-situ* ou micro-invasifs de la *lamina propria*. La résection englobe l'*épithélium*, le ligament vocal ainsi que les trois couches de la *lamina propria* (cf. figure 3.2).

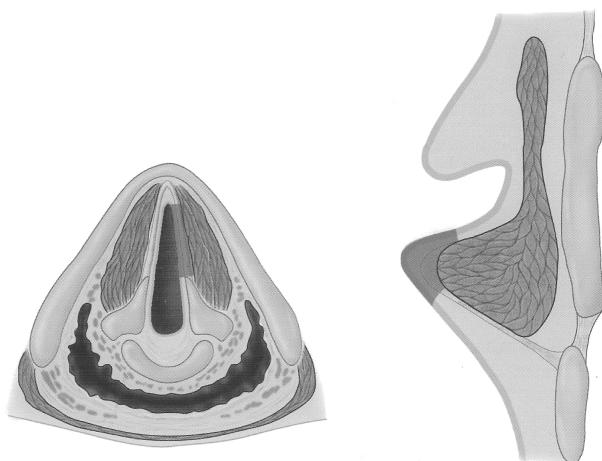
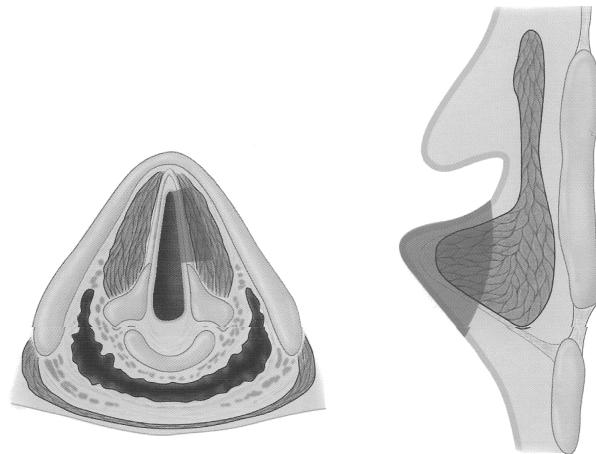


Figure 3.2 – Cordectomie Laser de type II (d'après Brasnu & al. 2005:153).

### 3.2.3 La cordectomie de type III (ou transmusculaire)

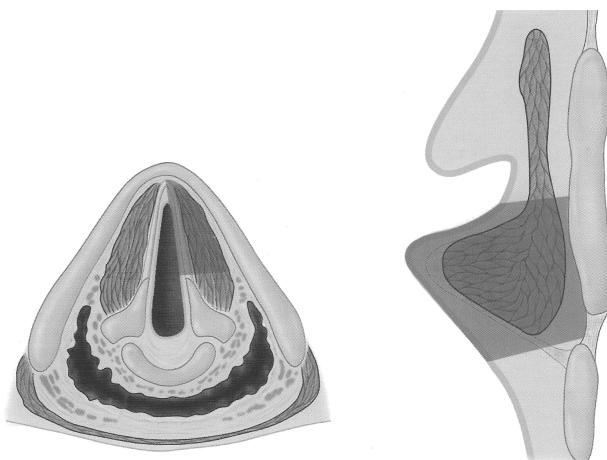
Cette chirurgie est indiquée pour les lésions de stade T1a du tiers moyen du pli vocal avec une mobilité cordale normale. Elle correspond à l'exérèse de l'*épithélium*, de la *lamina propria* et d'une partie du muscle vocal (cf. figure 3.3). Par ailleurs, l'intervention peut être étendue du processus vocal de l'aryténioïde à la commissure antérieure.



**Figure 3.3 – Cordectomie Laser de type III (d'après Brasnu & al. 2005:154).**

### 3.2.4 La cordectomie de type IV (ou totale)

Cette cordectomie est indiquée pour les lésions de stade T1a du pli vocal avec infiltration au muscle vocal et les lésions T2 avec diminution de la mobilité cordale. La tumeur peut atteindre la commissure antérieure, sans néanmoins l'envahir. Cette chirurgie correspond à l'exérèse complète du pli vocal, de la commissure antérieure jusqu'à l'apophyse vocale de l'aryténoïde. Elle emporte l'épithélium, la lamina propria, l'ensemble du muscle vocal TA et, éventuellement, une partie du cartilage thyroïde (cf. figure 3.4).



**Figure 3.4 – Cordectomie Laser de type IV (d'après Brasnu & al. 2005:155).**

### 3.2.5 Cordectomie de type V (ou étendue)

Cette chirurgie correspond à une cordectomie de type IV, c'est à dire à l'exérèse du pli vocal complet à laquelle s'ajoute quatre extensions selon l'étendue de la tumeur aux structures adjacentes :

### 3.2.5.1 Cordectomie de type Va

Cette chirurgie emporte la commissure antérieure, voire une partie plus ou moins importante du pli vocal opposé (cf. figure 3.5). Elle est indiquée pour les lésions infiltrant la commissure antérieure ou pour des cancers de type T1b et quelques T2 sélectionnés. La mobilité de l'aryténoïde doit toutefois être conservée.

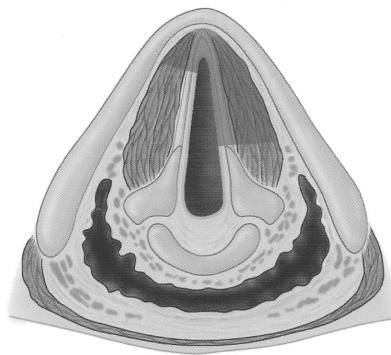


Figure 3.5 – Cordectomie Laser de type Va (d'après Brasnu & al. 2005:156).

### 3.2.5.2 Cordectomie de type Vb

Cette intervention correspond à l'exérèse partielle ou totale du cartilage aryténoïde — en plus de l'ablation du pli vocal — (cf. figure 3.6). Elle est indiquée pour les cancers atteignant le processus vocal de l'aryténoïde, voire sa face antérieure (lésions T1a, T1b et quelques T2 sélectionnés). L'aryténoïde doit rester mobile.

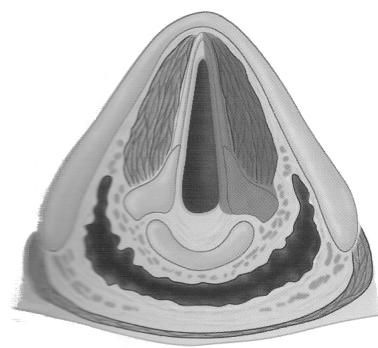


Figure 3.6 – Cordectomie Laser de type Vb (d'après Brasnu & al. 2005:157).

### 3.2.5.3 Cordectomie de type Vc

Elle emporte le pli vocal et la bande ventriculaire (cf. figure 3.7). Elle est indiquée pour les cancers glottiques et/ou sus-glottiques de type T2.

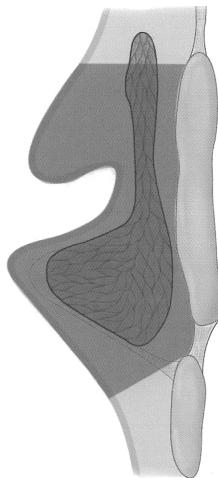


Figure 3.7 – Cordectomie Laser de type Vc (d'après Brasnu & al. 2005:158).

### 3.2.5.4 Cordectomie de type Vd

Celle-ci consiste en l'exérèse du pli vocal et de la sous-glotte jusqu'au bord supérieur du cartilage cricoïde (cf. figure 3.8). Ces interventions sont indiquées pour les cas sélectionnés de cancers de type T2. De manière générale, les chirurgies conservatrices par voie externe sont préférées à ces dernières puisque celles-ci donnent des marges d'exérèse plus importantes.

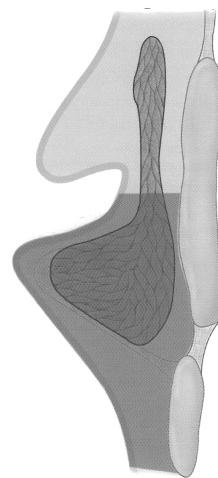


Figure 3.8 – Cordectomie Laser de type Vd (d'après Brasnu & al. 2005:159).

## 3.3 Revue de la littérature

Le développement récent des cordectomies laser par voie endoscopique a engendré quelques recherches portant sur les résultats fonctionnels, notamment en termes de qualité de voix et de qualité de vie. Cependant, l'hétérogénéité des populations étu-

diées associée aux diverses méthodologies utilisées rendent ces études difficilement comparables les unes par rapport aux autres (cf. tableau 3.1).

Dans cette section, nous réaliserons donc une revue de la littérature des différentes études concernant la voix après cordectomie laser en nous intéressant plus particulièrement aux populations, aux corpus utilisés et aux méthodologies employées.

### 3.3.1 Les populations étudiées

La majorité des études traitent de séries plus ou moins importantes composées ou non de plusieurs sous-groupes. La distinction entre les différents types de cordectomies est pourtant essentielle car la qualité de voix est corrélée à l'étendue de l'exérèse. Ainsi, les recherches de Zeitels & al. (2002), Peretti & al.(2003), Policarpo & al. (2004), Krengli & al. (2004), Haddad & al. (2006), Ledda & al. (2006), Motta & al. (2008), Claeys (2010) traitent des résultats fonctionnels sur la voix selon les types de cordectomies (selon le classement du « Committe of the European Laryngological Society ») ou les types de compensations mis en évidence par vidéolaryngoscopie (Motta & al. 2008). Les résultats ne sont ainsi pas biaisés car les groupes ne sont pas mélangés. De même, McGuirt & al. (1994), Tamura & al. (2003), Lopez Llames & al. (2004), Mirghani (2009) ont concentré leurs recherches sur un unique type de chirurgie. Par contre, certaines études couplent les chirurgies. Les populations ne sont donc plus homogènes et les résultats fonctionnels peuvent être influencés par les sujets dont la chirurgie est la plus importante (Sjögren & al. 2008, Sjögren & al. 2009, Roh & al. 2007, Vilaseca & al. 2007). Enfin, certaines publications ne définissent pas rigoureusement le type de cordectomie étudiée, ce qui est encore plus gênant dans l'interprétation des résultats (Zeitels 1995, Goor & al. 2006, Nunez-Batalla & al. 2008).

Par ailleurs, les comparaisons effectuées ne sont pas standardisées. Effectivement, dans un grand nombre de cas, les délais entre l'intervention et les enregistrements diffèrent. De ce fait, les qualités de voix ne sont pas identiques et donc pas comparables (voix enregistrées à quelques mois post-opératoires *vs* voix entraînées par des séances d'orthophonie). En outre, les types de groupes pris en parallèle s'opposent d'une recherche à une autre. Ainsi, certaines publications comparent les groupes de patients ayant subi une cordectomie laser avec des patients traités par radiothérapie (McGuirt & al. 1994, Delsupehe & al. 1999, Policarpo & al. 2004, Krengli & al. 2004, Sjögren & al. 2008, Goor & al. 2006, Nunez-Batalla & al. 2008), alors que d'autres comparent les patients avec un groupe contrôle de sujets sains (Zeitels 1995, Sittel & al. 1998, Peretti & al. 2003, Haddad & al. 2006, Ledda & al. 2006, Vilaseca & al. 2007, Motta & al. 2008, Mirghani 2009, Claeys 2010). D'autres recherches peuvent associer un groupe de patients ayant subi une radiothérapie et/ou un groupe de patients traités par cordectomie par voie externe (Schindler & al. 2004, Jotic & al. 2011). Certains travaux comparent à la fois des patients traités par cordectomie laser, des patients traités par radiothérapie et des sujets sains (Rydell & al. 1995, Tamura & al. 2003). Enfin, quelques publications comparent les résultats de leur cohorte à des « valeurs normatives ». Cette méthodologie semble problématique dans le sens où l'on ne peut pas décider de la normalité d'une voix. De plus, les valeurs liées au seuil de normalité que

les différents logiciels d'analyse de la voix définissent sont divergentes d'un logiciel à un autre, ce qui rend la comparaison difficile (Lopez Llames & al. 2004, Sjögren & al. 2009). D'autre part, certaines séries associent des hommes et des femmes, alors que les résultats acoustiques varient fortement d'une population à une autre (Sittel & al. 1998, Peretti & al. 2003, Policarpo & al. 2004, Krengli & al. 2004, Goor & al. 2006, Haddad & al. 2006, Roh & al. 2007, Sjögren & al. 2009).

### 3.3.2 Corpus

Les publications sont issues de plusieurs langues, ce qui rend la comparaison des différentes données difficile. La plupart des mesures acoustiques s'effectuent sur la voyelle tenue [a], mais le recours à d'autres voyelles n'est pas impossible, surtout pour les tracés des spectrogrammes et phonétogrammes (McGuirt & al. 1994, Peretti & al. 2003, Tamura & al. 2003, Policarpo & al. 2004, Krengli & al. 2004, Schindler & al. 2004, Ledda & al. 2006, Roh & al. 2007, Vilaseca & al. 2007, Motta & al. 2008, Sjögren & al. 2008, Sjögren & al. 2009, Mirghani 2009, Claeys 2010, Jotic & al. 2011). Ainsi, les voyelles [e], [i], [o] et [u] peuvent être analysées par certains auteurs en complément ou non du [a] (McGuirt & al. 1994, Tamura & al. 2003, Lopez Llames & al. 2004, Haddad & al. 2006, Vilaseca & al. 2007, Nunez-Batalla & al. 2008, Motta & al. 2008, Claeys 2010). Les corpus les plus complets peuvent également contenir des mots dans la langue cible (Policarpo & al. 2004, Krengli & al. 2004), de la parole spontanée (Vilaseca & al. 2007), des phrases (Vilaseca & al. 2007, Mirghani 2009, Jotic & al. 2011), des mots ou des textes lus (McGuirt 1994, Sittel & al. 1998, Mirghani 2009) et parfois même des voix chantées (Vilaseca & al. 2007).

### 3.3.3 Méthodologies

Comme nous l'avons évoqué précédemment, la description des voix pathologiques — et dans le cas présent, des voix après cordectomie laser — est une tâche difficile. La disparité entre les diverses publications, la frontière avec la question de la normalité et les différents critères (culturels, sociologiques, esthétiques etc.) liés au jugement perceptif sont autant d'arguments en faveur d'une approche multiparamétrique de la voix associant des analyses acoustiques, perceptives, aérodynamiques et morphologiques. La synthèse concernant le sujet confirme cette tendance. En effet, excepté les études de Delsupehe & al. (1999) et Goor & al. (2006), l'ensemble des recherches tendent à coupler les différentes mesures mises à leur disposition. Ainsi, la première recherche ne présente qu'une unique analyse perceptive basée sur un score spécifique pour 30 patients tandis que la seconde se base uniquement sur des questionnaires de qualité de vie pour 54 patients.

#### 3.3.3.1 Analyses acoustiques

La majorité des études présentées rapporte une analyse complète comportant de manière générale des valeurs de la *fréquence fondamentale*, du *jitter*, du *shimmer*, de

l'intensité, du *HNR*, du degré de *sub-harmoniques*, associées ou non à des *représentations spectrographiques* et/ou des *phonétogrammes* (McGuirt & al. 1994, Rydell & al. 1995, Zeitels 1995, Sittel & al. 1998, Peretti & al. 2003, Tamura & al. 2003, Pollicarpo & al. 2004, Schindler & al. 2004, Sjögren & al. 2008, Sjögren & al. 2009, Lopez Llames & al. 2004, Krengli & al. 2004, Ledda & al. 2006, Haddad & al. 2006, Roh & al. 2007, Vilaseca & al. 2007, Motta & al. 2008, Nunez-Batalla & al. 2008, Mirghani 2009, Claeys 2010 et Jotic & al. 2011). Bien que les différents logiciels utilisés ne soient pas identiques — ce qui soulève un problème de norme — et que la détection des paramètres ne soit pas toujours évidente en raison des irrégularités du signal, l'ensemble des auteurs notent des perturbations tant au niveau de la fréquence qu'au niveau de l'amplitude. Les valeurs des patients semblent plus élevées que celles des valeurs normatives ou des témoins. Les auteurs ajoutent que les paramètres acoustiques sont davantage perturbés avec l'étendue de la chirurgie. Par contre, en ce qui concerne les comparaisons acoustiques entre patients traités par cordectomie laser et patients traités par radiothérapie, les résultats divergent. McGuirt & al. (1994) et Tamura & al. (2003) n'ont trouvé aucune différence significative entre les deux groupes de patients bien que la qualité de voix des patients irradiés semble meilleure. Pour Pollicarpo & al. (2004), Sjögren & al. (2004), la cordectomie reste le traitement le plus adapté en terme de qualité de voix. Au contraire, les études de Rydell & al. (1995), Delsupehe & al. (1999), Krengli & al. (2004), Nunez-Batalla & al. (2008) et Jotic & al. (2011) mettent en évidence de meilleurs résultats acoustiques après radiothérapie.

### 3.3.3.2 Analyses perceptives

De nombreuses études ont couplé ces mesures acoustiques à des analyses perceptives et/ou des questionnaires de qualité de vie remplis par les patients eux-mêmes. D'un point de vue perceptif, l'échelle la plus courante est le *GRBAS* — où G= *grade of voice abnormal*, R= *roughness*, B= *breathiness*, A= *asthenic* et S= *strained quality* — (Rydell & al. 1995, Zeitels 1995, Peretti & al. 2003, Lopez Llames & al. 2004, Schindler & al. 2004, Ledda & al. 2006, Haddad & al. 2006, Roh & al. 2007, Vilaseca & al. 2007, Nunez-Batalla & al. 2008, Sjögren & al. 2008, Sjögren & al. 2009, Mirghani 2009 et Claeys 2010). Bien qu'incomplète sur bien des points, cette échelle est la plus reconnue dans l'analyse des voix pathologiques, ce qui lui confère un statut particulier. Les auteurs ont insisté sur le fait que le caractère de raucité avait tendance à s'estomper avec le temps en post-opératoire pour les cordectomies les plus minimes. Cependant, ces voix conservent une dysphonie persistante avec une impression de souffle (fermeture glottique incomplète). A l'instar des analyses acoustiques, les auteurs précisent que les grades, correspondant à la qualité de la voix, se détériorent en fonction de l'étendue de la cordectomie. Notons toutefois que ces résultats pourraient diverger d'une étude à une autre selon le choix des jurys (experts ou naïfs) et selon la nature de leur référent interne quant aux paramètres liés à la normalité d'une voix. En ce qui concerne les autoévaluations des patients, le questionnaire le plus utilisé est le *Voice Handicap Index* ou *VHI* (Peretti & al. 2003, Lopez Llames & al. 2004, Goor & al. 2006, Roh & al. 2007, Nunez-Batalla & al. 2008, Sjögren & al. 2008, Sjögren &

al. 2009, Mirghani 2009 et Claeys 2010). Cependant, il est important de noter que certaines publications ont utilisé d'autres questionnaires tels le *COOP/WONCA* (Goor & al. 2006), l'*échelle de Sittel* (Sittel & al. 1998, Peretti & al. 2003 et Haddad & al. 2006), le *RQOL* (Roh & al. 2007) ou des *scores spécifiques* (McGuirt & al. 1994, Rydell & al. 1995, Delsupehe & al. 1999, Haddad & al. 2006 et Vilaseca & al. 2007), ce qui rend les comparaisons difficiles. En effet, ces questionnaires ne sont, pour la plupart, pas reconnus par les professionnels. De plus, ceux-ci prennent pour cible différents critères liés au handicap vocal, ce qui soulève le problème de leur normalisation. De manière générale, les auteurs rapportent des scores améliorés en post-opératoire, ce qui est généralement en contradiction avec les analyses perceptives.

### 3.3.3.3 Analyses aérodynamiques

Des analyses aérodynamiques peuvent compléter l'ensemble des mesures décrites précédemment. Le *temps maximum de phonation (TMP)* est, sans aucun doute, le paramètre le plus utilisé (Sittel & al. 1998, Peretti & al. 2003, Tamura & al. 2003, Lopez Llames & al. 2004, Schindler & al. 2004, Ledda & al. 2006, Roh & al. 2007, Vilaseca & al. 2007, Nunez-Batalla & al. 2008, Sjögren & al. 2008, Sjögren & al. 2009 et Mirghani 2009). Au TMP peuvent s'ajouter la *capacité vitale* (Peretti & al. 2003, Sjögren & al. 2008 et Sjögren & al. 2009), le *quotient phonatoire* (Peretti & al. 2003, Sjögren & al. 2008 et Sjögren & al. 2009), le *débit d'air moyen* ou le *débit phonatoire* (Rydell & al. 1995, Zeitels 1995, Zeitels 2002, Tamura & al. 2003, Roh & al. 2007 et Mirghani 2009), la *résistance laryngée* (McGuirt & al. 1994), et la *pression intra-orale* (Zeitels 1995, Zeitels 2002 et Mirghani 2009). Les auteurs notent que l'ensemble de ces paramètres semblent perturbés après cordectomie laser puisque le TMP diminue alors que les débits d'air et la PIO augmentent par rapport aux témoins.

### 3.3.3.4 Analyses morphologiques

Enfin, des analyses morphologiques par *stroboscopie* permettent d'observer le comportement glottique (McGuirt & al. 1994, Zeitels 1995, Sittel & al. 1998, Peretti & al. 2003, Krengli & al. 2004, Policarpo & al. 2004, Schindler & al. 2004, Haddad & al. 2006, Ledda & al. 2006, Roh & al. 2007, Sjögren & al. 2008, Sjögren & al. 2009 et Mirghani 2009). Les auteurs ont montré l'intérêt de visualiser le larynx pour rendre compte des mécanismes de compensation après cordectomie. Policarpo & al. (2004) caractérisent trois types de compensations : (i) compensation par adduction du pli vocal valide avec la néocorde, (ii) compensation par adduction des fausses cordes vocales et (iii) compensation par une hyperadduction des arytenoïdes. Dans l'ensemble, ils notent une fermeture glottique incomplète associée à une ondulation muqueuse asymétrique et/ou à un serrage glottique.

### 3.3. Revue de la littérature

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures aérodynamiques	Mesures morphotologiques
<b>McGuirt (1994)/(11)</b>	III	Groupe irradié et valeurs normatives	VisiPitch : $F_0$ , jitter, intensité, % de voisement et analyses spectrales	Score spécifique	Résistance laryngée	Stroboscopie
<b>Rydell (1995)/(18)</b>	n.s.	Groupe irradié et groupe contrôle	SoundScope : shimmer, jitter, breathiness, HNR, $F_0$	GRBAS	Nbre de mots/min	
<b>Zeitels (1995)/(17)</b>	n.s.	groupe contrôle	$F_0$ , jitter, shimmer, HNR	GRBAS	PIO et débits d'air	Stroboscopie
<i>Classification</i>						
<b>Sittel (1998)/(72)</b>		Groupe contrôle	$F_0$ et phonétogramme	Echelle de Sittel + échelle d'autoévaluation	TMP	Stroboscopie
<b>Delsupehe (1999)/(30)</b>	n.s.	Groupe irradié		Score spécifique		
<b>Zeitels (2002)/(32)</b>	15 types I 8 types II 9 types IV (pas de sous-groupe)	Groupe avant et après reconstruction	$F_0$ , jitter, shimmer, HNR	PIO et débit phonatoire		Suite →

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures aérodynamiques	Mesures morphologiques
Tamura (2003)/(10)		Groupe irradié et groupe contrôle	SoundScope : $F_0$ , jitter, shimmer, HNR		TMP et débits d'air	Stroboscopie
Peretti (2003)/(51)	4 types I 22 types II 11 types III 8 types IV 6 types V	Groupe contrôle pour les analyses acoustiques	MDVP : $F_0$ , jitter, shimmer, HNR	GRBAS + VHI + Echelle de Sittel	TMP + capacité vitale + quotient phonatoire	Stroboscopie
Policarpo (2004)/(63)	23 types III 29 types IV 9 types Va 2 types Vb (seuls 20 patients ont finalement participé à l'étude)	Groupe irradié	MDVP : $F_0$ , jitter, shimmer et analyses spectrales		Stroboscopie	
Lopez Llames (2004)/(19)		Valeurs normatives issues d'autres articles	DrSpeech : $F_0$ , jitter, shimmer, HNR + analyses spectrales	GRBAS + VHI	TMP	
Krengli (2004)/(30)	10 types III 20 types IV	Groupe irradié	MDVP : $F_0$ , jitter, intensité, shimmer, HNR, degré de sub-harmonique et analyses spectrales			Stroboscopie

### 3.3. Revue de la littérature

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures aérodynamiques	Mesures morphotologiques
<b>Schindler (2004)/(30)</b>	n.s.	Groupe traité par cordectomie externe	MDVP CSL : $F_0$ moyenne, $F_0$ max, $F_0$ min, jitter, shimmer, HNR, nombre de segments non-voisés et analyses spectrales	GRBAS	TMP	Stroboscopie
<b>Goor (2006)/(54)</b>	majorité de II	Groupe irradié	VHI + COOP/WONCA	GRBASI (séparation en 2 sous-groupe pour l'analyse perceptive : I-II et III-IV) + Score spécifique + échelle de Sittel + RQOL		
<b>Haddad (2006)/(15)</b>	3 types I 5 types II 4 types III 3 types IV	Groupe contrôle	MDVP : $F_0$ , jitter, shimmer, HNR	Stroboscopie		
<b>Ledda (2006)/(141)</b>	16 types I 28 types II 31 types III 13 types IV 45 types V	Groupe contrôle	MDVP : $F_0$ , jitter, shimmer, HNR, valeurs des sub-harmoniques	GRBAS	TMP	Stroboscopie
						Suite →

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures dynamiques	Mesures aérodynamiques	Mesures morphotologiques
<b>Roh (2007)/(85)</b>	45 types I-II 17 types III-IV 13 types V	Comparaison inter-groupes	SpeechStudio : F <sub>0</sub> , jitter, shimmer, HNR	VHI + GRBAS + QOL questionnaire	TMP + débits d'air moyens	Stroboscopie	
<b>Vilaseca (2007)/(42)</b>	6 types I 11 types II 9 types III 16 types V 2 sous-groupes : I-II-III et V	Groupe contrôle	CSL Kay : F <sub>0</sub> , jitter, shimmer, HNR + phrases intonatives + voix chantée + phonétogramme	GRBAS + Score spécifique	TMP		
<b>Nunez-Batalla (2008)/(19)</b>	III	Groupe irradié	DrSpeech : F <sub>0</sub> , jitter, shimmer, HNR + analyses spectrales	GRBAS + VHI	TMP		Suite →

### 3.3. Revue de la littérature

63

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures aérodynamiques	Mesures morphotologiques
<b>Motta (2008)/(135)</b>	<p>5 types III 44 types IV 34 types Va 29 types Vb-d 23 n.s</p> <p>5 sous-groupes (selon la compensation) :</p> <ul style="list-style-type: none"> <li>- compensation glottique</li> <li>- compensation inférieure mixte</li> <li>- compensation vestibulaire</li> <li>- compensation supérieure mixte</li> <li>- compensation aditus laryngis</li> </ul>	Groupe contrôle	CSL : HNR+ intensité moyenne	TMP	Stroboscopie	
<b>Sjögren (2008)/(18)</b>	types I et II	Groupe irradié	GRBAS + VHI	Vitale + quotient phonatoire	TMP + capacité	Suite →

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures dynamiques	Mesures aérodynamiques	Mesures morphotologiques
Sjögren (2009)/(37)	types I-II	Valeurs normatives	Praat : $F_0$ , jitter, shimmer, HNR	GRBAS + VHI	TMP + quotient phonatoire	vitale + quotient phonatoire	Stroboscopie
Mirghani (2009)/(7)	I	Groupe contrôle	MDVP : $F_0$ , jitter, shimmer, HNR + intensité + phrases intonatives + texte lu + analyses	GRBAS + échelle visuelle analogique + VHI	TMP + débit phonatoire + PIO	Stroboscopie	
Claeys (2010)/(23)	5 types I 2 types II 6 types III 3 types IV 6 types V 1 type VI 3	Groupe contrôle	MDVP : $F_0$ , écart-type ( $F_0$ ), jitter, shimmer, HNR + phonétogrammes	GRBAS + VHI			Suite →

### 3.3. Revue de la littérature

Auteurs (année)/(nbre de patients)	Types de cordectomies	Types de comparaisons	Mesures acoustiques	Mesures perceptives	Mesures aérodynamiques	Mesures morphologiques
Jotic (2011)/(19)	types III-IV	Groupe irradié et groupe traité par cordectomie externe	Tiger DRS software : F <sub>0</sub> , écart-type (F <sub>0</sub> ), jitter, shimmer, F <sub>0</sub> trémor, amplitude tremor, NNE, intensité moyenne, écart-type (intensité) + phonétogrammes			

**Table 3.1** – Tableau synthétique des différentes études sur la voix après cordectomie. Le nom du premier auteur est précisé et ns est spécifié quand aucune information n'est disponible.

## 3.4 Les analyses perceptives

La méthode perceptive reste la méthode la plus ancienne pour juger de la qualité d'une voix. Les premiers essais de classification de voix datent du XIXème siècle (Rush 1859). Morsomme et Estienne (2006) précisent également que l'oreille reste le premier évaluateur de la qualité de la voix, avant l'utilisation d'une instrumentation plus complexe. En effet, l'évaluation perceptive, bien que subjective, représente une référence à laquelle sont confrontées les autres mesures vocales (Klatt & Klatt 1990, Kreiman & al. 1993, Giovanni & al. 1996, entre autres).

### 3.4.1 Définition

L'évaluation perceptive consiste en un jugement propre d'un individu sur les qualités acoustiques d'une voix et à les comparer à une représentation personnelle de la normalité, le référent interne (Fex 1992). Cette notion de normalité, propre à l'auditeur, varie selon l'expérience de l'expérimentateur, la langue ou les facteurs socio-culturels (Kreiman & al. 1992, 1993). C'est donc une notion subjective et variable d'une culture à une autre. En effet, par exemple, une voix soufflée peut être un indice de pathologie pour certains mais une qualité de voix particulière dans le traitement linguistique pour d'autres. Ainsi, nous ne pouvons pas dire qu'il existe véritablement une voix de référence dite « normale ». Laver (1994) précise néanmoins que d'une manière générale, l'auditeur doit prendre en compte quatre principales composantes de la perception auditive pour différencier les éléments de la parole : la qualité, la durée, la hauteur et l'intensité.

Les échelles perceptives se sont succédées au cours du temps pour tenter de classifier les voix pathologiques ou dysphoniques. Ces systèmes d'évaluation dépendent principalement des types de voix à analyser (normales ou pathologiques), de l'expérience de l'auditeur (naïf ou expert) et de l'objectif de l'expérience. A l'heure actuelle, nous savons qu'une analyse perceptive complète devrait comprendre une évaluation de la qualité de la voix, une évaluation de la qualité de la parole, son efficacité et son intelligibilité (Oates 2009).

### 3.4.2 Les types d'échelles

Afin de mieux se rendre compte de l'utilité des échelles proposées au fil du temps (points positifs et inconvénients), une rapide revue de la littérature de ces dernières sera réalisée avant de nous intéresser plus particulièrement à l'échelle GRBAS qui est la plus fréquemment utilisée de nos jours.

#### 3.4.2.1 L'échelle d'Osgood

Cette échelle est la plus ancienne. Celle-ci se base sur un concept de binarisme sémantique qui permettait une quantification de grandeur. Ainsi, l'auteur avait établi une liste de 50 paires d'adjectifs en opposition (propre-sale, dur-mou, grand-petit

etc.) regroupées en classes. Cette première échelle fut le point de départ des nombreuses échelles bipolaires suivantes et l'introduction aux méthodes sémantiques différentes. Néanmoins, cette échelle fut rapidement abandonnée car de nombreux adjectifs utilisés ne rendent pas compte clairement de la qualité d'une voix (Osgood 1957).

### 3.4.2.2 L'échelle de Voiers

Cette échelle est basée sur quatre facteurs principaux pour évaluer la qualité de la voix et de la parole : l'intelligibilité, la raucité, la sonie et le débit syllabique (Voiers 1964).

### 3.4.2.3 L'échelle GRBAS

Cette échelle est née des travaux d'Isshiki & al. (1969) qui, partant de la méthode différentielle sémantique d'Osgood (1957), a permis d'extraire quatre paramètres pertinents pour l'analyse de la voix : la raucité, le souffle, l'asthénie et le degré de sévérité de la dysphonie. En 1981, Hirano, à partir d'analyses spectrographiques supplémentaires, a affiné les travaux de ses comparses et a mis en évidence cinq facteurs principaux pour qualifier une voix pathologique : le facteur G (*Grade of voice abnormal*) correspond au degré de dysphonie, le facteur R (*Roughness*) correspond à la raucité, le facteur B (*Breathiness*) correspond au caractère soufflé de la voix, le facteur A (*Asthениc*) correspond au caractère hypotonique de la voix, et enfin le facteur S (*Strained quality*) correspond à l'impression de serrage de la voix (cf. tableau 3.2). Chacun des facteurs est évalué sur une échelle de sévérité allant de 0 à 3 (où 0= normal, 1= légèrement altéré, 2= moyennement altéré, 3= sévèrement altéré) (Hirano 1981). Cette échelle simple et rapide est largement utilisée dans la pratique clinique quotidienne car ces facteurs principaux restent les plus pertinents pour l'évaluation de la qualité des voix pathologiques. Néanmoins, cette dernière ne rend pas compte des fonctions supra-laryngées, des indices temporels (débit) ou prosodiques (intonation, pauses, etc.) ou encore de certains facteurs d'instabilités. Ces limites ont conduit certains auteurs à remanier l'échelle GRBAS selon d'autres aspects. Dejonckere & al. (1993) ont proposé des compléments à cette échelle en y ajoutant des paramètres tels que : le temps maximum de phonation (TMP), l'étendue vocale, la puissance vocale maximale (intensité maximale sur un [a] tenu), le comptage maximal sur une expiration (le nombre maximal et intelligible sur l'expiration), le temps de comptage de 1 à 25 (le plus rapidement en restant toujours intelligible) et un test d'intelligibilité. A la fin des années 90, ce même groupe de chercheurs a proposé l'échelle GIRBAS où I (*Instability*) prévoit les phases d'instabilités de qualité vocale dans le temps (Dejonckere & al. 1996, 1998). Enfin, dans sa thèse de doctorat, Crevier Buchman (1999) a démontré que des paramètres tels le voisement, la sonie, la hauteur, le débit et la mélodie semblent bien plus pertinents que l'échelle GRBAS dans l'étude perceptive post-opératoire de patients traités par LPSC-CHEP (Crevier Buchman 1999).

	Echelle	Description
G	<i>Grade</i>	Degré de sévérité de la dysphonie.
R	<i>Roughness</i>	Raucité de la voix, en rapport avec une fluctuation irrégulière et une aggravation de la hauteur
B	<i>Breathiness</i>	Impression de souffle dans la voix, en rapport avec une incompétence glottique
A	<i>Asthenic</i>	Asthénie, en rapport avec un comportement hypotonique. Voix faible, qui manque de puissance et d'intensité
S	<i>Strained</i>	Serrage laryngé et/ou supraglottique, en rapport avec un comportement hypertonique. Voix aiguë et/ou bruit dans les hautes fréquences

**Table 3.2 – L'échelle GRBAS (Hirano 1981).**

#### 3.4.2.4 L'échelle Laver's Voice Profil Analysis Sheme (VPAS)

C'est une échelle qui se base sur une description phonétique de la qualité de la voix. Les caractéristiques de la voix sont comparées à une position articulatoire neutre ayant un corrélat physiologique et acoustique. L'évaluation des positions phonétiques se fait à l'écoute et en relation avec l'observation du locuteur (Laver 1980, 1994).

#### 3.4.2.5 Le Stockholm Voice Evaluation Consensus Model (SVEC)

C'est une échelle qui est validée uniquement pour la langue suédoise. En effet, pour les auteurs, la qualité vocale dépend uniquement du contexte linguistique et culturel d'une population. Cette échelle se compose de 12 traits principaux (aphonie, souffle, tendue, détendue, rauque, raucité grave ou *creaky*, éraillée, cassures vocales, diplophonie, registre, hauteur et sonie) qui permettent de décrire parfaitement une voix pathologique suédoise. Cette méthode permet d'établir principalement des corrélations entre l'évaluation perceptive et certains paramètres acoustiques. Les auteurs préconisent que l'évaluation soit réalisée par des auditeurs experts, qui restent en permanence en contact avec des voix « de référence » afin de conserver une référence stable et identique pour tous les auditeurs (Hammarberg & Gauffin 1995).

### 3.4.3 Applications en pathologie

Nous avons donc démontré que l'ensemble de ces échelles met en évidence un nombre important de critères à prendre en considération dans l'analyse perceptive des voix pathologiques. Néanmoins, le nombre croissant d'échelles pose le problème de leurs subjectivités. En effet, le manque de fiabilité intra et inter-locuteurs associé au caractère aléatoire des choix des paramètres à prendre en compte sont des arguments contre la prise de position pour une échelle ou une autre. Dans ce sens, il est important de préciser qu'aucune étude à ce jour n'a jamais validé les traits les plus pertinents pour les voix pathologiques. En outre, nous avons montré que le jugement des audi-

teurs s'effectuait à partir d'une représentation personnelle de la normalité de la voix, à partir d'un référent interne (Fex 1992). Mais la question essentielle est de savoir à quel niveau se situe réellement la normalité et quelle est sa limite ? Finalement, l'absence de définition des traits perceptifs pertinents des voix normales peut être un biais pour la fiabilité des jugements des voix pathologiques. En outre, les passations de tests de perception ne sont pas toujours standardisées, ce qui entache une fois de plus leur validité. Certains facteurs qui influencent la fiabilité de l'évaluation sont contrôlables : les échantillons vocaux présentés sont à l'appréciation de l'expérimentateur (voyelle tenue, phrases ou parole spontanée), les auditeurs sont variables d'une étude à une autre (jury naïf ou expert), les tâches d'écoute sont différentes selon les paramètres perceptifs à juger. *A contrario*, d'autres facteurs dépendent des auditeurs et ne sont pas contrôlables par l'expérimentateur : phénomènes de mémoire, diminution d'attention, fatigue, etc.

De notre point de vue, il est important de souligner que ces échelles perceptives ont été créées à partir d'échantillons vocaux de voix pathologiques où les deux plis vocaux ont été conservés. Ainsi, l'utilisation de ces dernières sur des voix après chirurgies partielles laryngées avec résection d'un ou des deux plis vocaux n'est qu'une application secondaire. En conséquence, elles ne sont ni fiables, ni valides en ce qui concerne ces types de voix. Ainsi, comme nous l'avons montré plus haut, de nombreux auteurs se sont basés sur les paramètres principaux développés par l'échelle GRBAS (Hirano 1981) pour en critiquer la fiabilité pour les voix après LPS. En 1999, Crevier Buchman a mis en évidence l'inefficacité de cette échelle après LPS-CHEP. Cette recherche a démontré que seul le trait R était pertinent, et que l'échelle dans son intégralité ne permettait pas de qualifier l'évolution en postopératoire de 6 à 18 mois de ces voix (Crevier Buchman 1999). Aussi, de nombreux autres paramètres ont été évoqués pour une meilleure description perceptive de ces voix (Dejonckere & al. 1993).

## 3.5 Les autoévaluations

Le développement des échelles de qualité de vie orientées sur les handicaps vocaux des patients est récent. Face aux limites imposées par les échelles perceptives, des questionnaires d'autoévaluation permettent d'évaluer le point de vue subjectif des patients par rapport à leur propre voix. Elles sont utiles aux patients dans la description de leur vécu au quotidien, dans l'évaluation de leurs performances vocales face à leur entourage et aux situations de communication (Murry & al. 2004). Le patient doit s'interroger sur l'esthétique de sa voix, sur ce qu'il ressent par rapport à sa voix et sur ce que l'interlocuteur perçoit de sa voix.

### 3.5.1 Les types d'échelles

Plusieurs échelles de qualité de vie sont à disposition des patients. Nous exposerons ci-dessous les principales utilisées en clinique.

### 3.5.1.1 Le Medical Outcomes Study Short Form 36 (MOS SF 36)

Le MOS SF 36 a été introduit par Leplège & al. (1998). Ce questionnaire générique a été largement utilisé dans le domaine de la pathologie (Ware & Sherbourne 1992). Cette échelle recouvre 8 domaines, classés en 2 sous-groupes : « santé physique » (analyses de l'activité physique dans la vie quotidienne) et « santé psychique » (qualité de vie en relation avec un entourage).

### 3.5.1.2 Le Voice Handicap Index (VHI)

Le VHI a été développé par Jacobson & al. (1997). Ce questionnaire de qualité de vie est actuellement le plus utilisé dans le domaine de la pathologie (Benninger & al. 1998, Murry & Rosen 2000, Hogikyan & Rosen 2002). Il comprend 30 items regroupés en 3 classes (physique, émotionnel et fonctionnel). Une grille de réponse à 5 degrés de sévérité est proposée (0= non, pas de problème, 4= oui, toujours un problème). Les cliniciens considèrent que plus le chiffre final est élevé, plus le degré du handicap vocal est ressenti par le patient et plus la qualité de vie diminue.

### 3.5.1.3 Le Voice Related Quality Of Life (V-RQOL)

Le V-RQOL a été développé par Hogikyan et Sethuraman (1999). Ce questionnaire regroupe 10 items associés à une échelle de réponse de 1 (non, pas de problème) à 5 (oui, beaucoup de problèmes). Les domaines abordés par cet auto questionnaire sont la puissance vocale, l'essoufflement, l'instabilité vocale, l'anxiété et le sentiment dépressif lié au handicap vocal, les difficultés à tenir une conversation téléphonique ainsi que les répercussions sociales et professionnelles liées à la dysphonie. Comme pour le VHI, plus le score augmente et plus la qualité de vie diminue.

## 3.5.2 Applications en pathologie

Les échelles de qualité de vie, concomitantes aux évaluations perceptives et objectives, ont l'intérêt de compléter un bilan vocal et de mettre les patients face à leur handicap. Ainsi, les évaluations longitudinales de qualité de vie ont pour but de montrer aux patients les forces et faiblesses de leurs voix et d'accroître leur motivation quant au traitement à poursuivre.

Dans l'évaluation de la voix, plusieurs études ont inclus ces échelles en complément des évaluations perceptives et acoustiques (Peretti & al. 2003, Lopez Llames & al. 2004, Goor & al. 2006, Haddad & al. 2006, Roh & al. 2007, Sjögren & al. 2009, entre autres). Néanmoins, une fois encore, les échelles utilisées ne sont pas normalisées et les problèmes méthodologiques se posent toujours : les questions ne sont pas toujours très claires et compréhensibles pour les patients, ces échelles ne sont pas validées pour les résections d'un ou deux plis vocaux et le manque de fiabilité inter et intra-patients perdure.

Par ailleurs, une étude de Hsiung & al. (2002) a révélé une faible corrélation entre mesures acoustiques et VHI. A l'instar de cette recherche, Behrman & al. (2004) n'ont retrouvé aucune corrélation entre les échelles d'évaluation perceptive, les analyses acoustiques et la qualité de vie des patients. En effet, la perception du patient de la sévérité de son handicap vocal n'est pas toujours corrélée avec la réalité des mesures acoustiques. Aussi, il y a peu de corrélation entre auditeurs et locuteurs sur l'appréciation de la sévérité de la dysphonie. En effet, le jugement de la qualité vocale est subjectif et personnel et ne correspond pas toujours au référent du clinicien. Behrman & al. (2004) ajoutent que les autoévaluations des patients sur des dysphonies persistantes (plus d'1 an) vont souvent en s'améliorant, ce qui soulève les hypothèses d'un phénomène d'habituation ou de compensation.

## 3.6 Les analyses acoustiques

### 3.6.1 Introduction

L'analyse acoustique permet de faire des mesures objectives et de dégager des tendances générales pour le traitement de la voix et de la parole normale ou pathologique. Morsomme et Estienne (2006) précisent que l'oreille reste le premier évaluateur et indice pour qualifier une voix. Néanmoins, ces derniers ajoutent que l'audition est insuffisante car l'oreille ne peut percevoir les petites variations. En effet, le traitement auditif et perceptif établi par l'oreille est fait dans sa globalité et celle-ci n'est pas entraînée pour relever les variations minimes et imperfections. Baken (1987 : 2) ajoute que :

« [...] de telles méthodes d'évaluation perceptives sont inadéquates [...] incapables de mettre en évidence de petites différences. »

Finalement, l'analyse acoustique reste l'analyse la plus fiable permettant d'évaluer et de mesurer quantitativement les paramètres de la voix et la parole. Elle permet de mettre en évidence les particularités de celle-ci et de la situer par rapport au seuil de la pathologie. Cependant, il est toujours nécessaire de prendre les résultats avec précaution car ces analyses ne sont pas sans faille. Les corpus sont souvent réduits et les mesures peuvent être sujettes à des erreurs.

### 3.6.2 Qu'est ce qu'une onde acoustique ?

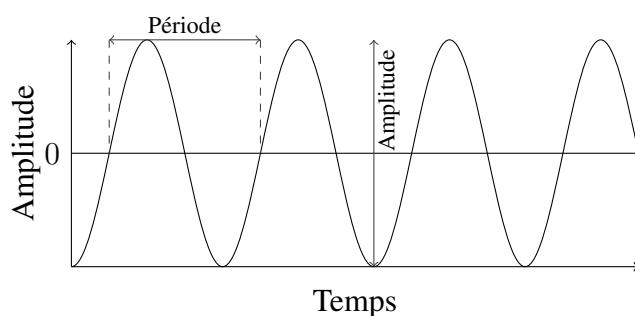
Un son correspond aux différentes variations de pression d'un milieu. Sous l'effet d'une excitation sonore, les molécules d'air vont se mettre en mouvement autour de la source. La rencontre des molécules forme des zones de compression qui tendent à s'étendre dans un milieu élastique. Finalement un son représente une mise en mouvement des molécules d'air, le phénomène de perturbation se propageant de façon répétitive sous l'effet de ce milieu environnant.

### 3.6.2.1 Le son pur

Un son pur est caractérisé par des molécules d'air qui oscillent en suivant le mode d'une sinusoïde pure, et vibrant à une période et à une amplitude données (cf. figure 3.9). Le son du diapason ou une sonnerie téléphonique sont des exemples d'ondes sinusoïdales. De manière générale, un son pur peut être défini mathématiquement par deux composantes : l'amplitude de vibration (symbole A), et la fréquence de vibration (symbole f). Ces paramètres évoluent selon une unité de temps qui est la seconde ou son sous-multiple, la milliseconde.

*La fréquence de vibration*, exprimée en Hertz, permet de quantifier les cycles complets de la sinusoïde en une seconde. Un son pur est par définition périodique, c'est-à-dire que les cycles se reproduisent à l'identique après une durée communément nommée *période* (symbole T). Elle correspond donc à la durée d'un cycle de vibration. En physique acoustique, la fréquence est l'inverse de la période, soit  $f = 1/T$  ou  $T = 1/f$ .

*L'amplitude de vibration* correspond à l'ampleur de la vibration et est directement liée aux variations de pression, elle s'exprime en Pascal (Pa). Certains auteurs précisent qu'elle peut également s'exprimer en newton/m<sup>2</sup> car la pression correspond à une force (symbole N) qui s'applique perpendiculairement sur une surface (m<sup>2</sup>). D'un point de vue acoustique, plus un son pur aura une amplitude importante et plus celui-ci sera perçu comme fort. Bien que corrélées, il est cependant essentiel de faire la distinction entre les notions d'amplitude de vibration et d'*intensité sonore* (Bourcet & Liénard 1987, Martin 2008).

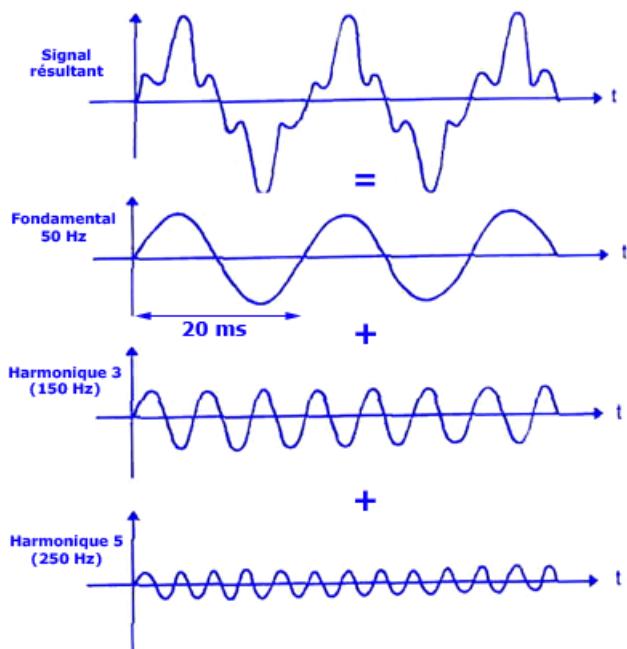


**Figure 3.9** – Représentation d'une onde périodique simple (sinusoïde) (schéma personnel réalisé avec TiKZ).

### 3.6.2.2 Le son complexe

Cette notion de sons purs est pédagogique mais reste très rare dans la nature. Selon la *loi de Fourier*, les sons complexes se définissent comme des sommes de divers sons purs d'amplitudes et de fréquences variables. En effet, le théorème de Fourier démontre qu'un signal de forme quelconque peut être divisé en plusieurs signaux purs d'amplitudes et de fréquences différentes. De même, l'addition de plusieurs sons purs produits simultanément, engendre un son complexe. Les fréquences de l'ensemble de

ces sons purs représentent des multiples entiers de celle du son pur qui a la fréquence la plus basse. Ce dernier est appelé « *le son fondamental* » et la fréquence de ce son est nommée « *fréquence fondamentale* ». Les sons purs dont les fréquences sont des multiples entiers de la fréquence fondamentale sont connus sous le nom d' « *harmoniques* » (cf. figure 3.10). Les sons peuvent néanmoins garder un aspect périodique et cyclique. Comme nous venons de l'évoquer, le cycle de base correspond à la fréquence fondamentale (notée  $F_0$ ) et ses harmoniques. Les exemples les plus évocateurs pour la parole sont les voyelles ou les sons voisés. A l'inverse, certains sons ne présentent pas cette répétition dans le temps et sont qualifiés d'apériodiques. Le bruit ou les consonnes non voisées sont des exemples de sons apériodiques.



**Figure 3.10** – Représentation d'un son périodique complexe (somme de trois sons simples dont le son fondamental a une  $F_0=50$  Hz et de ses harmoniques) (d'après [www.nquoilamusiqueaffectetellehomme.fr](http://www.nquoilamusiqueaffectetellehomme.fr)).

### 3.6.3 Analyses fréquentielles

#### 3.6.3.1 Préambules

**Le matériel utilisé.** Les prises de mesures, en particulier pour les voix pathologiques dont les variations acoustiques sont anormales, sont délicates. Les conditions d'enregistrement doivent être rigoureuses afin de garantir une reproductibilité et une fiabilité. Si le signal est très bruité, les indices acoustiques pertinents peuvent disparaître et ne pas révéler la véritable qualité vocale. La qualité du matériel d'enregistrement est donc primordial en amont de l'analyse acoustique. Ainsi, il est fortement déconseillé d'enregistrer un sujet directement sur le microphone de l'ordinateur ou à l'aide d'un microphone bas de gamme puisque le signal obtenu, bien que très correct à l'oreille,

ne pourra être analysé de façon précise par les logiciels d'analyse vocale. La démarche conseillée consiste à utiliser un microphone de qualité et, si possible, électrostatique. D'autre part, il est préférable de connecter le microphone sur un préamplificateur externe (carte son de type EDIROL par exemple) qui génère moins de bruit et qui permet une conversion directe du signal analogique en signal numérique. En outre, Titze & Winholtz (1993) précisent que la distance du microphone à la bouche peut être un élément de perturbation pour l'analyse des voix pathologiques. C'est pourquoi, ils recommandent l'utilisation d'un micro-casque où le micro, placé latéralement, serait réglé à égale distance pour chacun des sujets enregistrés, à savoir à environ 4 cm de la bouche.

**Le lieu d'enregistrement.** Tout comme le matériel, le lieu de la prise de son est déterminant dans la qualité de l'analyse acoustique. Pour un enregistrement idéal, c'est-à-dire non bruité, une chambre sourde est requise. Celle-ci a la particularité d'être totalement isolante. Effectivement, ses parois absorbent les buits extérieurs et les échos. Cependant, l'accès à ce type d'installation n'est pas évident et permis à tous. A défaut d'une chambre sourde, l'expérimentateur peut opter pour un local suffisamment isolé des bruits externes, c'est-à-dire qu'il présente une faible réverbération et qu'il soit hors d'atteinte de bruits parasites. Néanmoins, l'isolation totale reste problématique lors des enregistrements, ce qui peut par exemple être un réel biais dans les analyses acoustiques, et en particulier en pathologie.

### 3.6.3.2 Un outil : le spectrogramme

Le spectrogramme est une représentation tri-dimensionnelle de la parole en temps réel. En abscisse, nous trouvons le temps et en ordonnée, nous avons les fréquences échelonnées. L'intensité est, quant à elle, indiquée par la noirceur du tracé. Pour générer des spectrogrammes, la plupart des logiciels d'analyse vocale utilisent soit la méthode de la Transformée de Fourier (ou FFT Fast Fourier Transform) soit le codage par prédiction linéaire (ou LPC).

Les programmes tels que Praat (Boersma & Weenink 1996), qui s'appuient sur la FFT, isolent différentes parties du signal sonore à intervalles réguliers selon des « fenêtres » plus ou moins importantes. Pour chacune des fenêtres formées, les logiciels vont calculer les spectres d'amplitude correspondants. Le spectrogramme est donc la résolution de la juxtaposition de l'ensemble des spectres générés sur un axe temporel. L'apparence du spectrogramme va donc dépendre à la fois des types de fenêtres captées par les logiciels mais aussi de la taille de ces fenêtres. Boersma suggère que la fenêtre dite de GAUSS reste la plus adaptée à l'analyse spectrale.

Le problème posé par ce type de sélection des fenêtres est que dans la plupart des cas, les programmes ne respectent pas forcément la périodicité du signal sonore. Les sélections des fenêtres se font indépendamment les unes des autres avec des bords irréguliers. En conséquence, les pentes risquent d'être trop abruptes aux extrémités, ce qui entraînerait l'apparition de composantes fréquentielles nouvelles sur le spectrogramme par rapport au signal sonore original. Pour remédier à ce biais et atténuer l'amplitude du signal sur les bords, les programmeurs doivent faire intervenir des fonctions. Le but

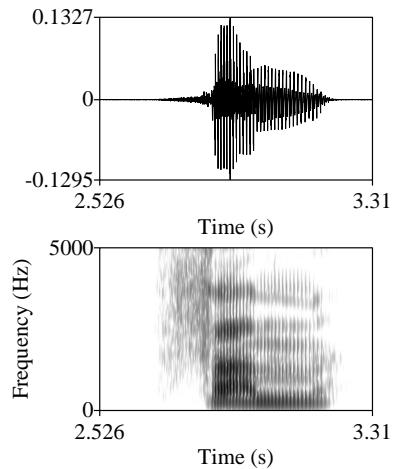
est de multiplier le contenu de la fenêtre par ces fonctions afin de moduler l'amplitude aux extrémités.

Comme nous le mentionnions précédemment, la taille de la fenêtre participe activement à la détermination de la résolution fréquentielle du spectrogramme. Ainsi, en fonction de la résolution fréquentielle, deux types de spectrogrammes sont couramment utilisés :

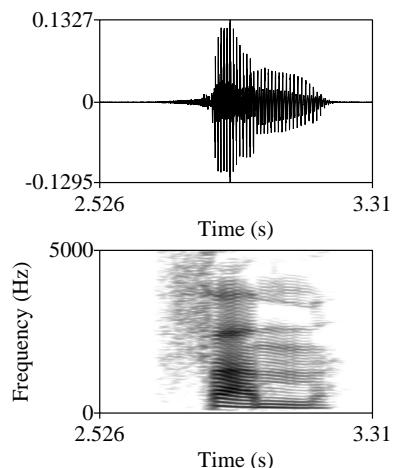
**Le spectrogramme à bandes larges.** Dans ce cas, la fenêtre est très courte. En conséquence, la résolution fréquentielle, de l'ordre de 300 Hz, devient très faible pour extraire la fréquence fondamentale et ses différents harmoniques. Ce type de spectrogramme permet de mieux repérer les formants — zones d'harmoniques renforcées — des voyelles par exemple. Par ailleurs, les différentes stries verticales qui apparaissent sur le tracé mettent en évidence la fermeture cyclique des plis vocaux. Ainsi, puisque le nombre de fermetures glottiques par secondes correspond à la  $F_0$ , il nous est quand même possible de déduire le fondamental en comptant le nombre de stries par seconde (cf. figure 3.11).

**Le spectrogramme à bandes étroites.** Dans ce cas, la fenêtre est beaucoup plus large. En conséquence, la résolution fréquentielle, de l'ordre de 45 Hz, est largement inférieure à la voix à analyser, ce qui permet de visualiser la  $F_0$  et ses composantes en harmoniques. Finalement, ce type de spectrogramme permet une meilleure représentation des caractéristiques glottiques. Par exemple, en comptant les dix premiers harmoniques visibles, il est possible de donner une valeur approximative de  $F_0$ . Celle-ci correspond à la valeur du dixième harmonique divisée par dix (cf. figure 3.12).

Les spectrogrammes restent des outils de choix pour les phonéticiens. En effet, quel que soit le type de tracé utilisé, un nombre important d'indices acoustiques émerge. Ainsi, nous pouvons observer la fréquence fondamentale et ses harmoniques, les différents formants avec leur intensité relative et leurs transitions et lire en temps réel les consonnes. Ces dernières sont repérables par des bruits de friction (fricatives), des silences associés à des explosions (occlusives) ou même par l'apparition de formants dans le cas des sonantes ou approximantes.



**Figure 3.11** – Exemple d'un spectrogramme à bandes larges du mot «femme» issu de Praat.



**Figure 3.12** – Exemple d'un spectrogramme à bandes étroites du mot «femme» issu de Praat.

### 3.6.4 Principaux paramètres utilisés en pathologie

#### 3.6.4.1 La fréquence fondamentale ( $F_0$ )

Comme nous l'avons mentionné précédemment, la *fréquence fondamentale* ( $F_0$ ) correspond au nombre de cycles effectués par les plis vocaux (en Hertz) par seconde. Celle-ci dépend de la pression sous-glottique (PSG), de la masse, de la raideur et de la taille des plis vocaux. Il n'existe pas de traits pertinents pour affirmer qu'une voix est normale. Cependant, « une périodicité stable » peut être un indice de normalité, même si de faibles variations d'un cycle à un autre apparaissent chez des locuteurs sains.

Mécaniquement, le contrôle de  $F_0$  est effectué par les muscles intrinsèques et extrinsèques du larynx. Le rôle des muscles intrinsèques ayant déjà été discuté en 1.3.2.1, nous n'aborderons pas cet aspect. De la même manière, nous avons vu que l'action des muscles extrinsèques était importante dans la production des fréquences très élevées et très basses. Ainsi, le larynx s'élève dans le registre aigu et s'abaisse dans le registre grave (Faaborg & Sonninen 1959). Il est également important de noter qu'il existe un lien étroit entre  $F_0$  et intensité : plus l'intensité augmente et plus la  $F_0$  a tendance à croître.

La détection informatique de  $F_0$  peut être effectuée de manière temporelle ou de manière fréquentielle. Dans le domaine temporel, la détection de la fréquence est basée sur les cycles (« zero-crossing », « autocorrélation »). L'analyse fréquentielle correspond, quant à elle, à la technique spectrale d'analyse des harmoniques (« technique du peigne ») ou au calcul du cepstre. Cette dernière technique, correspondant à l'inverse du spectre, permet de représenter l'amplitude des fluctuations des magnitudes spectrales. Cependant, la plupart de ces méthodes posent des problèmes de détection majeurs où des confusions voisé/non voisé apparaissent et des sauts d'octaves se manifestent. Dans le premier cas, il est fréquent d'observer, selon les réglages définis, que des parties du signal voisées soient éliminées par le détecteur ou au contraire que des parties non voisées soient considérées comme voisées, ce qui implique des valeurs de  $F_0$  erronées. Dans le second cas, nous constatons des problèmes de sauts d'octaves liés à des erreurs d'algorithme.

Dans le domaine de la pathologie, ces conflits de détection de  $F_0$  sont doublement présents. En effet, l'apériodicité du signal est un frein supplémentaire. Il est par exemple possible que l'algorithme du logiciel confonde du bruit avec une harmonique, créant ainsi des sous-harmoniques et des inter-harmoniques inexistant à la base, ce qui pose problème pour une détection correcte de  $F_0$  (Baken 1987, Nunez-Batalla & al. 2000).

### 3.6.4.2 L'intensité sonore

Comme nous l'avons souligné précédemment, il est essentiel de distinguer amplitude (cf. section 3.6.2.1) et intensité sonore. En physique, l'*intensité* est définie par un rapport de la puissance de vibration sur la surface sur laquelle elle s'applique (Bourcet & Liénard 1987, Martin 2008). Pour un son pur, cette définition revient à déterminer la quantité d'énergie fournie ou reçue (en Joule=  $J$ ) par unité de temps. De manière synthétique, nous pouvons dire qu'une source sonore diffuse de l'énergie mécanique  $E$  sous forme d'énergie cinétique et potentielle : une source sonore qui diffuse une énergie  $E$  en un temps  $t$  possède une puissance en Watts ( $W$ ) qui correspond à la formule  $W = E/t$ . Cette quantité d'énergie délivrée par seconde et mesurée en  $J/s$ , c'est-à-dire en Watts correspond à l'équivalence suivante :  $1W = 1J/s$ . Comme mentionné plus haut, l'intensité acoustique représente le flux moyen de la puissance par unité de surface, elle s'exprime donc en  $W/m^2$ . Ainsi, supposons qu'une surface  $S$  soit traversée par une énergie  $E$  pendant un temps  $t$  donné, l'intensité de cette surface sera calculée de la manière suivante :  $I = E/(t.S) = W/S$  (watts/ $m^2$ ).

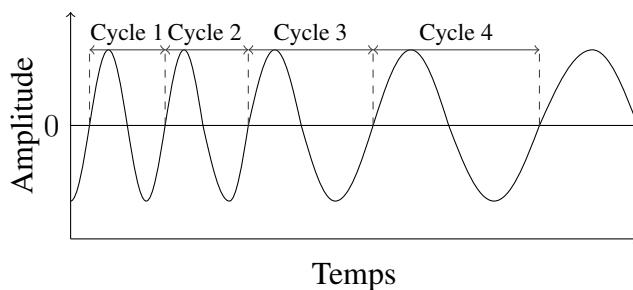
Nous avons donc vu que les unités pour représenter le niveau sonore étaient les suivantes : le Watt pour la puissance et le Watt/m<sup>2</sup> pour l'intensité acoustique. Néanmoins, ces valeurs sont très difficiles à manier et à représenter simplement car le rapport entre les plus fortes et les plus faibles mesures de puissance acoustique s'échelonne de 1 à plusieurs millions de watts. Aussi, afin de réduire cette plage de variation, les physiciens ont pris l'habitude d'exprimer la puissance et l'intensité en valeurs logarithmiques : les décibels (dB).

Au niveau glottique, les ouvertures et fermetures des plis vocaux donnent lieu à des bouffées d'air périodiques. Chacune de ces bouffées possède une vitesse  $V$  et une masse  $m$ , c'est-à-dire une énergie cinétique. La masse de l'air délivrée à chaque cycle varie faiblement tandis que la vitesse de sortie des ces impulsions change de manière non négligeable. Ces changements de vitesse — intensité vocale — dépendent de deux paramètres principaux : la pression sous-glottique (PSG) et la taille de l'orifice d'écoulement. Ainsi, pour que l'intensité augmente, la PSG doit croître de façon proportionnelle alors que la surface glottique doit diminuer (Isshiki 1964, 1965, Baken 1987, Titze 1989). Ces ajustements de pressions et de configurations du conduit vocal sont des arguments en faveur d'une corrélation positive entre intensité et fréquence fondamentale. Ainsi, plus la  $F_0$  augmente et plus l'intensité aurait également tendance à croître. Dans ce sens, une étude de Black (1961) a montré que la  $F_0$  de locuteurs s'amplifiait dans une tâche d'effort vocal. Par conséquent, il est intéressant d'observer les comportements des plis vocaux lors des variations d'intensité. Des expériences avec le glottographe ont permis de mettre en évidence certaines caractéristiques de la glotte lors d'une augmentation de l'intensité : la durée de la phase d'ouverture augmente tandis que la phase de fermeture diminue, ce qui traduit un accrolement plus rapide des plis vocaux. De plus, la durée d'occlusion glottique s'allonge alors que l'amplitude d'écartement des plis vocaux décroît (Baken 1987, Léothaud 2004).

Enfin, nous savons que la manipulation de l'intensité vocale permet au locuteur de transmettre un message plus efficacement par l'augmentation de cette dernière dans le but de focaliser sur les éléments clés par exemple (Finnegan & al. 2000). Un certain nombre d'expériences menées sur la voix parlée concernant le rétrocontrôle de l'intensité a permis de mieux appréhender les procédures de régulation (cf. Lombard 1911, Fletcher & al. 1918 par exemple). Les conclusions convergent toutes vers une hypothèse commune : le locuteur tend à s'adapter à son environnement et cherche à maintenir un niveau sonore adéquat en fonction de l'évolution des bruits qui l'entourent. Le but majeur de ce dernier étant de se faire entendre de ses auditeurs, ce qui suppose une adaptation aux conditions de parole (dimension du lieu, bruit, éloignement des auditeurs, style de parole, etc.).

### 3.6.4.3 Le jitter

Les fluctuations de durée d'un cycle glottique à un autre en terme de  $F_0$  sont communément appelées « *jitter* ». Elles sont exprimées parfois en valeurs absolues mais le plus souvent en pourcentages (Hollien & al. 1973, Schoentgen & Guchteneere 1991, 1995, entre autres). La figure 3.13 schématisé ces perturbations :

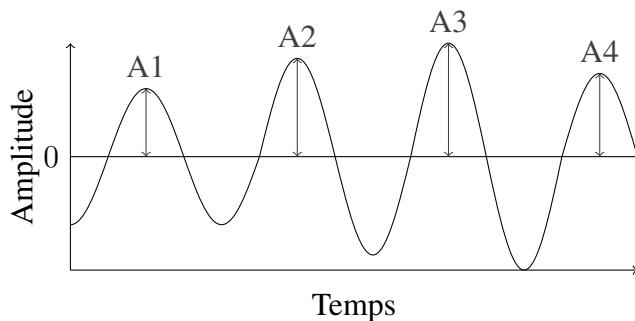


**Figure 3.13 – Représentation schématique du jitter (schéma réalisé avec TiKZ).**

Cependant, sa mesure peut être problématique du fait que les différents logiciels n'utilisent pas les mêmes références et qu'il existe plusieurs définitions de ce paramètre selon les techniques de mesure de  $F_0$ . De plus, ces mesures sont à prendre avec précaution du fait des nombreuses erreurs de détections dans l'analyse des voix pathologiques (Crevier Buchman & al. 1996, Crevier Buchman 1999, Hans 2003) Ainsi, le « *jitter absolu moyen* » représente la moyenne de la différence de fréquence entre deux cycles vibratoires consécutifs. « *Le jitter factor* » est le jitter moyen rapporté à la  $F_0$  moyenne du signal. Ce paramètre semble être le plus pertinent pour qualifier la stabilité de la  $F_0$  à court terme. Il est considéré comme l'indice le plus représentatif de la raucité de la voix, voire même de la présence de souffle (Eskenazi & al. 1990, Schoentgen & Guchteneere 1995 etc..). Le « *jitter ratio* » est la moyenne de la différence de période entre deux cycles consécutifs, rapportée à la période moyenne du signal. Enfin, le « *RAP* » est la moyenne de la différence de trois périodes consécutives, rapportée à la période moyenne du signal (Koike & al. 1977).

#### 3.6.4.4 Le shimmer

Pour les voix normales, le terme de « *shimmer* » correspond aux variations d'amplitude d'un cycle à un autre (cf. figure 3.14). De manière générale, le shimmer est exprimé en pourcentages. Aussi, pour les mêmes raisons évoquées précédemment, la mesure de ce paramètre s'avère toujours aussi problématique. Il existe en effet plusieurs définitions de ce dernier en fonction des calculs effectués et des logiciels utilisés. Le « *shimmer moyen* », exprimé en dB, caractérise les différences d'amplitude entre deux cycles consécutifs. Le « *shimmer factor* » représente, quant à lui, le shimmer moyen rapporté à l'amplitude moyenne du signal. Enfin, l'« *APQ* » calcule la moyenne des variations d'amplitude sur 11 cycles consécutifs rapportée à l'amplitude moyenne du signal (Koike & al. 1977). Comme nous l'avons précisé pour le jitter, le shimmer reste un bon indice dans le caractère *breathy* d'une voix (voire parfois pour la raucité) (Eskenazi & al. 1990 et Schoentgen 1995).



**Figure 3.14** – Représentation schématique du shimmer (schéma personnel réalisé avec TiKZ).

A l'instar du jitter, que nous avons décrit précédemment, l'intérêt de ces types de mesures est surtout d'établir des corrélations acoustico-perceptives avec les évaluations perceptives. Dans une synthèse de 2000, Woisard Bassols évoque ces interactions (Woisard Bassols 2000) :

	Forte corrélation	Faible corrélation
G	Shimmer et HNR	
R	Jitter	Shimmer
B	Shimmer, diminution du nombre d'harmoniques, hausse du débit phonatoire	Jitter
A	Jitter	Shimmer
S	Jitter et Quotient de fermeture (EGG)	PSG

**Table 3.3** – Synthèse des différentes corrélations acoustico-perceptives.

### 3.6.4.5 Les mesures de bruit de turbulence : le HNR et le NNE

Un troisième paramètre acoustique peut être mis en relation avec l'instabilité du signal pour les voix pathologiques. En effet, dans ce cas, l'instabilité du signal glottique peut se manifester par un bruit qui est surajouté (Teston 2004). Le but de cette mesure objective est de séparer la quantité de bruit pendant la phonation. Plusieurs méthodes ont été mises en oeuvre pour mesurer la partie apériodique du signal. La première consiste à mesurer le *HNR* (*harmonic noise ratio*). Ce paramètre, mesuré en dB, permet la recherche de bruit dans les harmoniques. Une forme ondulatoire moyenne du signal est obtenue à partir de 25 cycles voisins du signal. C'est la différence entre le signal de base et la forme ondulatoire qui permet d'obtenir un résidu nommé « bruit additif ». Si *H* correspond à l'énergie de l'onde moyennée et *N* est le bruit résiduel,

alors le rapport H/N calculé de manière régulière nous donne la valeur du HNR (Yumoto & al. 1982). La seconde méthode consiste à mesurer le *NNE* (*normalized noise energy*) à partir d'une méthode spectrale. Ce paramètre serait plus robuste dans les cas où les portions périodiques du signal acoustique ne seraient pas suffisamment longues. Il s'exprime par un rapport du bruit/composante totale du signal (Kasuya & al. 1986). Cependant, il est important de noter que la détection est également souvent problématique. En effet, les voix après LPS présentent des sous-harmoniques et des bruits de turbulence liés aux fuites glottiques, ce qui rend l'interprétation du spectre difficile.

## 3.7 Les analyses aérodynamiques

### 3.7.1 Le temps maximum de phonation (TMP)

Le *TMP* consiste à mettre en évidence l'utilisation de l'air pulmonaire en phonation. Celui-ci dépend, en effet, de la quantité d'air contenu dans les poumons, et par conséquent de la capacité vitale. On le mesure pendant l'émission d'une voyelle (le [a] tenu généralement) selon une hauteur et une intensité données après une inspiration forcée. Ces paramètres restent évidemment contrôlables bien que la majorité des études précisent que la voyelle tenue est produite à hauteur et intensité confortables (Yanagihara & al. 1966, Yanagihara & al. 1967 et Hirano & al. 1968). Traditionnellement, cette mesure est effectuée à l'aide d'un chronomètre, associé ou non à un sonomètre. Woodson & Cannito (1998) précisent qu'un TMP normal se situe aux alentours de 15 à 20 secondes. Ils ajoutent que, par définition, le TMP est plus court chez les femmes que chez les hommes. De nombreux auteurs ont néanmoins mis en évidence une grande variabilité inter et intra-individuelle du TMP de l'ordre de 30% (Isshiki & al. 1967, Hirano & al. 1968). En ce qui concerne les mesures après cancers du larynx, les diverses études montrent que les patients ont un TMP plus faible que la cohorte de témoins (Crevier Buchman & al. 1995, Crevier Buchman & al. 1998, Crevier Buchman 1999, Laccourreye & al. 1995, Hans 2003, Peretti & al. 2003, Lopez Llames & al. 2004, Vilaseca & al. 2007, Sjögren & al. 2009, entre autres). Les mesures du TMP associées aux valeurs de la capacité vitale (CV) d'un sujet permettent d'extraire un paramètre aérodynamique dérivé : *le quotient phonatoire*. Ce dernier correspond au rapport de la capacité vitale sur le TMP. Cette méthode est largement utilisée par les auteurs dans les études sur les voix pathologiques (Hirano & al. 1968, Peretti & al. 2003, Sjögren & al. 2008, Sjögren & al. 2009).

### 3.7.2 La pression sous-glottique (PSG)

Nous avons déjà évoqué le fait que l'air s'écoule d'une zone de pression plus haute vers une zone de pression plus basse. La pression, mesurée en Hecto-Pascal (hPa) ou en  $\text{cmH}_2\text{O}$ , correspond à la « force par unité d'air agissant perpendiculairement à une surface » (Baken 1987, Baken & Orlikoff 2000). En phonation, il est essentiel que la pression soit maintenue et qu'un équilibre avec la pression atmosphérique s'installe.

Ainsi, durant l'inspiration, la PSG est en-dessous de la pression atmosphérique. A l'inverse, durant l'expiration, la PSG est au-dessus.

Pour les études phonétiques, la mesure de la *PSG* peut être couplée avec d'autres paramètres acoustiques et/ou aérodynamiques et s'effectue au moyen d'investigations directes ou indirectes. La mesure directe de la *PSG* est invasive et correspond à une ponction trachéale (Netsell 1969). Les mesures indirectes peuvent être effectuées grâce à l'utilisation d'un pléthysmographe ou par estimation à l'aide des pics de pression intra-orale (*PIO*) pendant l'émission d'une occlusive non voisée. En effet, durant la phase de tenue des occlusives non voisées, la *PSG* est généralement égale à la *PIO* car les plis vocaux sont en abduction (Holmberg & al. 1988, Rothenberg 1982). Pour ce qui est de la tenue des occlusives voisées, la *PIO* est souvent inférieure à la *PSG* à cause du voisement (Warren 1996). Les diverses études mettent en évidence des valeurs de la *PSG* variables en raison des différentes corrélations existant entre *PSG*,  $F_0$  et intensité. De manière générale, en voix parlée à hauteur et intensité confortables, nous pouvons considérer que la *PSG* varie entre 7 et 10 cm $H_2O$ . Ces valeurs peuvent atteindre 15 cm $H_2O$  à forte intensité (Holmberg & al. 1988). Dans le domaine de la pathologie, les auteurs relèvent des valeurs de *PSG* plus élevées pour les patients dysphoniques par rapport aux sujets contrôles pour des intensités comparables (Giovanni & al. 2000, Teston 2004).

### 3.7.3 La pression intra-orale (*PIO*)

L'observation de la *PIO* permet de mieux appréhender le fonctionnement du conduit vocal, ce paramètre étant directement lié à l'articulation. Le recueil des données de ce paramètre peut s'effectuer de différentes manières sur un corpus commun du type CV, VCV ou CVC. La méthode développée par Subtelny & al. (1966) consistait à placer dans la bouche du sujet un transducteur de pression de 10 mm sur 3 mm collé au palais. Les auteurs ont ainsi pu établir une comparaison de valeurs entre une population féminine, une population masculine et une population d'enfants. Néanmoins, ce capteur a tendance à se décoller, ce qui peut être un biais dans la prise de mesures. Pour palier à ce problème, les différents auteurs ont, par la suite, proposé de placer un tube dans la cavité buccale. Une première méthode consiste à introduire un tube dans l'oropharynx par les fosses nasales jusqu'au voile du palais (Miller & Daniloff 1977). Une seconde méthode réside dans la fixation d'un tube en arrière des dents, le long de la gencive supérieure (Arkebauer & al. 1967). Enfin, la méthode la plus courante repose sur le placement d'un tube directement entre les lèvres du sujet (Smitheran & Hixon 1981, Holmberg & al. 1988, entre autres). Les résultats de la *PIO* restent très variables et dépendent largement des locuteurs, des corpus utilisés, des consonnes analysées, du matériel et de la méthode utilisés dans la prise de données. Ainsi, les auteurs notent que la *PIO* varie en fonction du voisement des consonnes : les occlusives non voisées ont une *PIO* plus importante que leurs correspondantes voisées (Subtelny & al. 1966, Arkebauer & al. 1967, Lisker 1970, Miller & Daniloff 1977, Stathopoulos 1986, Holmberg & al. 1988). Ils ajoutent que la *PIO* varie également en fonction du mode de production des consonnes : la *PIO* est plus élevée pour les occlusives non

voisées par rapport aux fricatives non voisées (Arkebauer & al. 1967). Aussi, la PIO varie plus ou moins avec le débit syllabique : Arkebauer & al. (1967) ont montré que la PIO augmentait avec le débit de parole alors que Brown & McGlone (1969) n'ont pas retrouvé cet effet dans leurs données. Certains précisent que la PIO varie avec le contexte syllabique : la PIO des occlusives semble plus élevée en attaque qu'en coda ( $\#_>V_C>\#$ ) (Arkebauer & al. 1967, Brown & al. 1970, Stathopoulos 1986). Aussi, certains auteurs ont remarqué que la PIO pourrait être dépendante de l'accentuation : Subtelny & al. (1966) et Malécot (1968) ont mis en évidence que les syllabes accentuées étaient produites avec une PIO supérieure. En revanche, Brown & al. (1970) n'ont pas retrouvé de différence significative entre ces deux types de syllabes (accentuées vs non accentuées). Certains auteurs notent également une influence du contexte vocalique sur les mesures de la PIO. Les consonnes occlusives suivies d'une voyelle haute — et en particulier le [u] — auraient des valeurs de PIO plus importantes que leurs homologues suivies d'une voyelle basse (Karnell & Willis 1982, Netsell & al. 1991, entre autres). Cette observation est également vérifiée pour la voix chuchotée (Klich 1982). Enfin, les auteurs ont également démontré une corrélation positive entre intensité vocale et PIO : la PIO augmente avec une intensité plus forte (Subtelny & al. 1966, Stathopoulos 1986, Holmberg & al. 1988, Hiss & al. 2001, Hans 2003). Cependant, malgré une grande variabilité inter et intra-locuteurs reconnue dans les mesures de la PIO, les auteurs s'opposent quant à l'effet du sexe dans ces variations. Ainsi, Subtelny & al. (1966) ont mis en évidence que les sujets féminins avaient une PIO supérieure par rapport aux sujets masculins. Au contraire, Holmberg & al. (1988) ont comparé un groupe de 20 femmes à un groupe de 25 hommes et n'ont pas trouvé de différence significative entre les deux. De la même manière, dans son étude portant sur les paramètres aérodynamiques chez 5 sujets masculins et 5 sujets féminins, Hans (2003) a montré que la PIO ne variait pas en fonction de l'âge et du sexe. Stathopoulos (1986), quant à elle, en comparant 10 femmes à 10 hommes, a montré que les sujets masculins avaient des valeurs plus hautes que les sujets féminins, sans que cela soit statistiquement significatif. Dans le domaine de la pathologie, très peu d'études font état des variations de la PIO après un cancer du larynx. Les auteurs s'accordent cependant sur le fait que les valeurs de la PIO sont plus élevées chez les patients par rapport aux témoins (Zeitels 1995, Hans 2003, Mirghani 2009). Dans sa thèse, Hans (2003) a également démontré que le paramètre de la PIO était le plus pertinent dans la distinction des différentes catégories de laryngectomies partielles. Ainsi, ce paramètre a tendance à augmenter proportionnellement à l'étendue de la résection.

### 3.7.4 Le débit d'air buccal (DAB)

Le *débit d'air buccal*, exprimé en ml/s ou en  $dm^3/s$ , est la quantité d'air expiré par la bouche par unité de temps. Ainsi, ce paramètre permet de déterminer la consommation d'air nécessaire à une production vocale (Baken 1987, Schutte 1992, Baken & Orlikoff 2000). En phonation, les différentes constrictions glottiques et supra-glottiques produisent un orifice déterminant ce débit. Pour Dejonckere (1985) et Schutte (1992),

le DAB permet d'établir le rendement énergétique du larynx, c'est à dire l'*efficience laryngée*.

La prise de mesures du DAB varie selon les études : spiromètre (mesure le volume d'air déplacé), pléthysmographe (mesure du volume intra-thoracique et des résistances ventilatoires) ou encore pneumotachographe (mesure du mouvement spécifique de chaque étage respiratoire). En voix parlée, à intensité et hauteur confortables, les valeurs du DAB définies par les différentes recherches sont extrêmement variables (entre 125 et 170 ml/s) en raison des conditions d'enregistrement (sujet assis/debout etc.), de la taille variable des populations, des matériels utilisés et de la grande variabilité intra et inter-individuelle (Hans 2003). De plus, à notre connaissance, il n'existe qu'une unique base de données chez les locuteurs français, la majorité des études étant japonaises (cf. Isshiki 1965, Yanagihara & Koike 1967, Hans 2003). Les auteurs notent toutefois un débit d'air plus élevé pour les hommes par rapport aux femmes (Holmberg & al. 1988, Schutte 1992, Sulter & Wit 1996). D'autres études mettent également en évidence la variation des valeurs du DAB en fonction du voisement des consonnes adjacentes, du mode d'articulation de ces consonnes, de la position dans le mot ou de la qualité propre de ces voyelles (Gilbert 1973, Trullinger & Emanuel 1983, Stathopoulos 1984, Stathopoulos & Weismer 1985, Higgins & al. 1998). En outre, la relation entre  $F_0$ , intensité, PSG et débit d'air reste complexe et souvent incomprise. En 1965, Isshiki avait démontré qu'il n'y avait pas de corrélation entre DAB,  $F_0$  et intensité. L'augmentation de la PSG entraînait une augmentation de l'intensité mais pas forcément des débits d'air (Isshiki 1965). Nous pouvons néanmoins affirmer aujourd'hui que le DAB reste davantage corrélé à l'intensité qu'à la  $F_0$  puisque le DAB a tendance à augmenter avec l'élévation de l'intensité sonore (Brown & McGlone 1974, Hans 2003). En ce qui concerne le domaine de la pathologie, Hans (2003) a mis en évidence une différence significative entre témoins et patients traités par laryngectomies partielles pour la voyelle tenue [i] comme pour la production du logatome [pipi]. Ainsi, le DAB semble statistiquement plus élevé pour les différents patients par rapport aux sujets contrôles de même classe d'âge.

### 3.8 Les analyses morphologiques

Nous avons souligné que plusieurs mesures peuvent être associées afin d'étudier une voix pathologique — et dans le cas présent — la voix après cordectomie. Les analyses morphologiques sont des approches qualitatives complémentaires qui permettent l'appréciation du comportement laryngé. Ces examens permettent d'analyser la dynamique du larynx et la présence ou non d'une ondulation muqueuse. Différentes méthodes d'investigations s'offrent aux chercheurs. Néanmoins, nous n'avons retenu que les deux techniques les plus couramment utilisées dans les études actuelles de la morphologie du larynx, à savoir la fibroscopie et l'électroglottographie.

### 3.8.1 La fibroscopie

C'est au japon, dans les années 1970, que le *nasofibroscope* prit son essor (Sawashima & Hirose 1968). Les progrès de l'informatique liés au développement technique de ce type de matériel ont permis une utilisation plus aisée et plus puissante de cette méthode. La prise de données s'effectuant directement par l'introduction d'une fibre optique dans les fosses nasales, l'examen n'entrave pas les mouvements articulatoires, ce qui représente un avantage certain dans l'observation des mouvements glottiques en phonation. Il est également possible de relier le nasofibroscope à un générateur de lumière stroboscopique. Le déclenchement d'un flash permet d'obtenir des images à un moment donné du cycle vibratoire. La succession de ces images prises à des instants différents d'un cycle vibratoire permet de reconstituer un mouvement ralenti des plis vocaux. Le mouvement lent ainsi capturé permet une observation affinée et précise de l'ondulation des plis vocaux, des mouvements d'abduction et d'adduction, et de manière générale du comportement du système laryngé. Ainsi, son application clinique réside particulièrement dans l'étude des mouvements des arytenoïdes, de la configuration glottique pendant les phases de respiration et de phonation, de l'existence d'une activité supraglottique et du comportement des plis vocaux. Par conséquent, cette technique permet de mettre en évidence les mécanismes de compensation établis après résection d'un et/ou des deux plis vocaux (Pech & al. 1988, Crevier Buchman & al. 1999, Mirghani 2009, entre autres). La fibroscopie peut également être couplée avec des enregistrements vidéos, acoustiques et EGG, ce qui permet d'établir des corrélations entre les différentes données.

### 3.8.2 L'électroglottographe (EGG)

Inventé par Fabre (1956), l'*EGG* représente un impédancemètre, c'est-à-dire qu'il permet la mesure de l'impédance — ou résistance — électrique entre deux électrodes placées de part et d'autre du cou. Cette méthode d'investigation non invasive est un outil précieux pour observer le degré d'accolement des plis vocaux au cours du temps. Un courant de haute fréquence et de faible intensité circule entre les deux électrodes et permet la mesure de la surface de contact entre les plis vocaux. Si la glotte est fermée, le courant passe facilement entre les deux électrodes, ce qui tend à éléver le signal EGG. Par contre, lorsque la glotte est ouverte, le courant a davantage de difficulté à passer d'une électrode à l'autre, ce qui amenuise le signal. Ainsi, l'*EGG* permet donc de renseigner le chercheur sur les phases d'ouverture et de fermeture des plis vocaux et par conséquent d'en déduire des mesures fiables de voisement, de quotient de fermeture et d'ouverture que nous aborderons dans la suite de cette thèse, de  $F_0$  et de ses variations. Contrairement aux idées reçues, le signal résultant ne correspond en rien au signal laryngé. Il n'est qu'une représentation des variations de surface d'accolement des plis vocaux et dépend largement du type d'appareil utilisé, des conditions d'enregistrement, de l'anatomie du sujet et bien évidemment de la connaissance de l'expérimentateur (Teston 2004). De manière générale, un signal brut se présente comme une onde constituée de deux composantes distinctes : une modulation rapide, synchrone

avec les cycles d'ouverture et de fermeture et une modulation lente, révélatrice des mouvements du larynx. Lors de l'analyse du voisement par exemple, il convient de « lisser » le signal originel par l'utilisation de filtrages.

Finalement, l'EGG est un outil relativement fiable pour l'étude de la périodicité du signal, non seulement d'un point de vue quantitatif mais aussi d'un point de vue qualitatif car il nous renseigne sur la qualité de voix des sujets (Fourcin 2000).

### 3.9 Conclusion

Dans ce chapitre, nous avons porté notre attention sur la notion même de cordectomie laser par voie endoscopique et son impact sur la qualité de voix. La première partie de ce chapitre était consacrée à la description des différentes techniques chirurgicales. Nous avons pu mettre en évidence que l'étendue du cancer des plis vocaux vers les structures adjacentes était un élément essentiel dans le choix de la chirurgie la plus adaptée. Dans une seconde partie, nous avons tenté d'établir une revue de la littérature détaillée des divers articles portant sur les résultats fonctionnels de la voix après cordectomie. Nous avons pu mettre en évidence l'hétérogénéité de ces recherches tant du point de vue des populations étudiées, des corpus utilisés ou des méthodologies adoptées. Ainsi, la plupart des travaux se base sur l'étude d'une cohorte de patients disparate du point de vue des traitements chirurgicaux, ce qui représente certainement un biais dans l'exposition des résultats. En effet, il semble évident que la voix sera d'autant plus perturbée que l'étendue de l'exérèse sera importante. De plus, malgré le fait que l'ensemble des auteurs soulignent l'importance d'une analyse multiparamétrique dans l'étude de ces voix, nous avons pu mettre en évidence que tous ne couplent pas les différentes analyses mises à leur disposition. Aussi, les corpus, les paramètres étudiés et les populations de référence divergent d'une recherche à une autre, ce qui rend les comparaisons et les discussions difficiles. Enfin, dans une dernière partie, nous avons voulu exposer les différentes analyses et mesures utilisées généralement en pathologie. Les paramètres évoqués ne sont pas exhaustifs et ne seront pas tous décrits dans cette thèse, mais nous avons néanmoins souhaité dresser un panorama le plus clair possible des outils disponibles. Ainsi, les analyses perceptives, les plus anciennes, restent des références majeures pour les phonéticiens et cliniciens dans l'analyse des voix pathologiques. Aujourd'hui, ces dernières sont généralement couplées avec des mesures plus objectives, représentées par les analyses acoustiques. Des mesures aérodynamiques et morphologiques permettent également de mieux comprendre le comportement laryngé de ces patients.

**En résumé...**

- Nous reconnaissons cinq types de cordectomies laser par voie endoscopique.
- Les recherches sur la qualité vocale en postopératoire mettent en évidence des populations hétérogènes et des méthodologies variées, ce qui rend les comparaisons transversales difficiles.
- Les auteurs reconnaissent, néanmoins, l'importance d'une analyse multiparamétrique pour l'évaluation de la qualité vocale.



# Population et Corpus

"Dans une langue que nous savons, nous avons substitué  
à l'opacité des sons la transparence des idées."

— Marcel Proust

## Résumé

Ce chapitre offre une description des populations étudiées, des corpus et des matériaux utilisés pour nos différentes études. La recherche en phonétique clinique pose de nombreux problèmes dans le recueil d'une grande quantité de données : (i) la fatigabilité physique et morale des patients, (ii) la durée des protocoles d'étude qui s'avère être un frein à l'expérimentation, (iii) des conditions d'enregistrement, qui ne sont, le plus souvent, pas optimales en milieu hospitalier et (iv) le nombre restreint de patients traités par cordectomie de type II-III. Dans notre travail de thèse, nous avons pu enregistrer 10 patients masculins francophones âgés de 47 à 80 ans (moyenne 66,7 ans). Les sujets ont été enregistrés entre 6 mois et 1 an postopératoire. Dix témoins masculins francophones, âgés de 34 à 67 ans (moyenne 49,2 ans), ont servi de référence. Nous avons constitué un corpus composé, d'une part, de trois voyelles tenues [a, i, u] et, d'autre part, de logatomes de type CVCVCVC, où C représente les 16 consonnes du français et V représente l'une des trois voyelles [a, i, u]. Ces logatomes ont été utilisés dans le but d'évaluer les qualités des consonnes dans deux positions syllabiques : attaque en position d'initiale absolue (#CV) et attaque en position intervocalique (V\_\_V). Tous les sujets ont été enregistrés dans des conditions similaires et le recueil de données a été effectué lors de deux sessions. La première session était dédiée aux enregistrements acoustiques et la seconde nous a permis d'acquérir des données aérodynamiques et EGG à l'aide de la station EVA2 (Galindo & al. 1989, Teston & al. 1999). Les analyses acoustiques ont été réalisées avec le logiciel Praat (Boersma & Weenink 1996) et les données aérodynamiques et EGG ont été traitées avec le logiciel Phonetid (Ghio 2002).

## 4.1 Introduction

La recherche clinique nécessite d'acquérir un grand nombre de données acoustiques, aérodynamiques ou électroglottographiques, afin de mettre en évidence la validité du matériel acoustique et de dégager des tendances statistiques les plus fiables possible. Nous verrons cependant que les prises de mesures sur des voix pathologiques s'avèrent difficiles. En outre, une étude complète de la voix et de la parole suggère l'utilisation de différents matériaux phonétiques. En effet, dans la littérature sur le sujet, deux types de matériaux sont couramment utilisés : les voyelles isolées, qui permettent de mettre en évidence les différents paramètres de perturbation glottique (*jitter* et *shimmer*) et la parole continue, qui donne un aperçu des phénomènes suprasegmentaux (Laver & al. 1992).

Ce chapitre s'articulera autour de trois axes majeurs. Dans une première partie, nous décrirons nos populations de patients et de témoins. Nous mettrons l'accent sur les motivations qui nous ont amené à choisir une cohorte de sujets dysphoniques et sur les difficultés rencontrées pour nos enregistrements acoustiques et aérodynamiques. Notre seconde partie sera consacrée à la motivation et la justification d'un protocole orienté pour la recherche clinique. Nous verrons qu'il est important de coupler la production de voyelles tenues à de la parole continue pour avoir une vision générale de la voix et de la parole de ces sujets pathologiques. Enfin, notre dernière partie présentera de manière globale notre corpus et nos outils de traitement et d'analyse. En effet, notre corpus comporte essentiellement des voyelles tenues produites en fonction des tâches demandées, ainsi que des logatomes. Ce corpus a été acoustiquement acquis en utilisant le même matériel et dans des situations similaires. Aussi, ces matériaux phonétiques ont été ensuite enregistrés à l'aide de la station EVA2 afin d'obtenir des données aérodynamiques et électroglottographiques (EGG). Tous les enregistrements ont été traités avec les mêmes logiciels : Praat pour les analyses acoustiques et Phonetid pour les analyses aérodynamiques et EGG.

## 4.2 Population

Les travaux de recherche requièrent une quantité importante de données afin d'établir des traitements statistiques mettant en évidence des variabilités inter et intralocuteurs et la validité des analyses effectuées. Néanmoins, la recherche en phonétique clinique pose de nombreux problèmes dans le recueil d'une grande quantité de données.

Premièrement, les patients enregistrés ne sont pas volontaires et ne sont pas aussi motivés que des sujets sains. Ce manque d'entrain est en grande partie lié à leur fatigabilité physique et morale.

Deuxièmement, la maladie reste un frein dans l'expérimentation. En effet, les protocoles ne doivent pas être trop longs et trop complexes. La compréhension des consignes est parfois difficile et la fragilité physique et psychologique prennent souvent le dessus.

Troisièmement, l'expérimentateur reste contraint aux conditions d'enregistrement. Les patients choisis sortaient toujours en amont de consultation de phoniatrice ou de chirurgie. La fatigue, la lassitude de la journée, le changement de salle, le transport étaient bien souvent des freins au bien-être de ces patients.

Quatrièmement, les patients habitent parfois loin, sont parfois pressés et sont parfois inintéressés pour un ultime enregistrement. De plus, la population ciblée (patients traités par cordectomie de type II-III) reste assez restreinte. Le choix des patients à enregistrer n'a donc pas été simple et les reconvocations d'autant plus difficiles.

Dans cette section, nous présenterons donc les patients qui ont contribué à nos différentes études et notre population témoin de référence.

## 4.2.1 Les patients

### 4.2.1.1 Période de recueil de données

Les enregistrements du signal de parole ont véritablement débuté en 2009 et se sont poursuivis jusqu'en 2011. Les données aérodynamiques ont été acquises durant la seconde moitié de l'année 2011. Le début tardif des enregistrements s'explique par la mise en place de notre protocole expérimental et par l'attente du matériel nécessaire. Le retard des enregistrements est dû à la recherche de patients correspondant à notre cible, à savoir des patients opérés par cordectomie laser CO<sub>2</sub> entre 6 mois et 1 an post-opératoire et aux reconvocations difficiles de ces derniers. Tous les enregistrements ont eu lieu dans le service ORL et de Chirurgie de la Face et du Cou de l'Hôpital Européen Georges Pompidou (HEGP) à Paris.

### 4.2.1.2 Le choix des patients

Notre étude concerne des sujets de sexe masculin et francophones. La contrepartie féminine de cette population reste mineure en comparaison à la population masculine. De plus, les femmes assument moins le traumatisme de l'intervention et la nouvelle voix engendrée. Pour ces raisons, nous avons écarté les patientes.

Au début des enregistrements, notre protocole n'était pas définitivement établi. Les enregistrements acoustiques se succédaient sans réellement prendre en considération les types de chirurgies. La prise de données a continué un temps dans ce sens. Lors du début des enregistrements aérodynamiques et surtout après les premières analyses, nous nous sommes rapidement aperçus que la sélection d'un unique type de chirurgie serait un atout pour ce travail. En effet, la classification des différents patients selon leur type de chirurgie était impossible. Nous avions un nombre très restreint de patients traités par cordectomie de type I ou de type IV/V par rapport aux patients traités par cordectomie de type II/III. En outre, le manque de temps ne nous permettait plus de reprendre d'autres enregistrements. En conséquence, nous avons fait le choix de restreindre notre population de patients aux cordectomies de type II/III (avec une prédominance pour le type II : nous avons 8 patients traités par une cordectomie de type II et 2 patients traités par une cordectomie de type III). Pour éviter les biais et garder

une cohérence dans notre analyse, nous avons gardé uniquement les patients à 6 mois à 1 an postopératoire. Ce choix a réduit considérablement notre effectif car seule une dizaine de patients de l'HEGP correspondait à tous ces critères, ce qui explique notre faible population de patients.

#### 4.2.1.3 L'âge

Le tableau 4.2 résume notre cohorte de patients. Notre population de 10 patients masculins et francophones est âgée de 47 à 80 ans (moyenne 66,7 ans ; médiane 66,5 ans).

#### 4.2.1.4 Les lésions

Tous les patients étaient venus consulter pour une dysphonie persistante de deux mois ou plus. Les diagnostics ont été effectués en consultation par nasofibroscopie. Les lésions du plan glottique étaient classées en stade pré-cancéreux ou en carcinomes *in situ* de type T1aN0. Seul un pli vocal était touché avec une mobilité cordale et de l'aryténioïde intacte. Aucun patient ne présentait d'envahissement ganglionnaire ou de métastase.

#### 4.2.1.5 Antécédents et facteurs de risque

D'après les dossiers médicaux de ces patients, seuls trois d'entre eux présentaient une intoxication tabagique. Nous n'avons pas de renseignement quant à leur consommation d'alcool. Les pathologies associées sont résumées dans le tableau 4.2. Trois des patients enregistrés avaient des antécédents de cancers (cancer de la thyroïde, mélanome, lymphome). Un patient présentait une papillomatose laryngée<sup>1</sup>. Enfin, un patient avait fait état d'une insuffisance rénale et de problèmes cardiaques.

#### 4.2.1.6 Traitement et suites postopératoires

Comme nous le précisions plus haut, tous nos patients ont subi une cordectomie au Laser CO2 de type II ou III sur l'un des plis vocaux. Ce type de chirurgie resèque l'épithélium, l'ensemble de la *lamina propria* et une partie du muscle vocal dans le cas d'une cordectomie de type III (cf. chapitre 3). Tous les patients ont été opérés en ambulatoire et aucune complication postopératoire n'a été constatée. Seul le patient n°7 a nécessité une reprise chirurgicale du fait de la récidive de sa papillomatose laryngée. Tous les patients ont été orientés vers une rééducation orthophonique ; le nombre de séances variait de cinq à une quinzaine.

1. La papillomatose laryngée est une maladie d'origine virale causée par le virus HPV. Des lésions appelées papillomes apparaissent. Ces papillomes représentent des proliférations tumorales benignes qui envahissent le larynx ou les poumons. Les papillomes doivent être retirés au laser mais les récidives sont fréquentes. De plus, la dégénérescence de ces tumeurs est possible et l'évolution en cancer n'est pas rare (cf. [www.papillomatose.com](http://www.papillomatose.com)).

#### 4.2. Population

93

N°	Age	Tabac	Profession	Antécédents	Lésions	Cordectomy	Intervention
1	74	ns	Retraité	Thyroïdectomie totale suivie d'une paralysie récurrentielle du pli vocal gauche	Lésions d'aspect leucoplasique de type T1aN0 (pli vocal droit)	Type II	le 6/01/10
2	77	ns	Retraité	ns	Etat précancéreux de type T1aN0 (pli vocal gauche)	Type II	le 22/01/10
3	60	+	Retraité	Pas d'antécédents particuliers	T1aN0 (pli vocal droit)	Type II	le 23/12/09
4	68	+	Retraité	ns	T1aN0 (pli vocal droit)	Type III	le 11/02/10
5	79	ns	Retraité	ns	Lésion tumorale de type T1aN0 (pli vocal droit)	Type II	le 17/02/10
6	80	+	Retraité	Insuffisance rénale, infarctus du myocarde, coronaropathie	Carcinome épidermoïde de type T1aN0 (pli vocal droit)	Type III	le 6/02/09
7	54	ns	ns	Mélanome et papillomatose laryngée	Carcinome épidermoïde in situ de type T1aN0 (pli vocal droit)	Type II	le 06/03/09 et reprise le 12/02/10
8	47	-	ns	Pas d'antécédents particuliers	Lésion bourgeonnant de type T1aN0 (pli vocal gauche)	Type II	le 4/11/09
9	65	ns	Retraité	ns	Lésion érythémateuse et leucoplasique de type T1aN0 (pli vocal gauche)	Type II	le 15/07/09
10	63	ns	Retraité	Lymphome	Lésion leucoplasique de type T1aN0 (pli vocal droit)	Type II	le 13/01/10

**Table 4.2** – Tableau récapitulatif des patients. La consommation de tabac est indiquée avec un "+" et un "-" si il n'y en a pas. L'abréviation "ns" suppose que nous n'avons pas de renseignements supplémentaires.

### 4.2.2 Les témoins

Nous avons étudié 10 sujets masculins francophones sains, sans pathologie vocale et n'ayant jamais été opérés du larynx. Les témoins sont âgés de 34 à 67 ans (moyenne 49,2 ans ; médiane 51 ans). Les enregistrements acoustiques ont débuté en décembre 2010 et se sont terminés en janvier 2011. En ce qui concerne les enregistrements aérodynamiques, ceux-ci ont été faits conjointement à ceux des patients, c'est-à-dire durant la seconde moitié de l'année 2011. Ces témoins ont été recrutés parmi les membres de nos proches et parmi notre entourage professionnel ou amical. Nous n'avons pas pu obtenir tous les renseignements concernant les antécédents médicaux du fait de notre connaissance partielle de certains témoins (cf. tableau 4.3).

N°	Age	Tabac	Profession
1	50	—	Technico-commercial
2	59	—	Cadre bancaire
3	34	—	Architecte
4	34	—	Ingénieur
5	51	—	Infirmier
6	55	—	Cadre bancaire
7	35	+	Ingénieur de recherche
8	67	+	Directeur de recherche
9	58	—	Retraité
10	54	—	Commercial

**Table 4.3** – Tableau récapitulatif des témoins. La catégorie professionnelle est précisée. La consommation de tabac est stipulée avec un “+” et un “—” s'il n'y en a pas.

### 4.3 Motivation d'un corpus clinique

Dans cette section, nous évoquerons les supports phonétiques disponibles dans l'étude des voix pathologiques. La plupart des recherches cliniques se basent essentiellement sur l'analyse de voyelles tenues, des syllabes ou des mots et, enfin, sur des phrases. Dans un premier temps, nous décrirons les études sur les voyelles tenues isolées, et en particulier les voyelles dites « cardinales » qui permettent principalement d'extraire certains paramètres essentiels puisque ce sont les voyelles les moins complexes d'un point de vue phonologique. Dans un second temps, nous montrerons la nécessité de coupler ces analyses avec des corpus plus larges portant sur la parole afin d'être au plus proche des situations de communication quotidiennes. Finalement, nous détaillerons le corpus enregistré, ainsi que les outils que nous avons utilisés pour son traitement.

### 4.3.1 Les voyelles tenues

Les voyelles longues et tenues représentent les performances vocales les plus stables. En effet, celles-ci peuvent être facilement réalisées, imitées et mesurées. D'un point de vue articulatoire, la production de voyelles nécessite un simple accrolement entretenu des plis vocaux afin de maintenir le voisement. Cependant, De Krom (1994) note, pour les voix normales, que les seules modifications des différents paramètres acoustiques peuvent être relevées au niveau de l'attaque (*onset*) et de la fin de la voyelle (*offset*). Le temps d'établissement et l'arrêt du voisement requièrent une mise en place précise des articulateurs et en particulier des plis vocaux. L'auteur précise par conséquent que les mesures directes doivent être prises pendant la période la plus stable du segment. Pour les voix pathologiques, de Krom (1994) ajoute que les altérations du signal vocal n'interviennent pas uniquement sur la phase de pré-voisement et pendant la fin de la voyelle mais sur l'ensemble de la production de cette dernière.

En outre, comme nous le précisons plus haut, les voyelles permettent une analyse acoustique relativement simple. Les paramètres les plus fréquemment analysés sont les variations de la fréquence fondamentale  $F_0$ , le jitter (c'est-à-dire l'instabilité vibratoire de  $F_0$  d'un cycle à un autre), le shimmer (c'est-à-dire l'irrégularité en amplitude des cycles vibratoires), le rapport harmoniques sur bruit (HNR), la richesse en harmoniques, etc. (Hollien & al 1973). Finalement, les analyses acoustiques informatisées porteront donc principalement sur les parties stables et voisées des voyelles pour éviter de prendre en compte les attaques particulières de chaque locuteur (attaques dures ou soufflées, cf. De Krom 1994), et pour vérifier la régularité ou non du voisement (Hollien & al. 1973).

### 4.3.2 Corrélates acoustico-articulatoires des voyelles : un argument dans le choix du protocole d'étude

Nous avons choisi d'étudier les voyelles [a], [i] et [u]. D'un point de vue articulatoire, [i] est produit avec une articulation antérieure et une faible ouverture de la bouche, [u] est produit avec une articulation postérieure et une faible ouverture de la bouche et [a] est produit avec une ouverture maximale de la bouche. Dans la lignée de ses travaux sur la théorie source-filtre qui ont permis la mise en évidence de l'interdépendance entre niveaux articulatoires et acoustiques, Fant (1960) a proposé un modèle du conduit à quatre tubes afin d'étudier la relation entre fréquences de résonances et cavités lors de la production de voyelles. Ce modèle prédit et explique le lien entre les configurations particulières du conduit vocal et les fréquences de résonances — appelées formants — des voyelles. Ainsi, l'auteur a démontré que la structure formantique des voyelles était corrélée aux variations de volumes et de formes des cavités du trac-tus vocal (Kent 1993). Selon l'auteur, les modélisations des configurations vocaliques du conduit vocal pourraient se résumer selon trois paramètres : *degré d'ouverture des lèvres, point d'articulation de la langue et labialisation*.

Selon les représentations articulatoires traditionnelles, seuls les deux premiers formants (F1/F2) sont caractéristiques des voyelles. De manière générale, les valeurs for-

mantiques sont modifiées par la forme du conduit vocal. Néanmoins, certains restent plus sensibles aux mouvements des différents articulateurs. Ainsi, l'ouverture de la mandibule et/ou l'abaissement de la langue ont pour conséquence une augmentation de F1. De même, le F2 varie en fonction de la position de la langue et de la protrusion des lèvres lorsque la langue est en arrière. Néanmoins, certains ajustements compensatoires permettent d'obtenir les mêmes résultats acoustiques. En ce sens, la théorie quantique de Stevens (1989) stipule qu'il existe des régions dans le conduit vocal — dites « quantiques » — où une imprécision de l'articulation n'entraîne pas de changements acoustiques majeurs. Au contraire, il existe également des régions dites « d'instabilité » où de légères modifications dans l'articulation entraînent de fortes perturbations dans le résultat acoustique.

D'un point de vue perceptif, Delattre (1958) pointe sur l'importance du F3 — en français du moins — pour établir des contrastes phonémiques. Ainsi selon lui, [i] et [y], deux voyelles fermées et antérieures, ne peuvent se distinguer par les valeurs uniques de F1 et F2. Seul le F3, qui est sensible à la longueur de la cavité antérieure lorsque la langue est en position antérieure, pourra les opposer. En effet, le jeu des lèvres est suffisant pour abaisser F3 et le regrouper avec F2 pour la production du [y].

Finalement, le français semble favoriser les voyelles ayant une forte concentration d'énergie dans une zone particulière du spectre. Cette proéminence d'énergie, conséquente à un rassemblement de deux formants, constitue la base de la théorie des *voyelles focales* (Schwartz & al. 1997). Ainsi, la voyelle [i] montre une convergence de F3/F4, la voyelle [y] montre un rapprochement de F2/F3 et la voyelle [u] est identifiable au regroupement de F1/F2.

Les voyelles [i] et [u] représentent des voyelles cardinales et focales, nous avons donc fait le choix de conserver ces segments pour ces études acoustiques. En ce qui concerne la voyelle [a], celle-ci représente la voyelle la plus ouverte et implique par conséquent un F1 très élevé.

### 4.3.3 La parole en contexte

La parole, bien que non spontanée puisque composée de non-mots lus en ce qui nous concerne, présente l'avantage de refléter des situations de phonation normale. Elle requiert des tâches dynamiques et successives mettant en jeu l'ensemble des muscles laryngés. Ainsi, les phases d'adduction et d'abduction des plis vocaux, associées aux variations mélodiques, aux rapides transitions formantiques et aux événements linguistiques propres aux locuteurs sont caractéristiques de ce type de corpus (Schoentgen & al. 1998). La prise en compte de ces différents paramètres semble toutefois difficile. En effet nombre de ces corrélats sont difficilement identifiables et intimement liés les uns aux autres. Aussi, d'un point de vue segmental, il n'est pas toujours aisé de déterminer le voisement d'un signal et l'articulation des sons prend souvent du temps dans la mise en place. De ce point de vue, Hammarberg & al. (1980) adoptent une vision plus large quant à l'utilisation de la parole comme support phonétique et préconisent l'étude unique des variations de hauteur et d'intensité pour qualifier la fonction vocale d'un sujet.

En ce qui concerne la pathologie, Schoentgen & al. (1998) précisent que les corpus de parole peuvent distinguer les voix normales des voix pathologiques, mais que ces derniers sont insuffisants pour montrer l'existence d'une pathologie laryngée particulière. Dans ce sens, De Krom (1994) précise que la situation de parole permet de mettre en évidence la pathologie de manière plus pertinente que l'analyse des voyelles tenues, mais l'analyse en est plus difficile à cause des nombreuses perturbations. Il ajoute que ces corpus de parole restent essentiels dans l'analyse clinique car les patients présentant un trouble de la voix ne peuvent cacher leur dysphonie ou adopter un comportement compensatoire.

## 4.4 Présentation succincte du corpus

Dans cette section, nous présenterons les différentes données que nous avons pu acquérir d'un point de vue acoustique, aérodynamique et EGG.

### 4.4.1 Données acoustiques

Lors de nos enregistrements acoustiques, nous avons pu recueillir des voyelles tenues, des logatomes, un texte et des phrases lus. Dans la logique de cette thèse, nous avons choisi d'analyser les voyelles tenues et des logatomes (non-mots). La production des voyelles tenues se décline de manière différente selon le but des expérimentations. Ces différentes méthodologies seront abordées dans la suite de ce travail. Dans cette section, nous décrirons uniquement le type de voyelles utilisées et la manière dont elles ont été enregistrées. Aussi, nous exposerons la constitution de nos logatomes et également la manière dont ils ont été recueillis. Les précisions quant aux traitements et aux analyses de ces logatomes seront également développées ultérieurement.

#### 4.4.1.1 Les voyelles tenues

Nous avons enregistré les voyelles [a, i, u] dans différentes conditions d'expérimentation que nous évoquerons dans la suite de cette thèse. Les situations et le matériel d'enregistrement restent similaires pour nos deux populations. Tous les enregistrements ont été effectués dans des pièces acoustiquement calmes, les patients à l'HEGP et les témoins au domicile personnel de l'expérimentateur. Les sujets étaient assis de manière confortable et portaient un micro-casque de type AKGC520 orienté à 45° de la bouche (soit à une distance de 3-4 cm des lèvres), relié à une carte son EDIROL FA 66 dont la fréquence d'échantillonnage était réglée par défaut à 44100 Hz (16 bits). Afin d'effectuer des comparaisons inter-groupes, nous avons réglé et déterminé dès le premier enregistrement des témoins et des patients le niveau du gain sur la carte. Il est donc important de préciser que ce niveau n'a jamais été retouché entre les différents enregistrements. Toutes les prises de données ont été réalisées avec le logiciel Soundforge et le traitement acoustique à l'aide du logiciel Praat.

#### 4.4.1.2 Les logatomes

Les logatomes étaient du type CVCVCVC où C représente les 16 consonnes du français et V représente l'une des trois voyelles [a, i, u]. Ces logatomes ont été constitués dans le but d'évaluer les qualités des consonnes dans deux positions syllabiques : attaque en position d'initiale absolue (#\_) et attaque en position intervocalique (V\_V) sur lesquelles nous reviendrons ultérieurement. Chaque sujet avait pour consigne de lire une fois la liste de logatomes, à un tempo modéré.

#### 4.4.2 Données aérodynamiques

Les données aérodynamiques, qui seront également évoquées plus loin, ont été recueillies en suivant exactement le même protocole. Ainsi, nous avons pu acquérir des voyelles tenues, des logatomes, un texte et des phrases lus. Une fois encore, nous n'avons retenu que les voyelles tenues, et en particulier [a, i, u] ainsi que les logatomes décrits précédemment pour nos analyses. Ces enregistrements ont été effectués à l'aide de l'appareillage d'Evaluation Vocale Assistée EVA2 (Galindo & al. 1989, Teston & al. 1999), à l'hôpital Pompidou, dans une pièce acoustiquement calme. La méthodologie étant un peu plus complexe que celle des enregistrements acoustiques, nous avons opté pour une description plus détaillée du matériel et de la prise de données (cf. section 4.5.2).

#### 4.4.3 Données EGG

Comme précédemment, nous avons pu enregistrer des voyelles tenues, des logatomes, un texte et des phrases lus. Pour ce travail de recherche, nous avons choisi de conserver pour les analyses les voyelles tenues et la parole en contexte (logatomes). L'acquisition EGG a été réalisée en même temps que la prise de données aérodynamiques. Pour des raisons logistiques, nous avons pu enregistrer l'ensemble de nos patients mais un unique sujet témoin, lequel nous servira de référence. Les signaux EGG ont été enregistrés à l'aide du laryngographe portable (Laryngograph Processor) développé par Laryngograph Ltd. Celui-ci était directement relié à la plateforme d'expérimentation EVA2, ce qui garantissait une synchronisation des différents signaux. Les électrodes, permettant de capter le mouvement des plis vocaux pendant la phonation, étaient placées de part et d'autre du cou, près de la glotte.

### 4.5 Outils d'analyses

Dans cette section, nous aborderons nos outils d'acquisition et de traitement des données, en particulier le logiciel de traitement de la parole Praat et la station d'Evaluation Vocale Assistée ainsi que son logiciel Phonetid pour la partie expérimentale dédiée à l'aérodynamique.

### 4.5.1 Présentation de Praat

Les analyses acoustiques ont été réalisées à l'aide de Praat, un logiciel libre d'analyse vocale développé par Boersma & Weenink (1996), de l'Institut des Sciences Phonétiques de l'Université d'Amsterdam. Celui-ci est disponible à la fois sous Windows, Macintosh et Unix. Il s'agit d'un programme largement utilisé par les professionnels de la voix et de la parole et qui bénéficie de mises à jour régulières de la part de ses concepteurs. L'avantage de ce logiciel est qu'il permet la programmation sous forme de scripts qui utilisent la syntaxe de Praat et qui permettent d'automatiser la recherche de paramètres acoustiques sur de larges bases de données et de générer l'ensemble des résultats dans des tableurs.

L'utilité principale de Praat est qu'il permet d'éditer un signal acoustique, d'étiqueter phonétiquement et orthographiquement ce signal, d'explorer de longs corpus, d'extraire le fondamental d'un son, de suivre les formants des voyelles, d'explorer les paramètres liés aux perturbations du signal (jitter ou shimmer), de synthétiser de manière très performante un son ou de rendre compte d'un certain nombre d'analyses spectrales. Ainsi, ce logiciel permet par exemple de visualiser des enveloppes d'un son<sup>2</sup>, des spectres<sup>3</sup> ou encore des spectrogrammes.

### 4.5.2 Présentation de EVA et Phonedit

Le recueil de données aérodynamiques a été réalisé grâce au système d'Evaluation Vocale Assistée EVA2 (cf. figure 4.1). Cette station de travail, destinée à l'analyse de la voix et de la parole dans le domaine de la phoniatrice a été développée par le laboratoire « Parole et Langage » d'Aix-en-Provence, en collaboration avec le centre hospitalier de la Timone de Marseille (Galindo & al. 1989 ; Teston & al. 1999). Cette plateforme permet la prise de mesures simultanées de données acoustiques grâce à la présence d'un microphone et d'un sonomètre, de données aérodynamiques grâce aux différents capteurs aérodynamiques et de données EGG grâce au raccordement d'un laryngographe portable (cf. figure 4.2). Le microphone, calibré à 120 dB et positionné à 30 cm de la bouche, se situe à l'arrière du capteur du débit d'air buccal. La mesure du débit d'air est possible grâce à la présence de pneumatachographes à grilles. L'ensemble des capteurs sont disposés sur une pièce à main, montée sur un pied réglable, positionné face au sujet. Cette pièce à main est directement reliée à un micro-ordinateur de type PC. La fréquence d'échantillonnage du signal acoustique est de 25000 Hz tandis que celle pour les données aérodynamiques est de 6250 Hz. Afin de recueillir le DAB, un masque de silicone adapté à la morphologie buccale des sujets est plaqué sur leurs bouches afin d'éviter toutes fuites d'air. Pour mesurer la PIO, nous avons placé un tube d'environ 4 cm de long et de 5 mm de diamètre à l'intérieur du masque. Les sujets devaient garder les dents serrées sur l'embout, sans l'écraser, pendant toute la séquence

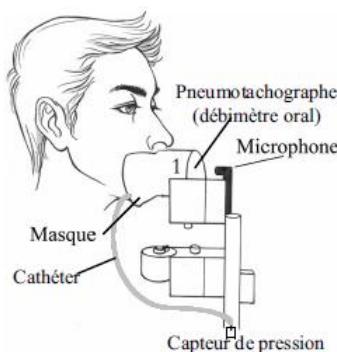
2. L'enveloppe d'un son est une courbe représentant l'amplitude ou énergie d'un son en fonction du temps. Cela correspond grossièrement à un oscilloscopie. L'enveloppe d'un son met particulièrement en évidence trois parties distinctes de ce dernier : l'attaque, la stabilité et la chute (phase pendant laquelle la vibration n'est plus entretenue. Dès lors, l'amplitude décroît).

3. Le spectre d'amplitude met en évidence l'amplitude du signal en fonction de la fréquence.

d'enregistrement. Pendant toute la session de recueil des données, les sujets étaient assis, le buste droit afin de faciliter la phonation. La plupart des sujets maintenaient la pièce à main par l'arrière afin de conserver une adhérence entre la bouche et le masque. Pour des raisons de confort des locuteurs, des pauses sont réglées entre les différentes séquences d'enregistrements.



**Figure 4.1 – Dispositif EVA2 (d'après de guide d'utilisation).**



**Figure 4.2 – Pièce à main du dispositif EVA2 permettant les prises de mesures simultanées de données aérodynamiques et acoustiques (d'après Basset & al. 2006).**

Les enregistrements, les commandes des diverses instrumentations, l'affichage des résultats et leur traitement sont possibles grâce à des logiciels spécifiques. La prise de données a été réalisée sous Phonetid (Ghio 2002). A ce sujet, il est important de préciser que les paramètres aérodynamiques sont calibrés de façon manuelle entre chaque session d'enregistrement. Ce logiciel, spécifique aux études acoustiques et aérodynamiques de la voix et de la parole, a été élaboré par la société SQLab (Aix en

Provence). Il permet d'afficher des fichiers acoustiques, EGG et aérodynamiques de manière synchronisée. Son atout principal est qu'il permet à la fois d'obtenir des mesures moyennes sur un segment donné mais aussi des valeurs exactes du DAB ou de la PIO sur n'importe quel point de la courbe. Il est également possible de procéder à un étiquetage manuel afin de faciliter les analyses.

## 4.6 Conclusion

Le but principal de ce chapitre consistait à présenter nos populations d'étude et notre corpus. Nous avons ainsi pu décrire nos cohortes de patients et de témoins et mettre en évidence les difficultés méthodologiques et logistiques que posent les enregistrements de patients dysphoniques. Ainsi, nous avons pu établir un corpus basique enregistré des points de vue acoustique et aérodynamique (avec utilisation d'un laryngographe portable branché). Les enregistrements acoustiques de chacun de nos sujets ont été réalisés dans des conditions similaires, avec un matériel commun. De la même manière, les méthodes d'enregistrements aérodynamiques sous EVA2 ont été identiques pour tous les sujets. Ces prises de données acoustiques, aérodynamiques et électroglottographiques sur des voyelles tenues — dépendantes des tâches à réaliser — associées aux productions de logatomes nous ont permis de réaliser diverses études présentées dans la suite de cette thèse.

### En résumé...

- Notre travail concerne les cordectomies de type II-III.
- Nous avons enregistré 10 patients masculins francophones entre 6 mois et 1 an post-opératoire. Ceux-ci sont comparés à une cohorte de 10 témoins.
- Nous avons couplé des voyelles tenues avec des logatomes lus de type CVCVCVC (C= Consonnes du français et V= Voyelles [a,i,u]).
- Les enregistrements ont été effectués lors de deux sessions : une session de recueil de données acoustiques et une session de recueil de données aérodynamiques et EGG.
- Les analyses acoustiques ont été réalisées avec le logiciel Praat et les données aérodynamiques et EGG ont été traitées avec le logiciel Phonetid.



# Cinq études relatives à la $F_0$ et à l'intensité de la voix

"There are three kinds of lies : lies, damned lies, and statistics."

– Mark Twain

## Résumé

Ce chapitre comprend trois études, augmentées d'un excursus, qui mettent en évidence l'impact de la chirurgie sur les paramètres de la fréquence fondamentale ( $F_0$ ), de l'intensité sonore et de leurs indices de perturbation (le jitter et le shimmer). Les plages de variation de  $F_0$  des sujets ont également été analysées. Dans l'étude 1, la  $F_0$  de nos patients et de nos témoins a été évaluée sur les voyelles tenues [a, i, u] à hauteur et intensité confortables ( $n = 60$ ). Nous avons également pu relever les mesures du jitter sur les parties stables de ces voyelles ( $n = 60$ ). Nous avons réalisé une série de tests-t non appariés sur les données issues des productions de ces trois voyelles. Les résultats montrent que :

- la hiérarchie des  $F_0$  intrinsèques des voyelles, à savoir  $F_0[a] < F_0[i] < F_0[u]$ , est conservée pour les deux groupes,
- la  $F_0$  moyenne des patients est toujours significativement plus élevée que celle des témoins,
- il n'existe pas de différence statistiquement significative entre les deux groupes pour le paramètre du jitter.

Dans l'étude 2, nous cherchons principalement à observer l'adaptation et le contrôle de la  $F_0$  des sujets lors d'une expérience d'imitation des voyelles [a, i, u] synthétisées à différentes  $F_0$ , de 90 Hz à 250 Hz selon des intervalles de 20 Hz ( $n = 540$ ). Le premier objectif de cette étude consistait à déterminer si les  $F_0$  moyennes des imitations des deux populations étaient significativement différentes des cibles théoriques attendues (i.e. les valeurs des stimuli créés). Nous avons donc conduit une série de tests-t univariés qui ont principalement montré une bonne imitation des témoins mais des difficultés dans les fréquences entre 90 et 150 Hz (voyelle [a]), entre 90 et 170 Hz (voyelle [i]), entre 90 et 230 Hz (voyelle [u]) pour les patients. Le second objectif de cette étude consistait en une comparaison inter-groupes des  $F_0$  moyennes des imitations de nos sujets. Les tests-t non appariés effectués montrent que les différences statistiquement significatives entre les deux groupes se situent principalement dans les moyennes fréquences (entre 90 et 130 Hz pour [a], 110 et 130 Hz pour [i], 110 et 190 Hz pour [u]), ce qui est prévisible puisque les écarts des patients aux cibles théoriques sont d'autant plus importants dans la zone fréquentielle de 90 à 190 Hz.

La troisième étude était consacrée aux mesures de l'intensité sonore ( $n = 60$ ) et du shimmer ( $n = 60$ ) de nos sujets sur les voyelles [a, i, u] tenues à hauteur et intensité conformables. Nous avons réalisé une série de tests-t non appariés sur les données issues de la production des trois voyelles. Les résultats sont les suivants :

- la hiérarchie des intensités intrinsèques des voyelles, à savoir  $I[a] > I[i,u]$ , est conservée pour les deux groupes,
- les intensités des patients sont toujours significativement plus élevées que celles des témoins,
- il n'existe pas de différence statistiquement significative entre les deux groupes pour le shimmer.

Finalement, afin d'évaluer les problèmes liés au corpus, nous avons confronté nos résultats préliminaires concernant les mesures de l'intensité et de  $F_0$  des voyelles tenues (études 1 et 3) à des mesures prises sur des voyelles en contexte, sur les positions 2 des logatomes de type CVCVCVC (où C= [p, t, k, b, d, g] et V=[a, i, u]) ( $n = 60$ ). Dans le cas où un effet du corpus était relevé, nous avons établi une comparaison intra-sujets afin de mettre en évidence les stratégies individuelles en production. En ce qui concerne l'intensité intrinsèque (étude 4), les tests-t non appariés conduits sur les différentes voyelles en contexte occlusif n'ont montré aucune différence significative entre les deux populations. De plus, l'ANOVA à 3 facteurs a particulièrement mis en évidence un effet significatif du corpus, entre autres, pour les mesures de l'intensité. La comparaison intra-sujets suggère que les patients ont tendance à produire des segments vocaliques plus forts à l'isolé par rapport aux voyelles des logatomes. En ce qui concerne la  $F_0$  (étude 5), les tests-t non appariés établis sur les différentes voyelles en contexte occlusif confirment les résultats de notre première étude, à savoir qu'il existe bel et bien une différence statistiquement significative entre les patients et les témoins. De plus, l'ANOVA à 3 facteurs ne montre aucun effet lié au corpus. Les patients présentent une  $F_0$  plus élevée quels que soient les contextes.

## 5.1 Introduction

La voix est un phénomène très riche et complexe. L’interaction entre ses différentes composantes qui sont d’ordre physique, physiologique, psychologique, social ou linguistique rendent son étude souvent délicate. Il est impossible de rendre compte de ces diverses dimensions au travers de quelques valeurs chiffrées. Les analyses acoustiques permettent néanmoins de quantifier certains aspects de la fonction vocale et elles sont très utiles aux professionnels dans leurs démarches diagnostiques et thérapeutiques.

Nous avons vu précédemment que la cordectomie laser — aussi minime qu’elle soit — est très dégradante d’un point de vue anatomico-physiologique. Elle mutile différentes couches du pli vocal opéré et par conséquent interfère sur les mécanismes vibratoires. La modification de la structure du pli vocal engendre une asymétrie et une diminution de la masse qui causent des perturbations du cycle en fréquence et en amplitude. Ce chapitre, qui concerne plusieurs études basées sur des protocoles cliniques classiques (cf. chapitre 4), sera entièrement consacré à l’étude de la fréquence fondamentale et de l’intensité sonore après chirurgie ainsi qu’aux indices de perturbation correspondants : le jitter et le shimmer.

Après avoir rappelé quelques éléments de statistique utiles pour la suite de ce chapitre, et plus largement dans ce document, nous nous intéresserons à différents aspects de ces paramètres au travers de quelques études originales.

Une première étude cherchera à rendre compte de la fréquence fondamentale et du jitter de nos patients et de nos témoins sur des voyelles tenues et isolées. Dans une seconde étude, nous étudierons les plages de variation de  $F_0$  de nos sujets à l’aide d’une expérience d’imitation de voyelles tenues et synthétisées à différentes fréquences. La troisième étude sera consacrée à l’évaluation de l’intensité sonore et du shimmer de nos sujets sur les voyelles tenues et isolées. Nous nous sommes confrontée à des problèmes de méthode concernant l’intensité sonore. La distance du micro-casque à la bouche influence l’intensité mesurée et peut représenter un biais pour ce type de recherche et invalider partiellement nos conclusions préliminaires. En conséquence, la dernière section de ce chapitre présente un *excursus* permettant d’examiner nos résultats précédents. Les valeurs d’intensité relevées dans notre troisième étude sur les voyelles tenues seront comparées avec des valeurs d’intensité mesurées dans des logatomes, pour chacun de nos sujets. Dans un souci de cohérence, nous avons adopté une démarche et une méthodologie similaire dans la comparaison de nos valeurs pour le paramètre de la fréquence fondamentale.

## 5.2 Notes préliminaires sur les tests statistiques

Avant d’aborder les différents tests statistiques que nous avons réalisés sur l’ensemble de notre corpus, nous croyons nécessaire d’apporter aux lecteurs les moins avisés quelques éléments rudimentaires de statistique.

## 5.2.1 Paramètres communs de statistique descriptive

### 5.2.1.1 La moyenne arithmétique

La notion sans doute la plus simple — mais aussi la plus intuitive — est celle de la moyenne. La moyenne *arithmétique*, notée  $\bar{x}$  peut être définie comme la somme des valeurs de la variable étudiée divisée par le nombre ( $n$ ) d'individus :

- (1) Calcul de la moyenne arithmétique d'un échantillon

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n n_i x_i$$

### 5.2.1.2 La variance

La deuxième notion importante est celle de la *variance*. En statistique, la variance d'une modalité d'un individu, que l'on note généralement  $\sigma^2$ , correspond à la moyenne des carrés des écarts entre la modalité d'un individu ( $x_i$ ) et la moyenne arithmétique  $\bar{x}$ . Celle-ci se calcule de la manière suivante :

- (2) Calcul de la variance d'un échantillon

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

### 5.2.1.3 L'écart-type

La troisième et dernière notion importante en ce qui nous concerne est celle de l'*écart-type* (ou encore déviation standard). Celle-ci est une mesure de la dispersion des valeurs autour de leur moyenne arithmétique ( $\bar{x}$ ). Par définition, l'écart-type est la moyenne quadratique des écarts à la moyenne  $\bar{x}$ . Il représente par conséquent la racine carrée de la variance. L'écart-type *empirique*, symbolisé habituellement par la lettre grecque  $\sigma$ , peut être défini de la manière suivante :

- (3) Calcul de l'écart-type d'un échantillon

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Néanmoins, l'écart-type sous-estime légèrement l'écart des données par rapport à la moyenne ; ce qui est d'autant plus vrai lorsque la taille de l'échantillon (ou de la population) est faible et lorsque l'écart-type  $\sigma$  est inconnu. Dans ce cas, l'écart-type — mais également la variance — est dit *biaisé*. Il est toutefois possible de corriger ce biais en calculant l'écart-type et la variance non biaisés, notés  $s$  et  $s^2$  respectivement, à l'aide des formules suivantes :

$$(4) \quad \begin{array}{ll} \text{a. Ecart-type corrigé} & \text{b. Variance corrigée} \\ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} & s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \end{array}$$

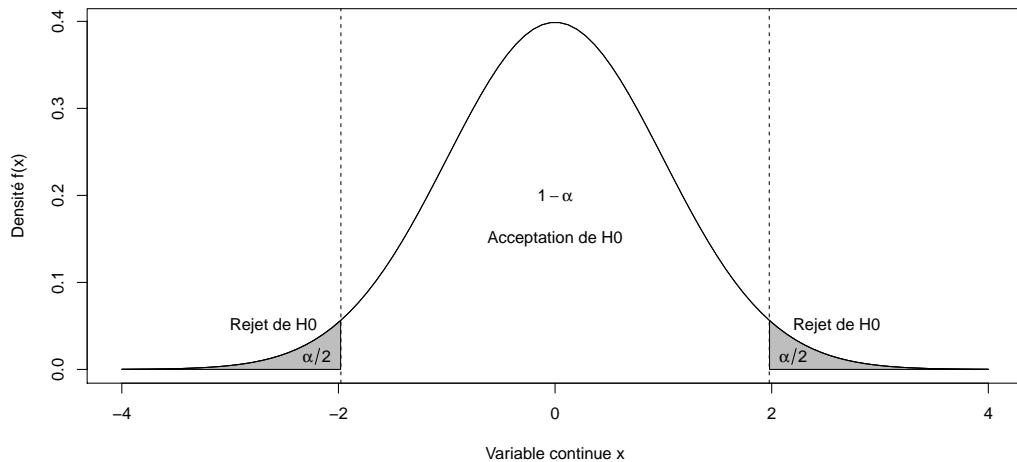
## 5.2.2 Les Tests statistiques

La plupart des tests statistiques dont il sera question dans cette thèse sont des tests de Student, également appelés tests  $t$ . Néanmoins, ces types de tests nécessitent certains tests préliminaires. En effet, l'application du test de Student est subordonnée, d'une part, à la *normalité* des échantillons — condition toutefois non essentielle lorsque les échantillons ont des effectifs supérieurs ou égaux à 30 — et, d'autre part, à *l'homogénéité* des variances des échantillons considérés.

### 5.2.2.1 Hypothèses et seuil de tolérance alpha

Un test statistique comporte généralement quatre étapes. La première d'entre elles consiste à formuler l'hypothèse nulle, notée  $H_0$ , que l'on désire contrôler. Celle-ci stipule simplement qu'il n'existe pas de différence entre les paramètres comparés ou que la différence observée entre ces paramètres n'est pas significative mais due aux fluctuations d'échantillonnage. De manière générale, cette hypothèse est principalement formulée dans le but d'être rejetée. Dans ce cas, l'hypothèse alternative, notée  $H_1$ , devra être acceptée. Deuxièmement, et afin de pouvoir apprécier le résultat d'un test statistique, il est nécessaire de choisir dans la loi considérée — loi normale en ce qui nous concerne — une région de rejet de l'hypothèse correspondant à une probabilité égale au seuil de signification  $\alpha$  que l'on accepte de prendre en rejetant l'hypothèse nulle. D'une manière générale, et en Sciences Humaines en particulier, il est courant d'admettre que l'éventuel rejet de l'hypothèse nulle peut avoir 5 chances sur 100 d'être faux (soit .05 sur 1). Autrement dit, s'il existe une différence entre une population A et une population B, celle-ci peut être due au hasard. Nous appellerons  $p$  le seuil de signification le plus bas qui permet de rejeter  $H_0$ . Cette valeur  $p$  exprime la probabilité d'obtenir par pur hasard une différence plus grande ou égale à celle observée. On dira par exemple que l'association entre deux distributions est statistiquement significative si  $p \leq .05$  (soit 5%). La figure 5.1 met en évidence les régions d'acceptation et de rejet de l'hypothèse nulle :

Troisièmement, selon le type d'indice statistique choisi pour estimer le paramètre étudié, divers types de distributions de probabilités seront possibles. Si l'on s'en tient à l'hypothèse nulle, le choix de cette distribution de référence permet de définir si la valeur calculée de l'estimateur est *probable*. Lorsque la valeur calculée est jugée très peu probable, nous décidons de rejeter l'hypothèse nulle parce que la distribution d'échantillonnage des estimateurs que celle-ci suppose ne semble pas appropriée. Par exemple, la plupart des tests d'hypothèse dont il sera question ici font appel à la distribution  $t$  de Student comme distribution de référence. Enfin, la quatrième et dernière étape consiste à rejeter ou non l'hypothèse  $H_0$  suivant que la valeur numérique de la fonction des observations tombe ou non dans la région de rejet  $\alpha$  définie précédemment.



**Figure 5.1** – Courbe de Gauss représentant les régions de rejet de l'hypothèse nulle pour un seuil  $\alpha = .05$ .

### 5.2.2.2 Test de normalité des échantillons

Afin de conduire nos tests statistiques, il est nécessaire de vérifier et de s'assurer que nos échantillons sont issus d'une distribution normale (i.e. qu'ils suivent une loi normale). Il existe un certain nombre de tests statistiques permettant de comparer une distribution donnée à une distribution standard. Le test de normalité sans doute le plus utilisé est celui de Kolmogorov-Smirnov. En effet, celui-ci est approprié pour vérifier l'hypothèse nulle selon laquelle un échantillon donné provient d'une distribution normale lorsque les paramètres  $\mu$  (moyenne théorique) et  $\sigma^2$  (variance) sont connus. Dans ce test non paramétrique, nous mesurons l'écart maximal qui existe entre la fonction de densité cumulée observée (fréquences cumulées) et la fonction de distribution théorique. Sous l'hypothèse  $H_0$ , l'écart est faible et la distribution des observations s'intègre relativement bien dans une distribution donnée. En ce qui nous concerne, l'utilisation du test de Kolmogorov-Smirnov ne semble pas approprié puisque la moyenne théorique ou la variance de nos distributions ne sont pas connues *a priori*. Dans ce type de configuration, il est nécessaire d'utiliser le test de normalité de Lilliefors, une modification du test de Kolmogorov-Smirnov. La statistique du test de Lilliefors consiste en une différence maximale absolue entre la fonction de distribution empirique et la fonction de distribution cumulative d'une distribution normale avec une moyenne et une variance estimées. Il peut être calculé de la manière suivante :

#### (5) Test de Lilliefors

$$D = \max_{i=1,\dots,n} \left( F_i - \frac{i-1}{n}, \frac{i}{n} - F_i \right)$$

Où  $F_i$  représente la fréquence théorique de la loi de distribution normale centrée et réduite associée à la fonction  $\Phi$  :  $F_i = \Phi\left(\frac{x_i - \bar{x}}{s}\right)$ . Ici  $\Phi$  est la fonction de distribution cumulative d'une distribution normale standard.

### 5.2.2.3 Test d'homogénéité des variances

Après avoir vérifié la normalité de la distribution de nos deux échantillons indépendants, il est nécessaire de savoir si les variances de ces deux échantillons sont statistiquement identiques (homoscédasticité). A ce titre, l'un des tests sans doute le plus adéquat est celui de Fisher-Snedecor. L'objectif de ce test, que l'on nomme également test F de Fisher, consiste à vérifier si les deux variances (non biaisées) de deux populations normalement distribuées sont différentes ou non. Sous  $H_0$ , les deux variances sont identiques (i.e.  $s_A^2 = s_B^2$ ) alors que sous  $H_1$ , l'hypothèse alternative, les variances sont différentes (i.e.  $s_A^2 \neq s_B^2$ ). Le calcul de ce test implique avant tout, d'une part, le calcul de la variance des deux échantillons ( $s_A^2$  et  $s_B^2$ ) et, d'autre part, la considération du degré de liberté à  $\nu = n - 1$ <sup>1</sup>. Ces éléments déterminés, le calcul du test F est le résultat du quotient de la plus grande des deux variances sur la plus petite :

(6)     Test F de Fisher-Snedecor (où  $s_A^2 > s_B^2$ )

$$F = \frac{s_A^2}{s_B^2}$$

Une fois la valeur de F calculée, l'étape suivante consiste à se reporter à la table de Fisher (table 5.1) afin d'y confronter la valeur de F obtenue à la valeur de F maximale pour un seuil critique donné.

Si la valeur de F obtenue est supérieure à la valeur de F reportée dans la table pour  $n$  degré de liberté, cela signifie que les variances des deux distributions ne sont pas statistiquement significatives. En d'autres termes, le résultat du test favorisera l'hypothèse alternative  $H_1$  et conclura à une non homoscédasticité des variances. Dans le cas contraire, c'est-à-dire si la valeur de F obtenue est inférieure à la valeur de F reportée dans la table, l'hypothèse nulle  $H_0$  sera favorisée et l'homoscédasticité des variances sera statistiquement prouvée. Par exemple, si la variance  $s^2$  d'un échantillon  $A$  composé de 10 valeurs ( $\nu = 9$ ) est égale à la variance d'un échantillon  $B$  composé d'un même nombre de valeur ( $\nu = 9$ ) alors la valeur de F obtenue doit être inférieure ou égale à  $F(.95, 9; 9) = 3.18$ , où .95 représente le pourcentage de confiance ( $1 - \alpha$ ) et 9; 9 le degré de liberté du nominateur et du dénominateur.

1. La forme d'échantillonnage des tests statistiques varie selon un paramètre que nous appelons le *degré de liberté* (ou ddl, noté  $\nu$ ). Le degré de liberté représente le nombre de classes pour lesquelles les effectifs peuvent varier librement. Pour prendre un exemple, lorsque nous disposons de dix nombres dont nous connaissons la somme, le degré de liberté est égal à  $\nu = 10 - 1 = 9$  puisqu'en connaissant, d'une part, la somme des dix nombres et, d'autre part, neuf de ces nombres, il est facile de déduire le dixième nombre ; celui-ci correspond à la somme des dix nombres diminuée de la somme des neuf nombres choisis. Le principe général qui permet de calculer le degré de liberté est le suivant (où  $l$  correspond au nombre de ligne et  $c$  au nombre de colonne) :  $\nu = (l - 1)(c - 1)$ .

		Nominateur (ddl)									
		1	2	3	4	5	6	7	8	9	10
Dénominateur (ddl)	1	161	199	216	225	230	234	237	239	241	242
	2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
	3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

Table 5.1 – Extrait de la table de Fisher pour un seuil  $\alpha = .05$ .

#### 5.2.2.4 Les tests de Student

Lorsque la normalité et l'homogénéité des variances ont été vérifiées, toutes les conditions nécessaires à l'application du test de Student sont alors réunies. Dans ce type de tests paramétriques, nous cherchons à comparer une moyenne d'un échantillon à une moyenne théorique, notée  $\mu$ , à comparer les moyennes de deux échantillons indépendants ou de deux échantillons appariés. Autrement dit, nous cherchons à vérifier l'hypothèse *nulle* — notée  $H_0$  — selon laquelle  $\bar{x} = \mu$  dans le cas d'une comparaison d'une moyenne observée à une moyenne théorique ou  $\bar{x}_1 = \bar{x}_2$  dans le cas d'une comparaison entre deux moyennes observées, appariées ou non. Dans le cas d'un test *t bivarié* (apparié ou non) par exemple, la valeur de *t* est obtenue en divisant la différence des deux moyennes d'échantillons aléatoires ( $\bar{x}_1$  et  $\bar{x}_2$ ) par la racine carrée de la variance (l'écart-type) corrigée ( $s^2$ ) divisée par l'effectif *n* de l'échantillon. La valeur de *t* ainsi obtenue est confrontée aux critères de la loi de Student, avec un paramètre *k(ddl)* équivalent à  $n - 1$ , soit l'effectif de l'échantillon - 1. En ce qui concerne le test *t univarié*, la valeur de *t* est obtenue en divisant la différence entre la moyenne d'un échantillon aléatoire ( $\bar{x}$ ) et une moyenne théorique ( $\mu$ ) par la racine carrée de la variance corrigée ( $s^2$ ) divisée par l'effectif *n* de l'échantillon. Les calculs de ce test figurent sous 7 (où  $\bar{x}$  = moyenne de l'échantillon,  $\mu$  = moyenne théorique,  $s^2$  = variance corrigée de l'échantillon et *n* = effectif de l'échantillon) :

- (7) Tests de Student (bivarié et univarié)  
 a. test *t bivarié*      b. test *t univarié*

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}} \quad t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}}$$

Lorsque la valeur de *t* est connue, il suffit, pour un degré de liberté égal à  $(n_1 + n_2) - 2$ , de se reporter au seuil de significativité  $\alpha$  choisi (normalement  $p = .05$ ) et d'y

lire la valeur inscrite sur la table  $t$  (cf. table 5.2). Si la valeur calculée de  $t$  excède la valeur figurant dans la table, nous pouvons en conclure que les moyennes sont significativement différentes au niveau de la probabilité. Une différence significative pour un seuil fixé à  $p = .05$  signifie que si l'hypothèse nulle était correcte, c'est-à-dire si les distributions ne différaient pas, alors nous devrions obtenir une valeur de  $t$  aussi grande que la valeur figurant dans la table  $t$  dans seulement 5% des cas. Dans une telle configuration, nous pouvons raisonnablement affirmer que les deux distributions comparées diffèrent l'une de l'autre mais que cette conclusion a environ 5% de chance d'être erronée.

<i>Degré de liberté (ddl)</i>	Seuil de risque <i>alpha</i> (bilatéral)			
	.1	.05	.01	.001
1	6.31	12.71	63.66	636.62
2	2.92	4.30	9.93	31.60
3	2.35	3.18	5.84	12.92
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.89	2.37	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
11	1.80	2.20	3.11	4.44
12	1.78	2.18	3.06	4.32
13	1.77	2.16	3.01	4.22
14	1.76	2.14	2.98	4.14
15	1.75	2.13	2.95	4.07
16	1.75	2.12	2.92	4.02
17	1.74	2.11	2.90	3.97
18	1.73	2.10	2.88	3.92
19	1.73	2.09	2.86	3.88
20	1.72	2.09	2.85	3.85

**Table 5.2** – Extrait de la table de la Loi de Student (test  $t$ )

### 5.2.2.5 L'ANOVA

De même, lorsque la normalité et la variance inter-groupes ont été prises en compte, il est possible d'utiliser une ANOVA ("ANalysis Of VAriance"). Ce test statistique permet de comparer non plus deux groupes de données — comme les test-t le faisaient — mais au moins trois. Il s'agit d'une analyse factorielle dans le sens où elle permet de vérifier si une ou plusieurs variables dépendantes (variables à expliquer) sont en relation avec une ou plusieurs variables indépendantes (variables explicatives). Finalement, l'ANOVA permet la comparaison de la variance inter-groupes à la variance

intra-groupes. Dans ce cas, c'est le test F de Fisher qui est utilisé. Celui-ci calcule le rapport entre la variance inter-groupes et la variance intra-groupes.

## 5.3 Etude n°1 : la $F_0$ intrinsèque et le jitter des voyelles tenues isolées

Cette première étude acoustique a pour objectif l'évaluation des paramètres de la  $F_0$  et du jitter. Nous avons voulu comparer la  $F_0$  et le jitter de notre population de patients avec la cohorte de témoins sur les voyelles tenues [a, i, u] isolées.

### 5.3.1 Méthodologie

#### 5.3.1.1 Recueil des données

Les témoins et patients ont été enregistrés dans les mêmes conditions : dans des pièces acoustiquement calmes, avec un même protocole et un matériel similaire (cf. chapitre 4). Les sujets avaient comme consigne d'émettre les trois voyelles [a, i, u] à hauteur et intensité confortables pendant au moins deux secondes.

#### 5.3.1.2 Mesures et paramètres étudiés

Nous avons analysé chacune des productions de nos deux populations à l'aide du logiciel Praat en réalisant des analyses spectrographiques à bandes larges avec une étendue de fréquences de 0 à 4000 Hz de l'ensemble de nos fichiers sons. Le logiciel calcule automatiquement les paramètres suivants : la  $F_0$  médiane, la  $F_0$  moyenne, l'écart-type, les  $F_0$  maximales et minimales, le nombre de périodes, l'écart-type des périodes, le degré de dévoisement, le jitter ratio, le jitter rap, le shimmer et le HNR. Après avoir pris soin de sélectionner uniquement les parties stables des voyelles tenues, en éliminant les attaques et les parties finales, nous nous sommes intéressée à la  $F_0$  moyenne ( $n = 60$ , soit 3 voyelles  $\times$  20 sujets) et au jitter moyen (soit  $n = 60$ , soit 3 voyelles  $\times$  20 sujets) de l'ensemble de ces productions. Toutes les mesures ont fait l'objet d'un traitement statistique avec Statview et R.

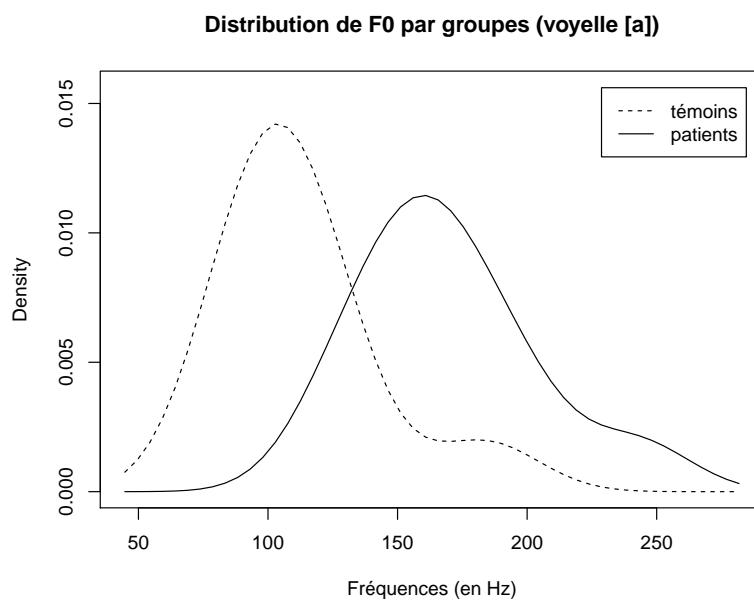
#### 5.3.1.3 Tests statistiques préalables

**Test de Lilliefors.** Le test de normalité de Lilliefors, qui permet de vérifier la normalité des distributions des populations, a été appliqué sur l'ensemble des productions des voyelles stables ([a, i, u]). L'hypothèse nulle ( $H_0$ ) pour ce test est que l'erreur est normalement distribuée, c'est-à-dire qu'il n'y a pas de différence entre la distribution observée de l'erreur et une distribution normale. L'hypothèse alternative ( $H_1$ ) est que l'erreur n'est pas normalement distribuée. Les résultats de ce test sont présentés dans le tableau suivant :

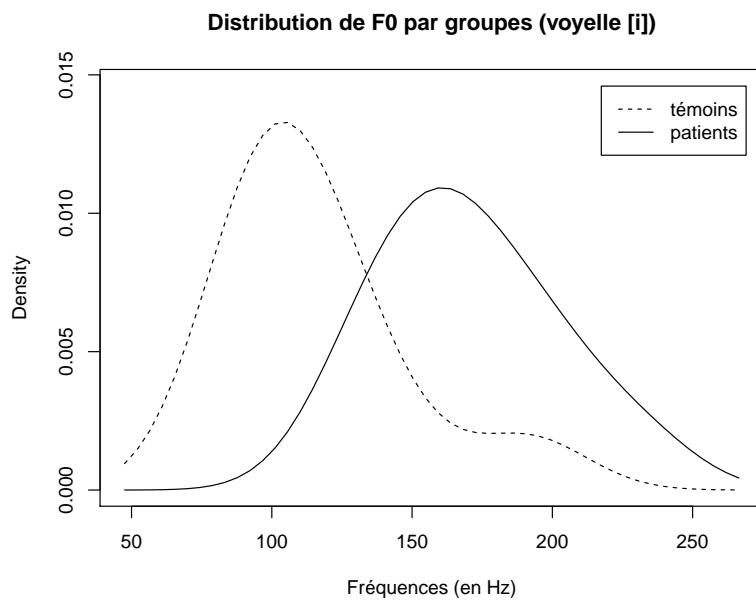
	[a]		[i]		[u]	
	D	p	D	p	D	p
Témoins	.2455	.0891	.2078	.2547	.2608	.0524
Patients	.2026	.2902	.14	.8365	.1251	.9286

**Table 5.3** – Test de normalité de Lilliefors appliqué sur la  $F_0$  moyenne des voyelles [a], [i] et [u] ( $n = 60$ ).

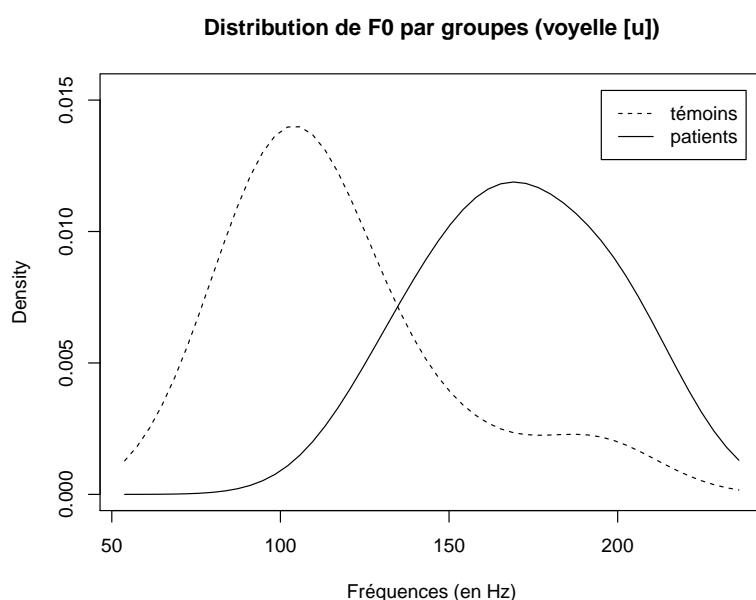
Dans la table des valeurs critiques du test de Lilliefors, la valeur de  $D$  ne doit pas être supérieure ou égale à .2616 pour accepter  $H_0$  avec un effectif  $n = 10$  et un seuil critique  $\alpha = .05$ . Les résultats de ces tests montrent que les distributions des trois voyelles selon les deux groupes s'intègrent à des distributions suivant une loi normale (cf. figures 5.2, 5.3 et 5.4). En effet, la valeur de  $p$  est supérieure à .05 dans tous les cas et, par conséquent, l'hypothèse nulle doit être acceptée. Dans nos données, la  $F_0$  d'un témoin se situe bien au-dessus des fréquences fondamentales moyennes des autres locuteurs témoins (aux alentours de 180 Hz), ce qui pourrait constituer un biais statistique.



**Figure 5.2** – Distribution des valeurs de  $F_0$  pour les groupes des témoins et des patients pour la voyelle [a] ( $n = 20$ ).



**Figure 5.3** – Distribution des valeurs de  $F_0$  pour les groupes des témoins et des patients pour la voyelle [i] ( $n = 20$ ).



**Figure 5.4** – Distribution des valeurs de  $F_0$  pour les groupes des témoins et des patients pour la voyelle [u] ( $n = 20$ ).

**Test F d'homogénéité des variances.** Dans un premier temps, nous avons démontré que les distributions de la variable étudiée étaient normalement distribuées pour nos deux populations. Dans un second temps, il est essentiel de vérifier que nos deux populations ont des variances homogènes afin de pouvoir les comparer deux à deux. Nous avons donc calculé le rapport des variances estimées ( $s^2$ ) entre nos deux populations sur l'ensemble de nos données de  $F_0$  pour les trois voyelles stables. Le tableau 5.4 résume les conclusions de ce test :

	Valeurs de $F$	Valeurs de $p$
[a]	1.402	.6227
[i]	.980	.9769
[u]	.624	.4931

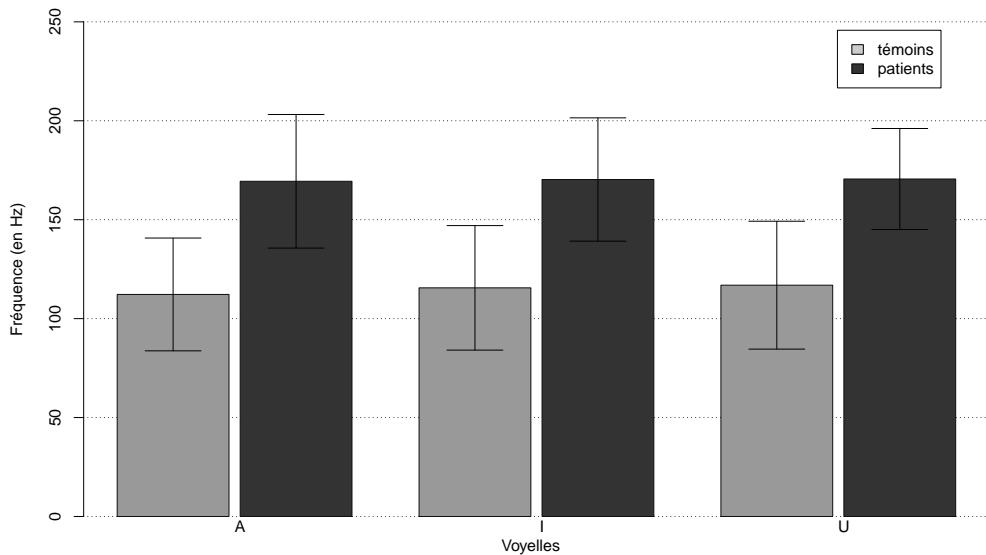
**Table 5.4** – Résultats du test F de Fisher-Snedecor sur les valeurs de  $F_0$  pour les trois voyelles [a], [i] et [u] ( $n = 60$ ).

Ainsi, nous notons que pour nos trois voyelles, les seuils de significativité sont toujours supérieurs à .05, ce qui implique que  $H_0$  n'est pas rejetée et que par conséquent les variances des deux distributions sont homogènes. Les populations forment bien deux groupes indépendants et comparables.

### 5.3.2 Présentation des résultats

#### 5.3.2.1 La fréquence fondamentale

La figure 5.5 met en évidence les différences inter-groupes pour le paramètre de la  $F_0$  pour les trois voyelles étudiées. Les valeurs de  $F_0$  moyennes des patients sont plus élevées que celles des témoins pour chacune des voyelles.



**Figure 5.5** – Valeurs moyennes de la  $F_0$  (en Hz) des voyelles [a], [i], [u] des témoins et des patients (les barres d'erreurs représentent les écarts-types) ( $n = 60$ ).

Afin de déterminer si les valeurs de  $F_0$  moyennes des patients sont significativement différentes de celles des témoins, nous avons conduit des tests-t non appariés sur les données issues de la production des trois voyelles. Les résultats sont considérés significatifs à partir d'un seuil alpha de 5% (soit  $p < .05$ ). Le tableau 5.5 met en évidence les résultats de cette expérience pour le paramètre de la fréquence fondamentale :

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	112.36	163.95	115.76	166.88	118.32	168.91
$s$	28.60	34.73	31.72	26.01	35.89	24.21
$t_{(18)}$		3.63		3.94		3.69
$p$		.0019		.001		.0017

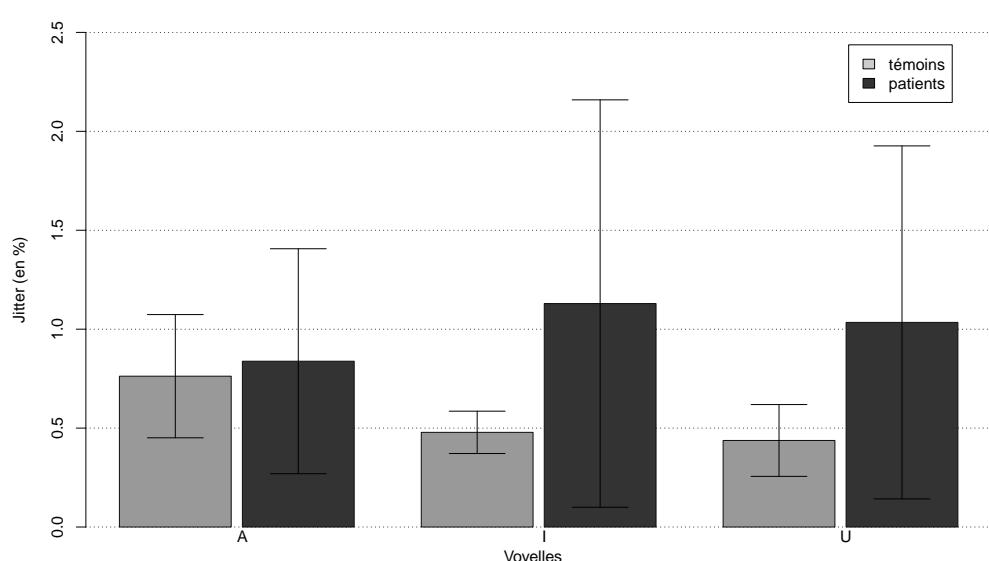
**Table 5.5** – Résultats des tests-t non appariés conduits sur les valeurs de  $F_0$  (en Hertz) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Nous remarquons que pour nos deux populations les valeurs de  $F_0$  de [u] sont plus élevées que les valeurs de  $F_0$  de [i] et les valeurs de  $F_0$  de [i] sont plus élevées que les valeurs de  $F_0$  de [a]. Ces observations peuvent être résumées de la manière suivante :

$F_0[a] < F_0[i] < F_0[u]$ . Néanmoins, il n'y a pas de différence statistiquement significative entre les  $F_0$  moyennes de chacune des voyelles, et ce pour nos deux populations. Par ailleurs, pour nos trois voyelles, nous trouvons une différence statistique hautement significative entre les valeurs moyennes de  $F_0$  des patients et celles des témoins. Pour [a], la p-value entre les deux groupes est de  $p = .0019$ ; pour [i], la p-value est de  $p = .0010$  et enfin pour [u], la p-value est de  $p = .0017$ .

### 5.3.2.2 Le jitter

La figure 5.6 montre les différences entre nos deux groupes pour le paramètre du jitter pour les trois voyelles. Une fois encore, nous remarquons que les pourcentages du jitter des patients sont toujours plus élevés que ceux des témoins.



**Figure 5.6** – Valeurs moyennes du jitter (en %) des voyelles [a], [i], [u] des témoins et des patients (les barres d'erreurs représentent les écarts-types) ( $n = 60$ ).

Comme pour le paramètre de la fréquence fondamentale, nous avons voulu vérifier si ces différences inter-groupes étaient significatives. Nous avons donc, à nouveau, conduit une série de tests-t non appariés sur nos données. Le tableau 5.6 résume les principaux résultats de ces tests :

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	.76	.84	.48	1.13	.44	1.03
$s$	.31	.57	.11	1.03	.18	.89
$t_{(18)}$		.37		1.99		2.07
$p$		.7166		.0621		.0528

**Table 5.6** – Résultats des tests-t non appariés conduits sur les valeurs du jitter (en pourcentages) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Nous pouvons observer que pour ces trois voyelles, il n'y a pas de différence statistiquement significative entre nos deux populations. Seules les voyelles [i] et [u] présentent un seuil de significativité proche de .05 entre les deux groupes avec, respectivement,  $p = .0621$  et  $p = .0528$ .

### 5.3.3 Discussion des résultats

#### 5.3.3.1 La fréquence fondamentale

**Fréquence fondamentale intrinsèque.** Dans un premier temps, il est important de constater que les fréquences fondamentales moyennes de nos deux populations suivent un ordre de grandeur caractéristique :  $F_0[a] < F_0[i] < F_0[u]$ . Cet ordre est prévisible et dépend de la  $F_0$  intrinsèque des voyelles, à savoir que chaque voyelle possède une fréquence fondamentale qui lui est propre (Whalen & Levitt 1995). A partir de différentes études incluant 31 langues de 11 familles différentes, ces auteurs ont montré que les fréquences fondamentales des voyelles fermées ont tendance à être plus hautes que celles des voyelles ouvertes<sup>2</sup>. Ce constat universel relève d'un effet très significatif lié aux voyelles elles-mêmes. En effet, la faible différence de  $F_0$  entre les voyelles hautes et les voyelles basses (de 4 à 25Hz qui varie selon les positions) serait due à l'aperture des voyelles et non pas à la dimension antérieure-postérieure. Finalement, les auteurs rapportent que les fréquences fondamentales des voyelles découlent de contraintes de production qui ne sont pas directement liées aux langues. Cette thèse est appuyée par les recherches de Whalen & al. (1995) sur le babillage d'enfants. Effectivement, pour ces auteurs, si la  $F_0$  intrinsèque dépend de la production des voyelles elles-mêmes, elle doit forcément apparaître dans les premières productions vocaliques. L'analyse du babil de six enfants français et six enfants américains montre une tendance similaire aux résultats cités précédemment sur des populations adultes. Ainsi, les mesures effectuées montrent une  $F_0$  plus haute pour les voyelles fermées antérieures et postérieures, des valeurs plus basses pour les voyelles mi-fermées antérieures et postérieures et des valeurs encore plus basses pour les voyelles ouvertes antérieures et postérieures.

2. Pour une revue de la littérature concernant la  $F_0$  intrinsèque des voyelles du français, voir Di Cristo & al. 1979, Di Cristo 1982, Rossi & Autesserre 1981 entre autres.

D'un point de vue physiologique, plusieurs théories se sont confrontées pour tenter d'expliquer les mécanismes liés à la fréquence fondamentale intrinsèque. A notre connaissance, le premier à avoir observé cette tendance est Taylor (1933), dans son étude sur l'anglais. Dans sa « théorie dynamogénétique », cet auteur note que l'élévation de la fréquence fondamentale pour les voyelles hautes serait la conséquence d'un transfert de la tension musculaire de la langue aux muscles du larynx par une sorte de résonance sympathique ou de radiation. Aujourd'hui, les chercheurs rejettent cette hypothèse car l'isolation entre muscles et nerfs a été prouvée, ce qui implique que la contraction d'un muscle ne peut déclencher la contraction d'un muscle voisin (Atkinson 1973). Mohr (1971) explique que cette différence de fréquence fondamentale intrinsèque est relative à des phénomènes aérodynamiques. En effet, selon celui-ci, le point de constriction derrière la langue lors de la production des voyelles serait à l'origine d'un phénomène de « rétropression » plus ou moins important au niveau des plis vocaux, ce qui influencerait la  $F_0$ . Ainsi, plus l'espace entre les plis vocaux et le point de constriction serait important, plus la « rétropression » serait grande et moins la  $F_0$  serait modifiée (voyelles hautes). Au contraire, plus l'espace entre les plis vocaux et le point de constriction serait court (constriction très postérieure), plus l'effet de « rétropression » serait bref et compact et plus la  $F_0$  serait abaissée (voyelles basses). Selon le niveau de la constriction, la théorie devrait normalement prédir  $F_0[a] < F_0[u] < F_0[i]$ . Or, Ohala (1972) relève de manière empirique  $F_0[a] < F_0[i] \leq F_0[u]$ , ce qui va à l'encontre de cette théorie. Une troisième théorie, liée au modèle de Flanagan & Landgraf (1968) (cf. section 1.10), stipule que la  $F_0$  intrinsèque dépend du couplage acoustique entre le conduit vocal et les plis vocaux, ces derniers devant maintenir un niveau d'impédance<sup>3</sup> assez bas pour permettre le bon fonctionnement de ce couplage. Ainsi, Atkinson (1973) suggère que les résonances du conduit vocal imposent la fréquence de vibration des plis vocaux. Ce couplage est d'autant plus important pour les voyelles hautes qui présentent un  $F_1$  et une  $F_0$  proches. Enfin, ces différences de fréquences fondamentales peuvent être expliquées d'un point de vue articulatoire. La théorie de « l'attraction linguale », par exemple, stipule que la langue est en position haute pour la production des voyelles hautes, ce qui exerce une pression plus importante sur les muscles et les structures du larynx, notamment sur l'os hyoïde. Cette tension verticale des structures laryngées peut être à l'origine d'un étirement plus important des plis vocaux, ce qui explique le réhaussement de  $F_0$  pour ces voyelles (Ohala 1972, Ohala 1973, Ohala & Eukel 1987, Di Cristo 1985, Honda 2004). Une autre explication, basée sur des expérimentations EMG, stipule que ces différences de fondamental sont liées à une activité plus importante du muscle crico-thyroïdien pour la production des voyelles hautes (Dyhr 1990, Honda & Fujimura 1991, Hoole & Honda 2011). Cependant, Whalen & al. (1998) restent sceptiques à propos de cette hypothèse.

**Différence de fréquence fondamentale entre nos deux populations.** Comme nous l'avons déjà expliqué précédemment, une oscillation entretenue et symétrique des plis

3. L'impédance acoustique caractérise la résistance d'un lieu, d'un matériau au passage de l'air. Physiquement, cela correspond au rapport entre la pression acoustique et la vitesse de propagation du son dans son milieu ([www.wikipedia.fr](http://www.wikipedia.fr)).

vocaux est réalisable grâce à l'interaction de deux phénomènes : (i) une déformation non uniforme du tissu des plis vocaux (cf. Modèle à deux masses) et (ii) un décollement des parois en un point de séparation. Cette notion de symétrie fait également référence à la structure verticale des plis vocaux. En effet, ces derniers peuvent et doivent être couplés de deux façons différentes afin de maintenir des cycles réguliers : un couplage élastique entre les parties hautes et basses des plis vocaux et un couplage par accollement entre les deux plis vocaux.

En outre, les locuteurs peuvent agir sur la masse, l'élasticité ou encore la longueur de leurs plis vocaux par un ajustement complexe des muscles laryngés. Ainsi, Titze (1994) ou Henrich (2001) entre autres, précisent que l'augmentation de la fréquence de vibration des plis vocaux, et donc l'élévation de la  $F_0$ , serait due à une augmentation de la PSG, à une contraction du muscle vocal (augmentation de la raideur des plis vocaux) ou à une diminution de la longueur vibrante.

Dans cette étude, nous avons pu mettre en évidence que la fréquence fondamentale moyenne des patients était statistiquement plus élevée que celle des témoins pour l'ensemble des voyelles analysées. Ces résultats corroborent les conclusions de McGuirt (1994), Tamura & al. (2003), Ledda & al. (2006), Haddad & al. (2006), Vilaseca & al. (2008), Mirghani (2009), Sjögren & al. (2009) et Claeys (2010) qui avaient déjà montré que la fréquence fondamentale de patients opérés par cordectomie laser au CO<sub>2</sub> était plus élevée que celle de leurs témoins. Nous confronterons nos résultats avec ceux de ces diverses recherches en insistant sur le fait qu'aucune étude présentée ne traite exclusivement des cordectomies de type II-III et que les voyelles analysées peuvent être différentes des nôtres.

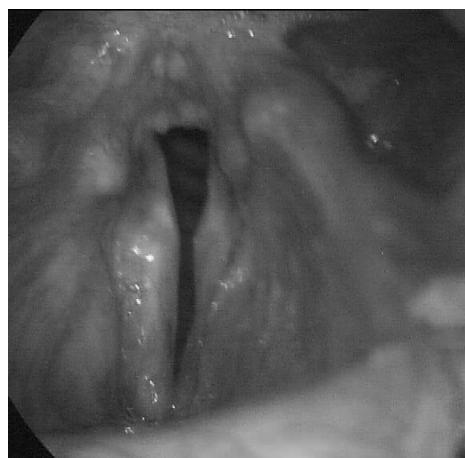
Les articles de McGuirt (1994), Tamura & al. (2003) et Sjögren & al. (2009) sont les seuls à ne pas avoir comme référence une population témoin. La première étude établit une comparaison acoustique entre un groupe de patients ayant subi une cordectomie laser (de type III majoritairement), un groupe de patients irradiés et des valeurs normatives issues de la littérature pour les voyelles [i] et [a]. Les résultats montrent que les patients opérés présentent, dans tous les cas, une  $F_0$  plus élevée. De même, la seconde recherche inclut 10 patients ayant subi une cordectomie de type III. Les résultats acoustiques sur les voyelles [a, e, i, o, u] sont comparés à des valeurs de référence issues de la littérature et à ceux d'une population de patients irradiés. Les auteurs relèvent une différence significative de  $F_0$  entre les deux groupes. La troisième étude présente les résultats acoustiques pour la voyelle [a] pour 34 patients traités par cordectomie de type I et II, comparés à des valeurs normatives issues du manuel de Praat. Les auteurs notent que la  $F_0$  des patients est plus élevée par rapport à la valeur de référence de 120 Hz dérivée du manuel.

Les études de Ledda & al. (2006), Haddad & al. (2006), Vilaseca & al. (2008), Mirghani (2009) et Claeys (2010) présentent leurs résultats en fonction d'un groupe contrôle sain. Ledda & al. (2006) ont étudié les productions de patients traités par cordectomie de type I, II, III, IV et V sur la voyelle [a]. Les auteurs n'ont noté aucune différence significative entre les patients traités par cordectomie de type I et II et leurs témoins. La différence significative de  $F_0$  entre le groupe de référence et les patients n'intervient qu'à partir du type III. Cet article met donc en évidence un écart signifi-

catif de  $F_0$  entre une cohorte de témoins et une certaine catégorie de patients, lesquels ne correspondent pas forcément au type de patients étudiés dans cette thèse. Haddad & al. (2006) présentent une confrontation de données acoustiques sur un ensemble hétérogène de 15 patients et un groupe témoin sur la voyelle [e], en ne faisant aucune association selon le type de cordectomie (types I, II, III et IV confondus). Ainsi, les résultats de Haddad & al. (2006), bien que similaires aux nôtres ( $F_0$  moyenne de 167,51 Hz), peuvent être considérés comme biaisés car les auteurs n'ont pas présenté leurs résultats en fonction du type de chirurgie. Or, il est clair que les écarts acoustiques observés pour les patients par rapport aux témoins sont proportionnels à l'étendue de l'exérèse, et il est donc utile de différencier les patients entre eux selon les types de chirurgies (Sittel & al. 1998, Peretti & al. 2003, Ledda & al. 2006, Roh & al. 2007, Vilaseca & al. 2007). Bien que ne traitant pas spécifiquement et exclusivement des cordectomies de type II-III, l'article de Vilaseca & al. (2008) et les mémoires de Mirghani (2009) ou Claeys (2010) sont, à mon sens, les plus proches de notre travail en termes de résultats et de méthodologie. Vilaseca & al. (2008) ont mis en parallèle les données acoustiques de 42 patients repartis en 2 sous-groupes (I-II-III et V) avec celles d'un groupe contrôle apparié pour les voyelles [a] et [i]. Les résultats de la  $F_0$  des deux sous-groupes de patients sont statistiquement supérieurs à ceux des témoins. De plus, les patients traités par cordectomie de type V présentent une  $F_0$  bien plus élevée que les patients traités par cordectomie de type I-II-III. Aussi, dans une analyse prospective de 7 patients traités par cordectomie de type I, Mirghani (2009) a montré que la  $F_0$  reste significativement plus élevée à 3 et 6 mois postopératoire (PO) par rapport aux témoins pour la voyelle [a]. Il ajoute que la  $F_0$  a tendance à s'améliorer avec le temps et que la différence statistique entre les deux groupes n'est plus significative à 12 mois PO. Selon lui, cette élévation de la  $F_0$ , associée à une augmentation de la PSG, serait exclusivement due à la fibrose cicatricielle qui augmenterait la rigidité du pli vocal. Dans sa recherche, Claeys (2010) a comparé les résultats acoustiques de 23 patients (regroupés en 3 groupes selon le type d'exérèse : I-II/ III-IV et V-VI) à ceux d'une cohorte de témoins appariés sur la voyelle [a]. Celle-ci a mis en évidence qu'à tous les temps d'évaluation postopératoire, la  $F_0$  était statistiquement plus élevée pour l'ensemble des sous-groupes par rapport au groupe contrôle, et plus nettement pour le troisième groupe. L'auteur explique ses résultats par la présence d'une fibrose cicatricielle que Mirghani (2009) avait déjà évoquée précédemment. Cette homogénéité dans la présentation des résultats met en évidence l'importance du type d'exérèse dans l'altération vocale : les paramètres acoustiques, et en particulier la  $F_0$ , sont dégradés proportionnellement à l'étendue de l'exérèse.

D'un point de vue physiologique, les auteurs interprètent l'augmentation de la  $F_0$  par la présence d'une fibrose qui serait responsable d'une augmentation de la tension du pli vocal et d'un mécanisme compensatoire d'hyperconstriction. Selon eux, la baisse de masse vibrante serait à l'origine de ce phénomène, ce que confirment Haddad & al. (2006). En ce qui nous concerne, l'hypothèse selon laquelle la fibrose cicatricielle entraîne une augmentation de la  $F_0$  nous semble être une explication convaincante. En effet, le laser employé dans ces types de cordectomies enlève la muqueuse, le ligament et une partie variable du muscle vocal du pli vocal opéré. Les tissus sont alors reséqués,

ce qui tend à diminuer la masse vibrante du pli vocal et à affiner et raidir ce dernier. En conséquence, plus la chirurgie est importante et plus la quantité de tissu emporté est grande. De ce fait, moins il y a de masse vibrante et plus le pli vocal est tendu en post-opératoire. Cette configuration morpho-physiologique particulière du pli vocal liée à l'intervention est donc directement mise en cause dans l'élévation de la fréquence fondamentale chez les patients. Les images 5.7 et 5.8 prises en nasofibroscopie chez deux patients différents à 12 mois postopératoire corroborent cette hypothèse liée à l'acte chirurgical. Malgré une glotte inflammatoire et un certain oedème du pli vocal droit, nous pouvons remarquer que le pli vocal gauche du patient n°2 est affiné par rapport à son homologue controlatéral. De ce fait, la masse vibrante est amoindrie et le pli vocal davantage tendu. De la même manière, le pli vocal gauche du patient n°9 semble suivre cette configuration post-chirurgicale.



**Figure 5.7** – Image prise par vidéo nasofibroscopie (25 images/sec) des plis vocaux en situation pré-phonatoire du patient n°2.



**Figure 5.8** – Image prise par vidéo nasofibroscopie (25 images/sec) des plis vocaux en situation respiratoire du patient n°9.

En outre, il semble intéressant de commenter brièvement les résultats des études qui opposent cordectomie laser et radiothérapie et/ou cordectomie par voie externe. En effet, bien que ces recherches divergent du point de vue de notre méthodologie en n'associant aucune valeur normative, ces conclusions pourraient valider la pertinence du traitement par cordectomie par voie endoscopique. Les recherches de Policarpo & al. (2004), Krengli & al. (2004), Nunez-Batalla & al. (2008), mettent en avant des valeurs de  $F_0$  plus élevées pour le groupe de patients irradiés par rapport au laser. Au contraire, McGuirt (1994) présente des conclusions où les  $F_0$  des patients opérés par voie endoscopique semblent statistiquement plus élevées que celles des patients irradiés. Dans ce sens, bien que les différences ne soient pas statistiquement significatives, Sjögren & al. (2008) notent que la  $F_0$  des patients ayant subi une cordectomie laser est plus importante que celle des patients irradiés (145 Hz contre 156 Hz). Aussi, dans une recherche longitudinale, Jotic & al. (2011) tentent de comparer les résultats acoustiques de trois groupes de patients : un groupe ayant subi une chirurgie laser (G1), un groupe ayant subi une chirurgie par voie externe (G2) et un groupe de patients irradiés (G3). Les auteurs ont pu mettre en évidence une  $F_0$  plus élevée à 12 mois postopératoire pour le G1 (168,5 Hz) par rapport aux deux autres groupes (147,8 Hz pour G2 et 161,4 Hz pour G3). En outre, une recherche de Schindler & al. (2004) compare un groupe de patients traités au laser et un groupe de patient ayant subi une cordectomie par voie externe : la  $F_0$  est plus haute lors d'une chirurgie par voie externe (168,4 Hz vs 153,63 Hz).

Finalement, les résultats semblent hétérogènes d'une étude à une autre mais avec un léger avantage pour la cordectomie laser du point de vue de la  $F_0$ . Certains auteurs rapportent une augmentation de  $F_0$  après radiothérapie du fait de l'augmentation de la raideur du pli vocal traité, conséquence directe de la fibrose post-traumatique (Dagli & al. 1997). D'autres remarquent une baisse de  $F_0$  après radiothérapie du fait de la présence d'un oedème muqueux (Stoicheff & al. 1983). Finalement, ces résultats variables restent totalement dépendants des mécanismes phonatoires adoptés par les patients et reposent largement sur les séries étudiées, sur les nombres de séances de radiothérapie et sur les suivis établis.

Cette étude a montré que les propriétés acoustiques des voyelles sont conservées après cordectomie laser de type II-III. L'ordre universel concernant la  $F_0$  intrinsèque des voyelles est préservé après cette chirurgie. L'élévation de la fréquence de vibration des plis vocaux est la conséquence directe des changements anatomico-physiologiques liés à l'acte chirurgical.

### 5.3.3.2 Le jitter

**Différence de jitter entre nos deux populations.** Comme nous l'évoquions plus haut, il n'existe aucune différence significative entre nos deux populations pour le paramètre du jitter : les valeurs du jitter des patients restent plus élevées que celles des témoins pour les trois voyelles sans que cela soit statistiquement significatif.

Les valeurs du jitter des voyelles hautes des patients sont plus élevées que celle de la voyelle basse [a]. En effet, les voyelles fermées des patients semblent plus perturbées que leurs voyelles ouvertes. Les écarts-types élevés observés pour chacune des voyelles dénotent d'importantes irrégularités pour la production de celles-ci. Nous ne pouvons donc pas tirer de conclusions claires. Au contraire, la valeur du jitter de [a] des témoins est plus élevée que celles des voyelles hautes. Aussi, il faut préciser que cette voyelle basse présente l'écart-type le plus important par rapport aux deux autres voyelles. En effet, les différentes productions restent irrégulières d'un sujet à un autre, ce qui tend à remettre en cause la fiabilité de cette hiérarchie. Finalement, nous pouvons considérer que le jitter des voyelles tenues à hauteur et intensité confortables n'est pas influencé par le type de voyelle réalisé (Orlikoff 1995).

Les conclusions de Haddad & al. (2006), Vilaseca & al. (2008), Mirghani (2009) et Sjögren & al. (2009) vont à l'encontre de nos résultats. Les résultats de Haddad & al. (2006) sur la voyelle [e] mettent en évidence une différence significative pour le jitter entre le groupe de patients (3,18%) et le groupe de référence (0,76%). Néanmoins, il est important de repréciser le fait que les patients ne sont pas catégorisés selon le type de chirurgie. Aussi, pour les deux voyelles [a] et [i] étudiées, Vilaseca & al. (2008) montrent que les pourcentages du jitter sont significativement plus altérés pour les deux groupes de patients par rapport à la cohorte de référence. Nous ne nous intéresserons pas au type V, mais les résultats pour le groupe I-II-III comparés aux témoins sont les suivants : 2,5% contre 1,5% pour [a] et 2,8% contre 1,1% pour [i]. Sjögren & al. (2009) notent que le jitter moyen de leur étude (0,57%) est plus élevé que le seuil de normalité ( $<0,4\%$ ) défini dans le manuel de Praat. Néanmoins, les auteurs précisent que ce pourcentage s'insère dans la moyenne de normalité (0,23-2,97) établie pour 68 sujets non dysphoniques (Wuyts & al. 2000). Comme nous l'évoquions précédemment, ces données sont à prendre avec précaution car aucune comparaison avec un groupe témoins n'existe et seules des valeurs théoriques sont prises comme valeurs de référence. Aussi, dans son mémoire, Mirghani (2009) a montré qu'en dépit d'une amélioration progressive de l'ensemble des paramètres acoustiques en postopératoire, les résultats du jitter restent toujours significativement altérés par rapport à la population témoin à 12 mois postopératoire.

Contrairement aux études que nous venons d'évoquer, les résultats de Tamura & al. (2003), Peretti & al. (2003), Lopez Llames & al. (2004), Ledda & al. (2006) et Claeys (2010) semblent corroborer nos résultats. L'article de Tamura & al. (2003), basé sur des valeurs acoustiques de référence issues d'un article antérieur, met en évidence des pourcentages du jitter équivalents entre le groupe de cordectomie (1,13%) et le groupe témoin (1,01%). Lopez Llames & al. (2004), quant à eux, montrent qu'il n'existe pas de différence statistiquement significative entre la valeur du jitter de leur propre étude (0,72%) et des valeurs issues d'articles de référence : population non dysphonique (0,26%) (Preciado & al. 1998), population traitée par cordectomie laser (1,13%) (Tamura & al. 2003) et population traitée par radiothérapie (0,93%) (Tamura & al. 2003). Peretti & al. (2003) ont mis en relation les résultats acoustiques de 51 patients répartis en deux groupes (G1 = type I-II, G2 = type III-IV-V) en préopératoire et en postopératoire avec un groupe contrôle sur la voyelle [a]. Le pourcentage du jit-

ter pour le G1, bien que plus élevé, n'est pas déviant statistiquement par rapport aux témoins (1,43%). Seul le G2 présente un pourcentage du jitter significativement plus haut (2,96% contre 0,56%). Ces résultats s'expliquent sûrement par l'hétérogénéité des types de chirurgie représentés dans le G2. En outre, les résultats de l'article de Ledda & al. (2006) corroborent les conclusions de Peretti & al. (2003). En effet, les auteurs ont mis en évidence que les pourcentages du jitter pour les cordectomies de type I et II ne semblent pas significativement différents de ceux du groupe témoin (type I = 1,14% ; type II = 1,15%). De la même manière, le mémoire de Claeys (2010) suit cette tendance. Effectivement, celle-ci a mis en évidence qu'à 6 mois postopératoire le jitter, bien que plus élevé, n'est pas significativement plus haut pour ses G1 (1,59%) et G2 (3,30%) par rapport aux témoins (0,60%). La détérioration de ce paramètre intervient significativement pour le G3, c'est-à-dire pour les types V-VI (3,34%). Finalement, ces travaux mettent en évidence que les résultats acoustiques, en particulier le jitter, sont significativement altérés à partir du type III. L'étendue de l'exérèse joue toujours un rôle majeur dans la dégradation des paramètres acoustiques.

Enfin, comme nous avons pu le faire pour le paramètre de  $F_0$ , nous commenterons dans les lignes suivantes les résultats du jitter pour quelques recherches secondaires à notre travail. Les études de McGuirt (1994) et de Krengli & al. (2004) mettent en évidence des pourcentages du jitter plus déviants pour les groupes de patients opérés par chirurgie laser par rapport aux patients irradiés. Au contraire, les recherches de Pollicarpo & al. (2004), Nunez-Batalla & al. (2008) et de Sjögren & al. (2008) rapportent des valeurs du jitter plus élevées pour les patients irradiés par rapport aux patients traités par cordectomie par voie endoscopique. Toutefois, ces recherches ne montrent pas de différence statistiquement significative entre les deux groupes. Aussi, le travail de Jotic & al. (2011) met en évidence que le groupe de patients ayant subi une cordectomie laser présente un jitter plus haut (0,68%) par rapport au groupe de patients traités par cordectomie traditionnelle (0,64%) mais égal au groupe de patients irradiés. Ces conclusions ne sont donc pas révélatrices du degré de perturbation et permettent d'affirmer qu'aucun des traitements étudiés ne prime par rapport aux autres. Ceci est en contradiction avec les résultats de Schindler & al. (2004) qui rapportent que le jitter semble être davantage perturbé pour le groupe des cordectomies laser (3,37%) par rapport aux cordectomies externes (2,48%).

Finalement cette discussion sur le jitter montre l'hétérogénéité des résultats d'une recherche à une autre. Les pourcentages du jitter dépendent des populations étudiées (nombre de sujets et type de chirurgie) mais également des logiciels utilisés qui ont leurs propres seuils de normalité. Toute chose égale par ailleurs, les différents auteurs notent des perturbations en terme de fréquence après cordectomie laser, ce qui implique une élévation du jitter par rapport à ces seuils ou à une population saine. Ces perturbations dépendent de l'étendue de l'exérèse bien que les patients traités par cordectomie de type II-III présentent une voix modérément altérée.

### Synthèse de l'étude n°1 : La $F_0$ intrinsèque et le jitter des voyelles tenues isolées

- Objectif de l'étude : chercher à évaluer la  $F_0$  et le jitter de nos patients et de nos témoins.
- Corpus : parties stables des voyelles tenues et isolées.
- Principaux résultats :
  - conservation d'un ordre caractéristique :  $F_0[a] < F_0[i] < F_0[u]$ ,
  - élévation de la  $F_0$  des patients par rapport aux témoins sur l'ensemble des voyelles,
  - pas de différence significative entre les 2 groupes pour le jitter.
- Conclusions :
  - résultats en adéquation avec les études antérieures bien que les populations et les méthodologies soient différentes,
  - voix modérément altérée après cordectomie de type II-III.

## 5.4 Etude n°2 : la plage de variation de $F_0$

Cette seconde étude a pour objectif principal l'étude des plages de variation de  $F_0$  des patients et des témoins. Après avoir comparé la pertinence d'une échelle hertzienne par rapport une échelle logarithmique dans l'étude de voix pathologiques en particulier, nous nous sommes intéressés aux étendues vocales de nos deux populations sur un corpus d'imitation des voyelles [a, i, u] synthétisées.

Ainsi, le premier objectif de cette étude consiste à déterminer si les  $F_0$  moyennes des imitations des deux populations sont significativement différentes des cibles théoriques attendues (les stimuli créées). Le second objectif de cette étude est de déterminer si les écarts de la  $F_0$  moyenne et du jitter des imitations des patients sont significativement différents de ceux des imitations des témoins.

### 5.4.1 Méthodologie

#### 5.4.1.1 Présentation du corpus synthétique

**La synthèse à formant.** Cette étude se base sur une tâche d'imitation des voyelles [a, i, u] synthétisées à l'aide du logiciel Klatt (Klatt 1980)<sup>4</sup>. Ce logiciel a été conçu sur les bases de la synthèse à formants.

L'un des principaux arguments en faveur de ce type de synthèse est l'élaboration d'une voix modélisée satisfaisante par la manipulation de la fréquence fondamentale, des trois ou quatre premiers formants, de l'intensité globale, des formants de nasalité etc. Ce descendant du « Pattern Playback »<sup>5</sup> propose une panoplie plus grande de paramètres et une précision plus importante dans le contrôle de ces derniers. Grâce à l'automatisation et aux progrès de l'informatique, la voix artificielle est produite à

4. logiciel téléchargeable à cette adresse : [www.asel.udel.edu/speech/tutorials/synthesis/vowels.html](http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html)

5. Synthèse à partir des formes spectrales où les variations de  $F_0$  étaient impossibles et le contrôle de l'intensité difficile, ce qui caractérisait des voix monocordes et insatisfaisantes (Chafcouloff 2004).

partir de données numériques correspondant aux différents paramètres — soit trente-neuf pour le logiciel Klatt — qui s'affichent sous la forme de tableaux (cf. table 5.7). La voix synthétisée est claire. Le simple jeu de combinaisons des paramètres, et en particulier le contrôle des formants et des zones de bruit, permet la création d'une multitude de stimuli.

Symbol	C/V	Min.	Typ.	Max.	Name
DU	C	30	200	5000	Durée du signal (ms)
NWS	C	1	5	20	Intervalle de mise à jour des paramètres (ms)
SR	C	5000	10000	20000	Fréquence d'échantillonnage du signal (Hz)
NF	C	1	5	6	Nombre de formants dans la configuration cascade
SW	C	0	0	1	0=Cascade, 1=Parallèle excité par AV
G0	C	0	47	80	Facteur de gain général du système (dB)
F0	V	0	100	500	Fréquence de la fondamentale (Hz)
AV	V	0	60	80	Amplitude du voisement (dB)
AVS	V	0	0	80	Amplitude du voisement quasi-sinusoidal (dB)
FGP	V	0	0	600	Fréquence du résonateur glottique "RGP" (Hz)
BGP	V	50	200	2000	Bande passante du résonateur glottique "RGP" (Hz)
FGZ	V	0	2500	5000	Fréquence de l'antirésonateur glottique "RGZ" (Hz)
BGZ	V	100	1500	9000	Bande passante de l'antirésonateur glottique (Hz)
BGS	V	100	200	1000	Fréquence du résonateur glottique "RGS" (Hz)
AH	V	0	0	80	Amplitude de l'aspiration (dB)
AF	V	0	0	80	Amplitude de la friction (dB)
F1	V	180	500	1300	Fréquence du 1er formant (Hz)
B1	V	30	60	1000	Bande passante du 1er formant (Hz)
F2	V	550	1500	3000	Fréquence du 2e formant (Hz)
B2	V	40	90	1000	Bande passante du 2e formant (Hz)
F3	V	1200	2500	4800	Fréquence du 3e formant (Hz)
B3	V	60	150	1000	Bande passante du 3e formant (Hz)
F4	V	2400	3800	4990	Fréquence du 4e formant (Hz)
B4	V	100	300	1000	Bande passante du 4e formant (Hz)
F5	V	3000	4500	6000	Fréquence du 5e formant (Hz)
B5	V	100	300	1500	Bande passante du 5e formant (Hz)
F6	V	4000	4990	6500	Fréquence du 6e formant (Hz)
B6	V	100	500	4000	Bande passante du 6e formant (Hz)
FNP	V	180	280	700	Fréquence du pôle nasal (Hz)
BNP	V	40	90	1000	Bande passante du pôle nasal (Hz)
FNZ	V	180	280	800	Fréquence du zéro nasal (Hz)
BNZ	V	40	90	1000	Bande passante du zéro nasal (Hz)
AN	V	0	0	80	Amplitude du formant nasal(dB)
A1	V	0	60	80	Amplitude du 1er formant (dB)
A2	V	0	0	80	Amplitude du 2e formant (dB)
A3	V	0	0	80	Amplitude du 3e formant (dB)
A4	V	0	0	80	Amplitude du 4e formant (dB)
A5	V	0	0	80	Amplitude du 5e formant (dB)
A6	V	0	0	80	Amplitude du 6e formant (dB)

Table 5.7 – Le synthétiseur de Klatt. Schéma fonctionnel (en Haut), paramètres de contrôle (en bas).

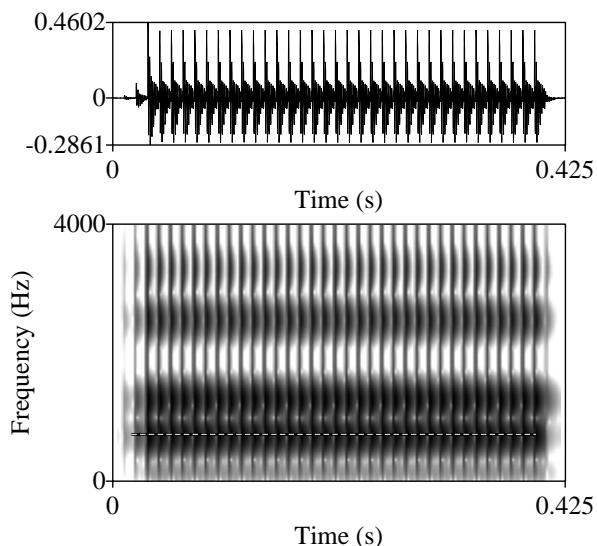
**Création des stimuli synthétiques.** Notre but principal consistait à établir une comparaison entre les plages de variation de  $F_0$  des témoins et des patients. Ainsi, nous avons choisi l'imitation de stimuli de synthèse créés avec le logiciel Klatt. Nous avons élaboré 10 stimuli synthétiques pour la voyelle [a], 10 stimuli pour la voyelle [i] et 10 stimuli pour la voyelle [u], soit 30 stimuli. Nous avons uniquement modifié le paramètre de la  $F_0$  en l'augmentant de façon croissante selon des intervalles de 20 Hz.

Aussi, pour chaque groupe de voyelles, l'intensité et les valeurs formantiques arrondies (selon Calliope 1989) restaient identiques : (i)  $F_1 = 800$  Hz,  $F_2 = 1500$  Hz,  $F_3 = 2700$  Hz,  $F_4 = 4000$  Hz pour la voyelle [a] ; (ii)  $F_1 = 300$  Hz,  $F_2 = 2400$  Hz,  $F_3 = 3400$  Hz,  $F_4 = 3400$  Hz pour la voyelle [i] et (iii)  $F_1 = 300$  Hz,  $F_2 = 800$  Hz,  $F_3 = 2500$  Hz,  $F_4 = 3600$  Hz pour la voyelle [u].

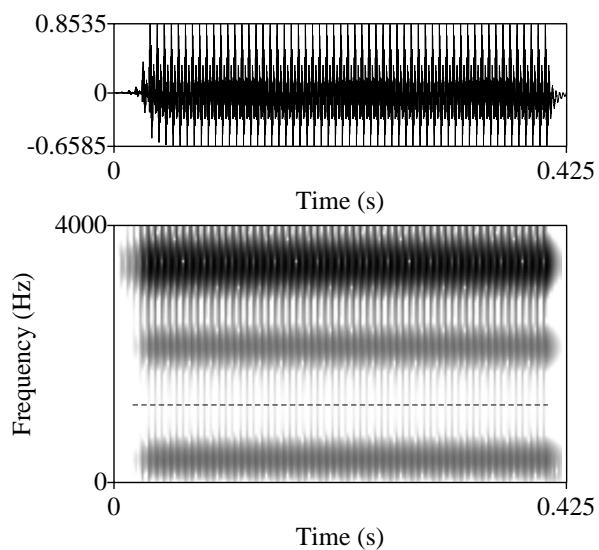
En ce qui concerne le paramètre de la durée, celui-ci n'était pas non plus exploité et la valeur par défaut du stimuli était de 425 ms. En ce qui concerne la  $F_0$ , et comme décrit précédemment, nous avons sélectionné une échelle graduelle selon des intervalles de 20 Hz dans laquelle la valeur la plus basse était 70 Hz et la valeur la plus haute était 250 Hz. Ainsi, nous avons pu synthétiser des voyelles à 70 Hz, 90 Hz, 110 Hz jusqu'à 250 Hz. L'avantage de cet échantillonnage de stimuli est qu'il permet l'analyse d'une plage de variation assez large, du registre bas vers un registre plus haut, en passant par des fréquences intermédiaires.

Les stimuli créés ont été par la suite enregistrés et stockés dans un format .wav sur l'ordinateur personnel de l'expérimentateur.

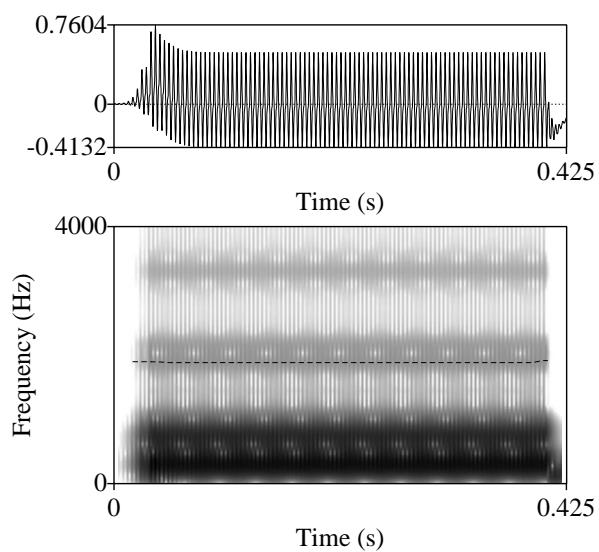
Les figures 5.9, 5.10, 5.11 ci-dessous représentent des exemples de spectrogrammes de voyelles que nous avons pu synthétiser à l'aide du logiciel Klatt :



**Figure 5.9** – Spectrogramme à bandes larges de la voyelle [a] synthétisée sous Klatt avec une  $F_0$  — ligne pointillée — de 90 Hz (Praat).



**Figure 5.10** – Spectrogramme à bandes larges de la voyelle [i] synthétisée sous Klatt avec une  $F_0$  — ligne pointillée — de 150 Hz (Praat).



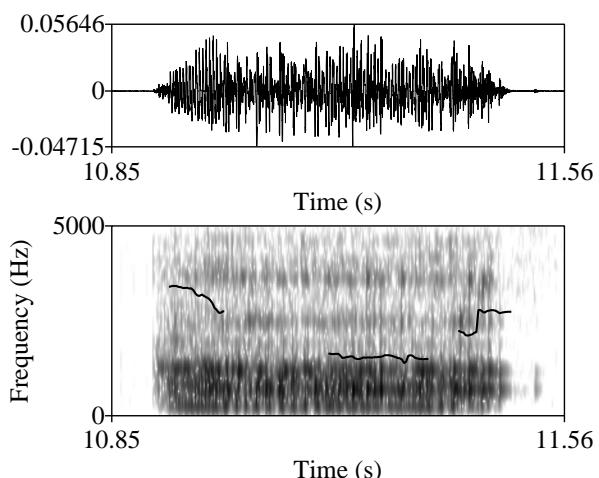
**Figure 5.11** – Spectrogramme à bandes larges de la voyelle [u] synthétisée sous Klatt avec une  $F_0$  — ligne pointillée — de 230 Hz (Praat).

### 5.4.1.2 Recueil de données

La méthodologie expérimentale reste identique à celle décrite précédemment (cf. chapitre 4). Pour la tâche d'imitation, les sujets avaient en plus un casque de la marque Sennheiser de type HP 518 pré-réglé à 74 dB sur les oreilles leur permettant d'écouter les stimuli de synthèse. Les voyelles artificielles étaient écoutées les unes après les autres avec une possibilité de répétition. Après chaque écoute, les sujets devaient imiter au mieux ce qu'ils venaient d'entendre.

### 5.4.1.3 Elimination des stimuli problématiques

En analysant nos données, nous nous sommes rapidement aperçus que les imitations à 70 Hz pouvaient poser problème. En ce qui concerne la plupart des témoins, ceux-ci atteignaient rarement la cible théorique. En effet, le fondamental moyen d'un homme est d'environ 120 Hz (Vaissière 2006, entre autres). En conséquence, les imitations à 70 Hz représentaient soit un effort vocal pour atteindre la cible, soit des écarts trop importants par rapport à ces stimuli. En ce qui concerne les patients, d'autres difficultés pouvaient apparaître. A l'instar des témoins, ces derniers ne pouvaient imiter correctement cette fréquence basse, leurs  $F_0$  étant intrinsèquement plus élevées. Aussi, cette complexité était repérable sur les spectrogrammes. Comme l'illustre la figure 5.12, ces productions étaient très instables en fréquence. Les voyelles étaient craquées et la  $F_0$  difficilement mesurable. Pour toutes ces raisons, nous avons décidé d'éliminer les imitations à 70 Hz de nos deux populations pour nos traitements statistiques. Seules neuf productions pour chacun des sujets ont été prises en compte, de 90 à 250 Hz.



**Figure 5.12** – Spectrogramme à bandes larges représentant l'imitation de la voyelle [a] à 70 Hz du patient n°1 (Praat).

### 5.4.1.4 Mesures et paramètres étudiés

Le premier objectif de cette étude consistait à déterminer si les F<sub>0</sub> moyennes des imitations des deux populations étaient significativement différentes des cibles théoriques attendues (i.e. les valeurs des stimuli créées). Nous cherchons principalement à observer l'adaptation et le contrôle de la F<sub>0</sub> des patients. Nous avons analysé chacune des imitations de nos deux populations à l'aide du logiciel Praat en réalisant des spectrogrammes à bandes larges avec une étendue de fréquences de 0 à 4000 Hz ( $n = 540$ , soit 3 voyelles  $\times$  20 sujets  $\times$  9 fréquences à imiter). Après avoir pris soin de sélectionner uniquement les parties stables des voyelles, en éliminant les attaques et les parties finales, nous nous sommes intéressée à la F<sub>0</sub> moyenne et au jitter moyen de l'ensemble de ces productions. Tous les résultats en Hertz ont été convertis en demi-tons à l'aide de la formule exprimée dans la section suivante. Toutes les mesures en Hertz et leurs équivalents selon l'échelle logarithmique ont fait l'objet d'un traitement statistique avec Statview et R.

### 5.4.1.5 Une échelle logarithmique : le demi-ton

Dans le cas présent, deux échelles s'offrent à nous : l'échelle linéaire et l'échelle logarithmique.

La première échelle, la plus courante, est la plus répandue en phonétique. Il s'agit d'une échelle fréquentielle, exprimée en Hertz. Les valeurs permettent principalement de distinguer entre voix d'hommes et voix de femmes et semblent pertinentes dans la mesure des registres et des différences inter et intra-locuteurs.

Une conversion en une échelle logarithmique — les demi-tons et les octaves en sont des exemples — est pertinente sur le plan perceptif. Ce type d'échelle rend compte de la distance entre deux valeurs tonales (intervalles musicaux), c'est-à-dire que chaque intervalle représente une proportion fréquentielle identique. Finalement, la conversion entre l'échelle fréquentielle et l'échelle en demi-ton se définit comme suit (Fant & al. 2002) :

$$(8) \quad \text{Conversion en échelle logarithmique (demi-ton)} \\ St = 12[\log(Hz/100)/\log 2]$$

où  $St$  représente l'unité et où le niveau de référence est fixé à 100 Hz. Cette formule permet de centrer l'échelle en demi-ton où 100 Hz est égal à 0 demi-ton.

## 5.4.2 Présentation des résultats

### 5.4.2.1 Comparaison moyenne théorique vs observée

Afin de déterminer si l'écart de la moyenne des imitations est significativement différent d'une moyenne théorique attendue (i.e. le stimulus créé), nous avons conduit une série de tests-t univariés sur l'ensemble de nos données. Les tableaux 5.8 et 5.9 résument nos résultats pour nos deux populations pour la voyelle [a]. Nous relevons

aucune différence entre les tests effectués sur les données en Hertz et les valeurs en demi-tons. En ce qui concerne les témoins, nous constatons qu'il existe une différence statistiquement significative dans les basses fréquences. La  $F_0$  moyenne des témoins durant l'exercice de l'imitation de [a] s'écarte de façon significative des valeurs cibles entre 90 et 110 Hz. L'imitation à 130 Hz reste encore quelque peu problématique pour notre cohorte saine car la différence entre la  $F_0$  moyenne et la cible atteint une p-value proche du seuil de significativité ( $p = .0826$ ). Dans les moyennes et hautes fréquences, c'est-à-dire entre 150 et 250 Hz, nous ne notons pas de différence statistiquement significative entre la  $F_0$  des productions des témoins et les valeurs cibles, ce qui dénote une bonne imitation de la part de notre groupe. En ce qui concerne les patients, nous constatons qu'il y a une différence statistiquement significative dans les basses et moyennes fréquences. La  $F_0$  moyenne des patients s'écarte significativement des cibles à atteindre entre 90 et 150 Hz. Nous considérons que la différence de la  $F_0$  moyenne des patients avec la moyenne théorique de 170 Hz est relativement significative ( $p = .0888$ ). Dans les hautes fréquences, à savoir entre 190 Hz et 250 Hz, nous ne relevons pas de différence statistiquement significative entre les imitations des patients et les moyennes théoriques (sauf cas isolé pour la cible à 210 Hz), ce qui dénote une bonne reproduction des voyelles.

	<b>a90</b>	<b>a110</b>	<b>a130</b>	<b>a150</b>	<b>a170</b>	<b>a190</b>	<b>a210</b>	<b>a230</b>	<b>a250</b>
<i>Témoins</i>	$\bar{x}$	111	125.9	142.1	157.5	170.6	192.5	211.6	234.6
	$s$	18.7	21.4	19.6	18.1	23.9	21	7	19.7
	$t_{(9)}$	3.547	2.351	1.953	1.309	.079	.380	.740	.733
	$p$	.0062	.0432	.0826	.2229	.9384	.7128	.4781	.4820
<i>Patients</i>	$\bar{x}$	154.5	183.7	180.1	188	200.5	216.6	241.3	230.7
	$s$	30.6	32.4	35.7	48.8	50.6	55.9	44.9	74
	$t_{(9)}$	6.663	7.200	4.438	2.465	1.907	1.504	2.200	.032
	$p$	<.0001	<.0001	.0016	.0358	.0888	.1669	.0554	.9755

**Table 5.8** – Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en hertz) de la voyelle [a] selon les témoins et les patients ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(9)}$  la valeur du test de Student pour un degré de liberté égal à 9 et  $p$  la probabilité) ( $n = 180$ ).

	<b>a90</b>	<b>a110</b>	<b>a130</b>	<b>a150</b>	<b>a170</b>	<b>a190</b>	<b>a210</b>	<b>a230</b>	<b>a250</b>	
<i>Témoins</i>	$\bar{x}$	3.42	2.13	1.41	.75	-.08	.14	.13	.29	-.29
	$s$	2.82	2.77	2.28	1.90	3.87	1.84	.57	1.41	3.08
	$t_{(9)}$	3.832	2.429	1.952	1.245	-.0112	.237	.698	.648	-.0302
	$p$	.0040	.0381	.0827	.2445	.9136	.8178	.5026	.5330	.7693
<i>Patients</i>	$\bar{x}$	9.01	8.59	5.33	3.28	2.32	1.61	2.13	-.92	-.14
	$s$	3.83	3.49	3.53	5.24	4.70	5.34	3.22	6.50	3.44
	$t_{(9)}$	7.446	7.790	4.775	1.977	1.558	.955	2.092	-.449	-.125
	$p$	<.0001	<.0001	.0010	.0795	.1536	.3645	.0659	.6639	.9033

**Table 5.9** – Résultats des tests de Student (tests-t univariés) pour l’imitation des différentes modalités (exprimées en **demi-tons**) de la voyelle [a] selon les témoins et les patients (où  $\bar{x}$  représente la moyenne arithmétique de l’échantillon,  $s$  l’écart-type non biaisé,  $t_{(9)}$  la valeur du test de Student pour un degré de liberté égal à 9 et  $p$  la probabilité)( $n = 180$ ).

En ce qui concerne la voyelle [i], les résultats sont présentés dans les tableaux 5.10 et 5.11. De même, Nous notons aucune différence entre les tests effectués sur les données en Hertz et les valeurs en demi-tons. Les témoins présentent un *pattern* similaire à celui de la voyelle [a]. Ainsi, il existe une différence statistiquement significative dans les basses fréquences. La  $F_0$  moyenne de notre population s’écarte significativement de la valeurs cible de 90 Hz. Nous considérons que la différence de la  $F_0$  moyenne et la valeur attendue de 110 Hz reste significative car la p-value reste proche du seuil de significativité ( $p = .0741$ ). Dans les moyennes et hautes fréquences, c'est-à-dire entre 130 et 250 Hz, il n'y a pas de différence statistiquement significative entre les  $F_0$  moyennes des voyelles produites et les cibles. En ce qui concerne les patients, nous relevons toujours une différence significative dans les basses et moyennes (voire hautes) fréquences. La  $F_0$  moyenne des patients s’écarte de façon significative de la moyenne théorique entre 90 et 170 Hz. Nous notons également des p-value proches du seuil de significativité pour la  $F_0$  moyenne des imitations des cibles de 190, 210 et 230 Hz. Enfin, nous ne notons pas de différence significative entre la  $F_0$  moyenne des patients et la valeur cible attendue de 250 Hz.

	<b>i90</b>	<b>i110</b>	<b>i130</b>	<b>i150</b>	<b>i170</b>	<b>i190</b>	<b>i210</b>	<b>i230</b>	<b>i250</b>	
<i>Témoins</i>	$\bar{x}$	113.4	130	152.2	167.1	194.2	195.7	222	244	260
	$s$	28	31.3	41	41.9	53.3	32.4	41.4	35.9	43.3
	$t_{(9)}$	2.652	2.020	1.714	1.292	1.438	.553	.922	1.236	.732
	$p$	.0264	.0741	.1206	.2286	.1843	.5935	.3803	.2477	.4826
<i>Patients</i>	$\bar{x}$	143.8	174.1	196.2	206.3	228	238.8	235.1	252.2	260.8
	$s$	58.2	48	43.8	59.9	51.3	71.7	39.4	33	32.2
	$t_{(9)}$	2.924	4.222	4.779	2.970	3.573	2.153	2.017	2.124	1.055
	$p$	.0169	.0022	.001	.0157	.006	.0597	.0744	.0626	.3188

**Table 5.10** – Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en **hertz**) de la voyelle [i] selon les témoins et les patients (où  $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(9)}$  la valeur du test de Student pour un degré de liberté égal à 9 et  $p$  la probabilité) ( $n = 180$ ).

	<b>i90</b>	<b>i110</b>	<b>i130</b>	<b>i150</b>	<b>i170</b>	<b>i190</b>	<b>i210</b>	<b>i230</b>	<b>i250</b>	
<i>Témoins</i>	$\bar{x}$	3.56	2.49	2.27	1.47	1.84	.32	.70	.86	.44
	$s$	4.09	3.79	4.03	3.76	4	2.65	3.21	2.55	3.12
	$t_{(9)}$	2.756	2.074	1.780	1.235	1.457	.376	.691	1.064	.447
	$p$	.0223	.0679	.1089	.2481	.1791	.7157	.5072	.3150	.6652
<i>Patients</i>	$\bar{x}$	6.61	7.25	6.71	4.80	4.71	3.25	1.73	1.46	.61
	$s$	7.93	5.45	4.04	5.37	3.78	5.28	2.95	2.28	2.08
	$t_{(9)}$	2.638	4.203	5.253	2.831	3.943	1.949	1.854	2.022	.932
	$p$	.0270	.0023	.0005	.0197	.0034	.0831	.0967	.0739	.3759

**Table 5.11** – Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en **demi-tons**) de la voyelle [i] selon les témoins et les patients (où  $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(9)}$  la valeur du test de Student pour un degré de liberté égal à 9 et  $p$  la probabilité) ( $n = 180$ ).

Enfin, les tableaux 5.12 et 5.13 mettent en évidence les résultats pour la voyelle [u]. Nous relevons aucune différence majeure entre les tests effectués sur les données en Hertz et les valeurs en demi-tons. Pour les témoins, nous notons toujours une différence statistiquement significative dans les basses fréquences. Une fois encore, la  $F_0$  moyenne de notre population s'écarte significativement de la valeurs cible de 90 Hz. Nous considérons que la différence de la  $F_0$  moyenne et la valeur attendue de 110 Hz reste significative car la p-value reste proche du seuil de significativité ( $p = .0644$ ). Entre 130 et 250 Hz, nous ne notons pas de différence significative entre la  $F_0$  des

voyelles produites et les moyennes théoriques, ce qui dénote une bonne capacité d'imitation. En ce qui concerne notre groupe de patients, nous notons une différence significative sur l'ensemble de l'échantillonnage de fréquences. Ainsi, la  $F_0$  moyenne des productions des patients s'écarte de façon significative des cibles à atteindre entre 90 et 230 Hz. En outre, nous ne notons pas de différence significative entre la  $F_0$  moyenne des sujets et la valeurs cible à atteindre de 250 Hz.

	<b>u90</b>	<b>u110</b>	<b>u130</b>	<b>u150</b>	<b>u170</b>	<b>u190</b>	<b>u210</b>	<b>u230</b>	<b>u250</b>	
<i>Témoins</i>	$\bar{x}$	111.8	126.8	144.2	164.7	188.3	198.4	217.8	236.9	257.4
	$s$	24.3	25.1	31.3	33.7	42.8	20.7	18.9	18.7	27.3
	$t_{(9)}$	2.836	2.107	1.436	1.384	1.349	1.288	1.309	1.168	1.852
	$p$	.0195	.0644	.1847	.1998	.2102	.2299	.2231	.2730	.4163
<i>Patients</i>	$\bar{x}$	140.5	181	216.7	201.9	225	255.9	243	259.7	260.9
	$s$	65.9	53	54.1	56.4	38.6	62.1	43.7	26.2	49.2
	$t_{(9)}$	2.426	4.324	5.064	2.910	4.504	3.354	2.390	3.588	.700
	$p$	.0382	.0019	.0007	.0173	.0015	.0085	.0406	.0059	.5014

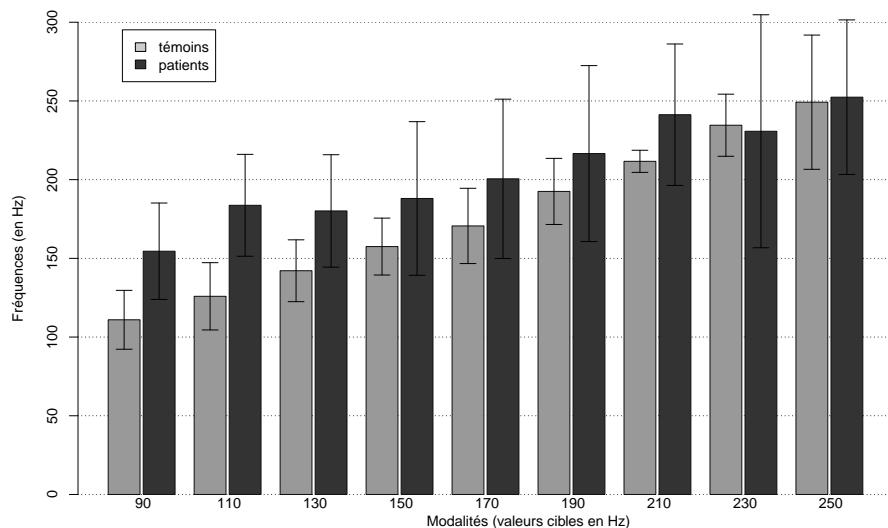
**Table 5.12** – Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en **hertz**) de la voyelle [u] selon les témoins et les patients (où  $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(9)}$  la valeur du test de Student pour un degré de liberté égal à 9 et  $p$  la probabilité) ( $n = 180$ ).

	<b>u90</b>	<b>u110</b>	<b>u130</b>	<b>u150</b>	<b>u170</b>	<b>u190</b>	<b>u210</b>	<b>u230</b>	<b>u250</b>	
<i>Témoins</i>	$\bar{x}$	3.43	2.19	1.48	1.36	1.45	.67	.58	.46	.41
	$s$	3.45	3.11	3.37	3.03	3.29	1.75	1.48	1.37	1.87
	$t_{(9)}$	3.146	2.227	1.388	1.418	1.391	1.210	1.233	1.069	.696
	$p$	.0118	.0529	.1985	.1900	.1975	.2570	.2489	.3131	.5037
<i>Patients</i>	$\bar{x}$	5.47	7.87	8.28	4.50	4.62	4.73	2.31	2.03	.422
	$s$	9.92	5.61	4.83	5.1	3	3.98	2.86	1.68	3.65
	$t_{(9)}$	1.744	4.434	5.420	2.789	4.864	3.759	2.555	3.821	.366
	$p$	.1151	.0016	.0004	.0211	.0009	.0045	.0310	.0041	.7231

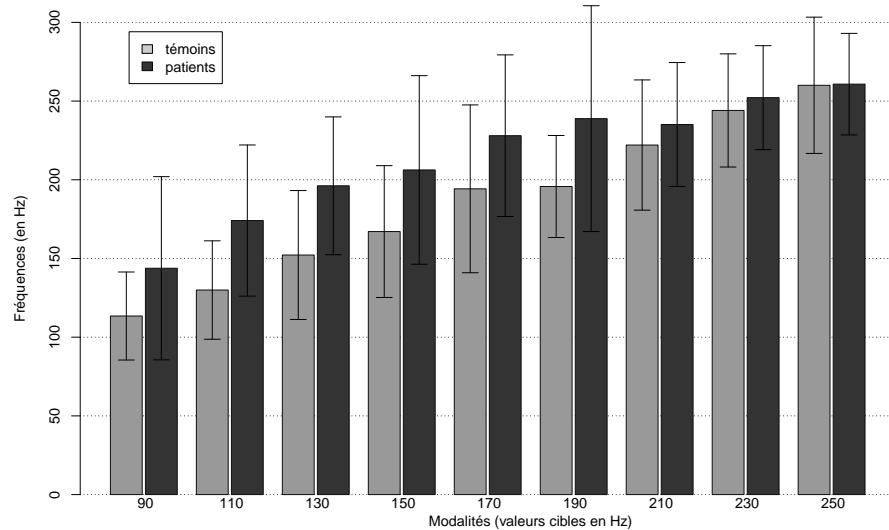
**Table 5.13** – Résultats des tests de Student (tests-t univariés) pour l'imitation des différentes modalités (exprimées en **demi-tons**) de la voyelle [u] selon les témoins et les patients (où  $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(9)}$  la valeur du test de Student pour un degré de liberté égal à 9 et  $p$  la probabilité) ( $n = 180$ ).

### 5.4.2.2 Comparaison inter-groupes

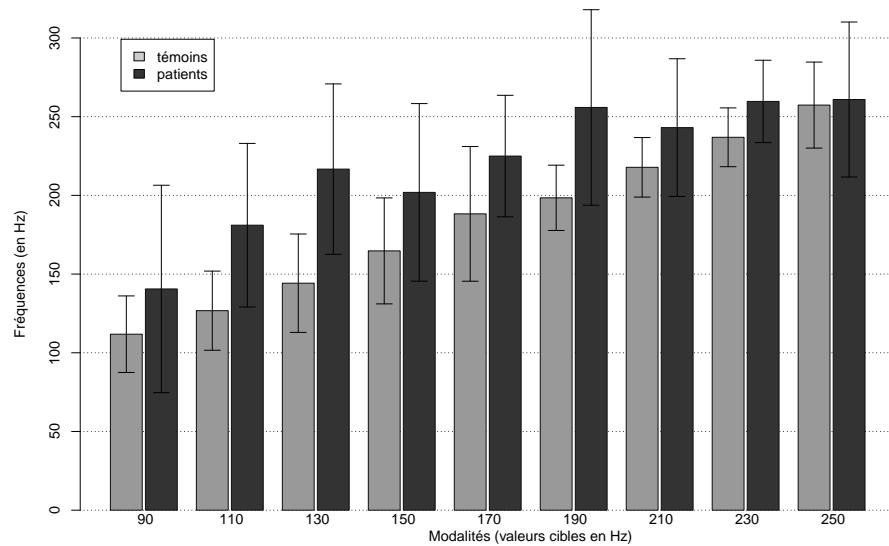
Les graphiques 5.13 , 5.14 et 5.15 mettent en évidence les différences inter-groupes pour chaque modalité et pour les trois voyelles étudiées. Visuellement, nous remarquons que les imitations des témoins suivent une certaine linéarité et l'augmentation de fréquences selon un intervalle de 20 Hz semble assez bien respectée. Au contraire, nous pouvons voir que les productions des patients sont toujours légèrement plus hautes que celles des témoins et que la linéarité évoquée précédemment est moins nette.



**Figure 5.13 –** Valeurs moyennes de la fréquence fondamentale des différentes modalités étudiées pour la voyelle [a] produite par les témoins et les patients (les barres d'erreurs représentent les écarts-types) ( $n = 180$ ).



**Figure 5.14** – Valeurs moyennes de la fréquence fondamentale des différentes modalités étudiées pour la voyelle [i] produite par les témoins et les patients (les barres d'erreurs représentent les écarts-types) ( $n = 180$ ).



**Figure 5.15** – Valeurs moyennes de la fréquence fondamentale des différentes modalités étudiées pour la voyelle [u] produite par les témoins et les patients (les barres d'erreurs représentent les écarts-types) ( $n = 180$ ).

Afin de déterminer si la  $F_0$  moyenne et le jitter moyen des patients pour chaque modalité sont significativement différents de ceux des témoins, nous avons conduit une série de tests-t non appariés sur l'ensemble de nos données. Le tableau 5.14 résume nos résultats pour la voyelle [a]. Dans un premier temps, il est important de noter qu'il n'existe pas de différence majeure entre les tests effectués sur nos données en Hertz et nos données en demi-tons. Dans un second temps, nous relevons une différence statistiquement significative entre nos deux groupes principalement dans les basses et moyennes fréquences entre 90 et 150 Hz. Par ailleurs, nous ne notons pas de différence significative sur le reste de l'échantillonnage (excepté pour la cible à 210 Hz pour laquelle nous avions mis en évidence un écart significatif des productions des patients). En ce qui concerne le jitter, les valeurs statistiques restent plus aléatoires. En effet, la différence inter-groupes est significative uniquement pour l'imitation à 250 Hz. Pour l'ensemble de notre échantillonnage de fréquences, nous ne trouvons pas de différence significative entre nos deux groupes. Il est cependant important de noter que les p-value se rapprochent souvent du seuil de significativité — en particulier pour 90, 130 et 170 Hz — ce qui pourrait provenir d'un problème lié à la dispersion des productions.

Modalités	$F_0$		$1/2$ tons		Jitter	
	$t(18)$	$p$	$t(18)$	$p$	$t(18)$	$p$
a90	3.839	.0012	-3.721	.0016	1.896	.0742
a110	4.714	.0002	-4.586	.0002	1.090	.2902
a130	2.948	.0086	-2.954	.0085	1.885	.0757
a150	1.857	.0797	-1.435	.1683	1.015	.3235
a170	1.691	.1081	-1.592	.1289	1.734	.0999
a190	1.274	.2189	-.826	.4197	1.329	.2005
a210	2.059	.0543	-1.939	.0684	1.648	.1168
a230	-.158	.876	.576	.5718	1.577	.1322
a250	-.155	.8788	-.108	.9150	2.456	.0244

**Table 5.14** – Résultats des différents tests de Student (tests-t non appariés) conduits sur les différentes modalités de la voyelle [a] ( $t_{(18)}$  représente la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité)( $n = 180$ ).

En ce qui concerne la voyelle [i], nos résultats sont présentés dans le tableau 5.15. Une fois encore, la différence en terme d'échelle n'est pas pertinente puisque les seuils de significativité suivent un *pattern* commun. Pour la  $F_0$ , il y a une différence statistiquement significative entre les deux groupes dans les fréquences moyennes, entre 110 et 130 Hz. Par contre, nous ne relevons pas de différence significative inter-groupes pour le reste de l'échantillonnage de fréquences. En ce qui concerne le jitter, nous notons une différence statistiquement significative entre patients et témoins dans les basses fréquences pour les valeurs de 90 et 130 Hz. En outre, il n'y a pas de différence significative inter-groupes pour le reste des fréquences.

	F <sub>0</sub>		1/2 tons		Jitter		
	t(18)	p	t(18)	p	t(18)	p	
Modalités	i90	1.487	.1544	-1.082	.2936	2.431	.0257
	i110	2.436	.0254	-2.265	.0361	.694	.4966
	i130	2.319	.0324	-2.466	.0239	2.184	.0424
	i150	1.694	.1075	-1.611	.1247	1.289	.2138
	i170	1.443	.1663	-1.646	.1172	.501	.6222
	i190	1.732	.1003	-1.573	.1332	.598	.5576
	i210	.722	.4794	-.749	.4637	1.048	.3083
	i230	.727	.6049	-.555	.5855	-.267	.7923
	i250	.043	.9659	-.146	.8856	1.592	.1287

**Table 5.15** – Résultats des différents tests de Student (tests-t non appariés) conduits sur les différentes modalités de la voyelle [i] ( $t_{(18)}$  représente la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité)( $n = 180$ ).

Enfin, le tableau 5.16 met en évidence nos résultats concernant la voyelle [u]. Pour la F<sub>0</sub>, nous observons une différence statistiquement significative entre les deux groupes dans les moyennes et hautes fréquences, entre 110 et 190 Hz. Nous considérons que les différences inter-groupes restent importantes sans être significatives à 150 et 170 Hz car les p-value sont proches du seuil de significativité ( $p = .0903$  et  $p = .0590$  respectivement). Nous pouvons également noter qu'il existe une différence significative inter-groupes pour l'imitation à 230 Hz. Pour le reste de l'échantillonnage de fréquences, il n'y a pas de différence statistique entre les deux groupes. En ce qui concerne le jitter, il existe une différence significative entre les deux groupes dans les moyennes fréquences, entre 130 et 170 Hz. Nous considérons toujours que la comparaison inter-groupes reste significative pour 130 Hz car la p-value reste proche du seuil de significativité ( $p = .0512$ ). Enfin, nous ne notons pas de différence significative entre les deux groupes pour le reste de l'échantillonnage de fréquences.

	F <sub>0</sub>		1/2 tons		Jitter		
	t(18)	p	t(18)	p	t(18)	p	
Modalités	u90	1.294	.2121	-.615	.5462	1.182	.2528
	u110	2.975	.0081	-2.801	.0118	.929	.3652
	u130	3.667	.0018	-3.652	.0018	2.089	.0512
	u150	1.790	.0903	-1.674	.1115	2.160	.0445
	u170	2.015	.0590	-2.25	.0372	2.353	.0302
	u190	2.773	.0125	-2.953	.0085	-.227	.8229
	u210	1.674	.1114	-1.70	.1063	-1.447	.1652
	u230	2.242	.0378	-2.284	.0347	-.537	.5980
	u250	.199	.8445	-.008	.9939	.571	.5752

**Table 5.16** – Résultats des différents tests de Student (tests-t non appariés) conduits sur les différentes modalités de la voyelle [u] ( $t_{(18)}$  représente la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité)( $n = 180$ ).

### 5.4.3 Discussion des résultats

Cette tâche d'imitation avait pour but d'exploiter trois problématiques distinctes. La première consistait à démontrer la pertinence d'une échelle en Hertz par rapport à une échelle de perception dite logarithmique. La deuxième consistait à vérifier les plages de variation des témoins et des patients par rapport à une valeur cible attendue et la dernière avait pour but de comparer les productions des deux groupes en termes de fréquence fondamentale et de jitter.

#### 5.4.3.1 Echelle hertzienne ou échelle logarithmique ?

De manière générale, les résultats présentés ci-dessus ne permettent pas de favoriser une échelle par rapport à une autre. Les seuils de significativité restent sensiblement les mêmes dans les deux cas, ce qui nous amène à des conclusions similaires. A ce titre, nous avons fait le choix de conserver l'échelle hertzienne — plus caractéristique des analyses phonétiques — dans l'exposition des résultats.

#### 5.4.3.2 Comparaison moyenne théorique vs observée

La première partie de la recherche montre clairement que l'exercice n'est pas aisé et que les groupes se comportent différemment face à cette tâche. Dans l'ensemble, les tableaux mettent en évidence une assez bonne imitation des voyelles par les témoins. Seules les fréquences basses entre 90 et 110 Hz semblent poser problème à notre groupe. En effet, de telles fréquences restent rares puisque la voix conversationnelle se situe entre La1 et Mi2, soit 110 et 164,8 Hz chez les hommes (Heuillet & al. 1995). Nous supposons donc qu'un effort vocal a dû être réalisé, ce qui implique des variations intra-sujets plus ou moins importantes et, par conséquent, un éloignement certain par rapport aux cibles attendues. En ce qui concerne les patients, nous constatons que les difficultés apparaissent dans les basses mais surtout les moyennes fréquences. Les moyennes de  $F_0$  des productions des patients sont souvent significativement plus hautes que les stimuli présentés. A titre de synthèse, le tableau 5.17 présente les écarts significatifs par rapport aux cibles attendues pour les trois voyelles selon les sujets. Pour les patients, nous observons que les difficultés en production dépendent des voyelles produites. Ainsi, pour [a] et [i], les différences se trouvent spécifiquement dans les basses et moyennes fréquences, entre 90 et 150 Hz ou 90 et 170 Hz respectivement. Quant à la voyelle [u], les imitations s'éloignent des stimuli de base sur la quasi totalité de l'échantillonnage (de 90 à 230 Hz). Cet ordre semble évoquer la hiérarchie des  $F_0$  intrinsèques des voyelles que nous avons évoquée précédemment (cf. section 5.3.3.1).

A notre connaissance, aucune étude ne s'est intéressée à la problématique de la plage de variation de  $F_0$  après cordectomie. En conséquence, nous ne pouvons qu'émettre des hypothèses relatives à ces résultats. En ce qui concerne les basses fréquences, l'éloignement des productions de nos deux échantillons par rapport à la cible attendue pourrait être lié à un effort, voire à un forçage vocal. Néanmoins, cette hy-

pothèse reste controversée par les conclusions de notre expérience préliminaire sur les voyelles stables. Nous avons pu mettre en lumière que la  $F_0$  des patients était statistiquement plus élevée que celle des témoins pour chacune des trois voyelles. Ainsi, les imitations dans les basses et moyennes fréquences restent difficiles pour nos patients car intrinsèquement leur fréquence fondamentale est plus élevée. Le cas de la voyelle [u] semble particulier. En effet, les imitations des patients s'écartent des cibles sur une portion plus importante de l'échantillonnage. Sachant que la  $F_0$  intrinsèque de cette voyelle est la plus haute, nous pouvons supposer que les patients ont davantage de difficulté à baisser la  $F_0$ .

D'un point de vue articulatoire, les voyelles [a, i, u] se distinguent par leur différence de hauteur et par une position linguale particulière. La production de la voyelle [a] ne nécessite pas de mouvement ample de la langue, ce qui implique une résistance laryngée et une tension du conduit vocal moindres. Celle-ci est centrale, ce qui confère à ce segment une articulation assez simple. Les voyelles [i] et [u] partagent une position haute de la langue, ce qui engendre une pression plus importante sur les structures laryngées, plus particulièrement sur l'os hyoïde, et une tension verticale plus grande du larynx (Netsell & al. 1991, entre autres). L'élévation du larynx pour ces voyelles est responsable d'une augmentation de l'impédance laryngée. La voyelle [i] est antérieure, c'est-à-dire que sa production nécessite un contact linguo-palatal dans la région coronaire. La voyelle [u], quant à elle, est postérieure et est réalisée par un contact linguo-palatal dans la région dorsale. Une vocalisation dépend donc d'une mise en vibration efficace des plis vocaux associée à une articulation précise, c'est-à-dire à la synchronisation de deux gestes distincts pour les voyelles fermées ou d'un geste unique — la vibration des plis vocaux — pour la voyelle basse. Ainsi, les voyelles hautes semblent plus complexes à réaliser car les sujets doivent faire coïncider un déplacement vertical et horizontal de la langue — ce qui est source d'une plus grande tension laryngée — avec une vibration glottique active. Les patients, quant à eux, sont confrontés à un défaut vibratoire et doivent faire face parallèlement à une articulation linguale précise pour la production de ces voyelles. L'association de l'ensemble de ces gestes et les difficultés vibratoires évidentes sont des arguments en faveur d'une production plus problématique de ces segments. Aussi, les plis vocaux et, plus largement le conduit vocal, sont davantage tendus pour la production des voyelles hautes, ce qui implique des résistances plus importantes (Netsell & al. 1991, entre autres) et, *in extenso*, rend la tâche d'imitation plus complexe.

		90	110	130	150	170	190	210	230	250
[a]	témoins									
	patients									
[i]	témoins									
	patients									
[u]	témoins									
	patients									

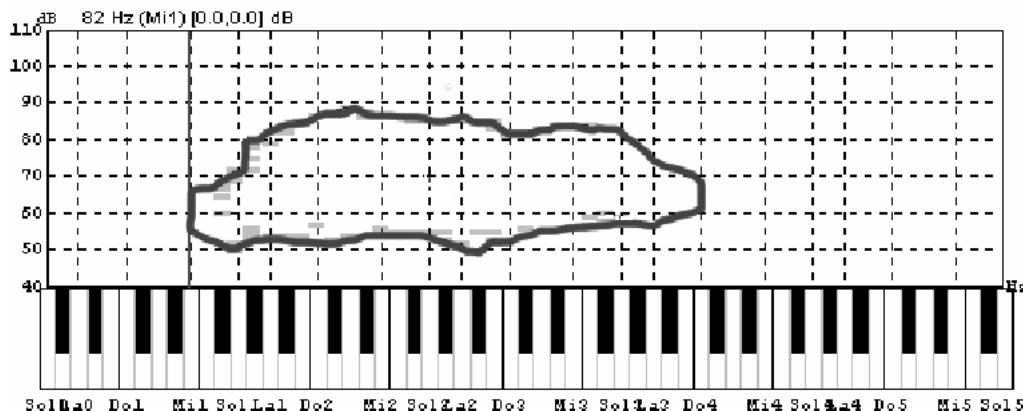
**Table 5.17** – Tables synthétiques des écarts significatifs (en gris) entre les cibles attendues et les imitations des voyelles [a], [i] et [u] selon les groupes des témoins et des patients (d'après les données exprimées en Hertz).

#### 5.4.3.3 Comparaison inter-groupes

La comparaison inter-groupes nous montre qu'une fois encore, la  $F_0$  des patients est statistiquement plus élevée que celle des témoins pour chacune des modalités à imiter et ce pour les trois voyelles. Les différences statistiques entre les deux groupes se trouvent principalement dans les moyennes fréquences, ce qui n'est pas étonnant puisque les écarts des patients avec les cibles se situent principalement dans cette zone fréquentielle. En ce qui concerne le jitter, les différences inter-groupes s'orientent également dans les basses et moyennes fréquences. Néanmoins, ces résultats sont à prendre avec précaution car les zones où apparaissent ces différences ne sont pas forcément nettes et définies. Aussi, le test sur les voyelles stables n'avait pas montré de différence significative entre nos deux populations pour le jitter.

Comme nous le précisons plus haut, aucune recherche ne s'est réellement attachée à définir les plages de variation de  $F_0$  de patients traités par cordectomie laser avec le type de méthodologie que nous avons mis en place. En conséquence, la discussion de nos résultats s'avère difficile. Nous pouvons mentionner les études de Sittel & al. (1998), Mirghani (2009) et Claeys (2010) qui utilisent le phonétogramme pour déterminer le profil vocal des patients. Néanmoins, une fois de plus, ces travaux n'incluent pas forcément les mêmes types de cordectomies que notre recherche, ce qui rend difficile la comparaison des résultats.

Le phonétogramme permet de tracer l'étendue vocale dans un plan avec les fréquences en abscisse et les intensités en ordonnée, en reportant sur un graphe les reproductions de notes de musique aux intensités les plus basses et les plus fortes. En reliant les points, nous obtenons un tracé qui renseigne sur la dynamique tonale et la dynamique de l'intensité du sujet (cf. figure 5.16).



**Figure 5.16** – Exemple d'un phonétogramme d'un locuteur de 58 ans sain. Sa dynamique tonale s'étend du Mi 1 (soit 82 Hz) au Do 4 (soit 523 Hz) et sa dynamique d'intensité varie de 50 à 90 dB (Teston 2004).

Celui-ci est rarement utilisé en recherche pour l'analyse de la voix mais est plutôt utilisé comme un outil de rééducation, ce qui explique le peu de travaux l'incluant. Sittel & al. (1998), dans une étude sur 72 patients ayant subi une cordectomie de type I, II, III et IV, ont mis en évidence que les dynamiques fréquentielles les plus altérées correspondaient au type IV. Les auteurs précisent que les résultats vocaux sont davantage corrélés aux mécanismes phonatoires adoptés après la chirurgie et ne sont pas forcément liés à l'étendue des tissus reséqués. Ils soulignent également que la préservation de la commissure antérieure est primordiale dans l'étendue des dynamiques fréquentielles et d'intensités. L'étude longitudinale de Claeys (2010) tend à confirmer ces résultats. Celle-ci a conclu que la linéarité anatomo-fonctionnelle n'a pas été retrouvée dans ses données. En effet, c'est le G2 (III-IV) qui présente la dynamique fréquentielle la plus restreinte par rapport au G1 (I-II) et au G3 (V-VI). Enfin, après avoir relevé les fréquences minimales et maximales et les intensités minimales et maximales pouvant être produites entre le 1er et le 12ème mois PO sur la voyelle [a] pour des cordectomies de type I, Mirghani (2009) a conclu que ce type de chirurgie n'avait finalement qu'une incidence limitée sur la qualité et l'étendue vocale.

### Synthèse de l'étude n°2 : la plage de variation de $F_0$

- Objectif de l'étude : observer les plages de variation de  $F_0$  des patients et des témoins
- Corpus : imitations des voyelles [a, i, u] synthétisées à différentes fréquences (de 90 Hz à 250 Hz selon un intervalle de 20 Hz).
- Principaux résultats :
  - bonne imitation des témoins,
  - difficultés dans les fréquences entre 90 et 150 Hz pour la voyelle [a], entre 90 et 170 Hz pour la voyelle [i] et entre 90 et 230 Hz pour la voyelle [u] des patients,
  - différences inter-groupes significativement différentes dans les moyennes fréquences, principalement entre 90 et 130 Hz pour [a], 110 et 130 Hz pour [i] et 110 et 190 Hz pour [u],
  - $F_0$  des productions des patients plus élevées que celles des témoins,
- Conclusions :
  - hiérarchie dans la difficulté de la tâche qui évoque l'ordre des  $F_0$  intrinsèques des voyelles,
  - résultats qui confirment l'élévation de la  $F_0$  en postopératoire.

## 5.5 Etude n°3 : l'intensité intrinsèque et le shimmer sur des voyelles tenues isolées

Cette troisième étude a pour objectif l'évaluation des paramètres de l'intensité sonore et de son facteur de perturbation, le shimmer. Nous avons voulu comparer les mesures de l'intensité et du shimmer de notre population de patients en relation avec une cohorte de témoins sur les voyelles tenues [a, i, u] isolées.

### 5.5.1 Méthodologie

#### 5.5.1.1 Recueil des données

Cette expérimentation sur l'intensité et le shimmer a été réalisée sur les mêmes voyelles tenues décrites précédemment. Les sujets avaient comme consigne d'émettre les trois voyelles [a, i, u] à hauteur et intensité confortables pendant au moins deux secondes (cf. section 5.3.1.1). Les situations d'enregistrement et le matériel restaient similaires à ce qui a été décrit dans le chapitre 4.

#### 5.5.1.2 Mesures et paramètres étudiés

Cette étude concerne l'évaluation des paramètres de l'intensité et du shimmer. Elle fait suite à la partie expérimentale sur la  $F_0$  et le jitter. Comme précédemment, nous avons analysé chacune des productions de nos deux populations à l'aide du logiciel Praat. Après avoir pris soin de sélectionner uniquement les parties stables des voyelles tenues, c'est-à-dire en éliminant les attaques et les parties finales, nous avons choisi

de nous intéresser à l'intensité moyenne ( $n = 60$ , soit 3 voyelles  $\times$  20 sujets) et au shimmer moyen ( $n = 60$ , soit 3 voyelles  $\times$  20 sujets) de l'ensemble de ces productions. Toutes les mesures ont fait l'objet d'un traitement statistique avec Statview et R.

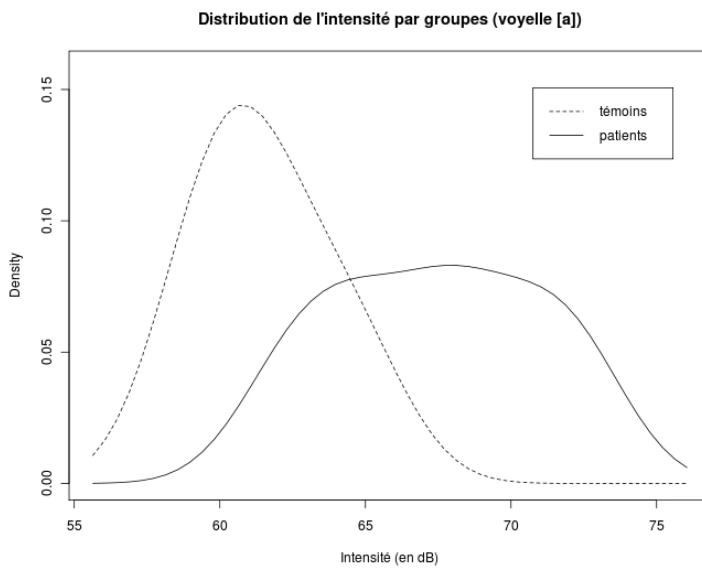
### 5.5.1.3 Tests statistiques préalables

**Test de Lilliefors.** Le test de normalité de Lilliefors a été appliqué sur l'ensemble des productions des voyelles stables ([a, i, u]). L'hypothèse nulle ( $H_0$ ) pour ce test est que l'erreur est normalement distribuée, c'est-à-dire qu'il n'y a pas de différence entre la distribution observée de l'erreur et une distribution normale. L'hypothèse alternative ( $H_1$ ) est que l'erreur n'est pas normalement distribuée. Les résultats de ce test sont présentés dans le tableau suivant :

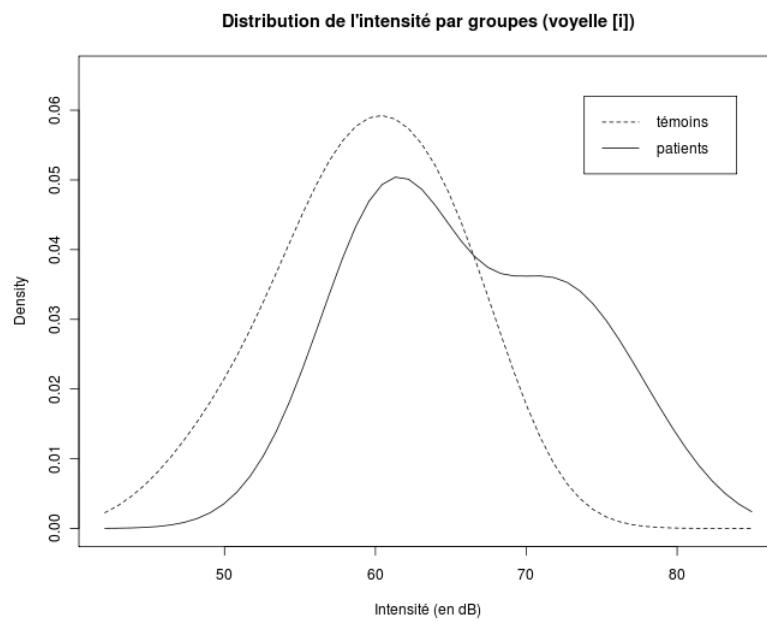
	[a]		[i]		[u]	
	$D$	$p$	$D$	$p$	$D$	$p$
Témoins	.1981	.323	.1433	.8111	.1301	.9021
Patients	.1244	.9317	.2183	.1937	.1557	.7038

**Table 5.18** – Test de normalité de Lilliefors appliqué sur les intensités moyennes des voyelles [a], [i], et [u] ( $n = 60$ ).

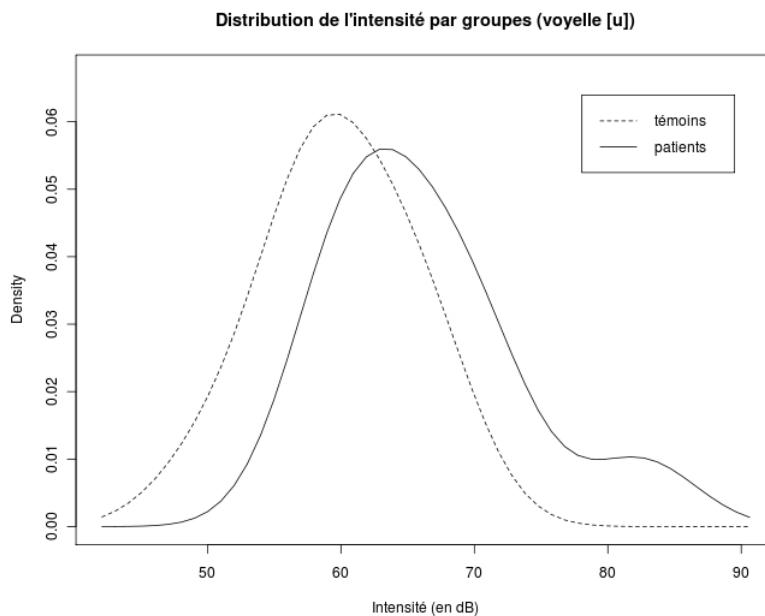
Comme nous l'avons mentionné précédemment, la valeur de  $D$  ne doit pas être supérieure ou égale à .2616 pour accepter  $H_0$  avec un effectif  $n = 10$  et un seuil critique  $\alpha = .05$ . Les résultats de ces tests montrent que les distributions des trois voyelles selon les deux groupes s'intègrent à des distributions suivant une loi normale (cf. figures 5.17, 5.18 et 5.19). En effet, la valeur de  $p$  est supérieure à .05 dans tous les cas et, par conséquent, l'hypothèse nulle doit être acceptée.



**Figure 5.17** – Distribution des valeurs de l'intensité pour les groupes des témoins et des patients pour la voyelle [a] ( $n = 20$ ).



**Figure 5.18** – Distribution des valeurs de l'intensité pour les groupes des témoins et des patients pour la voyelle [i] ( $n = 20$ ).



**Figure 5.19** – Distribution des valeurs de l'intensité pour les groupes des témoins et des patients pour la voyelle [u] ( $n = 20$ ).

**Test F d'homogénéité des variances.** Nous avons donc démontré que les distributions de la variable étudiée étaient normalement distribuées pour nos deux échantillons. Il est maintenant essentiel de vérifier que nos deux populations ont bien des variances homogènes afin de pouvoir les comparer deux à deux (homoscédasticité des variances). Nous avons donc calculé le rapport des variances estimées ( $s^2$ ) entre nos deux populations sur l'ensemble de nos valeurs d'intensité pour les trois voyelles stables. Le tableau 5.4 résume les conclusions de ce test :

	Valeurs de $F$	Valeurs de $p$
[a]	3.114	.1059
[i]	1.685	.4491
[u]	1.961	.3302

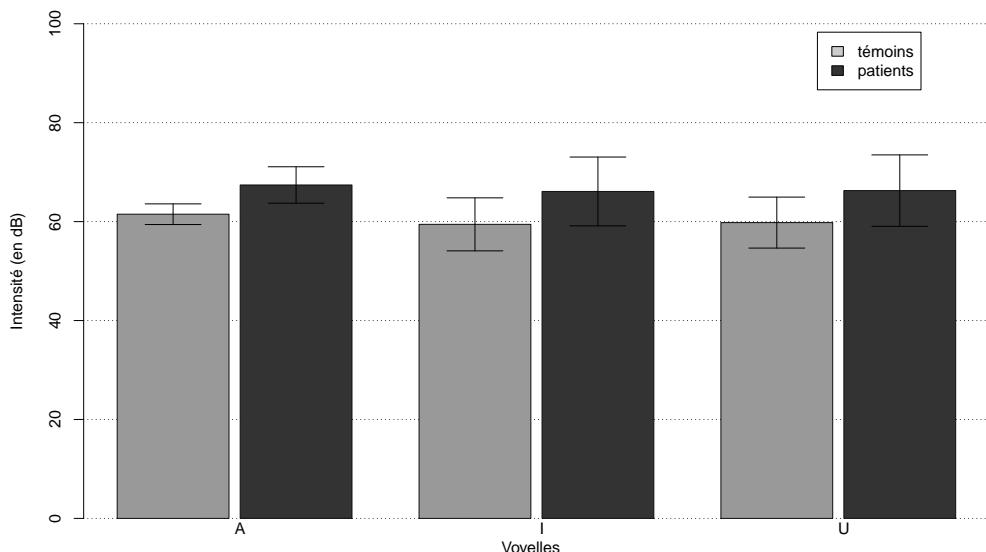
**Table 5.19** – Résultats synthétiques du test F de Fisher-Snedecor sur les valeurs de l'intensité pour les trois voyelles [a], [i] et [u] ( $n = 60$ ).

Ainsi, nous notons que pour nos trois voyelles, les seuils de significativité sont toujours supérieurs à .05, ce qui implique que  $H_0$  n'est pas rejetée et que par conséquent les variances des deux distributions sont homogènes. Les populations forment bien deux groupes indépendants et comparables.

## 5.5.2 Présentation des résultats

### 5.5.2.1 L'intensité

La figure 5.20 met en évidence les différences inter-groupes pour le paramètre de l'intensité pour les trois voyelles étudiées. Visuellement nous remarquons que les intensités moyennes des patients sont légèrement plus hautes que celles des témoins pour chacune des voyelles.



**Figure 5.20** – Valeurs moyennes de l'intensité (en dB) des voyelles [a], [i], [u] des témoins et des patients (les barres d'erreurs représentent les écarts-types) ( $n = 60$ ).

Afin de déterminer si les intensités moyennes des patients sont significativement différentes de celles des témoins, nous avons conduit une série de tests-t non appariés sur les données issues de la production de nos trois voyelles stables. Le tableau 5.20 met en évidence les résultats de cette expérience pour le paramètre de l'intensité moyenne :

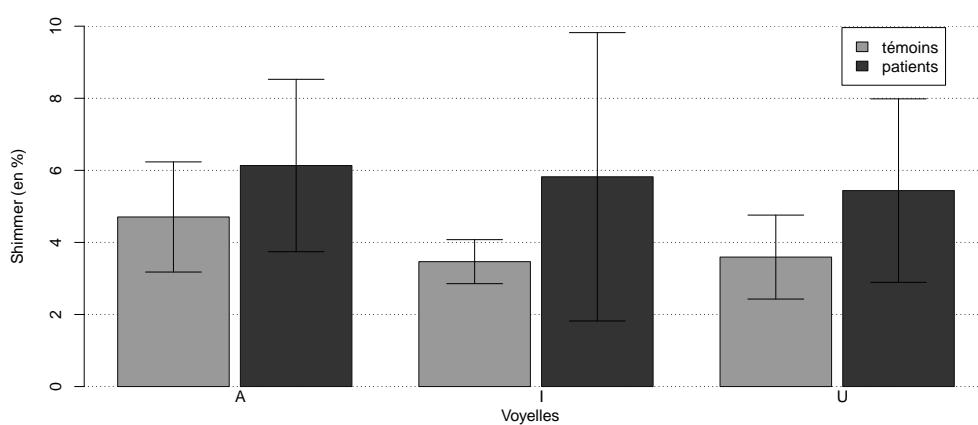
	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	62.20	68.4	59.65	66.2	59.90	66.67
$s$	2.5	4.1	5.12	5.4	4.72	5.84
$t_{(18)}$		4.12		2.78		2.85
$p$		.0007		.0122		.0106

**Table 5.20** – Résultats des tests-t non appariés conduits sur les valeurs de l'intensité (en dB) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Nous remarquons que pour nos deux populations les valeurs moyennes de l'intensité de la voyelle basse [a] sont plus élevées que celles des voyelles hautes [i] et [u]. Ainsi, cet ordre caractéristique peut être résumé de cette manière : I[a]>I[i,u]. Aussi, pour les trois voyelles, nous trouvons une différence statistiquement significative entre les valeurs moyennes de l'intensité des témoins et celles des patients. Pour [a], la p-value entre les deux populations est de  $p = .0007$ ; pour [i], la p-value est de  $p = .0122$  et enfin pour [u] la p-value est de  $p = .0106$ .

### 5.5.2.2 Le shimmer

En ce qui concerne le paramètre du shimmer, la figure 5.21 met en évidence les différences inter-groupes pour les trois voyelles étudiées. Visuellement nous remarquons que les pourcentages du shimmer des patients sont plus élevés que ceux des témoins pour chacune des voyelles.



**Figure 5.21** – Valeurs moyennes du shimmer (en %) des voyelles [a], [i], [u] des témoins et des patients (les barres d'erreurs représentent les écarts-types) ( $n = 60$ ).

Afin de déterminer si les pourcentages de shimmer moyen des patients sont significativement différents de ceux des témoins, nous avons conduit une série de tests-t non appariés sur les données issues de la production de nos trois voyelles stables. Le tableau 5.21 résume les principaux résultats de ces tests :

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	4.71	6.13	3.47	5.82	3.59	5.44
$s$	1.53	2.39	.61	4.001	1.17	2.55
$t_{(18)}$		1.59		1.84		2.08
$p$		.129		.0823		.0517

**Table 5.21** – Résultats des tests-t non appariés conduits sur les valeurs du shimmer (en pourcentages) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Bien que les valeurs moyennes du shimmer des patients soient plus élevées que celles des témoins, nous pouvons remarquer qu'il n'existe aucune différence statistiquement significative entre les deux populations. Seules les voyelles [i] et [u] présentent un seuil de significativité proche de .05 entre les deux groupes avec, respectivement,  $p = .0823$  et  $p = .0517$ .

### 5.5.3 Discussion des résultats

#### 5.5.3.1 L'intensité

**Intensité intrinsèque.** Nous avons pu remarquer que les intensités moyennes de nos deux populations suivent un ordre de grandeur caractéristique puisque l'intensité de [a] est supérieure à l'intensité de [i] et [u], soit  $I[a] > I[i,u]$ . Cet ordre est prévisible et reflète l'intensité intrinsèque — ou spécifique — des voyelles. Comme nous l'avons déjà évoqué dans le chapitre précédent, la notion d'intensité renvoie à l'ampleur d'une grandeur physique. C'est une énergie proportionnelle au carré de l'amplitude d'un son pur (Martin 2008). De manière générale, l'intensité vocale est directement liée aux variations de pression et aux conditions aérodynamiques autour de la glotte (Ladefoged & McKinney 1963 ; Isshiki 1964 ; Holmberg & al. 1988). Les chercheurs ont également pu mettre en évidence le rôle majeur du conduit vocal dans les ajustements de l'intensité, et plus spécifiquement les changements de dynamiques articulatoires (Dromey & al. 1995 ; Geumann & al. 1999 ; Geumann 2001). Dans ce sens, une étude physique récente se basant sur le modèle à deux masses (Ishizaka & Flanagan 1972) a cherché à démontrer l'influence du couplage entre la source et le conduit vocal pour la puissance intrinsèque des voyelles. Par un principe de simulation des différentes

configurations du tractus vocal liée à des coefficients de commandes préétablis (où  $P_s$  renvoie à la PSG et  $Q$  est une variable aléatoire qui renvoie aux divers réglages du filtre) pour la production des voyelles orales du français, les auteurs se sont intéressés aux caractéristiques intrinsèques de l'intensité. Selon eux, ces caractéristiques peuvent être imputables soit à la source soit aux configurations du conduit vocal. Après avoir mesuré l'intensité du signal de la source pour deux ensembles distincts de commandes et pour des couplages correspondants à toutes les voyelles, ces derniers ont pu relever de très faibles différences entre les niveaux d'intensité, ce qui leur a permis de rejeter la première hypothèse. Par contre, une seconde expérimentation concernant les simulations du conduit vocal a permis d'affirmer que la puissance intrinsèque dépend essentiellement de la fonction de transfert du tractus vocal (Teffahi & Kherouf 2010).

Néanmoins, peu de recherches se sont intéressées à ce paramètre car celui-ci semble complexe et requiert une méthodologie rigoureuse pour éviter les problèmes de calibrage. L'intensité sonore paraît être le paramètre le moins exploité dans les travaux contemporains. Cependant, ce déficit relatif n'exclut pas la présence d'études exploratoires dans plusieurs langues. Les premières études concernent majoritairement la langue anglaise. Ainsi, Fairbanks & al. (1950), Lehiste & Peterson (1959) et Peterson & Lehiste (1961) ont pu relever des différences de 3,5 à 5,1 dB entre les productions des voyelles [a] et [i]. Pour le hongrois, une étude de Fonagy (1966) a pu mettre en évidence des écarts de 10 dB entre les voyelles [a] et [i] en discours lu contre 6 à 7 dB en discours spontané, ce qui suggère un effet de surarticulation. À l'autre extrême, l'étude sur le japonais de Nishimura (1977) révèle un écart intrinsèque mineur entre ces voyelles, de l'ordre de 2,5 dB. Pour le français standard, les conclusions des études de Rossi (1971), Di Cristo (1985) ou Bartkova & al. (1993) montrent un écart significatif d'environ 3 dB entre les voyelles hautes et basses. Les résultats pour le français québécois suivent les conclusions du français standard et montrent un écart de 3 à 4 dB entre les deux types de voyelles (Ouellon & al. 1993). Finalement, même si nous remarquons des différences de valeurs relatives à l'intensité intrinsèque d'une recherche à l'autre, toutes les conclusions convergent néanmoins vers un même postulat : les voyelles basses ou ouvertes sont réalisées avec une intensité spécifique supérieure aux voyelles hautes ou fermées. À la lumière de ces résultats, Di Cristo (1985) a établi deux groupes distincts pour classifier les voyelles du français : les voyelles dites fortes (les voyelles moyennes et basses) qui ont une intensité intrinsèque élevée et les voyelles faibles (les voyelles hautes) dont l'intensité spécifique est moins importante. Ces différences d'intensité intrinsèque seraient liées majoritairement à l'aperture de la mâchoire et à la position verticale de la langue (Rossi 1971, Geumann & al. 1999, Geumann 200, Möbius 2003, entre autres). Dans ce sens, Rossi (1971 :129) précise qu'il existe :

« une corrélation très étroite entre le diamètre du conduit vocal aux incisives et l'intensité spécifique de chaque voyelle. En un mot, l'intensité spécifique est fonction du degré d'ouverture »

Sur le plan acoustique, des études ont montré que l'ouverture plus importante des mâchoires pour la production de voyelles en voix forte implique une augmentation de la  $F_0$  mais également du F1 (Stevens 1998). Cette élévation du F1 est aussi reconnue par

Teffahi & Kherouf (2010), qui précisent qu'il s'agit d'une caractéristique acoustique déterminante pour l'intensité intrinsèque des voyelles. De même, Vaissière (2001 :7) ajoute que ce lien étroit entre intensité et F1 joue un rôle dans la perception :

« les voyelles [i] et [u] sont faibles perceptivement alors que [a] est plus saillant ».

D'un point de vue physiologique, l'intensité vocale est directement liée au réglage de l'amplitude de la variation du débit d'air au cours de la vibration glottique. Ainsi, l'intensité dépend donc de la PSG et des débits d'air spécifiques, de la forme glottique, de la force d'adduction des plis vocaux et de la configuration du tractus (Teffahi & Kherouf 2010). Finalement, nous pouvons affirmer que la puissance intrinsèque des voyelles hautes est diminuée par la forme du conduit vocal. Ainsi, la production des voyelles hautes implique un élargissement de la cavité pharyngale, provoquant un effet modérateur de l'excitation du signal (Lehiste & Peterson 1959, Möbius 2003).

**Différences d'intensité entre nos deux populations.** Comme nous avons pu le démontrer dans cette recherche, les intensités intrinsèques des voyelles des patients sont significativement plus élevées que celles des témoins. La discussion de ces résultats semble difficile car, à notre connaissance, aucune étude ne s'identifie à la nôtre d'un point de vue méthodologique. Ce manque d'étude résulte vraisemblablement d'une connaissance partielle de ce paramètre et de ses corrélats ainsi qu'une difficulté dans le recueil des données. Les uniques recherches qui traitent de l'intensité après cordectomie par voie endoscopique sont des études comparatives avec un groupe de patients irradiés (McGuirt 1994, Sjögren & al. 2008 et Motta & al. 2008, Jotic & al. 2011). Ainsi, nous ne pouvons comparer ces valeurs avec notre étude car nous ne possédons pas de valeurs de référence. Néanmoins, nous décrirons les conclusions de ces quelques travaux afin de dégager une tendance générale du comportement de la puissance vocale des patients après de tels traitements. Seul le mémoire de Mirghani (2009) semble être le plus proche de notre étude tant du point de vue de la méthodologie que des valeurs d'intensité, bien que ses conclusions divergent par rapport à notre étude.

La recherche de McGuirt (1994) reste très évasive quant à la méthode de prise de mesures de l'intensité. En conséquence, nous ne pouvons pas en tirer de conclusions formelles. L'auteur précise toutefois qu'il existe une différence de 2 à 3 dB entre le groupe de patients traités par cordectomie laser et le groupe de patients irradiés. Pour la voyelle [a], il relève une moyenne de 49,7 dB pour le laser et une moyenne de 50,25 dB pour la radiothérapie. De la même manière, pour la voyelle [i], ce dernier relève une moyenne de 48,8 dB pour le laser et une moyenne de 51,35 dB pour la radiothérapie. Néanmoins, aucune valeur de référence pour ces voyelles n'est adoptée. Nous ne pouvons donc pas établir de comparaison avec une norme pour les voyelles tenues. Cependant, l'auteur réalise également une comparaison des intensités moyennes de ces deux groupes sur un paragraphe lu avec une référence issue de la littérature (70 dB). Les deux groupes présentent des intensités moyennes équivalentes (43,3 dB pour le laser et 43,5 dB pour la radiothérapie) pour cette tâche de lecture. Ce qui est important de noter ici, c'est l'importante différence qui existe entre les intensités moyennes

des deux groupes de patients et la valeur normative prise par McGuirt. Les matériels de prise de mesures n'étant pas spécifiés, nous pouvons nous interroger sur la validité d'une telle comparaison. L'auteur souligne, néanmoins, que les traitements sélectionnés dans le cadre de carcinomes des plis vocaux entraînent une baisse significative de la puissance vocale, ce qui est contraire à nos conclusions.

De ce point de vue, les résultats de Motta & al. (2008) contredisent également nos conclusions. En effet, les auteurs notent que quel que soit le sous-groupe analysé, les valeurs moyennes de l'intensité pour la voyelle [a] sont toujours inférieures à celles de la population de témoins. En effet, cette recherche présente une moyenne de 72,05 dB pour la cohorte de sujets contrôles alors que les moyennes d'intensité pour l'ensemble des sous-groupes mis en évidence s'échelonnent de 38 à 56,93 dB. Néanmoins, une fois encore, le matériel de prises de mesures n'est pas décrit dans cet article, ce qui peut porter à caution la différence de plus de 15 dB relevée entre les deux groupes.

Au contraire, les résultats de Sjögren & al. (2008) semblent tendre vers nos conclusions. En effet, bien qu'aucun groupe de témoins n'atteste d'une élévation de l'intensité sonore après cordectomie laser, cet article met en évidence une plus forte intensité sonore pour le groupe de patients traités par chirurgie laser par rapport au groupe de patients irradiés. Sans que la différence soit statistiquement significative, les auteurs rapportent une intensité moyenne de 60,53 dB pour la radiothérapie et 70,80 dB pour le laser sur la voyelle [a].

La recherche de Jotic & al. (2011) corrobore également partiellement nos résultats. Effectivement, les auteurs ont pu relever les intensités moyennes sur des phrases lues pour les groupes de cordectomies laser, cordectomies par voie externe et radiothérapie. Bien que le corpus soit différent du nôtre et qu'aucun groupe contrôle ne serve de comparaison à cette étude, Jotic & al. (2011) notent une intensité sonore plus élevée après cordectomie par voie endoscopique. Les auteurs relèvent une intensité moyenne de 76,4 dB pour le groupe de patients ayant subi une cordectomie laser, une intensité moyenne de 74,15 dB pour le groupe de patients ayant subi une cordectomie classique et une intensité moyenne de 70,9 dB pour le groupe de patients irradiés. Les résultats ne sont pas significatifs mais cette hiérarchie selon le traitement nous donne une certaine tendance quant à leur efficacité sur la puissance vocale. Nous pouvons tendre vers deux hypothèses principales : la cordectomie laser permet une meilleure conservation des qualités de la puissance vocale ou la cordectomie laser entraîne différentes compensations articulatoires ou aérodynamiques, responsables de cette élévation. A ce stade, nous ne pouvons choisir l'une ou l'autre des hypothèses car nous n'avons pas de valeurs normatives à disposition pour établir une comparaison.

Enfin, comme nous le mentionnions précédemment, le mémoire de Mirghani (2009) reste le plus proche de notre travail d'un point de vue méthodologique. Effectivement, bien que la prise de mesures soit différente de la nôtre — mesures aérodynamiques effectuées avec l'Aérophone II sur la voyelle [a] pour l'auteur *vs* mesures acoustiques sous Praat sur les trois voyelles [i, a, u] dans la présente recherche — l'auteur établit une comparaison inter-groupes entre une population témoin et une population de patients de type I. Celui-ci observe que les valeurs de l'intensité moyenne sur la voyelle tenue prononcée à hauteur et intensité confortables par l'ensemble des patients res-

tent très proches de la moyenne de référence (74,3 dB). Ce travail ne met donc pas en évidence de différence significative entre les deux groupes, ce qui est contraire à la présente étude même si nous sommes consciente que deux chirurgies distinctes sont concernées.

Cette discussion montre la difficulté pour l'étude de l'intensité de la voix. En effet, les différentes méthodologies utilisées (corpus, matériels, logiciels) rendent les comparaisons difficiles d'une recherche à une autre. De plus, les comportements individuels sont variables, ce qui rend l'interprétation des résultats complexe. Malgré tout, nos résultats préliminaires tendent à confirmer que la cordectomie laser de type II-III n'altère pas la puissance vocale des patients.

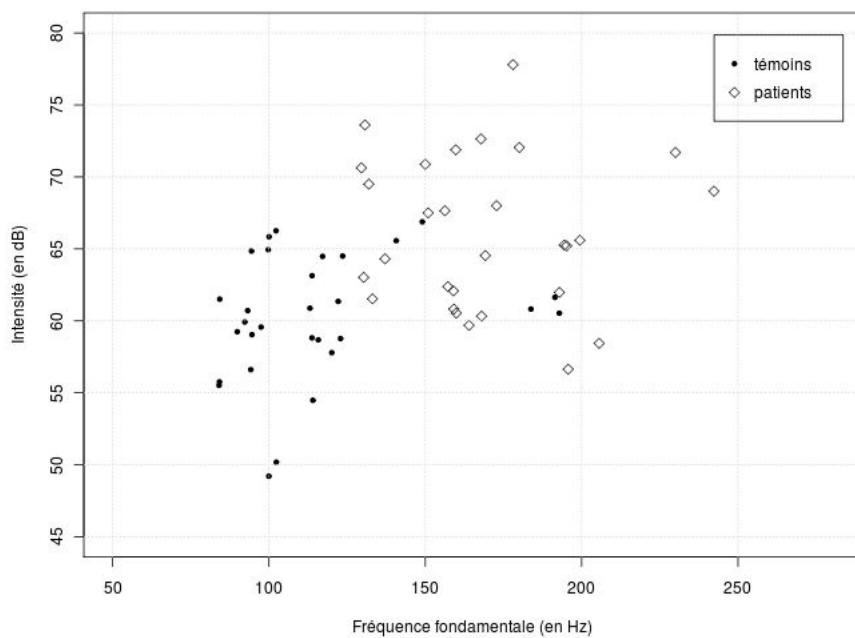
**Hypothétique relation entre intensité et fréquence fondamentale.** Nous avons déjà pu noter que les hausses de l'intensité sonore étaient étroitement corrélées à la PSG (Ishikki 1964 ou Sundberg & al. 1993, entre autres) et donc plus largement à la fréquence fondamentale (Titze 1992, Henrich 2001). Des études sur le forçage vocal ont pu mettre en évidence une certaine corrélation positive entre intensité et  $F_0$  : plus la  $F_0$  est élevée et plus l'intensité aurait tendance également à croître (Black 1961, Baken 1987 par exemple).

La première partie de ce chapitre met en évidence des valeurs de  $F_0$  intrinsèques statistiquement plus élevées pour notre groupe de patients par rapport aux témoins pour les trois voyelles analysées.

Notre but est donc de vérifier si les valeurs élevées de  $F_0$  des patients expliquent l'élévation de leurs intensités intrinsèques ou inversement. Nous avons donc effectué des tests de corrélation entre les différentes valeurs d'intensité et de  $F_0$  pour nos deux populations pour toutes les voyelles confondues et pour les voyelles indépendamment les unes des autres. Nous ne présenterons pas ces résultats en détails car le nombre limité de sujets représente, selon nous, un biais dans l'analyse statistique. Néanmoins, le tracé du graphe 5.22 résumant la répartition de l'ensemble des voyelles en fonction de leurs intensités et  $F_0$  spécifiques permet de dégager une tendance générale pour la corrélation. Nous observons que deux « paquets » se distinguent : les valeurs des patients regroupées en haut à droite et les valeurs des témoins davantage compactées en bas à gauche. Pour généraliser, nous remarquons que les mesures des voyelles des témoins se situent autour de 100 Hz et sont comprises entre 55 et 65 dB. En ce qui concerne les patients, les mesures des voyelles oscillent autour de 175 Hz et sont comprises entre 60 et 72 dB. Nous observons également une certaine linéarité dans le graphique (en particulier pour les témoins), ce qui nous laisse supposer une certaine relation de cause à effet entre ces deux paramètres acoustiques. Il existe potentiellement un lien entre l'intensité et la  $F_0$  : les valeurs plus élevées de  $F_0$  sont corrélées à des valeurs plus élevées de l'intensité et inversement.

Pour aller plus loin dans notre recherche, nous avons également effectué des tests de corrélation entre ces deux paramètres pour chacune des trois voyelles et pour nos deux populations confondues. Nos résultats semblent davantage concluants. Ainsi, pour la voyelle [a],  $r = .5$ , pour la voyelle [i],  $r = .41$  et pour la voyelle [u],  $r = .5$ . Ces valeurs suggèrent une corrélation moyenne entre  $F_0$  et intensité.

Du point de vue de la pathologie, nous avons déjà abordé le fait que l'élévation de la  $F_0$  était principalement due aux changements anatomiques ainsi qu'à la structure des plis vocaux, conséquences directes de la chirurgie. La diminution de la masse vibrante associée à l'affinement du pli vocal sont des arguments en faveur d'une vibration plus rapide des plis vocaux. Cette hypothèse devra faire l'objet d'une étude plus approfondie faisant appel à une analyse stroboscopique, ce qui permettrait de vérifier l'ondulation muqueuse et la vibration du pli vocal opéré. Nous avons également évoqué le fait que l'élévation de l'intensité dépend de la vitesse d'accolement des plis vocaux : plus ceux-ci s'accolent rapidement et plus l'intensité augmente proportionnellement (Baken 1987). Ainsi, selon nous, l'exérèse liée à la cordectomie est l'explication à cette relation. Effectivement, l'augmentation de la fréquence des cycles vibratoires pour notre groupe de patients fait varier à la fois les paramètres de  $F_0$  et de l'intensité de façon parallèle.



**Figure 5.22 – Corrélation entre les valeurs de  $F_0$  et de l'intensité intrinsèques des voyelles [a], [i], [u] des témoins et des patients ( $n = 60$ ).**

### 5.5.3.2 Le shimmer

**Différences de shimmer entre nos deux populations.** Nous avons pu mettre en évidence que les valeurs du shimmer restent plus élevées pour notre population de patients mais sans que cela soit statistiquement significatif par rapport à notre cohorte

de témoins. Nous pouvons également noter, que pour nos deux populations, les valeurs du shimmer de [a] sont plus importantes que celles des voyelles hautes. En outre, nous devons préciser que les importants écarts-types relevés pour les sujets pathologiques révèlent de fortes irrégularités intra-sujets en production.

Nos résultats sont en contradiction avec les travaux de Peretti & al. (2003), Haddad & al. (2006), Mirghani (2009) et Sjögren & al. (2009). Les travaux de Peretti & al. (2003), basés sur une répartition des patients en deux groupes, montrent des différences statistiquement significatives par rapport à la population contrôle (3,54%). Le pourcentage du shimmer du G1 (type I-II) s'élève à 6,52% en PO, tandis que celui du G2 (type III-IV-V) représente un taux de 8,05%. De la même manière, l'article de Haddad & al. (2006) tend vers ces mêmes conclusions. En effet, les auteurs ont mis en évidence un pourcentage du shimmer statistiquement plus haut pour le groupe des patients (7,79%) par rapport à la cohorte de témoins (3,95%). Comme nous l'avons déjà mentionné précédemment, ces résultats sont à prendre avec précaution car le groupe de patients reste hétérogène du fait que ceux-ci ne sont pas catégorisés selon l'étendue de l'exérèse. Dans leur étude, Sjögren & al. (2009) montrent que le shimmer moyen des patients traités par cordectomie I-II (4,6%) est plus élevé que le seuil de normalité (<3,8%) défini dans le manuel de Praat. Néanmoins, les auteurs précisent que ce pourcentage s'insère dans la moyenne de normalité (0,91-6,71) établie pour 68 sujets non dysphoniques (Wuyts & al. 2000). Une fois encore, il est important de noter que ces valeurs ne sont en aucun cas comparées avec une population de témoins mais que seules des valeurs théoriques servent de référence. Enfin, dans son mémoire sur les cordectomies de type I, Mirghani (2009) a démontré que malgré une amélioration effective des valeurs du shimmer au fil du temps, celles-ci restent significativement perturbées à 12 mois PO (6,35%) par rapport aux témoins (3,31%).

Au contraire, nos résultats semblent corroborer les conclusions des recherches de Tamura & al. (2003), Lopez Llames & al. (2004), Ledda & al. (2006), Vilaseca & al. (2008) et Claeys (2010). Dans une étude où les résultats acoustiques sont comparés avec des valeurs de référence relevées dans un article antérieur, Tamura & al (2003) ne relèvent pas de différence significative entre les deux groupes. Ainsi, le pourcentage du shimmer pour les patients s'élève à 3,80% alors que la valeur de normalité est de 3,34%. Lopez Llames & al. (2004), quant à eux, mettent en évidence qu'il n'existe pas de différence statistiquement significative entre la valeur du shimmer de leur propre étude (4,07%) et des valeurs issues d'articles de référence : population non dysphonique (2,44%) (Preciado & al. 1998), population traitée par cordectomie laser (3,8%) (Tamura & al. 2003) et population traitée par radiothérapie (2,81%) (Tamura & al. 2003). De la même manière, les résultats de Ledda & al. (2006) corroborent partiellement nos conclusions. En effet, selon ces derniers, bien que les pourcentages du shimmer soient plus élevés pour les différents groupes de cordectomies par rapport à la cohorte de référence, il n'existe pas de différence statistiquement significative entre les témoins (3,74%) et les groupes de cordectomies I (5,72%) et II (5,78%). Les différences inter-groupes apparaissent systématiquement à partir du type III. Selon les auteurs, les voix des patients ayant subi une cordectomie de type I ou II seraient relativement stables et proches de la normalité. Dans ce sens, Vilaseca & al. (2008) ne

relèvent pas de pourcentages déviants du shimmer pour le groupe hétérogène de cordectomie I-II-III par rapport aux sujets contrôles. Pour la voyelle [a], le shimmer des témoins s'élève à 0,8% alors que celui des patients est de 1,3%. Enfin, pour la voyelle [i], le shimmer des témoins est égal à 0,7% alors que celui des patients reste relativement stable, à 0,9%. Enfin, les résultats de Claeys (2010) sont à prendre avec précaution. En effet, l'auteur ne relève pas de différence statistiquement significative entre ses témoins (5,06%) et ses différents groupes (G1= I-II (5,61%), G2= III-IV (6,01%), G3= V-VI (5,60%)). Nous ne nous intéresserons pas aux G2 et G3 mais il est toutefois important de noter que le shimmer semble altéré de la même manière pour les trois groupes, ce qui est un fait étonnant compte tenu des différentes exérèses pratiquées. De plus, nous pouvons constater que le groupe de témoins présente un pourcentage du shimmer anormalement élevé par rapport aux différentes valeurs relevées dans la littérature, ce qui pourrait constituer un biais dans l'exposition de ces résultats.

Enfin, comme nous avons pu le faire lorsque nous avons abordé la  $F_0$  et le jitter, il nous semble pertinent d'examiner quelques résultats qui opposent la cordectomie laser avec la radiothérapie et/ou la cordectomie externe. Les travaux de Krengli & al. (2004) et de Nunez- Batalla & al. (2008) mettent en évidence des pourcentages du shimmer plus élevés, sans que cela soit statistiquement significatif, après une cordectomie laser par rapport au traitement par radiothérapie. Au contraire, Policarpo & al. (2004) et Sjögren & al. (2008) donnent l'avantage à la cordectomie. En effet, ces derniers observent que les patients irradiés présentent des valeurs du shimmer plus élevées que les patients traités par cordectomie laser. Dans une étude comparative entre cordectomie laser et cordectomie par voie externe, Schindler & al. (2004) précisent que les deux types de chirurgies présentent des pourcentages du shimmer équivalents (10,50% vs 10,32%). Finalement, Jotic & al. (2011) relèvent des valeurs plus déviantes pour le groupe des patients irradiés (6,99%) par rapport à la cordectomie par voie endoscopique (6,82%) et à la cordectomie traditionnelle (6%). Ces résultats ne sont pas statistiquement différents entre eux, ce qui ne permet pas de favoriser un traitement par rapport à un autre.

A l'instar de notre discussion sur le jitter, cette étude sur le shimmer démontre l'hétérogénéité des résultats d'une recherche à une autre. Les pourcentages du shimmer varient en fonction des populations étudiées, des types de traitements mais aussi des logiciels utilisés, ce qui rend les comparaisons délicates. Les auteurs notent, néanmoins, des perturbations en amplitude après cordectomie laser, ce qui implique une élévation du shimmer par rapport à des sujets sains ou à des seuils de normalité. En ce qui concerne les patients traités par cordectomie de type II-III, nous pensons que ces derniers conservent une voix relativement stable et peu altérée.

### Synthèse de l'étude n°3 : l'intensité intrinsèque et le shimmer sur des voyelles tenues isolées

- Objectif de l'étude : chercher à évaluer l'intensité sonore et du shimmer de nos patients et de nos témoins.
- Corpus : parties stables des voyelles tenues et isolées.
- Principaux résultats :
  - conservation d'un ordre caractéristique :  $I[a] > I[i, u]$ ,
  - élévation de l'intensité des patients par rapport aux témoins pour l'ensemble des voyelles,
  - pas de différence significative entre les 2 groupes pour le shimmer.
- Conclusions :
  - résultats difficilement comparables aux études antérieures compte tenu des différentes méthodologies utilisées,
  - puissance vocale conservée après cordectomie de type II-III.

## 5.6 Excursus : intensité intrinsèque et $F_0$ intrinsèque des voyelles tenues

Dans cette partie, nous cherchons à confirmer ou infirmer nos conclusions établies pour les paramètres de l'intensité intrinsèque et de la  $F_0$  intrinsèque des voyelle tenues. Ces résultats seront confrontés à des mesures prises sur les logatomes afin de mettre en évidence un effet significatif du corpus. Dans le cas où cet effet se confirmerait, nous tenterons une comparaison intra-sujets afin de mettre en évidence des stratégies individuelles en production.

### 5.6.1 Etude n°4 : l'intensité intrinsèque

Nous avons pu au préalable démontrer que les intensités intrinsèques des voyelles tenues isolées des patients étaient statistiquement plus élevées que celles des témoins, avec une différence de 6-7 dB environ entre nos deux populations. Nous avons pu voir que cette différence était majoritairement liée à une différence de  $F_0$ . Ainsi, l'élévation de l'un supposait l'élévation de l'autre.

Néanmoins, les mesures d'intensité demeurent problématiques et demandent une précaution particulière dans l'interprétation des résultats. En effet, ce paramètre reste très changeant et dépend largement de la prise de données et du matériel utilisé. Les mesures inter-locuteurs requièrent un gain réglé et inchangé pendant toutes les acquisitions de données et l'utilisation d'un micro-casque positionné à 3-4 cm de la bouche.

Dans notre cas, le matériel utilisé a été identique pendant tous les enregistrements (cf. chapitre 4). Le gain ayant été réglé lors de notre premier recueil de données et inchangé pendant les enregistrements suivants, il reste toutefois le problème du micro-casque. En effet, nous ne possédons pas de micro-casque calibré et le positionnement aux lèvres peut être aléatoire d'un sujet à un autre. Sans mesurer parfaitement la dis-

tance entre le micro et les lèvres, il est difficile de maintenir celui-ci à distance uniforme pour l'ensemble des locuteurs.

Ainsi, la loi  $1/R$  suivante permet de calculer l'atténuation du son/bruit selon la distance à laquelle le micro est placé de la bouche :

$$(9) \quad L_2 = L_1 - 20\log(r_2/r_1)$$

où  $L_2$  représente la variation recherchée en dB,  $L_1$  la valeur enregistrée en dB à distance  $r_1$ ,  $r_1$  la distance théorique du micro-casque et  $r_2$  la distance aléatoire du micro).

Si l'on considère que par défaut, le micro doit être placé à 4 cm de la bouche, une variation de 1 cm ou plus dans un sens ou dans l'autre peut aboutir à une atténuation ou un gain d'énergie sonore, ce qui peut représenter un éventuel biais dans nos conclusions précédentes. Le tableau 5.22 résume ces différentes variations :

Distance du micro à la bouche	Perte/gain d'énergie (en dB)
3 cm (décalage de 1 cm vers les lèvres par rapport à $r_1$ )	-2,5 dB (soit un gain de 2,5 dB par rapport à $L_1$ )
5 cm (décalage de 1 cm vers l'extérieur par rapport à $r_1$ )	1,94 dB (soit une perte de -1,94 dB par rapport à $L_1$ )
6 cm (décalage de 2 cm vers l'extérieur par rapport à $r_1$ )	3,52 dB (soit une perte de -3,52 dB par rapport à $L_1$ )
7 cm (décalage de 3 cm vers l'extérieur par rapport à $r_1$ )	4,86 dB (soit une perte de -4,86 dB par rapport à $L_1$ )
8 cm (décalage de 4 cm vers l'extérieur par rapport à $r_1$ )	6,02 dB (soit une perte de -6,02 dB par rapport à $L_1$ )

**Table 5.22** – Calculs de l'atténuation/gain (dB) en fonction de la distance du micro à la bouche.

Ces calculs symbolisent un modèle mathématique à partir d'une distance théorique mais doivent être considérés également avec prudence. En effet, il faut également prendre en compte le milieu de propagation de l'onde pour éviter tout écho susceptible de perturber les mesures. Ces calculs font référence à des variations sonores dans des chambres sourdes où tous les murs sont totalement absorbants aux ondes.

De notre point de vue, les enregistrements ont été effectués dans des lieux acoustiquement calmes — et non en chambre sourde —, ce qui n'exclue pas les échos et peut donc jouer sur la différence  $\Delta = L_1 - L_2$ . En outre, nous avons tenté, à chaque enregistrement, de conserver une distance de 3-4 cm entre le micro et la bouche. En conséquence, la marge d'erreur — si elle existe — reste minime, de l'ordre de - 1cm à 2 cm, par rapport à notre distance théorique. Par conséquent, nous devons admettre

une variation possible d'environ [-2 ; 3 dB] par rapport à nos conclusions présentées dans la section 5.5.2. Dans ces conditions, la significativité démontrée entre les deux groupes s'avère affaiblie et il est nécessaire de coupler notre première étude avec une étude intra-sujets — la distance de la bouche au micro restant inchangée pendant toute la durée de l'enregistrement d'un sujet — pour valider nos hypothèses.

### 5.6.1.1 Méthodologie

Pour cette étude, nous avons voulu essentiellement comparer les intensités intrin-  
sèques des voyelles tenues isolées aux valeurs d'intensité prises dans un contexte de  
parole (les logatomes) pour chaque sujet.

**Les voyelles isolées.** le recueil des données et les prises de mesures restent identiques  
à la méthodologie décrite dans la section 5.5.1.

**La parole en contexte ou logatomes.** Les témoins et patients ont été enregistrés dans  
les mêmes conditions. Tous les enregistrements acoustiques ont été effectués dans des  
pièces acoustiquement calmes avec un même protocole et un matériel similaire (cf.  
chapitre 4). Pour une question d'organisation et de logistique, nous avons découpé  
nos logatomes de type CVCVCVC en trois syllabes distinctes :  $CV_1$ ,  $CV_2$  et  $CVC_3$   
(pour une description complète de la méthodologie concernant les logatomes, voir  
le chapitre 6, section 6.6 ). Pour cette expérimentation nous avons choisi d'analyser  
uniquement les logatomes composés d'occlusives, soit 360 occurrences (18 logatomes  
de type CVCVCVC où  $C = [p, t, k, b, d, g]$  et  $V = [a, i, u] \times 1$  répétition  $\times 20$  sujets).

Dans cette étude, nous avons choisi de mêler l'ensemble des logatomes composés  
des différentes occlusives pour nos traitements statistiques en écartant volontairement  
les critères de voisement et de lieu d'articulation. Les prises de mesures de l'intensité  
moyenne sur chacun des logatomes ont été établies à l'aide d'un script Praat modifié  
spécialement pour notre étude<sup>6</sup>. Nous avons fait le choix de nous intéresser unique-  
ment aux voyelles de la position 2 des logatomes. En effet, nous savons que l'intensité  
a tendance à être plus forte en position initiale absolue d'un mot et à décroître sur la fin  
de l'énoncé. Cette position nous assure ainsi une intensité moyenne stable qui pourra  
être comparée avec celle des voyelles tenues isolées. Finalement, nous avons établi une  
moyenne des intensités des voyelles des position 2 des logatomes. Nous traitons donc  
60 items, soit une valeur moyenne de l'intensité en contexte  $\times 3$  voyelles  $\times 20$  sujets.

Nous sommes consciente de la possible présence d'infimes variations co-intrin-  
sèques de l'intensité des voyelles dues aux contextes consonantiques (Di Cristo 1985  
par exemple). Nous avons donc réalisé des tests-t appariés préliminaires pour nos  
deux populations afin de comparer les mesures d'intensité des voyelles [a, i, u] en  
contexte voisé/non voisé et selon le lieu d'articulation des consonnes environnantes.  
Nous constatons que de manière générale, les intensités des voyelles en contexte de  
nos deux populations tendent à être plus fortes — de 1 à 2 dB — en contexte voisé

6. Le script est une version plus basique du script « analyse tout » de C. Gendrot.

mais de manière non significative. Aussi, nous ne notons pas de tendance particulière quant à l'influence du lieu d'articulation sur l'intensité des voyelles. Néanmoins, nous avons choisi délibérément de ne pas exposer nos résultats car ce n'est pas le propos de ce travail. En effet, nous cherchons à mettre en évidence des écarts généraux entre les productions vocaliques tenues à l'isolé et les productions en contexte. Nous ne cherchons pas à démontrer l'influence des caractéristiques intrinsèques des consonnes — qui sont finalement peu pertinentes — sur les intensités moyennes des voyelles.

Afin d'établir un corrélat avec notre troisième étude sur l'intensité intrinsèque des voyelles tenues isolées, nous avons effectué, dans un premier temps, une série de tests-t non appariés préliminaires afin de vérifier l'existence d'une différence significative entre les voyelles en contexte de nos deux populations. Dans un second temps, nous avons réalisé une ANOVA à 3 facteurs — où la variable dépendante était l'intensité et les variables indépendantes représentaient le groupe, le type de voyelle et le corpus — afin d'évaluer, en particulier, un effet significatif de la tâche de production. Tous les tests ont été effectués à l'aide du logiciel Statview, les résultats étaient considérés comme significatifs à partir d'un seuil alpha de tolérance de 5%, soit  $p < .05$ . Dans la mesure où ces résultats généraux démontrent une significativité, nous avons voulu comparer les valeurs moyennes des intensités intrinsèques relevées sur les productions vocaliques isolées ( $n = 60$ ) et les valeurs moyennes des intensités des voyelles des positions 2 des logatomies ( $n = 60$ ) pour chacun des sujets. Un calcul du différentiel entre  $I_{V1} - I_{V2}$  (où  $I_{V1}$  représente la mesure de l'intensité moyenne de la voyelle isolée et  $I_{V2}$  représente la mesure de l'intensité moyenne de la voyelle en contexte) nous permettra d'observer les différences intra-individuelles de production dans les deux tâches.

### 5.6.1.2 Résultats

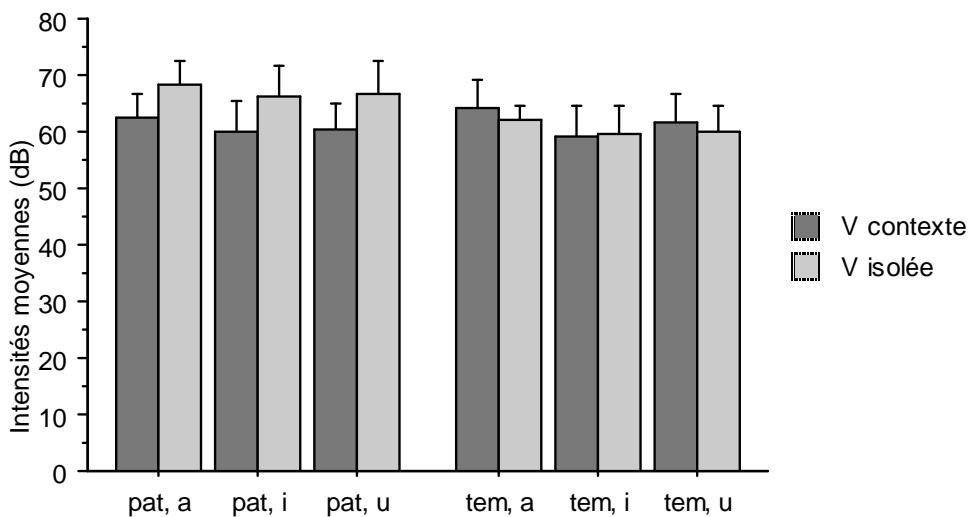
**Différences d'intensité en contexte occlusif entre nos deux populations.** Afin de déterminer si les intensités moyennes en contexte des patients sont significativement différentes de celles des témoins, nous avons conduit une série de tests-t non appariés sur les données issues des productions de nos trois voyelles en contexte occlusif. Le tableau 5.23 met en évidence les résultats de ces tests :

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	64	62.62	59.22	60.12	61.6	60.62
$s$	4.99	3.95	5.17	5.18	5	4.56
$t_{(18)}$		-1.56		0.87		-1.05
$p$		.1209		.3844		.2965

**Table 5.23** – Résultats des tests-t non appariés conduits sur les valeurs de l'intensité (en dB) des voyelles [a], [i] et [u] en contexte ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Nous remarquons que les voyelles en contexte occlusif des patients sont moins intenses que celles des témoins. Néanmoins, nous ne relevons aucune différence statistiquement significative entre nos deux populations pour ces productions. En outre, il semble pertinent de noter que l'ordre établi précédemment concernant les intensités intrinsèques reste inchangé pour nos deux populations : I[a]>I[i,u].

**Résultats de l'ANOVA.** La figure 5.23 présente les valeurs moyennes de l'intensité en fonction de nos deux populations, des contextes vocaliques et des corpus (voyelles isolées vs. voyelles en contexte occlusif). De manière générale, nous constatons qu'il existe une différence significative entre nos deux populations ( $F(1, 108) = 11,898, p = .0008$ ). Cette différence s'explique principalement par l'écart hautement significatif entre nos deux groupes pour les voyelles isolées et tenues (cf. section 5.5.2). En effet, comme nous venons de le montrer, les valeurs d'intensité moyennes en contexte occlusif ne sont pas significativement différentes entre témoins et patients. Aussi, l'ANOVA met en évidence un effet significatif des contextes vocaliques ( $F(2, 108) = 4,175, p = .0179$ ). Ainsi, le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [a, u], avec des p-value respectivement égales à  $p = .0058$  et  $p = .0513$ . Ces résultats étaient attendus dans le sens où ils révèlent l'ordre établi concernant les intensités spécifiques des voyelles, à savoir que les voyelles basses tendent à être plus intenses que les voyelles hautes. Enfin, ce qui semble pertinent pour cette étude est l'effet significatif des corpus ( $F(1, 108) = 8,054, p = .0054$ ). Ainsi, les valeurs moyennes d'intensité des témoins et des patients semblent différentes selon la tâche de production. Cette hypothèse est confirmée par la significativité de l'interaction entre les facteurs « sujets » et « tâche » ( $F(1, 216) = 16,062, p = .0001$ ), c'est-à-dire que l'un des facteurs présentés varie en fonction des modalités du second facteur. En conséquence, nous tenterons, dans la suite de cette étude, d'observer les stratégies intra-individuelles liées à ces deux différentes tâches de production.



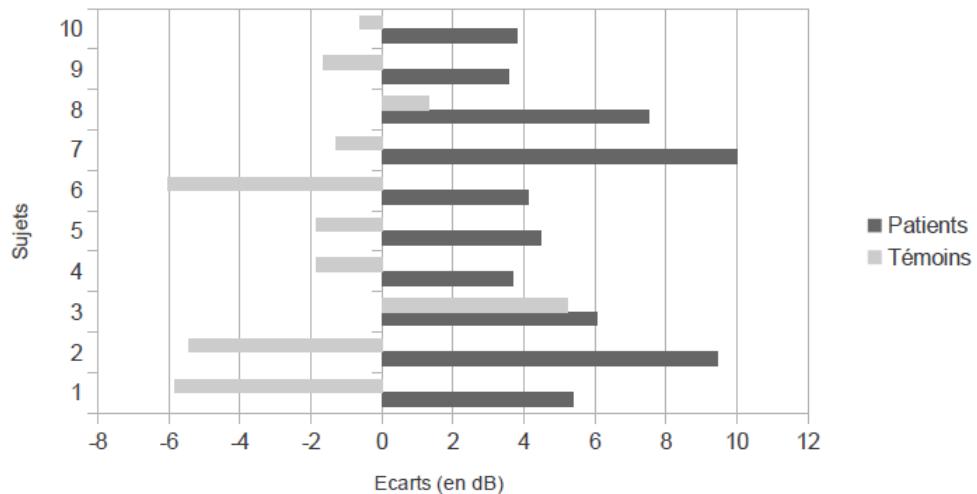
**Figure 5.23** – Valeurs moyennes de l'intensité (en dB) en fonction des sujets, des voyelles et des corpus (les barres d'erreurs représentent les écarts-types) ( $n = 120$ ).

**Ecarts entre  $I_{V1} - I_{V2}$ .** Nous venons de démontrer que le corpus semblait influencer les valeurs d'intensité relevées. Ainsi, il nous semble pertinent d'observer pour chacun des sujets les stratégies mises en oeuvre pour les deux types d'exercices.

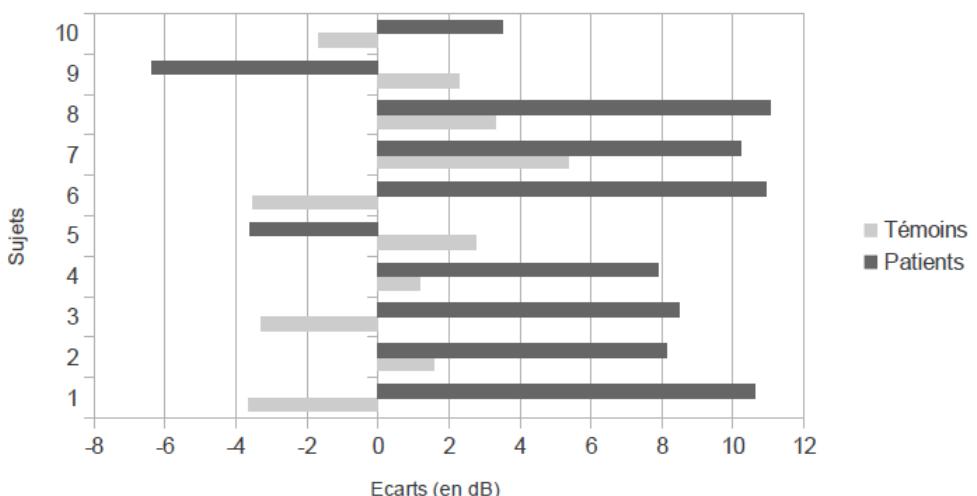
Nous avons donc choisi le calcul de la différence entre les intensités moyennes relevées à l'isolé ( $I_{V1}$ ) et les moyennes des intensités des positions 2 de chaque contexte occlusif ( $I_{V2}$ ) :  $\Delta = I_{V1} - I_{V2}$ . Les tableaux 5.24 et les graphiques 5.24, 5.25, 5.26 présentent nos résultats pour chacun des sujets et pour chacune des trois voyelles :

	[a]	[i]	[u]		[a]	[i]	[u]
Témoins	-5.82	-3.65	-7.07	Patients	5.38	10.6	8.76
	-5.43	1.59	-3.89		9.47	8.13	12.37
	5.24	-3.29	-2.82		6.05	8.49	5.21
	-1.86	1.17	1.09		3.68	7.87	8.05
	-1.83	2.74	-0.95		4.48	-3.59	-1.62
	-6.01	-3.54	-3.47		4.13	10.94	9
	-1.29	5.35	-0.79		10	10.21	8.33
	1.31	3.31	4.99		7.52	11.03	12.64
	-1.65	2.27	-2.04		3.58	-6.36	-3.39
	-0.62	-1.65	-2.02		3.81	3.5	1.19

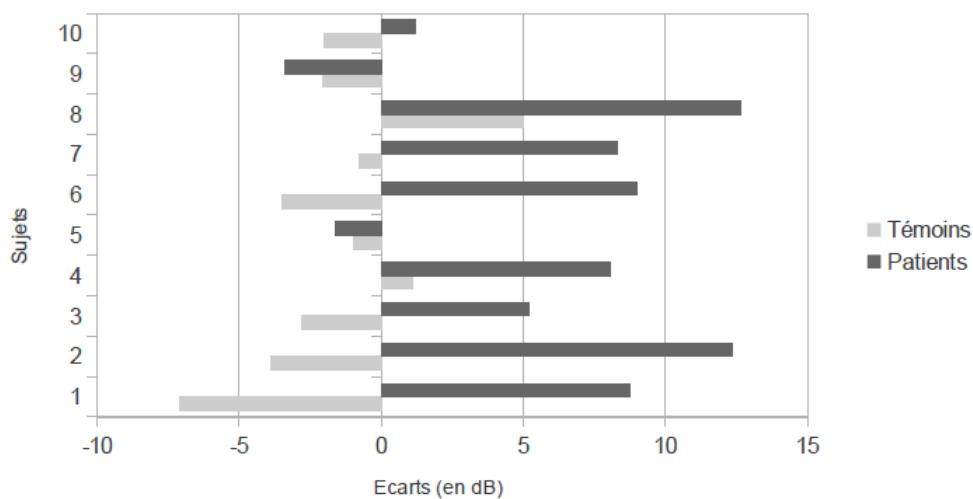
**Table 5.24** – Tableaux récapitulatifs des différences  $I_{V1} - I_{V2}$  (en dB) des voyelles [a], [i], [u] des témoins et des patients (où  $I_{V1}$  représente l'intensité moyenne sur les voyelles isolées et  $I_{V2}$  représente l'intensité moyenne des voyelles en contexte) ( $n = 120$ ).



**Figure 5.24** – Différences intra-sujets  $I_{V1} - I_{V2}$  (en dB) pour la voyelle [a] ( $n = 40$ ).



**Figure 5.25** – Différences intra-sujets  $I_{V1} - I_{V2}$  (en dB) pour la voyelle [i] ( $n = 40$ ).



**Figure 5.26** – Différences intra-sujets  $I_{V1} - I_{V2}$  (en dB) pour la voyelle [u] ( $n = 40$ ).

Nous remarquons que les écarts des témoins représentent majoritairement des valeurs négatives plus ou moins importantes pour les trois voyelles. Nous notons, toutefois, quelques exceptions où les différences restent positives. Ainsi, pour la voyelle [a], les témoins 3 et 8 présentent des écarts positifs avec un maximum de 5,24 dB pour le témoin 3 (cf. figure 5.24). Pour la voyelle [i], le profil semble s'inverser légèrement, avec des écarts positifs — qui restent malgré tout assez bas — pour les témoins 2, 4, 5, 7, 8 et 9. Dans ce cas, c'est le témoin 7 qui présente la différence la plus importante (5,35 dB) (cf. figure 5.25). Enfin, pour la voyelle [u], nous notons que les écarts des témoins 4 et 8 restent positifs avec un pic maximal pour le témoin 8 à 4,99 dB (cf. figure 5.26). En conséquence, nous pouvons en déduire que, dans la plupart des cas, les témoins ont tendance à produire des voyelles en contexte plus intenses que leurs voyelles isolées.

Au contraire, les écarts des patients présentent, la plupart du temps, des valeurs positives assez élevées pour les trois voyelles. Nous relevons, néanmoins, quelques exceptions où les différences sont négatives. En ce qui concerne la voyelle [a], nous ne notons aucune exception pour notre groupe de dysphoniques, toutes les différences restent positives (cf. figure 5.24). Pour les voyelles [i] et [u], nous remarquons que seuls les patients 5 et 9 présentent des valeurs négatives, avec respectivement, des différences maximales de -6,36 dB et -3,39 dB (cf. figures 5.25 et 5.26). Ainsi, nous pouvons en conclure que les patients semblent produire des voyelles isolées beaucoup plus intenses que les voyelles en contexte.

### 5.6.1.3 Discussion des résultats

Cet *excusus* représente un complément essentiel à notre étude sur l'intensité des voyelles tenues isolées. En effet, il nous a permis de mettre en évidence des problèmes

méthodologiques quant aux prises de mesures de l'intensité des voyelles et nous apporte surtout des conclusions qui sont en contradiction avec celles exposées dans la section 5.5.2.

Ainsi, nous avons pu montrer que malgré un gain inchangé sur notre carte son entre les différents enregistrements, le maintien du micro-casque à distance uniforme entre tous les sujets reste difficile, ce qui peut représenter un biais d'un ordre de -2 à 3 dB dans nos résultats inter-groupes sur les voyelles tenues isolées.

Il nous a donc semblé préférable de comparer cette étude inter-groupes préliminaire sur l'intensité des voyelles tenues isolées à une étude intra-groupes sur l'intensité des voyelles en contexte occlusif afin de dégager des tendances qui confirmeraient ou infirmeraient nos hypothèses initiales. Dans un premier temps, nous avons voulu établir un parallèle avec l'étude qui a été réalisée en 5.5.2. Nous avons donc effectué une série de tests-t non appariés sur l'ensemble des données issues des productions des trois voyelles en contexte occlusif afin de déterminer si les intensités moyennes en contexte des patients étaient significativement différentes de celles des témoins. Nous attestons qu'il existe bien une intensité intrinsèque propre aux voyelles, à savoir que les voyelles basses sont plus intenses que les voyelles hautes. Néanmoins, nos résultats montrent qu'il n'existe pas de différence significative de l'intensité entre les productions en contexte des deux populations. Ces résultats sont en opposition totale avec les conclusions établies en 5.5.2. Ainsi, ces contradictions montrent qu'une étude sur des voyelles stables reste insuffisante et doit être associée à une étude sur des voyelles en contexte pour valider nos observations. Dans notre cas, cette seconde expérimentation a démontré que les productions vocaliques des deux groupes s'équilibrent en terme d'intensité en situation de parole. Nous ne pouvons donc pas confirmer le fait que les patients présentent des intensités intrinsèques plus élevées par rapport à celles des témoins. Les deux groupes semblent adopter des intensités équivalentes — voire légèrement plus faibles pour les patients — en situation de parole, ce qui semble davantage aller dans le sens des conclusions de McGuirt 1994, Sjögren & al. 2008 et Motta & al. 2008, Jotic & al. 2011 et Mirghani 2009.

Dans un second temps, nous avons réalisé une ANOVA qui a révélé que le corpus influençait de façon significative nos résultats. Partant de cette hypothèse, nous avons calculé les écarts individuels d'intensités entre voyelles isolées et voyelles en contexte ( $I_{V1} - I_{V2}$ ) afin de mettre en évidence des stratégies intra-locuteurs. Nous avons pu observer que les différences entre les deux productions étaient majoritairement négatives pour l'ensemble des témoins, ce qui suppose des productions plus intenses en contexte par rapport aux voyelles isolées. Au contraire, nous avons pu relever des écarts majoritairement positifs pour l'ensemble des patients, ce qui suppose des productions plus intenses à l'isolé par rapport aux voyelles des logatomes. Nous avons donc deux stratégies totalement différentes qui émergent au sein de nos deux groupes et qui apparaissent comme significatives pour l'ensemble des sujets. Ainsi, nous pouvons établir deux hypothèses quant à ces conclusions : (i) soit les témoins se limitent pour la réalisation de voyelles tenues à hauteur et intensité confortables et produisent des segments « sous-évalués », (ii) soit les patients compensent leur handicap en produisant des voyelles tenues trop « poussées » ou cherchent simplement à réaliser

correctement l'exercice, ce qui correspond à des segments « sur-évalués »<sup>7</sup>. Dans sa thèse sur différents types de laryngectomies partielles, Hans (2003) avait déjà mis en évidence la tendance des patients à produire des voyelles à l'isolé plus fortes que celles des témoins, ce qui pouvait être interprété en terme de forçage vocal (comportement supraglottique).

Quoi qu'il en soit, nous ne pouvons pas confirmer le fait que les patients présentent des intensités intrinsèques plus élevées que celles des témoins. Les intensités des voyelles relevées en contexte montrent qu'il n'y a pas de différence entre les deux populations. Les écarts mis en lumière dans notre étude 5.5.2 sur les voyelles tenues et isolées résultent uniquement de stratégies individuelles, et en particulier d'une intensité plus poussée pour chacun des patients dans cette tâche. Les patients tendent à produire des segments vocaliques plus forts à l'isolé, ce que nous interprétons comme un effort supplémentaire afin de bien faire l'exercice ou de surpasser le handicap vocal qu'ils pensent perdurer. Les différences d'intensité mises en évidence entre voyelles en contexte et voyelles isolées résultent donc davantage de stratégies individuelles liées aux corpus plutôt qu'à un défaut méthodologique.

#### Synthèse de l'étude n°4 : l'intensité intrinsèque

- Objectif de l'étude : écarter les problèmes liés au recueil de données pour les mesures de l'intensité sonore et écarter un effet du corpus en comparant les valeurs d'intensité des voyelles tenues aux valeurs d'intensité prises sur les voyelles des positions 2 des logatomes.
- Corpus : (i) voyelles tenues et isolées et (ii) logatomes de type CVCVCVC où C= [p, t, k, b, d, g] et V= [a, i, u].
- Principaux résultats :
  - pas de différence significative de l'intensité entre nos 2 groupes en contexte occlusif,
  - effet important du corpus pour les mesures de l'intensité,
  - voyelles isolées des patients plus intenses que les voyelles des logatomes,
  - voyelles isolées des témoins moins intenses que les voyelles des logatomes,
- Conclusions :
  - pas de différence entre les 2 groupes pour l'intensité : les patients conservent bien une puissance sonore après la chirurgie,
  - la production des voyelles isolées résulte de stratégies individuelles : importante variabilité,
  - conclusions en adéquation avec les études antérieures qui montrent que les patients tendent à produire des voyelles isolées plus intenses pour compenser leur handicap vocal.

#### 5.6.2 Etude n°5 : la $F_0$ intrinsèque

Dans l'étude que nous avions présentée en 5.3, nous avions mis en évidence que les  $F_0$  intrinsèques des voyelles tenues et isolées des patients étaient significativement

7. ce cas de figure est souvent présent dans les études sur les voyelles tenues à l'isolé où les locuteurs ont tendance à « sur-produire » les segments en terme d'intensité.

plus élevées que celles des témoins. Contrairement à notre *excursus* concernant les problèmes liés au recueil de données pour l'évaluation de l'intensité sonore, nous ne rencontrons pas de difficultés méthodologiques pour les recherches sur la  $F_0$ . Cet *excursus*, développé surtout dans un souci de cohérence, cherche essentiellement à confirmer ou infirmer nos conclusions préliminaires dans le domaine du mot (logatomes).

### 5.6.2.1 Méthodologie

**Les voyelles isolées.** le recueil des données et les prises de mesures restent identiques à la méthodologie décrite dans la section 5.5.1.

**La parole en contexte ou logatomes.** Le corpus et le matériel d'enregistrement restent identiques par rapport à la méthodologie décrite précédemment. Les logatomes de type CVCVCVC ont été coupés en trois syllabes distinctes  $CV_1$ ,  $CV_2$  et  $CVC_3$ , que nous commenterons dans le chapitre 6, section 6.6.

A l'instar de l'*excursus* sur l'intensité sonore, nous avons donc choisi de nous concentrer sur les logatomes composés d'occlusives, soit 360 occurrences (18 logatomes de type CVCVCVC où C= [p, t, k, b, d, g] et V= [a, i, u] x 1 répétition x 20 sujets).

Nous avons délibérément choisi de mêler l'ensemble des logatomes composés des différentes occlusives pour nos traitements statistiques en écartant volontairement les critères de voisement et de lieu d'articulation. Les prises de mesures des  $F_0$  moyennes sur chacun des logatomes ont été extraites à l'aide d'un script Praat, modifié spécialement pour cette étude<sup>8</sup>. Comme précédemment, les relevés de valeurs ont été effectués sur les voyelles des positions 2 de nos logatomes. En effet, cette position nous assure une stabilité dans le maintien du fondamental des voyelles. Finalement, nous avons calculé une moyenne des  $F_0$  des voyelles des position 2 des logatomes. Nous traitons donc 60 items, soit une valeur moyenne des  $F_0$  en contexte x 3 voyelles x 20 sujets.

Nous sommes consciente de la présence de variations co-intrinsèques possibles de la  $F_0$  des voyelles selon les contextes consonantiques, nous optons pour un traitement plus global de nos données. Une fois encore, nous avons réalisé des tests-t appariés préliminaires pour nos deux populations afin de comparer les mesures de  $F_0$  de nos voyelles en contexte voisé/non voisé et selon le lieu d'articulation. Nous avons pu mettre en évidence que les mesures de  $F_0$  des voyelles en contexte tendent à être plus importantes en contexte non voisé. Au contraire, nous n'avons pas retrouvé de tendance générale quant à l'influence du lieu d'articulation sur le fondamental des voyelles. Ces résultats ne sont pas exposés ici car ils sont en dehors du champ de notre recherche. En effet, notre propos ne concerne pas l'influence des caractéristiques consonantiques sur la  $F_0$  des voyelles adjacentes.

Cette étude s'articule autour de deux modes de traitement. Dans un premier temps, nous avons voulu faire un parallèle avec les mesures des  $F_0$  intrinsèques des voyelles

8. Le script est une version plus basique du script « analyse tout » de C. Gendrot.

tenues et isolées ( $n = 60$ ). Ainsi, nous avons réalisé des tests-t non appariés préliminaires entre nos deux populations afin de vérifier l'existence d'une différence quant aux valeurs de  $F_0$  des voyelles en contexte occlusif ( $n = 60$ ). Dans un second temps, nous avons réalisé une ANOVA à 3 facteurs — où la variable dépendante était la  $F_0$  et les variables indépendantes représentaient le groupe, le type de voyelle et le corpus — afin d'évaluer, en particulier, un effet significatif de la tâche de production. Tous les tests ont été effectués à l'aide du logiciel Statview, les résultats étaient considérés comme significatifs à partir d'un seuil alpha de 5%, soit  $p < .05$ . Dans la mesure où ces résultats globaux démontreraient une significativité, nous calculerons la différence  $F_{0V1} - F_{0V2}$  (où  $F_{0V1}$  représente la mesure de  $F_0$  moyenne de la voyelle isolée et  $F_{0V2}$  représente la mesure de  $F_0$  moyenne de la voyelle en contexte) qui nous permettra d'observer les différences de production de chacun des sujets dans les deux tâches.

### 5.6.2.2 Résultats

**Différences de  $F_0$  en contexte occlusif entre nos deux populations.** Afin de déterminer si les  $F_0$  moyennes en contexte des patients sont significativement différentes de celles des témoins, nous avons conduit des tests-t non appariés sur les données issues des productions de nos trois voyelles en contexte occlusif. Le tableau 5.25 met en évidence les résultats de ces tests :

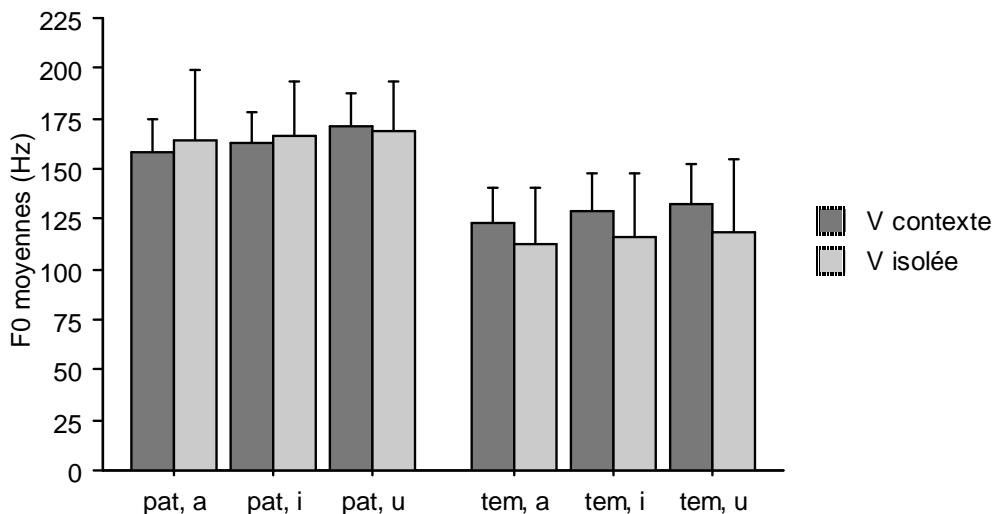
	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	122.75	158.45	129.48	163.4	131.9	171
$s$	17.53	16.5	17.78	14.70	20.42	16.59
$t_{(18)}$		10.35		9.73		10.14
$p$		<.0001		<.0001		<.0001

**Table 5.25** – Résultats synthétiques des tests-t non appariés conduits sur les valeurs de la  $F_0$  (en Hz) des voyelles [a], [i] et [u] en contexte ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Nous remarquons que les  $F_0$  moyennes des voyelles en contexte occlusif des patients sont toujours significativement plus élevées que celles des témoins. De plus, il est important de noter que l'ordre précédemment mis en évidence concernant les  $F_0$  spécifiques des voyelles semble conservé pour les deux groupes puisque la  $F_0$  de [a] est plus basse que celle de [i] et de [u], soit :  $F_0[a] < F_0[i] < F_0[u]$ .

**Résultats de l'ANOVA.** La figure 5.27 présente les valeurs moyennes de la  $F_0$  en fonction de nos deux populations, des contextes vocaliques et corpus (voyelles isolées vs. voyelles en contexte occlusif). Nous remarquons que seule la variable « sujet »

montre un effet significatif sur nos résultats ( $F(1, 108) = 92,987, p < .0001$ ). Cette tendance est prévisible dans le sens où les  $F_0$  des voyelles des patients sont significativement plus élevées que celles des témoins quels que soient les contextes de production. Aussi, l'ANOVA ne montre pas d'effet significatif lié aux contextes vocaliques ( $F(2, 108) = 1,087, p = .3410$ ). Enfin, et contrairement à l'*excursus* sur l'intensité sonore, notre test ne montre aucun effet de la tâche de production sur les valeurs de la  $F_0$  de nos sujets ( $F(1, 108) = 1,287, p = .2592$ ). En conséquence, cette étude confirme totalement nos conclusions sur la  $F_0$  des voyelles isolées. Les patients présentent des  $F_0$  significativement plus importantes que celles des témoins dans les deux tâches. Par conséquent, il ne nous paraît pas justifié d'établir des comparaisons intra-individuelles.



**Figure 5.27** – Valeurs moyennes de  $F_0$  (en Hz) en fonction des sujets, des voyelles et des corpus (les barres d'erreurs représentent les écarts-types) ( $n = 120$ ).

### 5.6.2.3 Discussion des résultats

Ce second *excursus* représente un complément à notre étude sur la  $F_0$  des voyelles tenues isolées. Nous avions pu mettre en évidence que les  $F_0$  intrinsèques de ces voyelles étaient statistiquement plus élevées pour notre population de patients par rapport à nos témoins (cf. section 5.3.2). Contrairement au premier *excursus* détaillé précédemment et lié aux problèmes de recueil de données pour l'évaluation de l'intensité sonore, la présente recherche a uniquement été réalisée dans un souci de cohérence méthodologique. Rappelons que nous cherchons uniquement à comparer les valeurs de  $F_0$  intrinsèques des voyelles tenues isolées aux valeurs de  $F_0$  relevées dans les logatomes en contexte occlusif afin de confirmer ou infirmer nos conclusions établies en 5.3.2.

Dans un premier temps, nous avons voulu établir un parallèle avec notre étude réalisée en 5.3.2. Nous avons réalisé une série de tests-t non appariés sur l'ensemble des données issues des productions des trois voyelles en contexte occlusif afin de déterminer si les  $F_0$  moyennes des patients étaient significativement différentes de celles des témoins. La significativité des résultats montre une différence en production entre nos deux groupes. Les  $F_0$  spécifiques des voyelles en contexte des patients sont toujours statistiquement plus élevées que celles des témoins. De plus, nos résultats suggèrent qu'il existe bel et bien une hiérarchie pour la  $F_0$  intrinsèque de chacune des voyelles analysées. Pour l'ensemble de nos sujets, nous constatons que la voyelle [a] est produite à une fréquence plus basse que les voyelles [i] et [u].

Dans un second temps, nous avons réalisé une ANOVA afin de montrer l'effet du corpus, en particulier, sur nos résultats de  $F_0$ . Contrairement à l'intensité, ce test montre qu'il n'existe aucun effet significatif lié à la tâche de production. En conséquence, nous pouvons en conclure que, intrinsèquement, les patients présentent une  $F_0$  plus élevée que la cohorte de témoins. Dans ce sens, nous avons jugé qu'il n'était pas intéressant d'observer les stratégies individuelles de nos sujets. Nos résultats confirment totalement nos observations préliminaires présentées en 5.3.2. En effet, nous relevons des  $F_0$  spécifiques plus élevées chez les patients, aussi bien pour les voyelles tenues et isolées que pour les voyelles en contexte, par rapport aux témoins. La cordectomie laser est directement liée à cette élévation du fondamental, du fait des modifications anatomico-physiologiques qu'elle implique.

#### Synthèse de l'étude n°5 : la $F_0$ intrinsèque

- Objectif de l'étude : écarter un effet de la tâche de production en comparant les valeurs de  $F_0$  des voyelles tenues aux valeurs de  $F_0$  prises sur les voyelles des positions 2 des logatomes.
- Corpus : (i) voyelles tenues et isolées et (ii) logatomes de type CVCVCVC où C= [p, t, k, b, d, g] et V= [a, i, u].
- Principaux résultats :
  - différence significative des  $F_0$  des voyelles entre nos 2 groupes en contexte occlusif,
  - pas d'effet du corpus pour les mesures de  $F_0$ ,
- Conclusions :
  - la  $F_0$  des patients est toujours plus élevée que celle des témoins et ce quels que soient les contextes,
  - cette élévation est directement liée aux changements anatomico-physiologiques liés à la chirurgie.

## 5.7 Conclusion

Nous avons pu mettre en évidence que la modification de la structure interne du pli vocal touché est responsable d'une asymétrie et d'une diminution de la masse qui ont pour conséquence des perturbations en termes de fréquence et d'amplitude. Dans ce chapitre nous nous sommes intéressée aux résultats acoustiques — principalement à la

fréquence fondamentale et à l'intensité sonore, associées à leurs facteurs de perturbation — liés aux modifications anatomiques après cordectomie laser de type II-III.

Si l'on se réfère à nos observations quant au jitter et au shimmer, nous pouvons conclure que ce type de chirurgie ne présente qu'un impact mineur sur la qualité vocale. En effet, nos expérimentations ont montré qu'il n'existe pas de différence statistiquement significative entre nos cohortes de témoins et de patients pour l'analyse de ces deux paramètres. Ces conclusions, bien que pertinentes dans la mesure où elles valident l'intérêt de ces interventions chirurgicales du point de vue de la conservation vocale, sont à prendre avec précaution. La grande variabilité intra-sujets peut constituer un contre-argument aux résultats que nous avons présentés.

L'un des aspects sans doute le plus pertinent de ce chapitre concerne les études fines que nous avons pu entreprendre concernant la fréquence fondamentale. Nous avons ainsi pu démontrer que quels que soient les supports phonétiques utilisés — voyelles tenues ou logatomes —, les  $F_0$  des patients étaient systématiquement plus élevées que celles des témoins. Aussi, les difficultés majeures de nos sujets dysphoniques se situent principalement dans les moyennes fréquences (voir toutes les fréquences pour la voyelle [u]), ce qui pourraient éventuellement poser problème pour les modulations intonatives. Ces résultats, en adéquation avec les précédentes études cliniques, découlent d'une recherche originale qui tente d'apporter une explication physiologique à ce phénomène. La cordectomie de type II emporte la totalité de la *lamina propria* et la cordectomie de type III s'étend jusqu'au muscle vocal, ce qui tend — dans les deux cas — à affiner la surface du pli vocal touché. Ainsi, quelle que soit la quantité de tissu reséqué, nous observons une diminution de la masse vibrante, une tension et une raideur plus importantes de ce même pli vocal en postopératoire. Cette nouvelle configuration glottique est responsable d'un changement dans le geste vocal, et plus particulièrement dans l'ondulation muqueuse. En effet, il est démontré que l'augmentation de raideur des plis vocaux, la diminution de la masse vibrante ou la diminution de la longueur vibrante sont des paramètres qui jouent en faveur d'une augmentation de la fréquence des cycles glottiques, et donc d'une élévation de la  $F_0$  (Titze 1994, Henrich 2001 entre autres). Il est évident que cette hypothèse devra faire l'objet d'investigations stroboscopiques supplémentaires afin d'être validée.

Un second versant de ce chapitre concerne l'étude de l'intensité sonore. Il s'agit d'un sujet qui reste délicat de part les difficultés méthodologiques qu'il implique. En effet, après une première expérimentation sur les voyelles tenues qui s'est conclue par une différence significative entre nos deux groupes, il nous est très rapidement apparu nécessaire de confronter ces résultats à des données issues des logatomes afin de valider ces observations préliminaires. Cette seconde étude n'a montré aucune différence entre les groupes et a surtout contribué à infirmer totalement nos premières conclusions. Elle nous a également permis de mettre en exergue quelques hypothèses quant aux prises de données et surtout quant aux stratégies individuelles : (i) un micro-casque ne garantit pas une prise de données correcte de l'intensité sonore, nous ne pouvons établir des conclusions sans en vérifier l'exactitude, (ii) les intensités spécifiques des patients ne diffèrent pas spécialement de celles des témoins, les différences sont davantage d'ordre individuelles et (iii) les patients ont tendance à « sur-produire » des

segments à l'isolé alors que les témoins semblent plus réservés dans la réalisation de cette tâche.

Enfin, il nous semble intéressant de souligner le fait que pour toutes les présentes études — et donc pour toutes les tâches réalisées —, les patients présentent une hiérarchie dans la production des voyelles. Bien qu'altéré par rapport aux témoins, le [a] est souvent mieux réalisé que les voyelles hautes [i] et [u] sur le plan acoustique. Ces voyelles hautes et fermées semblent donc être plus sensibles à la chirurgie. La cordectomie mutile uniquement le pli vocal, elle n'a donc aucun effet sur le conduit supralaryngé. Les organes de résonance restent intacts et seules les vibrations peuvent être perturbées. Ainsi, les hypothèses concernant ce phénomène peuvent être liées éventuellement à l'articulatoire ou à l'aérodynamique, mais en aucun cas à un défaut de configuration du conduit vocal. Nous avons mis en évidence que les configurations articulatoires plus complexes pour les voyelles hautes pouvaient expliquer cette différence. En effet, les plis vocaux, et plus largement le conduit vocal, sont davantage tendus pour la production de ces voyelles, ce qui implique des résistances plus importantes. La voyelle [a] semble donc être la seule voyelle qui se rapproche des productions des témoins, ce qui suggère qu'elle reste la plus stable.

Nous avons donc souligné que la cordectomie laser garantit des résultats vocaux satisfaisants mais que cette dernière n'est pas totalement non invasive. En effet, deux sons — les voyelles hautes principalement — restent altérés en postopératoire.

Aussi, la résection de tissus est responsable d'une asymétrie vibratoire et d'un accolement partiel des plis vocaux, ce qui pourrait poser problème dans l'établissement du voisement. Dans le prochain chapitre, nous aborderons cette question des points de vue acoustique, aérodynamique, électroglottographique et perceptif.

#### En résumé...

- Ce chapitre montre l'élévation de la  $F_0$  des patients par rapport aux témoins.
- Dans une tâche d'imitation de stimuli à différentes  $F_0$ , nous avons mis en évidence des difficultés dans les fréquences entre 90 et 150 Hz pour la voyelle [a], entre 90 et 170 Hz pour la voyelle [i] et entre 90 et 230 Hz pour la voyelle [u] des patients.
- L'élévation de la  $F_0$  est directement liée aux changements de la structure des plis vocaux après cordectomie de type II-III.
- L'intensité sonore n'est pas altérée après ces chirurgies, malgré une variabilité importante en production : les patients ont tendance à produire des segments vocaliques plus forts à l'isolé par rapport aux voyelles des logatomes.
- Finalement, la chirurgie garantit des résultats vocaux assez satisfaisants, malgré une altération plus importante des paramètres acoustiques des voyelles hautes.



# Quatre études relatives au voisement

"Il importe peu quand la voix parle haut, quelle langue elle parle."  
– Victor Hugo (*Hernani*)

## Résumé

Ce chapitre est consacré au maintien du trait de voisement des occlusives après cordectomie de type II-III. Celui-ci se compose de quatre études : aerodynamique (étude 6), acoustique (étude 7), EGG (étude 8) et perceptive (étude 9). Dans une première étude aérodynamique, nous mesurons la présence de fuites glottiques sur les voyelles tenues [a, i, u] à hauteur et intensité confortables ( $n = 60$ ). Cet indice se calcule à partir des valeurs du débit d'air buccal (DAB) et de l'intensité sur une portion stable des voyelles. Le rapport DAB/intensité, exprimé en  $cm^3/dB/s$ , représente l'indice de fuites glottiques (Giovanni & al. 1995). Nous avons réalisé une série de tests-t non appariés sur les valeurs du DAB issues des productions des trois voyelles et sur les indices de fuites glottiques (Fg) dérivés. Les résultats de ces analyses montrent que :

- pour les témoins, la hiérarchie est la suivante :  $Fg[a] < Fg[i] < Fg[u]$ ,
- pour le groupe des patients, la hiérarchie est différente :  $Fg[i] < Fg[a] < Fg[u]$ ,
- les valeurs de DAB et les indices de fuites glottiques des patients sont toujours significativement plus élevés que ceux des témoins.

En outre, nous avons entrepris une analyse quantitative des débits d'air et des pressions nécessaires et suffisantes pour la production des occlusives labiales des logatomes CVCVCVC (où C= [p, b] et V=[a, i, u]) pour les positions 1 et 2 ( $n = 240$ ). Nous avons réalisé une série d'ANOVA à 4 facteurs où les variables dépendantes étaient les pics de PIO des consonnes labiales, le pic maximal de DAB (Umax) et le débit phonatoire moyen (DPM) et où les variables indépendantes représentaient le groupe, le type de voyelle, le contexte de voisement et la position dans les logatomes. Nos résultats montrent :

- une élévation de l'ensemble des paramètres chez les patients,
- pour les deux groupes, une élévation des paramètres aérodynamiques en contexte non voisé,
- la PIO est plus élevée en contexte de voyelles fermées.

Pour finir, nous avons calculé l'estimation de la pression transglottique à partir des différences de mesures de la PIO des consonnes, soit  $\text{PIO}(/p-b/)$ . Nous notons :

- une augmentation de cette pression en contexte de voyelles fermées chez les deux groupes,
- une pression transglottique estimée plus élevée chez les patients.

Dans une étude acoustique (étude 7), nous étudions les paramètres suivants : la durée totale des occlusives orales, l'intensité relative du relâchement, la durée des voyelles adjacentes, la  $F_0$  moyenne au début de ces segments vocaliques et le Voice Onset Time (VOT) des positions 1 et 2 des logatomes de type CVCVCVC (où C= [p, t, k, b, d, g] et V= [a, i, u]) ( $n = 342$  pour les mesures de durées des consonnes,  $n = 684$  pour les mesures de durées vocaliques et VOT et  $n = 720$  pour les autres mesures). Nous avons donc conduit une série d'ANOVA à 4 facteurs où les variables indépendantes représentaient le groupe, le type de voyelle, le contexte de voisement et la position dans les logatomes. Nos résultats montrent :

- les durées segmentales et la  $F_0$  au début des voyelles sont significativement plus élevées pour les patients,
- pour nos deux groupes, une élévation de l'ensemble des paramètres acoustiques (durées consonantiques, intensité des relâchements,  $F_0$  au début des voyelles et VOT) en contexte non voisé, excepté pour la durée vocalique.

Dans une étude pilote EGG (étude 8), nous analysons les quotients ouverts (Oq) des voyelles tenues [a, i, u] à hauteur et intensité confortables des patients et les comparons avec les valeurs d'un unique témoin de référence. Après avoir écarté les signaux EGG et DEGG atypiques de deux patients, révélateurs de cas de diplophonie ou de triplophonie, nous avons conduit des tests-t univariés pour lesquels les moyennes théoriques correspondaient aux valeurs de notre témoin de référence ( $n = 27$ ). Cette première expérimentation montre que la majorité des patients présentent des valeurs de Oq supérieures à celles du témoin, et ce pour les trois voyelles. Un second objectif de cette étude consistait à mettre en évidence l'intérêt des mesures de Oq pour l'opposition de voisement des occlusives. Nous nous sommes donc intéressée aux positions 1 et 2 des logatomes de type CVCVCVC (où C= [p, t, k, b, d, g] et V=[a, i, u]) ( $n = 288$ ). Nous avons conduit une ANOVA à 3 facteurs qui nous a permis d'évaluer l'influence du voisement des occlusives sur les voyelles environnantes en fonction des contextes vocaliques, de la position syllabique et du voisement de la consonne. Nos résultats démontrent que les valeurs de Oq au début des voyelles, en particulier des voyelles fermées, sont plus élevées dans un contexte non voisé et ce, pour tous les sujets.

Enfin, la dernière étude perceptive (étude 9), nous a permis d'évaluer l'intelligibilité des consonnes occlusives produites par les patients dans les deux positions syllabiques. Nous avons donc élaboré deux tests perceptifs sous Praat pour chacune des positions que nous avons expérimentés sur 22 auditeurs naïfs ( $n = 7920$ ). Dans les deux positions, l'intelligibilité globale est très élevée (93,59% pour la position initiale et 93,79% pour la position intervocalique). Nous observons deux types de confusions mineures dans la perception des occlusives : des confusions de voisement et de lieu d'articulation. Les confusions de voisement sont plus fréquentes pour les consonnes non voisées perçues comme voisées, principalement pour la position initiale (la labiale [p] est perçue [b] dans 9,55% des cas, la dentale [t] est perçue [d] dans 5% des cas et la vélaire [k] est perçue [g] dans 2,58% des cas). Les confusions de lieu concernent principalement la vélaire voisée [g] qui est confondue avec [b] en contexte de voyelle [u] ou [d] en contexte de voyelle [i].

## 6.1 Introduction

Nos études précédentes ont particulièrement mis en évidence l'altération de la fréquence fondamentale après cordectomie de type II-III. En effet, nous avons pu montrer que ces chirurgies engendraient une diminution de la masse vibrante, une tension et une raideur plus importantes du pli vocal resqué, ce qui tendait à perturber les mécanismes vibratoires.

Le contrôle laryngé intervient au niveau segmental pour la production des voyelles et — en français tout du moins — pour maintenir un contraste distinctif entre consonnes voisées et non voisées. Le mécanisme peut différer selon les langues. Selon Ladefoged (1983) et Stevens (1991), le voisement correspond à un état de détente des structures laryngées, augmenté par un abaissement du larynx qui a pour effet de raccourcir la longueur des plis vocaux, une adduction de ces derniers et une pression transglottique suffisante afin d'entretenir les cycles vibratoires. Or, l'altération de la source vocale en postopératoire, caractérisée par une asymétrie, une désynchronisation vibratoire, un accollement partiel des plis vocaux et une tension laryngée, pourrait contribuer à dégrader le voisement. Ce chapitre, composé de diverses études basées sur des protocoles cliniques classiques (cf. chapitre 4), nous permettra d'explorer ce paramètre et de confirmer ou d'inflimer cette hypothèse.

Dans la première section de ce chapitre, nous tenterons d'apporter une définition du voisement des points de vue phonologique, aérodynamique et articulatoire. Nous nous attacherons principalement à la description des indices acoustiques, primaires et secondaires, utilisés dans les langues pour l'opposition de voisement. La suite de ce chapitre sera consacrée à plusieurs études originales qui nous éclaireront sur le contrôle des structures laryngées permettant le maintien du voisement chez les sujets ayant subi une cordectomie de type II-III, en comparaison à une cohorte de témoins. La première étude, axée sur des paramètres aérodynamiques, nous permettra d'explorer les débits d'air et les pressions nécessaires ou suffisantes pour la production des occlusives voisées et non voisées. A ce titre, nous associerons des descriptions qualitatives des *patterns* de pression intra-orale (PIO) observés avec des données quantitatives afin de caractériser au mieux le mode vibratoire des patients. La seconde étude, basée sur des paramètres acoustiques, mettra en évidence la coordination des gestes articulatoires des patients à travers l'analyse de paramètres tels le VOT, les durées consonantiques et vocaliques, l'intensité relative des explosions des consonnes ou la  $F_0$  au début des voyelles adjacentes. Notre troisième section présentera une étude EGG pilote sur les voyelles tenues des patients. Après avoir tenté de caractériser leur qualité vocale à travers l'étude du quotient ouvert, nous tenterons de montrer que l'EGG peut être considéré comme un outil pertinent pour explorer la différence entre occlusives voisées et non voisées. Enfin, la dernière étude, perceptive, nous permettra de confirmer ou d'inflimer le maintien du contraste de voisement chez les patients.

## 6.2 Définitions du voisement

Dans cette section, nous apporterons une définition du voisement des points de vue phonologique, aérodynamique et articulatoire.

### 6.2.1 Définition phonologique

D'un point de vue phonologique, le voisement est un trait pertinent servant à assurer un contraste minimal entre deux segments, tels que [p] et [b] par exemple. A l'instar des autres traits qui rendent compte des différentes configurations laryngales, celui-ci s'insère dans la classe des traits laryngaux. Le trait de voisement est impliqué dans un certain nombre de processus phonologiques qui seront évoqués par la suite.

#### 6.2.1.1 Les traits laryngaux

Les langues peuvent être décomposées en mots qui eux-mêmes sont segmentables en morphèmes (unités porteuses de sens)<sup>1</sup>. Les morphèmes peuvent également être découpées en phonèmes<sup>2</sup>. Les phonèmes n'ont pas de contenu sémantique à proprement parlé. D'un point de vue phonologique, les phonèmes peuvent être décrits par un ensemble de traits distinctifs souvent considérés comme binaires, c'est-à-dire qui fonctionnent en termes d'oppositions (+ vs -) (Jakobson & al. 1952).

Jakobson (1932) définit le phonème comme étant :

« [...] la somme des propriétés phoniques (traits) simultanés par les-  
quelles un son d'une langue se distingue de ses autres sons en tant que  
moyen pour différencier la signification des mots »

Ainsi, cette définition suppose l'existence d'un ensemble de propriétés cognitives <T> qui composent le phonème telles que : (i) les membres de <T> sont capables de distinguer tous les segments des langues du monde, (ii) les membres de <T> permettent d'établir des classes naturelles et (iii) chaque trait possède sa propre définition phonétique.

Jakobson & al. (1952) avaient mis en évidence 12 traits distinctifs universels, dont les trois traits laryngaux [voiced], [tense] et [checked] servaient, respectivement, à rendre compte du voisement, de l'aspiration et de la glottalisation. Néanmoins, les définitions articulatoires et aérodynamiques utilisées pour le classement des consonnes excluaient, en partie, les liquides [l, r]. En effet, Jakobson & al. (1952 :19) considéraient ces segments comme étant à la fois consonantiques et vocaliques, ce qui réduisait considérablement la puissance et la validité de ce modèle.

Sur la même base, Chomsky & Halle (1968) proposèrent un nouveau modèle composé de 22 traits distinctifs parmi lesquels figuraient, non plus trois, mais quatre traits laryngaux : [heightened subglottal pressure], [voice], [glottal constriction] et [tense].

1. Par exemple, en français standard, le mot « refaire » est composé des deux morphèmes suivants : /ʁə-/ (préfixe) et /feʁ/ (lémème).

2. Les deux morphèmes du mot « refaire » sont composés des cinq phonèmes /ʁ/, /ə/, /f/, /ɛ/ et /ʁ/.

Halle & Stevens (2002 [1971]) affinèrent la classe des traits laryngaux en introduisant les traits [stiff vocal folds], [slack vocal folds], [spread glottis] et [constricted glottis]. A l'heure actuelle, les traits [stiff vf.] et [slack vf.] ont été abandonnés puisque ces derniers impliquent de la redondance. En effet, la spécification positive de l'un suppose la spécification négative de l'autre (les plis vocaux ne peuvent pas être tendus et relâchés à la fois). A ce titre, la plupart des travaux actuels en phonologie utilisent les traits [voice] (ou [voix]<sup>3</sup>) [spread glottis] et [constricted glottis] pour caractériser l'activité laryngale.

**Le trait [voix].** Les segments [+voix] sont produits avec une configuration laryngale qui permet une vibration périodique des plis vocaux (voyelles, nasales, obstruantes voisées), ce qui n'est pas le cas pour les segments [–voix].

Le contraste de voisement implique une activité musculaire du larynx et une coordination entre l'articulateur glottique et les articulateurs supraglottiques. Ainsi, les gestes spécifiques supralaryngaux, qui dépendent de stratégies individuelles, doivent être couplés avec des cycles périodiques d'adduction et d'abduction des plis vocaux, soit un accrolement régulier, complet et actif de ces derniers pour la production des segments [+voix]. Au contraire, la production des segments [–voix] nécessite une ouverture glottique variable selon les segments.

**Le trait [spread glottis].** Les segments [+spread glottis] sont produits avec les plis vocaux écartés, produisant un bruit apériodique de friction (consonnes aspirées, consonnes murmurées, voyelles non voisées). Ce type de segments implique un écartement croissant des plis vocaux pendant la réalisation de la consonne qui atteint souvent un maximum au moment du relâchement. Ainsi, le degré d'ouverture glottique et/ou le timing de l'ouverture sont corrélés au degré d'aspiration. Au contraire, les segments [–spread glottis] ne sont pas réalisés avec une telle configuration. Il est intéressant de noter que le trait [spread glottis], qui est utilisé pour décrire les phénomènes d'aspiration, correspond en réalité à un voisement partiel, c'est-à-dire à une déperdition d'air (*breathiness*).

Selon Lisker & Abramson (1964), l'aspiration se définit comme un retard des vibrations glottiques après le relâchement de la consonne, c'est-à-dire correspondant à un VOT (cf. section 6.3.1 pour une définition) plus long.

De rares cas de langues tels que le hindi, le changana ou le nepali, par exemple, opposent occlusives non voisées aspirées ([p<sup>h</sup>]) et occlusives voisées aspirées ([b<sup>h</sup>]). En ce qui concerne la réalisation des occlusives voisées aspirées en hindi, Dixit (1989) note que l'ouverture glottique reste modérée et inférieure à celle de leurs contreparties non voisées. Le degré d'ouverture semble être plus important à l'initiale absolue par rapport aux postions intervocaliques. De plus, au contraire de leurs homologues non voisées, les auteurs notent que l'ouverture glottique atteint son maximum souvent bien après le relâchement de la consonne. Il est toutefois important de souligner que ce

---

3. Désormais, et dans le reste de ce document, nous utiliserons la traduction française de ce trait (i.e. [voix]) dans nos études concernant le voisement.

genre de configuration n'est pas stable et dépend fortement des langues étudiées, de la position syllabique et des variations intra-locuteurs (pour une revue de la littérature, voir Ridouane 2006).

**Le trait [constricted glottis].** Les segments [+constricted glottis] sont produits avec les plis vocaux étroitement rapprochés, ce qui implique une vibration anormale ou déphasée de ces derniers. Les segments [−constricted glottis] ne présentent pas ce type de configuration laryngale.

Ce trait permet de rendre compte des consonnes glottalisées ou laryngalisées, des implosives (langues asiatiques et africaines) et des éjectives (présentes dans l'inventaire de certaines langues du Caucase, les langues mayas, les langues asiatiques et les langues africaines).

### 6.2.1.2 Typologie des processus phonologiques impliquant le trait de voisement

**Dévoisement final.** Le dévoisement final correspond à la perte du trait [+voix] d'un segment en position finale. Dans une telle configuration, il est impossible de distinguer les deux termes d'une opposition et celle-ci est *neutralisée*<sup>4</sup>. Ainsi, un archiphonème non voisé émerge et provoque une alternance morphophonologique régulière entre allomorphes non voisés et voisés en position finale et intervocalique respectivement. Contrairement au processus inverse, qui consiste en un voisement final, Lombardi (1991) et Cho (1990) notent que le dévoisement final est un processus très répandu à travers les langues. Partant de ce constat typologique, les auteurs soulignent que le trait [−voix] représente la valeur non marquée du trait.

Ce type de processus est actif dans des langues telles que l'allemand (O'Dell & Port 1983, Port & O'Dell 1985, entre autres), le néerlandais (Grijzenhout & Krämer 2000, Warner & al. 2004), le catalan (Charles-Luce 1993) ainsi que dans la majorité des langues slaves (Chen 1970 pour le russe, Jassem & Richter 1989 pour le polonais).

En néerlandais par exemple, le voisement est distinctif parmi les obstruantes. Dans certains environnements, cette distinction est neutralisée. Ainsi, toutes les obstruantes de la langue sont non voisées en position finale de mot. Ce processus est illustré par les exemples sous (1), où l'occlusive coronale /d/ et la fricative coronale /z/ sont réalisées comme des obstruantes non voisées [t] et [s] respectivement en position finale (1a). Néanmoins, lorsque ces obstruantes apparaissent en position intervocalique, comme c'est le cas au pluriel (1b), celles-ci ne subissent aucun changement et sont réalisées [d] et [z] respectivement :

- (1) Dévoisement final en néerlandais (Grijzenhout & Krämer 2000 :56)

4. En français standard, l'opposition entre [o] et [ɔ] se neutralise en fin de mot, c'est-à-dire que seule la voyelle [o] peut apparaître en fin de mot.

	<i>Singulier</i>		<i>Pluriel</i>	
a.	pad → pat	pad-ən → padən	"crapaud(s)"	
	pu:z → pu:s	pu:z-ən → pu:zən	"chat(s)"	
b.	lat → lat	lat-ən → latən	"latte(s)"	
	vos → vɔ:s	vɔ:s-ən → vɔ:sən	"renard(s)"	

**Assimilation de voisement.** L'assimilation de voisement peut se définir comme la propagation du trait de voisement d'un segment à un segment adjacent ou non. Par conséquent, un segment spécifié [–voix] peut devenir [+voix] sous l'influence d'un segment voisé adjacent de la même manière qu'un segment [+voix] peut devenir [–voix] sous l'influence d'un segment non voisé adjacent. Dans un certain sens, l'assimilation de voisement pourrait être interprétée comme une neutralisation, bien qu'incomplète d'un point de vue phonétique, de la première consonne sous l'influence de la seconde (Lombardi 1999, Wetzels & Mascaro 2001).

L'assimilation peut être strictement locale, c'est-à-dire qu'elle peut concerner deux consonnes strictement adjacentes, ou elle peut être limitée à un domaine particulier tel que la syllabe, le pied ou le mot. Dans ce cas, il s'agit d'une *harmonie* consonantique puisque le trait concerné se propage à travers les segments intervenants. De même qu'il existe une notion de localité dans l'assimilation, il existe également une notion de directionnalité dans l'application de ce processus. Ainsi, certaines assimilations sont progressives, c'est-à-dire qu'elles s'appliquent de gauche à droite (anglais) ; d'autres sont régressives, c'est-à-dire qu'elles s'appliquent de droite à gauche (français) et d'autres s'appliquent enfin de manière bidirectionnelle (suédois).

D'un point de vue typologique, l'assimilation de voisement est un processus très répandu à travers les langues. Celle-ci s'applique par exemple dans des langues telles que le russe, le polonais, le slovaque et le serbe pour les langues slaves, le néerlandais, l'anglais et le yiddish pour les langues germaniques ou encore le latin, le roumain et le français pour les langues romanes (pour une revue la littérature voir Wetzels & Mascaro 2001).

En français moderne par exemple (Rigault 1967, Dell 1995), deux obstruantes adjacentes séparées par une frontière de mot ( $C_1 \# C_2$ ) doivent s'accorder sur le trait de voisement. Il s'agit d'une assimilation régressive symétrique déclenchée par la spécification du trait [ $\pm$ voix] de la seconde obstruante ( $C_2$ ). Ainsi, lorsque la première consonne ( $C_1$ ) est une obstruante voisée et que la seconde est une obstruante non voisée, la première consonne va se dévoiser sous l'influence du trait [–voix] de la seconde (2a). De la même manière, lorsque la première consonne ( $C_1$ ) est une obstruante non voisée et que la seconde est une obstruante voisée, la première consonne va se voisier sous l'influence du trait [+voix] de la seconde (2b) :

- (2) Assimilation régressive de voisement en français
- a. buʒ pa → buʃ pa "bouge pas"
  - kʁab kʁi → kʁap kʁi "crabe cuit"
  - b. buʃ be → buʒ be "bouche bée"
  - ʒyʁ gʁiz → ʒub gʁiz "jupe grise"

En anglais, la formation du pluriel implique une assimilation de voisement progressive. En effet, la suffixation du morphème pluriel /-z/ à une base nominale entraîne une allomorphie gouvernée par le dernier segment de la racine. Ainsi, lorsque la base nominale se termine par une obstruante voisée (3a) ou une voyelle (3b), le suffixe pluriel se réalise tel quel dans les représentations de surface. Par contre, lorsque la base nominale se termine par une obstruante non voisée (3c), le suffixe pluriel se réalise comme une fricative coronale non voisée [-s] :

(3) Assimilation progressive de voisement en anglais

	<i>Singulier</i>	<i>Pluriel</i>	
a.	gləʊb	gləʊb- <u>z</u>	"globe(s)"
	raɪd	raɪd- <u>z</u>	"promenade(s)"
	dɒg	dɒg- <u>z</u>	"chien(s)"
	weɪv	weɪv- <u>z</u>	"onde(s)"
	laɪən	laɪən- <u>z</u>	"lion(s)"
b.	bɪ:	bɪ:- <u>z</u>	"abeille(s)"
	flaɪ	flaɪ- <u>z</u>	"mouche(s)"
	aɪdiə	aɪdiə- <u>z</u>	"idée(s)"
c.	lɪp	lɪp- <u>s</u>	"lèvre(s)"
	kæt	kæt- <u>s</u>	"chat(s)"
	bʊk	bʊk- <u>s</u>	"livre(s)"
	stɑ:f	stɑ:f- <u>s</u>	"membre(s)"

L'exemple de l'anglais montre que le trait [–voix] de la dernière obstruante de la base nominale se propage sur le suffixe pluriel, ce qui crée une allomorphie [z-] ~ [s-].

Le dernier exemple que nous citerons ici est extrait du kera (Pierce 2005), une langue Tchadique parlée au Tchad. Dans cette langue, le préfixe du pluriel /k-/ se réalise [k-] lorsque la première consonne de la syllabe suivante est une occlusive non voisée(4a) et [g-] lorsque la première consonne de la syllabe suivante est une occlusive voisée (4b) :

## (4) Harmonie de voisement en kera (Pierce 2005 :347)

	<i>Singulier</i>	<i>Pluriel</i>	
a.	kúmná	<u>k</u> í-kámñí	"chef"
	tāatá	<u>k</u> í-tāatá-w	"gros pot"
	táasā	<u>k</u> ō-táasā-w	"coupe"
b.	bìrwá	<u>g</u> ì-bìrwá-ŋ	"blanc"
	dáarà	<u>g</u> ò-dàarà	"ami"
	àzrá	<u>g</u> -àzrá-w	"gazelle"

Contrairement aux exemples précédents, où une consonne assimilait le trait de voisement d'une consonne strictement adjacente, l'exemple du kera montre un bel exemple d'harmonie de voisement où une consonne d'un préfixe assimile le trait de voisement d'une consonne située dans la syllabe suivante. Selon Pierce (2005), l'harmonie de voisement apparente est en réalité un effet de la propagation du ton bas de la voyelle de la racine sur la consonne du préfixe. Elle note en effet qu'il existe une corrélation entre la  $F_0$  et le VOT : plus un ton est haut, c'est-à-dire plus la  $F_0$  est élevée, et plus le VOT est élevé (cf. section 6.3.1 pour une définition). Par conséquent, Pierce (2005) conclue que le VOT dans cette langue à pour principale fonction de renforcer les indices tonals.

**Dissimilation.** La dissimilation est un processus à travers lequel deux segments similaires dans un domaine particulier (syllabe, pied, mot) sont différenciés d'une manière ou d'une autre. Généralement, la dissimilation est réalisée à travers le changement de la valeur d'un trait. La dissimilation du trait de voisement est un processus assez rare à travers les langues. Dans ce type de dissimilation, deux obstruantes plus ou moins éloignées et qui partagent la même spécification du trait [voix] peuvent se différencier par l'effacement de la valeur du trait de l'une ou l'autre.

Par exemple, dans un dialecte du bade (Schuh 1997), une langue Tchadique (Afro-Asiatique) parlée au Nigéria, les obstruantes voisées — exceptées les sonantes — se dévoisent lorsque la syllabe suivante commence par une obstruante voisée, excepté lorsque celle-ci est une sonante ou une implosive. Par exemple, le clitique sujet de la seconde personne du masculin /gə-/ se réalise respectivement [gə-] lorsque la première consonne du verbe est une obstruante non voisée (5a) et [kə-] lorsque la première consonne du verbe est une obstruante voisée (5b) :

## (5) Dissimilation de voisement en Bade (Schuh 1997 :7)

a.	gò-kórū	→ <u>g</u> ò-kérū	"tu as volé"
	gò-dâbdú	→ <u>g</u> ó-dâbdú	"tu as vendu"
	gò-lágú	→ <u>g</u> ó-lágú	"tu as arrêté"
	gò-ná:jú	→ <u>g</u> ó-ná:jú	"tu as grandi"
b.	gò-vérú	→ <u>k</u> ó-vérú	"tu es sorti"
	gò-dèk <sup>w</sup> ú	→ <u>k</u> ó-dèk <sup>w</sup> ú	"tu as entendu"
	gò-gàfó	→ <u>k</u> ó-gàfó	"tu as pris"
	gò-bà:sú	→ <u>k</u> ó-bà:sú	"tu as parlé"

Ce type de configuration montre un exemple clair de processus de dissimilation de voisement par changement de trait où une séquence [+voix]...[+voix] illicite est réalisée [−voix]...[+voix] dans les représentations de surface.

Le processus inverse se retrouve en kinyarwanda (Shimamungu 1998), une langue Bantu parlée au Rwanda. Dans cette langue, les occlusives non voisées des classes nominales se voisent devant une racine consonantique à initiale non voisée (loi de Dahl). Ainsi, le préfixe infinitif /ku-/ (classe 5) se réalise tel quel devant les racines qui débutent par une obstruante voisée (6a) mais se réalise [gu-] devant les racines verbales qui débutent par une obstruante non voisée (6b) :

- (6) Dissimilation de voisement en Kinyarwanda (Shimamungu 1998 :90)
- a. ku-vúg-a → ku-vúga "parler"
  - ku-geend-a → ku-geenda "marcher"
  - b. ku-kór-a → gu-kóra "travailler"
  - ku-térur-a → gu-térura "soulever"
  - ku-híng-a → gu-hínga "cultiver"

Contrairement à l'exemple précédent, les données du kinyarwanda montrent une dissimilation de voisement à travers laquelle une séquence [−voix]...[−voix] illicite est réalisée [+voix]...[−voix] dans les représentations de surface.

## 6.2.2 Définition aérodynamique

D'un point de vue aérodynamique, le voisement dépend des pressions sous-glottique (PSG), supraglottique (PO) et donc intra-orale (PIO). Rothenberg (1968) distingue ainsi trois types de pressions qui interagissent dans la production de la parole : la pression sous-glottique (PSG), qui correspond à la pression exercée par le flux pulmonaire sous les plis vocaux ; la pression supraglottique ou intraorale (PIO), qui correspond à la pression au-dessus des plis vocaux et la pression atmosphérique (PA), qui représente la pression extérieure.

Entre la PSG et la PIO, la résistance glottique (RG) contrôle le débit d'air. De même, entre la PIO et la PA, la résistance articulatoire (RA) au niveau de l'occlusion contrôle le débit d'air.

Pour la production des occlusives non voisées, la glotte est ouverte et ne fait donc pas obstacle au passage de l'air. En conséquence, un équilibre entre la PSG et la PIO — inférieur à la PA — s'établit après le point de constriction dans le conduit vocal et ce jusqu'au relâchement. Finalement, la fermeture entraîne une augmentation rapide de la PIO qui atteint un plateau et une décroissance progressive du flux pulmonaire. Les parois laryngées et les joues se gonflent légèrement pour lutter contre cette pression. La phase de relâchement, qui correspond à une évacuation brutale et instantanée de l'air, implique une chute de la PIO et une décroissance plus rapide du volume pulmonaire (Stevens 1991).

Pour la production des occlusives voisées, deux gestes antagonistes doivent être associés, ce qui rend ces consonnes difficiles à réaliser. Ohala (1983) met en évidence une incompatibilité certaine entre le voisement et la phase d'occlusion de ces consonnes.

Ainsi, la production d'une occlusive voisée nécessite une vibration des plis vocaux entretenue pendant toute la durée de la consonne. Le larynx oppose une résistance au passage de l'air pulmonaire, ce qui tend à diminuer la PIO par rapport à la PSG. Or, le voisement est principalement dû à un flux d'air transglottique caractérisé par le maintien d'une différence de pression entre la PSG et la PIO. Ce différentiel est préservé principalement par (i) une expansion passive, à savoir un faible gonflement des tissus qui maintient la détente du conduit vocal et (ii) une expansion active caractérisée par un élargissement de la cavité supraglottique qui permet un maintien de la PIO dans le tractus vocal et donc une perdurance de l'appel d'air. Ces contraintes articulatoires ont une conséquence directe sur la durée du voisement des consonnes. Ainsi, les occlusives voisées sont plus brèves que leurs contreparties non voisées. Ohala (1983) ajoute que plus la constriction est antérieure, plus la cavité buccale est large et plus la PIO augmente lentement. En conséquence, la durée de la labiale voisée est supérieure à celle de la dentale voisée, qui elle-même est supérieure à celle de la vélaire voisée (durées d'occlusions voisées relevées par Ohala & Riodan (1979) pour [b, d, g] : 82, 63 et 52 ms respectivement). De plus, l'auteur remarque que la durée de la labiale voisée est allongée lorsqu'elle est adjacente à une voyelle haute. En effet les voyelles hautes sont produites avec un élargissement de la cavité pharyngale. L'anticipation de ce type de voyelle permet un élargissement de la cavité buccale plus rapide en contexte labial et donc un voisement plus long.

A partir de ce modèle aérodynamique, les études de Rothenberg (1982), Holmberg & al. (1988) ou Kitajima & al. (1992), entre autres, ont démontré qu'il était possible de caractériser la fonction vibratoire des plis vocaux à partir d'une mesure indirecte de la pression intra-orale (PIO) lors de la production de logatomes contenant des labiales, soit [ipi] vs [ibi]. En effet, nous savons que les plis vocaux sont écartés pendant la production de l'occlusive labiale non voisée [p]. Ainsi, nous notons un équilibre dans le conduit vocal entre la PSG et la PIO. De ce fait, la PIO augmente très lentement après le point de constriction, ce qui nous permet de conclure que la PSG est à peu près égale à la PIO mesurée dans la cavité buccale pendant la phase d'occlusion. Au relâchement, la PIO chute brutalement tandis que le débit d'air buccal (DAB) augmente progressivement. A l'inverse, nous savons que les plis vocaux sont en adduction pour la production de l'occlusive voisée [b]. La fermeture glottique oppose une résistance au flux pulmonaire, ce qui bloque l'augmentation de la PIO derrière la constriction. En conséquence, la PSG reste toujours supérieure à la PIO. Finalement, la différence des pics de PIO lors des occlusives [p] et [b] à une intensité identique, permet d'estimer assez bien la pression transglottique nécessaire au maintien du voisement.

En ce qui concerne le DAB, Fisher-Jorgensen (1972) a montré qu'il existe une différence significative dans les valeurs du DAB entre occlusives voisées et non voisées. Ainsi, l'auteur rapporte des valeurs du DAB plus élevées pour les non voisées [p, t, k] par rapport à leurs homologues voisées [b, d, g]. Selon elle, cette différence découle du degré d'ouverture de la glotte au moment du relâchement de la consonne. L'auteur souligne également l'existence évidente d'une forte variabilité intra-individuelle.

Enfin, Ohala (1983) précise également que les effets de coarticulation interviennent directement sur le voisement des consonnes. Ainsi, la position des occlusives dans les

mots est un paramètre pertinent dans le voisement ou le dévoisement de ces consonnes. En effet, les locuteurs cherchent la facilité articulatoire, c'est-à-dire qu'ils minimisent les changements de paramètres aérodynamiques et/ou articulatoires dans un souci d'économie. De manière générale, les occlusives simples ont tendance à se voiser en contexte intervocalique car la PIO tend à se maintenir à un niveau constant pendant toute la durée du cluster. Au contraire, les occlusives géminées vont se dévoiser sur leurs parties finales car PSG et PIO auront eu le temps de s'équilibrer. Aussi, le voisement est plus difficile à maintenir à l'initial et en final. Effectivement, le différentiel de pression entre PSG et PIO semble être moins important qu'en position intervocalique, ce qui tend à affaiblir le voisement de ces consonnes dans ces deux positions.

### 6.2.3 Définition articulatoire

Le voisement est défini en fonction de la vibration ou non des plis vocaux. Cette vibration est dépendante de deux facteurs essentiels : (i) un accrolement actif et régulier des plis vocaux dont la tension est appropriée grâce aux ajustements des arytenoïdes et (ii) une pression transglottique suffisante qui permet le maintien des vibrations laryngées.

La différence majeure entre les consonnes voisées et les consonnes non voisées résulte essentiellement de la tension des plis vocaux. En effet, Halle & Stevens (2002 [1971]) précisent que la réalisation du voisement des consonnes dépend de la tension des plis vocaux. Ainsi, les consonnes non voisées ou les voyelles à ton haut seraient corrélées à la raideur (*stiffness*) des plis vocaux, tandis que les consonnes voisées ou les voyelles à ton bas seraient produites avec des plis vocaux relâchés (*slack*). En effet, Stevens (1991) précise qu'un état neutre du larynx est impossible et que les plis vocaux doivent être soit tendus, soit relâchés pour marquer cette opposition de voisement.

La production des consonnes non voisées [p, t, k] — segments non marqués phonologiquement car les plus répandus dans les langues — est liée à un état de raideur des plis vocaux et du conduit vocal qui entrave la vibration. Stevens (1998) ajoute que la vibration est inhibée durant la phase de fermeture par une augmentation de la rigidité des parois du conduit vocal, ce qui empêche l'augmentation des volumes des cavités. Faaborg-Andersen (1957) ajoute que l'abduction des plis vocaux est principalement due à l'arrêt de l'activité des muscles CAL, IA et TA associé à une action antagoniste du muscle CAP.

En ce qui concerne la production des consonnes voisées [b, d, g], elle nécessite un état relâché des plis vocaux. Ce relâchement permet le maintien d'une vibration glottique même à faible pression transglottique pendant la fermeture buccale. Le voisement nécessite le maintien d'un différentiel entre PSG et PIO. A ce titre, des ajustements glottiques et supraglottiques sont nécessaires pour permettre à l'air de continuer de s'écouler. Westbury (1983) explique le processus par une expansion de la paroi pharyngée et du volume des cavités supraglottiques, caractérisées par une baisse du larynx, un avancement de la racine de la langue et de l'épiglotte, une remontée du palais mou. Un relâchement actif et un épaississement des plis vocaux, lié au muscle vocal et au muscle CT, peuvent également entretenir la vibration. Ces ajustements restent

beaucoup plus simples pour la production des labiales par rapport aux consonnes dentales ou vélaires, ce qui explique que le [b] est plus fréquent dans les langues que le [g] (Maddieson 1984).

La raideur des plis vocaux et un fait difficile à déterminer car aucun indice ne prouve que la production d'une consonne voisée est plus relâchée qu'une consonne non voisée. Seule la  $F_0$  au début de la voyelle adjacente peut être un indice de la tension des plis vocaux. Cette observation avait déjà été faite par Halle & Stevens (2002 [1971]) ou Hyman & Schuh (1974), qui avaient constaté que les consonnes voisées avaient tendance à abaisser les tons voisins alors que les consonnes non voisées élevaient les tons adjacents. Ainsi, la raideur des plis vocaux lors de la production des consonnes non voisées engendre une  $F_0$  plus élevée à l'initial de la voyelle suivante. Pour Stevens (1991) ce postulat est lié au relâchement des plis vocaux pendant l'explosion de la consonne. Vaissière (1997) interprète cette élévation de la  $F_0$  après des consonnes non voisées comme une « assimilation progressive de la perte de la tension des plis vocaux ». Pour Maddieson (1997), cette hausse de  $F_0$  est due à l'apparition d'un grand flux transglottique au relâchement de la consonne.

Stevens (1977) ajoute que les mouvements verticaux du larynx participent au contraste de voisement et ce plus particulièrement lors de l'implémentation du trait [slack] spécifié pour les consonnes voisées. Ainsi, celui-ci note qu'une position haute du larynx permet d'allonger les plis vocaux et d'accroître leur raideur. À l'inverse, une position basse du larynx permet un plus grand relâchement des plis vocaux et des parois laryngées, ce qui facilite le maintien du voisement. Stevens (1998) a démontré que le larynx est abaissé de 4 à 6 mm pour la production des occlusives voisées par rapport à leurs homologues non voisées. La conséquence de ce mouvement est que les plis vocaux se raccourcissent de 2-3%, ce qui tend à diminuer leur rigidité de 9 à 15% et à abaisser de 5 à 7% la  $F_0$  au début de la voyelle suivante. Dans cette logique, Honda & al. (1999) ont montré dans une étude IRM sur le [a] japonais que la montée du larynx permettait un allongement des plis vocaux, ce qui se traduisait par une hausse de  $F_0$ .

## 6.3 Les corrélats acoustiques du voisement

Dans cette section, nous mettrons en évidence les principaux corrélats acoustiques liés au voisement dans les langues, et en particulier le Voice Onset Time (VOT).

### 6.3.1 Le temps d'établissement de voisement

La plupart des langues disposent dans leurs systèmes consonantiques de deux ou plusieurs occlusives homorganiques. Les indices qui permettent de distinguer ces occlusives sont nombreux et dépendent des langues à analyser. Néanmoins, nous notons que le *timing* laryngé reste l'indice principal — associé au mode de vibration des plis vocaux — lié à ces oppositions. Le *timing* laryngé permet de mettre en évidence la relation existante entre les mouvements de la glotte et la mise en place des articulateurs supra-laryngés (Lisker & Abramson 1964). Pour les occlusives en position initiale, le

corrélat acoustique principal du *timing laryngé* est le *Voice Onset Time* (VOT), à savoir « l'intervalle de temps entre le relâchement de la consonne et le début des premières vibrations laryngées, correspondant au début du voisement » (Lisker & Abramson 1964). Plus tard, Klatt (1975) a redéfini le VOT comme « l'intervalle de temps entre le relâchement consonantique et l'apparition d'une structure formantique stable de la voyelle adjacente ». Cho & Ladefoged (1999) comparent le VOT à un trait phonologique qui résulte de la différence entre la durée de l'établissement du geste articulatoire (relâchement) et la durée de la mise en vibration des plis vocaux. De manière générale, les auteurs notent un VOT négatif lorsque le début des vibrations intervient avant le relâchement de l'occlusion et un VOT positif quand le voisement démarre après le relâchement de la consonne.

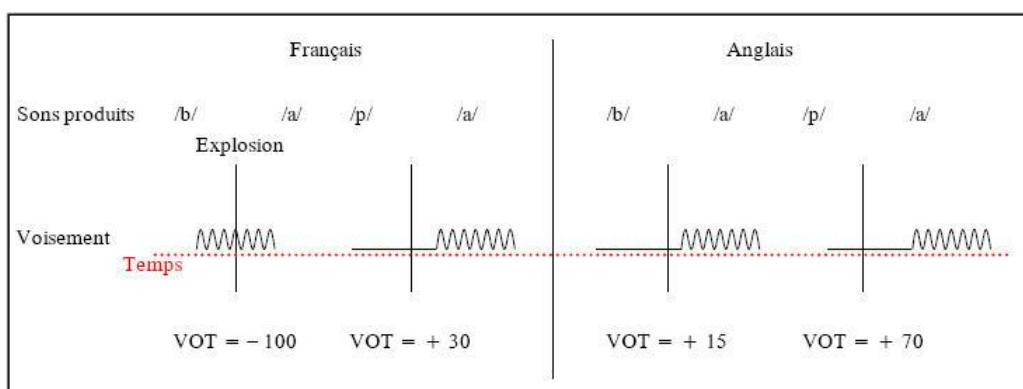
Dans une étude sur 11 langues, Lisker & Abramson (1964) ont cherché à mettre en évidence des tendances générales dans l'utilisation du *timing laryngé*. Plus tard, Cho & Ladefoged (1999) ont confirmé ces résultats préliminaires dans une étude sur 18 langues différentes qui opposent des consonnes occlusives aspirées à leurs contreparties non aspirées et voisées. Les auteurs montrent qu'il existe trois modes distincts de productions de ces occlusives. Ainsi, nous relevons les occlusives voisées [b, d, g], les occlusives dévoisées [p, t, k] et les occlusives non voisées aspirées [p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>].

Les auteurs précisent que les durées des VOT sont variables et très sensibles aux lieux d'articulation. Dans chaque langue, le VOT des vélaires semble plus long que celui des dentales ou des labiales (Peterson & Lehiste 1961, Lisker & Abramson 1964, Cho & Ladefoged 1999). Plusieurs explications liées à l'articulatoire et à l'aérodynamique ont été avancées pour généraliser cette tendance. La première explication, d'ordre aérodynamique, prédit que plus l'aire de contact linguo-palatal est étendue et plus le VOT sera allongé. Or, l'occlusive vélaire présente un large contact entre le dos de langue et la palais, ce qui engendre une déperdition d'air plus lente par rapport aux labiales. En effet, les auteurs précisent que la PIO après le relâchement change à une vitesse qui est en relation directe avec l'augmentation de l'aire à la constriction. En l'espèce, les vélaires sont caractérisées par une diminution plus lente de la PIO après la fermeture par rapport aux autres consonnes (Stevens 1998). La deuxième explication, d'ordre articulatoire, prédit quant à elle que plus le mouvement des articulateurs est rapide et plus court sera le VOT, ce qui explique pourquoi les labiales ont un VOT bien plus court que les dentales et les vélaires car les mouvements des lèvres et/ou de l'apex sont plus rapides que les mouvements de l'arrière de la langue (Hardcastle 1973). Enfin la troisième explication, également d'ordre articulatoire, prédit que la durée du VOT est dépendante soit de la taille de la cavité supraglottique, soit du lieu de la constriction. Ainsi, la cavité derrière une occlusion vélaire a un volume plus petit que lorsqu'il s'agit d'une occlusion dentale ou labiale. À l'inverse, la cavité en avant d'une occlusion vélaire a un volume plus important par rapport aux dentales ou aux labiales. D'un point de vue aérodynamique, la petitesse de la cavité en arrière du point de constriction pour les vélaires implique une élévation de la pression pendant le relâchement de la consonne. Or, cette augmentation aura pour conséquence d'élargir également la pression de l'air au niveau des plis vocaux. Si la pression est plus importante, il est logique qu'elle mette plus de temps à se rééquilibrer pour la mise

en vibration des plis vocaux. Parallèlement, en avant de la constriction, la masse d'air accumulée doit se vider progressivement. Cette cavité étant plus importante pour les vélaires, le temps d'évacuation sera plus long pour ces consonnes. Les deux cavités doivent se vider totalement pour que l'activité glottique se mettent en place, ce qui explique pourquoi les occlusives vélaires présentent un VOT plus long par rapport aux dentales et aux labiales (Maddieson 1997).

Cette tendance se retrouve pour l'ensemble des langues étudiées, excepté pour la mesure du VOT de l'occlusive non voisée non aspirée [p] qui semble plus longue que celles du [t] ou du [k] en Tamil par exemple ; ou la mesure de l'occlusive non voisée aspirée [ $p^h$ ] qui est allongée par rapport à [ $t^h$ ] ou [ $k^h$ ] en cantonais (Abramson & Lisker 1964). Cho & Ladefoged (1999) ont également mis en évidence une langue — le dahalo — pour laquelle la mesure du VOT des labiales est supérieure à celle de la dentale, et pour laquelle la mesure du VOT de la vélaire est supérieure à celle de l'alvéolaire. Les auteurs précisent que ces locuteurs ont peut être opté pour des vitesses d'élocution inhabituelles, ce qui constituerait un biais dans leurs résultats. En effet, il est évident que la vitesse d'élocution est directement corrélée à la durée du VOT. L'augmentation de la vitesse d'élocution est responsable d'une accélération des mouvements des articulateurs et la lenteur de la vitesse d'élocution implique également un ralentissement de l'activité des articulateurs.

Serniclaes (1987) ajoute que la mesure du VOT est également dépendante de l'accentuation et de la position de l'occlusive dans la syllabe. Pour l'anglais, l'auteur précise que les occlusives non voisées sont aspirées en position initiale ou en syllabe accentuée, ce qui allonge le VOT puisque celui-ci est de 70 ms. Les occlusives voisées sont, quant à elles, souvent dévoisées (voir aussi Lisker & Abramson 1967 et Klatt 1975). *A contrario*, le français suit une opposition stricte de voisement, et ce quelle que soit la position de l'occlusive. Ainsi, les occlusives voisées ont un VOT négatif, ce qui induit une vibration glottique antérieure à la détente de la consonne et les occlusives non voisées sont produites avec un VOT positif, c'est-à-dire que le voisement intervient au début de la voyelle suivante (cf. figure 6.1).



**Figure 6.1 – Mesures en ms du VOT en français et en anglais (valeurs extraites de Serniclaes 1987).**

### 6.3.2 Paramètres acoustiques secondaires liés au voisement

La perception du trait phonologique de voisement est principalement catégorielle. Ainsi, cette perception catégorielle présuppose que les auditeurs sont capables de faire une distinction entre les catégories ([+voix] vs [-voix]) et non à l'intérieur de ces catégories. Ainsi plusieurs auteurs se sont intéressés à la perception du trait de voisement du point de vue de la variation de ses corrélats acoustiques. Plusieurs études dans diverses langues ont ainsi démontré que la modification du VOT sur des stimuli synthétiques ou la modification de la durée de la vibration des plis vocaux en parole engendraient des perturbations dans les tâches d'identification de consonnes non voisées ou voisées (Lisker & Abramson 1964, Williams 1977, Van Dommelen 1983, Van Dommelen 1985, Serniclaes 1987, entre autres). Par déduction, nous suggérons que le VOT — corrélé principalement à la durée de vibration des plis vocaux — reste l'indice principal à la présence du voisement.

Pour l'anglais, Lisker (1986) précise que le VOT est l'indice primaire du voisement dans une tâche de reconnaissance des mots anglais « *rabid* » et « *rapid* ». Il ajoute qu'une combinaison de 16 indices acoustiques indépendants les uns des autres peuvent contribuer à identifier le trait de voisement. Selon lui, les indices les plus pertinents restent ceux perçus durant l'occlusion, à savoir la durée de l'occlusion, la durée du signal glottal et l'intensité du signal glottal. Des indices secondaires liés à la phase de pré-occlusion (durée de la voyelle, durée de la transition du F1, fréquence de la fin du F1, temps de la fin de la transition du F1, *timing* de la fin de voisement, contour de la F<sub>0</sub>, temps de disparition du signal) et à la phase de post-occlusion (intensité du relâchement, temps d'établissement de la transition du F1, fréquence de l'établissement du F1, durée de la transition du F1, contour de la F<sub>0</sub>) peuvent être associés. De manière générale, les auteurs notent qu'en position initiale, les occlusives non voisées aspirées sont associées à une F<sub>0</sub> plus élevée à la fin du voisement par rapport aux occlusives voisées (Lisker & Abramson 1971). De même, les voyelles qui suivent des occlusives voisées sont plus longues que celles qui suivent des occlusives non voisées (Peterson & Lehiste 1961). Enfin, les durées d'occlusions semblent plus longues pour les occlusives non voisées par rapport à leurs contreparties voisées.

Pour le français, nous pouvons par exemple citer les expériences de Dommelen (1983, 1985). Dans une tâche d'identification du trait de voisement en final, cet auteur a fait varier la durée de vibration des plis vocaux dans l'occlusion, le bruit d'explosion, la durée de l'occlusion et la durée de la voyelle précédant l'occlusive dentale dans la paire « *rate/rade* » issue du contexte phrastique « *C'est une rate/rade* ». Les résultats — pour la position finale — prouvent que la durée de vibration des plis vocaux reste un paramètre essentiel dans la distinction entre le [t] et le [d], mais qu'il doit être associé à la durée vocalique. En effet, une voyelle allongée permet une meilleure identification des occlusives voisées. Dans une seconde expérience, l'auteur a fait varier ces mêmes paramètres en position non finale dans la phrase cadre « *C'est une rate/rade noire* » afin de valider ses conclusions préliminaires. Celui-ci note que la durée de vibration des plis vocaux dans l'occlusion, tout comme la durée de l'occlusion, restent les indices les plus pertinents. La qualité de la voyelle précédant la consonne ne doit pas être

négligée dans la mesure où une occlusion voisée précédée d'une voyelle issue d'un contexte non voisé ne conduit pas forcément à la perception d'un [d]. De la même manière, une occlusion non voisée précédée d'une voyelle issue d'un contexte voisé ne conduit pas forcément à la perception d'un [t]. Cette observation est également partagée par Luce & Charles-Luce (1985), qui soulignent qu'une voyelle précédée par une consonne voisée aura tendance à être allongée. En outre, nous notons que la durée de l'occlusion est plus courte pour les occlusives voisées par rapport aux occlusives non voisées. De même, l'intensité de l'explosion est plus faible pour les occlusives voisées par rapport aux occlusives non voisées (Serniclaes 1987). Par ailleurs, Stevens & Klatt (1974) insistent sur l'importance d'associer VOT et transition du F1 pour la distinction de voisement. Selon eux, l'existence ou non d'une transition du F1 après de début du voisement serait vraisemblablement un indice bien plus cohérent que la valeur absolue du VOT. Toutefois, les auteurs restent prudents quant à la prise en compte des transitions de F2 et F3 pour l'implémentation du voisement, réflexion partagée par Summerfield & Haggard (1977). Enfin, bien que différentes théories s'opposent quant aux variations de la  $F_0$ , les valeurs de la  $F_0$  à la fin du voisement restent également un indice de voisement non négligeable.

## 6.4 Les occlusives orales du français

Les occlusives orales étant les composantes principales de notre corpus en contexte, cette section sera entièrement dédiée à leur description des points de vue articulatoire, aérodynamique et acoustique.

Une occlusive est une consonne produite avec une constriction totale dans le conduit vocal. Les occlusives orales, par opposition aux nasales, sont définies par une obstruction complète dans le conduit oral. En français standard, les occlusives orales ne sont ni aspirées, ni affriquées, ni implosives, ni éjectives. Seule une opposition de voisement classique permet d'opposer les occlusives orales non voisées [p, t, k], aux occlusives orales voisées [b, d, g]. Ainsi, ces consonnes sont produites soit avec les plis vocaux en abduction (non voisées) ou en adduction (voisées). D'autres configurations intermédiaires de la glotte peuvent apparaître — *breathy voice* ou *creaky voice* — mais ne sont pas phonologiquement distinctives en français.

### 6.4.1 Caractéristiques articulatoires

Une occlusive est une consonne qui présente une fermeture complète à l'intérieur du conduit vocal, entre la glotte et les lèvres. Ces consonnes sont classées selon l'endroit où se réalise la constriction (le lieu d'articulation). Le système des occlusives orales du français est composé de la manière suivante :

	- voix	+ voix
Labiales/Bilabiales	p	b
Coronales/Labio-dentales	t	d
Dorsales/Dorso-vélaires	k	g

**Table 6.1** – Système des occlusives orales du français standard

A partir d'une étude radiocinématographique des consonnes du français, Simon (1967) a confirmé trois phases articulatoires distinctes dans la production des occlusives : (i) une mise en position des organes qui changent de position pour la fermeture, (ii) une phase de maintien des articulateurs dans la position adoptée et (iii) une phase de relâchement durant laquelle les organes abandonnent leurs positions pour se préparer à l'articulation suivante ou se détendre.

#### 6.4.2 Caractéristiques acoustiques

Les phases articulatoires correspondent à des indices acoustiques caractéristiques.

La mise en place des articulateurs se traduit par des mouvements des formants. L'occlusion durant les consonnes non voisées est représentée par un silence sur un spectrogramme. Seul le relâchement sera identifiable dans cette position particulière.

La phase de relâchement correspond à l'apparition d'une barre d'explosion (relâchement /impulsion). Ce relâchement apparaît comme une libération brusque de l'air comprimé par la constriction (bruit impulsif). Dans un second temps, un bruit de friction est créé au lieu de constriction, identique à celui de la fricative correspondante. Enfin, un bruit d'aspiration peut apparaître. De manière générale, ces trois différents types de bruits sont rarement séparables dans les spectrogrammes et représentent une unité complète : le bruit d'explosion. Le lieu d'articulation de l'occlusive a une incidence sur la distribution du bruit d'explosion (et de friction) car cette impulsion dépend du volume de la cavité en avant de la constriction. Les occlusives labiales orales sont identifiables à un bruit d'explosion faible et diffus et prédominant dans les basses fréquences. Les occlusives dentales orales sont caractérisées par un bruit d'explosion assez intense et diffus dans les hautes fréquences (au delà de 5kHz). Enfin, les fréquences renforcées du bruit d'explosion des vélaires dépendent largement du contexte vocalique. Pour des vélaires palatalisées telles [ki] ou [gi], le bruit compact et intense est concentré majoritairement dans les hautes fréquences, autour du F3 de la voyelle suivante. Au contraire, pour les vélaires pures telles [ku] ou [gu], le bruit compact et intense est localisé dans les basses fréquences, autour du F2 de la voyelle adjacente.

La qualité du relâchement dépend également de la position dans le mot (Benguerrel & al. 1978). L'occlusive dentale voisée [d] aura un relâchement plus intense à l'initial absolu — la glotte étant fermée plus légèrement, la PIO monte plus vite — par rapport à la position intervocalique. Cette configuration glottique a été observée également pour le hindi (Dixit 1989) ou le suédois (Lindqvist 1972).

En outre, comme nous l'évoquions, les occlusives en position pré-vocalique sont également suivies de transitions formantiques de la voyelle adjacente. La durée et la

transition de F1 sont des indices perceptivement liés au trait de voisement : la vitesse de transition de F1 est plus rapide pour les consonnes non voisées. Quant au F2, sa transition est directement liée au lieu d'articulation de la consonne.

Sur un spectrogramme, les occlusives voisées sont repérables à la présence d'une barre de voisement qui se caractérise par une succession de stries verticales dans les basses fréquences (100-300 Hz). L'intervalle régulier entre ces stries correspond aux cycles glottiques. Les occlusives non voisées sont produites avec un retard de voisement, c'est-à-dire que la vibration glottique intervient après le relâchement de l'occlusion de la consonne. Le VOT est alors positif, de l'ordre de 10 à 30 ms. Au contraire, la production des occlusives voisées nécessite une mise en vibrations des plis vocaux pendant la phase d'occlusion, c'est-à-dire pendant le silence acoustique. Il est toutefois important de noter que le voisement a pour conséquence l'atténuation des bruits d'explosions. Dans ce cas, le VOT est négatif, de l'ordre de -130 ms environ.

### 6.4.3 Caractéristiques aérodynamiques

Les occlusives orales du français sont des consonnes pulmonaires (contrairement aux éjectives et implosives par exemple), c'est-à-dire que leur production nécessite un flux d'air qui provient directement des poumons. Le cheminement du flux pulmonaire est dit égressif, dans la mesure où celui-ci part des poumons pour traverser la glotte puis les cavités supraglottiques.

Les caractéristiques aérodynamiques des occlusives non voisées et voisées ayant déjà été abordées dans la section 6.2.2, nous nous permettons simplement de rappeler que les premières présentent un équilibre entre PIO et PSG alors que les secondes nécessitent une différence entre ces pressions afin de maintenir une pression transglottique suffisante pour la vibration des plis vocaux.

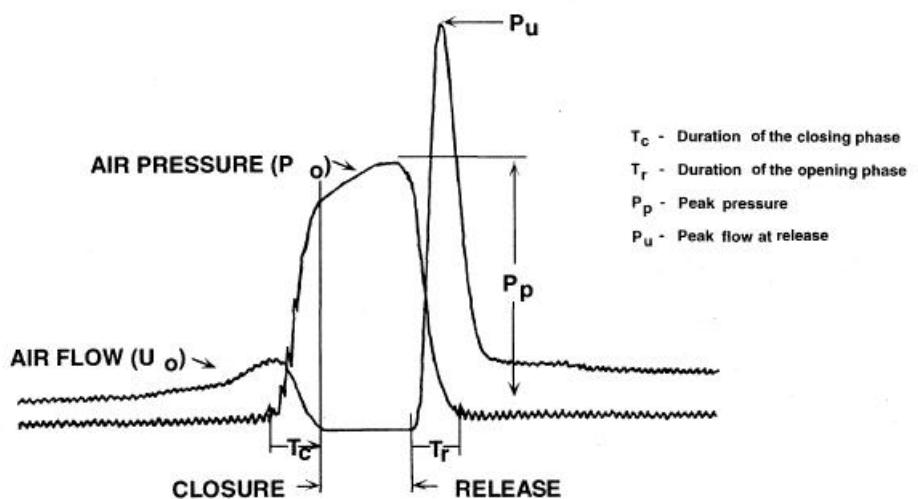
## 6.5 Etude n°6 : études aérodynamiques

Après avoir mis en évidence la présence de fuites glottiques sur des voyelles tenues et isolées, nous couplerons une analyse qualitative des *patterns* de PIO observés avec une analyse quantitative de certains paramètres aérodynamiques sur des logatomes afin de caractériser au mieux les stratégies adoptées par les patients.

### 6.5.1 Les phases aérodynamiques des occlusives

Comme nous l'avons mentionné précédemment, les occlusives se composent de trois phases distinctes : une fermeture, une phase de tenue de l'occlusion et un relâchement. Dans une étude sur 5 hommes anglophones qui avaient pour tâches les répétitions de dissyllabes de type VCV (où C= [p, t, b, d] et V= [a, i]), Müller & Brown (1980) ont mis en évidence les liens entre les courbes de PIO et de DAB pour la distinction de ces événements (cf. figure 6.2). Premièrement, la phase de fermeture pendant laquelle l'occlusion se met en place est caractérisée par une PIO qui augmente progressivement derrière le point de constriction pendant que le DAB diminue pour atteindre

une valeur de zéro lorsque l’occlusion est réalisée. Deuxièmement, la phase de tenue de l’occlusion pendant laquelle les articulateurs maintiennent une position de fermeture est repérable au fait que la PIO continue d’augmenter pour atteindre sa valeur maximale tandis que le DAB reste stable et nul car la fermeture empêche le passage de l’air. Troisièmement, la phase de relâchement ou d’ouverture de l’occlusion correspondant à une émission rapide et précipitée de l’air marquée par une décroissance de la PIO, qui atteint une valeur nulle à la fin de l’ouverture. Au contraire, le DAB va augmenter très rapidement pour atteindre sa valeur maximale puis redescendre progressivement.



**Figure 6.2 – Courbes de PIO et de DAB pendant la production d’un [p] (Müller & Brown 1980).**

Les auteurs proposent une segmentation particulière selon l’évolution des deux courbes. La phase  $T_c$  correspond à la fermeture. Celle-ci démarre au moment où la PIO quitte son point zéro pour commencer son ascension et s’achève quand le DAB atteint sa ligne de base. Dès lors, les auteurs mettent en évidence une phase de maintien de cette occlusion durant laquelle le DAB reste nul et la PIO continue d’augmenter pour atteindre sa valeur maximale ( $P_p$ ). La phase  $T_r$  correspond au relâchement de la consonne. Celle-ci débute lorsque la PIO commence à décroître tandis que le DAB quitte sa ligne de base pour atteindre sa valeur maximale ( $P_u$ ).

Müller & Brown (1980) ont également pu décrire qualitativement les formes des courbes de PIO à partir de normalisations des phases  $T_c$ . Les auteurs suggèrent que les formes des courbes peuvent être qualifiées de convexes, concaves, linéaires, bimodales ou retardées. Les changements de direction, quant à eux, peuvent être continus ou discontinus. Ils précisent que ces formes particulières ne sont ni influencées par la qualité de la consonne ni par la voyelle associée. Cependant, quelques tendances semblent émerger. Ainsi, les occlusives non voisées seraient majoritairement du type convexe

alors que les occlusives voisées auraient tendance à être non convexes. Les occlusives non voisées seraient également plus influencées par les contextes vocaliques que leurs correspondantes voisées.

## 6.5.2 Etude préliminaire sur les fuites glottiques des voyelles isolées

### 6.5.2.1 Tests statistiques préalables

Nous avons déjà souligné le fait qu'il était nécessaire de montrer que nos deux populations (témoins vs patients) étaient distinctes avant de pouvoir les comparer statistiquement. Contrairement au chapitre précédent — et principalement pour des raisons de gain de place — nous ferons l'économie des graphiques dans la présentation des résultats des tests statistiques préliminaires.

**Test de Lilliefors.** Le test de normalité de Lilliefors a été appliqué sur l'ensemble des valeurs de DAB des voyelles stables ([a, i, u]) ( $n = 60$ , soit 3 voyelles  $\times$  20 sujets). L'hypothèse nulle( $H_0$ ) pour ce test est que l'erreur est normalement distribuée, c'est-à-dire qu'il n'y a pas de différence entre la distribution observée de l'erreur et une distribution normale. L'hypothèse alternative ( $H_1$ ) est que l'erreur n'est pas normalement distribuée. Les résultats de ce test sont présentés dans le tableau suivant :

	[a]		[i]		[u]	
	D	p	D	p	D	p
Témoins	.232	.1325	.2271	.1521	.1919	.3716
Patients	.2111	.2346	.1738	.5321	.1879	.4049

Table 6.2 – Test de normalité de Lilliefors appliqué sur les valeurs moyennes du DAB des voyelles [a], [i], et [u] ( $n = 60$ ).

Comme nous l'avons mentionné précédemment, la valeur de  $D$  ne doit pas être supérieure ou égale à .2616 pour accepter  $H_0$  avec un effectif  $n = 10$  et un seuil critique  $\alpha = .05$ . Les résultats de ces tests montrent que les distributions des trois voyelles selon les deux groupes s'intègrent à des distributions suivant une loi normale. En effet, la valeur de  $p$  est supérieure à .05 dans tous les cas et, par conséquent, l'hypothèse nulle doit être acceptée.

**Test F d'homogénéité des variances.** Le test précédent a pu mettre en évidence le fait que nos deux groupes suivaient une distribution normale. Il est maintenant essentiel de vérifier que nos deux populations forment bien deux groupes distincts afin de pouvoir les comparer deux à deux. Nous avons donc calculé le rapport des variances estimées entre nos deux populations sur l'ensemble des valeurs de DAB pour les trois voyelles stables. Le tableau 6.3 résume les conclusions de ce test :

	Valeurs de $F$	Valeurs de $p$
[a]	1.699	.4419
[i]	3.552	.0728
[u]	1.548	.5255

**Table 6.3** – Résultats synthétiques du test F sur les valeurs du DAB pour les trois voyelles [a], [i], [u] ( $n = 60$ ).

Ainsi, nous notons que pour les trois voyelles, les seuils de significativité sont toujours supérieurs à .05, ce qui implique que  $H_0$  n'est pas rejetée et que par conséquent les variances des deux distributions sont semblables. Les populations forment bien deux groupes indépendants et comparables.

#### 6.5.2.2 Méthodologie

Comme nous l'avons déjà mentionné dans le chapitre 4, nos mesures aérodynamiques ont été acquises à l'aide de l'appareillage EVA2 (Galindo & al. 1989, Teston & al. 1999), qui permet d'enregistrer simultanément des paramètres acoustiques classiques ( $F_0$ , intensité) et des paramètres aérodynamiques tels que le DAB et la PIO. Pour cette étude préliminaire, il était demandé aux sujets d'inspirer profondément afin d'émettre les voyelles tenues [a, i, u] à hauteur et intensité confortables dans le masque buccal ajusté de manière étanche pendant au moins 2 secondes ( $n = 60$ ).

Le traitement des résultats a été effectué à l'aide du logiciel Phonetedit. Pour chacun des enregistrements, nous avons obtenu trois courbes distinctes : le signal acoustique, la variation de l'intensité et la variation du DAB (cf. figure 6.3). Les prises de mesures ont été réalisées manuellement en plaçant le curseur de sélection sur la parties à analyser. Afin d'éviter les variations importantes intra-sujets concernant le début et la fin des voyelles, nous nous sommes concentrée sur les parties les plus stables des voyelles. Effectivement, le pic initial de DAB représente un biais dans la mesure où il peut amener à une surestimation de ce DAB. Notre sélection manuelle représente donc une fenêtre de 50 ms sur les signaux acoustiques des segments vocaliques. Nous avons ainsi pu extraire les paramètres suivants : l'intensité sonore moyenne (en dB), le débit d'air buccal moyen ( $dm^3/s$ ), que nous nommerons désormais Débit Phonatoire Moyen (cf. figure 6.12). A l'instar de Giovanni & al. (1995), nous avons fait appel à ces deux paramètres pour calculer un indice secondaire lié aux fuites glottiques.

Cet indice a été introduit pour quantifier le souffle nécessaire dans le production d'un [a] tenu chez 89 sujets normaux et 159 patients dysphoniques. En parallèle d'une étude perceptive, les auteurs ont tenté d'évaluer la fuite glottique de ces patients — dont les plis vocaux ne s'accroient pas sur la totalité de leurs parties vibrantes — par l'introduction de la formule suivante : le rapport de la mesure du DAB par la mesure de l'intensité pendant un même temps, soit :  $DAB/INTENSITE$ . Ainsi, le résultat

exprimé, en  $cm^3/dB/s$ ,<sup>5</sup> correspond à la quantité d'air consommée pour produire 1 dB en 1 seconde sur une même portion vocalique. Les auteurs mettent en évidence qu'il existe une corrélation positive quasi-parfaite entre ce calcul de « fuites » et les valeurs de DAB relevées malgré une variabilité intra-individuelle persistante.

Toutes nos prises de mesures, ainsi que nos calculs de « fuites glottiques » ont fait l'objet d'un traitement statistique avec Statview et R.

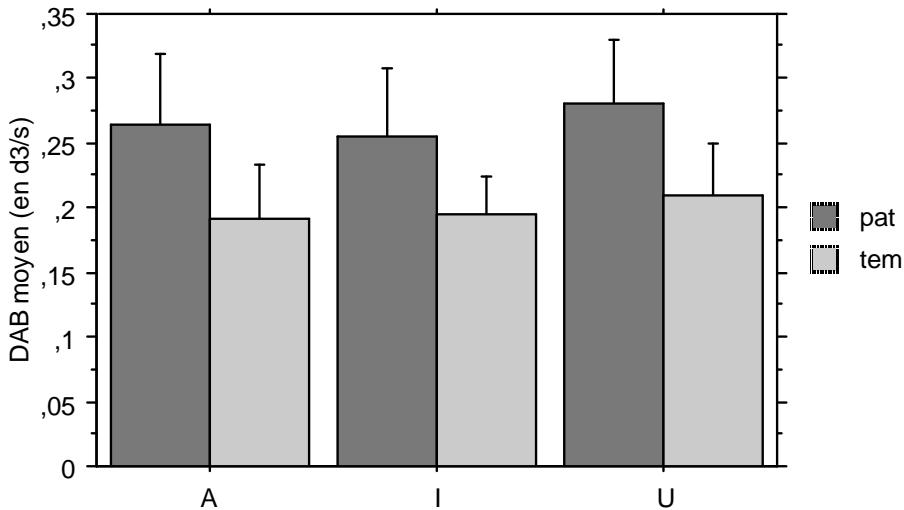


**Figure 6.3** – Courbes obtenues pour les voyelles tenues [a, i, u] sur Phonedit (exemple du témoin n°1). En haut, l'intensité sonore émise (dB) ; au milieu, l'oscillogramme ; en bas, le DAB en  $dm^3/s$ .

### 6.5.2.3 Présentation des résultats

**Définition des valeurs de DAB entre nos deux groupes.** La figure 6.4 met en évidence les différences inter-groupes pour le paramètre du DAB pour les trois voyelles étudiées. Visuellement, nous remarquons que les valeurs moyennes de DAB des patients sont plus hautes que celles des témoins pour chacune des voyelles.

5. Notons que les mesures de volume ne sont pas les mêmes entre les valeurs données par Phonédit pour le DAB ( $dm^3/s$ ) et le résultat attendu après l'application de notre formule ( $cm^3/dB/s$ ). Nous avons donc eu recours à une conversion où  $1 cm^3 = 0,001 dm^3$ .



**Figure 6.4** – Valeurs moyennes du DAB (en  $\text{dm}^3/\text{s}$ ) des voyelles [a], [i], [u] des témoins et des patients (les barres d'erreurs représentent les écarts-types) ( $n = 60$ ).

Afin de déterminer si les valeurs moyennes de DAB des patients sont significativement différentes de celles des témoins, nous avons conduit des tests-t non appariés sur les données issues de la production des trois voyelles. Les résultats sont considérés significatifs à partir d'un seuil alpha de 5% (soit  $p < 0.05$ ). Le tableau 6.4 met en évidence les résultats de cette expérience pour ce paramètre aérodynamique.

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	.191	.265	.195	.255	.210	.281
$s$	.042	.054	.029	.054	.04	.049
$t_{(18)}$	3.407		3.098		3.596	
$p$	.0031		.0062		.0021	

**Table 6.4** – Résultats des tests-t non appariés conduits sur les valeurs de DAB (en  $\text{dm}^3/\text{s}$ ) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Pour les témoins, nous remarquons la hiérarchie suivante :  $\text{DAB}[a] < \text{DAB}[i] < \text{DAB}[u]$ . En ce qui concerne les patients, nous notons la hiérarchie suivante :  $\text{DAB}[i] <$

$DAB[a] < DAB[u]$ . Néanmoins, il n'existe pas de différence statistiquement significative entre les valeurs moyennes du DAB de chacune des voyelles, et ce pour les deux populations. Il est également important de souligner que pour les trois voyelles, nous relevons des différences hautement significatives entre les valeurs moyennes du DAB des témoins et celles des patients. Pour [a], la p-value entre les deux groupes est de  $p = .0031$ ; pour [i], la p-value est de  $p = .0062$  et enfin pour [u], la p-value est égale à  $p = .0021$ .

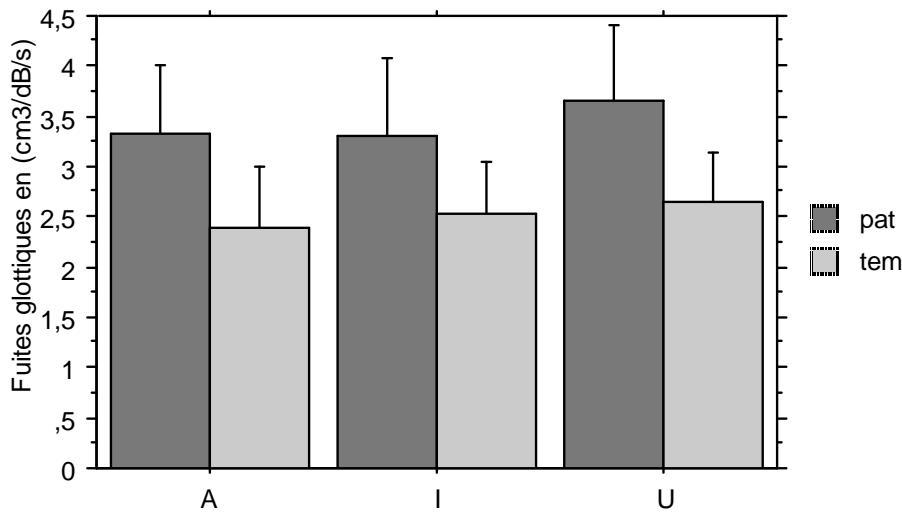
**Différences des valeurs d'intensité entre les deux groupes.** Il est important de rappeler que cette recherche n'a pas le but de mettre en évidence une différence inter-groupes (cf. section 5.5.2). Nous avons simplement relevé les intensités moyennes de chacune des trois voyelles afin de justifier notre calcul des indices de fuites glottiques. Le tableau suivant présente les résultats des tests-t non appariés appliqués sur les valeurs de l'intensité des voyelles stables des deux groupes :

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	80.800	79.705	77.894	77.445	79.555	77.580
$s$	6.536	7.344	5.494	7.482	5.893	7.251
$t_{(18)}$	-.352		-.153		-.668	
$p$	.7287		.8801		.5124	

**Table 6.5** – Résultats des tests-t non appariés conduits sur les valeurs d'intensité (en dB) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Le tableau 6.5 montre qu'il n'existe aucune différence significative entre les deux groupes pour les trois voyelles puisque les p-values pour [a, i, u] sont respectivement égales à  $p = .7287$ ,  $p = .8801$  et  $p = .5124$ .

**Différences d'indices de fuites glottiques entre les deux groupes.** La figure 6.5 met en évidence les différences inter-groupes pour les indices de fuites glottiques et ce pour les trois voyelles étudiées. Visuellement, nous remarquons que les indices calculés pour le groupe des patients sont plus élevés que ceux des témoins pour chacune des voyelles.



**Figure 6.5** – Valeurs moyennes des Indices de fuites glottiques en ( $cm^3/dB/s$ ) des voyelles [a], [i], [u] des témoins et des patients (les barres d'erreurs représentent les écarts-types) ( $n = 60$ ).

Afin de déterminer si les valeurs du rapport des mesures du DAB par les mesures de l'intensité du groupe des patients sont significativement différentes de celles des témoins, nous avons conduit une série de tests-t non appariés sur les données issues des productions des trois voyelles. Les résultats sont considérés significatifs à partir d'un seuil alpha de 5% (soit  $p < 0.05$ ). Le tableau 6.6 met en évidence les résultats de cette expérience pour ce paramètre aérodynamique.

	[a]		[i]		[u]	
	témoins	patients	témoins	patients	témoins	patients
$\bar{x}$	2.387	3.332	2.534	3.315	2.644	3.658
$s$	.621	.680	.521	.753	.507	.749
$t_{(18)}$		3.244		2.697		3.544
$p$		.0045		.0147		.0023

**Table 6.6** – Résultats des tests-t non appariés conduits sur les valeurs des indices de fuites glottiques (en  $cm^3/dB/s$ ) des voyelles [a], [i] et [u] ( $\bar{x}$  représente la moyenne arithmétique de l'échantillon,  $s$  l'écart-type non biaisé,  $t_{(18)}$  la valeur du test de Student pour un degré de liberté égal à 18 et  $p$  la probabilité) ( $n = 60$ ).

Ces résultats restent cohérents avec la comparaison inter-groupes que nous avons effectuée sur le paramètre du DAB. En effet, il est certain que les indices de fuites glottiques sont directement corrélés aux mesures de DAB (Giovanni & al. 1995), ce qui tend à conserver l'ordre pré-établi précédemment, à savoir, pour nos témoins,  $Fg[a] < Fg[i] < Fg[u]$  et, pour les patients,  $Fg[i] < Fg[a] < Fg[u]$ . Néanmoins, il n'existe pas de différence statistiquement significative entre les indices de fuites glottiques calculés pour chacune des trois voyelles, et ce pour les deux populations. En outre, il est important de souligner que pour les trois voyelles, nous notons des différences hautement significatives entre les valeurs moyennes des indices de « fuites glottiques » des témoins et celles des patients. Pour [a], la p-value entre les deux groupes est de  $p = .0045$ ; pour [i], la p-value est de  $p = .0147$  et enfin pour [u], la p-value est égale à  $p = .0023$ .

#### 6.5.2.4 Discussion

**Paramètre du Débit d'air buccal (DAB).** Il est intéressant de constater que les valeurs moyennes de DAB relevées pour les patients et les témoins ne suivent pas un ordre identique, ce qui pourrait poser problème dans nos interprétations. Pour les témoins, nous relevons la hiérarchie suivante :  $DAB[a] < DAB[i] < DAB[u]$ . Par contre, pour le groupe de patients, nous notons la hiérarchie suivante :  $DAB[i] < DAB[a] < DAB[u]$ . Nos résultats semblent être, malgré tout, en accord avec les conclusions de Vaughn (1965) ou Bucella & al. (2000) qui retrouvent une hiérarchie telle que  $DAB[i] < DAB[a] < DAB[u]$  dans leurs recherches. Dans sa thèse basée sur 20 hommes américains, Vaughn (1965) a montré les effets des changements de  $F_0$  et d'intensité sur les débits d'air buccaux des voyelles [i, u, ae, a]. Celle-ci montre que l'augmentation de  $F_0$  ou de l'intensité est corrélée à une hausse des débits d'air mais que l'ordre exposé précédemment reste stable. Aussi, dans une étude sur un unique locuteur francophone — ce qui pourrait biaiser l'exposition des résultats —, Bucella & al. (2000) ont montré que pour trois intensités et trois hauteurs différentes, les valeurs moyennes du DAB des voyelles [a, i, u] suivent cette même hiérarchie.

Néanmoins, nos conclusions semblent être en contradiction avec les résultats de Delvaux & Harmegnies (2011). Cette étude, basée sur les mesures des débits d'air buccaux et nasals des voyelles orales et nasales en contexte de nasalité, montre qu'un seul ordre n'est possible concernant le DAB : les voyelles hautes présentent toujours des valeurs de DAB plus basses que les voyelles ouvertes, et ce indépendamment du contexte. Les auteurs précisent que cette régularité est directement liée aux différences de configurations du conduit vocal pendant la production de ces voyelles et donc aux différences de résistance lors du passage de l'air dans le tractus vocal. En effet, il est démontré que la résistance liée à l'aperture est plus faible pour les voyelles ouvertes par rapport aux voyelles fermées, ce qui permet de prédire un DAB plus important pour la voyelle [a] par rapport à [i] ou [u].

Finalement, nous avons pu mettre en évidence une variabilité importante dans les conclusions des diverses études, et ce malgré une théorie aérodynamique qui semble robuste. Ces différences sont majoritairement dues à une évidente variabilité intra-sujets dans les prises de mesures aérodynamiques, en particulier pour les voix patho-

logiques (Schutte 1992 ; Higgins & al. 1998 ; Goozée & al. 1998). Ces derniers ont montré la variabilité des productions des sujets, y compris avec plusieurs répétitions, et appellent à la prudence dans l'interprétation des résultats. Goozée & al. (1998) restent incertains quant à l'intérêt des paramètres aérodynamiques pour l'identification des voix pathologiques dans la mesure où les relevés des débits d'air, d'intensité ou les calculs de la résistance laryngée et de l'efficience phonatoire présentent trop de variabilités, c'est-à-dire d'instabilités, chez les sujets témoins. Cette variabilité peut être due à des différences en production en termes de hauteur et d'intensité (Isshiki 1964, 1965 ; Vaughn 1965) ou à des différences physiologiques entre les locuteurs qui ne sont pas toujours contrôlables (Fant 1960, Hixon 1966, entre autres).

En ce qui concerne la présente étude, nos mesures n'échappent pas à cette réalité, ce qui pourrait être une explication à nos conclusions. Le tableau 6.4 met en évidence des écarts-types assez importants pour les deux populations et pour chacune des trois voyelles. De plus, le tableau 6.7 ci-dessous montre également des coefficients de variation (en %)<sup>6</sup> assez élevés pour les témoins et les patients, et ce pour les trois voyelles :

	[a]	[i]	[u]
témoins	21,7	14,6	18,8
patients	20,5	21,1	17,5

**Table 6.7** – Coefficients de variation (exprimés en pourcentages) des mesures de DAB pour nos deux populations

**Paramètre du DAB et indices des fuites glottiques.** Il est couramment admis que les valeurs de DAB sont révélatrices du comportement de la fonction laryngée. Ainsi, Iwata & al. (1972) ont étudié les paramètres aérodynamiques de 191 sujets présentant une pathologie laryngée. Ils ont mis en évidence que l'hyperfonctionnement laryngé était corrélé à des valeurs basses du DAB. Au contraire, l'hypofonctionnement laryngé était lié à des mesures de DAB élevées. De la même manière, une étude de Netsell & al. (1984) portant sur 18 sujets dysphoniques et 30 sujets témoins a révélé que la voix « breathy » était marquée par une insuffisance glottique, laquelle est caractérisée par une hausse du DAB. Au contraire, la voix « creaky » est marquée par une hyperadduction des plis vocaux et est liée à des valeurs basses du DAB. Finalement, l'hypofonctionnement laryngé correspond à un accolement insuffisant des plis vocaux (hausse du DAB) tandis que l'hyperfonctionnement laryngé se rapporte à un serrage glottique (baisse du DAB). Notre étude a mis en évidence des valeurs du DAB significativement plus hautes pour le groupe des patients par rapport aux témoins, et ce pour les trois voyelles. Ces résultats mettent en évidence le fait que les patients présentent une insuffisance laryngée liée à la chirurgie, caractérisée par un accolement partiel des plis vocaux. Ces conclusions vont dans le sens des études de Zeitels & al. (2002), Tamura & al. (2003) et Mirghani (2009) sur les cordectomies. Ces auteurs ont également mis en

6. Le coefficient de variation correspond au rapport de l'écart type à la moyenne. Ainsi, plus le pourcentage du coefficient de variation est élevé et plus la dispersion autour de la moyenne est importante.

évidence des valeurs du DAB plus élevées pour leurs groupes de patients — indépendamment du fait que ces derniers ne font pas tous partie d'un même type de chirurgie — par rapport à des témoins. Ces résultats portent majoritairement sur la voyelle tenue [a] en phonation confortable ou soutenue. Mirghani (2009) note par exemple que ses patients traités par cordectomie de type I présentent des valeurs qui restent, malgré tout, proches des valeurs extrêmes relevées chez les témoins. Celui-ci souligne que ce constat est principalement lié à la variabilité importante constatée pour la cohorte de témoins, ce que nous avons également pu relever dans nos données.

La recherche de Giovanni & al. (1995) a mis en évidence une forte corrélation positive entre les mesures de DAB et les calculs d'indices de fuites glottiques. Ainsi, plus la valeur de DAB est élevée et plus l'indice sera important et révélateur d'une fuite glottique. En conséquence, il semble normal que les hiérarchies observées précédemment concernant le paramètre du DAB — gardons en mémoire l'importante variabilité intra-individuelle — soient conservées pour nos indices de fuites glottiques. Ainsi, pour les témoins, nous relevons  $Fg[a] < Fg[i] < Fg[u]$ . Par contre, pour le groupe des patients, nous notons :  $Fg[i] < Fg[a] < Fg[u]$ .

Teston (2004) note que le calcul des indices de fuites glottiques est pertinent pour quantifier une voix pathologique. Il précise que la frontière supérieure de la normalité se situe autour de  $4cm^3/dB/s$  et la limite maximale considérée comme pathologique est autour de  $8cm^3/dB/s$ . Or, nos sujets n'ont jamais atteint le seuil supérieur de la normalité fixé par l'auteur. Selon nous, ce seuil se situe davantage autour de  $3cm^3/dB/s$ . Ainsi, nous constatons que les indices de fuites glottiques des patients sont toujours supérieurs à ce seuil supérieur de la normalité. De plus, nous avons montré qu'il existe bien une différence hautement significative des indices de fuites glottiques entre les témoins et les patients, ce qui implique une déperdition d'air pour ces derniers.

Ces résultats confirment que nos patients présentent une fuite glottique liée à une insuffisance laryngée, bien que cette fuite semble « minime » par rapport à ce que l'on pourrait attendre de la chirurgie. L'insuffisance laryngée est caractérisée par un accoulement peu efficace des plis vocaux qui entraîne une fermeture glottique incomplète et donc une déperdition d'air à travers l'espace glottique. Baken & Orlikoff (2000) ajoutent que l'insuffisance d'adduction des plis vocaux est à mettre en relation avec une constriction plus faible du conduit vocal, ce qui implique une baisse de la résistance des flux d'air et une hausse du DAB. Ce constat est directement imputable aux modifications anatomiques liées aux cordectomies. En effet, la chirurgie enlève la muqueuse, le ligament et une partie variable du muscle vocal, ce qui tend à affiner le pli vocal touché et à réduire sa masse vibrante. En conséquence, une asymétrie et une adduction variable des plis vocaux sont constatées, ce qui explique l'installation de fuites glottiques lors des productions vocaliques.

Par ailleurs, la présence d'une fuite glottique persistante liée à un défaut d'adduction des plis vocaux est une raison qui nous pousse à nous intéresser à l'implémentation du trait de voisement chez nos patients. En effet, un voisement nécessite un accoulement efficace des plis vocaux, ce que cette population ne peut maîtriser totalement.

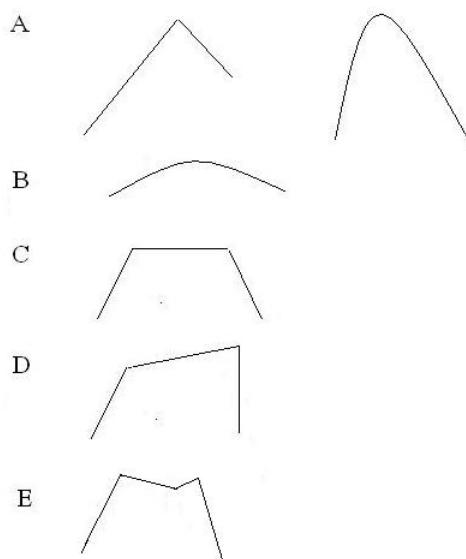
### 6.5.3 Etudes qualitatives des courbes de la PIO et du DAB des logatomes

Avant toute étude quantitative des paramètres aérodynamiques liés au trait de voisement, il nous a semblé intéressant d'étudier les différentes courbes obtenues sous Phonedit. Cette section cherche donc à décrire le plus fidèlement possible les courbes de PIO et de DAB dérivées des productions de différents logatomes. La méthodologie des conditions d'enregistrement sera exposée plus particulièrement dans la section relative à l'étude de l'implémentation du voisement d'un point de vue aérodynamique (cf. section 6.5.4). Nous pouvons, néanmoins, souligner que nos observations se basent sur 720 occurrences (soit 6 logatomes de type CVCVCV où  $C=[p, b]$  et  $V=[a, i, u] \times 3$  répétitions  $\times 20$  sujets  $\times 2$  positions). Notre choix s'est porté principalement sur les occlusives labiales car ces consonnes facilitent le recueil des données aérodynamiques. En effet, celles-ci se réalisent avec un pincement des lèvres durant l'occlusion, ce qui permet de recueillir plus facilement la PIO à l'aide d'une sonde placée entre les dents, ainsi que le DAB à la sortie de la bouche.

#### 6.5.3.1 Description des patterns de PIO observés

L'observation de l'ensemble de nos données nous a permis de mettre en évidence cinq *patterns* caractéristiques concernant les courbes de PIO. La figure 6.6 représente les modélisations de ces *patterns* et montre particulièrement la succession d'événements liée à ces courbes.

Premièrement, la PIO des occlusives labiales est caractérisée par une montée plus ou moins rapide de la courbe — le *pattern* B ne présente pas de montée aussi rapide que les autres — correspondant à la fermeture complète du conduit vocal. Deuxièmement, nous remarquons que cette montée peut être suivie (i) d'un pic maximalement haut (*patterns* A), (ii) d'un pic plus ovale (*pattern* B), (iii) d'un plateau stable (*pattern* C), (iv) d'une montée plus lente (*pattern* D) ou (v) d'un plateau montant-descendant (*pattern* E). Enfin, nous observons une diminution plus ou moins rapide de la PIO, correspondant au relâchement de la consonne.



**Figure 6.6 –** Modélisations des patterns de PIO retrouvés dans nos données pour les logatomes contenant les labiales [p] et [b].

Après avoir mis en évidence les *patterns* de PIO les plus caractéristiques de notre corpus, il nous a semblé intéressant d'observer leur répartition selon le groupe, la position syllabique et le type de consonne attendue. Le tableau 6.8 résume nos conclusions pour la position d'attaque en initiale absolue. Dans tous les cas, nous remarquons que le *pattern* A est privilégié, en particulier lorsque les logatomes sont composés d'une occlusive labiale non voisée (41,1% pour les témoins et 43,9% pour les patients). Ce *pattern* est également largement répertorié pour les débuts de mots débutant par leur homologue voisé (34,5% pour les témoins et 30,5 % pour les patients). Cependant, nous notons que le *pattern* A alterne avec le *pattern* B dans ce même cas de figure où le logatome est composé d'une occlusive voisée (15,5% pour les témoins et 15,6% pour les patients). Nous pouvons également souligner le fait que les patients utilisent les *patterns* C et D — quels que soient les consonnes — tandis que les témoins peuvent avoir recours au *pattern* E — lorsque le logatome contient un [p] — sans que cela soit réellement pertinent.

Patterns	Témoins			Patients		
	[p]	[b]	$\sum$	Patterns	[p]	[b]
A	41,1	34,5	75,6	A	43,9	30,5
B	6,7	15,5	22,2	B	2,2	15,6
C	0	0	0	C	2,2	1,7
D	0	0	0	D	1,7	2,2
E	2,2	0	2,2	E	0	0
$\sum$	50	50	100	$\sum$	50	50
						100

**Table 6.8** – Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction des patterns de PIO observés pour la position d’attaque **en initiale absolue (#\_)** ( $n = 360$ ).

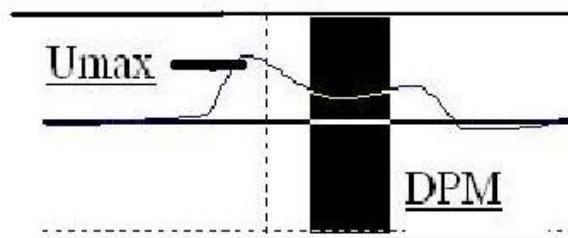
Enfin le tableau 6.9 synthétise nos observations pour la position d’attaque en intervocalique. Une fois encore, les patients et témoins semblent favoriser le *pattern A*. Cependant les comportements semblent différents par rapport au tableau précédent. Ainsi, nous pouvons constater une répartition plus hétérogène dans l’utilisation des différents *patterns* pour notre groupe de témoins. Ces derniers privilégié le premier *pattern* tant pour [p] (18,9%) que pour [b] (19,4%). Néanmoins, nous remarquons que tous les *patterns* décrits précédemment apparaissent dans une moindre mesure. Le *pattern B* est présent à 19,5%, avec une prédominance lorsque la consonne est voisée (16,7%). De même, le *pattern C* est retrouvé dans 20% des cas, en particulier quand la consonne est non voisée. Enfin, les *patterns D* et *E* se partagent les 22% restant du corpus. A l’inverse, les *patterns* des patients restent plus homogènes et semblent davantage en adéquation avec les comportements mis en évidence pour la position initiale. Ainsi, le *pattern A* est favorisé (60,6%) et plus particulièrement lorsque la consonne est non voisée (38,4%) tandis que pour les consonnes voisées, ce *pattern* alterne avec le *pattern B* (17,7%). Nous notons également que les *patterns C* et *D* totalisent les 19% restant du corpus. Néanmoins, contrairement aux témoins pour lesquels le *pattern E* est présent dans 12,2% des cas, les patients adoptent que très rarement une telle configuration de la PIO (0,6%).

Patterns	Témoins			Patients		
	[p]	[b]	$\sum$	Patterns	[p]	[b]
A	18,9	19,4	38,3	A	38,4	22,2
B	2,8	16,7	19,5	B	3,9	17,2
C	15	5	20	C	4,4	3,9
D	6,1	3,9	10	D	3,3	6,1
E	7,2	5	12,2	E	0	0,6
$\sum$	50	50	100	$\sum$	50	50
						100

**Table 6.9** – Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction des patterns de PIO observés pour la position d’attaque **en intervocalique (V\_V)** ( $n = 360$ ).

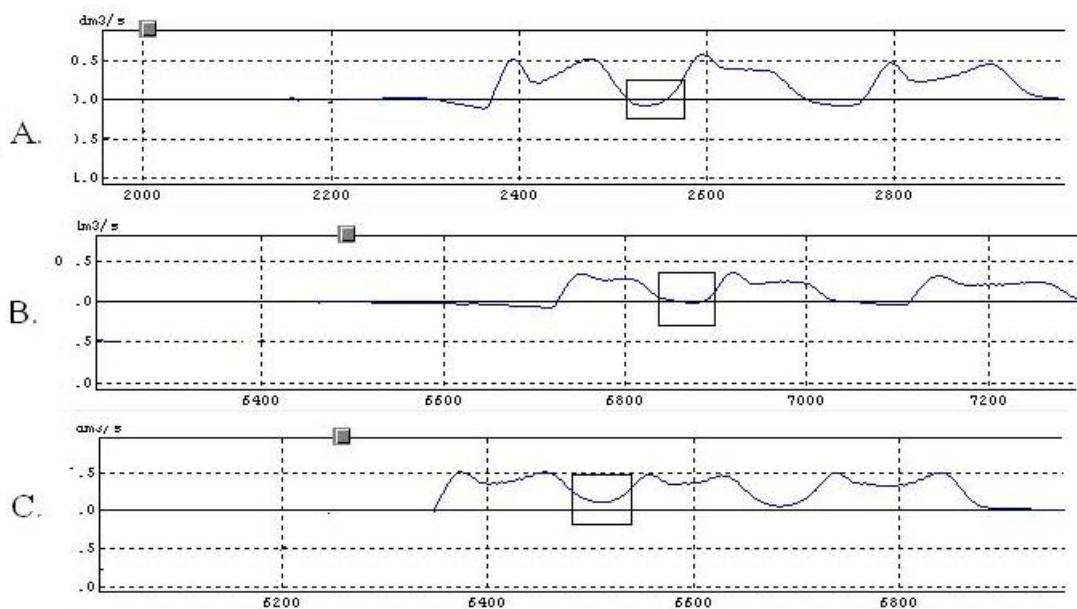
### 6.5.3.2 Observation des courbes générales du DAB des logatomes

Pour tous les logatomes du corpus, les courbes de DAB sont caractérisées par divers événements successifs : une montée rapide associée à la mise en place des articulateurs pour le relâchement de la consonne, un pic maximal ( $U_{max}$ ) qui représente l'ouverture brutale de l'occlusion et une descente progressive de cette courbe. Nous avons nommé Débit Phonatoire Moyen (DPM) la partie stable descendante de la courbe qui correspond à la partie stable de la voyelle pendant une fenêtre temporelle de 50 ms (cf. figure 6.7).



**Figure 6.7 –** Courbe du DAB d'une voyelle [i] où  $U_{max}$  correspond à la valeur du débit phonatoire maximal lié à l'ouverture de la consonne et DPM correspond à la valeur du débit d'air phonatoire moyen pendant une fenêtre temporelle de 50 ms, soit la portion la plus stable de la voyelle.

En outre, nous avons observé le comportement de ces courbes de DAB pendant la réalisation des logatomes sur l'ensemble de notre corpus. Théoriquement, après le relâchement de la consonne, il est attendu que le DAB diminue pour atteindre une valeur nulle. Cependant, l'examen de nos données met en évidence trois comportements distincts de la courbe : (i) le DAB rejoint sa ligne de base et atteint une valeur égale à 0, (ii) le DAB descend au-delà de cette ligne de base et devient négatif ou (iii) le DAB n'atteint pas cette même ligne et reste dans des valeurs positives. La figure 6.8 donne des exemples de ces différentes configurations.



**Figure 6.8** – Exemples de courbes de DAB pour le logatome [papapap] (témoin n°10). La courbe A représente un évènement où le débit d'air phonatoire est négatif (2nde répétition), la courbe B montre une courbe du DAB qui atteint sa ligne de base (1ère répétition) et la courbe C indique un débit d'air positif (répétition non prise en compte).

Enfin, après cette analyse qualitative du comportement des courbes de DAB, il nous a semblé intéressant d'observer leur répartition selon les groupes et les consonnes produites. Dans cette étude, les positions syllabiques ne semblent pas pertinentes car les événements nuls, positifs ou négatifs tendent à se reproduire sur l'ensemble du logatome. Le tableau 6.10 résume nos calculs. Il apparaît que les deux groupes présentent majoritairement un *pattern* de DAB attendu, soit un retour nul à la ligne de base (65% pour les témoins et 52,3% pour les patients). Le second cas de figure représente des courbes où les valeurs de DAB sont négatives (26,7% pour les témoins et 34,4% des patients). Les témoins montrent une répartition assez équivalente entre [p] (13,9%) et [b] (12,8%). Par contre, ces courbes sont davantage prédominantes pour les logatomes comportant une occlusive non voisée pour notre groupe de patients (23,3%). D'un point de vue articulatoire, l'interprétation de ces configurations reste assez difficile. Enfin, les événements où le DAB reste positif sont plus rares et sont souvent liés à la présence de l'occlusive labiale voisée. Il est toutefois important de noter que ces cas de figure sont davantage présents chez les patients (13,3%) par rapport aux sujets témoins (8,3%). En effet, la présence d'un débit positif est à mettre en rapport avec une ouverture faible de la glotte ou à un hypofonctionnement glottique. Ainsi, ce débit positif témoigne d'une baisse de la pression transglottique et est donc souvent caractéristique des voix chuchotées ou *breathy*. Finalement, ces débits phonatoires positifs peuvent s'expliquer soit par un matériel défaillant où le masque bouge pendant l'enregistrement, ce qui modifie la sortie de l'air, soit par une déperdition d'air liée à un

hypofonctionnement glottique — en particulier pour les patients —, soit par un phénomène de spirantisation spontanée des occlusives qui induit une fermeture incomplète du conduit vocal ([ɸ] et [β]).

<i>Témoins</i>				<i>Patients</i>			
Patterns	[p]	[b]	Σ	Patterns	[p]	[b]	Σ
DAB nul	33,3	31,7	65	DAB nul	22,8	29,5	52,3
DAB négatif	13,9	12,8	26,7	DAB négatif	23,3	11,1	34,4
DAB positif	2,8	5,5	8,3	DAB positif	3,9	9,4	13,3
Σ	50	50	100	Σ	50	50	100

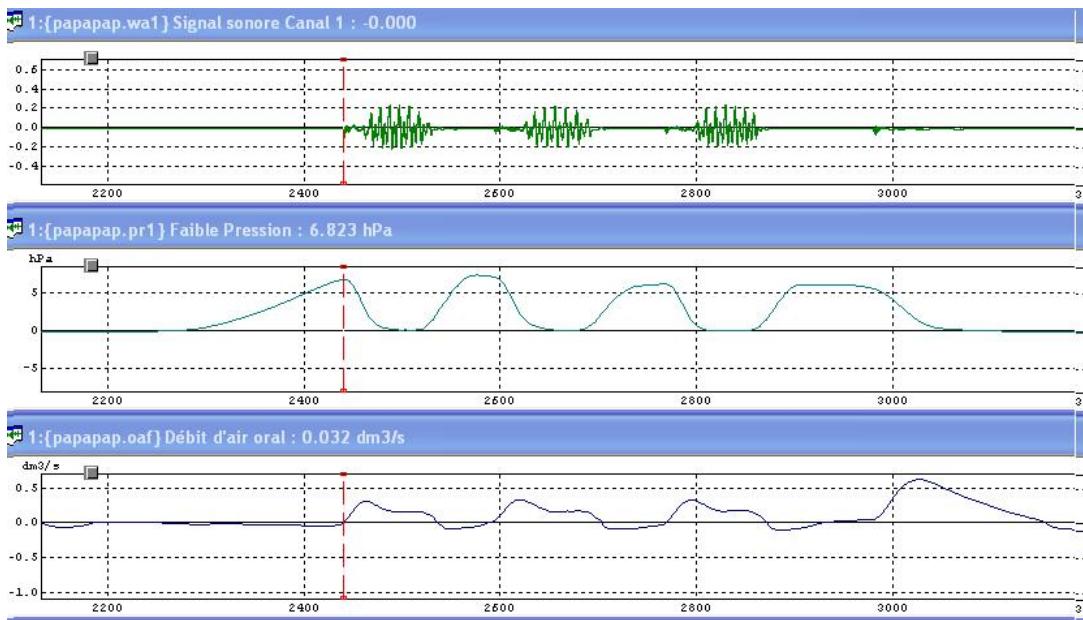
**Table 6.10** – Répartitions (en %) des différentes courbes de DAB des logatomes contenant les labiales [p] et [b] selon qu’elles passent par le point zéro, qu’elles passent sous la ligne de base (valeurs négatives du DAB) ou qu’elles restent au-dessus de cette ligne de base (valeurs positives du DAB) ( $n = 720$ ).

### 6.5.3.3 Synchronisation des courbes de PIO et de DAB

Comme nous l’avons mentionné précédemment, les occlusives sont caractérisées par trois évènements distincts — occlusion, tenue et relâchement — marqués par des phases aérodynamiques spécifiques. Idéalement, nous avons noté que le relâchement de la consonne est repérable à la synchronisation entre la chute de la courbe de la PIO et la hausse concomitante de la courbe du DAB. Cependant, la réalité peut être toute autre. Ainsi, comme nous l’avons déjà discuté, les productions intra-locuteurs et les différentes répétitions restent variables, ce qui tend à désorganiser la succession de ces évènements. De ce fait, il nous a paru intéressant de mettre en évidence les différentes configurations de l’intervalle entre l’évolution des deux courbes afin d’illustrer la manière dont ces évènements peuvent se coordonner. Nos données montrent trois types d’exemples : (i) la PIO et le DAB sont synchronisés, (ii) le DAB commence à augmenter avant que la PIO ne diminue et (iii) la PIO commence sa diminution avant que le DAB n’augmente.

**Synchronisation des courbes.** Sur les 180 productions des patients et les 180 productions des témoins, nous avons compté le nombre d’occlusives — tout contexte vocalique confondu — dont la coordination entre PIO et DAB est parfaite au moment du relâchement pour chacune des deux positions étudiées. La figure 6.9 illustre ce premier cas de figure.

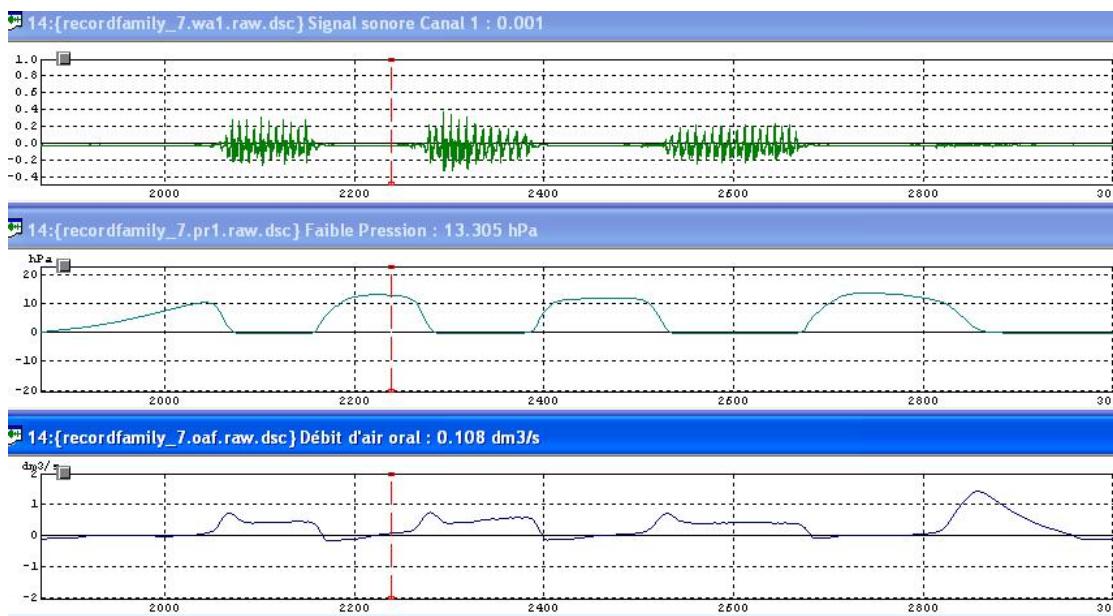
Pour les deux populations et les deux positions (#\_\_ et V\_\_V), nous constatons que les pourcentages d’occlusives qui présentent une synchronisation parfaite sont maximums (entre 76,7% et 78,9%). De même, nous pouvons noter que pour les deux cohortes et pour les deux positions, les pourcentages sont légèrement supérieurs pour l’occlusive voisée [b]. La synchronisation semble donc plus aisée pour les occlusives voisées par rapport aux occlusives non voisées. Les tableaux 6.11 et 6.12 résument ces observations.



**Figure 6.9** – Exemple d'un logatome [pipipip] où les événements de PIO et DAB sont synchronisés pour la position 1 (témoin n°2).

**Désynchronisation des courbes : le DAB augmente avant la diminution de la PIO.** Comme énoncé plus haut, nous avons également compté le nombre d'occlusives — hors contextes vocaliques — qui présentent une désynchronisation des deux événements aérodynamiques, caractérisée par une montée du DAB antérieure à la diminution de la PIO pour les deux populations et pour les deux positions syllabiques. Ce second cas de figure est représenté dans la figure 6.10.

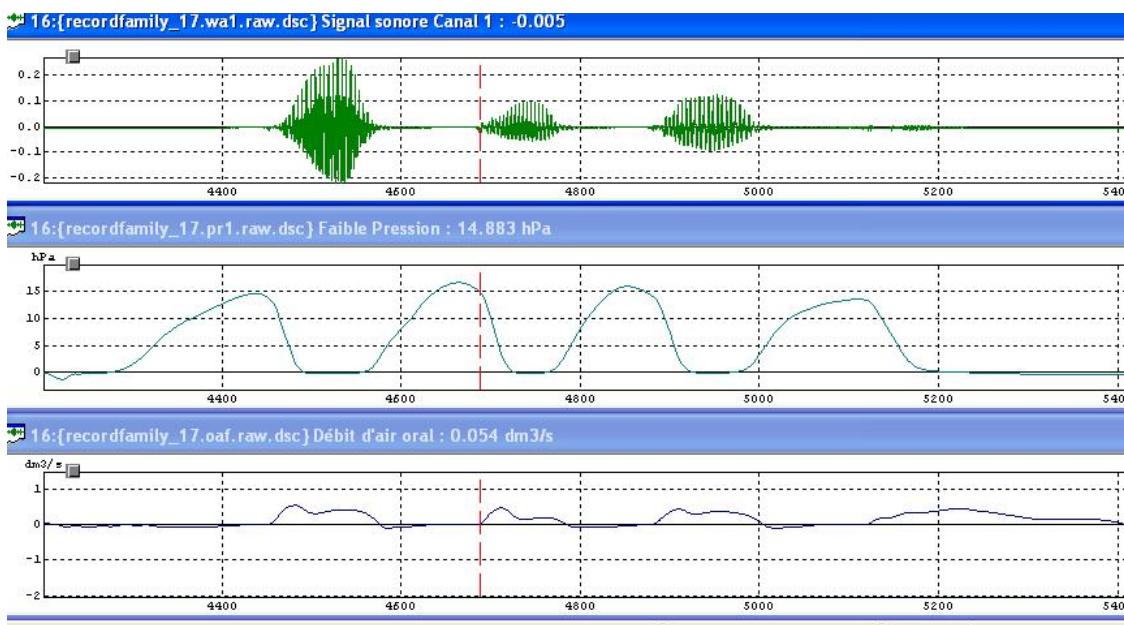
Nous observons des pourcentages assez similaires pour les deux populations, et ce quelle que soit la position. Cette désynchronisation ne semble donc pas être une tendance spécifique à l'une ou l'autre de nos cohortes. Aussi, nous pouvons noter que pour l'ensemble des sujets et pour les deux positions, les pourcentages sont toujours légèrement supérieurs pour l'occlusive voisée [b] par rapport à sa contrepartie non voisée [p]. Les tableaux 6.11 et 6.12 synthétisent ces conclusions.



**Figure 6.10** – Exemple d'un logatome [papapap] où les évènements de PIO et DAB sont désynchronisés, à savoir que la courbe de DAB commence à augmenter avant la diminution de la PIO pour la position 2 (patient n°10).

**Désynchronisation des courbes : la PIO diminue avant l'augmentation du DAB.** Enfin, nous avons également compté le nombre d'occurrences — toujours en dehors de tout contexte vocalique — qui montrent une désynchronisation des gestes aérodynamiques lors du relâchement des occlusives. Ce dernier type d'interaction est déterminé par une diminution de la PIO qui survient avant la hausse du DAB pour les deux groupes et pour les deux positions syllabiques. Ce troisième évènement est représenté par la figure 6.11.

Nous remarquons que les pourcentages sont relativement bas mais restent assez semblables d'une population à l'autre. Toutefois, la position 2 semble être la plus propice à cette configuration et en particulier lorsque le logatome est composé d'occlusives labiales non voisées. Les tableaux 6.11 et 6.12 présentent ces observations.



**Figure 6.11** – Exemple d'un logatome [pupupup] où les évènements de PIO et DAB sont désynchronisés, à savoir que la courbe de PIO commence à descendre avant le début de l'augmentation de DAB pour la position 2 (patient n°2).

Nous avons donc pu mettre en évidence trois comportements distincts pour la coordination des courbes de DAB et PIO lors du relâchement des occlusives. Les tableaux ci-dessous montrent la répartition des logatomes selon que ces évènements soient synchronisés ou non en fonction de la position syllabique. De manière générale, nous notons une excellente tendance à la synchronisation pour les deux populations. Par contre, lorsque les évènements aérodynamiques ne sont pas synchronisés, il est plus fréquent que le DAB commence à augmenter avant que la PIO ne commence à baisser plutôt que le contraire.

Patterns	<i>Témoins</i>			<i>Patients</i>		
	[p]	[b]	$\sum$	[p]	[b]	$\sum$
S	37,8	39,5	77,3	38,9	40	78,9
NS1	10	9,4	19,4	7,2	8,3	15,5
NS2	2,2	1,1	3,3	3,9	1,7	5,6
$\sum$	50	50	100	50	50	100

**Table 6.11** – Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction de la synchronisation des évènements aérodynamiques pour la position d'attaque **en initiale absolue (#\_\_)** (où S représente une bonne synchronisation entre les courbes de DAB et PIO, NS1 représente une désynchronisation des courbes où le DAB commence à monter avant que la PIO diminue et NS2 représente une désynchronisation des courbes où la PIO commence à diminuer avant que le DAB n'augmente) ( $n = 360$ ).

Témoins				Patients			
Patterns	[p]	[b]	Σ	Patterns	[p]	[b]	Σ
S	40	38,3	78,3	S	37,2	39,4	76,6
NS1	5,6	7,8	13,4	NS1	5,6	10,6	16,2
NS2	4,4	3,9	8,3	NS2	7,2	0	7,2
Σ	50	50	100	Σ	50	50	100

**Table 6.12** – Répartitions (en %) des différents logatomes contenant les labiales [p] et [b] en fonction de la synchronisation des événements aérodynamiques pour la position d'attaque **en intervocalique** (V—V) (où S représente une bonne synchronisation entre les courbes de DAB et PIO, NS1 représente une désynchronisation des courbes où le DAB commence à monter avant que la PIO diminue et NS2 représente une désynchronisation des courbes où la PIO commence à diminuer avant que le DAB n'augmente) ( $n = 360$ ).

## 6.5.4 Etude du trait de voisement des occlusives en contexte

### 6.5.4.1 Méthodologie

Nos mesures aérodynamiques ont été acquises à l'aide de l'appareillage EVA2 qui permet d'enregistrer simultanément des paramètres acoustiques classiques tels que la  $F_0$  et l'intensité ainsi que des paramètres aérodynamiques tels que le DAB et la PIO. Notre protocole se compose de logatomes de type CVCVCVC où C représente les consonnes du français et V= [a, i, u]. Pour les mesures de PIO dans l'analyse du voisement, nous avons restreint nos mesures aux logatomes composés d'occlusives labiales, soit [p, b]. Les sujets devaient répéter trois fois ces logatomes à hauteur et intensité confortables, en essayant de conserver une intensité *constante*. Comme nous l'avons déjà expliqué précédemment, les logatomes ont été divisés en trois syllabes distinctes : CV<sub>1</sub>, CV<sub>2</sub> et CV<sub>3</sub> afin de faciliter le traitement de nos données. Afin de conserver une cohérence avec les analyses acoustiques, nous avons choisi de nous intéresser aux positions 1 et 2, soit les positions d'attaque en initiale absolue (#—) et d'attaque en intervocalique (V—V). Nous avons pu analyser 720 occurrences, soit 2 consonnes × 3 voyelles × 3 répétitions × 2 positions × 20 sujets.

Les locuteurs devaient maintenir fermement le pied de l'appareillage afin d'éviter les déperditions d'air latérales et garder la sonde intra-buccale entre les lèvres en évitant de la mordre. Plusieurs essais étaient réalisés par les sujets afin qu'ils se familiarisent avec le matériel et que l'expérimentateur soit certain de leur compréhension de la tâche. De plus, un contrôle visuel pendant les sessions d'enregistrement nous a permis de valider les répétitions où les pics de PIO étaient corrects. Dans les cas où ceux-ci apparaissaient comme inexistant, les sujets devaient reprendre les logatomes afin d'obtenir trois répétitions exploitables. Les sujets pouvaient respirer normalement par le nez et chaque session d'enregistrement se concluait par une pause.

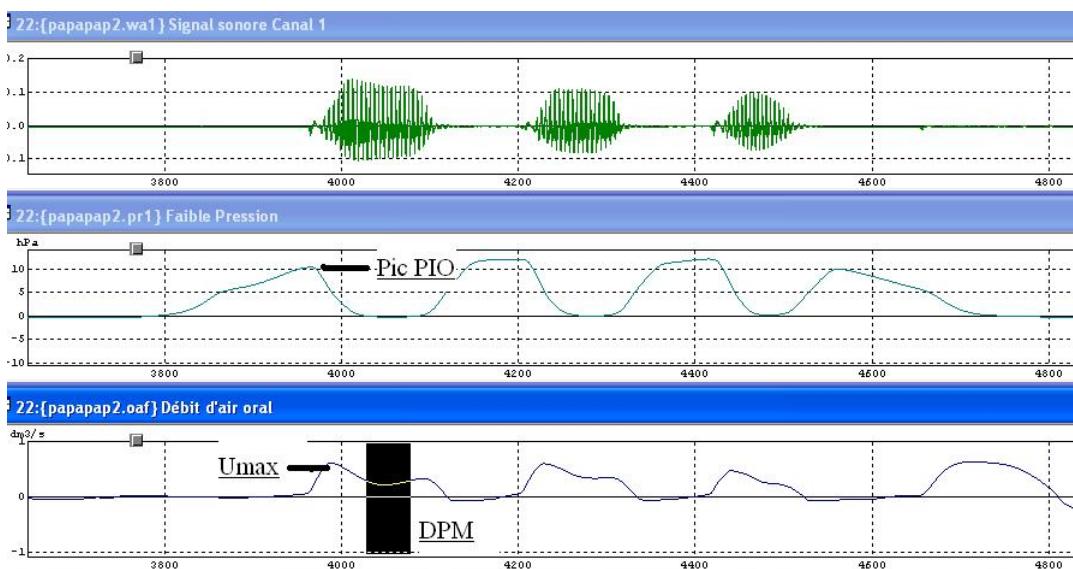
Le traitement des résultats a été effectué à l'aide du logiciel Phonetedit. Pour chaque sujet, nous pouvons extraire les courbes d'intensité et de DAB pendant les voyelles et les courbes de PIO durant la production des occlusives labiales. Les différentes

mesures dont nous avons eu besoin ont été réalisées manuellement à l'aide d'un curseur mobile.

Nous avons donc pu relever les pics maximums de la PIO sur les occlusives [p] et [b]. En ce qui concerne le DAB, nous avons réalisé deux mesures différentes : (i) le pic (Umax) pendant le relâchement de la consonne et (ii) le débit phonatoire moyen (DPM) sur la partie la plus stable de la voyelle pendant une fenêtre temporelle de 50 ms afin d'éviter les variations intra-individuelles au début et à la fin de ces segments (cf. figure 6.12). Une moyenne des mesures des trois répétitions a été calculée pour chaque paramètre étudié et pour tous les sujets. Finalement, nos analyses statistiques ont porté sur 240 occurrences pour chaque paramètre, soit une moyenne des mesures  $\times$  2 consonnes  $\times$  3 voyelles  $\times$  2 positions  $\times$  20 sujets.

Ces différentes mesures nous ont donc permis d'évaluer l'implémentation du trait de voisement en fonction du groupe, des contextes vocaliques et des positions syllabiques. Nous avons donc conduit des ANOVA à 4 facteurs où les variables dépendantes étaient la PIO, le DAB maximal (Umax) et le DPM et les variables indépendantes représentaient le groupe, le type de voyelle, le contexte de voisement et la position dans les logatomes à l'aide du logiciel Statview. Les résultats étaient considérés comme significatifs à partir d'un seuil alpha de 5%, soit  $p < .05$ .

De plus, nous avons estimé la pression transglottique nécessaire au voisement en établissant un calcul de la différence des pics moyens de PIO de [p] et des pics moyens de PIO de [b] à intensité identique.

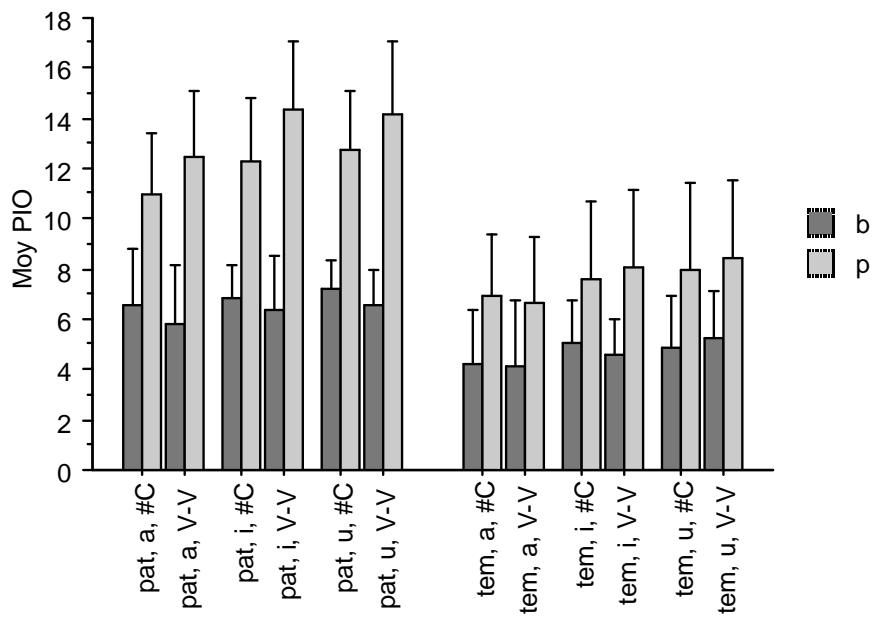


**Figure 6.12** – Courbes obtenues pour le logatome [pipipip] sur Phonedit (Exemple du témoin n°1). En haut, l'oscillogramme ; au milieu, la PIO (en hPa) où le pic de PIO correspond à la tenue de l'occlusive ; en bas, le DAB en  $dm^3/s$  où Umax représente la valeur du débit d'air maximal lié au relâchement de la consonne et DPM représente la valeur du débit d'air moyen pendant une fenêtre temporelle de 50 ms (portion stable de la voyelle [i]).

### 6.5.4.2 Présentation des résultats

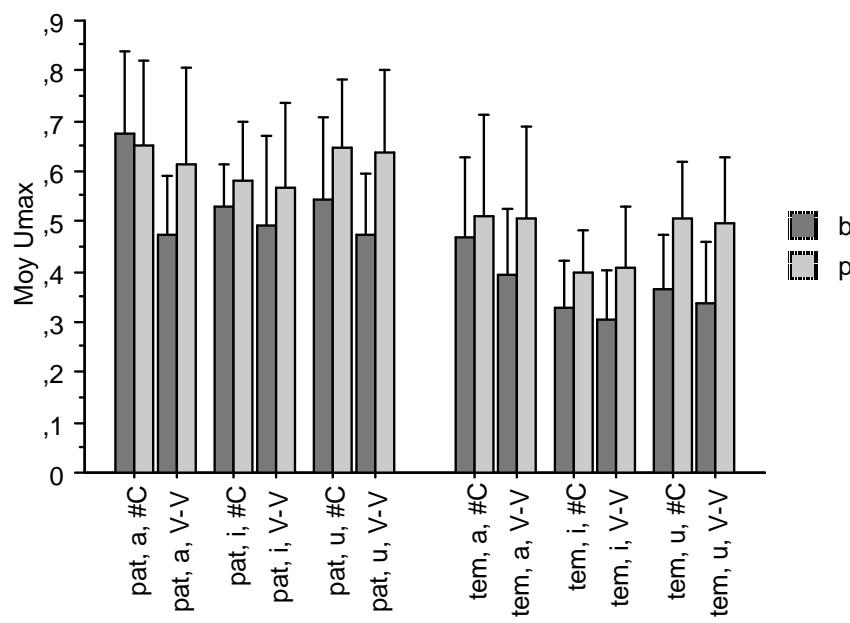
**Valeurs moyennes de la PIO.** La figure 6.13 présente les valeurs moyennes de PIO en fonction des deux populations, des contextes vocaliques, de la différence de voisement entre [p] et [b] et de la position dans le mot. Nous constatons que quel que soit le voisement de la consonne, les valeurs des patients sont toujours supérieures à celles des témoins. Le test ANOVA montre une différence hautement significative entre les deux groupes ( $F(1, 216) = 132,731, p < .0001$ ). En outre, nous remarquons que pour les deux groupes, les valeurs de PIO sont plus élevées lorsque la consonne est une occlusive labiale non voisée ( $F(1, 216) = 222,684, p < .0001$ ). Aussi, les contextes vocaliques semblent avoir une influence sur les valeurs moyennes de PIO de nos sujets ( $F(2, 216) = 5,46, p = .0049$ ). Les valeurs de PIO des occlusives tendent à être plus hautes lorsque celles-ci sont entourées de voyelles hautes et plus basses au contact d'une voyelle basse. Ainsi, le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [a, u], avec des p-value respectivement égales à  $p = .0152$  et  $p = .0019$ . Néanmoins, bien que nous remarquions que la PIO tende à être plus élevée en position intervocalique (V\_V), la position dans les logatomes ne présente finalement pas d'effet significatif sur les valeurs de PIO ( $F(1, 216) = 0,902, p = .3434$ ).

En outre, nous pouvons noter que deux interactions sur l'ensemble de nos variables indépendantes présentent un seuil de significativité inférieur à 5%. Premièrement, l'interaction entre les facteurs « sujets » et « voisement » est significative ( $F(1, 216) = 29,267, p < .0001$ ), c'est à dire que l'un des facteurs présentés varie en fonction des modalités du second facteur. Deuxièmement, l'interaction entre les facteurs « voisement » et « position » est également significative ( $F(1, 216) = 4,338, p = .0385$ ).



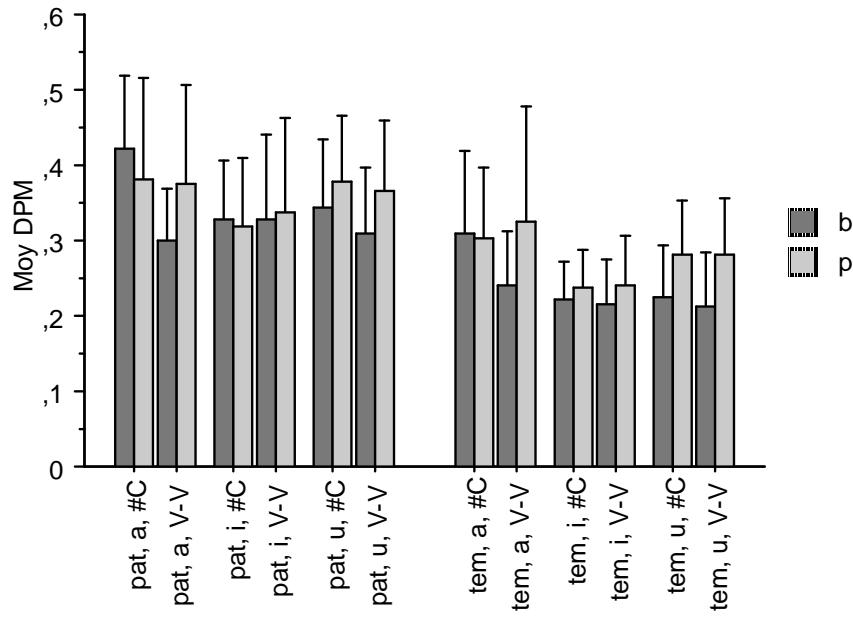
**Figure 6.13** – Valeurs moyennes de la PIO (en hPa) en fonction des sujets, du type de voyelle, du voisement et de la position dans le logatome (les barres d'erreurs représentent les écarts-types) ( $n = 240$ ).

**Valeurs moyennes de Umax.** La figure 6.14 présente les valeurs moyennes des pics de DAB (soit Umax) en fonction des deux populations, des contextes vocaliques, de la différence de voisement entre [p] et [b] et de la position dans le mot. Nous remarquons que quel que soit le voisement de la consonne, les valeurs de Umax des patients sont toujours supérieures à celles de notre cohorte de témoins ( $F(1, 216) = 71,363, p < .0001$ ). Nous constatons également pour les deux groupes, que les valeurs moyennes des pics de DAB des voyelles ont une amplitude plus importante en contexte non voisé par rapport aux valeurs moyennes en contexte voisé. L'ANOVA présente une différence hautement significative quant à l'effet du voisement sur les valeurs de Umax ( $F(1, 216) = 26,426, p < .0001$ ). De la même manière, le contexte vocalique semble avoir une influence sur les valeurs moyennes Umax de nos sujets ( $F(2, 216) = 7,25, p = .0009$ ). En effet, les valeurs de Umax des voyelles semblent être plus élevées pour [a] par rapport aux voyelles hautes. Ainsi, le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [i, u], avec des p-value respectivement égales à  $p = .0002$  et  $p = .0267$ . En outre, et contrairement à la PIO, la position dans le logatome semble avoir une influence sur les valeurs moyennes de Umax ( $F(1, 216) = 5,343, p = .0217$ ). Les moyennes de Umax semblent être plus élevées à l'initial absolu. Néanmoins, pour cette analyse, nous ne retrouvons aucune interaction entre les différents facteurs.



**Figure 6.14** – Valeurs moyennes de Umax — pic du DAB — (en  $dm^3/s$ ) en fonction des sujets, du type de voyelle, du voisement et de la position dans le logatome (les barres d'erreurs représentent les écarts-types) ( $n = 240$ ).

**Valeurs moyennes de DPM.** La figure 6.15 présente les valeurs moyennes des débits phonatoires moyens (DPM) en fonction des deux populations, des contextes vocaliques, de la différence de voisement entre [p] et [b] et de la position dans le mot. Encore une fois, nous pouvons remarquer que quel que soit le voisement la consonne, les valeurs de DPM des patients sont toujours supérieures à celles de notre cohorte de témoins ( $F(1, 216) = 57,819, p < .0001$ ). Par ailleurs, les valeurs du DPM restent dépendantes du contexte de voisement. Ainsi, comme pour les valeurs de Umax, les valeurs moyennes du DPM tendent à être plus élevées en contexte non voisé par rapport au contexte voisé ( $F(1, 216) = 6,731, p = .0101$ ). Nous notons également un effet significatif des contextes vocaliques sur les valeurs moyennes de DPM ( $F(2, 216) = 6,538, p = .0018$ ). Comme pour Umax, les valeurs de DPM semblent être plus importantes pour [a] par rapport aux voyelles hautes. En effet, le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [a, u], avec des p-value respectivement égales à  $p = .0004$  et  $p = .0297$ . Enfin, la position dans le logatome ne semble pas avoir d'influence sur les valeurs moyennes du DPM ( $F(1, 216) = 2,212, p = .1384$ ). Nous pouvons noter, malgré tout, que les valeurs semblent suivre les résultats exposés pour les pics maximaux de DAB. Effectivement, la position initiale absolue semble favoriser l'élévation du DPM. Néanmoins, pour cette analyse, nous ne retrouvons aucune interaction entre les différents facteurs.



**Figure 6.15** – Valeurs moyennes du DPM — relevées sur une fenêtre temporelle de 50 ms, représentant la portion stable de la voyelle — (en  $dm^3/s$ ) en fonction des sujets, du type de voyelle, du voisement et de la position dans le logatome (les barres d'erreurs représentent les écarts-types) ( $n = 240$ ).

Le tableau 6.13 résume les principaux résultats des tests ANOVA pour les paramètres de la PIO, de Umax et du DPM :

		PIO	Umax	DPM
Groupe	patients > témoins	**	**	**
Type de voyelle	[a] > [i,u] (excepté pour PIO)	*	**	*
Contexte	[–voix] > [+voix]	**	**	**
Position	[V—V] > [#—] (excepté pour PIO)	ns	*	ns

**Table 6.13** – Tableau récapitulatif des différents tests ANOVA pour les paramètres aérodynamiques étudiés (où « \*\* » note des différences hautement significatives ( $p < .0001$ ), « \* » note des différences significatives ( $p < .05$ ), « ns » signifie « non significatif » et « na » signifie « non applicable »).

**Estimation de la pression transglottique.** Comme nous l'avons mentionné précédemment, la différence des pics de PIO lors de la production des occlusives labiales [p] et [b] à une même intensité permet une estimation de la pression transglottique nécessaire au voisement. Nous avons donc calculé ces différences de PIO pour chacun

des contextes vocaliques ( $\text{PIO}/p - b/$ ) et également effectué les rapports de ces différences sur la PIO de [p], soit  $(\text{PIO}/p - b/)/(\text{PIO}/p/)$ . Les tableaux 6.14, 6.15 et 6.16 résument nos résultats :

	<b>PIO/pa/</b>	<b>PIO/ba/</b>	<b>P/pa-ba/</b>	<b>(P/pa-ba/)/(P/pa/)</b>	<b>PIO/apa/</b>	<b>PIO/aba/</b>	<b>P/apa-aba/</b>	<b>(P/apa-aba/)/(P/apa/)</b>
<b>TEM</b>	6,911	4,223	2,688	0,39	6,69	4,123	2,567	0,38
<b>PAT</b>	11,01	6,572	4,438	0,40	12,469	5,821	6,648	0,53

**Table 6.14** – Valeurs moyennes de la PIO (en hPa) et calculs de la pression transglottique estimée (en hPa) selon les deux positions syllabiques (attaque en initiale absolue et attaque en intervocalique) pour les témoins et les patients en contexte vocalique [a] (où TEM= témoins et PAT= patients) ( $n = 80$ ).

	<b>PIO/pi/</b>	<b>PIO/bi/</b>	<b>P/pi-bi/</b>	<b>(P/pi-bi/)/(P/pi/)</b>	<b>PIO/ipi/</b>	<b>PIO/ibi/</b>	<b>P/ibi-ibi/</b>	<b>(P/ibi-ibi/)/(P/ibi/)</b>
<b>TEM</b>	7,626	5,072	2,554	0,33	8,085	4,574	3,511	0,43
<b>PAT</b>	12,32	6,825	5,495	0,45	14,311	6,383	7,928	0,55

**Table 6.15** – Valeurs moyennes de la PIO (en hPa) et calculs de la pression transglottique estimée (en hPa) selon les deux positions syllabiques (attaque en initiale absolue et attaque en intervocalique) pour les témoins et les patients en contexte vocalique [i] (où TEM= témoins et PAT= patients) ( $n = 80$ ).

	<b>PIO/pu/</b>	<b>PIO/bu/</b>	<b>P/pu-bu/</b>	<b>(P/pu-bu/)/(P/pu/)</b>	<b>PIO/upu/</b>	<b>PIO/ubu/</b>	<b>P/upu-ubu/</b>	<b>(P/upu-ubu/)/(P/upu/)</b>
<b>TEM</b>	7,975	4,917	3,058	0,38	8,441	5,25	3,191	0,38
<b>PAT</b>	12,719	7,24	5,479	0,43	14,166	6,605	7,561	0,53

**Table 6.16** – Valeurs moyennes de la PIO (en hPa) et calculs de la pression transglottique estimée (en hPa) selon les deux positions syllabiques (attaque en initiale absolue et attaque en intervocalique) pour les témoins et les patients en contexte vocalique [u] (où TEM= témoins et PAT= patients) ( $n = 80$ ).

### 6.5.4.3 Discussion

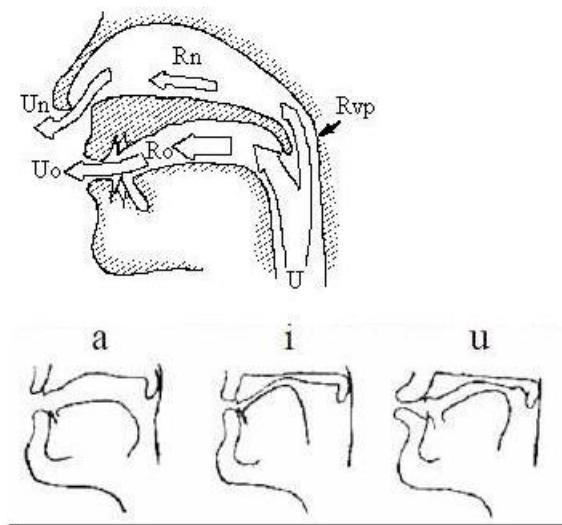
**Valeurs moyennes de PIO.** Nous avons pu mettre en évidence que les valeurs moyennes de PIO étaient principalement dépendantes du contexte de voisement, des voyelles environnantes et des deux populations. Bien que nous n'argumenterons pas les propos suivants, il est toutefois important de savoir que d'autres paramètres sont susceptibles de faire varier les mesures de PIO. Certains peuvent être contrôlés, comme l'intensité ou la hauteur (Isshiki 1964, Isshiki 1965, Vaughn 1965), et d'autres restent

dépendants des locuteurs. Parmi ceux-ci, citons la capacité vitale, les dimensions du conduit vocal (Fant 1960) ou encore la largeur buccale (Hixon 1966).

Nous avons montré que les valeurs de PIO, quels que soient les contextes vocaux pour les deux groupes de témoins et patients, étaient statistiquement plus élevées pour les occlusives non voisées. Ces résultats sont en adéquation totale avec les études de Subtelny & al. (1966), Arkebauer & al. (1967), Lisker (1970), Miller & Daniloff (1977), Stathopoulos (1986), Holmberg & al. (1988), entre autres. Nous avons déjà évoqué le fait que pendant la production des occlusives non voisées la glotte reste ouverte, ce qui permet le maintien du différentiel entre PSG et PIO. Au contraire, la réalisation des occlusives voisées nécessite deux gestes antagonistes : une occlusion dans le conduit vocal associée à une vibration entretenue des plis vocaux. La fermeture glottique est responsable d'une résistance au passage de l'air, ce qui tend à diminuer la PIO dans le tractus vocal par rapport à la PSG. Néanmoins, le voisement est principalement lié à un flux d'air transglottique caractérisé par la préservation d'une différence de pression entre la PSG et la PIO. Ce différentiel est généralement conservé par des principes de compensations articulatoires plus ou moins actifs. Ainsi, les différences de PIO moyennes observées entre [p] et [b] sont principalement dues aux volumes de la cavité supraglottique. Nous retenons qu'une expansion active, caractérisée par l'augmentation du volume de la cavité supraglottique liée à l'élargissement des parois, permet une augmentation de la PIO dans le tractus et par conséquent une perdurance d'air pour la production des occlusives voisées. En outre, la détente du conduit vocal, associée à un léger gonflement des tissus, reste favorable au maintien du voisement. Aussi, ces différences de PIO moyennes peuvent s'expliquer par la largeur de la glotte, qui s'amenuise pendant la production de [b].

Nous avons également noté que les valeurs moyennes de PIO de nos deux groupes tendent à être plus élevées en contexte de voyelles hautes. Nos résultats sont en adéquation avec les études de Karnell & Willis (1982), Klich (1982), Netsell & al. (1991) et Higgins & al. (1998). Karnell & Willis (1982) avaient déjà conclu que les valeurs de PIO du [p] étaient plus importantes quand la consonne était entourée de la voyelle postérieure et haute [u] par rapport à la voyelle basse [a]. Klich (1982) a retrouvé cette même dichotomie en voix chuchotée. Enfin, dans une étude sur 11 hommes et 10 femmes portant sur divers paramètres acoustiques et aérodynamiques — qui inclut la pression sous-glottique estimée (PSGE) — Higgins & al. (1998) ont également montré l'effet du contexte vocalique sur les mesures de PSGE à partir des mesures de PIO des occlusives. Les valeurs moyennes de PIO des occlusives semblent également être plus élevées devant [i] que devant [a] et ce, quelles que soient les tâches de parole. Karnell & Willis (1982) avaient émis deux hypothèses pour justifier ces changements : (i) une plus forte impédance de la cavité orale liée à la constriction des voyelles hautes et (ii) une tension plus importante des plis vocaux pour ces voyelles qui nécessite une augmentation de la pression afin d'entretenir la vibration. Néanmoins, le fait que cet effet se retrouve à la fois en parole modale et en voix chuchotée suppose que la vibration des plis vocaux n'est pas dépendante de ce constat. Finalement, l'élévation de la PIO au contact des voyelles hautes par rapport aux voyelles basses est uniquement due aux différences d'impédance de la cavité orale qui existe entre ces deux types de

voyelles. Ainsi, pour la production des voyelles hautes, le conduit vocal est plus tendu, ce qui a pour effet d'augmenter l'impédance du conduit vocal (Higgins & al. 1998). De plus, les différentes articulations pour la production des voyelles ont pour effet de modifier la taille et les volumes des cavités orales et pharyngales, ce qui implique des différences en terme de résistance ou d'impédance dans le tractus vocal. Le premier schéma de la figure 6.16 présente la mise en connexion des différentes cavités. L'air total mis à disposition et venant du bas du pharynx (U) est susceptible de rencontrer trois types de résistances : une résistance au niveau du voile du palais (R<sub>vp</sub>), une résistance dans la cavité nasale (R<sub>n</sub>) et une résistance dans la cavité orale (R<sub>o</sub>). Dans notre cas, les voyelles [a, i, u] sont orales et seule la résistance dans la cavité orale est pertinente. Il est évident que la résistance dans la cavité orale (R<sub>o</sub>) est dépendante de l'articulation des sons et plus précisément du degré de constriction des segments. La voyelle [i] est produite avec une constriction à l'avant du conduit vocal (environ 13 cm de la glotte). L'aperture de la mâchoire est minimale tandis que la langue effectue une forte constriction dans la région palatale. De ce fait, nous constatons l'existence d'une grande cavité pharyngale en opposition avec une cavité orale réduite. A l'instar de la voyelle [i], la production de la voyelle [u] nécessite une fermeture de la mâchoire. Par contre, la constriction vélaire (située à environ 7 cm de la glotte) a pour conséquence de rétrécir la cavité pharyngale. Aussi, une seconde constriction labiale est un second argument à l'augmentation de l'impédance. En opposition, la voyelle [a] nécessite une constriction dans la région pharyngale (environ 8 cm de la glotte) mais la langue est totalement abaissée. Nous ne notons donc pas de constriction linguo-palatale pour ce type de voyelle et l'aperture de la mâchoire est maximale, ce qui a pour conséquence de rétrécir la cavité pharyngale. Finalement, l'abaissement de la langue associé à l'importante aperture pour cette voyelle sont des arguments en faveur d'un passage libre de l'air dans la cavité orale (cf. figure 6.16). En définitive, nous pouvons considérer que l'impédance dans la cavité orale est directement liée à la qualité des voyelles : les voyelles basses présentent une faible impédance tandis que les voyelles hautes sont caractérisées par une résistance plus importante. En conséquence, l'écoulement d'air mécanique est moins bien absorbé pour ces voyelles hautes, ce qui tend, par un effet de coarticulation, à augmenter la PIO des consonnes adjacentes.



**Figure 6.16** – Coupe sagittale montrant (i) la propagation du flux d'air (U) et (ii) la localisation des différentes résistances (R) dans le conduit vocal (d'après Baken 1987 :408) ainsi que les profils sagittaux des voyelles [a, i, u] (d'après Ghio 2007).

Enfin, nous avons pu mettre en évidence que quelles que soient les consonnes et leurs environnements vocaliques, les mesures moyennes de PIO des patients étaient toujours significativement plus élevées que celles des témoins. Nos conclusions sont en adéquation avec les études de Zeitels & al. (2002), Hans (2003) et Mirghani (2009). Pour différentes catégories de laryngectomies partielles, Hans (2003) a montré que les valeurs de PIO étaient toujours plus importantes que celles des témoins et que ce paramètre était le plus pertinent pour établir les distinctions entre ces catégories. Ainsi, la PIO avait tendance à augmenter avec l'étendue de l'exérèse. De même, pour différents types de cordectomies, Zeitels & al. (2002) et Mirghani (2009) ont montré que les valeurs moyennes de PIO étaient significativement plus élevées — à intensité constante — pour les patients en comparaison avec des témoins de même âge et de même sexe. Cette hausse de la PIO pour notre cohorte de patients pourrait s'expliquer par un comportement supraglottique correspondant à un phénomène compensatoire. Koufman (1991) met en évidence deux types de comportements supraglottiques : (i) une hyperadduction des bandes ventriculaires et (ii) une contraction antéro-postérieure du larynx caractérisée par une bascule antérieure des arytenoïdes vers l'épiglotte. Nous ne pouvons, néanmoins, pas faire d'hypothèse quant au type de compensation adoptée par nos patients, un examen vidéolaryngostroboscopique devant être pratiqué pour valider ce postulat. Finalement, l'élévation de la PIO chez les patients peut être liée à deux systèmes antagonistes : (i) le développement d'un forçage vocal, lié à un comportement supraglottique caractérisé par une tension plus importante des muscles laryngés et, plus largement, une tension extrême du conduit vocal entraînant secondairement

cette hausse de la pression ou (ii) la hausse de la PIO liée à l'intervention chirurgicale qui permet le développement d'un comportement supraglottique.

**Valeurs moyennes de Umax et de DPM.** Nous avons choisi de discuter les résultats de ces deux mesures de manière simultanée car celles-ci représentent un unique paramètre : le débit d'air buccal (DAB) pendant et après le relâchement de la consonne. Nous avons pu mettre en évidence que les valeurs moyennes de Umax et du DPM étaient principalement dépendantes du voisement des consonnes, de la qualité même des voyelles, de la position dans le mot (en particulier Umax) et de nos deux populations. A l'instar de la PIO, les mesures de Umax et du DPM peuvent être soumises aux variations de paramètres secondaires tels que les dimensions du conduit vocal (Fant 1960) ou la largeur buccale (Hixon 1966) par exemple, lesquels ne seront pas abordés dans cette discussion.

Nous avons montré que les valeurs de Umax et DPM, quels que soient les contextes vocaliques et ce, pour nos deux populations, étaient statistiquement plus élevées en contexte d'occlusives non voisées. Ces résultats confirment les études de Emanuel & Counihan (1970), Gilbert (1973), Trullinger & Emanuel (1983), Stathopoulos & Weismeyer (1985) ou Higgins & al. (1998) entre autres. Par exemple, dans une étude sur 25 hommes et 25 femmes, Emanuel & Counihan (1970) ont mis en évidence que les débits d'air maximums (Umax) des occlusives combinées à [i] ou [a] et en contextes CV et VCV variaient de façon significative avec le voisement, la qualité des voyelles, la syllabe et le sexe. Pour les auteurs, le voisement tend à réduire les mesures de débits d'air des occlusives voisées de 26 à 53% par rapport à leurs homologues non voisés. L'effet du voisement sur les pics de débits d'air des occlusives se retrouve également dans l'étude de Trullinger & Emanuel (1983). Basé sur un corpus similaire à Emanuel & Counihan (1970), cet article retrouve les mêmes tendances pour une population enfantine de 15 garçons et 15 filles. De la même manière, Gilbert (1973) a également mis en évidence cet effet du voisement sur les pics maximums de débits d'air sur des logatomes de type [CΛC] et [ΛC] chez 9 adultes masculins. En ce qui nous concerne, les différences de Umax pendant le relâchement de [p, b] et du DPM de la voyelle adjacente peuvent être imputées à la résistance laryngée imposée par la vibration des plis vocaux pendant la production des occlusives voisées. En effet, nous savons que les occlusives non voisées impliquent une ouverture glottique, caractérisée par l'abduction des plis vocaux. Par conséquent, le flux d'air pulmonaire n'est pas ralenti et peut s'échapper directement dans le conduit vocal. Un équilibre entre PSG et PIO s'établit après le point de constriction dans le tractus vocal. A l'opposé, la production des occlusives voisées nécessite une vibration entretenue des plis vocaux, ce qui augmente la résistance au passage de l'air. Ainsi, la PSG est toujours plus importante que la PIO dans le conduit vocal (Gilbert 1973, Stathopoulos & Weismeyer 1985).

Nous avons également pu noter que, de manière générale, les valeurs moyennes de Umax pendant le relâchement de la consonne tendent à être plus élevées lorsque la voyelle est basse. Cette conclusion va dans le sens des études de Emanuel & Counihan (1970), Trullinger & Emanuel (1983) ou Higgins & al. (1998). Nous avons déjà expliqué précédemment que l'élévation de la langue pour les voyelles hautes implique une

impédance plus importante dans la cavité orale, ce qui tend à réduire les débits d'air. Au contraire, l'articulation associant une aperture maximale et une position basse de la langue pour la réalisation de la voyelle ouverte [a] implique une faible résistance au débit d'air. Pour Higgins & al. (1998), l'aire et la résistance glottique peuvent également participer à ces différences de débits d'air maximums. En effet, l'élévation du larynx pour la tension des plis vocaux et, plus précisément, l'ajustement antéro-vertical du cartilage thyroïde et de l'os hyoïde, a pour effet principal de réduire l'aire glottique et donc d'augmenter la résistance au passage de l'air pour l'anticipation des voyelles hautes. Néanmoins, ces mêmes auteurs ne retrouvent pas de différences concernant les mesures des débits d'air moyens (DPM). Ainsi, selon eux il n'existe pas de différences significatives entre les débits phonatoires moyens de [a] et [i], ce qui est en contradiction avec nos résultats. Ils expliquent que les ajustements laryngés effectués en prévision d'une voyelle haute sont totalement établis pendant la partie stable de la voyelle, ce qui tend à équilibrer les résistances au niveau de la glotte entre ces deux types de voyelles. De notre côté, nous pensons que la réduction de l'aire glottique et la tension du conduit vocal sont maintenues pendant toute la production des voyelles hautes. De plus, la résistance dans la cavité orale est toujours plus élevée pour ces sons, ce qui participe à la réduction des DPM. En outre, il faut garder en mémoire la grande variabilité de ces mesures et considérer donc les interprétations des résultats des différentes études avec la plus grande prudence.

Nous avons également mis en évidence que la position dans les logatomes influençait les mesures de Umax en particulier. Ces dernières sont plus importantes en initiale absolue par rapport à la position intervocalique. Nos observations vont dans le sens des conclusions de Emanuel & Counihan (1970), Gilbert (1973) ou Trullinger & Emanuel (1983). Comme nous l'avons évoqué précédemment, l'étude de Emanuel & Counihan (1970) s'est intéressée à des syllabes de type CV et VCV, l'étude de Gilbert comportait des syllabes de type CVC et VC et l'étude de Trullinger & Emanuel (1983) comportait des échantillons de syllabes de type CV, VC et VCV. Il est également intéressant de préciser que les auteurs soulignent l'interaction de la position avec les voyelles adjacentes et le voisement de la consonne. Ainsi, l'effet de la syllabe sur les pics de débits d'air n'est pas indépendant du contexte vocalique et/ou de la vibration des plis vocaux. Il existe une échelle graduelle de hauteur des pics de débits qui prend en compte ces différents paramètres. Toutefois, pour les deux populations et ce, quelles que soient les occlusives analysées, tous les auteurs s'accordent sur le fait que les pics de débits d'air sont plus importants en initial ou en position prévocalique par rapport à la position intervocalique. En ce qui concerne la position finale — que nous n'avons pas prise en compte — la plupart des auteurs notent un moindre débit d'air. Ces conclusions suggèrent que les pressions et les résistances semblent plus faibles en position initiale par rapport à la position intervocalique. Le conduit vocal et la cavité buccale opposent une faible impédance pendant le relâchement de la consonne, ce qui tend à augmenter les débits d'air (Emanuel & Counihan 1970). En outre, notre corpus se base sur la lecture de logatomes dont l'énergie aérodynamique décline au cours du temps. Il est évident que le flux pulmonaire diminue pendant la phonation. Ainsi, il est attendu que l'éner-

gie soit maximale en début de logatome — en initiale absolue — et qu'elle s'amenuise jusqu'à la fin du mot.

Enfin, nous avons pu mettre en évidence que quelles que soient les consonnes et leurs environnements vocaliques, les mesures moyennes de DAB (Umax et DPM) des patients étaient toujours significativement plus élevées que celles des témoins. Nos conclusions corroborent les études de Zeitels & al. (2002), Tamura & al. (2003) et Mirghani (2009) pour les cordectomies et correspondent aux conclusions que nous avons établies pour les voyelles tenues (cf. section 6.5.2). Cette élévation des débits d'air est directement liée à la fuite glottique engendrée par la chirurgie. Ainsi, la résection d'une partie du pli vocal est responsable d'une asymétrie vibratoire et d'un accrolement incomplet des muqueuses. En conséquence, nous suggérons que l'intervention diminue la résistance glottique de ces patients en phonation, ce qui implique une déperdition permanente du flux d'air entre l'espace glottique et explique cette hausse de Umax et du DPM à la sortie.

**Estimation de la pression transglottique.** Pour nos deux populations et selon les deux positions étudiées, nous avons établi des calculs permettant une étude fine du trait de voisement : (i) la différence entre les PIO moyennes de [p] et [b] qui correspond à la pression transglottique de [b] et (ii) le rapport de cette différence sur la PIO de [p] qui permet de standardiser nos résultats initiaux par rapport à l'intensité sonore émise pendant la lecture du corpus.

Comme nous l'avons déjà mentionné, la pression sous-glottique et, plus particulièrement, la pression transglottique permettent d'initier et de maintenir une vibration entretenue des plis vocaux pour les sons voisés. Ainsi, Titze (1988) précise que la vibration des plis vocaux doit être initiée par une pression pulmonaire suffisante. Dans cette perspective, Verdolini & al. (1994) ajoutent qu'une PSG minimale permet de déclencher le voisement mais qu'elle contribue surtout à maintenir une oscillation permanente des plis vocaux. Chan & al. (1997), Luceno (1996) et Luceno & Koenig (2005) précisent que cette pression est différente entre le début du voisement et la phase d'entretien des vibrations. En effet, la pression doit être plus importante au début du son voisé car l'énergie aérodynamique transférée entre les plis vocaux est plus vite absorbée. En l'espèce, si l'énergie dynamique conduite par cette pression transglottique ne se perpétue pas au cœur du système vibratoire, alors les cycles vibratoires seront amortis.

Pour Kitajima & Fujita (1992) et Baken & Orlikoff (2000), la différence des pics de PIO lors des consonnes [p] et [b] représente une estimation de la pression transglottique et peut être identifiée à la pression de seuil phonatoire (PSP) (cf. Smitherson & Hixon 1981). Finalement, cette PSP est définie comme la pression minimale nécessaire à l'initiation et au maintien du voisement pour des configurations glottiques particulières. Ainsi, elle permet de compenser la perte d'énergie liée à l'adduction des plis vocaux. Ce seuil phonatoire est de l'ordre de 2 à 4 hPa avec une pression sous-glottique usuelle de 7 hPa.

Le niveau de la PSP dépend de la raideur de la partie vibrante des plis vocaux, de leur épaisseur, de leur viscosité, de la largeur de la fente glottique préphonatoire et de

la qualité de l'ondulation de la muqueuse pendant la phonation. Ainsi, Titze (1988) a mis en évidence une corrélation positive entre ces différents facteurs. Si l'ondulation muqueuse, la largeur glottique et la raideur des plis vocaux diminuent, alors la PSP suivra cette tendance. Chan & al. (1994) précisent, quant à eux, qu'une augmentation des tissus, une épaisseur des plis vocaux ou une augmentation de la raideur de la membrane épithéliale sont en faveur d'une hausse de la PSP. Par contre, une glotte préphonatoire de forme rectangulaire aura tendance à abaisser la PSP (Luceno 1996). De même, un lien avec la fatigue vocale, l'hydratation des muqueuses, le contexte vocalique ou la  $F_0$  a été démontré. Par exemple, Koenig & al. (2011) tirent la conclusion qu'une plus grande tension des plis vocaux a pour conséquence une augmentation de la  $F_0$  et, par extension de la PSP. Les voyelles hautes, qui sont caractérisées par une  $F_0$  intrinsèque élevée par rapport aux voyelles basses, requièrent un effort phonatoire supplémentaire, ce qui tend à augmenter la PSP. Titze (1988) et Verdolini & al. (1994) avaient déjà constaté ce lien étroit entre  $F_0$  et PSP. Ainsi, dans une étude sur 12 adultes, Verdolini & al. (1994) avaient souligné que la PSP moyenne dans les basses fréquences était de  $3,2 \text{ cmH}_2\text{O}$ , la PSP en voix modale était de  $3,19 \text{ cmH}_2\text{O}$  et la PSP dans les hautes fréquences s'élevait à  $8,24 \text{ cmH}_2\text{O}$ . Enfin, les auteurs ajoutent que cette mesure/calcul permet d'objectiver la facilité ou la difficulté à produire des sons voisés. La PSP permet de diagnostiquer une pathologie vocale ou une fermeture glottique inefficace (Titze 1988, Verdolini & al. 1994, Chan & al. 1997).

En ce qui nous concerne, les résultats pour la population témoin permettent d'affirmer que la différence de pression nécessaire au voisement en contexte vocalique [a] est de 2,69 hPa et 2,57 hPa pour les séquences /pa/, /ba/ et /apa/, /aba/ respectivement. De même, la différence essentielle au voisement en contexte [i] est de 2,55 hPa et 3,51 hPa pour les séquences /pi/, /bi/ et /ipi/, /ibi/ respectivement. Enfin, cette différence est de 3,06 hPa et 3,19 hPa pour les séquences /pu/, /bu/ et /upu/, /ubu/. Nous constatons que nos résultats suivent bien l'intervalle donné de 2 à 4 hPa pour l'initiation et le maintien du voisement. Aussi, ces calculs sont en adéquation avec les PSP déterminées par Titze (1988) et Verdolini & al. (1994) selon les contextes vocaliques. De plus, nous notons que les pressions minimales sont plus élevées en contexte de voyelles hautes par rapport à la voyelle basse. Pendant le relâchement de la consonne, des ajustements liés à la coarticulation s'effectuent en amont. Pour la production des voyelles hautes, les plis vocaux vont se tendre, s'étirer davantage, ce qui va réduire leur masse vibrante et augmenter l'impédance glottique. En conséquence, la vibration va être plus difficile à maintenir, ce qui va favoriser, par un phénomène de compensation, l'élévation de la PSP. Finalement, nos résultats sont plus ou moins identiques aux résultats des études de Subtelny & al. (1966), Murry & Brown (1976), Stathopoulos (1986), Kitajima & Fujita (1992) et Hans (2003) qui appellent, malgré tout, à la considération des variations intra-individuelles entre les sujets. Le tableau 6.17 dans lequel apparaissent uniquement les résultats pour les sujets masculins sains, résume ces principales observations :

Auteurs	Corpus	$PIO/p - b/$	$(PIO/p - b)/(PIO/p/)$
Subtelny (1966)	/ipi/, /ibi/	2,06	0,32
Murry (1976)	/apa/, /aba/	3,48	0,42
Stathopoulos (1986)	/pap/, /bab/	2,68	0,23
Kitajima (1992)	/ipi/, /ibi/	4,7	0,44
Hans (2003)	/pi/, /bi/ et /ipi/, /ibi/	2,4 et 2,8	0,35 et 0,29

**Table 6.17** – Résultats des principales études concernant les estimations de la pression transglottique chez des sujets sains

En ce qui concerne les résultats de notre population de patients, ces derniers montrent que la différence de pression nécessaire au voisement en contexte vocalique [a] est de 4,44 hPa et 6,65 hPa pour les séquences /pa/, /ba/ et /apa/, /aba/ respectivement. De même, la différence essentielle au voisement en contexte [i] est de 5,5 hPa et 7,93 hPa pour les séquences /pi/, /bi/ et /ipi/, /ibi/ respectivement. Enfin, cette différence est de 5,48 hPa et 7,56 hPa pour les séquences /pu/, /bu/ et /upu/, /ubu/ respectivement. Dans un premier temps, nous remarquons qu'à l'instar des témoins, les pressions restent plus élevées dans des contextes de voyelles hautes. Dans un second temps, nous relevons des pressions transglottiques de [b] bien plus hautes que celles des témoins et ce, pour l'ensemble des contextes vocaliques. Ce constat est en adéquation avec l'étude de Hans (2003), qui avait également mis en évidence des différences de pression plus importantes pour les patients traités par laryngectomie partielle par rapport à une cohorte de témoins, ce que l'auteur avait déjà interprété comme une difficulté de voisement.

Nous avons déjà mentionné le fait que la PSP est à mettre en relation avec la raideur de la partie vibrante des plis vocaux, leur épaisseur, leur viscosité, la largeur de la fente glottique préphonatoire et la qualité de l'ondulation de la muqueuse pendant la phonation. Or, la cordectomie modifie largement la configuration glottique en postopératoire. Ainsi, une ou plusieurs couches du pli vocal sont réséquées lors de l'intervention, ce qui a pour conséquence une augmentation de la raideur du pli vocal, une tension plus importante de ce dernier et une réduction de la masse vibrante. De plus, une asymétrie vibratoire est constatée, ce qui implique une ondulation muqueuse imprécise. Le modèle « body-cover » de Hirano (1977), basé sur trois coupes histologiques du pli vocal caractérisées par le « body » (ligament et muscle vocal) et le « cover » (muqueuse du pli vocal) séparés par un espace de glissement (espace de Reinke) est à l'origine d'un triple mouvement horizontal, vertical et muqueux des plis vocaux en phonation. Or, les modifications des structures anatomiques, en particulier du « cover », engendrent une réduction de l'aire glottique et une amplitude de vibrations amortie. En conséquence, les patients auront tendance à augmenter leur PSP par rapport aux témoins afin d'initier et de maintenir l'oscillation glottique pour compenser cette mutilation cordale. De notre point de vue, nous sommes totalement en accord avec l'étude de Titze (1988). En effet, comme nous l'avons montré précédemment, celui-ci précise qu'il existe un lien étroit entre  $F_0$  et PSP. Selon lui, une forte tension des plis vocaux est liée à une augmentation de leur longueur, à leur amincissement, à une hausse de leur rigidité et à

une diminution de leur masse vibrante, ce qui tend à augmenter les cycles vibratoires et, *in extenso*, la  $F_0$ . Ainsi, plus les plis vocaux sont tendus et plus la PSP doit être importante pour vaincre l'importante résistance laryngée et initier ces oscillations. Dans le chapitre 5, nous avions déjà mis en évidence que la chirurgie avait pour conséquence directe l'augmentation des  $F_0$  intrinsèques des voyelles de nos patients par rapport à notre cohorte de témoins, ce qui explique la hausse de la PSP constatée pour ces sujets et corrobore les conclusions de Titze (1988). En outre, nos résultats vont dans le sens de Titze (1988), Verdoloni & al. (1994) et Chan & al. (1997), qui considèrent que le calcul de la PSP est un indice d'efficience de vibrations des plis vocaux et qu'il permet donc de classifier les voix des sujets sains par rapport aux sujets dysphoniques présentant une pathologie vocale. Ainsi, dans une étude sur des patients japonais atteints de diverses pathologies laryngées, Kitajima & Fujita (1992) ont montré que la différence  $PIO/p - b/$  permettait d'objectiver la pathologie : si la différence est inférieure à la normale, la force d'adduction des plis vocaux est considérée comme insuffisante, c'est le cas typique des paralysies récurrentielles unilatérales. Au contraire, si la différence est supérieure à la normale, une rigidité anormale des plis vocaux doit être considérée et une fermeture glottique inefficace est admise, c'est le cas typique des cancers du pli vocal.

Finalement, cet indice de PSP — caractérisé par la différence  $PIO/p - b/$  pour laquelle nous avons pu mettre en évidence des valeurs moyennes de PIO bien plus élevées chez nos patients par rapport à nos témoins — indique la persistance de la rigidité et de la tension du pli vocal opéré, une fermeture partielle des plis vocaux (cf. fuite glottique) et implique donc une difficulté certaine pour réaliser le trait de voisement chez nos patients d'un point de vue aérodynamique.

**Synthèse de l'étude n°6 : études aérodynamiques**

- Objectifs de l'étude : confirmer l'existence de fuites glottiques et analyser de manière quantitative certains paramètres aérodynamiques.
- Corpus : (i) parties stables des voyelles tenues et isolées et (ii) logatomes de type CVCVCVC où C= [p, b] et V= [a, i, u].
- Principaux résultats :
  - les indices de fuites glottiques des patients sont significativement plus élevés que ceux des témoins (incompétence laryngée),
  - synchronisation satisfaisante des courbes de PIO et de DAB des 2 groupes,
  - élévation des valeurs de la PIO, de Umax et du DPM en contexte non voisé,
  - PIO plus importante en contexte de voyelles hautes, Umax et DPM plus élevés pour la voyelle [a],
  - les mesures des paramètres aérodynamiques sont toujours supérieures pour les patients par rapport aux témoins,
  - les patients présentent des pressions transglottiques estimées plus élevées que les témoins, en particulier pour les voyelles hautes.
- Conclusions :
  - l'inefficacité glottique est confirmée chez les patients,
  - l'élévation des mesures des paramètres aérodynamiques traduit une difficulté pour l'opposition de voisement après cordectomie,
  - néanmoins, le fait qu'il existe des différences significatives dans les mesures entre contexte voisé et non voisé suggère que l'opposition de voisement est maintenue chez les sujets dysphoniques.

## 6.6 Etude n°7 : études acoustiques

Dans la section précédente, nous avons mis en évidence une fermeture irrégulière des plis vocaux impliquant des fuites glottiques et une altération partielle du trait de voisement d'un point de vue aérodynamique. Cette section est dédiée à l'étude du voisement d'un point de vue acoustique. Après avoir rappelé la méthodologie générale utilisée, nous étudierons plus particulièrement les paramètres suivants : la durée totale des occlusives orales, l'intensité relative du relâchement, la durée des voyelles adjacentes, la  $F_0$  moyenne au début de ces segments vocaliques et le VOT.

### 6.6.1 Méthodologie

#### 6.6.1.1 Procédure initiale

Tous nos sujets ont été enregistrés dans les mêmes conditions, selon un protocole commun et un matériel similaire (cf. chapitre 4). Notre protocole se compose de logatomes de type CVCVCVC où C représente les consonnes du français et V= [a, i, u]. Pour cette étude acoustique sur l'implémentation du trait de voisement, nous avons restreint notre analyse aux occlusives orales [p, t, k, b, d, g].

Néanmoins, en observant nos données, nous nous sommes rapidement rendue compte que notre carte son était défectueuse et qu'un bruit de fond lié à l'alimentation rendait l'interprétation de nos résultats acoustiques aléatoires. En conséquence, un script Matlab Butterworth a été utilisé pour filtrer ces bruits parasites. Il s'agit d'un filtre passe haut qui prend en compte la largeur de bande et la valeur de la coupure. Ainsi, plus on augmente la largeur de bande et plus le filtre coupe. Dans notre cas, nous avons implémenté un filtre à 50 Hz (bruit de fond) et une zone de filtrage assez large pour avoir un meilleur rendu final. Finalement, la segmentation des données et les analyses ont, ainsi, pu être effectuées sur ces fichiers filtrés à partir de spectrogrammes à bandes larges avec une étendue de fréquences de 0 à 5000 Hz.

### 6.6.1.2 Segmentation des données

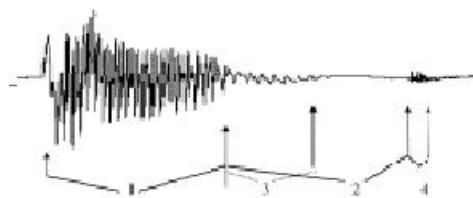
Les données que nous avons recueillies ont été segmentées de manière semi-automatique avec EasyAlign (Goldman 2008), une extension de Praat (Boersma & Weenink 1996) uniquement disponible sous Windows. Ce *plugin* permet de générer automatiquement des annotations sur différentes tires contenant une segmentation alignée de phonèmes, de syllabes, de mots ou d'énoncés à partir de la transcription orthographique d'un enregistrement de parole. Les résultats se présentent sous la forme d'un TextGrid composé de cinq tires : phonèmes, syllabes, mots, phonologie et orthographe. Néanmoins, cet alignement reste dans certains cas aléatoire, ce qui constraint à une inspection manuelle et visuelle des spectrogrammes. Ainsi, des ajustements à cet alignement initial peuvent être apportés en manipulant conjointement les spectrogrammes et les oscillogrammes. Les segments sont écoutés attentivement et la réalité perceptive est associée à la réalité acoustique.

Dans le cas des occlusives intervocaliques, la segmentation dépend principalement de trois évènements, lorsqu'ils sont visibles. Premièrement, nous repérons la fin du F2 de la voyelle précédente qui concorde avec une phase de silence (pour les occlusives non voisées) ou une barre de voisement (pour les occlusives voisées). Deuxièmement, nous considérons que la fin de l'occlusion correspond au début de l'explosion, celle-ci pouvant prendre différentes formes selon les locuteurs. Par ailleurs, le relâchement peut ne pas apparaître de façon nette sur les spectrogrammes. Nous avons donc tenté de le déterminer à l'aide des changements de formes du signal — principalement pour les occlusives voisées — ou par le suivi formantique proposé par Praat. Enfin, le début du F1 de la voyelle suivante représente la fin de ces occlusives.

Dans le cas des occlusives non voisées en position initiale, un problème de segmentation des phases d'occlusions nous a été posé. En effet, il est impossible de rendre compte, avec certitude, du démarrage de ces phases pour ces types de consonnes. Ainsi, par convention certes arbitraire, nous avons déterminé et segmenté celles-ci à 50 ms.

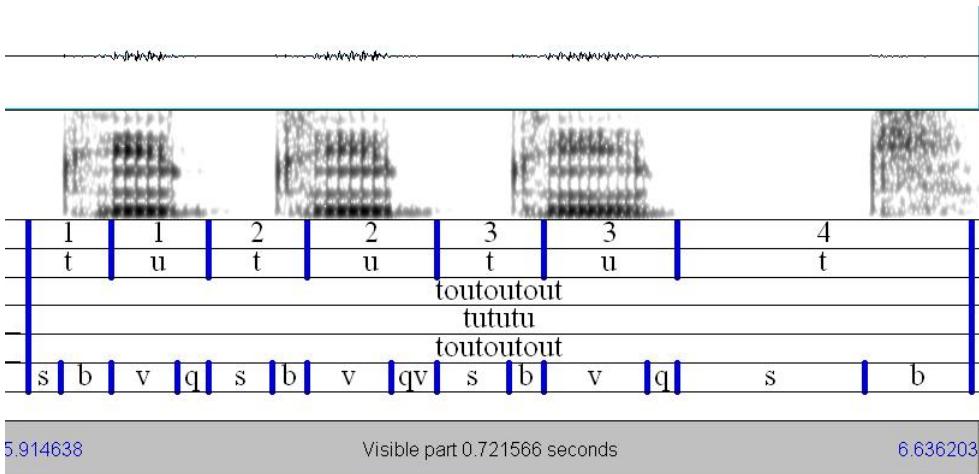
Le début des voyelles était, quant à lui, déterminé principalement par le changement de forme du signal et par l'apparition du F2 de cette voyelle. La fin des voyelles était repérée par la fin de la visibilité du F2 et par l'apparition du début de l'occlusion de la consonne suivante.

De notre point de vue, nous avons opté pour une segmentation fine de ces différents évènements — à l'aide des changements d'amplitude du signal et du suivi formantique — afin de faciliter l'extraction automatique de quelques paramètres liés à notre étude sur le voisement. La figure 6.17 nous a servi de base pour notre découpage (Smith & al. 2009) : le premier intervalle, lié à la présence d'un noyau vocalique, a été codé « v » dans notre corpus ; le troisième intervalle qui indique la présence de cycles glottiques persistants pendant le début de l'occlusion a été codé « qv » (queue de voyelle) et correspond au *Voice Decay Time* ; le deuxième intervalle qui correspond à la phase d'occlusion a été identifié « s » (silence) dans notre TextGrid et le quatrième intervalle qui représente la phase de relâchement a été annoté « b » (relâchement).



**Figure 6.17** – Exemple de forme du signal pour une segmentation fine. L'intervalle 1 indique la durée de la voyelle, l'intervalle 2 indique la durée de l'occlusion de la consonne, l'intervalle 3 indique le *Voice Decay Time*, l'intervalle 4 indique le relâchement de la consonne (Smith & al. 2009).

Finalement, notre objectif majeur étant de vérifier l'implémentation phonétique du trait de voisement selon la position dans le mot pour les patients et les témoins, nous avons fait le choix de découper l'ensemble de nos logatomes en trois syllabes distinctes afin de faciliter leur traitement. Ainsi, une tire supplémentaire a été ajoutée au TextGrid original dans laquelle chaque phonème était associé à une position. Nous obtenons donc le schéma suivant :  $C_1V_1C_2V_2C_3V_3C_4$  qui renvoie plus largement à la distinction de trois syllabes  $CV_1$ ,  $CV_2$  et  $CVC_3$ . Par conséquent, nous considérons que la première syllabe est composée d'une occlusive en position d'attaque en initiale absolue ( $\#$ ), soit  $C_1$  ; la seconde syllabe est composée d'une consonne en position d'attaque intervocalique ( $V_V$ ), soit  $C_2$  et la troisième syllabe est constituée à la fois d'une consonne en position d'attaque intervocalique ( $C_3$ ) et d'une coda en position finale ( $\#$ ), soit  $C_4$ . La figure 6.18 nous donne un aperçu des fichiers étiquetés que nous avons pu exploiter pour cette étude.



**Figure 6.18** – Segmentation du logatome [tututut] sous Praat (exemple du témoin n°1). La tire que nous devons prendre en compte et qui représente une segmentation correcte est la dernière où « s » correspond au silence de l’occlusive, « b » correspond au relâchement de la consonne, « v » correspond à la voyelle et « qv » correspond au *Voice Decay Time*.

### 6.6.1.3 Mesures acoustiques

Toutes les mesures ont été effectuées sur les positions 1 et 2 de nos logatomes, soient  $C_1$ ,  $V_1$ ,  $C_2$  et  $V_2$ . Notre choix se justifie principalement par l’intérêt d’exploiter une position syllabique, caractérisée par des attaques dans deux contextes distincts. Par ailleurs, nous avons délibérément écarté la troisième syllabe pour des raisons essentiellement prosodiques. En effet, la lecture d’un corpus composé de logatomes n’est pas naturelle, ce qui induit souvent des *patterns* de  $F_0$  non canoniques par rapport à une assertion type. Certains locuteurs inséraient des pauses associées à des montées de continuation entre les positions 2 et 3. De plus, la tâche de lecture de logatomes est souvent liée au principe d’énumération, ce qui implique également des montées de continuation finales. Aussi, nous avons constaté, dans nos données, qu’il était fréquent que le relâchement final ne soit pas réalisé à cause de l’enchaînement avec le logatome suivant qui débute par une occlusive homorganique ou qu’il soit associé à un schwa final. En conséquence, nous nous sommes concentrée sur les positions 1 et 2, qui représentaient les positions les plus stables en termes d’amplitude, de fréquence et d’articulation.

Pour cette étude liée à l’implémentation du trait de voisement, nous avons choisi d’analyser les logatomes composés d’occlusives, soit 720 logatomes (18 logatomes de type CVCVCVC où  $C = [p, t, k, b, d, g]$  et  $V = [a, i, u] \times 2$  positions  $\times 1$  répétition  $\times 20$  sujets). A l’exception du VOT, pour lequel le lieu d’articulation des consonnes semble pertinent, nous avons fait le choix de regrouper les différents logatomes selon leur contexte de voisement pour les traitements statistiques des divers paramètres étudiés.

Selon les critères de Lisker & Abramson (1964) ou Repp (1979), nous avons sélectionné, dans notre étude, les paramètres suivants : la durée totale des occlusives orales avec une attention particulière portée aux durées des phases de tenue et de relâchement, l'intensité relative du relâchement, la durée des voyelles subséquentes, la  $F_0$  moyenne au début des segments vocaliques et la mesure du VOT (cf. section 6.3.1). Toutes nos prises de mesures ont été extraites à l'aide d'un script semi-automatique Praat créé spécialement pour cette étude et vérifiées manuellement<sup>7</sup>. Aussi, avant d'expliquer en détail les méthodologies adoptées pour l'évaluation de ces divers paramètres, il est important de noter que pour l'ensemble des mesures de durées, y compris le VOT, nous avons volontairement écarté les productions du témoin n°6. En effet, nous nous sommes rapidement aperçus que les logatomes composés de l'occlusive vélaire [k] et les logatomes composés des occlusives voisées étaient hyperarticulés. Ainsi, ce locuteur a cherché à syllaber en adoptant un débit ralenti certains logatomes au détriment d'autres, ce qui rend notre corpus hétérogène en terme de production. De ce fait, des durées exagérées — de l'ordre de 250 ms parfois — pour les consonnes voisées ont été relevées, ce qui pouvait représenter un biais assez important par rapport aux occlusives non voisées, et plus largement par rapport aux productions des autres témoins. Pour ces mesures de durées, nous avons donc analysé 684 occurrences, soit  $18 \text{ logatomes} \times 2 \text{ positions} \times 1 \text{ répétition} \times 19 \text{ sujets}$ .

Les durées des consonnes — tenue et relâchement compris — et des voyelles ont été extraites par notre script Praat en se basant sur l'étiquetage initial, c'est-à-dire sur notre dernière tire où « s » et « b » représentent la consonne et « v » correspond à la voyelle. Pour la mesure des durées consonantiques, nous nous sommes principalement intéressés à la position intervocalique ( $n = 342$ , soit  $18 \times 1 \text{ position} \times 1 \text{ répétition} \times 19 \text{ sujets}$ ). En effet, comme nous le précisons précédemment, la durée de la fermeture des occlusives non voisées initiales n'est pas détectable. Par convention, nous avons segmenté cette phase à 50 ms, mais nous sommes consciente que cette valeur, bien qu'arbitraire, ne représente pas forcément la réalité. En conséquence, un biais peut apparaître dans cette unique position syllabique pour l'opposition de durées des occlusives voisées et non voisées.

En ce qui concerne la  $F_0$  des voyelles adjacentes, nous avons sélectionné une fenêtre temporelle de 15 ms au début du segment dans laquelle une moyenne des différentes valeurs a pu être calculée de manière automatique par le script.

L'intensité relative du relâchement est calculée dans le but de normaliser les différences entre les locuteurs ainsi que les conditions d'enregistrement. Dans un premier temps, notre script nous a permis d'identifier les pics d'amplitude maximums dans les relâchements des logatomes des différents sujets. Une tire supplémentaire a été ajoutée au TexGrid original avec des points « B » — vérifiés manuellement — qui représentent ces pics. Dans un second temps, et afin d'éviter les effets de la  $F_0$  dans les occlusives voisées, nous avons filtré nos enregistrements à l'aide d'un filtre passe-bande entre 400 et 12000 Hz. Le filtre utilisé était de type « Hanning » avec un lissage de 100 Hz. Après filtrage, l'intensité relative du relâchement a été calculée dans une fenêtre de 10

7. Le script a été créé par JM. Beltzung

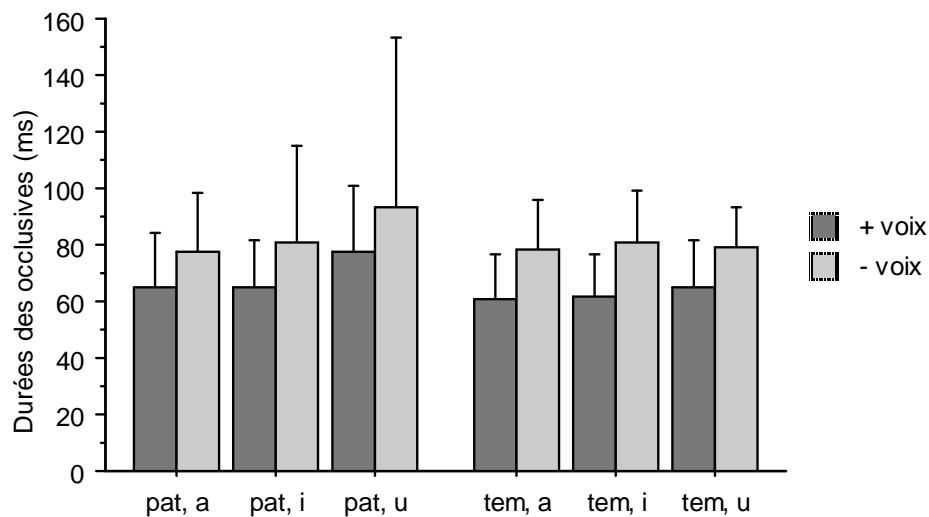
ms, centrée sur le pic « B », à savoir 5 ms avant et 5 ms après ce point. Finalement, l'intensité relative du relâchement représente la différence entre la valeur calculée dans le relâchement et le pic d'intensité de la voyelle suivante (Flack 2007).

Nous avons mesuré le VOT des consonnes selon la définition de Lisker & Abramson (1964). Nous avons donc calculé, pour chaque consonne, la durée entre le relâchement de l'occlusion et l'apparition des premières vibrations laryngées. Ce paramètre sera le seul à être présenté en fonction des lieux d'articulation des consonnes.

Finalement, ces différentes mesures nous ont permis d'évaluer la réalisation du voisement en fonction de nos deux groupes, des contextes vocaliques et des positions syllabiques. Nous avons donc conduit des ANOVA à 4 facteurs où les variables dépendantes représentaient les paramètres évoqués et les variables indépendantes correspondaient au groupe, au type de voyelle, au type de consonne et à la position dans les logatomes (sauf pour la durée des occlusives) à l'aide du logiciel Statview. Les résultats étaient considérés comme significatifs à partir d'un seuil alpha de 5%, soit  $p < .05$ .

## 6.6.2 Présentation des résultats

**Durée des occlusives.** La figure 6.19 présente les durées moyennes des occlusives en position intervocalique en fonction de nos deux groupes, du type de voyelle et du contexte de voisement des consonnes. Nous remarquons un effet significatif du voisement sur les durées des segments consonantiques ( $F(1, 330) = 30,444, p < .0001$ ). Les occlusives non voisées tendent à être plus longues que leurs contreparties voisées. De même, nous notons une différence significative entre nos deux groupes. Les durées des occlusives des patients semblent un peu plus longues que celles des témoins ( $F(1, 330) = 3,992, p = .0465$ ). De même, un effet du contexte vocalique semble émerger. Les occlusives ont une durée supérieure en contexte de voyelles hautes par rapport à la voyelle basse [a] ( $F(1, 330) = 3,127, p = .0452$ ). Le test PLSD de Fisher montre une différence significative entre les voyelles [a, u] et une différence quasi significative entre les voyelles [i, u], avec des p-value respectivement égales à  $p = .0141$  et  $p = .0506$ . Par ailleurs, nous ne retrouvons aucune interaction entre les différents facteurs.

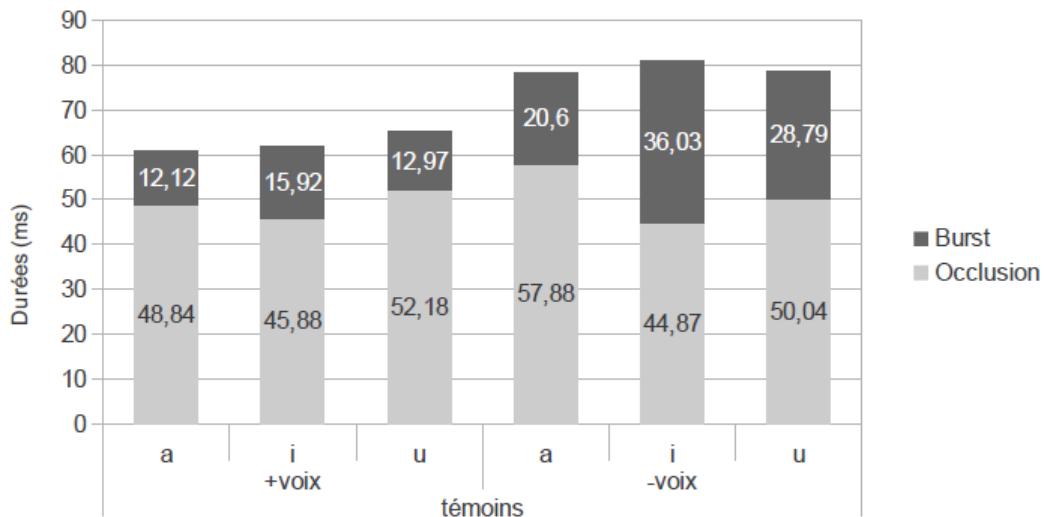


**Figure 6.19** – Valeurs moyennes des durées consonantiques (ms) en position intervocalique (V\_V) en fonction des sujets, du type de voyelle et du contexte de voisement (les barres d'erreurs représentent les écarts-types) ( $n = 342$ ).

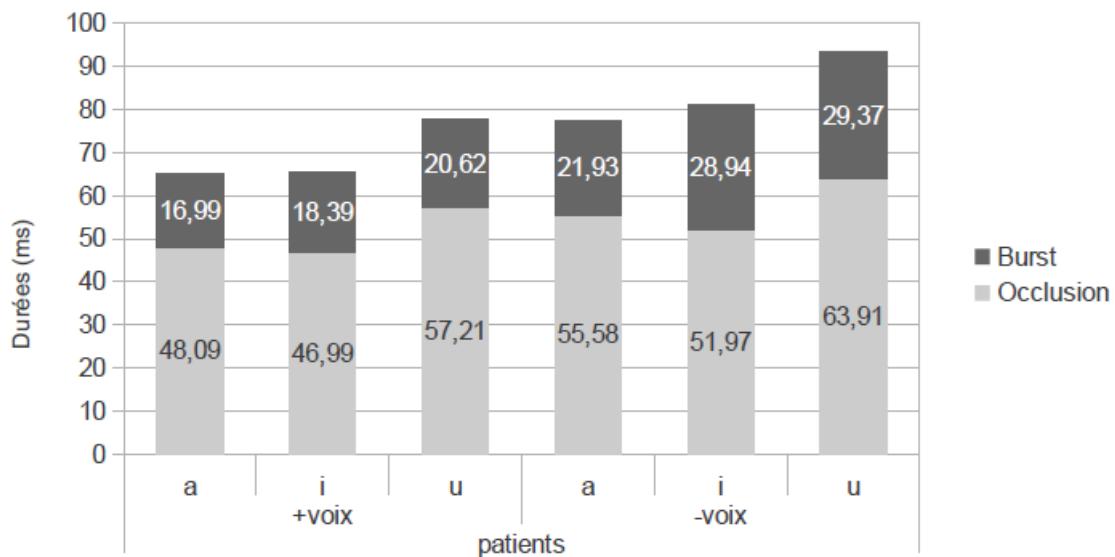
Pour un examen plus approfondi de ces segments consonantiques, nous avons également analysé les durées des phases de tenue et de relâchement des occlusives. Les figures 6.20 et 6.21 présentent les répartitions des durées moyennes des deux phases articulatoires des consonnes intervocaliques de nos sujets, en fonction du contexte de voisement et du type de voyelle. En ce qui concerne les durées moyennes des phases de tenue, nous constatons que le voisement ne semble pas avoir d'effet significatif sur ce paramètre. Bien que les durées des occlusions tendent à être plus longues en contexte non voisé, les résultats pour notre cohorte de témoins en contexte de voyelles hautes ne suivent pas la même tendance ( $F(1, 330) = 2,194, p = .1395$ ). De même, le test ANOVA ne montre aucune différence significative entre nos deux groupes ( $F(1, 330) = 2,022, p = .1560$ ). Seul le contexte vocalique semble influencer, dans une moindre mesure, les durées moyennes des occlusions. La voyelle [u] tend à allonger ce paramètre ( $F(2, 330) = 2,914, p = .0557$ ). Le test PLSD de Fisher montre une unique différence entre les voyelles [i, u], avec une  $p$ -value égale à  $p = .0154$ . Néanmoins, nous ne retrouvons aucune interaction entre les facteurs.

Les résultats pour les durées moyennes des phases de relâchement sont bien différents. Ainsi, pour ce paramètre, le contexte de voisement montre une influence hautement significative. Pour les deux populations et dans tous les contextes vocaliques, les durées moyennes des relâchements sont beaucoup plus importantes en contexte non voisé ( $F(1, 330) = 132,385, p < .0001$ ). De la même manière, l'ANOVA montre un effet significatif du facteur « voyelle ». Comme précédemment, nous constatons que les voyelles hautes tendent à avoir un effet allongeant sur les relâchements des

consonnes de nos sujets ( $F(2, 330) = 17,346, p < .0001$ ). Le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [a, u] avec des p-value égales à  $p < .0001$ . Au contraire, nous ne retrouvons aucune différence significative entre nos deux populations ( $F(1, 330) = 2,638, p = .1053$ ). Enfin, nous pouvons noter que trois interactions sur l'ensemble de nos variables indépendantes présentent un seuil de significativité inférieur à 5%. Premièrement, l'interaction entre les facteurs « voyelle » et « sujet » est significative ( $F(2, 330) = 3,998, p = .0194$ ). Deuxièmement, l'interaction entre les facteurs « sujet » et « voisement » est aussi significative ( $F(1, 330) = 11,339, p=.0008$ ). Troisièmement, l'interaction entre les facteurs « voyelle » et « voisement » est également significative ( $F(2, 330) = 6,444, p = .0018$ ).

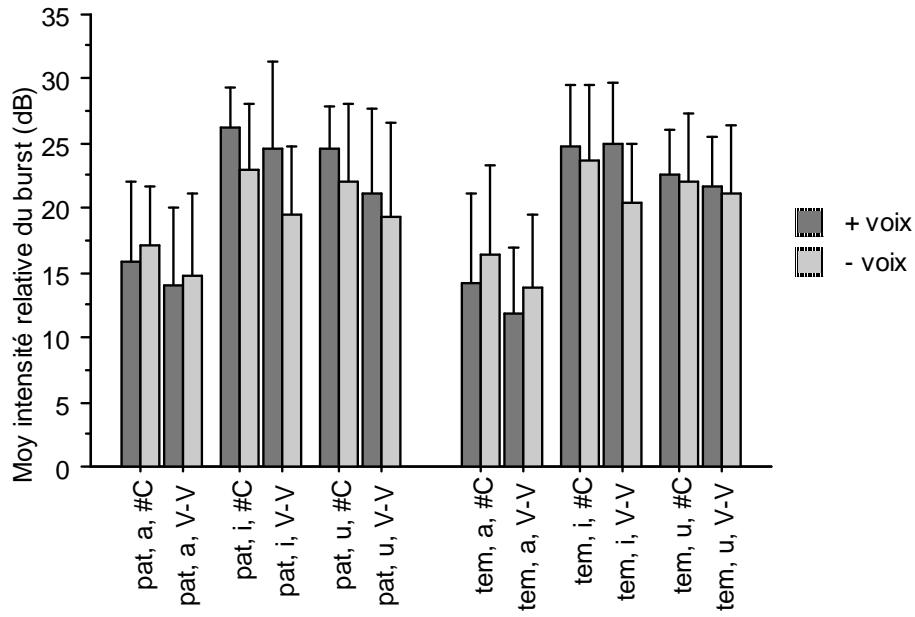


**Figure 6.20** – Durées moyennes de la tenue et du relâchement (ms) des consonnes en position intervocalique (V\_V) des témoins en fonction du contexte de voisement et du type de voyelle ( $n = 162$ ).



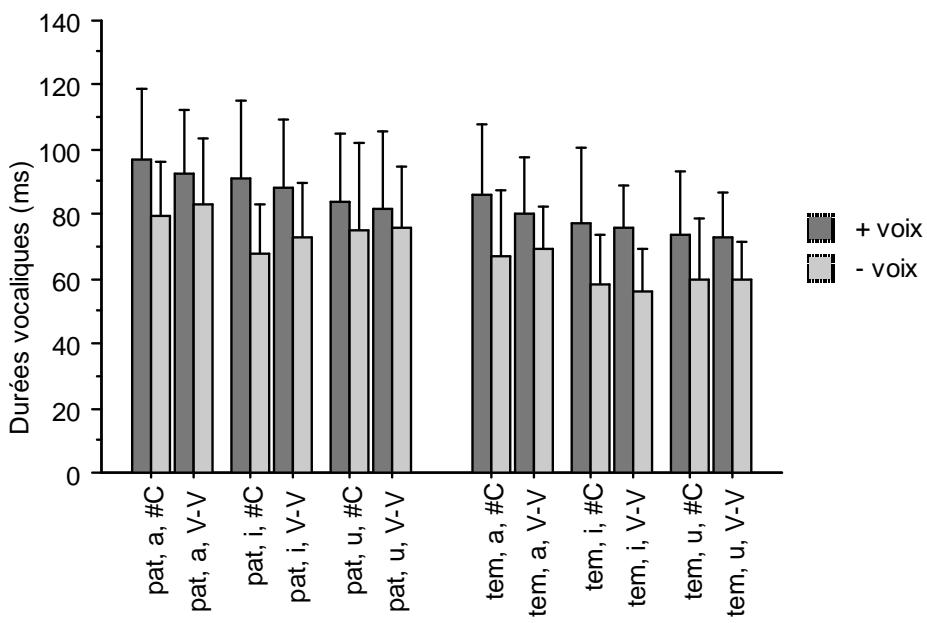
**Figure 6.21** – Durées moyennes de la tenue et du relâchement (ms) des consonnes en position intervocalique (V\_V) des patients en fonction du contexte de voisement et du type de voyelle ( $n = 180$ ).

**Intensité relative du relâchement.** Comme nous l'avons déjà évoqué, l'intensité relative du relâchement représente la différence entre le pic d'intensité de la voyelle adjacente et le pic maximal d'intensité du relâchement de la consonne. La figure 6.22 présente les valeurs moyennes de ce paramètre en fonction de nos deux populations, du type de voyelle, du contexte de voisement et de la position dans les logatomes. Nous pouvons remarquer un effet significatif du contexte de voisement sur les valeurs moyennes de l'intensité relative du relâchement ( $F(1, 696) = 7,661, p = .0058$ ). Ainsi, pour les deux groupes et à l'exception du contexte vocalique [a], nous notons des valeurs plus élevées en contexte voisé, quelle que soit la position dans le mot. De même, le type de voyelle semble avoir une influence significative sur les valeurs moyennes de l'intensité relative du relâchement ( $F(2, 696) = 174,024, p < .0001$ ). Dans ce sens, nous remarquons que les valeurs moyennes de l'intensité relative du relâchement tendent à être plus importantes en contexte de voyelles hautes par rapport à la voyelle basse [a]. Le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i], [a, u] et [i, u] avec des p-value respectivement égales à  $p < .0001$ ,  $p < .0001$  et  $p = .0017$ . Nous notons également un effet de la position dans le mot sur cette mesure ( $F(1, 696) = 27,047, p < .0001$ ). La position initiale semble favoriser des valeurs plus importantes de l'intensité relative du *relâchement*. Enfin, notre test ANOVA ne montre aucun effet du facteur « sujet » sur cette mesure ( $F(1, 696) = 0,929, p = .3356$ ). Finalement, nous notons qu'une unique interaction entre nos variables indépendantes présente un seuil de significativité inférieur à 5%. Ainsi, seule l'interaction entre les facteurs « voyelle » et « voisement » est significative ( $F(2, 696) = 12,053, p = < .0001$ ).



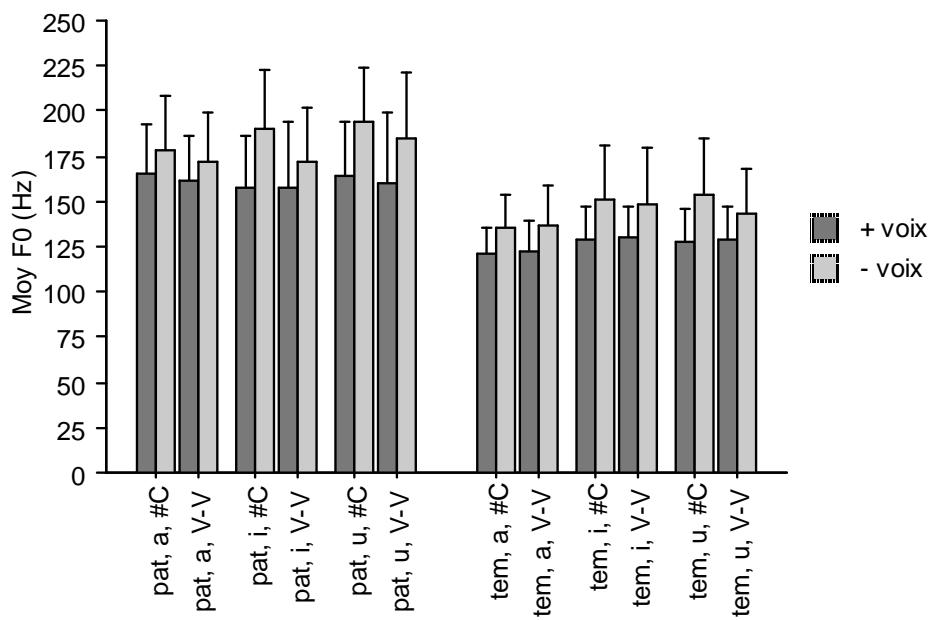
**Figure 6.22** – Valeurs moyennes de l'intensité relative du relâchement (en dB) des consonnes en fonction des sujets, du type de voyelle, du contexte de voisement et de la position dans le logatomes (les barres d'erreurs représentent les écarts-types) ( $n = 720$ ).

**Durée des voyelles adjacentes.** La figure 6.23 présente les durées moyennes des voyelles en fonction de nos deux populations, du type de voyelle, du contexte de voisement des occlusives et de la position dans le mot. Nous constatons que quel que soit le voisement de la consonne, les durées vocaliques des patients sont hautement supérieures à celles des témoins. Le test ANOVA montre une différence significative entre les deux groupes ( $F(1, 660) = 76,023, p < .0001$ ). Nous remarquons également que le type de voyelle semble avoir une influence sur les durées des segments ( $F(2, 660) = 15,448, p < .0001$ ). La voyelle ouverte [a] tend à être plus longue que les voyelles fermées [i] et [u]. Le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [a, u], avec des p-value égales à  $p < .0001$ . En outre, nous notons que pour nos deux groupes, les durées vocaliques sont plus longues en contexte d'occlusives voisées ( $F(1, 660) = 100,874, p < .0001$ ). Néanmoins, notre ANOVA ne montre aucun effet de la position dans le logatome sur la durée des voyelles ( $F(1, 660) = 0,172, p = .6783$ ). Finalement, nous notons qu'une unique interaction entre nos variables indépendantes présente un seuil de significativité inférieur à 5%. Ainsi, seule l'interaction entre les facteurs « voyelle » et « voisement » est significative ( $F(2, 660) = 3,216, p = .0408$ ).



**Figure 6.23** – Valeurs moyennes des durées vocaliques (ms) en fonction des sujets, du type de voyelle, du contexte de voisement et de la position dans le logatomes (les barres d'erreurs représentent les écarts-types) ( $n = 684$ ).

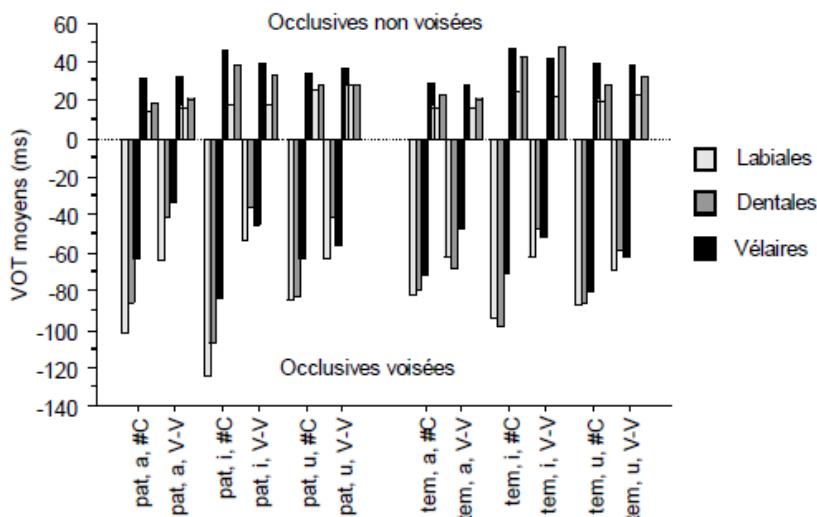
**Valeurs moyennes de  $F_0$  au début des voyelles adjacentes.** La figure 6.24 présente les valeurs moyennes de  $F_0$  au début des voyelles en fonction des deux groupes, du type de voyelle, du contexte de voisement et de la position dans les logatomes. Nous constatons que dans tous les cas, les valeurs moyennes de  $F_0$  des voyelles des patients sont toujours supérieures à celles de notre cohorte de témoins ( $F(1, 696) = 304, 07, p < .0001$ ). De plus, nous remarquons que le type de voyelle semble avoir une influence majeure sur les valeurs moyenne de  $F_0$  ( $F(2, 696) = 5, 139, p = .0061$ ). La  $F_0$  des voyelles hautes est toujours plus élevée que la  $F_0$  de la voyelle basse. Ainsi, le test PLSD de Fisher montre des différences significatives entre les voyelles [a, i] et [a, u] avec des p-value respectivement égales à  $p = .0361$  et  $p = .0017$ . En outre, nous remarquons que pour nos deux groupes, les valeurs moyennes de  $F_0$  sont plus importantes en contexte non voisé ( $F(1, 696) = 89, 957, p < .0001$ ). Par ailleurs, nous notons un effet minime (non significatif) de la position dans le mot sur les valeurs moyennes de  $F_0$  ( $F(1, 696) = 3, 756, p = .0530$ ). Néanmoins, pour cette analyse, nous ne retrouvons aucune interaction entre les différents facteurs.



**Figure 6.24** – Valeurs moyennes de  $F_0$  (en Hz) au début des voyelles en fonction des sujets, du type de voyelle, du contexte de voisement et de la position dans le logatomes (les barres d'erreurs représentent les écarts-types) ( $n = 720$ ).

**Voice Onset Time.** La figure 6.25 présente les valeurs moyennes du VOT des occlusives en fonction de nos deux populations, des contextes vocaliques, du caractère voisé ou non des consonnes, de leurs lieux d'articulation et de la position dans les logatomes. Comme attendu, et ce pour l'ensemble de nos sujets, les valeurs moyennes du VOT des occlusives non voisées restent positives, ce qui implique un établissement du voisement post-relâchement. Au contraire, les occlusives voisées de nos sujets présentent un VOT négatif, ce qui montre que le voisement précède l'explosion. Ainsi, l'ANOVA met en évidence une différence hautement significative entre les deux catégories de consonnes, et ce pour nos deux populations ( $F(1, 662) = 1789, 162, p < .0001$ ). En outre, nous constatons que pour nos deux groupes, les valeurs du VOT augmentent à mesure que l'on recule dans la cavité buccale, en particulier pour les occlusives non voisées. En ce qui concerne les occlusives voisées, cette tendance n'est pas vérifiée. Quoi qu'il en soit, notre test ANOVA montre un effet hautement significatif des lieux d'articulation des consonnes occlusives sur les valeurs moyennes du VOT ( $F(5, 612) = 511, 026, p < .0001$ ). Par ailleurs, nous remarquons que les VOT des consonnes initiales tendent à être plus longs que ceux de la position intervocalique. Ainsi, l'ANOVA met en évidence une différence significative liée à la position dans le mot ( $F(1, 612) = 101, 949, p < .0001$ ). Néanmoins, bien que les valeurs moyennes du VOT semblent légèrement supérieures en contexte de voyelles hautes, le test statistique ne montre aucun effet significatif des voyelles environnantes sur les durées

du VOT des consonnes ( $F(2, 612) = 0,652, p = .5213$ ). De la même manière, nous ne retrouvons aucun effet significatif du groupe sur les valeurs moyennes du VOT des consonnes ( $F(1, 612) = 1,946, p = .1635$ ). Enfin, nous pouvons noter que deux interactions sur l'ensemble de nos variables indépendantes présentent un seuil de significativité inférieur à 5%. Premièrement, l'interaction entre les facteurs « voyelle » et « lieu d'articulation » est significative ( $F(10, 612) = 3,065, p = .0008$ ). Deuxièmement, l'interaction entre les facteurs « position » et « lieu d'articulation » est également significative ( $F(5, 612) = 21,722, p < .0001$ ).



**Figure 6.25** – Valeurs moyennes du VOT des occlusives (ms) en fonction des sujets, du type de voyelle, du lieu d'articulation et de la position dans le logatomes ( $n = 684$ ).

Le tableau 6.18 résume les principaux résultats des tests ANOVA pour les paramètres des durées consonantiques — y compris les durées des tenues et des relâchements —, des intensités relatives des relâchements, des durées vocaliques, des  $F_0$  au début des voyelles adjacentes et des VOT.

			Durée Cons.	Durée Tenue	Durée relâchement	Int. Relative	Durée Voy.	F <sub>0</sub> Onset-V	VOT
Groupe	patients > témoins		*	ns	ns	ns	**	**	ns
Type de voyelle	[i,u] > [a] (excepté pour Durée-V)		*	ns	**	**	**	*	ns
Contexte	[–voix] > [+voix] (excepté pour Durée-V)		**	ns	**	*	**	**	**
Position	[#—] > [V—V] (excepté pour IR)		na	na	na	**	ns	ns	**

**Table 6.18** – Tableau récapitulatif des différents tests ANOVA pour les paramètres acoustiques étudiés (où « \*\* » note des différences hautement significatives ( $p < .0001$ ), « \* » note des différences significatives ( $p < .05$ ), « ns » signifie « non significatif » et « na » signifie « non applicable »).

### 6.6.3 Discussion

**Les durées segmentales.** Pour cette partie de la discussion, nous regrouperons nos résultats concernant les durées consonantiques et vocaliques car nous pensons qu'il existe une influence mutuelle entre ces deux paramètres. Dans cette étude, nous avons pu mettre principalement en évidence que les durées segmentales moyennes étaient dépendantes du contexte de voisement, du type de voyelle et de nos deux populations.

Nous avons montré que les voyelles étaient allongées devant une consonne voisée alors que les occlusives — et plus largement les phases de tenue et de relâchement — étaient plus brèves lorqu'elles étaient voisées. Pour résumer, les consonnes [p, t, k] sont longues mais les voyelles adjacentes sont brèves tandis que les consonnes [b, d, g] sont brèves mais les voyelles subséquentes sont plus longues. Ces résultats rejoignent les études de Lisker (1972), Port (1979), Wajskop (1979) ou Abdelli-Beruh (2004), entre autres. Selon les auteurs, ce constat peut-être lié soit à une interprétation perceptive, soit à une interprétation d'ordre articulatoire ou à une contrainte aérodynamique. La première théorie repose sur la co-variance des durées consonantiques et vocaliques pour optimiser la discrimination perceptive du trait de voisement ou, plus globalement de la sonorité des segments (cf. Kavitskaya 2001 pour l'exploitation de cette hypothèse dans le cadre de l'allongement compensatoire). Ainsi, les modifications des durées vocaliques en fonction du voisement des consonnes serviraient à augmenter le caractère distinctif des durées de l'occlusion pour l'opposition de voisement. Les auteurs partisans de cette théorie précisent que la durée vocalique est systématiquement inversée par rapport à la durée de l'occlusion de la consonne afin de rendre cette dernière perceptivement plus saillante. Il y a donc là un rapport mutuel de cause à effet qui permettrait une meilleure discrimination du trait de voisement (Lisker 1957, Raphael 1972, Kluender & al. 1988). Luce et Charles-Luce (1985), quant à eux, s'accordent sur la variation systématique des durées vocaliques pour l'opposition de voisement en position finale. Néanmoins, ces derniers précisent que la durée de l'occlusion est un

paramètre moins pertinent pour distinguer les deux catégories de consonnes en syllabes finales et dans un environnement dévoisé, ce qui tend à aller à l'encontre de cette première théorie de co-variance puisqu'ils prônent l'importance des durées vocaliques au détriment des durées des occlusions. La seconde théorie explique ces variations par des différences de tensions articulatoires pour la production des occlusives voisées et non voisées. Pour certains auteurs, le fait que les consonnes non voisées soient plus longues que leurs contreparties voisées s'explique par une augmentation de la tension musculaire au niveau supraglottique pour la production des consonnes non voisées (Fisher-Jorgensen 1972). Pour d'autres, c'est la vitesse de fermeture du conduit vocal qui influe sur les durées segmentales. Dans une étude sur le français, l'anglais, le russe et le coréen, Chen (1970) a montré que la fermeture du conduit vocal était plus rapide lors de la production des consonnes [p, t, k]. L'auteur ajoute que la vitesse de fermeture a également une répercussion sur les durées des voyelles environnantes. Finalement, ces deux explications convergent vers une conclusion similaire. La tension musculaire des articulateurs requise pour la production des occlusives non voisées est plus élevée, ce qui résulte d'une fermeture du conduit vocal plus rapide mais d'un accroissement de la durée de l'occlusion associé à une réduction de la durée vocalique. La troisième théorie, concernant la durée des consonnes, est directement liée à une contrainte aérodynamique. Nous avons déjà montré que la différence entre occlusives voisées et non voisées est essentiellement due à une différence de pression transglottique. Ainsi, les consonnes voisées nécessitent le maintien d'un différentiel de pression entre la PSG et la PIO afin que la pression transglottique, qui permet la vibration entretenue des plis vocaux, perdure (Ohala 1983). En effet, sous le seuil de 1 à 2 hPa, la vibration cesse du fait d'un flux d'air trop peu puissant. Cette contrainte joue donc en défaveur d'une longue durée des consonnes voisées, contrairement aux consonnes non voisées, pour lesquelles le passage de l'air dans le conduit vocal est libre du fait de l'abduction des plis vocaux.

En ce qui concerne les durées consonantiques, nous avons établi une analyse plus détaillée des phases de tenue et de relâchement des consonnes de nos deux populations. Ces résultats sont remarquables car ils révèlent une différence notable avec les résultats de la littérature. En effet, les principales études mettent en évidence un rôle majeur de la durée de l'occlusion dans l'opposition acoustique et perceptive du trait de voisement (Lisker 1972, Port 1979, Wajskop 1979, Ohala 1983, Fowler 1991, Abdelli-Beruh 2004, entre autres). Nous retrouvons, de manière non significative, cette tendance à l'allongement de cette partie de l'occlusive dans un contexte non voisé. Selon nous, le paramètre qui semble le plus robuste pour la distinction du voisement est la durée du relâchement. Effectivement, nos données montrent un allongement significatif de cette dernière en contexte non voisé et ce, plus particulièrement pour les témoins. En effet, pour ces derniers, lorsque la phase de d'occlusion est abrégée, la phase de relâchement est allongée par compensation. Serniclaes & Bejster (1974) avait déjà proposé qu'en position intervocalique, la durée de l'occlusion soit allongée pour pallier un relâchement plus bref. Il n'est donc pas exclu que le contraire soit également attesté. Comme nous le précisons précédemment, peu d'études font référence à l'importance de cet indice pour la distinction entre occlusives voisées et occlusives non

voisées. Ainsworth (1973) fut l'un des premiers à souligner l'intérêt de la durée du relâchement pour l'opposition de voisement. Après avoir présenté des bruits synthétiques de différentes durées à une cohorte d'auditeurs, l'auteur a conclu à trois phases perceptives : le bruit le plus long était perçu comme la fricative [s], un bruit moyen était identifié comme un [t] et un bruit bref était perçu comme une voisée [d]. Aussi, Van Alphen (2007) montre qu'en néerlandais, les mesures du relâchement (durée, centre de gravité, énergie) constituent de robustes indices acoustiques de voisement. Selon l'auteur, les occlusives phonologiquement voisées présentent des mesures toujours significativement plus basses que leurs contreparties phonologiquement non voisées. De même, Shrager (2006), qui étudie la neutralisation des occlusives voisées du russe en position finale, souligne que la durée du relâchement des occlusives voisées est statistiquement différente de la durée du relâchement des occlusives non voisées. Celle-ci explique cette tendance par un mécanisme aérodynamique : lorsque les plis vocaux sont ouverts, l'air peut s'échapper librement. En conséquence, nous observons une accumulation plus importante de la PIO derrière l'occlusion, ce qui affecte directement l'intensité et la durée du relâchement. Puisque l'augmentation de la durée du relâchement est liée à l'ouverture des plis vocaux, celle-ci peut être associée à la production et à la perception des segments non voisés.

En outre, nous avons démontré que les durées segmentales — consonnes et voyelles — de nos patients étaient statistiquement plus élevées que celles de nos témoins. De manière générale, une augmentation des durées syllabiques est directement imputable à un débit ralenti (pour une revue de la littérature, voir Zellner 1998). Même si l'âge plus avancé des patients pourrait être une explication à ce constat, nous tenterons d'expliquer ces résultats d'une manière différente. Bien que la voix après cordectomie et la voix chuchotée supposent deux modes phonatoires totalement différents, nous proposerons une comparaison transversale. En effet, nos conclusions et les différentes études sur la voix chuchotée semblent converger vers des hypothèses similaires. Malgré le fait que les conclusions concernant les valeurs de PIO en voix chuchotée soient largement discutées par les auteurs, ces derniers s'accordent sur le fait que le DAB est toujours supérieur par rapport à la voix modale (Schwartz 1972, Stathopoulos & al. 1991). Cette hausse du DAB est liée à la configuration glottique particulière adoptée, à savoir une fermeture glottique avec une ouverture triangulaire entre les arytenoïdes. Dans notre étude aérodynamique, nous avons également mis en évidence des hausses du DAB des patients par rapport aux témoins, en raison des fuites glottiques conséquentes aux changements anatomico-physiologiques engendrés par la chirurgie (cf. section 6.5.2). De plus, et en relation avec la présente étude, les auteurs notent une augmentation générale des durées segmentales en voix chuchotée, avec un maintien des différences de durées segmentales pour l'opposition de voisement (Schwartz 1972, Parnell & al. 1977, Vercherand 2010, Meynadier & Gaydina 2012). Dans une étude sur les productions de phrases cadres de type « VCV saw VCV with VCV » où V= [a, i] et C=[p, b, m] de 12 hommes anglais en phonation normale et chuchotée, Schwartz (1972) a montré que les labiales orales étaient plus longues de 14 ms en voix chuchotée par rapport à la phonation normale. Selon lui, cet allongement consonantique résulterait d'un principe d'économie de conservation respiratoire (« *Respiratory Conservation*

*Hypothesis* ») afin de compenser les importants débits d'air mis en évidence pour ce mode phonatoire. Ainsi, la prolongation des gestes articulatoires serait favorable à un ralentissement de la déperdition d'air. Par ailleurs, dans une étude sur 20 phrases prononcées par 10 locuteurs anglais américains et composées des occlusives alvéolaires [t, d, n, s, z] en contexte [a, i] en phonation normale et chuchotée, Parnell & al. (1977) ont démontré une augmentation des durées consonantiques associée à une augmentation des durées vocaliques en voix chuchotée. Dès lors, les auteurs ont rejeté l'hypothèse de Schwartz (1972), jugée contradictoire. Sur les principes de cette théorie, Parnell & al. (1977) précisent que les voyelles — qui présentent les débits d'air les plus importants — devraient être réduites afin de conserver un maximum d'air phonatoire. Finalement, pour ces auteurs, l'augmentation des durées segmentales serait la conséquence d'une adaptation des locuteurs qui ralentiraient leurs discours afin de les rendre plus intelligibles (« *Intelligibility Hypothesis* »). En ce qui nous concerne, nous pensons que cette seconde hypothèse peut être une explication valable aux différences de durées mises en évidence entre nos deux populations. En effet, il n'est pas exclu que nos patients ralentissent volontairement leur débit afin de rendre les logatomes plus intelligibles, que ce soit pour la distinction du voisement ou des lieux d'articulation. Cette stratégie serait mise en place dans le but de compenser un handicap vocal qu'ils imaginent perdurer.

Enfin, nous avons mis en évidence le fait que les durées vocaliques variaient en fonction du type de voyelle pour nos deux populations. La voyelle ouverte tend à être plus longue que les voyelles fermées, soit  $D[a] > D[i, u]$  (où  $D$ = durée). Cet ordre est prévisible et dépend de la durée intrinsèque des voyelles. Cette notion universelle a été largement exploitée dans les langues et prévoit que ces variations sont liées au degré d'aperture et à la présence du trait de nasalité dans les systèmes étudiés. En anglais par exemple, nous pouvons citer les travaux de House & Fairbanks (1953) ou Peterson & Lehiste (1961), qui retrouvent une différence de durée de 50% entre les voyelles hautes et les voyelles basses. Pour le français standard, Di Cristo (1985) relève un rapport de 20% environ entre les durées des voyelles [i, u, y] et [ɛ, ɔ, a]. L'auteur note que l'écart augmente de 22% si l'on restreint l'analyse à la comparaison des voyelles hautes [i, u] avec la voyelle basse [a]. Les résultats pour le français québécois suivent des conclusions similaires et montrent une augmentation des durées de 32% entre [i, u] et [ɛ] ainsi qu'une augmentation de 70% si l'on compare les voyelles hautes avec la voyelle basse [a] (O'Shaughnessy 1981). En ce qui concerne le trait de nasalité, les voyelles nasales tendent à être plus longues que leurs correspondantes orales. Ainsi, Di Cristo (1985) met en évidence un écart de 36% entre la durée des nasales et [a] et un écart de 73% entre les nasales et [i]. De même, Bartkova & Sorin (1987) retrouvent, dans une moindre mesure, cette tendance avec un écart de 17% entre voyelles nasales et voyelles orales. Finalement, nous remarquons que malgré des différences d'une recherche à l'autre, toutes les conclusions convergent vers une même tendance : les durées sont proportionnelles à l'aperture buccale et inversement proportionnelles à la hauteur de la langue (House & Fairbanks 1953). Ainsi, les voyelles hautes sont plus brèves que les voyelles basses, et le trait de nasalité engendre un allongement des durées vocaliques.

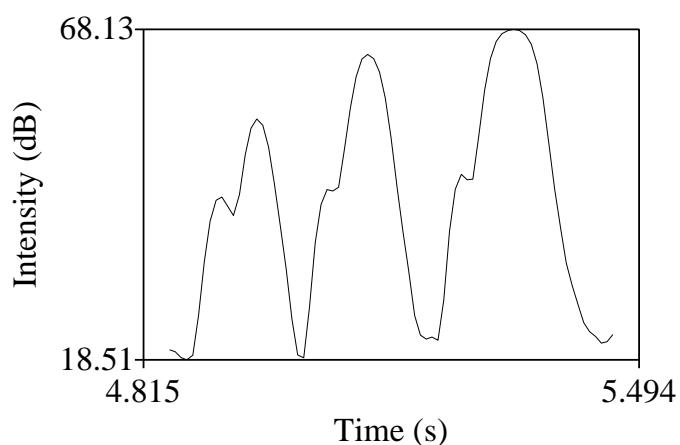
**Intensité relative du relâchement.** Nous avons déjà précisé que l'intensité relative du relâchement représentait la différence entre le pic d'amplitude maximal pendant le relâchement et le pic d'intensité de la voyelle suivante. Des valeurs proches de 0 indiquent que les intensités du relâchement et de la voyelle sont proches, c'est-à-dire que l'intensité du relâchement de la consonne est importante. Au contraire, des valeurs positives et élevées indiquent que le relâchement de l'occlusive est largement moins intense que la voyelle subséquente. En d'autres termes, plus une valeur est positive et plus l'intensité du relâchement de la consonne est faible par rapport à la voyelle suivante. Pour ce paramètre, nous avons pu mettre en évidence une influence du contexte de voisement, du type de voyelle et de la position. Au contraire, le facteur « groupe » ne semblait pas interférer sur les valeurs de l'intensité relative du relâchement, ce qui implique un contrôle total de ce paramètre de nos patients pour l'opposition de voisement.

De manière générale, nous notons des valeurs de l'intensité relative du relâchement plus élevées en contexte voisé. Toute chose égale par ailleurs, ce constat implique une intensité du relâchement plus importante pour les occlusives non voisées par rapport à leurs contreparties voisées. Ces résultats vont dans le sens des conclusions de Zué (1976), Repp (1979), Flack (2007) ou Röttger & al. (2011). Par exemple, Zué (1976), qui examine, entre autres, l'intensité relative du relâchement des consonnes de l'anglais américain, a montré une différence d'environ 2 dB entre occlusives voisées et non voisées. De même, Flack (2007), qui étudie les différences de mesures de VOT et d'intensité relative du relâchement des occlusives labiales et dentales du français en position initiale et médiane, a également mis en évidence cette élévation de l'intensité pour les non voisées. En position initiale, l'auteur relève une différence pour les labiales et les dentales de 2 et 4 dB respectivement. En position intervocalique, cette différence se porte à 4 dB pour les labiales et à 5 dB pour les dentales. Röttger & al. (2011) notent également un effet significatif du voisement sur l'intensité du relâchement en allemand puisque celui-ci tend à être 1,24 dB plus haut pour les occlusives non voisées. Comme nous l'avons évoqué plus haut, cette différence est directement liée à une contrainte aérodynamique. Pour la production des occlusives non voisées, les plis vocaux sont en abduction, ce qui permet à l'air de s'échapper plus librement. En conséquence, nous notons une accumulation de la PIO plus importante en arrière de la constriction. Ainsi, l'intensité plus importante au relâchement de l'occlusive non voisée est directement liée à une pression buccale plus importante.

Nous remarquons également un effet significatif du type de voyelle. Les voyelles hautes tendent à éléver les valeurs de l'intensité relative du relâchement. Comme nous l'expliquions précédemment, il faut inverser ce constat pour obtenir les tendances naturelles. Finalement, les relâchements des occlusives sont plus intenses en contexte [a], aussi bien pour les témoins que pour les patients. Nous pouvons corrélérer ces observations avec les notions d'intensité intrinsèque des voyelles ou de Umax que nous avons discutées dans les sections 5.5.3.1 et 6.5.4.3. La voyelle [a] est produite avec une faible impédance par rapport aux voyelles hautes, ce qui lui confère d'importants débits d'air et plus largement une intensité supérieure. En conséquence, il n'est pas étonnant que l'anticipation des voyelles basses s'effectue de manière plus relâchée,

avec une accumulation plus importante de l'air derrière la constriction, ce qui tend à augmenter l'intensité globale du logatome dans ce contexte.

Enfin, nous avons observé que la position initiale engendrait des valeurs d'intensité relative du relâchement des consonnes plus importantes par rapport à la position inter-vocalique. Suivant la logique selon laquelle des valeurs élevées impliquent de faibles relâchements et *vice versa*, nous pouvons en conclure que les relâchements tendent véritablement à être plus forts en position médiane. Ce constat est contraire à nos attentes dans le sens où la position initiale reste une position forte tant du point de vue articulatoire que de l'amplitude. De plus, il nous semble normal de supposer que la PSG tend à décroître pendant le logatome, ce qui suggère une pression plus importante et, *in extenso*, une intensité plus forte à l'initiale absolue par rapport aux autres positions. Nos conclusions peuvent néanmoins être expliquées par le phénomène de lecture énumérative qui favorise un *pattern* inversé de l'intensité par rapport à l'assertion type (cf. figure 6.26).



**Figure 6.26** – Exemple d'une courbe d'intensité (dB) pour le logatome [tititit] du témoin n°2 (Praat).

**Valeurs moyennes de  $F_0$  au début des voyelles adjacentes.** Nous avons mis en évidence que les valeurs moyennes de  $F_0$  au début des voyelles variaient en fonction du contexte de voisement, de nos deux groupes et du type de voyelle.

A l'instar des études de House & Fairbanks (1953), Mohr (1968), Ohde (1984) ou Hanson (2009), entre autres, nous avons montré que le voisement de la consonne prévocalique avait une influence sur la  $F_0$  au début des voyelles suivantes. Ainsi, la  $F_0$  tend à être significativement plus élevée quand la voyelle est précédée d'une occlusive non voisée. Comme nous l'avons évoqué précédemment, cette tendance est également attestée en tonologie où les consonnes non voisées élèvent les tons des voyelles adjacentes alors que les consonnes voisées abaissent les tons voisins (Hyman & Schuh 1974, Hombert & al. 1979). Cette dichotomie peut s'expliquer soit par une contrainte

articulatoire, soit par une contrainte aérodynamique. Notre première hypothèse se justifie par une différence de tension des plis vocaux pour la production des occlusives voisées et non voisées. Stevens (1991) note qu'un état neutre du larynx est impossible et que les plis vocaux doivent être tendus ou relâchés pour marquer l'opposition de voisement. Dans ce sens, Halle & Stevens (2002 [1971]) précisent que la  $F_0$  des voyelles est modifiée car les plis vocaux doivent être tendus pour la production des occlusives non voisées afin d'inhiber les vibrations laryngées alors que ces derniers doivent être relâchés pour maintenir la vibration glottique nécessaire au voisement des occlusives voisées. Stevens (1977, 1998) ajoute qu'une position verticale particulière du larynx doit être associée afin de faciliter l'état de raideur ou de relâchement des plis vocaux. Ainsi, celui-ci note qu'une position haute du larynx permet une diminution de volume des cavités supraglottiques, ce qui permet d'allonger les plis vocaux et d'accroître leur raideur. A l'inverse, une position basse du larynx permet un plus grand relâchement des plis vocaux et des parois laryngées, ce qui facilite le maintien du voisement. D'un point de vue musculaire, Löfqvist & al. (1989) ont mis en évidence une augmentation de l'activité du muscle crico-thyroïdien pour la production des occlusives non voisées par rapport aux occlusives voisées. Ce muscle a pour principale fonction d'accroître la tension des plis vocaux et d'inhiber le voisement. La seconde hypothèse que nous soulevons est directement liée aux contraintes aérodynamiques. Lors de la production des occlusives voisées, les plis vocaux opposent une résistance au passage de l'air, ce qui tend à réduire la pression transglottique. Au relâchement de la consonne, cette diminution de pression transglottique est responsable d'un abaissement de la  $F_0$  pour ce type de consonne et, par la même, au début de la voyelle subséquente. En ce qui concerne la production des occlusives non voisées, nous avons déjà expliqué que l'air ne rencontre aucun obstacle et peut donc s'écouler naturellement dans le conduit vocal. Ainsi, le flux d'air au relâchement est beaucoup plus élevé, ce qui est à l'origine d'un effet de Bernouilli plus important et donc d'une vibration plus rapide des plis vocaux. Cette augmentation des vibrations glottiques est responsable de l'élévation de la  $F_0$  au début de la voyelle suivante (Homber & al. 1979).

En outre, nous avons montré que ces valeurs étaient influencées par le type de voyelle. La  $F_0$  des voyelles hautes est plus importante que celle de la voyelle basse [a]. Ces résultats ne sont pas étonnantes car directement liés à la  $F_0$  intrinsèque des voyelles que nous avons déjà pu discuter (cf. section 5.3.3.1).

Enfin, nous remarquons que les valeurs de  $F_0$  diffèrent d'une population à une autre. Nous notons que les mesures de  $F_0$  de nos patients sont significativement plus élevées que celles de notre cohorte de témoins. Une fois encore, ces résultats étaient attendus car directement imputables aux changements anatomico-physiologiques engendrés par la chirurgie. En effet, nous avions déjà expliqué que le laser reséquait une partie du pli vocal, ce qui tendait à diminuer sa masse vibrante et à rendre ce dernier plus tendu. Or, une tension accrue des plis vocaux a pour conséquence directe une augmentation de la  $F_0$  (cf. section 5.3.3.1).

**Voice Onset Time.** Comme défini précédemment, le VOT représente le délai de la mise en mouvement des plis vocaux. Pour Lisker & Abramson (1964), cet indice met

en évidence la synchronisation entre les gestes glottiques et supraglottiques. Pour les principaux auteurs, le VOT dépend de règles aérodynamiques, de la coordination des articulateurs, de l'étendue du contact des articulateurs mais est également influencé par le débit et le contexte vocalique.

De notre point de vue, nous avons uniquement montré un effet des lieux d'articulation des consonnes et de la position dans les logatomes. Contrairement à ce que nous attendions, même si les voyelles hautes tendent à allonger légèrement les valeurs du VOT, nous ne retrouvons aucun effet significatif du contexte vocalique. En effet, nous avions déjà précisé que l'anticipation d'une voyelle haute permettait un élargissement plus rapide de la cavité buccale et, par extension, un voisement plus long. De même, les valeurs moyennes du VOT de la cohorte des témoins et des patients ne semblent pas différentes. La majorité des recherches qui s'intéressent à l'évolution de ce paramètre pour les voix pathologiques reportent des valeurs plus importantes ou instables par rapport aux sujets témoins. Ces instabilités sont directement liées à une difficulté dans le contrôle de la coordination des événements laryngés et supralaryngés (voir par exemple Hirsch 2007 pour les sujets bégues, Duez 2007 pour les parkinsoniens). Nos patients semblent maîtriser totalement cet aspect du *timing* laryngé nécessaire à l'opposition du voisement, malgré un handicap vocal certain lié à un accolement partiel des plis vocaux consécutif à la chirurgie. Nous pouvons émettre l'hypothèse que ce voisement est initié et maintenu par une compensation aérodynamique liée à une pression transglottique plus élevée (cf. section 6.5.4).

Comme nous l'évoquions précédemment, le lieu d'articulation des consonnes présente un effet hautement significatif sur les valeurs du VOT et ce, pour nos deux populations. Les valeurs du VOT augmentent avec le recul de l'articulation dans la cavité buccale. Ainsi, le VOT est plus important pour les consonnes vélaires par rapport aux dentales et aux labiales (Peterson & Lehiste 1961, Lisker & Abramson 1964, Cho & Ladefoged 1999). Ce constat est vérifié pour les occlusives non voisées mais plus variable pour les occlusives voisées, ce qui avait déjà été souligné par Serniclaes (1987). Nous ne rentrerons pas dans les détails concernant les explications articulatoires et aérodynamiques qui justifient ces conclusions car celles-ci ont déjà été discutées dans la section 6.3.1. Rappelons juste brièvement que les vélaires présentent un VOT plus important par rapport aux autres consonnes du fait d'une production nécessitant un contact linguo-palatal plus étendu (Stevens 1998), un mouvement des articulateurs — en l'occurrence la masse linguale — plus lent (Hardcastle 1973) et qui entraîne ainsi une cavité réduite en arrière de la constriction (Maddieson 1997).

Enfin, nous avons pu noter que l'ensemble des sujets présentent des VOT plus longs en position initiale par rapport à la position intervocalique. Nos conclusions rejoignent celles des études de Pierrehumbert & Talkin (1992), Jun (1993) ou Flack (2007), entre autres. En effet, ces différents auteurs mettent en évidence que les consonnes initiales et médianes sont prononcées différemment, tant du point de vue du domaine du syntagme prosodique le plus large que du simple mot. Ainsi, Pierrehumbert & Talkin (1992) ont montré que l'occlusive dentale non voisée [t] du mot anglais « tomahawk » avait un VOT plus long en début de syntagme intonatif par rapport à une position médiane dans le syntagme. De même, Jun (1993) a mis en évidence que le VOT de

l’occlusive sourde aspirée [p<sup>h</sup>] du coréen était plus long à l’initial par rapport à la position médiane. Flack (2007), dans une étude sur le français, relève une différence de 5 ms entre les VOT de l’occlusive dentale [t] selon les deux types de positions. Finalement, les consonnes initiales semblent être plus longues et plus fortes d’un point de vue articulatoire (Byrd 1994, Keating & al. 1999). Jun (1993) ajoute que le geste glottique est renforcé dans cette position, ce qui implique une plus grande et une plus longue ouverture glottique pour la production des consonnes initiales.

#### **Synthèse de l’étude n°7 : études acoustiques**

- Objectifs de l’étude : analyser de manière quantitative certains paramètres acoustiques.
- Corpus : logatomes de type CVCVCV où C= [p, t, k, b, d, g] et V= [a, i, u].
- Principaux résultats :
  - élévation des durées consonantiques, de l’intensité du relâchement des consonnes et de la F<sub>0</sub> au début des voyelles en contexte non voisé. Les durées vocaliques sont plus importantes en contexte voisé,
  - les paramètres sont plus importants en contexte de voyelles hautes, excepté les durées vocaliques qui sont plus longues pour la voyelle [a],
  - les durées segmentales et la F<sub>0</sub> au début des voyelles sont plus importantes pour les patients par rapport aux témoins,
- Conclusions :
  - résultats acoustiques des patients satisfaisants et qui traduisent une bonne synchronisation entre les gestes laryngés et supralaryngés,
  - le fait qu’il existe des différences significatives dans les mesures entre contexte voisé et non voisé suggère que l’opposition de voisement est maintenue chez les sujets dysphoniques.

## **6.7 Etude n°8 : études électroglottographiques (EGG)**

Après avoir défini l’EGG et les paramètres qui lui sont corrélés, nous analyserons le quotient ouvert des voyelles tenues des patients. Notre seconde étude sera consacrée à la pertinence de l’utilisation de cet outil pour la distinction du trait de voisement des occlusives orales.

### **6.7.1 Préliminaires**

#### **6.7.1.1 Analyse du signal EGG**

L’électroglottographie a été mise au point par Fabre (1956). Les études acoustiques et médicales se sont rapidement développées avec l’apparition des premiers modèles commercialisés au début des années 1970 (Fourcin 1974)

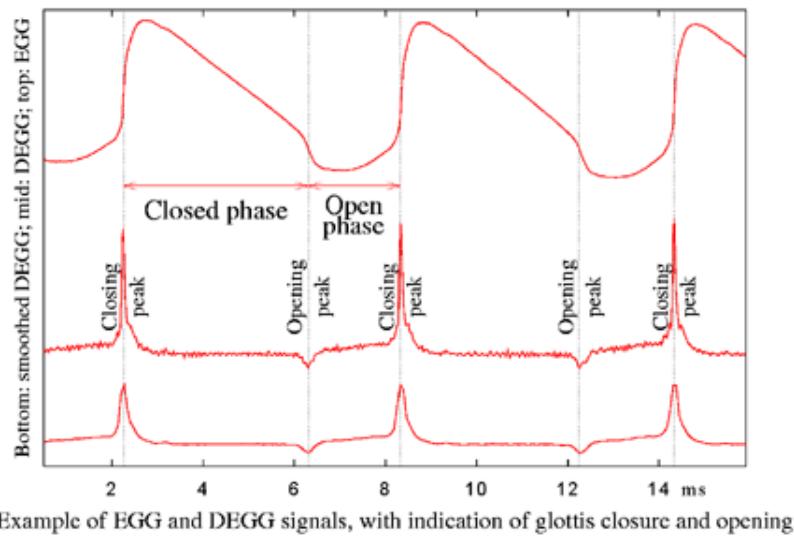
Comme nous l’avons vu dans le chapitre 3, l’EGG est un moyen d’investigation simple et non invasif qui nous apporte des renseignements quant à la surface d’accolement des plis vocaux. Deux électrodes placées de part et d’autre du cou permettent

de mesurer la différence de potentiel liée à la résistance que le courant rencontre lorsqu'il passe d'une électrode à l'autre. Ainsi, ce signal nous renseigne non seulement sur la périodicité de la vibration glottique mais aussi sur la qualité vocale du locuteur (Childers & Krishnamurthy 1985).

Une période du signal EGG est composée de deux phases distinctes qui représentent la fermeture et l'ouverture des plis vocaux. Dans une étude multiparamétrique incluant des mesures de PSG, des données EGG, photoglottographiques et stroboscopiques sur le chien, Berke & al. (1987) ont pu mettre en évidence que les variations dans la courbe EGG traduisaient des phases d'ouverture des plis vocaux le long de leur bord supérieur et des phases de fermeture le long de leur bord inférieur. De manière générale, nous pouvons noter que le signal augmente lorsque la glotte se ferme et qu'il décroît lorsqu'elle s'ouvre.

Le signal EGG renseigne donc sur les instants de fermeture et d'ouverture glottique qui sont caractérisés par des discontinuités de la pente du signal. Certains paramètres liés à l'aire de contact entre les plis vocaux peuvent être extraits à partir de l'EGG (Titze 1990, Baken 1992). Néanmoins, les auteurs précisent que ce signal ne nous renseigne qu'indirectement sur le flux d'air qui passe à travers la glotte, ce qui rend son interprétation délicate (Orlikoff 1998). Par conséquent, les auteurs préconisent l'utilisation de la dérivée de l'EGG (la DEGG) pour le calcul de certains paramètres glottiques — tels le quotient d'ouverture et le quotient de fermeture, les indices de voisement ou la  $F_0$  et ses variations — de façon précise et automatisée (Henrich 2001).

Le signal DEGG montre deux pics distincts par période : un pic positif qui rend compte de l'accélération de la fermeture des plis vocaux, et un pic négatif qui met en évidence l'accélération de l'ouverture de la glotte (cf. figure 6.27). La durée entre deux pics positifs de fermeture glottique, qui correspond à une période, permet le calcul précis de la fréquence fondamentale de la voix. La durée entre le pic négatif d'ouverture glottique et le pic positif de fermeture suivant renvoie à la phase d'ouverture des plis vocaux et permet donc la mesure du quotient ouvert (Oq).



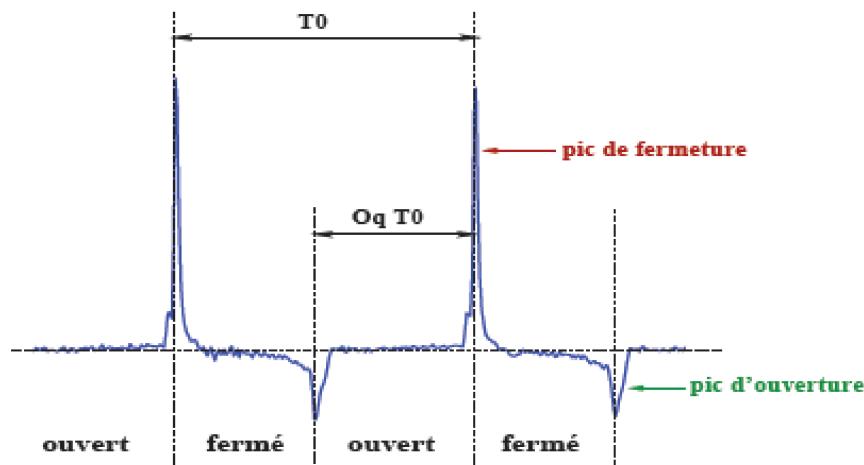
**Figure 6.27** – Exemple d'un signal EGG et de sa dérivée. En haut, le signal EGG, au milieu le signal DEGG et en bas le signal DEGG lissé. Les pics positifs de DEGG montrent une augmentation de l'aire de contact des plis vocaux et les pics négatifs une diminution de l'aire de contact (<http://voiceresearch.free.fr/egg/>).

### 6.7.1.2 Les paramètres glottiques

**Définitions.** Comme nous le précisions plus haut, la fréquence fondamentale, le quotient ouvert ( $Oq$ ) et sa contrepartie inverse, le quotient fermé ( $Cq$ ), sont des paramètres glottiques que nous pouvons mesurer à partir de la dérivée DEGG du signal EGG.

La mesure du  $Oq$  représente le rapport entre la durée de la phase d'ouverture et la période fondamentale (cf. figure 6.28). Ses valeurs varient entre 0 (cycle sans ouverture) et 1 (fermeture incomplète). Au contraire, la mesure du  $Cq$  correspond au rapport entre la durée de la phase de fermeture et la durée du cycle. On le détermine à partir de la mesure de  $Oq$ , soit  $Cq = 1 - Oq$ .

Ces paramètres sont directement liés à l'appréciation de la qualité vocale dans la mesure où ils sont corrélés au caractère pressé ou relâché d'une voix. Toute chose égale par ailleurs, une montée des valeurs de  $Oq$  est liée à un relâchement musculaire, caractérisé par un accolement mou des plis vocaux. Une baisse des valeurs de  $Oq$ , liée à une fermeture très rapide de la glotte, représente un serrage vocal (Sundberg 1995).



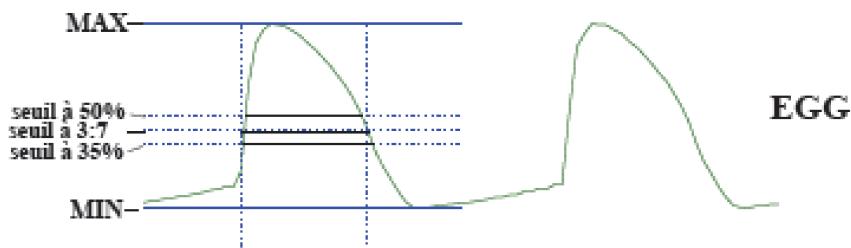
**Figure 6.28** – Signal DEGG présentant des pics d’ouverture et de fermeture, ce qui permet la mesure du fondamental ( $T_0$ ) et du quotient ouvert ( $Oq$ ) (Henrich 2001).

**Estimation du Oq.** A partir de la connaissance des durées d’ouverture de la période, plusieurs méthodes ont été décrites afin d’estimer le quotient ouvert :

La première méthode est basée sur le signal EGG uniquement (cf. figure 6.29) : cette technique nécessite une ligne de base représentant un seuil  $H$  qui détermine la frontière entre les phases de fermeture et d’ouverture. Scherer & al. (1995) ont proposé un seuil de 25 à 35%, validé par des observations stroboscopiques. D’autres études ont établi les instants d’ouverture et de fermeture lorsque le signal équivaut de 35 à 50% de son amplitude totale, la phase d’ouverture se situant pendant la phase de décroissance de l’onde et la phase de fermeture pendant la phase de croissance (Rothenberg & Mahshie 1988)

La seconde méthode, nommée « DECOM » (DEgg Correlation-based method for Open quotient Measurement) est basée sur le signal DEGG : c’est la méthode utilisée par Henrich (2001) dans son travail de thèse. Elle nécessite un signal DEGG propre avec des pics d’ouverture et de fermeture clairs, ce qui n’est pas toujours évident. Cette technique permet une détection de la fréquence fondamentale et des  $Oq$  dans la mesure où elle repose sur la détermination des intervalles entre les pics positifs et négatifs.

La troisième méthode est, quant à elle, basée sur les signaux EGG et DEGG (cf. figure 6.29) : la dérivée permet de détecter de manière fiable les instants de fermeture via les pics positifs. L’ouverture, quant à elle, est déterminée à 3/7 de l’amplitude totale de l’EGG sur une période (Howard 1995).



**Figure 6.29 –** Méthode de mesure du Oq sur le signal EGG par détection des seuils (35, 50%) (Henrich 2001).

## 6.7.2 Etude préliminaire sur le quotient ouvert des voyelles isolées

### 6.7.2.1 Méthodologie

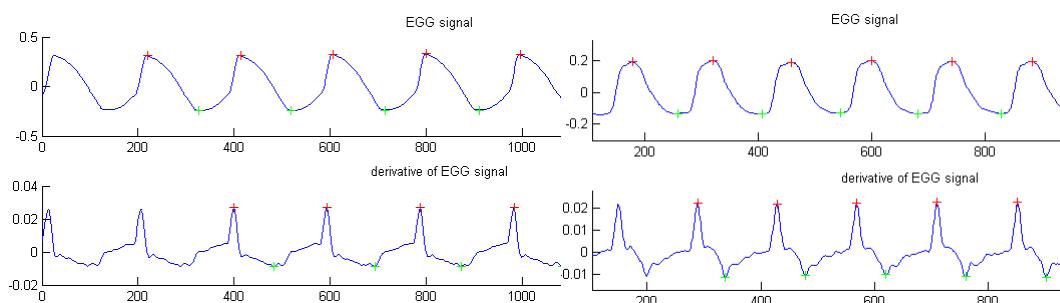
Comme nous le mentionnions dans le chapitre 4, nos données EGG ont été acquises pendant nos enregistrements aérodynamiques à l'aide d'un laryngographe branché à l'appareillage EVA2. Notre idée d'utiliser l'EGG dans cette thèse ayant émergé tardivement, au moment des premières prises de données aérodynamiques des patients, il nous a été impossible de reconvoquer l'ensemble des témoins. Finalement, seul un témoin a pu revenir pour une seconde prise de données, ce dernier nous servira de référence. Pour cette étude préliminaire, il était demandé aux sujets d'inspirer profondément afin d'émettre les voyelles tenues [a, i, u] à hauteur et intensité confortable. Les enregistrements aérodynamiques et EGG ont été effectués à l'aide du logiciel Phone-dit. L'intérêt de cette prise de données réside dans les enregistrements synchronisés du signal audio et du signal EGG.

Les signaux audio et EGG ont été reconvertis dans un format .wav afin d'établir une première segmentation sous Praat. Ce logiciel nous a permis, notamment, de combiner les deux signaux afin d'obtenir une synchronisation visuelle parfaite entre les deux signaux. Afin d'éviter les variations intra-individuelles concernant le début et la fin des voyelles, nous avons étiqueté nos voyelles sur leurs portions les plus stables, c'est-à-dire en sélectionnant une fenêtre représentant 50 ms. Les signaux EGG étiquetés ont ensuite été analysés à l'aide d'un script Matlab utilisant la méthode « DECOM », c'est-à-dire qu'il détecte les pics de la DEGG. Ce script a la particularité d'établir un filtre passe-bande de 5 à 1200 Hz pour éliminer les fréquences basses et les fréquences hautes non associées aux phases d'ouverture et de fermeture des plis vocaux. Le signal est, par la suite, lissé à l'aide d'un filtre de 3ème ordre Butterworth (-18 dB/octave) avec une fréquence de coupure normalisée à 0.054. Ce script permet, notamment, l'extraction des valeurs de Oq et de Cq, ainsi que d'autres paramètres secondaires que nous n'aborderons pas ici. Par mesure de prudence, les fichiers résultats ont été inspectés visuellement. Ainsi, lorsqu'une valeur s'écartait de plus de 10% des valeurs adjacentes, celle-ci était écartée (Di Canio 2009).

Toutes nos prises de mesures ont été statistiquement analysées à l'aide du logiciel Statview. Nous avons effectué un test-t univarié sur les moyennes des valeurs de Oq de nos patients pour chacune des trois voyelles en prenant comme références les moyennes de notre sujet témoin.

### 6.7.2.2 Difficultés rencontrées

**Le problème des doubles pics.** En observant nos données, et particulièrement pour les voix pathologiques, nous nous sommes rendus compte que les signaux EGG et DEGG sont parfois différents des *patterns* standards évoqués précédemment ou illustrés par la figure 6.30 ci-dessous :



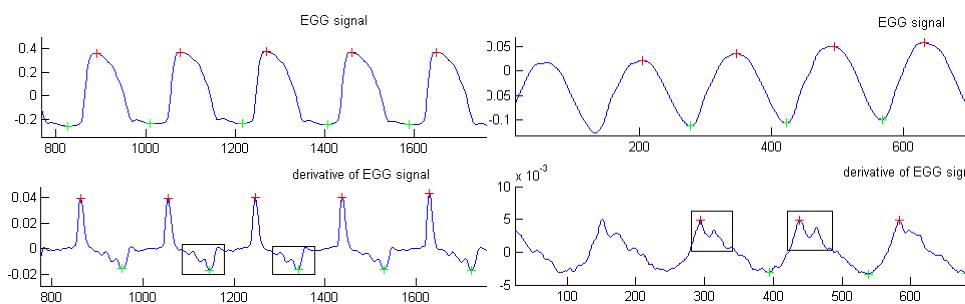
**Figure 6.30 – Exemples typiques de 6 périodes sur des signaux EGG et DEGG pour la voyelle [u].** A gauche, la voyelle de notre témoin de référence et à droite la voyelle du patient n°3.

Les deux sujets présentent bien deux phases distinctes, caractérisées par des pics de fermeture et d'ouverture, qui correspondent aux changements de configurations glottiques pendant l'acte phonatoire. Les cycles semblent périodiques et symétriques, ce qui traduit une adduction suivie d'une abduction des plis vocaux. Néanmoins, le signal EGG du patient a une forme plus arrondie et son amplitude semble moins importante, ce qui correspond à une ouverture glottique peut-être plus longue. De plus, le signal DEGG de notre témoin présente des transitions moins abruptes, plus lisses en comparaison avec le signal DEGG de notre patient. Les phases de fermeture des signaux EGG et DEGG de notre témoin semblent plus longues que celles du patient. Aussi, l'amplitude des pics de fermeture et des pics d'ouverture de la dérivée du patient est quasi identique. Au contraire, l'amplitude des pics de fermeture de la DEGG de notre témoin est beaucoup plus marquée que celle liée aux pics d'ouverture. Finalement, ces signaux se rapprochent des signaux présentés par Henrich (2001) et Henrich & al. (2004) pour la distinction des mécanismes I et II<sup>8</sup>. Les signaux de notre témoin

8. En mécanisme I, les plis vocaux sont épais et vibrent sur toute leur longueur. La masse vibrante et l'amplitude de vibration sont importantes. La durée d'ouverture glottique est souvent plus brève que la phase de fermeture. Au contraire, en mécanisme II, les plis vocaux sont fins et ne vibrent plus sur la totalité de leur longueur. Les phases d'ouverture sont plus importantes (Henrich 2001).

semblent s'identifier aux *patterns* mis en évidence pour le mécanisme I, tandis que les signaux des patients s'assimilent davantage à ceux établis pour le mécanisme II.

Néanmoins, il existe des cas fréquents où ces deux phases ne semblent pas aussi bien déterminées. La figure 6.31 met en évidence deux voyelles pour lesquelles soit le pic d'ouverture est multiple (double), soit le pic de fermeture est doublé (cf. Henrich 2001, Henrich & al. 2004).



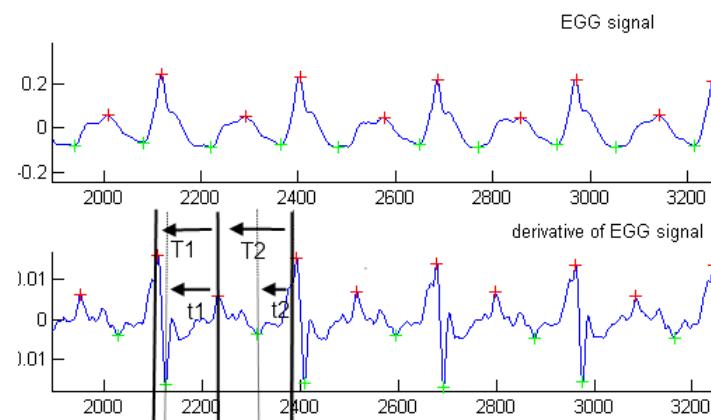
**Figure 6.31 –** Exemples atypiques de 5 périodes sur des signaux EGG et DEGG pour la voyelle [u] qui montrent (i) un double pic d'ouverture et (ii) un double pic de fermeture (patient n°5 et n°8).

Dans les deux cas, nous notons que les amplitudes des deux pics pendant les phases d'ouverture ou de fermeture sont bien distinctes l'une de l'autre. Selon Henrich (2001) ou Henrich & al. (2004), l'apparition de doubles pics peut être liée à deux phénomènes différents : (i) une ouverture ou une fermeture sur deux parties de la glotte de manière simultanée ou déphasée ou (ii) une ouverture rapide des plis vocaux sur leur épaisseur, suivie d'une ouverture sur leur longueur.

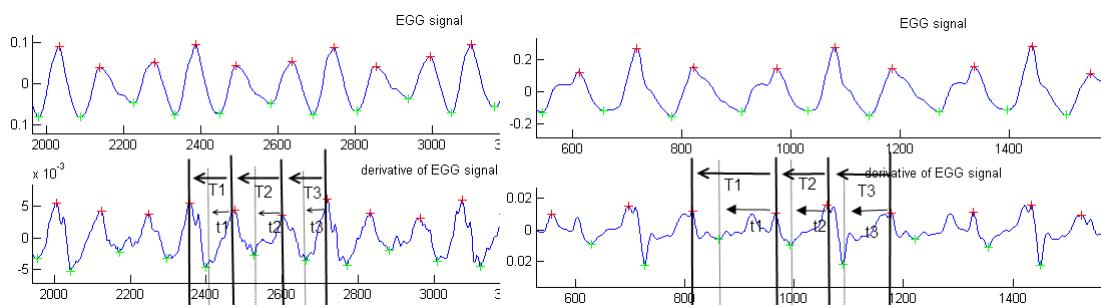
Pour ces exemples imprécis, la méthode « DECOM », qui est basée sur une méthode de corrélation pour la détermination de la distance entre deux pics de fermeture et la distance entre un pic d'ouverture et le pic de fermeture suivant, est préférée. En effet, les scripts Matlab qui utilisent cette méthode, possèdent un algorithme qui détecte automatiquement les doubles pics, ce qui permet une analyse de ces signaux particuliers au même titre que les signaux typiques. En ce qui nous concerne, notre script Matlab élimine les pics erronés, c'est-à-dire ceux dont l'amplitude représente moins de 10% des pics maximums et minimums les plus hauts. Aussi, lorsque deux pics maximums ou minimums consécutifs sont détectés, celui dont l'amplitude est la plus importante est sélectionné par le script. Par conséquent, cette méthode nous a permis d'analyser les valeurs de Oq de huit de nos patients.

**Signaux atypiques.** Nous avons délibérément éliminé les signaux EGG et DEGG de deux de nos patients (patients n°1 et 4) pour notre analyse quantitative. En effet, ces signaux s'écartaient largement du *pattern* commun, ce qui engendrait des problèmes de détection de pics maximums. Ainsi, les figures 6.32 et 6.33 montrent les signaux

DEGG particulièrement inhabituels du patient n°1. De plus, le cas de ce patient semble atypique dans le sens où ses productions vocaliques (enregistrées à l'aide d'un laryngographe sous EVA2) semblent différentes de celles enregistrées pendant la première session acoustique. Les mesures acoustiques de ce patient correspondaient à celles des autres patients (en particulier une  $F_0$  élevée) alors que le signal EGG montre des fréquences assez basses, ce qui va à l'encontre des autres sujets et de notre première étude. En conséquence, nous considérons ces productions comme des cas isolés liés à l'instabilité du vibrateur ou au sujet lui-même. Nous préférons donc une analyse qualitative de ces signaux particuliers dans laquelle nous tentons d'apporter des éléments de réponses quant à ces singularités. Ces hypothèses doivent être, toutefois, considérées avec prudence car nous n'avons pas de données fibroscopiques ou kymographiques pour valider nos postulats.



**Figure 6.32 – Signaux EGG et DEGG acquis sur 9 cycles pour la voyelle [a] qui montrent un doublement de période (T1 et T2) et par conséquent deux temps d'ouverture glottique (t1 et t2) (patient n°1).**

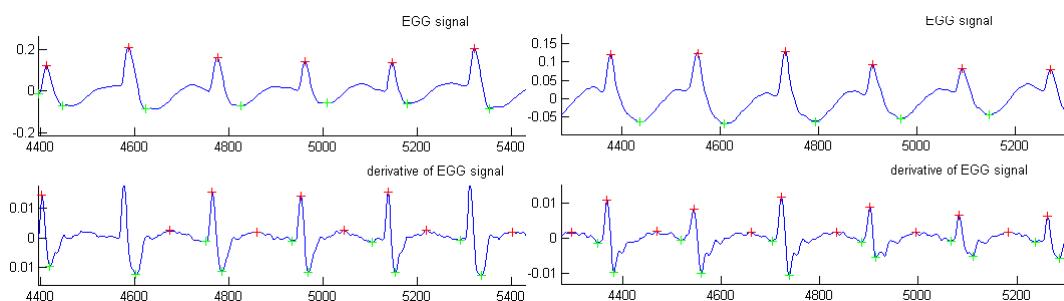


**Figure 6.33 – Signaux EGG et DEGG acquis sur 10 cycles pour la voyelle [i] (à gauche) et sur 8 cycles pour la voyelle [u] (à droite) qui montrent un triplement de période (T1, T2 et T3) et par conséquent trois temps d'ouverture glottique (t1, t2 et t3) (patient n°1).**

Ces signaux EGG et DEGG du patient n°1 mettent en évidence des doublements ou triplements de période. Dans le cas d'une diplophonie ou triplophonie, nous constatons la présence de deux ou trois fréquences distinctes et simultanées — c'est-à-dire de deux ou trois périodes différentes — qui se répètent de façon périodique sur ces signaux. Notre première hypothèse quant à l'apparition de ces doubles ou triples périodes est liée au phénomène de désynchronisation du mode vibratoire. Comme le soulignaient Titze (1994) ou Ishizaka & Flanagan (1972) dans leur modèle à deux masses, une oscillation régulière des plis vocaux repose sur leurs mouvements latéraux, sur une ondulation muqueuse régulière et sur une pression sous-glottique constante. Cette relation entre mode vibratoire et ondulation muqueuse implique une vibration synchronisée des deux plis vocaux à des fréquences identiques. Ainsi, partant de ce modèle, plusieurs études ont démontré qu'un changement de la structure de ces masses pouvait altérer les cycles glottiques et entraîner des asymétries, responsables de ces doubles ou triples périodes. Svec & al. (1996) ont observé que des transitions incontrôlées — ou bifurcations — pouvaient apparaître lors du passage du mécanisme I au mécanisme II. Les auteurs notent, ainsi, l'apparition de sous-harmoniques secondaires liée soit à une augmentation involontaire de la PSG, soit aux interactions entre le flux d'air et les irrégularités des plis vocaux, à savoir des asymétries de tension, de masse ou de longueur. Une désynchronisation vibratoire des faces supérieures et inférieures des plis vocaux est constatée, ce qui implique l'apparition de deux fréquences différentes. Neubauer & al. (2001) ont également mis en évidence que des changements qualitatifs de la dynamique des plis vocaux liés à des variations myoélastiques ou aérodynamiques pouvaient engendrer des bifurcations. Les auteurs montrent que la bitonalité pourrait résulter soit d'une asymétrie antéro-postérieure, soit d'une asymétrie transverse. Dans ce sens, nous ne pouvons exclure que ces asymétries puissent s'effectuer simultanément sur les deux axes verticaux et horizontaux, ce qui pourrait être une explication à la tritonalité. Enfin, Jiang & al. (2001) ont démontré qu'à une PSG constante, une diminution de l'aire glottique était responsable d'une importante baisse du seuil de pression vibratoire, et par conséquent de l'apparition de vibrations irrégulières et d'asymétries. Notre seconde hypothèse pour expliquer la bitonalité ou tritonalité est liée à l'intervention de structures supralaryngées et, plus particulièrement, aux bandes ventriculaires. Dans ce système à double vibrateur, les plis vocaux demeurent le vibrateur principal et donnent la fréquence fondamentale du son tandis que les bandes ventriculaires deviennent co-vibrateurs. Des études sur la voix chuchotée (Crevier Buchman & al. 2008), sur la voix pathologique (Lindestad & al. 2004) et sur la voix chantée (Sakakibara & al. 2001, Lindestad & al. 2001, Chevaillier & al. 2008) ont mis en évidence que ces structures pouvaient se rapprocher, rentrer en contact ou vibrer durant ces gestes vocaux particuliers. Lorsqu'elles entrent en vibrations, les bandes ventriculaires produisent des sons secondaires — identifiables aux sous-harmoniques — qui sont des multiples entiers de la  $F_0$  produite par les plis vocaux, soit  $F_{0BV} = 1/nF_{0PV}$  (où BV désigne les bandes ventriculaires et PV représente les plis vocaux). Ainsi, lorsque le contact se produit périodiquement tous les deux cycles,  $F_{0BV} = F_{0PV}/2$ , ce qui implique le phénomène de double période. Pour la voyelle [a], en particulier, nous observons les mêmes phénomènes décrits dans la thèse de Bailly (2009). Nous

pouvons supposer que les pics positifs de plus grande amplitude sur le signal DEGG représentent une phase durant laquelle les bandes ventriculaires se rapprochent mais ne se touchent pas tandis que la glotte se referme. Les pics positifs de plus faible amplitude représentent, quant à eux, l'adduction des bandes. Durant les pics négatifs, les bandes ventriculaires sont écartées alors que les plis vocaux initient leur décollement. Nous constatons également que la vitesse de contact des plis vocaux est dépendante de l'intervention des bandes ventriculaires et, plus largement, que les cycles glottiques sont perturbés par ce déphasage. Nous notons que  $t_1 > t_2$ , ce qui implique que les cycles glottiques où les bandes sont accolées sont allongés par rapport aux cycles durant lesquels les bandes ventriculaires sont décollées.

En ce qui concerne notre patient n°4, bien que les  $F_0$  de ses productions vocales soient régulières d'une session d'enregistrement à une autre, nous avons décidé d'écartier ses signaux EGG et DEGG que nous jugeons atypiques et biaisés (cf. figure 6.34).



**Figure 6.34 – Signaux EGG et DEGG acquis sur 5 cycles pour la voyelle [a] (à gauche) et sur 6 cycles pour la voyelle [i] (à droite) (patient n°4).**

Les signaux EGG des voyelles du patient n°4 montrent une phase arrondie de fermeture des plis vocaux qui anticipe la fermeture visible et réelle de la glotte, caractérisée par un large pic positif. Si nous analysons, d'un point de vue qualitatif, ces cycles, nous observons une phase d'ouverture allongée suivie d'un contact des plis vocaux irrégulier qui aboutit finalement à une fermeture complète. Ce déphasage dans le temps de la fermeture serait responsable d'une fuite glottique importante. Néanmoins, la forme inappropriée du signal EGG pourrait également être liée à un problème méthodologique lors du recueil de nos données. Certaines personnes ont une physiologie (cou trop gras ou trop maigre) qui entrave la prise de données EGG. Le collier n'est pas adapté car trop relâché ou trop serré, ce qui interfère dans la mesure de la résistance rencontrée par le courant d'une électrode à l'autre. De plus, des mouvements involontaires du locuteur au cours de sa phonation peuvent également perturber les mesures.

### 6.7.2.3 Présentation des résultats

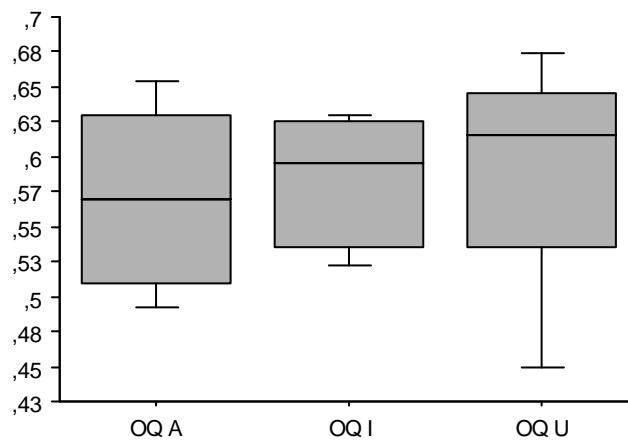
Le tableau 6.19 met en évidence les valeurs moyennes de Oq pour chacune de nos trois voyelles et ce, pour les huit patients retenus et notre témoin et référence ( $n = 27$ , soit  $3$  voyelles  $\times$   $9$  sujets). Comme nous le mentionnions précédemment, nous comparons ces valeurs aux mesures de référence établies pour notre témoin à l'aide d'un test-t univarié. Les résultats sont considérés comme significatifs à partir d'un seuil alpha de 5% (soit  $p < .05$ ).

	[a]	[i]	[u]
Témoin de référence	.54	.54	.56
P1	.66	.54	.55
P2	.64	.63	.68
P3	.50	.53	.52
P4	.62	.63	.66
P5	.54	.61	.61
P6	.49	.52	.42
P7	.52	.62	.63
P8	.60	.58	.62
Résultats du test-t univarié	$t_{(7)} = 1.326, p = .2264$	$t_{(7)} = 2.584, p = .0363$	$t_{(7)} = .868, p = .4140$

**Table 6.19** – Quotients moyens ouverts des huit patients et de notre témoin de référence pour les voyelles [a], [i], [u] et résultats des tests-t univariés ( $t_{(7)}$  la valeur du test de Student pour un degré de liberté égal à 7 et  $p$  la probabilité) ( $n = 27$ ).

Nous remarquons que cinq de ces patients présentent des mesures moyennes de Oq supérieures ou égales aux valeurs de référence pour la voyelle [a] (exceptés P3, P6 et P7). En ce qui concerne la voyelle [i], nous notons que six de nos patients montrent également des mesures supérieures ou égales à notre témoin (exceptés P3 et P6). Enfin, pour la voyelle [u], nous constatons que cinq de nos sujets dysphoniques ont également des valeurs moyennes de Oq supérieures ou égales à notre référence (exceptés P1, P3, P6). En ce qui concerne le test-t univarié, seule la voyelle [i] du témoin est significativement différente des valeurs moyennes relevées pour les patients, avec une p-value égale à  $p = .0363$ .

La figure 6.35 présente, quant à elle, les valeurs moyennes de Oq de notre population de patients pour chacune des trois voyelles.



**Figure 6.35** – Valeurs moyennes des quotients ouverts selon le type de voyelle pour huit de nos patients (les barres d'erreurs représentent les écarts-types) ( $n = 24$ ).

Nous observons que les voyelles hautes tendent à avoir des valeurs de Oq légèrement supérieures à celles de la voyelle [a], sans que cela soit significatif. Par ailleurs, il est important de prendre en compte les écarts-types, en particulier pour la voyelle [u], ce qui dénote une grande variabilité en production.

#### 6.7.2.4 Discussion

Dans un premier temps, nous observons que les valeurs moyennes de Oq des voyelles hautes des patients tendent à être légèrement plus élevées par rapport à celles prises pour la voyelle basse [a]. Ces résultats — que nous devons considérer avec prudence étant donné la forte variabilité individuelle et le nombre restreint de données — sont en opposition avec les conclusions de Henrich (2001) pour la voix chantée en mécanisme I. Celle-ci stipule que le Oq dépendrait de l'ouverture mandibulaire. Ainsi, la voyelle [a] serait caractérisée par une ouverture maximale de la bouche, ce qui favoriserait le relâchement de la compression laryngée et impliquerait un quotient d'ouverture légèrement supérieur par rapport aux voyelles [e] et [u]. Néanmoins, l'étude de Cookman & Verdolini (1999) qui met en relation l'ouverture de la bouche avec des paramètres glottiques, contredit cette thèse et valide nos résultats préliminaires. Ces derniers notent que l'ouverture mandibulaire entraîne une diminution du quotient d'ouverture aussi bien chez une population masculine que chez une population féminine, ce qui est contraire aux données de Henrich (2001). D'autres études qui cherchent à démontrer l'effet des voyelles sur les valeurs de Oq s'accordent quant à leur indépendance. Ainsi, Chen & al. (2002) montrent qu'il n'y a aucun effet significatif de la voyelle quant aux mesures du Oq en voix modale ou craquée, quelle que soit la population. De même, Lim & al. (2006) soulignent les effets significatifs des voyelles sur l'ouverture de la bouche et certains paramètres laryngés, à l'exception du

quotient ouvert. En ce qui nous concerne, bien que les différentes études démontrent que l'effet des voyelles sur Oq reste variable, nous pensons que le fait que les voyelles hautes présentent des valeurs moyennes de Oq plus élevées par rapport à la voyelle [a] est logique et explicable. En effet, nous avons déjà montré que la production des voyelles hautes nécessite une tension et une élongation plus importante des plis vocaux et, plus largement, du larynx, ce qui est finalement à l'origine de l'élévation de leur  $F_0$  intrinsèque. L'augmentation de la tension et/ou de la longueur des plis vocaux est responsable d'une plus grande raideur, ce qui a pour conséquence une réduction de leur amplitude de vibration. De ce fait, l'adduction laryngée diminue, ce qui implique des phases d'ouverture allongées et, par définition, des valeurs de Oq un peu plus importantes pour les voyelles [i] et [u] (Higgins & al. 1998). Dans ce sens, Bickley & Stevens (1986) expliquent que d'importantes constrictions orales — cas typiques pour les voyelles hautes — engendrent des phases d'ouverture plus longues dans le cycle glottique.

Dans un second temps, nous notons que de façon générale, six de nos patients présentent des valeurs moyennes de Oq supérieures ou égales à celles de notre témoin, c'est-à-dire qu'ils présentent des phases d'ouverture glottique plus longues. Seuls les patients P3 et P6 semblent adopter des comportements laryngés contradictoires. Le premier patient présente des valeurs moyennes de Oq proches de 0.5, ce qui le rapproche de la voix modale de notre témoin. Au contraire, le second patient présente des valeurs moyennes de Oq inférieures aux valeurs de notre référence, ce qui implique des phases d'ouverture laryngée raccourcies. Ces différences entre patients tendent finalement à montrer les différentes stratégies liées aux cordectomies laser. Comme nous le mentionnions précédemment, une montée des valeurs de Oq est liée à un relâchement musculaire, tandis que des valeurs basses du Oq dénotent un serrage laryngé. Selon Laver (1994), les tensions peuvent être laryngées ou supralaryngées et ces différences de configuration musculaire permettent d'opposer une phonation « lax » à une phonation « tense ». Les tensions laryngées peuvent être dues à une adduction trop importante des plis vocaux ou à une forte compression médiane. Les tensions supralaryngées, quant à elles, résultent d'une tension musculaire de l'ensemble du conduit vocal. Ainsi, les patients P3 et P6 montrent des valeurs de Oq moyennes ou basses, ce qui est lié à des voix davantage pressées. Laver (1980) ajoute que ces voix tendues ou pressées sont également caractérisées par un *pitch* important, des irrégularités tant en fréquence qu'en amplitude, une montée du larynx et une hausse de la PSG, ce qui correspond à nos données. En outre, la relation entre débit d'air et Oq n'est plus à démontrer puisqu'il existe une corrélation positive entre ces deux paramètres (Lofqvist & McGowan 1992, Löfqvist & al. 1995). Or, nos mesures aérodynamiques montrent que malgré une tendance générale aux fuites glottiques, ces patients présentent les indices les plus bas, ce qui implique une déperdition d'air dans une moindre mesure. Au contraire, les patients P1, P2, P4, P5, P7 et P8 présentent des valeurs moyennes de Oq légèrement plus élevées — sans que cela soit significatif —, ce qui traduit un certain relâchement et une voix plus détendue ou soufflée. Nous avons déjà pu mettre en évidence que la chirurgie était responsable de changements anatomico-physiologiques, et en particulier la réduction de la masse vibrante. Cette nouvelle configuration glottique

est défavorable à une fermeture complète des plis vocaux, ce qui explique la déperdition d'air à travers la glotte et les fuites persistantes en phonation. Ceci est donc un argument en faveur d'une mauvaise adduction des plis vocaux, ce qui se traduit par des temps d'ouverture allongés. Laver (1980) précise que les voix « lax » ou « breathy » possèdent les caractéristiques opposées aux voix « tense », à savoir une baisse de la  $F_0$ , une détente des différentes structures, un abaissement du larynx, etc. Or, nos données montrent le caractère soufflé et tendu de ces voix, associé à une élévation et à des irrégularités du fondamental. En effet, il n'est pas antinomique de présenter également une voix soufflée par hyperfonction. De ce fait, nous ne pouvons pas catégoriser ces voix selon un type de phonation prédéfini, à savoir « tense/harsh/whispered voice » ou « lax/breathy » (Laver 1980). Plusieurs caractéristiques acoustiques, aérodynamiques et glottiques contribuent à donner à ces voix pathologiques une qualité qui leur est propre car les plis vocaux sont atypiques. Finalement, les défauts d'accolement des plis vocaux que nous venons de mettre en évidence doivent nous alerter sur la maîtrise du voisement de ces patients.

### 6.7.3 Le quotient ouvert peut-il être considéré comme un paramètre robuste pour l'opposition de voisement ? Le cas des occlusives en contexte

#### 6.7.3.1 Méthodologie

Comme nous le précisions précédemment (cf. section 6.7.2), nos données EGG ont été acquises lors des enregistrements aérodynamiques à l'aide d'un laryngographe relié à la station EVA2. Nous avons utilisé le logiciel Phonetik qui nous a permis une prise de données synchronisée entre le signal acoustique et le signal EGG. Pour les raisons exposées plus haut, seul un témoin a pu être enregistré, ce dernier nous servira de référence. Nos 10 patients ont pu être enregistrés mais en raison des difficultés rencontrées et des patterns particuliers que nous avons pu mettre en évidence, les productions de deux de ces sujets ont été écartées (cf. section 6.7.2.2). Pour cette étude, nous avons choisi d'analyser les logatomes composés d'occlusives. Afin de garder une cohérence dans notre travail et de faciliter le traitement de nos données, les logatomes ont été divisés en trois syllabes distinctes  $CV_1$ ,  $CV_2$  et  $CVC_3$  et seules les positions 1 et 2 ont été traitées. Nous avons donc traité 288 logatomes, soit  $18 \text{ logatomes} \times 1 \text{ répétition} \times 2 \text{ positions} \times 8 \text{ patients}$ .

Les signaux audio et EGG ont été reconvertis dans un format .wav afin d'établir une segmentation sous Praat. L'intérêt de cette démarche est de pouvoir combiner parfaitement les deux signaux et d'avoir ainsi une meilleure synchronisation visuelle. La segmentation a été réalisée manuellement en manipulant les spectrogrammes obtenus et les signaux synchronisés. Les segments sont inspectés, écoutés et la réalité perceptive est associée avec la réalité acoustique. Cette segmentation se base sur la méthodologie exposée dans la section 6.6. Néanmoins, nous avons fait le choix d'une segmentation moins fine où seule une unique tire segmentale est exploitée. Les signaux EGG étiquetés ont été, par la suite, analysés à l'aide du script Matlab que nous avons pu décrire

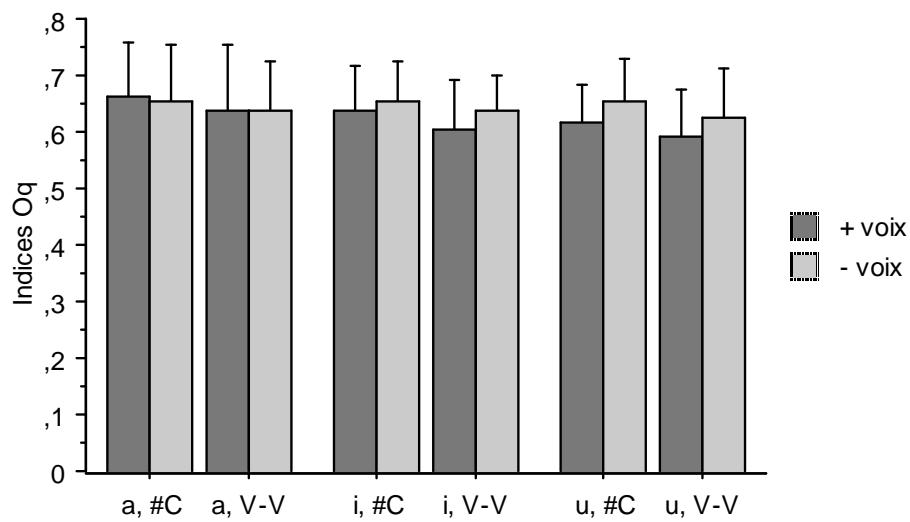
dans notre étude sur les voyelles isolées (cf. section 6.7.2.1). Les fichiers textes résultants ont également été inspectés visuellement et les valeurs de Oq qui s'écartaient de plus de 10% des valeurs adjacentes ont été écartées (Di Canio 2009).

Pour cette étude, nous avons relevé les valeurs de Oq des 5 premiers cycles complets des voyelles après le relâchement des consonnes occlusives et nous avons calculé leur moyenne. En effet, nous avons pu montrer qu'il existe une corrélation entre le quotient ouvert et la qualité vocale, les valeurs de Oq étant plus importantes pour des voix « breathy » du fait d'un moindre contact des plis vocaux (Laver, 1980). Notre but étant de vérifier que le voisement avait une influence sur la qualité vocale du début des voyelles adjacentes, nous avons opté dans notre traitement statistique pour un regroupement des valeurs moyennes en fonction du voisement des occlusives, soit [p, t, k] vs [b, d, g] en écartant volontairement le critère du lieu d'articulation.

Finalement, cette étude nous a permis d'évaluer l'influence du voisement des occlusives sur les voyelles environnantes en fonction des contextes vocaliques, de la position syllabique et évidemment du voisement de la consonne. Nous avons donc conduit une ANOVA à 3 facteurs où la variable dépendante était la mesure moyenne de Oq des voyelles et les variables indépendantes représentaient le type de voyelle, la position dans le logatome et le voisement de la consonne à l'aide du logiciel Statview. Les résultats étaient considérés comme significatifs à partir d'un seuil alpha de 5%, soit  $p < .05$ .

### 6.7.3.2 Présentation des résultats

La figure 6.36 présente les valeurs moyennes de Oq des voyelles en contexte occlusif pour huit de nos patients en fonction des voyelles, de la position dans les logatomes et du voisement. Le test ANOVA montre que les contextes vocaliques ne semblent avoir aucune influence sur nos résultats ( $F(2, 276) = 2,063, p = .1290$ ). Au contraire, nous notons une tendance à l'abaissement des valeurs de Oq en contexte intervocalique par rapport à la position initiale. Notre test statistique met en évidence un effet significatif de la position sur les moyennes de Oq ( $F(1, 276) = 5,953, p = .0153$ ). En ce qui concerne le voisement, bien que ces résultats ne soient pas très pertinents pour la voyelle [a], nous notons que les valeurs moyennes de Oq tendent à être plus basses en contexte voisé et à s'élever en contexte non voisé. Le résultat du test ANOVA corrobore ce constat ( $F(1, 276) = 3,778, p = .0529$ ). Pour cette analyse, nous ne retrouvons aucune interaction entre les différents facteurs.



**Figure 6.36** – Valeurs moyennes des quotients ouverts des voyelles en contexte occlusif pour huit de nos patients en fonction des voyelles, de la position dans le logatome et du voisement (les barres d'erreurs représentent les écarts-types) ( $n = 288$ ).

Par ailleurs, nous avons également établi ce même test ANOVA sur les valeurs moyennes de Oq de notre témoin de référence ( $n = 36$ , soit  $18$  logatomes  $\times$  1 répétition  $\times$  2 positions  $\times$  1 témoin). Nous ne présenterons pas ces résultats en détails car un unique sujet est insuffisant pour avancer des conclusions. Toutefois, nous notons des tendances qui semblent converger vers les observations préliminaires que nous avons pu émettre concernant les patients, en particulier pour l'effet du voisement sur les valeurs de Oq. Nous retrouvons un abaissement des valeurs de Oq en contexte voisé et une élévation de celles-ci en contexte non voisé ( $F(1, 24) = 12, 062, p = .002$ ). Aussi, et contrairement aux patients, nous retrouvons un effet significatif des contextes vocaux ( $F(2, 24) = 14, 15, p = < .0001$ ). Les voyelles hautes semblent éléver les valeurs moyennes de Oq, ce que nous avions déjà pu mettre en évidence dans la section précédente sur les voyelles isolées. Enfin, et toujours en opposition aux patients, nous ne retrouvons aucun effet significatif de la position dans le logatome sur notre variable dépendante ( $F(2, 24) = 2, 215, p = .1497$ ). Notre test ne montre, une fois n'est pas coutume, aucune interaction entre les facteurs.

### 6.7.3.3 Discussion

En l'absence de données électroglottographiques, la qualité vocale peut être évaluée à partir de la pente spectrale, et plus particulièrement en observant les différences d'amplitude entre le premier harmonique (H1) et l'harmonique supérieur (H2) ou l'amplitude du premier formant (A1) ou l'amplitude du second formant (A2). De manière générale, les auteurs précisent que le paramètre H1-H2 peut être corrélé au quotient

d'ouverture des voyelles (Holmberg & al. 1988). Ainsi, ce paramètre acoustique est largement utilisé pour distinguer les voyelles soufflées, modales ou craquées qui sont phonémiquement distinctives dans plusieurs langues. Les voyelles soufflées sont caractérisées par un maximum d'énergie au niveau de la fréquence fondamentale, ce qui implique une grande différence entre H1 et H2. Au contraire, les voyelles craquées ont davantage d'énergie au niveau des premiers harmoniques, dans la région du F1 ou F2. De ce fait, la différence d'amplitude entre H1 et H2 est très basse (Ladefoged & Maddieson 1996).

Cependant, les études qui utilisent les paramètres H1-H2 ou Oq — qui nous intéresse plus particulièrement pour cette étude — pour mettre en évidence les effets de la coarticulation sur les catégories phonologiques des consonnes restent très minoritaires. Basée sur les travaux préliminaires de Löfqvist & McGowan (1992), la recherche de Löfqvist & al. (1995) représente, à notre connaissance, la première étude qui met en relation les écoulements d'air buccaux avec des paramètres glottiques tels que le Oq lors des transitions VC et CV chez 6 locuteurs américains. Malgré une variabilité individuelle, les auteurs rapportent d'importantes variations de l'écoulement de l'air et des quotients d'ouverture proches de 1 dans les transitions voyelles-consonnes non voisées. Dans ce sens, l'étude de Ni Chasaide & Gobl (1999) représente la première recherche comparative qui s'est intéressée à l'effet du contexte de voisement des occlusives sur les voyelles environnantes en allemand, anglais, suédois, français et italien. Les auteurs ont également mis en évidence que le voisement des consonnes influençait le mode vibratoire des voyelles adjacentes. Sans parler de tendances universelles, ces derniers ont montré que le début d'une voyelle qui suit une consonne non voisée sera davantage soufflé et aura un quotient d'ouverture plus haut. Des recherches plus récentes sur le coréen (Cho & al. 2002), le persan (Heselwood & Mahmoodzade 2007), le sundanais (Kulikov 2010) par exemple tendent vers des conclusions similaires. Les différents auteurs notent des différences d'amplitude entre H1 et H2 plus importantes, des quotients d'ouverture plus élevés et inversement des quotients de fermeture plus bas au début des voyelles au contact d'occlusives non voisées. Au contraire, le début des voyelles en contexte d'occlusives voisées serait caractérisé par des différences H1-H2 moindre, des quotients d'ouverture abaissés ou des quotients de fermeture plus élevés.

Ces différences acoustiques ou EGG entre les voyelles produites en contexte voisé ou non voisé suggèrent que le voisement des consonnes occlusives interfère dans le mode vibratoire de ces segments vocaliques. Les valeurs de Oq des voyelles indiquent que la tension des plis vocaux est différente après une consonne non voisée par rapport à une consonne voisée, avec une réduction de la tension additive et de la compression médiane au début du voisement de la voyelle suivant une occlusive non voisée. Dans cet environnement C[-voix]V, nous observons donc une diminution du contact des plis vocaux au début de la voyelle, ce qui explique leur caractère soufflé. Cette explication est en adéquation avec les conclusions de Löfqvist & al. (1995) qui expliquent qu'au relâchement de l'occlusion des consonnes non voisées la glotte est ouverte mais que le *timing* laryngé se met progressivement en place pour la production de la voyelle suivante. Ainsi, les vibrations glottiques sont initiées lorsque l'aire glottique décroît

alors que les écoulements d'air persistent. Par conséquent, des valeurs importantes du Oq sont observées au début des voyelles du fait du retard de la mise en place des articulateurs. Les valeurs de ces paramètres tendent à décroître durant la première partie de la voyelle, ce qui dénote une adduction plus complète des plis vocaux.

En outre, nous avons pu noter que les valeurs de Oq avaient tendance à s'abaisser de la position initiale à la position intervocalique. En position initiale absolue, la glotte est largement ouverte pour la respiration. En conséquence, le *timing laryngé* peut être plus long à se mettre en place et l'adduction des plis vocaux plus molle, ce qui implique des valeurs de Oq plus importantes. Au contraire, en position intervocalique, l'aire glottique est réduite et les articulateurs peuvent se synchroniser un peu plus rapidement, ce qui engendre une réduction de la phase d'abduction et donc un abaissement des valeurs de Oq.

Finalement, la qualité vocale associée au contrôle de la tension des plis vocaux, pourrait être considérée comme un paramètre robuste dans l'opposition du trait de voisement. Nos résultats pour notre population de patients et pour notre témoin de référence vont dans le sens de ces observations et doivent être complétés par des investigations supplémentaires afin de valider nos conclusions préliminaires et de vérifier l'existence de différences inter-groupes.

#### Synthèse de l'étude n°8 : études électroglottographiques (EGG)

■ Objectifs de l'étude : analyser les valeurs de Oq des voyelles des patients et montrer l'intérêt de l'EGG pour l'opposition de voisement des occlusives.

■ Corpus : (i) parties stables des voyelles tenues et isolées et (ii) logatomes de type CVCVCVC où C= [p, t, k, b, d, g] et V= [a, i, u].

■ Principaux résultats :

- mise en évidence de cas de diplophonie ou triplophonie,
- la majorité des patients présentent des valeurs de Oq des voyelles supérieures à celles de notre témoin référence (voix breathy),
- valeurs de Oq au début des voyelles plus importantes en contexte occlusif non voisé, en particulier pour les voyelles hautes.

■ Conclusions :

- les valeurs de Oq élevées des patients traduisent une fermeture glottique incomplète,
- différences des valeurs de Oq au début des voyelles en contexte voisé et non voisé suggèrent que le voisement des occlusives interfère dans le mode vibratoire des voyelles adjacentes,
- intérêt de ce type de méthodologie pour la distinction de voisement des consonnes et argument supplémentaire pour le maintien du contraste après cordectomie.

## 6.8 Etude n°9 : étude perceptive

Afin de compléter nos études acoustiques, aérodynamiques et EGG sur le maintien du voisement après cordectomie de type II-III, nous avons réalisé une étude perceptive sur l'intelligibilité des occlusives du français en situation syllabique. Notre but est de

mettre en évidence ou non des confusions de voisement qui seraient directement liées aux modifications laryngées engendrées par la chirurgie.

### 6.8.1 Méthodologie

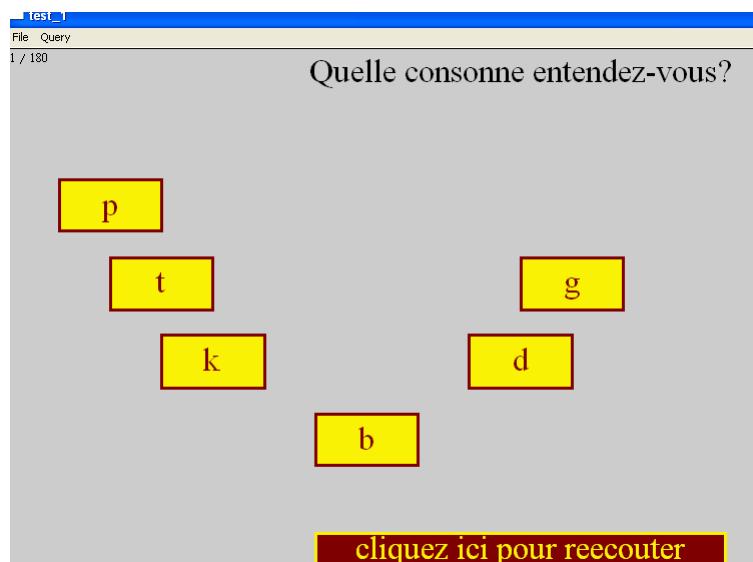
Nos tests de perception se basent sur les enregistrements acoustiques des logatomes CVCVCVC de nos 10 patients. Dans une logique de cohérence, nous avons choisi de conserver uniquement les logatomes composés d'occlusives, soit 180 occurrences (18 logatomes de type CVCVCVC où C= [p, t, k, b, d, g] et V= [a, i, u] × 1 répétition × 10 sujets). Un découpage syllabique de type CV<sub>1</sub>, CV<sub>2</sub> et CVC<sub>3</sub> a été effectué et seules les positions 1 et 2 de ces logatomes ont été traitées. En effet, nous avons déjà mis en évidence que celles-ci étaient pertinentes d'un point de vue de leur robustesse articulatoire et prosodique, ce qui nous a conduit à l'élaboration de deux tests distincts en fonction des positions syllabiques.

Pour ces analyses perceptives d'identification des consonnes, nous avons constitué des fichiers sons indépendants à partir des échantillons vocaux enregistrés au préalable. Les spectrogrammes et les oscillogrammes de ces derniers étaient inspectés visuellement sous Praat afin de couper les parties qui nous intéressaient pour nos tests. Ainsi, les syllabes en initiale absolue de type CV et les syllabes en intervocalique de type VCV ont été segmentées et stockées séparément dans un format .wav dans deux dossiers différents. Finalement, nous avons pu constituer 360 stimuli, soit 180 sons pour la position où la consonne est en attaque absolue et 180 sons où la consonne est en attaque en intervocalique.

Nous avons donc élaboré deux tests d'identification sous Praat. Ces derniers prennent la forme de la figure 6.37. Ce sont des tests à choix forcé où les auditeurs doivent identifier parmi l'ensemble des occlusives orales du français le son qu'ils pensent avoir entendu. En cas de doute, les auditeurs avaient la possibilité de réécouter plusieurs fois le stimulus entendu en cliquant sur le champs prévu à cet effet. La validation d'une des six occlusives permet le passage au stimulus suivant et le décompte des sons est repérable en haut à gauche. Les syllabes étaient présentées dans un ordre aléatoire avec une seule répétition de chaque syllabe. En effet, après un pré-test sur deux auditeurs où deux répétitions étaient proposées — soit 180 sons × 2 répétitions × 2 tests = 720 occurrences —, nous nous sommes rapidement rendue compte que la lassitude et le caractère répétitif de la tâche biaisaient nos résultats, et en particulier la partie finale des tests. En conséquence, nous avons volontairement privilégié le nombre d'auditeurs plutôt que les répétitions. Finalement, les sessions d'écoute durraient 30 minutes durant lesquelles les auditeurs pouvaient faire trois pauses tous les 60 stimuli environ.

Pour cette étude sur l'intelligibilité des consonnes et, en particulier, du contraste de voisement, nous avons donc sollicité 22 auditeurs âgés de 25 à 65 ans (moyenne 38 ans) et de langue maternelle française, n'ayant aucun problème auditif ou autres pathologies langagières, neurologiques ou communicationnelles. Les auditeurs n'ont pas eu de séance d'entraînement pour les syllabes testées. Chaque auditeur était seul face à l'ordinateur dans une pièce acoustiquement calme. L'écoute se faisait au casque avec une intensité pré-réglée (74 dB) et identique pour tous nos auditeurs. Finalement,

pour chaque test, nous avons traité 3960 stimuli (soit 180 syllabes  $\times$  22 auditeurs), répartis par série de 660 occlusives orales. Au total, nous avons donc analysé 7920 stimuli (180 syllabes  $\times$  2 tests  $\times$  22 auditeurs).



**Figure 6.37** – Exemple d'un test de perception de Praat. L'auditeur clique sur la consonne qu'il pense reconnaître, ce qui déclenche le stimulus suivant. En cas de doute, l'auditeur peut réécouter le son en cliquant dans le champ prévu à cet effet.

## 6.8.2 Présentation des résultats

Les résultats de nos deux tests sont présentés sous la forme d'une matrice de confusion à six entrées (les six occlusives orales) qui regroupe les résultats des 10 patients évalués par nos 22 auditeurs. Nous présenterons donc deux matrices qui mettent en évidence les taux d'identification globaux selon les positions syllabiques. Nous montrerons également la répartition de ces taux selon les contextes vocaliques.

L'analyse des données se fait verticalement et horizontalement. La colonne verticale correspond aux phonèmes cibles, c'est-à-dire aux phonèmes réalisés par les patients. La colonne horizontale correspond aux phonèmes perçus par les auditeurs. Par exemple, le tableau 6.20 montre que sur 660 [p] traités, le phonème a été perçu 593 fois [p], 2 fois [t], 63 fois [b], 1 fois [d] et 1 fois [g]. Finalement, nous avons calculé les pourcentages des taux d'identification correcte pour chacune des consonnes.

### 6.8.2.1 La position initiale

Pour cette position, le taux d'identification global est de 93,59% pour nos 22 auditeurs. Le tableau 6.20 présente la matrice de confusions générale, c'est-à-dire tous contextes vocaliques confondus, pour l'ensemble des six occlusives orales.

	p	t	k	b	d	g	% d'identification
p	593	2		63	1	1	89.85
t	3	610	13		33	1	92.43
k		18	624		1	17	94.55
b	2			645	1	12	97.73
d		1	1	12	633	13	95.91
g	8	1	14	25	11	601	91.06

**Table 6.20** – Matrice de confusion globale pour la position **initiale absolue**. Les identifications correctes ont été converties en pourcentages d'identifications ( $n = 3960$ ).

Les résultats montrent que l'opposition de voisement semble être maintenue par nos patients. En effet, le taux général d'identification reste assez élevé. Les quelques confusions de voisement concernent les occlusives homorganiques, dans le sens d'un voisement des occlusives non voisées. Ainsi, nous remarquons que la labiale [p] est perçue dans 9,55% des cas comme [b], mais l'inverse ne représente que 0,3% des cas. Ce constat est également valable pour la dentale [t] qui est perçue dans 5% des cas comme [d], mais l'inverse ne représente que 0,15% des cas. La vélaire [k] présente une confusion de voisement (2,58%) associée à une confusion de lieu d'articulation avec la dentale [t]. Les consonnes voisées [b] et [d] présentent majoritairement des confusions de lieu d'articulation avec un maintien de la perception du trait de voisement. La consonne [g] semble être la moins robuste puisqu'elle montre à la fois des confusions de lieu d'articulation et de voisement avec son homologue non voisée dans 2,12% des cas.

Les tableaux 6.21, 6.22 et 6.23 présentent les matrices de confusions selon les contextes vocaliques [a], [i] et [u]. Les taux d'identification généraux représentent 92,5%, 94,02% et 94,24% respectivement.

	p	t	k	b	d	g	% d'identification
p	185			34		1	84.09
t	1	186	11		21	1	84.55
k		8	201			11	91.36
b	2			216	1	1	98.18
d				2	215	3	97.73
g	1				1	218	99.09

**Table 6.21** – Matrice de confusion pour le contexte [a], soit [#Ca] ( $n = 1320$ ).

	p	t	k	b	d	g	% d'identification
p	198	1		20	1		90
t	1	217	2				98.64
k	9	209			1	1	95
b			218			2	99.09
d		1	3	208	8		94.55
g	3	1	14	10	191		86.82

**Table 6.22** – Matrice de confusion pour le contexte [i], soit [#Ci] ( $n = 1320$ ).

	p	t	k	b	d	g	% d'identification
p	210	1		9			95.45
t	1	207			12		94.09
k		1	214			5	97.27
b			211			9	95.90
d		1	7	210	2		95.45
g	4		13	11	192		87.27

**Table 6.23** – Matrice de confusion pour la voyelle [u], soit [#Cu] ( $n = 1320$ ).

Ces matrices confirment les tendances générales que nous avons pu mettre en évidence précédemment, à savoir un voisement des occlusives non voisées. Néanmoins, nous notons que de manière générale, le contexte vocalique [a] semble défavorable à la reconnaissance des occlusives labiales et dentales sourdes. Au contraire, les erreurs que nous avons pu relever pour l'occlusive vélaire voisée [g] apparaissent majoritairement dans les contextes [i] et [u].

### 6.8.2.2 La position intervocalique

Pour cette position, le taux d'identification global est de 93,79% pour nos 22 auditeurs. Le tableau 6.24 présente la matrice de confusion générale pour l'ensemble des six occlusives orales.

	p	t	k	b	d	g	% d'identification
p	644	2		14			97.58
t	8	627	7		16	2	95
k	5	13	608	1	1	32	92.2
b	6		1	633	3	17	95.91
d	2	13		20	604	21	91.52
g		1	21	15	25	598	90.61

**Table 6.24** – Matrice de confusion globale pour la position **intervocalique**. Les identifications correctes ont été converties en pourcentages d'identifications ( $n = 3960$ ).

Encore une fois, nos résultats montrent une très bonne intelligibilité des consonnes avec une opposition de voisement intacte en position intervocalique. Comme pour la position initiale, les quelques confusions de voisement relevées concernent les occlusives homorganiques, avec un voisement des occlusives non voisées. Ainsi, dans une moindre mesure par rapport à nos observations précédentes, la labiale [p] est perçue [b] dans 2,12% des cas, mais l'inverse ne représente que 0,91% des cas. De même, la dentale [t] est perçue [d] dans 2,42% des cas, mais l'inverse ne représente que 1,97% des cas. Enfin, la vélaire [k] présente à la fois une confusion de voisement (4,85%) et une confusion de lieu d'articulation majeure avec la dentale [t]. Les confusions prédominantes pour les occlusives voisées concernent les lieux d'articulation avec un maintien de la perception du trait de voisement. La vélaire voisée [g] semble être la moins robuste avec des confusions de lieu, en particulier avec la dentale [d], associées à des confusions de voisement avec son homologue non voisée et ce dans 3,18% des cas.

Les tableaux 6.25, 6.26 et 6.27 présentent les matrices de confusions selon les contextes vocaliques [a], [i] et [u]. Les taux d'identification généraux représentent 96,29%, 93,03% et 92,05% respectivement.

	p	t	k	b	d	g	% d'identification
p	215			5			97.73
t	5	204	6		4	1	92.73
k	1	4	203			12	92.27
b			1	218	2		99.09
d	1	2		3	212	2	96.36
g				1		219	99.55

Table 6.25 – Matrice de confusion pour le contexte [a], soit [aCa] ( $n = 1320$ ).

	p	t	k	b	d	g	% d'identification
p	210	2		8			95.45
t		216	1		2	1	98.18
k	1	9	201		1	8	91.36
b	4			213	2	1	96.82
d		1		12	194	13	88.18
g		1			25	194	88.18

Table 6.26 – Matrice de confusion pour le contexte [i], soit [iCi] ( $n = 1320$ ).

	p	t	k	b	d	g	% d'identification
p	219			1			99.55
t	3	207			10		94.09
k	3		204	1		12	92.73
b	2			202		16	91.82
d	1	10		5	198	6	90
g			20	15		185	84.09

Table 6.27 – Matrice de confusion pour le contexte [u], soit [uCu] ( $n = 1320$ ).

Contrairement à nos observations précédentes, le contexte [aCa] semble favoriser l'identification des consonnes. Les confusions de voisement concernant les occlusives non voisées se répartissent selon les trois contextes vocaliques. Néanmoins, les erreurs d'identification sur les lieux d'articulation des consonnes voisées se situent majoritairement dans les contextes [iCi] et [uCu]. D'autre part, seul le contexte vocalique [uCu] semble être propice au dévoisement de la vélaire [g].

### 6.8.3 Discussion

Dans la présente, les tests de perception ont montré que les consonnes réalisées par les patients ont été correctement identifiées par les auditeurs dans 93,59% des cas en position initiale et 93,79% des cas en position intervocalique. Ces importants taux mettent en évidence le maintien de l'opposition de voisement chez nos sujets dysphoniques.

Nous avons observé deux types de confusions mineures dans la perception des occlusives produites : des confusions de voisement et de lieux d'articulation. Les confusions de voisement étaient plus fréquentes pour les consonnes non voisées perçues comme voisées, principalement pour la position initiale. Ces confusions affectaient surtout la labiale [p] perçue [b] et la dentale [t] reconnue [d] à l'initial et particulièrement dans le contexte vocalique [a]. Flack (2007) explique la difficulté à percevoir correctement les sons en position initiale, en particulier pour établir l'opposition entre les occlusives labiales voisées et non voisées en français. Dans une première partie perceptive basée sur l'identification des consonnes [p, b, t, d], l'auteur a montré que l'occlusive labiale non voisée [p] en position initiale était reconnue avec un temps de réaction significativement plus long que sa contrepartie voisée [b]. Cette différence ne se retrouve pas pour la position intervocalique et pour l'identification des consonnes dentales dans les deux positions. Néanmoins, l'auteur note qu'il n'existe aucune différence en termes de réponses correctes entre les consonnes [p] et [b] initiales, avec des pourcentages d'identification respectivement à 93% et 94%. L'auteur justifie l'hésitation des auditeurs pour l'identification de la consonne [p] initiale par sa fréquence d'apparition, moins importante dans le lexique français par rapport à l'occlusive [b] initiale. Dans une seconde étude, l'auteur justifie ses conclusions préliminaires par la proximité acoustique que peuvent avoir les consonnes [p] et [b] initiales par rapport à ces mêmes consonnes en position intervocalique ou par rapport aux occlusives [t] et [d]. Selon elle, les [p] initiaux ressemblent davantage à des [b] mais le contraire

n'est pas vérifié. Dans ce sens, Flack s'est intéressée aux VOT et à l'intensité relative des relâchements de 376 occlusives françaises. Ainsi, contrairement à la paire [t-d] ou à la position intervocalique, les consonnes [p] et [b] initiales ne présentent pas de différence statistiquement significative pour les valeurs d'intensité relative des relâchements. De plus, la consonne labiale non voisée initiale possède un VOT plus court que cette même consonne en position intervocalique ou que la dentale non voisée, ce qui tend à la rapprocher de sa contrepartie voisée dans cette position. En outre, il est clair que la position intervocalique est favorable à l'identification des consonnes à cause du phénomène de coarticulation. Les informations acoustiques telles que le voisement ou le lieu d'articulation se répartissent sur l'ensemble de la syllabe. Ainsi, nous notons un transfert de ces indices phonétiques de la consonne intervocalique vers les segments vocaliques adjacents et *vice-versa*. De ce point de vue, Ohman (1966) a mis en évidence l'effet anticipatoire d'une voyelle initiale à une voyelle finale dans une structure de type VCV. L'auteur précise que le premier segment vocalique influence la transition initiale de la dernière voyelle à travers le segment consonantique intermédiaire. Ainsi, les gestes articulatoires de la consonnes, influencés par la voyelle initiale, sont également synchronisés sur les mouvements de la voyelle finale.

Enfin, nous avons montré que les erreurs de lieu d'articulation touchaient principalement la consonne vélaire [g] dans les deux positions. Celle-ci était majoritairement confondue avec l'occlusive [b] dans les contextes [#Cu] et [uC] ou [d] dans les contextes [#Ci] et [iCi]. Nous tenterons d'apporter une brève explication acoustique à ce constat bien que ce n'est pas le propos de notre démarche. Stevens (1998) a montré que la perception des lieux d'articulation des consonnes était principalement liée à leurs explosions. Ainsi, il est connu que les occlusives postérieures présentent des relâchements d'amplitude plus importante et de plus longue durée car le conduit vocal en avant de la constriction est plus long, ce qui suggère davantage d'informations sur l'articulation de ces consonnes (Ohala 1983). De plus, nous savons que les transitions vocaliques influencent la fréquence centrale du bruit de relâchement : celui-ci est plus grave devant les voyelles postérieures, moyen devant [a] et aigu devant les voyelles antérieures. Stevens (1998) explique que cette fréquence de résonance est liée à la résonance de la cavité antérieure. La voyelle [i] est plus antérieure que la voyelle [a], ce qui tend à antérioriser la consonne vélaire et à la rapprocher, d'un point de vue acoustique, de l'occlusive dentale [d]. De plus, la voyelle [u] nécessite un arrondissement labial, ce qui coïncide avec les occlusives labiales. Du point de vue de la coarticulation, il est donc plus aisé acoustiquement et perceptivement d'assimiler une labiale à cette voyelle qu'une occlusive vélaire.

**Synthèse de l'étude n°9 : étude perceptive**

- Objectifs de l'étude : confirmer le maintien du contraste de voisement des occlusives à travers une étude perceptive.
- Corpus : logatomes de type CVCVCVC où C= [p, t, k, b, d, g] et V= [a, i, u].
- Principaux résultats :
  - intelligibilité globale très élevée (93,59% pour la position initiale et 93,79% pour la position intervocalique),
  - confusions mineures de voisement pour les consonnes non voisées (perçues voisées), principalement en position initiale (la labiale [p] est perçue [b] dans 9,55% des cas, la dentale [t] est perçue [d] dans 5% des cas et la vélaire [k] est perçue [g] dans 2,58% des cas),
  - les confusions de lieu concernent principalement la vélaire voisée [g] qui est confondu avec [b] en contexte de voyelle [u] ou [d] en contexte de voyelle [i]
- Conclusions : résistance du trait de voisement des occlusives après cordectomie de type II-III.

## 6.9 Conclusion

Nous avons déjà évoqué les conséquences anatomico-physiologiques engendrées par les cordectomies de type II-III, en particulier une asymétrie cordale et un accoulement incomplet des plis vocaux. Or ces perturbations de la source vocale pourraient être responsables d'une mauvaise gestion du trait de voisement. Ainsi, nous avons voulu montrer l'impact de la chirurgie sur le maintien de l'opposition de voisement à travers des études aérodynamiques, acoustiques, électroglottographiques et perceptives.

Notre première étude aérodynamique nous a permis de confirmer l'hypothèse de l'existence de fuites glottiques que nous avions postulée dans notre chapitre précédent. Bien que nos résultats doivent être considérés avec prudence du fait d'une importante variabilité individuelle, nous pouvons confirmer l'incompétence laryngée caractérisée par une hausse des débits d'air chez ces patients. Cette fermeture glottique insuffisante mise en évidence était la principale raison qui nous a poussé à nous intéresser à l'implémentation du voisement chez ces sujets. Après avoir démontré la bonne synchronisation des paramètres de la pression intra-orale et du débit d'air buccal dans nos prises de mesures, il nous a semblé pertinent de caractériser de manière quantitative le comportement de ces derniers chez nos sujets dysphoniques. De manière générale, les patients semblent marquer le contraste de voisement avec des valeurs de PIO, de Umax et de DPM plus élevées en contexte non voisé. Néanmoins, nous notons des valeurs toujours supérieures par rapport à notre cohorte de témoins, ce qui traduit soit un forçage vocal, soit une difficulté de voisement. Cette seconde hypothèse est validée par des mesures estimées de pression transglottique assez élevées qui démontrent une difficulté certaine à initier et maintenir le voisement. Ainsi, nous avons émis l'hypothèse que les patients tendent à éléver naturellement leur pression transglottique afin de faciliter l'initiation des vibrations glottiques et ce, pour compenser la mutilation

cordale. Ces résultats vont dans le sens des conclusions établies sur l'élévation de la  $F_0$  après cordectomie (cf. chapitre 5). En effet, les changements anatomiques des plis vocaux induits par la chirurgie impliquent une hausse de la  $F_0$ . Or, plus les plis vocaux sont tendus et plus les sujets doivent éléver leur pression transglottique afin de maîtriser l'importante résistance laryngée (Titze 1988).

Dans notre analyse acoustique, nous nous sommes particulièrement intéressés aux paramètres des durées segmentales, de l'intensité relative du relâchement, de la  $F_0$  au début des voyelles et du VOT. Une fois encore, et à l'instar des témoins, les patients maintiennent une opposition entre occlusives voisées et non voisées. Ainsi, leurs consonnes non voisées sont plus longues que leurs contreparties voisées, les voyelles sont plus longues devant des consonnes voisées, l'intensité des relâchements est plus forte en contexte non voisé, la  $F_0$  des voyelles tend à être plus importante en contexte non voisé et les valeurs du VOT restent comparables à celles de la littérature. Outre les mesures de  $F_0$  et un débit ralenti, nous ne retrouvons aucune différence statistiquement significative entre nos deux groupes, ce qui suggère une bonne maîtrise du *timing* laryngé par nos patients pour l'opposition de voisement. Ces résultats acoustiques satisfaisants mettent en évidence la bonne synchronisation entre les gestes laryngés et supralaryngés, facilitée par une hypothétique compensation aérodynamique.

Une troisième étude EGG pilote, nous a permis de caractériser la qualité de voix des patients cordectomisés et de montrer que cet outil pouvait s'avérer efficace dans l'étude des différences entre occlusives voisées et non voisées. Ainsi, malgré des variations intra-sujets, nous avons démontré que la majorité des patients présentait des valeurs de quotient d'ouverture (Oq) élevées, ce qui implique des phases d'ouverture glottique allongées. Ces résultats attendus corroborent nos conclusions quant à l'existence de fuites glottiques dues à une fermeture partielle des plis vocaux. De même, nous avons pu mettre en évidence une certaine élévation des valeurs de Oq des patients au début des voyelles en contexte non voisé, ce qui représente à la fois un argument en faveur de cette instrumentation pour la distinction de voisement et une justification supplémentaire du maintien du contraste chez nos sujets pathologiques. En effet, les différences de Oq des voyelles produites en contexte voisée ou non voisé impliquent que le voisement des consonnes influence le mode vibratoire de ces voyelles. Ainsi, la tension des plis vocaux est plus importante après une occlusive non voisée, ce qui induit une réduction de l'adduction au début du voisement de la voyelle adjacente, définie par une élévation des valeurs de Oq. De plus, le fait que les patients adoptent ces différentes stratégies montre que ces derniers sont sensibles aux ajustements articulatoires nécessaires à l'opposition de voisement des occlusives.

Enfin, notre dernière étude perceptive nous a permis de confirmer la résistance du voisement des occlusives des patients pour les deux positions étudiées. En effet, malgré des confusions mineures, les importants taux d'identification des consonnes montrent une bonne intelligibilité de ces dernières et donc un voisement robuste. Ces résultats sont en adéquation avec nos précédentes études, qui soulignent une certaine difficulté dans l'initiation du voisement mais qui prouvent surtout la préservation de ce trait distinctif pour l'opposition des consonnes occlusives non voisées et voisées.

Finalement, comme nous l'évoquions dans le chapitre précédent, les cordectomies laser de type II-III, bien qu'invasives sur le plan physiologique, garantissent d'excellents résultats vocaux avec une implémentation du trait de voisement intacte.

#### En résumé...

- Ce chapitre met en évidence l'insuffisance laryngée, caractérisée par l'existence de fuites glottiques se traduisant par des voix "breathy" après de telles chirurgies.
- L'élévation des paramètres aérodynamiques des patients traduit une certaine difficulté dans l'initiation et le maintien du voisement mais les différences de mesures en contexte voisé et non voisé témoignent du maintien du contraste.
- Les résultats acoustiques satisfaisants des patients sont des arguments supplémentaires en faveur d'un trait robuste.
- Le caractère soufflé des voix des patients est confirmé par l'EGG. Cette instrumentation est pertinente car elle permet de montrer l'influence du voisement des consonnes sur les voyelles adjacentes.
- Nos hypothèses sont confirmées par des tests de perception qui montrent une importante intelligibilité des occlusives, et donc une implémentation intacte du trait de voisement après cordectomie de type II-III.



# Conclusion générale

"C'est le commencement qui est le pire, puis le milieu  
puis la fin ; à la fin, c'est la fin qui est le pire."  
— Samuel Beckett

Cette thèse propose une analyse de la voix et de la parole dans un contexte pathologique. L'originalité de ce travail réside d'une part, dans le choix d'une population homogène et d'autre part, dans la constitution de corpus acoustiques, aérodynamiques et électroglottographiques (EGG) spécifiques. Ainsi, nous avons pu recueillir différentes données physiologiques auprès d'une population de 10 patients ayant subi une cordectomie laser de type II-III et appariée avec une cohorte de 10 sujets sains. Les problématiques centrales de ce travail sont (i) le contrôle de la maîtrise de la fréquence fondamentale, de l'intensité et de leurs facteurs de perturbation et (ii) le contrôle du maintien du trait de voisement après une telle chirurgie.

Afin de répondre à cette première question, nous avons analysé les productions vocaliques — produites à l'isolé ou, selon une tâche d'imitation de segments synthétisés — de nos sujets. Nous avons pu mettre en évidence qu'il n'existant aucune différence significative entre nos deux groupes pour le jitter et le shimmer, ce qui justifie ces chirurgies. Néanmoins, l'altération de la fréquence fondamentale, engendrée par une modification de la structure anatomique du pli vocal touché, démontre le principe invasif des cordectomies. Le laser est responsable d'une augmentation de la raideur des plis vocaux, d'une diminution de la masse vibrante et de la longueur vibrante, ce qui concourent à une élévation de la fréquence vibratoire. Un second versant intéressant de cette étude concerne le paramètre de l'intensité sonore. Nous avons principalement montré que ce paramètre reste difficile à interpréter car la prise de données est délicate d'un point de vue méthodologique et la variation individuelle est omniprésente. Ainsi, une comparaison entre l'intensité des voyelles isolées et l'intensité des voyelles en contexte occlusif nous a particulièrement éclairé sur les stratégies mises en oeuvre selon les tâches de production. Pour résumer, les intensités spécifiques des voyelles des patients ne sont pas différentes de celles des témoins. Les patients tendent à produire des segments vocaliques plus forts à l'isolé, ce que nous interprétons comme un effort supplémentaire afin de bien faire l'exercice ou de surpasser le handicap vocal qu'ils pensent perdurer. Enfin, ces différentes études ont mis en évidence la difficulté des patients pour la production des voyelles hautes, en particulier. En effet, ces segments vocaliques impliquent des contraintes articulatoires et aérodynamiques plus complexes par rapport à la voyelle [a].

Afin d'observer l'implémentation du trait de voisement après cordectomie de type II-III, nous avons réalisé des études aérodynamiques, acoustiques, EGG et perceptives à partir de logatomes composés d'occlusives, en particulier pour la position d'attaque

en initial absolu et la position d'attaque en intervocalique. Nous avons confirmé l'existence de fuites glottiques, caractérisées par une fermeture glottique insuffisante, ce qui pourraient représenter un obstacle pour la réalisation d'un voisement complet. Cependant, l'étude de divers paramètres aérodynamiques — PIO, Umax et DPM — et acoustiques — durées segmentales, intensité relative du relâchement,  $F_0$  au début des voyelles adjacentes et VOT — révèlent le maintien d'un contraste de voisement intact entre occlusives voisées et non voisées. L'augmentation de la pression intraorale et de la pression transglottique estimée chez les patients démontrent un mécanisme compensatoire afin de lutter contre l'importante résistance laryngée et ainsi, de faciliter l'initiation et le maintien du voisement quand cela est nécessaire. Aussi, une étude EGG pilote nous a permis, à la fois, de confirmer la qualité de voix « *breathy* » de ces patients et de valider l'intérêt de ce type de méthodologie pour la distinction de voisement. En effet, le voisement des occlusives influent sur le mode vibratoire des voyelles adjacentes, ce qui se traduit par des valeurs de quotients d'ouverture différentes au début de ces voyelles. Le fait que les patients adoptent ces différences est un indice supplémentaire dans leur maîtrise du trait de voisement. Enfin, malgré quelques confusions mineures, nos tests perceptifs confirment l'intelligibilité des consonnes dans les deux positions, ce qui est la preuve d'un voisement intact et robuste après ce type de chirurgie.

Finalement, ce travail tend à montrer l'efficacité de la cordectomie laser de type II-III tant du point de vue des résultats carcinologiques que des résultats fonctionnels. Ainsi, malgré la perte de qualités esthétiques, ces voix restent intelligibles avec préservation de la fonction linguistique du larynx, des points de vue des fréquences intrinsèques des voyelles ou de l'identification du trait de voisement, entre autres. La chirurgie reste mutilante d'un point de vue physiologique mais tend à conserver une fonction de communication intacte, à travers les paramètres que nous avons pu étudier.

Cependant, de nombreuses questions restent en suspens. Nous sommes consciente que nos hypothèses physiologiques quant à l'élévation de la  $F_0$  des patients devront faire l'objet d'investigations fibroscopiques supplémentaires. De même, le comportement de la muqueuse et du vibrateur seraient mieux évalués à travers des études en caméra ultra-rapide. Nos études quant au trait de voisement pourraient également être approfondies grâce à l'utilisation de l'ePGG (external photoglottography) qui nous permettrait d'évaluer plus simplement le contact des plis vocaux, à travers l'estimation des variations de l'aire glottique. En outre, nos mesures de pressions pourraient être couplées avec des mesures directes de la PSG qui confirmeraient les hypothèses développées dans cette thèse. Enfin, nous suggérons l'utilisation de nouveaux protocoles, incluant les fricatives par exemple, ainsi que le recours à de nouvelles populations afin de comprendre au mieux les mécanismes mis en jeu et de caractériser le plus fidèlement possible ces voix particulières.

# Bibliographie

- Abdelli-Beruh, NB. (2004). The stop voicing contrast in French sentences : contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica* **61**, 201–219.
- Ainsworth, WA. (1973). Durational cues in the perception of certain consonants. In *Proceedings of the British Acoustical Society*, volume 2, 1–4.
- Airas, M. (2008). *Methods and studies of laryngeal voice quality analysis in speech production*. Ph.D. thesis, University of Technology, Helsinki.
- Arkebauer, H.J., T. Hixon & J.C. Hardy (1967). Peak intraoral air pressure during speech. *Journal of Speech and Hearing Research* **10**, 196–208.
- Arnoux-Sindt, B. (1992). Articulation crico-aryténoïdienne et laryngectomie reconstructive. *Revue de Laryngologie, rhinologie, otologie* **113**, 339–342.
- Atkinson, J. (1973). *Aspects of intonation in speech : Implications from an experimental study of fundamental frequency*. Ph.D. thesis, University of Connecticut, Storrs.
- Avanzini, F., P. Alku & M. Karjalainen (2001). One-delayed-mass model for efficient synthesis of glottal flow. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, 51–54. Aalborg, Danemark.
- Bailly, L. (2009). *Interaction entre cordes vocales et bandes ventriculaires en phonation : exploration in-vivo, modélisation physique, validation in-vitro*. Ph.D. thesis, Université du Maine.
- Baken, R. (1992). Electroglossography. *Journal of Voice* **6**, 98–110.
- Baken, RJ. (1987). *Clinical measurement of speech and voice*. Boston : College-Hill Press.
- Baken, R.J. & R.F. Orlikoff (2000). *Clinical measurement of speech and voice*. San Diego : Singular Publishing Group.
- Bartkova, K., P. Haffner & D. Larreur (1993). Intensity prediction for speech synthesis in French. In *Proceedings of ESCA Workshop on Prosody 1993*, 281–283. University of Lund : Sweden.
- Bartkova, K. & C. Sorin (1987). A model of segmental duration for speech synthesis in French. *Speech Communication* **6**, 245–260.

- Basset, P., A. Amelot & L. Crevier Buchman (2006). Etude multiparamétrique des consonnes nasales du français : prise de données simultanées aérodynamiques et fibroscopiques. *Parole* **39-40**, 113–136.
- Behrman, A., L. Sulica & T. He (2004). Factors predicting patient perception of dysphonia caused by benign vocal fold lesions. *Laryngoscope* **114**, 1693–1700.
- Benguerrel, AP., H. Hirose, M. Sawashima & T. Ushijima (1978). Laryngeal control in French stop production : a fibrescopic, acoustic and electromyographic study. *Folia Phoniatrica* **30**, 175–198.
- Benninger, MS., AS. Ahuja, G. Gardner & C. Grywalski (1998). Assessing outcomes for dysphonic patients. *Journal of voice* **12**, 540—550.
- Benveniste, E. (1966). *Problèmes de linguistique générale (tome 1)*. Paris : France : Gallimard.
- Berke, G., DM. Moore, DR. Hantke, DG. Hanson & B. Gerratt (1987). Laryngeal modeling : Theoretical, in vitro, in vivo. *Laryngoscope* **97**, 871–881.
- Biacabe, B., L. Crevier Buchman, S. Hans, O. Laccourreye & D. Brasnu (1999). Vocal function after vertical partial laryngectomy with glottic reconstruction by false vocal fold flap : durational and frequency measures. *Laryngoscope* **109**, 678–704.
- Biacabe, B., L. Crevier Buchman, S. Hans, O. Laccourreye & D. Brasnu (2001). Phonatory mechanisms after vertical partial laryngectomy with glottic reconstruction by false vocal fold flap. *Ann Otol Rhinol Laryngol* **110**, 935–940.
- Biacabe, B., L. Crevier Buchman, O. Laccourreye & D. Brasnu (1998). Laryngectomie partielle verticale avec reconstruction glottique : Résultats carcinologiques et fonctionnels. *Ann Otolaryngol Chir Cervicofac* **115**, 195–198.
- Bickley, CA. & KN. Stevens (1986). Effects of vocal-tract constriction on the glottal source : experimental and modelling studies. *Journal of Phonetics* **14**, 373–382.
- Black, J. (1950). The pressure component in the production of consonants. *Journal of Speech and Hearing Research* **15**, 20 –210.
- Black, J.W. (1961). Relationships among fundamental frequency, vocal sound pressure and rate of speaking. *Language and Speech* **4**, 196–199.
- Boersma, P. & D. Weenink (1996). Praat, a system for doing phonetics by computer, version 3.4. Technical Report n°132, Institute of Phonetic Sciences of the University of Amsterdam. Téléchargeable à partir de [www.praat.org](http://www.praat.org).
- Bourcet, P. & P. Liénard (1987). Acoustique fondamentale. In D. Mercier (ed.), *Le livre des techniques du son, vol 1 : Notions fondamentales*, 13–43. Eyrolles.

- Brasnu, D., D. Ayache, D. Hartl, S. Hans & JF. Papon (2008). *Traité d'ORL*. Flammarion.
- Brasnu, D., O. Laccourreye, S. Hans, M. Menard, E. De Mones & E. Behm (2005). *La chirurgie conservatrice des cancers du larynx et du pharynx*. Paris : Amplifon.
- Brown, W. & R. McGlone (1974). Aerodynamic and acoustic study of stress in sentence productions. *Journal Acoustic Society of America* **56**, 971–974.
- Brown, W., R. McGlone, A. Tarlow & T. Shipp (1970). Intraoral air pressures associated with specific phonetic positions. *Phonetica* **22**, 202–212.
- Brown, WS. & RE. McGlone (1969). Constancy of intraoral air pressure. *Folia Phoniatrica* **21**, 332–339.
- Bucella, F., S. Hassid, R. Beeckmans, A. Soquet & D. Demolin (2000). Pression sous-glottique et débit d'air buccal des voyelles en français. In *Actes des 23ème Journées d'Etude sur la Parole (JEP)*, 449–452. Aussois : France.
- Burgess, L. (1993). Laryngeal reconstruction following partial laryngectomy. *Laryngoscope* **103**, 109–132.
- Byrd, DM. (1994). *Articulatory timing in English consonant sequences*. Ph.D. thesis, UCLA.
- Calliope (1989). *La parole et son traitement automatique*. Paris : Masson.
- Chafcouloff, M. (2004). Voir la parole. *Travaux Interdisciplinaires du Laboratoire Parole et Langage* **23**, 23–65.
- Chan, R., IR. Titze & M. Titze (1997). Further studies on phonation threshold pressure in a physical model of the vocal fold mucosa. *Journal Acoustic Society of America* **101**, 3722–3727.
- Charles-Luce, J. (1993). The effects of semantic context on voicing neutralization. *Phonetica* **50**, 28–43.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* **22**, 129–159.
- Chen, Y., MP. Robb & H. Gilbert (2002). Electroglossographic evaluation of gender and vowel effects during modal and vocal fry phonation. *Journal of Speech and Hearing Research* **45**, 821–829.
- Chevallier, G., R. Guilbault, R. Renard & P. Tran Ba Huy (2008). La voix saturée du chanteur de rock métal, un mécanisme performant. In *Actes des 64ème congrès de la Société Française de Phoniatrice et des Pathologies de la Communication*. Paris : France.

- Childers, D. & AK. Krishnamurthy (1985). A critical review of electroglottography. *CRC Critical Reviews in Biomedical Engineering* **12**, 131–151.
- Cho, T., S. Jun & P. Ladefoged (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* **30**, 193–228.
- Cho, T. & P. Ladefoged (1999). Variations and universals in VOT : evidence from 18 languages. *Journal of Phonetics* **27**, 207–229.
- Cho, Y. (1990). A typology of voicing assimilation. In A. Halpern (ed.), *Proceedings of the Ninth West Coast Conference on Formal Linguistics*, 141–155. Stanford : CSLI.
- Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New-York : Harper and Row.
- Claeys, F. (2010). *Evaluation multiparamétrique de la voix après cordectomie endoscopique au laser CO<sub>2</sub>*. Master's thesis, Université Paris 6.
- Cookman, S. & K. Verdolini (1999). Interrelation of mandibular and laryngeal functions. *Journal of Voice* **13**, 11–24.
- Cornut, G. (1959). *La mécanique respiratoire dans la parole et le chant*. Presses Universitaires de France.
- Cornut, G. (1983). *La voix*. Presses Universitaires de France.
- Crevier Buchman, L. (1999). *La voix et la parole sans cordes vocales après laryngectomie partielle supracricoidienne*. Ph.D. thesis, Université Paris 3.
- Crevier Buchman, L., S. Brihaye & C. Tessier (2003). *La voix après chirurgie partielle du larynx*. Paris : Solal.
- Crevier Buchman, L., O. Laccourreye, JF. Papon, MC. Monfrais-Pfauwadel & D. Brasnu (1996). Apports et limites de l'analyse acoustique de la voix et de la parole alaryngée au moyen d'un système informatique. *Annales d'Otolaryngologie et de Chirurgie Cervicofaciale* **113**, 61–68.
- Crevier Buchman, L., O. Laccourreye, G. Weinstein, D. Garcia, V. Jouffre & D. Brasnu (1995a). Evolution of speech and voice following supracricoid partial laryngectomy. *The Journal of Laryngology and Otology* **109**, 410–413.
- Crevier Buchman, L., O. Laccourreye, FL. Wuyts, M. Monfrais-Pfauwadel, C. Pillot & D. Brasnu (1998). Comparison and evolution of perceptual and acoustic characteristics of voice after partial laryngectomy with crico-hyoido-epiglottopexy. *Acta Otolaryngologica (Stockh)* **118**, 594–599.
- Crevier Buchman, L., S. Maeda, N. Bely, O. Laccourreye, J. Vaissière & D. Brasnu (2001). Compensations articulatoires après laryngectomie partielle supracricoidienne avec crico-hyoido-epiglottopexy. *Ann Otolaryngol Chir Cervicofac* **118**, 81–88.

- Crevier Buchman, L., E. Martigny, C. Gate, C. Pillot, C. Tessier, MC. Monfrais-Pfauwadel & D. Brasnu (1995b). Evaluation subjective de la voix et de la parole après laryngectomie partielle supra-cricoidienne. *Revue de Laryngologie, rhinologie, otologie* **116**, 273–276.
- Crevier Buchman, L., J. Vaissière, S. Maeda & D. Brasnu (2002). Etude de l'intelligibilité des consonnes du français après laryngectomie partielle supracricoidienne. *Revue de Laryngologie, rhinologie, otologie* **123**, 307–310.
- Crevier Buchman, L., C. Vincent, S. Maeda, S. Hans & D. Brasnu (2008). Comportements laryngés en voix chuchotée, étude en caméra ultra rapide. In *Actes des 64ème congrès de la Société Française de Phoniatrie et des Pathologies de la Communication*. Paris : France.
- Dagli, AF., HF. Mahieu & JM. Festen (1997). Quantitative analysis of voice quality in early glottic laryngeal carcinomas treated with radiotherapy. *Eur Arch Otorhinolaryngol* **254**, 78–80.
- De Krom, G. (1994). Consistency and reliability of voice quality ratings for different types of speech fragments. *Journal of Speech and Hearing Research* **37**, 985–1000.
- Dejonckere, P., C. Obbens, GM. De Moor & GH. Wieneke (1993). Perceptual evaluation of dysphonia : reliability and relevance. *Folia Phoniatrica* **45**, 76–83.
- Dejonckere, P.H. (1981). *Theorie oscillo-impédantielle de la vibration des cordes vocales*. Ph.D. thesis, Université Catholique de Louvain, Belgique.
- Dejonckere, P.H. (1985). *Techniques de base d'évaluation de la voix*. Louvain-La-Neuve : Cabay.
- Dejonckere, PH., M. Remacle, E. Fresnel-Elbaz, V. Woisard, L. Crevier Buchman & B. Millet (1996). Differentiated perceptual evaluation of pathological voice quality : reliability and correlation with acoustic measurements. *Revue de Laryngologie, rhinologie, otologie* **117**, 219–224.
- Dejonckere, PH., M. Remacle, E. Fresnel-Elbaz, V. Woisard, L. Crevier Buchman & B. Millet (1998). Reliability and clinical relevance of perceptual evaluation of pathological voices. *Revue de Laryngologie, rhinologie, otologie* **119**, 247–248.
- Delattre, PC. (1958). Les indices acoustiques de la parole. *Phonetica* **2**, 108–118, 226–251.
- Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua* **95**, 5–26.
- Delsupehe, KG., I. Zink, M. Lejaegere & RW. Bastian (1999). Voice quality after narrow-margin laser cordeectomy compared with laryngeal irradiation. *Otolaryngol Head Neck Surg* **121**, 528–533.

- Delvaux, V. & B. Harmegnies (2011). Nasalité contextuelle et aperture vocalique en français : un étude aérodynamique. In M. Embarki & C. Dodane (eds.), *La coarticulation : des indices à la représentation*. Paris : France : L'Harmattan.
- Descartes, R. (1840). *Discours de la méthode*. Paris : France : Hachette.
- Di Canio, C. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association* **39**, 162–188.
- Di Cristo, A. (1982). *Prolegomenes à l'étude de l'intonation : micromélodie*. Paris : Editions du Centre National de la Recherche Scientifique.
- Di Cristo, A. (1985). *De la microprosodie à l'intonosynataxe*. Université de Provence.
- Di Cristo, A., DJ. Hirst & Y. Nishinuma (1979). L'estimation de la F0 intrinsèque des voyelles : étude comparative. *Travaux de L'Institut de Phonétique d'Aix-en-Provence* **6**, 149–176.
- Dixit, RP. (1989). Glottal gestures in Hindi plosives. *Journal of Phonetics* **17**, 213–237.
- Draper, M., P. Ladefoged & D. Whitteridge (1960). Expiratory pressures and air flow during speech. *British Medical Journal* **118**, 1837–1843.
- Draper, MH., P. Ladefoged & D. Whitteridge (1959). Respiratory muscles in speech. *Journal of Speech and Hearing Research* **2**, 16–27.
- Dromey, C., L. Raming & A. Johnson (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease : A case study. *Journal of Speech and Hearing Research* **38**, 751–763.
- Duez, D. (2007). Effets de la maladie de Parkinson sur la réalisation acoustique des occlusives du français lu. *Travaux Interdisciplinaires du Laboratoire Parole et Langage* **26**, 15–31.
- Dyhr, N. (1990). The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica* **47**, 141–154.
- Emanuel, F. & D. Counihan (1970). Some characteristics of oral and nasal air flow during plosive consonant production. *Cleft Palate Journal* **7**, 249–260.
- Eskenazy, L., DG. Childers & DM. Hicks (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research* **33**, 298–306.
- Ewald, J.R. (1898). Die Physiologie der Kehlkopfes und der Luftröhre, Stimmbildung. In P. Heymann (ed.), *Handbuch der Laryngologie und Rhinologie*, 165–205. Vienne : Alfred Hölder.

- Faaborg-Andersen, K. (1957). Electromyographic investigation of intrinsic laryngeal muscles in humans. *Physiol Scand* **41**, 1–148.
- Faaborg-Andersen, K. & A. Sonninen (1959). The function of extrinsic laryngeal muscles at different pitch. *Acta Otolaryngol* **51**, 89–93.
- Fairbanks, G., AS. House & EL. Stevens (1950). An experimental study of vowel intensities. *Journal Acoustic Society of America* **22**, 457–459.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague : Mouton.
- Fant, G., A. Kruckenberg, K. Gustafson & J. Liljencrants (2002). A new approach to intonation analysis and synthesis of Swedish. In *Proceedings of the 1st International Conference on Speech Prosody*, 283–286. Aix en Provence.
- Fant, G. & Q. Lin (1987). Glottal source - Vocal Tract Interaction. *STL-QPSR* **1**, 13–27.
- Fex, S. (1992). Perceptual evaluation. *Journal of voice* **6**, 155–158.
- Fink, B.R. & R.J. Desmarest (1978). *Laryngeal biomechanics*. Cambridge : Harvard University Press.
- Finnegan, E.M., E.I. Luschei & H.T. Hoffman (2000). Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech. *Journal of Speech and Hearing Research* **43**, 934–950.
- Fisher-Jorgensen, E. (1972). Ptk et bdg français en position intervocalique accentuée. In A. Valdman (ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre*, 143–200. La Haye : Pays-Bas : The Hague : Mouton.
- Flack, K. (2007). *The sources of phonological markedness*. Ph.D. thesis, University of Massachusetts, Amherst.
- Flanagan, J.L. & L. Landgraf (1968). Self-oscillating source for vocal tract synthesizers. *IEEE Trans-audio Electroacoustic* **16**, 57–64.
- Fletcher, H., GM. Raff & F. Parmley (1918). Study of the effects of different sidetones in the telephone set. Technical Report n° 19412, Case n° 120622, Western Electrical Company.
- Fonagy, I. (1966). Electrophysiological and acoustic correlates of stress perception. *Journal of Speech and Hearing Response* **9**, 231–244.
- Fourcin, A. (2000). Voice quality and electrolaryngography. In RD. Kent & MJ. Ball (eds.), *Voice quality measurement*, 285 –306. Singular Publishing Group.
- Fourcin, A.J. (1974). Laryngographic examination of vocal fold vibration. In B. Wyke (ed.), *Ventilatory and phonatory function*, 315–326. London : Oxford University Press.

- Galindo, B. & B. Teston (1989). Physiologia, une station de travail qui étudie des paramètres physiologiques de la voix. *Travaux interdisciplinaires du laboratoire Parole et Langage d'Aix -en-Provence* **13**, 197–217.
- Geumann, A. (2001). Vocal intensity : acoustic and articulatory correlates. In *Proceedings of 4th Conference on Motor Control*, 70–73. Nijmegen.
- Geumann, A., C. Kroos & HG. Tillmann (1999). Are there compensatory effects in natural speech ? In *Proceedings of 14th International Congress of Phonetic Sciences*, 399–402. San Francisco.
- Ghio, A. (2002). Phonedit : Multiparametric speech analysis. LPL (Université d'Aix-en-Provence), [www.lpl.univ-aix.fr](http://www.lpl.univ-aix.fr).
- Ghio, A. (2007). L'onde sonore : réalités physiques et perception. In P. Auzou, V. Rolland, S. Pinto & C. Ozsancak (eds.), *Les dysarthries*, 81–90. Marseille : Solal.
- Ghio, A., A. Giovanni, B. Teston, J. Révis, P. Yu, M. Ouaknine, D. Robert & T. Legou (2008). Bilan et perspectives de quinze ans d'évaluation vocale par méthodes instrumentales et perceptives. In *Journées d'Etude sur la Parole (JEP)*. Avignon : France.
- Gilbert, H. (1973). Oral airflow during stop consonant production. *Folia Phoniatrica* **25**, 288–296.
- Giovanni, A., N. Estublier, D. Robert, B. Teston, M. Zanaret & M. Cannoni (1995). Evaluation vocale objective des dysphonies par la mesure simultanée de paramètres acoustiques et aérodynamiques à l'aide de l'appareillage EVA. *Ann Otolaryngol Chir Cervicofac* **112**, 85–90.
- Giovanni, A., B. Guelfucci, R. Gras, P. Yu & M. Zanaret (2001). Partial frontolateral laryngectomy with epiglottic reconstruction for management of early-stage glottic carcinoma. *Laryngoscope* **111**, 663–668.
- Giovanni, A., B. Guelfucci, P. Yu, D. Robert & M. Zanaret (2002). Acoustic and aerodynamic measurements of voice production after near-total laryngectomy with epiglottoplasty. *Folia Phoniatrica and Logopedia* **54**, 304–311.
- Giovanni, A., C. Heim, D. Demolin & JM. Triglia (2000). Estimated subglottic pressure in normal and dysphonic subjects. *Annals of Otology, Rhinology and Laryngology* **109**, 500–504.
- Giovanni, A., D. Robert, B. Teston, MD. Guarelia & M. Zanaret (1996). Etude préliminaire des paramètres acoustiques et aérodynamiques après laryngectomie frontale antérieure de Tucker. *Annales d'Otolaryngologie et de Chirurgie Cervicofaciale* **113**, 277–284.

- Goldman, JP. (2008). Easy Align : a semi-automatic phonetic alignment tool under Praat. Technical report, Université de Genève. Téléchargeable à partir <http://latl-cui.unige.ch/phonetique/easyalign>.
- Goor, KM., AJ. Peeters, HM. Mahieu, JA. Langendijk, CR. Leemans, IM. Verdonck de Leeuw & M. Van Agthoven (2006). Cordeectomy by CO<sub>2</sub> laser or radiotherapy for small T1a glottic carcinomas : costs, local control, survival, quality of life and voice quality. *Head Neck* **29**, 128–136.
- Goozée, JV., BE. Murdoch Head, DG. Theodoros & EC. Thompson (1998). The effects of age and gender on laryngeal aerodynamics. *International Journal of Language & Communication Disorders* **33**, 221–238.
- Grijzenhout, J. & M. Krämer (2000). Final devoicing and voicing assimilation in Dutch derivation and cliticization. In B. Stiebels & D. Wunderlich (eds.), *Lexicon in Focus. Studia Grammatica 45*, 55–82. Berlin : Akademie-Verlag.
- Guerrier, B., A. Giovanni & M. Remacle (2004). *Pathologie de la corde vocale chez l'adulte*. Société française d'ORL et de chirurgie de la face et du cou.
- Haddad, L., M. Abrabao, O. Cervantes, FP. Ceccon, I. Gtelow, JR. Carvalho & FD. Leonhardt (2006). Vocal assessment in patients submitted to CO<sub>2</sub> laser cordeectomy. *Rev Bras Otorrinolaringol* **72**, 295–302.
- Halle, M. & K. Stevens (2002). A note on laryngeal features. In M. Halle (ed.), *From memory to speech and back : papers on phonetics and phonology 1954-2002*. Berlin : Mouton de Gruyter.
- Hammarberg, G., B. Fritzell, J. Gauffin, J. Sundberg & L. Wedin (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngologica* **90**, 441–451.
- Hammarberg, G. & J. Gauffin (1995). Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In O. Fujimura & M. Hirano (eds.), *Vocal fold physiology : voice quality control*. San Diego : Singular Publishing Group.
- Hans, S. (2003). *Etude aérodynamique comparée de logatomes chez 120 sujets normaux et 38 patients avec résection d'un ou des deux plis vocaux*. Ph.D. thesis, Université Paris 3.
- Hanson, HM. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal Acoustic Society of America* **125**, 425–441.
- Hardcastle, WJ. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics* **1**, 263–272.

- Henrich, N. (2001). *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception.* Ph.D. thesis, Université de Paris 6.
- Henrich, N., C. D'Alessandro, B. Doval & M. Castellengo (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal Acoustic Society of America* **115**, 1321–1332.
- Heselwood, B. & Z. Mahmoodzade (2007). Vowel onset characteristics as a function of voice and manner contrasts in Persian coronal stops. *Leeds working Papers in Linguistics and Phonetics* **12**, 125–142.
- Heuillet, G., H. Garson Bavard & A. Legré (1995). *Une voix pour tous, la voix normale et comment l'optimiser.* Solal.
- Higgins, MB., R. Netsell & L. Schulte (1998). Vowel-related differences in laryngeal articulatory and phonatory function. *Journal of Speech and Hearing Research* **41**, 712–724.
- Hirano, M. (1977). Structure and vibratory behavior of the vocal folds. In M. Sawashima & FS. Cooper (eds.), *Dynamic aspects of speech production*, 13–27. Tokyo : University of Tokyo Press.
- Hirano, M. (1981). *Clinical examination of voice.* Vienne : Springer Verlag.
- Hirano, M., S. Hibi, R. Terasawa & M. Fujii (1986). Relationship between aerodynamic vibratory, acoustic and psychoacoustic correlates in dysphonia. *Journal of Phonetics* **14**, 445–456.
- Hirano, M., Y. Koike & H. Von Leden (1968). Maximum phonation time and air usage during phonation. *Folia Phoniatrica* **20**, 185–201.
- Hirsch, F. (2007). *Le bégaiement : perturbation de l'organisation temporelle de la parole et conséquences spectrales.* Ph.D. thesis, Université Marc Bloch.
- Hiss, S.G., K. Treole & A. Stuart (2001). Effect of age, gender and repeated measures on intraoral air pressure in normal adults. *Journal of Voice* **15**, 159–164.
- Hixon, T. (1966). Turbulent noise sources for speech. *Folia Phoniatrica* **19**, 168–182.
- Hixon, TJ., MD. Goldman & J. Mead (1973). Kinematics of the chest wall during speech production : Volume displacements of the rib cage, abdomen and lung. *Journal of Speech and Hearing Research* **16**, 78–115.
- Hixon, TJ., J. Mead & MD. Goldman (1976). Dynamics of the chest wall during speech production : Function of the thorax, rib cage, diaphragm and abdomen. *Journal of Speech and Hearing Research* **19**, 297–356.

- Hogikyan, N. & C. Rosen (2002). A review of outcome measurements for voice disorders. *Otolaryngol Head Neck Surg* **126**, 562–572.
- Hogikyan, N. & G. Sethuraman (1999). Validation of an instrument to measure voice related quality of life (V-RQOL). *Journal of voice* **13**, 557–569.
- Hollien, H., J. Michel & E.T. Doherty (1973). A method for analyzing vocal jitter in sustained phonation. *Journal of Phonetics* **1**, 85–91.
- Holmberg, E., R. Hillman & J. Perkell (1988a). Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal and high pitch. *Journal of Voice* **3**, 294–305.
- Holmberg, EB., R. Hillman & JS. Perkell (1988b). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *Journal Acoustic Society of America* **84**, 511–529.
- Hombert, JM., J. Ohala & W. Ewan (1979). Phonetic explanations for the development of tones. *Language* **55**, 37–58.
- Honda, K. (2004). Physiological factors causing tonal characteristics of speech : from global to local prosody. In *Proceedings of Speech Prosody 2004*, 739–744. Nara : Japon.
- Honda, K. & O. Fujimura (1991). Intrinsic vowel F0 and phrase final F0 lowering : Phonological vs. biological explanations. In J. Gauffin & B. Hammarberg (eds.), *Vocal fold physiology : acoustic, perceptual and physiological aspects of voice mechanisms*, 149–157. San Diego : Singular Publishing Group.
- Honda, K., H. Hirai, Masakin S. & Y. Shimada (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* **42**, 401–411.
- Hoole, P. & K. Honda (2011). Automaticity vs. feature-enhancement in the control of segmental F0. In N. Clements & R. Ridouane (eds.), *Where do phonological features come from ? Cognitive, physical and developmental bases of distinctive speech categories*, 131–171. John Benjamins Publishing Company.
- House, A. & G. Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal Acoustic Society of America* **25**, 105–113.
- Howard, DM. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice* **9**, 163–172.
- Hsiung, M. W., L. Pai & H. W. Wang (2002). Correlation between voice handicap index and voice laboratory measurements in dysphonic patients. *Eur Arch Otorhinolaryngol* **259**, 97–99.

- Husson, R. (1950). *Etude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée*. Ph.D. thesis, Faculté des sciences de Paris, Paris.
- Hyman, L. & RG. Schuh (1974). Universals of tone rules : Evidence from West Africa. *Linguistic Inquiry* **5**, 81–115.
- Ishizaka, K. & J.L. Flanagan (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Sys Tech Journal* **51**, 1233–1268.
- Isshiki, N. (1964). Regulatory mechanism of voice intensity variation. *Journal of Speech and Hearing Research* **7**, 17–29.
- Isshiki, N. (1965). Vocal intensity and air flow rate. *Folia Phoniatrica* **17**, 92–104.
- Isshiki, N., H. Okamura & M. Morimoto (1967). Maximal phonation time and air flow rate during phonation : simple clinical tests for vocal function. *Annals of Otology, Rhinology and Laryngology* **76**, 998–1007.
- Isshiki, N., H. Okamura, M. Tanabe & M. Morimoto (1969). Differential diagnosis of hoarseness. *Folia Phoniatrica (Basel)* **21**, 9–19.
- Iwata, S., H Von Leden & D. Williams (1972). Air flow measurement during phonation. *Journal of Communication Disorders* **5**, 67–79.
- Jacobson, BH., A. Johnson & C. Grywalski (1997). The voice handicap index (VHI) : development and validation. *Am J Speech Lang Pathol* **6**, 66–70.
- Jakobson, R. (1932). Phoneme and Phonology. *Selected Writing* **1**, 231–233.
- Jakobson, R., CM. Fant & M. Halle (1952). *Preliminaries to speech analysis. The distinctive features and their correlates*. Cambridge Massachussets : MIT Press.
- Jassem, W. & L. Richter (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics* **17**, 317–325.
- Jiang, JJ., Y. Zhang & J. Stern (2001). Modeling of chaotic vibrations in symmetric vocal folds. *Journal Acoustic Society of America* **110**, 2120–2128.
- Jotic, A., P. Stankovic, S. Jesic, J. Milovanovic, M. Stojanovic & V. Djukic (2012). Voice quality after treatment of early glottic carcinoma. *Journal of Voice* **26**(3), 381–389.
- Jun, SA. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. thesis, Ohio State University.
- Karnell, MP. & CR. Willis (1982). The effect of vowel context on consonantal intraoral air pressure. *Folia Phoniatrica* **34**, 1–8.

- Kasuya, H., S. Ogawa, K. Mashima & S. Ebihara (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. *Journal Acoustic Society of America* **80**, 1329–1334.
- Kavitskaya, D. (2001). *Compensatory lengthening : phonetics, phonology, diachrony*. Ph.D. thesis, University of California, Berkeley.
- Keating, P., T. Cho, C. Fougeron & CS. Hsu (1999). Domain-initial strengthening in four languages. *UCLA Working Papers in Phonetics* **97**, 137–151.
- Kent, RD. (1993). Vocal tract acoustics. *Journal of Voice* **7**, 97–117.
- Kitajima, K. & F. Fujita (1992). Clinical report on preliminary data on intraoral pressure in the evaluation of laryngeal pathology. *Journal of Voice* **6**, 79–85.
- Klatt, D. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* **18**, 686–706.
- Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal Acoustic Society of America* **67**, 971–995.
- Klatt, DH. & LC. Klatt (1990). Analysis, synthesis and perception of voice quality variations among female and men talkers. *Journal Acoustic Society of America* **87**, 820–857.
- Klich, RJ. (1982). Effects of speech level and vowel context on intraoral air pressure in vocal and whispered speech. *Folia Phoniatrica* **34**, 33–40.
- Kluender, KR., RL. Diehl & BA. Wright (1988). Vowel-length differences before voiced and voiceless consonants : an auditory explanation. *Journal of Phonetics* **16**(2), 153–169.
- Koenig, L., S. Fuchs & J. Lucero (2011). Effects of consonant manner and vowel height on intraoral pressure and articulatory contact at voicing offset and onset for voiceless obstruent. *Journal Acoustic Society of America* **129**, 3233–3244.
- Koike, Y., H. Takahashi & TC. Calcaterra (1977). Acoustic measures for detecting laryngeal pathology. *Acta Otolaryngologica* **84**, 105–117.
- Koufman, J. (1991). Approach to the patient with a voice disorder. *Otolaryngol Clin North Am* **24**, 989–998.
- Kreiman, J., B. Gerratt, C. Precoda & G. Berke (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research* **35**, 512–520.
- Kreiman, J., BR. Gerratt, G. Kempster, A. Erman & GS. Berke (1993). Perceptual evaluation of voice quality : Review, tutorial, and a framework of future research. *Journal Speech Hearing Research* **36**, 21–40.

- Krengli, M., M. Policarpo, I. Manfredda, P. Aluffi, G. Gambaro, M. Panella & F. Pia (2004). Voice quality after treatment for T1a glottic carcinoma : radiotherapy vs laser cordectomy. *Acta oncologica* **43**, 284–289.
- Kulikov, V. (2010). Voicing and vowel raising in Sundanese. In *Proceedings of the 17th annual meeting of the Austronesian Formal Linguistics Association*. Stony Brook, NY.
- Labayle, JBR. (1971). Laryngectomie totale avec reconstruction. *Ann Otolaryngol Chir Cervicofac* **88**, 219–228.
- Laccourreye, H. (1993). La chirurgie partielle des épithéliomas de l'étage glottique. Evolution de nos indications. *Bulletin de l'Académie Nationale de Médecine* **177**, 1333–1341.
- Laccourreye, H. (2000). Traitement chirurgical du cancer du larynx au cours du XXème siècle. *Ann Otolaryngol Chir Cervicofac* **117**, 237–247.
- Laccourreye, O., L. Crevier Buchman, B. Weinstein, G. and Biacabe, H. Laccourreye & D. Brasnu (1995). Duration and frequency characteristics of speech and voice following supracycoid partial laryngectomy. *Ann Otol Rhinol Laryngol* **104**, 516–519.
- Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics* **1**, 73–83.
- Ladefoged, P. (1983). The linguistic use of different phonation types. In D. Bless & J. Abbs (eds.), *Vocal fold physiology : Contemporary research and clinical issues*, 351–360. San Diego : College Hill Press.
- Ladefoged, P. & I. Maddieson (1996). *The sounds of the world's languages*. Blackwell Publishers.
- Ladefoged, P. & N. McKinney (1963). Loudness, sound pressure and subglottal pressure in speech. *Journal Acoustic Society of America* **35**, 454–460.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.
- Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.
- Laver, J., S. Hiller & BJ. Mackenzie (1992). Acoustic waveform perturbations and voice disorders. *Journal of Voice* **6**, 115 – 126.
- Le Huche, F. & A. Allali (1991). *La voix : Anatomie et physiologie des organes de la voix et de la parole (Tome 1)*. Paris : Masson.
- Le Huche, F. & A. Allali (2001). *La voix : Pathologie vocale d'origine organique (Tome 3)*. Paris : Masson.

- Lebrun, Y. (1966). Sur l'activité du diaphragme au cours de la phonation. *La Linguistique* **2**, 71–78.
- Ledda, G., N. Grover, V. Pundir, E. Masala & R. Puxeddu (2006). Functional outcomes after CO<sub>2</sub> laser treatment of early glottic carcinoma. *Laryngoscope* **116**, 1007–1011.
- Lefebvre, JL. & D. Chevalier (2005). Cancers du larynx. *Oto-rhino-laryngologie* **20-710-A-10**.
- Lehiste, I. & GE. Peterson (1959). Vowel amplitude and phonemic stress in American English. *Journal Acoustic Society of America* **31**, 428–435.
- Leothaud, G. (2004). Théorie de la phonation. Ms. Université Paris 4.
- Leplège, A., A. Ecosse, A. Verdier & T. Perneger (1998). The french SF36 health survey : translation, cultural adaptation and preliminary psychometric evaluation. *Journal Clin Epidemiol* **11**, 1013–1023.
- Lim, M., E. Lin & P. Bones (2006). Vowel effect on glottal parameters and the magnitude of jaw opening. *Journal of Voice* **20**, 46–54.
- Lindestad, PA., V. Blixt, J. Pahlberg-Olsen & B. Hammarberg (2004). Ventricular fold vibration in voice production : a high-speed imaging study with kymographic, acoustic and perceptual analyses of a voice patient and a vocally healthy subject. *Logoped. Phoniatr. Vocol* **29**, 162–170.
- Lindestad, PA., M. Sodersten, B. Merker & S. Granqvist (2001). Voice source characteristics in mongolian "throat singing" studied with high-speed imaging technique, acoustic spectra and inverse filtering. *Journal of Voice* **15**, 78–85.
- Lindqvist, J. (1972). Laryngeal articulation studied on Swidish subjects. In *Quarterly Progress and Status Report, Speech Transmission Laboratory*, 2/3, 10–27. Stockholm : Royal Institute of Technology.
- Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* **33**, 42–49.
- Lisker, L. (1970). Supraglottal air pressure in the production of English stops. *Language and Speech* **13**, 215–230.
- Lisker, L. (1972). Stop duration and voicing in English. In A. Valdman (ed.), *Papers on linguistics and phonetics in memory of Pierre Delattre*, 339–343. La Haye : Mouton.
- Lisker, L. (1986). "Voicing" in English : A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* **29**, 3–11.
- Lisker, L. & AS. Abramson (1964). A cross-language study of voicing in initial stops : acoustical measurements. *Word* **20**, 384–422.

- Lisker, L. & AS. Abramson (1971). Distinctive features and laryngeal control. *Language* **47**, 776–783.
- Lofqvist, A., T. Baer, NS. McGarr & RS. Story (1989). The cricothyroid muscle in voicing control. *Journal Acoustic Society of America* **85**, 1314–1321.
- Lofqvist, A., L. Koenig & R. McGowan (1995). Vocal tract aerodynamics in /aCa/ utterances : Measurements. *Speech Communication* **16**, 49–66.
- Lofqvist, A. & RS. McGowan (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics* **20**, 93–110.
- Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des maladies de l'oreille et du larynx* **37**, 101–119.
- Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization*. Ph.D. thesis, University of Massachusetts, Amherst.
- Lombardi, L. (1999). Positional faithfulness and voicing assimilation in optimality theory. *NLLT* **17**, 267–302.
- Lopez Llames, A., F. Nunez Batalla, JL. Llorente Pendas, M. Puente Verez, P. Al-dama Barahona & C. Suarez Nieto (2004). Laser cordectomy : oncologic and functional results. *Acta Otorrinolaryngol Esp* **55**, 34–40.
- Luce, PA. & J. Charles-Luce (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal Acoustic Society of America* **78**, 1949–1957.
- Luceno, J. & L. Koenig (2005). Phonation threshold as a function of laryngeal size in a two-mass model of the vocal folds. *Journal Acoustic Society of America* **118**, 2798–2801.
- Luceno, JC. (1996). Relation between the phonation threshold pressure and the pre-phonatory glottal width in a rectangular glottis. *Journal Acoustic Society of America* **100**, 2551–2554.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge : Cambridge University Press.
- Maddieson, I. (1997). Phonetic universals. In J. Laver & Hardcastle W. (eds.), *Handbook of phonetic sciences*, 619–639. Oxford : Blackwell Publishers.
- Majer, H. & W. Reider (1959). Technique de laryngectomie permettant de conserver la perméabilité respiratoire : la cricohyoido-pexie. *Ann Otolaryngol Chir Cervicofac* **76**, 677–683.
- Makeieff, M., E. Barbotte, A. Giovanni & S. Guerrier (2005). Acoustic and aerodynamic measurements of speech production after supracricoid partial laryngectomy. *Laryngoscope* **115**, 545–551.

- Makeieff, M., A. Giovanni & B. Guerrier (2007). Laryngostroboscopic evaluation after SCPL. *Journal of Voice* **21**, 507–515.
- Malécot, A. (1968). The force of articulation of American stops and fricatives as a function of position. *Phonetica* **18**, 95–102.
- Malmberg, B. (1954). *La phonétique*. Presses Universitaires de France.
- Martin, P. (2008). *Phonétique acoustique : introduction à l'analyse acoustique*. Armand Colin.
- Martinet, A. (1968). *Eléments de linguistique générale*. Colin.
- Mcfarland, D.H. (2006). *L'anatomie en orthophonie : Parole, voix et déglutition*. Paris : Masson.
- McGuirt, WF., D. Blalock, JA. Koufman, RS. Feehs, AJ. Hilliard, Greven. K. & M. Randall (1994). Comparative voice results after laser resection or irradiation of T1 vocal cord carcinoma. *Arch Otolaryngol Head Neck Surg* **120**, 951–955.
- Meynadier, Y ; & Y. Gaydina (2012). Contraste de voisement en parole chuchotée. In *Actes de JEP-TALN-RECITAL*, volume 1, 361–368. Grenoble.
- Miller, C.J. & R. Daniloff (1977). Aerodynamics of stops in continuous speech. *Journal of Phonetics* **5**, 351–360.
- Mirghani, H. (2009). *Evolution de la voix après cordectomie de type I par voie endoscopique au laser*. Master's thesis, Université Paris 3.
- Mobius, B. (2003). Gestalt psychology meets phonetics : An early experimental study of intrinsic F0 and intensity. In *Proceedings of 15th ICPHS*, 2677–2680. Barcelone : Espagne.
- Mohr, B. (1968). Intrinsic fundamental frequency variation. In *Monthly Internal Memorandum*. University of California, Berkeley.
- Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica* **23**, 65–93.
- Morsomme, D. & F. Estienne (2006). *Les bilans de langage et de voix : fondements théoriques et pratiques*. Paris : Masson.
- Motta, S., U. Cesari, M. Mesolella & G. Motta (2008). Functional vocal results after CO<sub>2</sub> laser endoscopic surgery for glottic tumours. *The journal of laryngology and Otology* **122**, 948–951.
- Muller, EM. & WS. Brown (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. In N. Lass (ed.), *Speech and language : Advances in basic research and practice*, volume 4. New-York : Academic Press.

- Murry, T. & WS. Brown (1976). Intraoral air pressure in whispered stop consonants. *Journal of Phonetics* **4**, 183–187.
- Murry, T., R. Medrado, ND. Hogikyan & JE. Aviv (2004). The relationship between ratings of voice quality and quality of life measures. *Journal of voice* **18**, 183–192.
- Murry, T. & CA. Rosen (2000). Outcome measurements and quality of life in voice disorders. *Otolaryngol Clin North Am* **33**, 905–916.
- Negus, V.E. (1949). *The comparative anatomy and physiology of the larynx*. London : W.Heinemann Medical Books.
- Netsell, R. (1969). Subglottal and intraoral pressures during the intervocalic contrast of /t/ and /d/. *Phonetica* **20**, 68–73.
- Netsell, R., W. Lotz & Shaughnessy AL. (1984). Laryngeal aerodynamics associated with selected voice disorders. *American Journal of Otolaryngology* **5**, 397–403.
- Netsell, R., WK. Lotz, AS. Duchane & SM. Barlow (1991). Vocal tract aerodynamics during syllable productions : normative data and theoretical implications. *Journal of Voice* **5**, 1–9.
- Neubauer, J., P. Mergell, U. Eysholdt & H. Herzl (2001). Spatio-temporal analysis of irregular vocal fold oscillations : Biphonation due to desynchronization of spatial modes. *Journal Acoustic Society of America* **110**, 3179–3192.
- Nguyen, H., H. Person & B. Vallée (1999). *Nouveaux dossiers d'anatomie P.C.E.M : tête (tome 1)*. Heures de France.
- Ni Chasaide, A. & C. Gobl (1999). Voice source variation in the vowel as a function of consonantal context. In Hardcastle W. & N. Hewlett (eds.), *Coarticulation*, 122–143. Cambridge : Cambridge University Press.
- Nishinuma, Y. (1977). *Contribution à l'étude prosodique du japonais*. Ph.D. thesis, Université d'Aix-en-Provence.
- Norton, NS. (2009). *Précis d'anatomie clinique de la tête et du cou*. Elvèsier Masson.
- Nunez Batalla, F., MJ. Caminero Cueva, B. Senaris Gonzales, J. Llorente Pendas, C. Gorri Gil, A. Lopez Llames, R. Alonso Pantiga & C. Suarez Nieto (2008). Voice quality after endoscopic laser surgery and radiotherapy for early glottic cancer : objective measurements emphasizing the Voice Handicap Index. *Eur Arch Otorhinolaryngol* **265**, 543–548.
- Nunez Batalla, F, C. Suarez Nieto, C. Munoz Pinto, L. Baragano Rio, MJ. Alvarez Zapico & A. Martinez Ferreras (2000). Spectrographic study of voice disorders : subharmonics. *Acta Otorrinolaringologia Esp* **51**, 52–56.

- O' Dell, ML. & RF. Port (1983). Discrimination of word final voicing in German. *Journal Acoustic Society of America* **73**, Suppl.1.
- O' Shaughnessy, D. (1981). A study of French vowel and consonant durations. *Journal of Phonetics* **9**, 385–406.
- Oates, J. (2009). Auditory perceptual evaluation of disordered voice quality. *Folia Phoniatrica and Logopedia* **61**, 49–56.
- Ohala, J. (1972). How is pitch lowered ? *Journal Acoustic Society of America* **52**, 124.
- Ohala, J. (1973). Explanations for the intrinsic pitch of vowels. Monthly Internal Memorandum (Berkley).
- Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In PF. MacNeilage (ed.), *The production of speech*, 189–216. New-York : Springer Verlag.
- Ohala, J. & BW. Eukel (1987). Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockley (eds.), *In honor of Ilse Lehiste*, 207–215. Dordrecht : Foris.
- Ohala, J. & CJ. Riodan (1979). Passive vocal tract enlargement during voiced stop. In J. Wolf & DH. Klatt (eds.), *Speech communication papers*, 89–92. New-York : Acoustical Society of America.
- Ohde, RN. (1984). Fundamental frequency as an acoustic correlate of stop correlate of stop consonant voicing. *Journal Acoustic Society of America* **75**, 224–230.
- Ohman, SE. (1966). Coarticulation in VCV utterances : Spectrographic measurements. *Journal Acoustic Society of America* **39**, 151–168.
- Orlikoff, R. (1998). Scrambled EGG : The uses and abuses of electroglottography. *Phonoscope* **1**, 37–53.
- Orlikoff, RF. (1995). Vocal stability and vocal tract configuration : an acoustic and electroglottographic investigations. *Journal of Voice* **2**, 174–181.
- Ormezzano, Y. (2000). *Le guide de la voix*. Paris : Odile Jacob.
- Osgood, CE. (1957). *The measurement of meaning*. Urbana : University of Illinois Press.
- Ouellon, C., J. Dolbec & M. Ouellet (1993). La question de l'intensité en français québécois. In *Actes du 15th Congrès International des Linguistes*, volume 2, 75–78. Québec : Presses de l'Université de Laval.
- Parnell, M., JD. Amerman & GB. Wells (1977). Closure and constriction duration for alveolar consonants during voiced and whispered speech. *Journal Acoustic Society of America* **86**, 1678–1683.

- Pech, A., M. Cannoni, S. Abdul, M. Zanaret, JM. Thomassin & JL. Gourbet (1982). Laryngectomie reconstructive frontale anté. *Ann Otolaryngol Chir Cervicofac* **99**, 141–146.
- Pech, C., A. Giovanni & N. Henin (1988). La voix sans corde vocale. *Revue de Laryngologie, rhinologie, otologie* **109**, 373–377.
- Perello, J. (1962). Théorie muco-ondulatoire de la phonation. *Annales Oto-rhino-laryngologiques* **79**, 722–725.
- Peretti, G., C. Piazza, C. Balzanelli, MC. Mensi, M. Rossini & AR. Antonelli (2003). Preoperative and postoperative voice in TIs-T1 glottic cancer treated by endoscopic cordectomy : an additional issue for patient counseling. *Ann Otol Rhinol Laryngol* **112**, 759—763.
- Peterson, GE. & I. Lehiste (1961). Duration of syllable nuclei in English. *Journal Acoustic Society of America* **32**, 693–703.
- Pierce, M. (2005). Kera tone and voicing. In M. Pearce & N. Topintzi (eds.), *Working papers in Linguistics 17*. UCLA : Department of Phonetics and Linguistics.
- Pierrehumbert, J. & D. Talkin (1992). Lenition of /h/ and glottal stop. In G. Docherty & DR. Ladd (eds.), *Laboratory phonology II : Gesture, segment, prosody*, 90–117. Cambridge : Cambridge University Press.
- Piquet, JJ., A. Desaulty & G. Decroix (1974). Crico-Hyoido-épiglottopexie : Technique opératoire, et résultats fonctionnels. *Annales d'Otolaryngologie et de Chirurgie Cervicofaciale* **91**, 681–689.
- Policarpo, M., P. Aluffi, G. Brovelli & F. Borello (2004). Oncological and functional results of CO<sub>2</sub> laser cordectomy. *Acta Otolaryngologica* **24**, 267–274.
- Port (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics* **7**, 45–56.
- Port, RF. & M. O' Dell (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics* **13**, 455–471.
- Portmann, M. & Y. Guerrier (1987). *Traité de technique chirurgicale ORL et cervico-faciale (Tome 3)*. Paris : Masson.
- Preciado, JA., R. Garcia & JC. Infante (1998). Analisis de la funcion vocal. Estudio de casos y controles. *Acta Otorrinolaringología Esp* **49**, 467–474.
- Raphael, LJ. (1972). Preceding vowel duration as a cue to the perception of voicing of word-final consonants in American English. *Journal Acoustic Society of America* **51**, 1296–1303.

- Remacle, M., HE. Eckel & A. Antonelli (2000). Endoscopic cordectomy. A proposal for a classification by the working committee, European Laryngological society. *Eur Arch Otorhinolaryngol* **257**, 227–231.
- Repp, BH. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* **22**, 2173–2189.
- Ridouane, R. (2006). Voisement, aspiration et pré-aspiration : le comportement glottal. Une revue de la littérature.
- Rigault, A. (1967). L'assimilation consonantique de sonorité en français. In *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague.
- Roh, JL., DH. Kim, SY. Kim & CI. Park (2007). Quality of life and voice in patients after laser cordectomy for Tis and T1 glottic carcinomas. *Head Neck* **29**, 1010–1016.
- Rossi, M. (1971). L'intensité spécifique des voyelles. *Phonetica* **24**, 129–161.
- Rossi, M. & D. Autesserre (1981). Movements of the hyoid and the larynx and the intrinsic frequency of vowels. *Journal of Phonetics* **9**, 233–249.
- Rothenberg, M. (1968). Breath-stream dynamics of simple-released-plosive production. *Bibliotheca Phonetica* **6**, 6–22.
- Rothenberg, M. (1982). Interpolating subglottal pressure from oral pressure. *Journal of Speech and Hearing Research* **47**, 218–224.
- Rothenberg, M. & JJ. Mahshie (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal Speech Hearing Research* **31**, 338–351.
- Rottger, TB., B. Winter & S. Grawunder (2011). The robustness of incomplete neutralization. In *Proceedings of the 17th ICPHS*, 1722–1725. Hong Kong.
- Rush, J. (1859). *The phylosophy of the human voice*. Philadelphia : Lippincott.
- Rydell, R., L. Schalén, S. Fex & A. Elner (1995). Voice evaluation before and after laser excision vs. radiotherapy of T1A glottic carcinoma. *Acta Otolaryngol* **115**, 560–565.
- Sakakibara, KI., H. Imagawa, T. Konishi, K. Kondo, EZ. Murano, M. Kurnada, H. Imagawa & S. Niima (2001). Vocal fold and false vocal fold vibrations in throat singing and synthesis of Khöömei. In *Proceedings of the International Computer Music Conference*, 135–138. Havana : Cuba.
- Saussure, F. (1916). *Cours de linguistique générale*. Payot.
- Sawashima, M. & H. Hirose (1968). New laryngoscopic technique by use of fiber optics. *Journal Acoustic Society of America* **43**, 168—173.

- Scherer, RC., VJ. Vail & B. Rockwell (1995). Examination of the laryngeal adduction measure EGGW. In F. Bell-Berti & BN. Raphael (eds.), *Producing speech : Contemporary issues*, 269–289. Woodbury, NY : American Institute of physics.
- Schindler, A., F. Palonta, G. Preti, F. Ottaviani, O. Schindler & AL. Cavalot (2004). Voice quality after carbon dioxide laser and conventional surgery for T1a glottic carcinoma. *Journal of Voice* **18**, 545–550.
- Schoentgen, J., M. Bensaid, F. Bucella & S. Ciocea (1998). Issues in the acoustic evaluation of laryngeal disorders. In *Proceedings VOICEDATA98, Symposium on Databases in Voice Quality Research and Education*, 16–21. Université of Utrecht, Pays-Bas.
- Schoentgen, J. & D. Guchteneere (1991). An algorithm for the measurement of jitter. *Speech Communication* **10**, 533–538.
- Schoentgen, J. & De Guchteneere (1995). The series analysis of jitter. *Journal of Phonetics* **23**, 189–201.
- Schuh, RG. (1997). Changes in obstruent voicing in Bade/Ngizim. manuscript, UCLA.
- Schutte, H.K (1992). *The efficiency of voice production*. Singular Publishing Group.
- Schwartz, JL., LJ. Boë, N. Vallée & C. Abry (1997). The dispersion- focalisation theory of vowel systems. *Journal of Phonetics* **25**, 255–286.
- Schwartz, MF. (1972). Bilabial closure durations for /p/, /b/ and /m/ in voiced and whispered vowel environments. *Journal Acoustic Society of America* **51**, 2025–2029.
- Serniclaes, W. (1987). *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Ph.D. thesis, Université Libre de Bruxelles, Belgique.
- Serniclaes, W. & P. Bejster (1974). Influence du contexte vocalique sur la perception de voisement des occlusives. In *Actes des 5èmes Journées d'Etude sur la Parole*, volume 1, 10–18. Orsey.
- Shimamungu, E. (1998). *Le kinyarwanda : Initiation à une langue Bantu*. Paris : L'Harmattan.
- Shrager, M. (2006). Neutralization of word-final voicing in Russian. In *proceedings of the first conference of the Slavic linguistics society*. Indiana University, Bloomington.
- Simon, P. (1967). *Les consonnes françaises : mouvements et positions articulatoires à la lumière de la radiocinématographie*. Paris : C. Klincksieck.
- Sittel, C., H. Eckel & C. Eschenburg (1998). Phonatory results after laser surgery for glottic carcinoma. *Otolaryngol Head Neck Surg* **119**, 418–424.

- Sjogren, EV., MA. Van Rossum, TP. Langeveldt, MS. Voerman, VAH. Van de Kamp, MOW. Friebel, R. Wolterbeek & RJ. Baatenburg de Jong (2008). Voice outcome in T1a midcord glottic carcinoma : laser surgery vs radiotherapy. *Arch Otolaryngol Head Neck Surg* **134**, 965–972.
- Sjogren, EV., MA. Van Rossum, TP. Langeveldt, MS. Voerman, VAH. Van de Kamp & RJ. Baatenburg de Jong (2009). Voice profile after type I or II laser cordectomy for T1a glottic carcinoma. *Head Neck* **31**, 1502–1510.
- Smith, BL., R. Hayes-Harb, M. Bruss & A. Harker (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics* **37**, 257–275.
- Smitheran, JR. & T. Hixon (1981). A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech and Hearing Research* **46**, 138–146.
- Stathopoulos, E. (1984). Oral airflow during vowel production of children and adults. *Cleft Palate Journal* **21**, 277–285.
- Stathopoulos, E., JD. Hoit, T. Hixon, PJ. Watson & N. Solomon (1991). Respiratory and laryngeal function during whispering. *Journal of Speech and Hearing Research* **34**, 761–767.
- Stathopoulos, E. & G. Weismer (1985). Oral airflow and air pressure during speech production : A comparative study of children, youths and adults. *Folia Phoniatrica* **37**, 152–159.
- Stathopoulos, E.T. (1986). Relationship between intraoral air pressure and vocal intensity in children and adults. *Journal of Speech and Hearing Research* **29**, 71–74.
- Stevens, KN. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica* **34**, 264–279.
- Stevens, KN. (1989). On the quantal nature of speech. *Journal of Phonetics* **17**, 3–45.
- Stevens, KN. (1991). Vocal fold vibration for obstruent consonants. In J. Gauffin & B. Hammarberg (eds.), *Vocal fold physiology : acoustic, perceptual and physiological aspects of voice mechanisms*, 29–36. CA : Singular Publishing Group.
- Stevens, KN. (1998). *Acoustic phonetics*. Cambridge : MA : MIT Press.
- Stevens, KN. & DH. Klatt (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal Acoustic Society of America* **55**, 653–659.
- Stoicheff, ML., A. Ciampi & JE. Passi (1983). The irradiated larynx and voice : a perceptual study. *Journal of Speech and Hearing Research* **26**, 482–485.

- Story, H. (2002). An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology* **23**, 195–206.
- Story, H. & I.R. Titze (1995). Voice simulation with a body-cover model of the vocal folds. *Journal of Acoustic Society of America* **97**, 1249–1260.
- Strong, M.S. (1975). Laser excision of carcinoma of the larynx. *Laryngoscope* **85**, 1286–1289.
- Subtelny, J.D., J.H Worth & M. Sakuda (1966). Intraoral pressure and rate of flow during speech. *Journal of Speech and Hearing Research* **9**, 498–518.
- Sulter, A.M. & H.P. Wit (1996). Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age. *Journal Acoustic Society of America* **100**, 3360–3373.
- Summerfield, Q. & M. Haggard (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal Acoustic Society of America* **62**, 435–448.
- Sundberg, J. (1995). Vocal fold vibration patterns and modes of phonation. *Folia Phoniatrica* **47**, 218–228.
- Sundberg, J., I. Titze & R. Scherer (1993). Phonatory control in male singing : a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *Journal of Voice* **7**, 15–29.
- Svec, JG., HK. Schutte & DG. Miller (1996). A subharmonic vibratory pattern in normal vocal folds. *Journal of Speech and Hearing Research* **39**, 135–143.
- Sylvestre, J.L. & P. MacLeod (1968). Le muscle vocal humain est-il asynchrone ? *Journal Physiol* **5**, 373–389.
- Tamura, E., S. Kitahara, M. Oqura & N. Kohno (2003). Voice quality after laser surgery or radiotherapy for T1a glottic carcinoma. *Laryngoscope* **113**, 910–914.
- Taylor, HC. (1933). The fundamental pitch of English vowels. *Journal of Experimental Psychology* **16**, 565–582.
- Teffahi, H. & F. Kherouf (2010). Simulation de la source vocale : effet du couplage avec le conduit vocal. In *Proceedings of 8th Conférence Internationale de Modélisation et Simulation - MOSIM'10*. Hammamet : Tunisie.
- Teston, B. (2004). L'évolution instrumentale des dysphonies. Etat actuel et perspectives. In A. Giovanni (ed.), *Le bilan d'une dysphonie*, 105—169. Marseille : Solal.
- Teston, B., B. Galindo & A. Ghio (1999). A multisensor data acquisition and processing system for speech production investigation. In *Proceedings of 14th ICPHS*, volume 3, 2251–2254. San Francisco.

- Titze, I. (1990). Interpretation of the electroglottographic signal. *Journal of Voice* **4**, 1–9.
- Titze, I. (1992). Vocal intensity in speakers and singers. *Journal Acoustic Society of America* **91**, 2936–2946.
- Titze, IR. (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal Acoustic Society of America* **83**, 1536–1552.
- Titze, I.R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal Acoustic Society of America* **85**, 901–906.
- Titze, I.R. (1994). *Principles of voice production*. NJ, USA : Prentice-Hall.
- Titze, IR & WS. Winholtz (1993). Effect of microphone type and placement on voice perturbation measurements. *J Speech Hear Res* **36**, 1177–1190.
- Trullinger, RW. & FW. Emanuel (1983). Airflow characteristics of stop-plosive productions of normal-speaking children. *Journal of Speech and Hearing Research* **26**, 202–208.
- Tucker, HM., BG. Wood, H. Levine & R. Katz (1979). Glottic reconstruction after near-total laryngectomy. *Laryngoscope* **89**, 609–618.
- Vaissière, J. (1997). Phonological use of the larynx. In *Proceedings of Larynx-1997*, 115–126. Marseille : France.
- Vaissière, J. (2001). Changements de sons et changements prosodiques : du latin au français. *Parole* **15/16**, 53–88.
- Vaissière, J. (2006). *La phonétique*. Presses Universitaires de France.
- Van Alphen, PM. (2007). Prevoicing in Dutch initial plosives. Production, perception and word recognition. In J. Van de Weijer & EJ. Van der Torre (eds.), *Voicing in Dutch. (De)voicing- phonology, phonetics and psycholinguistics*, 99–124. Amsterdam/Philadelphie : Benjamins.
- Van Den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* **1**, 227–244.
- Van Den Berg, J. (1968). Mechanism of the larynx and the laryngeal vibrations. In B. Malmberg (ed.), *Manual of phonetics*, 278–308. Amsterdam : North Holland Publishing.
- Van Dommelen, WA. (1983). Parameter interaction in the perception of French plosives. *Phonetica* **40**, 32–62.
- Van Dommelen, WA. (1985). Further experiments in the perception of French plosives. *Phonetica* **42**, 100–115.

- Vaughn, A. (1965). *An experimental study of oral and nasal airflow during sustained vowel production*. Ph.D. thesis, University of Oklahoma.
- Vercherand, G. (2010). *Production et perception de la parole chuchotée en français : analyse segmentale et prosodique*. Ph.D. thesis, Université Paris VII.
- Verdolini, K., I. Titze & A. Fennell (1994). Dependance of phonatory effort on hydration level. *Journal of Speech and Hearing Research* **37**, 1001–1007.
- Vilaseca, I., P. Huerta, JL. Blanch, AM. Fernandez Planas, C. Jimenez & M. Bernal-Sprekelsen (2008). Voice quality after CO<sub>2</sub> laser cordectomy—What can we really expect? *Head Neck* **30**, 43–49.
- Voiers, W. (1964). Perceptual bases of speaker identity. *Journal Acoustic Society of America* **36**, 1065–1073.
- Von Frisch, K. (1965). *Tanzsprache und Orientierung der Bienen*. Springer Verlag.
- Von Frisch, K. (1967). *The dance language and orientation of bees*. Harvard University Press.
- Wajskop, M. (1979). Segmental durations of French intervocalic plosives. In B. Lindblom & S. Ohman (eds.), *Frontiers of speech communication research*, 109–123. New-York : Academic Press.
- Wallet, L. (2007). *La prosodie des modalités assertives et interrogatives en français après laryngectomies partielles*. Master's thesis, Université Paris 3.
- Wallet, L. (2008). *Etude perceptive des modalités interrogatives et assertives après laryngectomies partielles verticales*. Master's thesis, Faculté de Médecine d'Aix-Marseille 2.
- Wallet, L., L. Crevier Buchman, J. Vaissière & S. Hans (2009). A perceptual study of question and assertive modalities after vertical partial laryngectomies. *Revue de Laryngologie, rhinologie, otologie* **130**, 65–68.
- Ware, JE. & CD. Sherbourne (1992). The MOS 36-item short-form health survey (SF-36). *Medical Care* **30**, 473–483.
- Warner, N., A. Jongman, J. Sereno & R. Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception : Evidence from Dutch. *Journal of Phonetics* **32**, 251–276.
- Warren, DW. (1996). Regulation of speech aerodynamics. In NJ. Lass (ed.), *Principles of Experimental Phonetics*, 46–92. St-Louis : Mosby.
- Westbury, J. & P. Keating (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics* **22**, 145–166.

- Westbury, JR. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal Acoustic Society of America* **73**, 1322–1336.
- Wetzels, WL. & J. Mascaro (2001). The typology of voicing and devoicing. *Language* **77**, 207–244.
- Whalen, D., B. Gick & M. Kumada (1998). Cricothyroid activity in high and low vowels : exploring the automaticity of intrinsic F0. *Journal of Phonetics* **27**, 125–142.
- Whalen, DH. & AG. Levitt (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics* **23**, 349–366.
- Whalen, DH., AG. Levitt, PL. Hsiao & I. Smorodinsky (1995). Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French and English learning infants. *Journal Acoustic Society of America* **97**, 2533–2539.
- Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics* **5**, 169–184.
- Woisard-Bassols, V. (2000). Bilan clinique de la voix. In *Encyclopédie Médicale Chirurgie*. Editions Scientifiques et Médicales Elsevier SAS.
- Woodson, GE. & M. Cannito (1998). Voice analysis. In CW. Cummings (ed.), *Otolaryngology-Head and Neck Surgery*, 1876—1890. St. Louis : Mosby-Year Book.
- Wuyts, FL., MS. De Bodt, G. Molenberghs, M. Remacle, L. Heylen & B. et al. Miller (2000). The dysphonia severity index : an objective measure of vocal quality based on a multiparameter approach. *Journal of Speech, Language and Hearing Research* **43**, 796–809.
- Yanagihara, N. & Y. Koike (1967). The regulation of sustained phonation. *Folia Phoniatrica* **19**, 1–18.
- Yanagihara, N., Y. Koike & H. Von Leden (1966). Phonation and respiration. *Folia Phoniatrica* **18**, 323–340.
- Yanagihara, N. & H. Von Leden (1967). Respiration and phonation. *Folia Phoniatrica* **19**, 153–166.
- Yumoto, E, W.J. Gould & T. Baer (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am* **71**, 1544–1549.
- Zeitels, SM. (1995). Premalignant epithelium and microinvasive cancer of the vocal fold : the evolution of phonosurgical management. *Laryngoscope* **105**, 1–51.
- Zeitels, SM., RE. Hillman, RA. Franco & GW. Bunting (2002). Voice and treatment outcome from phonosurgical management of early glottic cancer. *Ann Otol Rhinol Laryngol* **190**, 3–20.

Zellner, B. (1998). *Caractérisation et prédition du débit de parole en français : une étude de cas*. Ph.D. thesis, Université de Lausanne.

Zué, V. (1976). Acoustic characteristics of stop consonants : A controlled study. Technical Report n°523, Lincoln Laboratory, Lexington, MA.

## **Etudes multiparamétriques de la voix et de la parole après cordectomie laser par voie endoscopique de type II-III**

Cette thèse est dédiée à l'étude des caractéristiques phonétiques de la voix et de la parole après cordectomie laser de type II-III, proposées pour des tumeurs laryngées limitées à un seul pli vocal. Les cordectomies de type II-III consistent en la résection de l'épithélium, du ligament vocal, de la lamina propria et d'une partie variable du muscle vocal. La nouvelle configuration anatomico-physiologique engendre un accolement des plis vocaux moins efficace et une asymétrie vibratoire, impliquant des perturbations en fréquence et en amplitude du fondamental ainsi qu'une possible altération du trait de voisement. Il s'agit d'une étude prospective de 10 patients masculins et francophones âgés de 47 à 80 ans. Les sujets ont été enregistrés entre 6 mois et 1 an postopératoire. Une population saine de 10 témoins masculins, âgés de 34 à 67 ans nous a servi de référence. Le premier chapitre est un rappel anatomique et biomécanique de la phonation. Il permet de mieux comprendre et appréhender les mécanismes de production de la voix et de la parole. Le second chapitre présente succinctement les chirurgies partielles du larynx. Il nous permet de déterminer, en particulier, les conséquences d'apparition et de pratique des cordectomies laser. Le troisième chapitre est une description des différentes cordectomies laser. Nous présentons également une revue de la littérature qui met en évidence les résultats fonctionnels après ces chirurgies, mais surtout qui insiste sur la dimension multiparamétrique des approches scientifiques. Le quatrième chapitre est une description détaillée de nos populations étudiées, des corpus et des matériaux utilisés pour nos différentes études. Le cinquième chapitre comprend quatre études qui montrent l'impact de la chirurgie sur les paramètres de la  $F_0$ , de l'intensité sonore et de leurs indices de perturbation (jitter et shimmer). Les plages de variation de  $F_0$  des sujets y sont analysées. Enfin, le dernier chapitre est consacré à l'implémentation du trait de voisement. Celui-ci se compose de quatre études (acoustique, aérodynamique, EGG et perceptive) qui nous permettent d'étudier la réalisation du contraste de voisement des occlusives après ces cordectomies.

**Mots-clés : phonétique clinique, cordectomie laser, acoustique, aérodynamique, EGG, perception**

---

## **Multiparametric studies of voice and speech after laser endoscopic cordectomy of types II-III**

This thesis concerns the study of the phonetic characteristics of the voice and speech after laser cordectomy of types II-III, proposed for laryngeal tumors limited to the vocal fold. Cordectomies of types II-III consist of resection of the epithelium, vocal ligament, lamina propria and a variable part of the vocal muscle. This new anatomico-physiological configuration creates an apposition of the vocal folds less effective and a vibratory asymmetry, which involve perturbations in frequency and amplitude and a possible alteration of the voicing. This is a prospective study of 10 male French patients aged 47-80 years. The subjects were recorded between 6 months and 1 year postoperatively. A healthy population of 10 males, aged 34-67 years was the control group. The first chapter is a anatomical and biomechanical reminder of the phonation. It provides a better understand of the mechanisms of voice production and speech. The second chapter is a brief presentation of partial laryngeal surgeries. It allows us to determine, in particular, the emergence of laser cordectomies. The third chapter is a description of the different laser cordectomies. We also present a review of the literature which shows the functional results after these surgeries, but also emphasizes on the multiparametric dimension of the scientific approaches. The fourth chapter is a detailed description of the populations studied, the corpus and the materials used for our various studies. The fifth chapter includes four studies which show the impact of surgery on the parameters of the  $F_0$ , loudness and their perturbation factors (jitter and shimmer). The ranges of variation of  $F_0$  of the subjects will be analyzed. The last chapter is devoted to the implementation of the voicing feature. It consists of four studies (acoustic, aerodynamic, EGG and perceptual) which will allow us to observe the preservation of voicing contrast of plosives after these cordectomies.

**Keywords : clinical phonetics, laser cordectomy, acoustic, aerodynamic, EGG, perception**

Discipline : phonétique

Institut de Linguistique et Phonétique Générales et Appliquées, 19, rue des Bernardins, 75005 Paris