HAL
archives-ouvertes.fr

Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception

Nathalie Henrich

▶ **To cite this version:**

Nathalie Henrich. Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception. Acoustique [physics.class-ph]. Université Pierre et Marie Curie - Paris VI, 2001. Français. <tel-00123133>

HAL Id: tel-00123133

https://tel.archives-ouvertes.fr/tel-00123133

Submitted on 8 Jan 2007

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

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

THÈSE de DOCTORAT de l'UNIVERSITÉ PARIS 6

Spécialité :
Acoustique

Sujet de la thèse :

Etude de la source glottique en voix parlée et chantée :
modélisation et estimation, mesures acoustiques et électroglottographiques, perception

Présentée par : **Nathalie Henrich**

pour obtenir le titre de Docteur de l'Université Paris 6

Soutenue le 30 Novembre 2001 devant le jury composé de :

Madame	Michèle Castellengo	Directrice de thèse (LAM)
Monsieur	Christophe d'Alessandro	Directeur de thèse (LIMSI)
Monsieur	Didier Demolin	Rapporteur
Monsieur	Thierry Dutoit	Rapporteur
Monsieur	Johan Sundberg	Examinateur
Monsieur	Xavier Rodet	Examinateur
Monsieur	Bernard Roubeau	Examinateur

Remerciements

C'est une chance inouïe de pouvoir se consacrer à ses passions et cette chance m'a été offerte par Michèle Castellengo et Christophe d'Alessandro. Ils ont ouvert pour moi les portes d'un monde dont je rêvais mais que je pensais alors inacessible : le monde de la voix. Ils ont été mes initiateurs et mes guides et je les en remercie du fond du coeur. J'associe à ces remerciements un autre guide, Boris Doval, qui n'a pas ménagé ni son temps ni son énergie pour m'aider. J'ai également une pensée affectueuse pour Joe Wolfe, qui m'a fait découvrir l'Acoustique Musicale.

Afin que cette étude sur la voix ne reste pas sans voix, de nombreux chanteurs ont accepté de franchir la porte de la cabine d'enregistrement pour offrir leur talent avec tant de gentillesse et de spontanéité : je tiens donc à remercier Marie-Agnès Faure, Bernard Roubeau, Anne-Julia Audray, Daniel Blanchard, Bertrand Chuberre, Agnès-Laure Bigorie, Jean-Michel Ankaoua, Robert Expert, Renaud Derrier, Magali Léger, Caroline Marçot, Renaud Tripathi, Nicolas Kern, Philippe Jarousski, Blandine Foliot, Elise Dabrowsky, Olivier Willemin, Alain Zürcher, Emmanuel Bouquey, Coriandre Vilain et Tran Quang Hai. Merci également à tous ceux qui ont accepté de passer un peu de leur temps dans cette même cabine, non pas pour chanter mais pour écouter. Cela n'avait rien d'une partie de plaisir et j'en suis d'autant plus reconnaissante à ceux qui ont renouvelé l'expérience plusieurs fois.

Derrière ce travail, il y a une multitude de personnes qui m'ont offert un peu de leur temps et de leur savoir : merci à tous, en particulier Gwénaël Robin, Julien Boissinot et Laurent Ghys, des informaticiens de génie dont les conseils et les coups de main m'ont été précieux, Vincent Rioux qui m'a permis d'utiliser le logiciel de test perceptif qu'il avait développé et m'a aidé, par ses conseils, à y insérer le module dont j'avais besoin, Vu Ngoc Tuan pour les nombreuses discussions et pour les programmes en C, Daniel Ambroise pour ses conseils en statistique, Juergen Neubauer pour les précieuses données qu'il m'a fait parvenir, Christel Beaujard Amato, Yannick de Kercadio et Bertrand David dont les conseils m'ont aidée à avancer. Un grand merci à mon petit lutin préféré, Gunilla Sundin, dont la présence a embelli mes derniers mois de thèse.

L'étude de la voix demande une approche pluridisciplinaire. J'ai ainsi pu bénéficier des expertises complémentaires de deux laboratoires : le Laboratoire d'Acoustique Musicale (LAM) et le Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI). L'étude de la perception de la voix chantée et des mécanismes laryngés est une des compétences du Laboratoire d'Acoustique Musicale, qui, de plus, a développé depuis de nombreuses années des contacts avec les chanteurs et les professionnels de la voix. L'équipe « Analyse et synthèse audio » au LIMSI étudie depuis quelques années les propriétés temporelles et spectrales de la source glottique à partir de modèles de signaux. Je voudrais donc remercier ces deux laboratoires de m'avoir accueillie et d'avoir mis à ma disposition tous les moyens nécessaires au bon déroulement de ces recherches. J'aimerais exprimer mon amitié à toutes ces personnes que j'y ai rencontrées et avec qui j'ai partagé des moments de joies pendant ces trois années. La liste serait longue et j'espère donc qu'ils se reconnaîtront. J'ai été touchée que les membres qui composent le jury de thèse acceptent spontanément d'y participer, montrant ainsi leur intérêt pour ce travail et je les en remercie chaleureusement.

Je voudrais également exprimer toute ma gratitude à ma famille proche et à mes amis pour leur patience et leur compréhension. Je voudrais remercier du fond du coeur mon ami Paul Bernardoni, dont

le soutien a été quotidien, allant du ménage à la relecture de la thèse avec toujours le soleil au coeur, le sourire dans les yeux et une patience infinie face à mes humeurs. Merci à toi pour tout ! Je dédie cette thèse à sa mère Annie qui nous avait appris le sens des mots *joie de vivre* et *courage* avant de s'éteindre, trop jeune. Je la dédie également à mes parents, qui m'ont toujours soutenue de leur amour et guidée sur le chemin, parfois difficile, de la vie.

à *Annie*,

Son regard est pareil au regard des statues,
Et pour sa voix, lointaine et calme, et grave elle a
L'inflexion des voix chères qui se sont tues.
(Paul Verlaine, « Poèmes saturniens », Melancholia VI, Mon rêve familier)

à *mes parents*,

On ne voit bien qu'avec le coeur,
L'essentiel est invisible pour les yeux ...
(Antoine de Saint-Exupery, « Le Petit Prince »)

Sommaire

INTRODUCTION **1**

Partie préliminaire : Caractérisation de la source glottique **7**

1 Description physiologique **11**

1.1 Description statique et dynamique du larynx 11

1.2 Les mécanismes laryngés 14

1.3 Méthodes exploratoires 20

1.4 Conclusion 21

2 Modélisation et paramètres de source glottique **23**

2.1 Le modèle de signal source-filtre 23

2.2 Les paramètres de source glottique 26

2.3 Conclusion 32

I Première partie : Analyse temporelle et spectrale de la source glottique **33**

3 Modélisation de la source glottique **37**

3.1 Unification des modèles temporels de source glottique et formalisme 37

3.2 Effet spectral des paramètres de source glottique 42

3.3 Conclusion 50

4 Estimation spectrale des paramètres de source glottique **55**

4.1 Etude théorique de la différence spectrale $(H_1 - H_2)$ 55

4.2 Estimation de la source glottique 63

4.3 Mesures expérimentales de $(H_1 - H_2)$ 72

4.4 Conclusion 80

II Deuxième partie : Mesures acoustiques et électroglottographiques 83

5 Principe et application de l'électroglottographie 87

5.1 Description du dispositif . . . 87
5.2 Caractérisation des signaux électroglottographiques . . . 88
5.3 Apport des signaux électroglottographiques . . . 94
5.4 Conclusion . . . 95

6 Constitution d'une base de données 97

6.1 Le dispositif expérimental . . . 97
6.2 Les chanteurs . . . 99
6.3 Le protocole d'enregistrement . . . 101
6.4 Conclusion . . . 103

7 Etude des signaux DEGG 105

7.1 Des signaux électroglottographiques à leurs dérivées . . . 105
7.2 Analyse des signaux DEGG de la base de données . . . 107
7.3 Mesure de la fréquence fondamentale et du quotient ouvert . . . 119
7.4 Conclusion . . . 127

8 Mesures du quotient ouvert en voix parlée et chantée 129

8.1 Revue bibliographique . . . 129
8.2 Effet des mécanismes laryngés . . . 132
8.3 Fréquence fondamentale et intensité vocale . . . 138
8.4 Quotient ouvert et intensité vocale . . . 139
8.5 Quotient ouvert et fréquence fondamentale . . . 149
8.6 Effet de la voyelle . . . 151
8.7 Comparaison parlé - chanté . . . 155
8.8 Conclusion . . . 160

III Troisième partie : Etude perceptive des paramètres de source glottique 165

9 Seuils de perception des paramètres de source glottique 169

9.1 Problématique . . . 169
9.2 Méthode adaptative à choix forcé . . . 171
9.3 Mise en place du test . . . 176
9.4 Résultats . . . 183
9.5 Conclusion . . . 193

Exemples sonores 196

CONCLUSION GENERALE DE L'ETUDE 201

ANNEXES 207

A Expressions temporelles et spectrales des modèles d'ODG 209

B Analyse par prédiction linéaire 213

C Protocole d'enregistrement 221

D Un peu de statistique ... 225

E Variations de l'intensité vocale en fonction du quotient ouvert 229

F Visualisation des glissandos 237

G Etude des premières mesures de l' Ave Maria de Gounod 253

BIBLIOGRAPHIE 273

TABLE DES MATIERES 283

Abréviations

Signaux étudiés :	**EGG**	signal électroglottographique (Ω)
	DEGG	dérivée du signal électroglottographique ($\Omega.s^{-1}$)
	ODG	onde de débit glottique ($\mathrm{l.s}^{-1}$)
	ODGD	dérivée de l'onde de débit glottique ($\mathrm{l.s}^{-2}$)

Mécanismes laryngés :	**m0**	mécanisme 0
	mI	mécanisme I
	mII	mécanisme II
	mIII	mécanisme III

Paramètre acoustique :	**I**	intensité vocale (dB)
Paramètres de source glottique :	$\mathbf{f_0}$	fréquence fondamentale (Hz)
	$\mathbf{O_q}$	quotient ouvert
	$\mathbf{C_q}$	quotient fermé
	$\mathbf{S_q}$	quotient de vitesse ("speed quotient")
	$\alpha_\mathbf{m}$	coefficient d'asymétrie
	$\mathbf{A_v}$	amplitude de voisement ($\mathrm{l.s}^{-1}$)
	E	vitesse de fermeture ($\mathrm{l.s}^{-2}$)
	$\mathbf{Q_a}$	quotient de phase de retour

Description spectrale :	$\mathbf{F_i}$	i^{eme} formant
	$\mathbf{H_i}$	i^{eme} harmonique : $H_i = i * f_0$ (Hz)

Ecriture phonétique des voyelles :	[**a**]	comme dans *âme*
	[**e**]	comme dans *année*
	[**u**]	comme dans *jour*
	[**i**]	comme dans *vie*

Notation musicale : la3 (notation française) = A4 (notation anglo-saxonne) correspond à $f_0 = 440$ Hz

Pour noter les chiffres, nous utiliserons la notation décimale anglo-saxonne.

Introduction

Notre société se tourne vers l'information, et la communication parlée en est un élément essentiel. Depuis 40 ans, le domaine du traitement de la parole est en pleine expansion : nous pouvons transmettre les signaux de parole à distance et nous nous prenons à rêver de machines qui pourraient nous parler, nous entendre et nous comprendre. Ce rêve est même déjà partiellement réalisé en synthèse de la parole et en reconnaissance vocale. La parole se synthétise de façon très réaliste à partir d'un texte par concaténation d'unités vocales pré-enregistrées (diphone, syllabe, mot ou même phrase complète). Les systèmes de reconnaissance d'un locuteur par apprentissage d'un vocabulaire ciblé sont actuellement très efficaces. Néanmoins, un élément essentiel de la production vocale humaine est encore mal maîtrisé : la qualité vocale. Nous savons faire parler des machines mais nous voudrions qu'elles s'expriment avec le naturel d'un être humain. Ces machines peuvent reconnaître le message phonétique que nous lui transmettons par la parole, mais nous aimerions qu'elles soient capables également de reconnaître les particularités qui distinguent une voix d'une autre et les émotions que nous véhiculons par notre voix.

Si la qualité vocale est un aspect de la phonation qui reste encore peu exploré, c'est essentiellement parce qu'elle traduit toute la complexité de la production vocale humaine. Nous pensons les sons que nous émettons et nous les percevons en retour. Nous pouvons agir sur ces sons que nous produisons, mais ces actions sont généralement inconscientes, résultats de notre apprentissage du langage. En parole, nous nous préoccupons peu de la qualité vocale de la phonation puisque l'intelligibilité du message phonétique est prioritaire : même avec une voix très rauque, une voix soufflée ou une voix pressée (trois qualités vocales que nous détaillerons par après), nous pouvons nous exprimer et nous faire comprendre. Dans le chant au contraire, la qualité de la voix prévaut et l'apprentissage vocal d'un chanteur consiste à en prendre conscience et à savoir la maîtriser indépendamment de l'articulation. Les productions vocales chantées constituent alors un support idéal pour l'étude de la qualité vocale, puisque les modifications de qualité y sont contrôlées, plus lentes et plus marquées que dans la parole, ainsi que l'illustre le crescendo chanté par un contre-ténor (CD, piste [1]), présenté sur la figure 1.

Cet exemple nous amène à définir plus précisement ce que nous entendons par le terme "qualité vocale". Selon le domaine d'étude (traitement de la parole, phoniatrie et orthophonie, enseignement du chant), il peut prendre des significations très variées. Ainsi, les phoniatres entendent par la notion de qualité vocale une terminologie précise permettant de définir les caractéristiques perceptives d'une voix pathologique : voix *hyperfonctionnelle*, *hypofonctionnelle*, *aphone*, *diplophonique*, *rugueuse*, ...(Hammarberg & Gauffin [95], 1995). Ces caractéristiques perceptives sont liées à des aspects fonctionnels du mouvement vibratoire des cordes vocales, comme par exemple, une compression trop importante des cordes vocales l'une contre l'autre (voix hyperfonctionnelle) ou au contraire, un relâchement excessif (voix hypofonctionnelle). Elles traduisent souvent la présence de pathologies (nodule, paralysie d'une corde, ...). Dans le cadre des voix non-pathologiques, la qualité vocale traduit essentiellement les variations d'intensité et de timbre d'un son produit. Les termes de voix *faible*, *normale* ou *forte* traduisent les nuances d'intensité dans le cadre de la parole, tandis que la terminologie *piano*, *mezzo-forte* et *forte* est plus adaptée dans le cadre du chant. Les termes de voix *timbrée* ou *détimbrée* traduisent la richesse spectrale d'une voix. Dans le cadre du chant, cet aspect de la qualité vocale est souvent un concept très subjectif au chanteur et à l'enseignant, basé sur une description imagée du timbre perçu (voix *blanche*, *sourde*) ou des sensations ressenties lors de la production (voix *engorgée*, *nasillarde*). Des termes imagés sont également employés en parole (Titze [205],

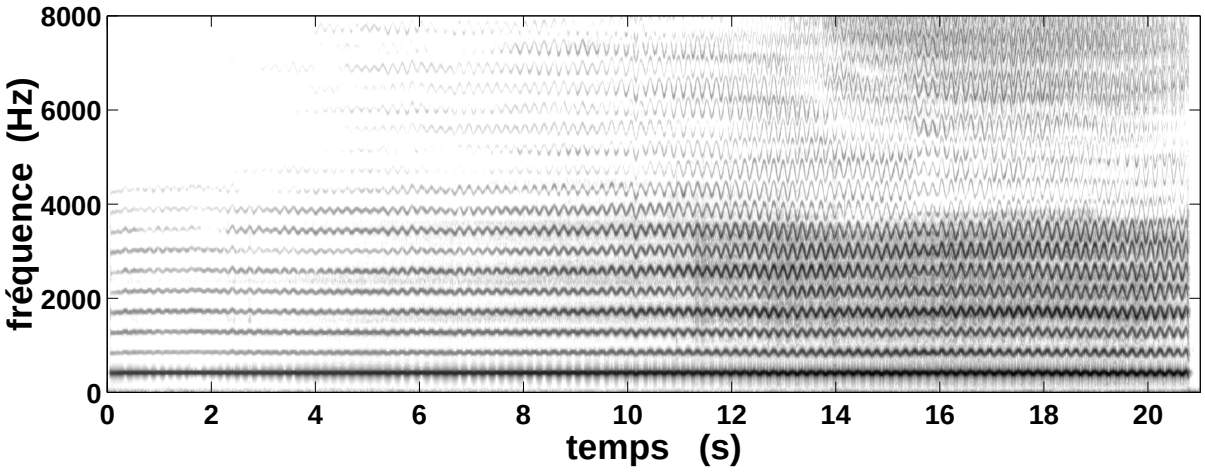


FIG. *1 – Crescendo chanté par le contre-ténor CT2. (CD, piste [1])*

1995) : voix *relâchée* ou *pressée* (impression ou non d'une tension dans la voix), voix *soufflée* (impression de souffle sur la voix). La qualité vocale peut également être le fait d'un changement de mécanisme laryngé [1] et l'on peut être amené à utiliser des termes empruntés au chant : voix *de fry*, *de poitrine*, *de tête*. La qualité vocale est également fortement modifiée par la présence de bruit (souffle) et d'irrégularités dans le mouvement vibratoire glottique (jitter, shimmer) (Pabon & Plomb [163], 1988 ; Klatt & Klatt [132], 1990 ; Childers & Lee [35], 1991 ; d'Alessandro & al [45], 1998).

Comme nous le voyons, l'étude de la qualité vocale est un sujet très vaste que nous n'avons pas prétention à couvrir dans son ensemble ici. La qualité vocale dépend principalement du mouvement vibratoire des cordes vocales (Miller [157], 1959). Nous nous limiterons donc à l'étude de la partie périodique des sons voisés (générés par vibration des cordes vocales), dans la parole et dans le chant, à travers la source glottique. La source glottique correspond au débit d'air qui traverse la glotte au cours du temps. Elle est représentée sur la figure 2, dans le cadre de la théorie source-filtre (Fant [64], 1960). Cette figure illustre le lien très fort entre la production acoustique d'un son et la perception à la fois auditive et proprioceptive que nous en avons. Ainsi, l'écoute de l'exemple sonore (CD, piste [1]) associé à la figure 1 soulève une question d'importance : comment un chanteur fait-il pour augmenter ainsi l'intensité sonore perçue sans marquer d'effort vocal apparent ? Des études comme celles menées par Monsen & Engebretson en 1977, par Holmberg & al [109, 110] en 1988 ou par Gauffin & Sundberg [85] en 1989 montrent que la forme temporelle du signal de source glottique varie en fonction des modifications d'intensité et de timbre. La façon plus ou moins rapide dont les cordes vocales se ferment est fortement corrélée à l'intensité vocale émise. La durée pendant laquelle la glotte est ouverte relativement à la période fondamentale de voisement est liée à une impression de tension dans la voix. De nombreuses questions se posent alors. Est-il possible de quantifier les variations de qualité en définissant des paramètres de source glottique pertinents d'un point de vue acoustique et perceptif ? Quelles sont alors les variations de ces paramètres dans un contexte vocal ? Est-il possible de caractériser une qualité vocale donnée par un jeu unique de paramètres ? Est-il possible de reproduire des changements de qualité vocale sur des sons synthétiques par seule modification de ces paramètres ? Telles sont les questions auxquelles cette étude se propose d'apporter des réponses.

[1] Nous reviendrons sur cette notion dans le chapitre 1 de la partie préliminaire et nous définirons précisement la terminologie utilisée pour décrire les mécanismes laryngés.

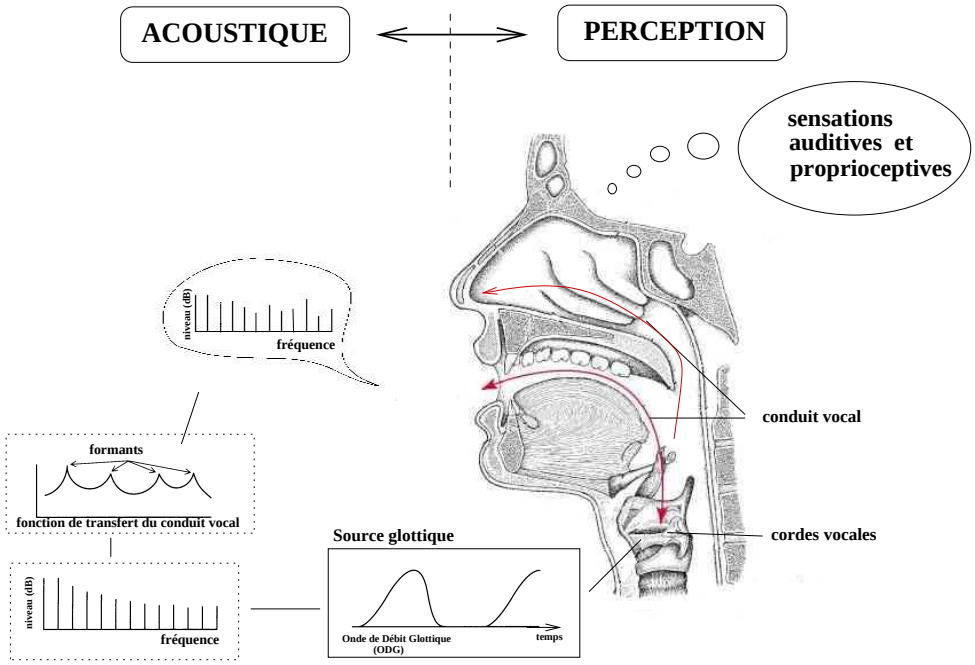


FIG. *2 – Principe de la production d'un son voisé, d'après la théorie source-filtre (Fant,* [64]*) : une onde est générée par la vibration des cordes vocales (source glottique). Cette onde se propage dans le larynx, le pharynx, les cavités buccales et nasales, dont l'ensemble constitue le conduit vocal, et elle est rayonnée vers l'extérieur par les lèvres et les narines. Le conduit vocal fait office de résonateur et certaines composantes fréquentielles de l'onde acoustique qui s'y propage sont alors accentuées (formants). L'acoustique et la perception sont étroitement liées : l'émission d'un son génère des sensations auditives et proprioceptives qui, en retour, agissent sur la production vocale. Certains des schémas de la partie acoustique sont extraits de Sundberg* [187] *(1987).*

L'articulation de l'étude

Trois approches complémentaires sont menées pour caractériser le lien entre la source glottique et la qualité vocale du son émis.

Première approche (partie I) : par modélisation du signal de source glottique et estimation par filtrage inverse

Cette première approche consiste à décrire le signal de source glottique en s'appuyant sur un modèle capable d'appréhender et de reproduire ce signal. Ce modèle peut être élaboré sur des aspects physiques du mouvement vibratoire des cordes vocales, comme la modélisation du mouvement mécanique des cordes par des systèmes masse-ressort par exemple. Il peut également s'attacher à décrire aussi fidèlement que possible le signal lui-même à l'aide de fonctions mathématiques qui ne se déduisent pas nécessairement des lois de la physique. Dans cette étude, nous avons choisi d'appliquer des modèles de signaux. Dans ce cadre, de nombreux modèles temporels de source glottique sont proposés dans la littérature : Rosenberg [169], 1971 ; Fant [65], 1979 ; Hedelin [100], 1984 ; Fant & al [72], 1985 ; Fujisaki & Ljungqvist [146], 1986 ; Klatt & Klatt [132], 1990 ; Milenkovic [153] 1993 ; Childers & Hu [32], 1994 ; Fant [69], 1995 ; Veldhuis [213] 1998. Ces modèles présentent des formes temporelles de source glottique très similaires, mais ne s'appuient pas forcément sur les mêmes paramètres, ni sur le même nombre de paramètres. Dans un souci d'unification, Doval & d'Alessandro [55] (1999) ont alors exploré certains de ces modèles afin d'en déterminer une description commune. Nous présenterons ce formalisme dans le chapitre 3. Ces modèles de source glottique sont exprimés dans le domaine temporel. Pourtant, la forme spectrale d'un signal est souvent plus intéressante à étudier que sa forme temporelle, car la perception humaine s'appréhende

mieux par une description spectrale. Peut-on caractériser spectralement les effets perceptifs de la source glottique ? Pour répondre à cette question, nous analyserons dans le chapitre 3 les propriétés spectrales des paramètres de source glottique d'un point de vue théorique, dans le but de comprendre les effets perceptifs de ces paramètres sur un signal voisé. Mais peut-on réellement estimer la source glottique à partir des théories et des modèles actuels ? Nous tenterons de répondre à cette question au chapitre 4 en étudiant la différence en amplitude entre les deux premiers harmoniques du spectre de source glottique. Plusieurs études expérimentales ont montré une corrélation forte entre cette différence spectrale et le quotient ouvert, un des paramètres de source glottique auquel nous nous intéresserons tout particulièrement tout au long de l'étude (Holmberg & al [111], Fant [69, 70], Sundberg & al [190]) et avancé que cette différence spectrale permettrait une mesure indirecte du quotient ouvert (Stevens & Hanson [184], Hanson [97, 98], Fant [69, 70]). Nous étudierons cette relation de façon théorique en partie 4.1, à la lumière du formalisme développé au chapitre 3. Cette approche théorique se doit d'être complétée par une analyse de signaux voisés réels de parole et de chant. Pour ce faire, des méthodes d'estimation de la source glottique par filtrage inverse proposées dans la littérature seront présentées et évaluées en partie 4.2. Elles seront appliquées en partie 4.3 pour l'étude de la relation possible entre cette différence spectrale et les paramètres de source glottique dans le cas de signaux voisés réels de parole et de chant.

Seconde approche (partie II) : par des mesures acoustiques et électroglottographiques

L'approche que nous venons de développer se base sur la théorie source-filtre exposée par Fant [64] en 1960. Cette théorie fondamentale de la production vocale fait l'hypothèse d'une non-interaction entre le signal d'excitation source et le filtre modélisant le conduit vocal. En particulier dans le chant, de telles hypothèses sont remises en question et les méthodes d'analyse employées peuvent alors paraître inadaptées. Des mesures directes des paramètres de source glottique permettraient de contourner ce problème, mais elles sont difficilement réalisables en raison du caractère souvent très invasif qu'elles présentent. Certaines méthodes indirectes et non-invasives permettent d'appréhender des aspects du mouvement vibratoire glottique de façon découplée du système supraglottique. L'électroglottographie se place en avant de ces méthodes. Elle renseigne sur le degré d'accolement des cordes vocales au cours de la phonation. Nous présenterons donc dans le chapitre 5 le principe de l'électroglottographie et son application à la mesure des paramètres de source glottique. Comme il n'existait aucune base de données disponible pour cette étude, nous avons été amenés à la constituer. Des enregistrements acoustiques et électroglottographiques simultanés ont donc été effectués sur 18 chanteurs entraînés, à partir d'un protocole défini pour l'étude. Ils seront présentés au chapitre 6. La période fondamentale de vibration glottique et le quotient ouvert sont des paramètres de source glottique que l'on peut mesurer par électroglottographie, mais les signaux électroglottographiques ne permettent qu'une mesure approximative du quotient ouvert. Pourtant, la dérivée de ces signaux apporte des informations précises quant aux instants d'ouverture (initialisation du débit glottique) et de fermeture glottique (arrêt du débit). Ces signaux n'ont été que très peu exploités dans la littérature. Nous discuterons ce point dans le chapitre 7 et nous montrerons que de tels signaux sont tout à fait exploitables pour peu que certaines précautions soient prises lors de l'analyse. Nous montrerons qu'ils apportent des informations intéressantes et complémentaires aux signaux électroglottographiques. Dans le chapitre 8, nous appliquerons ces méthodes pour analyser le quotient ouvert en fonction des mécanismes laryngés, de l'intensité vocale et de la fréquence fondamentale, sur des voyelles tenues, des sons filés et des phrases parlées et/ou chantées extraites de la base de données.

Troisième approche (partie III) : par la perception auditive

Quelle soit théorique ou expérimentale, toute étude de la source glottique se doit de rester fondée sur la perception que nous avons du son produit, ainsi que le montre la figure 2. Nous illustrerons donc tout ce travail par des exemples sonores, afin de marquer à chaque fois le lien entre ce que nous analysons ou synthétisons et ce que nous percevons. Aussi précises et fines soient les analyses que nous pouvons effectuer sur les sons voisés, l'oreille humaine reste seule experte et juge de la qualité vocale. Les variations mesurées d'un paramètre de source glottique ne deviennent pertinentes que si leur influence sur la qualité vocale du son émis est perceptible. La marge d'erreur d'une méthode de mesure de ces paramètres ne sera significative que si elle est supérieure à un seuil différentiel qui correspond à la plus petite différence de qualité perceptible. De même, deux méthodes de mesure seront similaires en pratique si la différence de

mesure entre ces deux méthodes est inférieure au seuil différentiel. Nous mesurerons au chapitre 9 les seuils différentiels associés à certains paramètres de source glottique, afin de déterminer dans quelle mesure la modification de ces paramètres est perceptible. Nous nous intéresserons en particulier à l'effet possible du contexte (nature de la voyelle, fréquence fondamentale, paramètre d'amplitude associé, vibrato).

Le but de cette thèse est d'apporter de nouvelles connaissances sur la production et la perception de la voix, en particulier une meilleure connaissance objective des paramètres de source glottique liés à la qualité vocale. Les applications d'une telle étude sont multiples, tant en analyse et synthèse de la parole que dans le domaine médical ou pour l'enseignement du chant. Une meilleure compréhension du lien entre qualité vocale et source glottique permettrait de pouvoir apporter aux voix de synthèse le naturel qui leur fait encore défaut. Elle permettrait de modifier la qualité d'un son vocal par analyse-synthèse et modification du signal de source glottique. L'amélioration de l'analyse de la source glottique pourrait permettre de caractériser les émotions véhiculées par la voix. De manière plus générale, cette étude pourrait également s'appliquer à la description et la caractérisation des voix dans le cadre d'une rééducation vocale ou de l'apprentissage du chant.

Partie préliminaire

Caractérisation de la source glottique

Si l'on fait une Analyse exacte des lèvres de la Glotte, si on les dépouille des parties qui les environnent, et qu'après on porte un oeil curieux sur elles, on observera qu'elles font une espèce de rubans larges d'une ligne, tendus horizontalement, arrêtés par les deux bouts, susceptibles de plusieurs degrés de tension, et de différentes vibrations, ou oscillations, et séparés l'un de l'autre par l'intervalle de la Glotte, de sorte que l'air ne sauroit sortir sans déployer contr'eux son action.

(Jean-Antoine Bérard, « L'Art du Chant », 1755)

Sommaire de la partie

Dans cette partie introductive, nous nous proposons de présenter brièvement les caractéristiques générales de la source glottique sur le plan physiologique et physique.

Etudier la source glottique nécessite en effet de connaître l'anatomie des cordes vocales, de comprendre les actions musculaires mis en jeu lors du mouvement vibratoire glottique et de pouvoir caractériser celui-ci pour les émissions vocales parlées et chantées. En particulier, la production vocale humaine repose sur l'utilisation de quatre mécanismes laryngés distincts, qui permettent la production d'une gamme très étendue de fréquence. Dans le chapitre 1, nous ferons donc une revue rapide des connaissances physiologiques que nous avons de la source glottique. Nous évoquerons également les principales méthodes actuelles d'exploration de l'activité glottique.

Etudier la source glottique, c'est aussi comprendre les modèles théoriques qui permettent de la décrire. La théorie source-filtre exposée par Fant [64] en 1960 est à la base des modèles de production des sons vocaux. Elle sera donc introduite au chapitre 2. Les paramètres de source glottique y seront explicités et leur influence sur la qualité du son produit y sera détaillée.

Cette partie préliminaire constitue la base de notre étude et nous serons souvent amenés à nous y référer par la suite.

Chapitre 1

Description physiologique

Il y a quelques siècles à peine, les hommes pensaient que la voix était l'expression de l'âme et qu'un corps privé de vie ne pouvait donc plus émettre de son. Mais quand on eût réussi à produire un son en insufflant de l'air avec un soufflet dans un larynx humain excisé, on comprit que le travail vocal est principalement musculaire et respiratoire. Les cordes vocales, source de la phonation, sont avant tout deux muscles qui se contractent ou se dilatent, et qui s'éloignent ou se rapprochent sous la pression de l'air expulsé hors des poumons. Nous allons nous attacher à décrire dans ce chapitre l'anatomie du larynx, à préciser les actions musculaires mis en jeu et à décrire les mécanismes laryngés qui caractérisent le mouvement vibratoire glottique. Nous explorerons ensuite les méthodes qui permettent d'étudier la source glottique de façon qualitative et quantitative.

1.1 Description statique et dynamique du larynx

1.1.1 Anatomie du larynx et vibration laryngée

L'anatomie de l'appareil phonatoire est détaillée sur la figure 1.1, qui présente des coupes schématiques de l'appareil phonatoire et des cordes vocales. Dans un plan de coupe horizontal, le côté avant (partie antérieure des cordes vocales) correspond à la partie avant du cou, caractérisée par une proéminence (pomme d'Adam) chez les hommes. Le côté arrière (partie postérieure des cordes vocales) correspond à la partie médiane du cou, avant l'oesophage. On note sur la coupe (3) que la corde vocale n'est pas simplement un muscle (le corps), mais qu'elle est également recouverte de plusieurs couches de fibres (la transition) et d'une couche de muqueuse (la couverture). La figure 1.2 présente deux vues de dessus du larynx dans le plan de la glotte, en position de respiration et en phonation. La glotte constitue l'espace situé entre les deux muscles vocaux par où l'air est expulsé. Lors d'un mouvement respiratoire, les aryténoïdes sont écartés, laissant le flux d'air circuler sans contrainte. En début de phonation, ces cartilages pivotent pour s'accoler et entraîner une fermeture de la glotte.

La théorie aérodynamique-myoélastique développée par Van den Berg [210] en 1958 explique le principe de la vibration laryngée. La glotte fermée oppose une résistance au passage de l'air venant des poumons et la pression en-dessous des cordes (pression sous-glottique) s'accroît, jusqu'à vaincre cette résistance et forcer l'ouverture glottique. Une fois la glotte ouverte, la pression sous-glottique décroît et les cordes vocales s'accolent à nouveau sous l'effet de leur élasticité propre et d'un effet mécanique de succion (effet de Bernoulli, Van den Berg & al [211]) . En une seconde, les cordes vocales vont ainsi s'accoler et se décoller quelques centaines de fois (440 fois pour produire un la3, $f_0 = 440$ Hz), générant par ce mouvement vibratoire des ondes acoustiques qui vont se propager dans le conduit vocal.

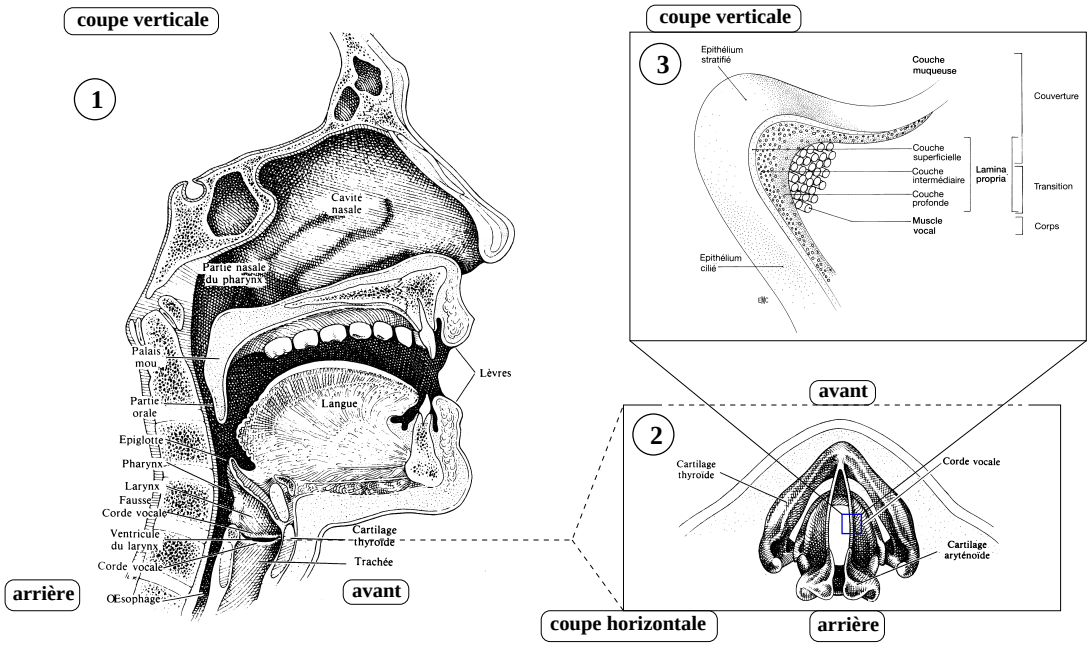


FIG. *1.1 – Description physiologique de l'appareil phonatoire :* **(1)** *vue d'ensemble (coupe verticale),* **(2)** *les cordes vocales (coupe horizontale, au niveau de la pomme d'Adam),* **(3)** *visualisation de l'épaisseur d'une corde vocale (coupe verticale). (*(**1**) *et* **(2)** *d'après Sundberg* [186] *et Miller* [156]*,* **(3)** *d'après Lacau St Guily & Roubeau* [92]*)*

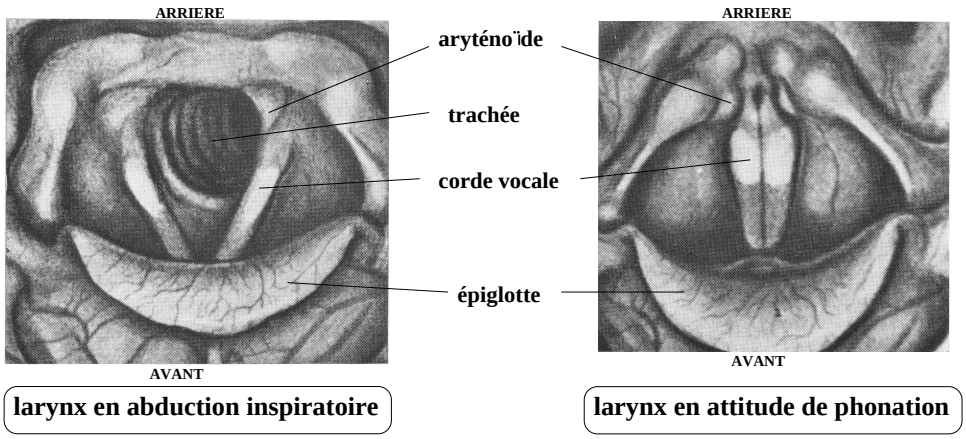


FIG. *1.2 – Images laryngoscopiques d'un larynx en position de respiration (abduction inspiratoire) et en phonation (d'après Husson* [114]*, 1962)*

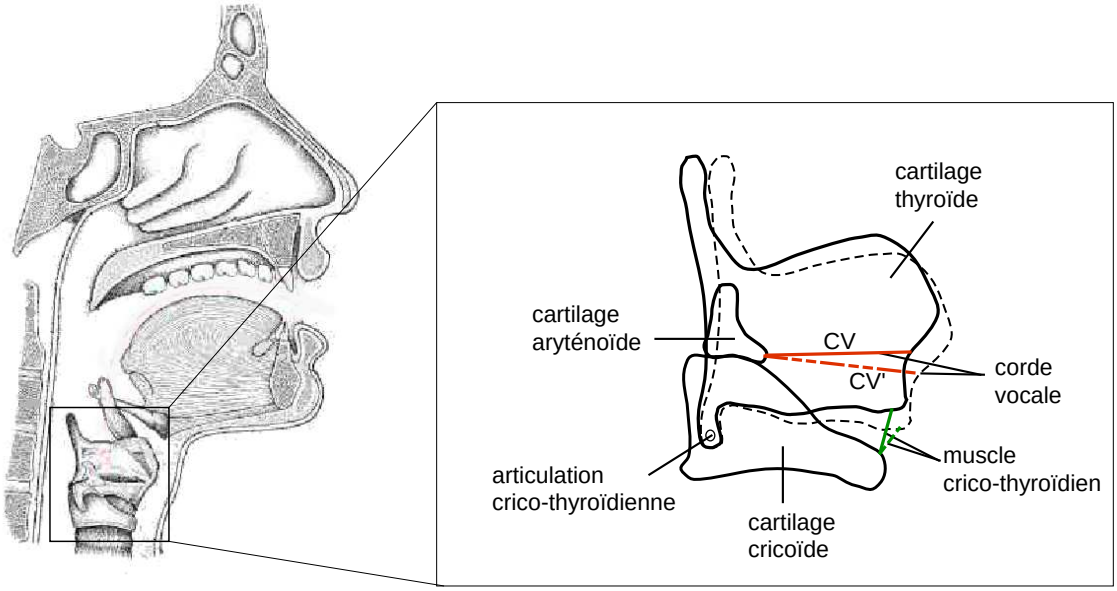


Fig. *1.3 – Sous l'effet de la contraction du muscle crico-thyroïdien, le cartilage thyroïde bascule par rapport au cartilage cricoïde, entraînant par ce mouvement un allongement des cordes vocales (CV)*

1.1.2 Actions musculaires

Le mouvement vibratoire des cordes vocales met en action un ensemble de muscles, dont l'étude est possible par électromyographie, une technique très invasive qui consiste à mesurer l'activité musculaire à l'aide d'électrodes insérées dans les muscles grâce à des aiguilles.

La musculature laryngée se sépare en deux catégories :

1. les muscles extrinsèques (sous- et sus-hyoïdiens), qui agissent sur la mobilité du larynx et de la mâchoire inférieure (abaissement, élévation),
2. les muscles intrinsèques (thyro-aryténoïdien, inter-aryténoïdien, crico-thyroïdien, crico-aryténoïdien), qui vont agir directement sur les cordes vocales (position, longueur, tension) et sur leur mouvement vibratoire.

Modification de longueur et d'épaisseur des cordes vocales

Le muscle thyro-aryténoïdien interne constitue le muscle vocal, encore appelé vocalis. Il peut se contracter, entraînant un rétrecissement et un épaississement des cordes vocales. Il est étiré sous l'action des autres muscles intrinsèques, ce qui entraîne une élongation et un amincissement des cordes vocales. En particulier, le cartilage thyroïde peut basculer ou se translater par rapport au cartilage cricoïde sous l'action du muscle crico-thyroïdien, ainsi que l'illustre la figure 1.3 dans le cas du basculement. Ces actions musculaires (contraction du thyro-aryténoïdien, contraction du crico-thyroïdien) entraînent une augmentation de la tension des cordes vocales.

Mouvements d'adduction et d'abduction

Comme l'illustre la figure 1.4, les mouvements d'adduction (fermeture glottique) ou d'abduction (ouverture glottique) s'effectuent par déplacement des cartilages aryténoïdes (pivotement, glissement) sous

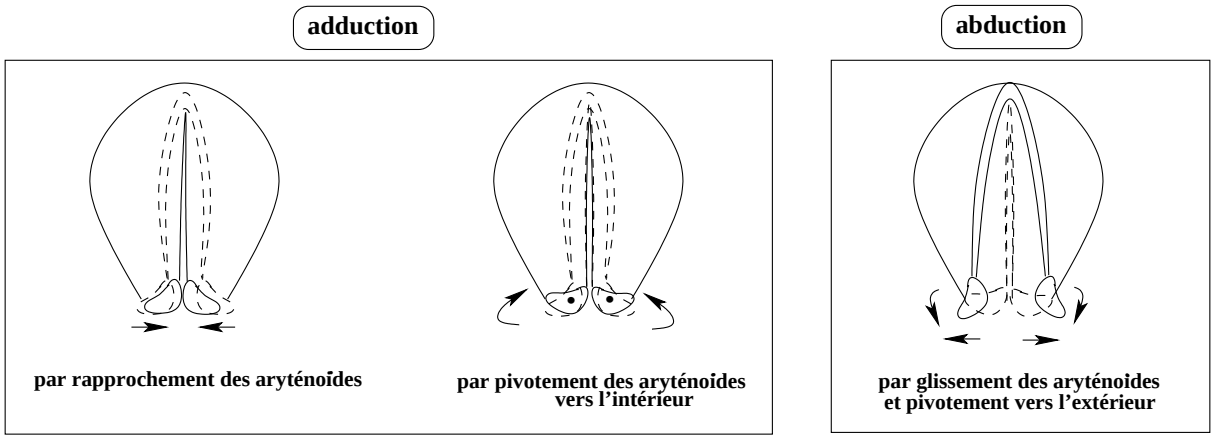


FIG. *1.4 – Schématisation des mouvements d'adduction (fermeture glottique) ou d'abduction (ouverture glottique) par déplacement des cartilages aryténoïdes.*

l'action des muscles intrinsèques, principalement le crico-aryténoïdien et l'inter-aryténoïdien. L'adduction s'effectue par rapprochement des aryténoïdes sous l'action de l'inter-aryténoïdien ou par rotation des aryténoïdes vers l'intérieur sous l'action du crico-aryténoïdien latéral. L'abduction s'effectue par rotation des aryténoïdes vers l'extérieur sous l'action du crico-aryténoïdien postérieur.

1.1.3 Quelques ordres de grandeurs

Les cordes vocales ont une longueur de l'ordre de 13 à 17 mm chez les femmes et 17 à 24 mm chez les hommes et elles peuvent s'étirer sur 3 - 4 mm (Titze [204], Childers [29]). L'amplitude de l'ouverture glottique est d'environ 3 mm (Lacau St Guily & Roubeau [92]).

La pression sous-glottique varie en moyenne entre 10 cm d'eau (1 kPa soit 30 dB) (cas d'une conversation normale) et 360 cm d'eau (120 dB) (ténor au maximum de sa puissance) (Husson [114]).

La fréquence fondamentale des vibrations laryngées entretenues varie de 60 Hz à plus de 1500 Hz. La figure 1.5 présente l'étendue vocale moyenne en voix parlée et chantée.

Pour couvrir une telle étendue de fréquence, l'appareil phonatoire dispose de plusieurs modes vibratoires. Nous allons donc maintenant aborder un aspect fondamental de la production vocale, la notion de mécanisme laryngé.

1.2 Les mécanismes laryngés

1.2.1 Illustration

La figure 1.6 présente le glissando ascendant effectué par une soprano légère (soprano S3, voir chapitre 6) depuis les fréquences les plus graves jusqu'aux fréquences les plus aiguës de sa tessiture. A l'écoute de cet exemple ((CD, piste [2])), nous remarquons que la montée en fréquence n'est pas continue et nous pouvons distinguer 3 ruptures marquant les changements de mécanisme laryngé. Ces ruptures correspondent à des brusques sauts de fréquence fondamentale, qui se traduisent visuellement par une discontinuité sur le tracé sonagraphique du glissando. Ces mécanismes laryngés ne sont pas propres au chant mais sont une caractéristique principale de la production vocale humaine. Dans le chant lyrique en particulier, les chanteurs évitent même de produire de telles ruptures dans la ligne mélodique et apprenent par la

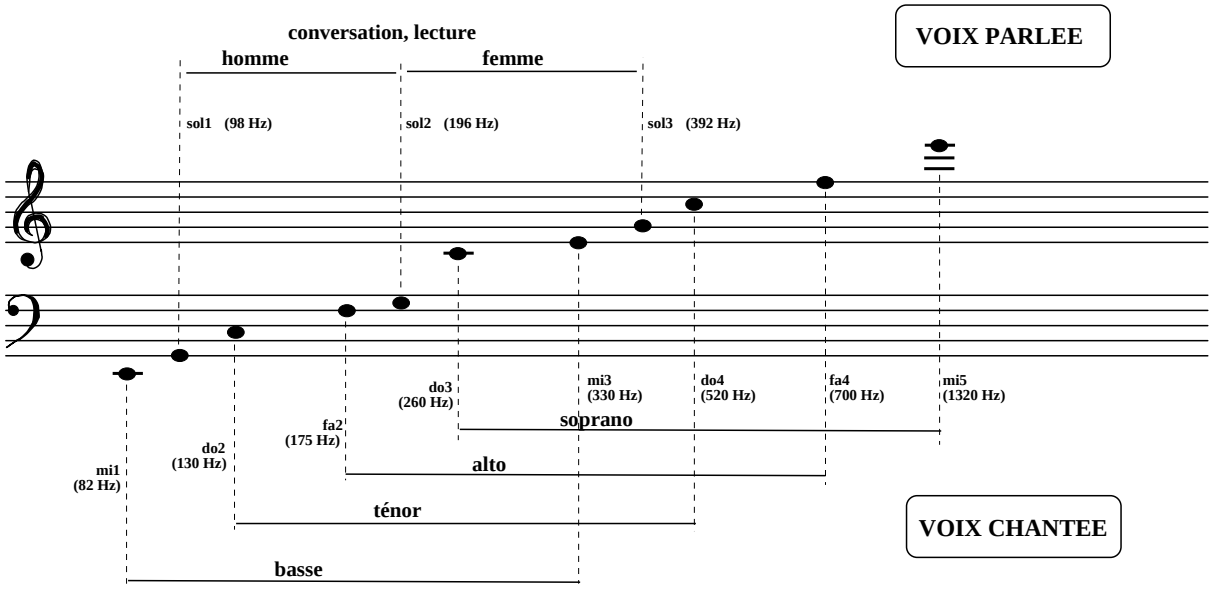


Fig. *1.5 – Etendue vocale moyenne en voix parlée et en voix chantée (d'après Le Huche & Allali* [136]*, Sundberg* [189]*)*

technique vocale à lisser les passages entre mécanismes pour les rendre inaudibles.

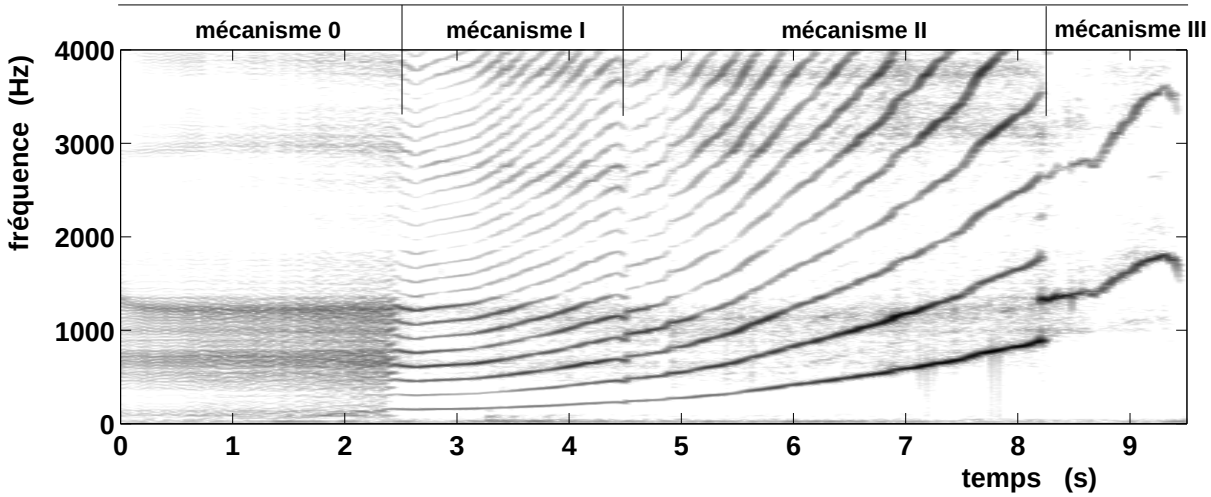


Fig. *1.6 – Glissando effectué par une chanteuse (soprano légère) qui couvre l'ensemble des* 4 *mécanismes laryngés utilisés en phonation : m0, mI, mII et mIII. (CD, piste [2])*

1.2.2 Définition

Les chanteurs parlent des registres de la voix. Ce terme désigne une réalisation complexe englobant tout à la fois les mécanismes laryngés et les adaptations du conduit vocal. Un travail de longue haleine a été conduit au Laboratoire d'Acoustique Musicale (Castellengo & al [28], 1983 ; Roubeau & al [177], 1987 ; Roubeau & al [176], Castellengo [26], 1991 ; Roubeau [174], Roubeau & Castellengo [175], 1993 ; Castellengo & Roubeau [27], 2001) pour caractériser d'un point de vue acoustique et physiologique les « registres

laryngés », qui seuls nous concernent ici. En accord avec ces auteurs, nous appelons mécanisme laryngé une configuration glottique particulière, caractérisée par la forme des cordes vocales (longueur, épaisseur) ainsi que par la tension musculaire mise en jeu. La production vocale humaine non-pathologique se caractérise par l'utilisation de 4 mécanismes laryngés :

* **Le mécanisme 0** ou voix de fry : les cordes vocales sont courtes, très épaisses et peu tendues (Hollien & al [108]). La vibration est caractérisée par une durée d'ouverture glottique très brève relativement à la période fondamentale. Ce mécanisme permet de réaliser les fréquences les plus basses. Il se rencontre essentiellement en parole, mais très peu en chant lyrique occidental.

* **Le mécanisme I** : les cordes vocales sont épaisses et vibrent sur toute leur longueur. La masse vibrante est importante, ainsi que l'amplitude de la vibration. La phase fermée est souvent plus importante que la phase ouverte [1]. En voix parlé, ce mécanisme est utilisé par l'homme sur l'ensemble de sa tessiture et par la femme dans le grave de sa tessiture. En voix chantée, il constitue le mécanisme principal de la production vocale des basses, des barytons et des ténors et des altos féminins, ainsi que des chanteurs et chanteuses de variétés en général.

* **Le mécanisme II** : les cordes vocales sont fines et ne vibrent plus que sur les 2/3 de leur longueur, du fait d'une compression accentuée entre les aryténoïdes. La phase ouverte est très importante. En voix parlée, ce mécanisme peut être employé de façon occasionnelle par l'homme pour émettre des sons aigus, tandis que la femme l'utilise soit exclusivement, soit en alternance avec le mécanisme I. En voix chantée, il est principalement utilisé par les altos masculins, les mezzo-sopranos et les sopranos. Dans d'autres cultures (Asie, Afrique), son utilisation par l'homme est courante comme moyen d'expression parlée ou chantée.

* **Le mécanisme III** ou voix de sifflet : les cordes vocales sont fines, très tendues et l'amplitude de vibration est très réduite. Il n'y a quasiment pas de fermeture. Ce mécanisme permet de réaliser les fréquences les plus aiguës. [2]

Les mécanismes I et II sont très courants, utilisés dans 90 % des productions vocales. La figure 1.7 illustre la configuration glottique mise en jeu dans ces deux principaux mécanismes (mI et mII) par une vue de dessus du plan de la glotte et une vue verticale de l'épaisseur des cordes vocales.

Il est important de noter que ces deux mécanismes couvrent des plages de fréquence qui ne sont pas disjointes et qu'il n'existe donc pas une unique fréquence fondamentale en-dessous de laquelle se situe le mécanisme I et au-dessus de laquelle se situe le mécanisme II. La figure 1.8 illustre les zones fréquentielles couvertes par l'homme et la femme en mécanisme I et en mécanisme II. En moyenne, l'homme peut chanter entre si0 ($f_0 = 62$ Hz) et do4 (523 Hz) en mécanisme I et entre mi2 (164 Hz) et do5 (1046 Hz en mécanisme II. La femme peut chanter entre ré2 (147 Hz) et do4 (523 Hz) en mécanisme I et entre fa2 (174 Hz) et la5 (1760 Hz) en mécanisme II. Dans la zone commune aux deux mécanismes (i.e. entre mi2/ré2 et do4), le chanteur (homme ou femme) pourra produire des sons voisés en utilisant soit l'un soit l'autre des deux mécanismes. Le passage d'un mécanisme à un autre peut même devenir un ornement, à l'exemple du chant yodlé (Castellengo [26]).

1.2.3 Terminologie

Dans la littérature, une grande confusion règne dès qu'il s'agit de qualifier ces mécanismes laryngés. Roubeau [174] (1993) en présente une revue bibliographique très détaillée. Cette confusion provient du décalage qu'il peut y avoir entre ce que le chanteur ressent et exprime et la réalité physiologique du mouvement vibratoire des cordes vocales. Une distinction est à faire entre les « *mécanismes laryngés* », qui résultent de configurations glottiques spécifiques, et les « *registres résonantiels* », qui, comme le nom

[1] La terminologie que nous appliquons ici est précisée par après, en partie 2.2.1.

[2] Il est à noter que le son est bien dû aux vibrations des cordes vocales et non à un système jet d'air/biseau, comme pourrait le laisser croire le terme « sifflet ».

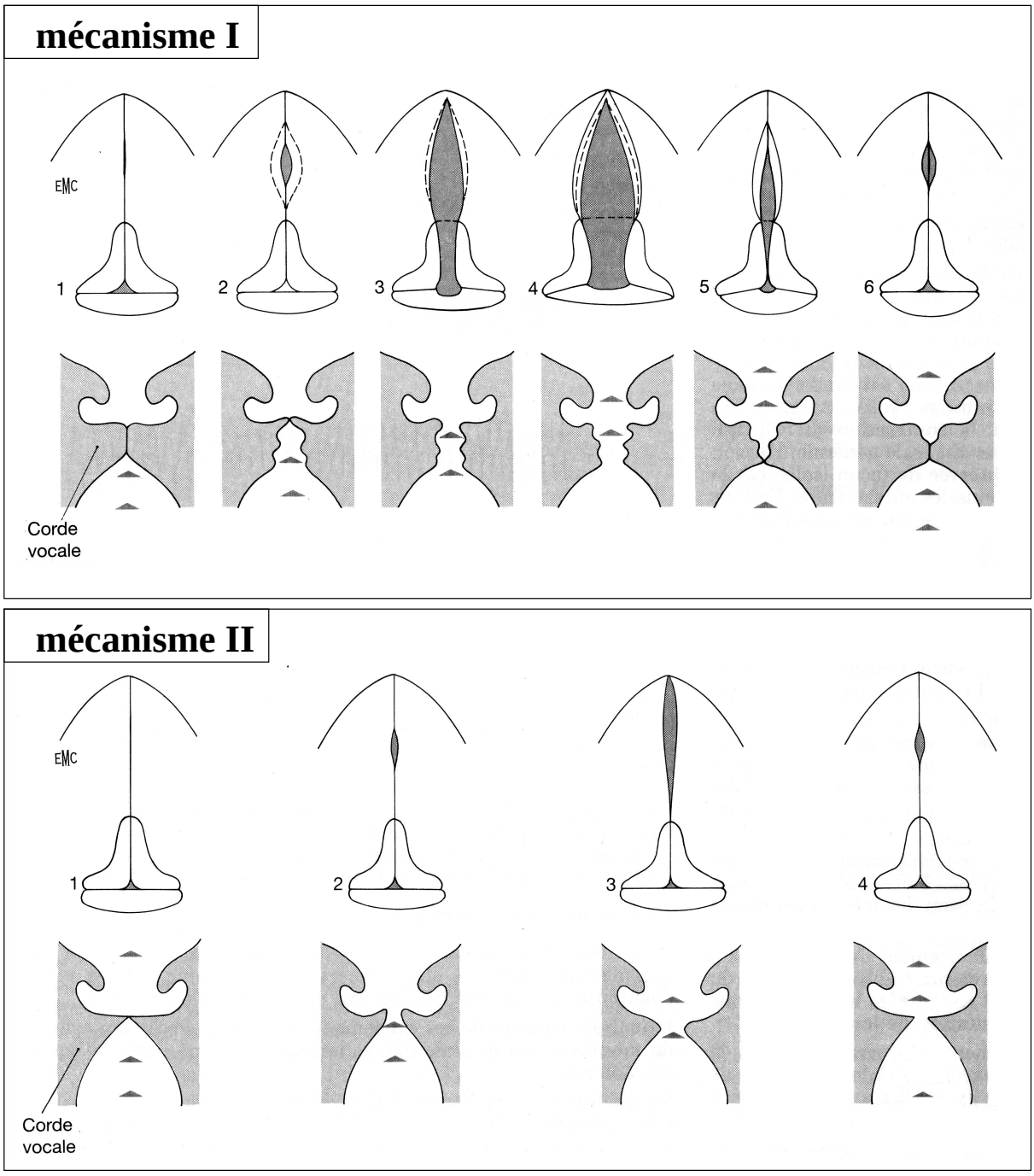


Fig. *1.7 – Visualisation de la configuration glottique associée aux mécanismes I et II (Lacau St Guily & Roubeau [92], d'après Vennard [214], 1967). En mécanisme I, les cordes vocales sont épaisses et vibrent sur toute leur longueur. En mécanisme II, l'épaisseur et la longueur vibrante sont réduites.*

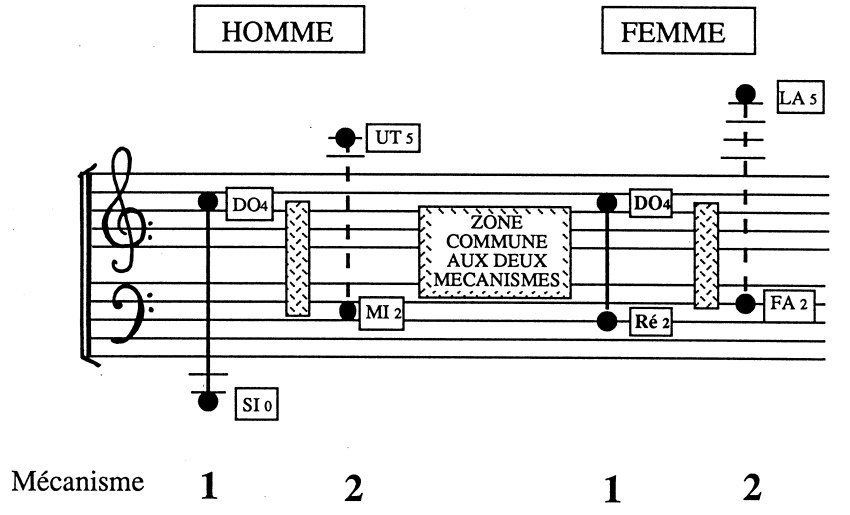


FIG. *1.8 – Zones fréquentielles moyennes couvertes par les mécanismes I et II, pour l'homme et la femme en voix chantée. Il existe une plage de fréquence importante où les deux mécanismes peuvent coexister. D'après Castellengo* [26]

l'indique, résultent de l'ajustement des cavités buccales et nasales qui servent de résonateur (Roubeau [174]). Au sein d'un même mécanisme laryngé, un chanteur peut donc ressentir plusieurs registres résonantiels. Ainsi, ce que l'on nomme « voix de poitrine » et « voix de tête » chez l'homme correspond simplement à deux registres résonantiels distincts, tous deux produits en mécanisme I (Roubeau [174], Chuberre [39]). Par contre, chez la femme, la « voix de poitrine » correspond au mécanisme I et la « voix de tête » au mécanisme II. Ceci explique qu'une définition précise de la terminologie et de ce qui s'y réfère soit nécessaire. Afin d'éviter ces possibles confusions et pour insister sur le fait qu'un mécanisme laryngé donné est associé à une configuration glottique identique chez l'homme et chez la femme, nous avons choisi d'utiliser les termes de « mécanisme 0 », « mécanisme I », « mécanisme II » et « mécanisme III » proposés par Castellengo ([25] 1986, [26] 1991) et Roubeau ([174] 1991). Nous les jugeons plus explicites et moins à même de porter à confusion que des termes comme « registre ou voix de poitrine », « registre ou voix de tête », « registre ou voix de fausset ».

1.2.4 Les transitions entre mécanismes

Le passage d'un mécanisme à un autre correspond à un ajustement physiologique, dont la durée n'excède pas quelques centaines de *ms* (Roubeau [174]). Il s'accompagne d'une perte de contrôle momentanée de la fréquence fondamentale, résultat d'une modification brutale de la masse vibrante. Ce saut en fréquence est particulièrement perceptible lors de la production d'un glissando (Castellengo & al [28]), comme le montre l'exemple illustratif précédent (figure 1.6). Ces transitions sont audibles, mais elles peuvent être rendues imperceptibles par un travail vocal. C'est en particulier le cas dans le chant lyrique occidental, où l'homogénéité du timbre prévaut sur l'ensemble de la tessiture. Par contre, d'autres cultures vocales s'appuient justement sur ces transitions et les développent (Castellengo [26]).

Dans le cas de la transition entre mécanisme I et II, deux théories sont proposées pour expliquer ce changement brutal de configuration glottique. La plus ancienne et plus communément acceptée pose que celui-ci est le fait d'une modification soudaine de l'activité des muscles laryngés intrinsèques, en particulier une chute de l'activité du vocalis qui atteint alors une limite physiologique de contraction (Hirano & al [107], Titze [204]). Plus récemment, une hypothèse alternative, basée sur la théorie de la dynamique non-

linéaire, pose que cette transition peut avoir lieu spontanément, sans s'accompagner nécessairement d'un ajustement des muscles laryngés et qu'elle correspond à un phénomène de « bifurcation » du système vibratoire (Titze & al [207], Herzel [105], Svec & al [194]). Le terme « bifurcation » signifie que la nature du mouvement vibratoire change brutalement, tandis que les ajustements phonatoires sont plus graduels.

1.2.5 Contrôle de la fréquence fondamentale

Une augmentation de la fréquence fondamentale de vibration des cordes vocales est principalement le fait des actions suivantes (Titze, [204]) :

1. une augmentation de la pression d'air dans les poumons (pression sous-glottique),
2. une contraction du vocalis (augmentation de la raideur des cordes vocales),
3. une diminution de la masse vibrante,
4. une diminution de la longueur vibrante.

Ces actions peuvent être effectuées de façon conjointe. En particulier, les cordes vocales s'ouvrent quand la pression sous-glottique est supérieure à une pression seuil (Titze [203]), qui va dépendre de la raideur des cordes vocales. Une accentuation de la raideur des cordes amène à une augmentation de la fréquence fondamentale et la pression sous-glottique nécessaire pour initier la vibration devient plus importante. Ainsi, la pression sous-glottique et la fréquence fondamentale seront bien souvent corrélés (Titze & Sundberg [209]).

Nous avons vu précédemment que la raideur, la longueur et l'épaisseur des cordes vocales étaient modifiées par l'action principale des muscles intrinsèques. De nombreuses recherches ont été effectuées pour déterminer les activités musculaires mises en jeu dans le contrôle de fréquence fondamentale (pour une revue détaillée, voir Titze [204], p. 191-209). Elles montrent que ce contrôle n'est pas uniquement le fait de l'action d'un ou deux muscles intrinsèques, mais d'un ajustement complexe des activités musculaires laryngés. Cela dépend également du mécanisme laryngé mis en jeu (St Guily & Roubeau [92]). Nous allons résumer brièvement les actions musculaires majeures mises en jeu lors d'une augmentation de la fréquence fondamentale.

En mécanisme 0 : une augmentation de f_0 est principalement le fait d'une augmentation de la pression sous-glottique. Le rôle des muscles intrinsèques semble limité.

En mécanisme I : le vocalis se contracte, ce qui provoque un raccourcissement et un épaississement des cordes vocales. La raideur des cordes vocales augmente. En toute logique, la longueur des cordes vocales devrait diminuer, or justement, elle augmente sous l'effet antagoniste des muscles crico-thyroïdiens qui se contractent. La fréquence fondamentale augmente par accentuation de la raideur, alors même que les cordes vocales s'allongent.

En mécanisme II : au moment du passage en mécanisme II, l'activité musculaire du vocalis chute brutalement car ce muscle a alors atteint une limite physiologique de raideur. Sa participation au mouvement vibratoire de la corde est réduite comparativement au mécanisme I. Par contre, les muscles crico-thyroïdiens continuent de se contracter et entraînent une augmentation de la tension des cordes. En parallèle, la compression entre les aryténoïdes s'accentue sous l'action des muscles crico-aryténoïdiens latéraux, ce qui implique un raccourcissement de la partie vibrante. L'augmentation de la tension des cordes et la diminution de la masse vibrante contribuent à l'augmentation de la fréquence fondamentale.

En mécanisme III : les cordes vocales sont très tendues et cette tension peut encore s'accroître sous l'action des muscles crico-aryténoïdiens latéraux. L'augmentation de la tension des cordes, ainsi que l'augmentation de la pression sous-glottique induite par le rétrécissement de l'ouverture glottique peuvent alors expliquer une augmentation de la fréquence fondamentale.

Selon le mécanisme laryngé considéré, le lien entre fréquence fondamentale et pression sous-glottique sera également modifié. En particulier, le mécanisme I se traduit par des variations de fréquence fondamentale en fonction de la pression sous-glottique de l'ordre de 1 à 3 Hz / cm H_2O dans le bas de la tessiture et 2 à 6 Hz / cm H_2O dans le haut de la tessiture, tandis que les variations de fréquence fondamentale sont plutôt de l'ordre de 5 à 10 Hz / cm H_2O en mécanisme II (Titze [202]).

1.3 Méthodes exploratoires

Méthodes directes d'observation du mouvement vibratoire des cordes vocales

Comment explorer cet appareil vocal très secret parce qu'inaccessible ? Poussé par la curiosité, le chanteur et professeur de chant Manuel Garcia réussit à observer les cordes vocales à l'aide d'un petit miroir placé au bout d'une tige (Behnke [18], 1880, Garcia [84], 1884). C'est l'avènement de la laryngoscopie qui, de 1845 à nos jours, va évoluer et se complexifier. Actuellement, une multitude de méthodes de visualisation directe se sont mises en place. Elles permettent de nous renseigner très précisement sur le comportement vibratoire des cordes vocales (Baken & Orlikoff [17], Childers [29]) :

laryngoscopie stroboscopique : sur le principe de la stroboscopie, cette méthode permet de ralentir visuellement le mouvement vibratoire des cordes vocales en les éclairant de façon discontinue à une période proche de la période fondamentale de vibration. On peut ainsi reconstruire un cycle glottique à partir d'images prises sur plusieurs cycles consécutifs à des instants décalés d'un peu plus (ou moins) d'une période du cycle.

cinématographie à vitesse normale et ultra-rapide : la glotte est éclairée en continu et des images sont enregistrées à l'aide d'une caméra tournant à la vitesse de 25 à 50 images par seconde (cinématographie à vitesse normale). A une fréquence fondamentale de 100 Hz, cela signifie qu'un cliché de la glotte est pris tous les 4 à 2 cycles glottiques. Cette méthode permet donc de filmer le mouvement glottique, mais elle ne permet pas d'appréhender les détails d'un cycle glottique. Mise au point au *Bell Telephone Laboratories* [20] en 1957, la cinématographie ultra-rapide permet d'enregistrer des images de la glotte à des vitesses nettement plus élevées, de 4000 à 10000 images par seconde, ce qui fournit de 40 à 100 images par cycle glottique pour une fréquence fondamentale de 100 Hz. Cette méthode, bien qu'actuellement encore très lourde à appliquer et très coûteuse, permet de visualiser des détails du mouvement glottique qui passeraient inaperçus avec des méthodes plus simples et plus classiques, comme les phénomènes apériodiques (en particulier les transitoires).

vidéokymographie : cette méthode renseigne sur le mouvement d'un point particulier de la glotte au cours du temps. Exploitant les propriétés d'affichage par lignes horizontales d'un écran de télévision, elle permet, par sélection d'une unique ligne du plan de la glotte, d'accroître la vitesse de visualisation du mouvement de cette ligne. Une caméra standard, qui enregistre 50 images glottiques par seconde, peut alors enregistrer quelques 8000 lignes par seconde, ce qui permet, à $f_0 = 100\ Hz$, de visualiser, à un point donné de la glotte, 80 instants d'un même cycle glottique. Si cette méthode permet l'observation à moindre coût des apériodicités, elle est limitée par la perte de la vue d'ensemble de la glotte.

Pour toutes ces méthodes, une source lumineuse éclaire les cordes vocales par le dessus, au moyen d'un miroir laryngé (source lumineuse externe), d'un endoscope rigide (endoscopie) ou souple (fibroscopie, introduction par les cavités nasales) (source lumineuse interne, véhiculée par fibre optique). Actuellement, la fibroscopie est la seule méthode de visualisation directe qui puisse s'appliquer à toute forme de production vocale, les autres méthodes ne permettant que la production de certaines voyelles, en raison de l'introduction d'une tige rigide dans la partie supérieure du pharynx. Ces méthodes permettent la visualisation du mouvement des cordes vocales dans un plan horizontal (fermeture et ouverture sur la longueur des cordes). Le mouvement vertical des cordes vocales (fermeture et ouverture sur l'épaisseur) peut être

visualisé à l'aide de la radiographie par rayons X et de l'imagerie par résonance magnétique.

Méthodes indirectes d'exploration

La glottographie

La glottographie peut être optique (photoglottographie) ou électrique (électroglottographie). Dans le cas de la photoglottographie, une source lumineuse est appliquée à l'extérieur, sur le cou et en-dessous de la glotte, et un capteur photosensible placé à l'intérieur du conduit vocal, en haut du larynx, mesure l'intensité lumineuse qui traverse la glotte au cours de la phonation. Le signal ainsi obtenu peut être généralement relié à l'aire glottique. Cette méthode est facile à mettre en oeuvre, mais elle est relativement invasive, comme toutes les méthodes qui nécessitent l'introduction d'un capteur dans la cavité buccale. Dans le cas de l'électroglottographie, deux électrodes appliqués sur le cou au niveau de la glotte délivrent un courant de faible intensité et mesurent les variations d'impédance engendrées par le mouvement vibratoire des cordes vocales. Très simple à mettre en oeuvre, cette méthode est non-invasive, portable et peu coûteuse. Nous y reviendrons très en détail dans le chapitre 5.

Mesures de débit et de pression

La mesure du débit s'effectue en sortie des lèvres, grâce à un masque appliqué sur le visage, le masque de Rothenberg (Rothenberg [170], 1973). La mesure de pression peut s'effectuer à l'aide d'un capteur (microphone de pression) placé en sortie des lèvres (signal acoustique), à l'intérieur de la cavité buccale (pression intra-orale) ou en-dessous des cordes vocales (pression sous-glottique, mesurée par ponction trachéale).

Dans le cas des mesures extra-orales, le signal obtenu renseigne sur les propriétés acoustiques du son émis (intensité, fréquence fondamentale, contenu spectral). Il permet également d'estimer l'onde de débit glottique par application de techniques de filtrage permettant d'annuler (ou à défaut d'atténuer) les effets du conduit vocal. Nous reviendrons en détail sur ces techniques en partie 4.2 du chapitre 4.

1.4 Conclusion

Les caractéristiques physiologiques de la source glottique ont été décrites dans ce chapitre. Les actions musculaires mises en jeu ont été évoquées. Elles soulignent la complexité du mouvement vibratoire glottique, car plusieurs muscles peuvent agir simultanément sur les cordes vocales, parfois de façon antagoniste. Ces actions musculaires, associées à l'action de la pression sous-glottique, sont des facteurs majeurs de la vibration laryngée, dont le rôle n'est toujours pas totalement compris. Il n'est évidemment pas simple de les caractériser par des mesures directes, à moins d'appliquer des méthodes expérimentales très invasives, à l'exemple de l'électromyographie. L'approche expérimentale, que nous aborderons au chapitre 4 de la partie I et en partie II de cette étude, se veut non-invasive, laissant au sujet étudié la liberté de son expression vocale. Nous appliquerons donc deux méthodes indirectes et totalement non-invasives : la mesure du signal acoustique en sortie des lèvres et l'électroglottographie.

Encore à l'heure actuelle, une grande confusion règne à propos des mécanismes laryngés. De plus, de nombreuses études ne prennent pas soin de déterminer le mécanisme laryngé dans lequel a lieu la production vocale étudiée. Pourtant, cette information est fondamentale, puisque la configuration glottique est très différente selon le mécanisme laryngé considéré. Nous définirons donc toujours précisément, quand cela sera possible, le mécanisme laryngé sous-jacent et nous mettrons en avant les aspects qui distinguent un mécanisme d'un autre.

Chapitre 2

Modélisation et paramètres de source glottique

L'approche d'un phénomène physique passe par l'élaboration de modèles qui vont essayer de rendre compte de ce phénomène et de le reproduire fidèlement. Le modèle à la base de la production vocale humaine est un modèle linéaire source-filtre que nous allons maintenant présenter. Il met en jeu un ensemble de paramètres que nous préciserons.

2.1 Le modèle de signal source-filtre

2.1.1 Présentation

Exposée par Fant [64] en 1960, la théorie source-filtre propose de modéliser l'acoustique de la production vocale par un système linéaire et invariant dans le temps. La linéarité se traduit par le principe de superposition : la réponse du système à une somme d'entrées est la somme des réponses du système à chacune de ces entrées. L'invariance dans le temps se traduit par le fait que la réponse du système à une entrée décalée dans le temps est décalée dans le temps de la même façon.

Le modèle de signal source-filtre est illustré sur la figure 2.1 dans le cas d'un son voisé. Il se décompose en trois parties supposées linéairement indépendantes, comme suit :

la source de voisement $u_g(t)$ ou $U_g(f)$: sous la force de la pression sous-glottique exercée par l'air contenu dans les poumons, les cordes vocales se séparent, puis se rejoignent sous l'effet d'une force de rappel et par effet Bernoulli du fait du gradient entre pression sus- et sous-glottique. Ce mouvement d'ouverture-fermeture génère une onde acoustique qui se propage à travers le conduit vocal (larynx, pharynx, cavités buccales et nasales).

la fonction de transfert du conduit vocal $h(t)$ ou $H(f)$: le conduit vocal est composé du pharynx, des cavités buccales et nasales. Du point de vue acoustique, il constitue un tube non-uniforme et peut être modélisé par l'accolement de tubes cylindriques de diamètre variable. Des ondes planes stationnaires vont alors s'établir dans ce conduit, générant des résonances. Certaines fréquences sont accentuées, ce qui se traduit sur la fonction de transfert du conduit vocal par des maxima locaux. Ces maxima sont appelées les formants du conduit vocal, et leur position fréquentielle détermine en particulier la voyelle produite.
En première approximation, le conduit vocal peut être assimilé à un tube acoustique uniforme, fermé à un bout (glotte) et ouvert à l'autre bout (lèvres). Les fréquences propres des ondes stationnaires

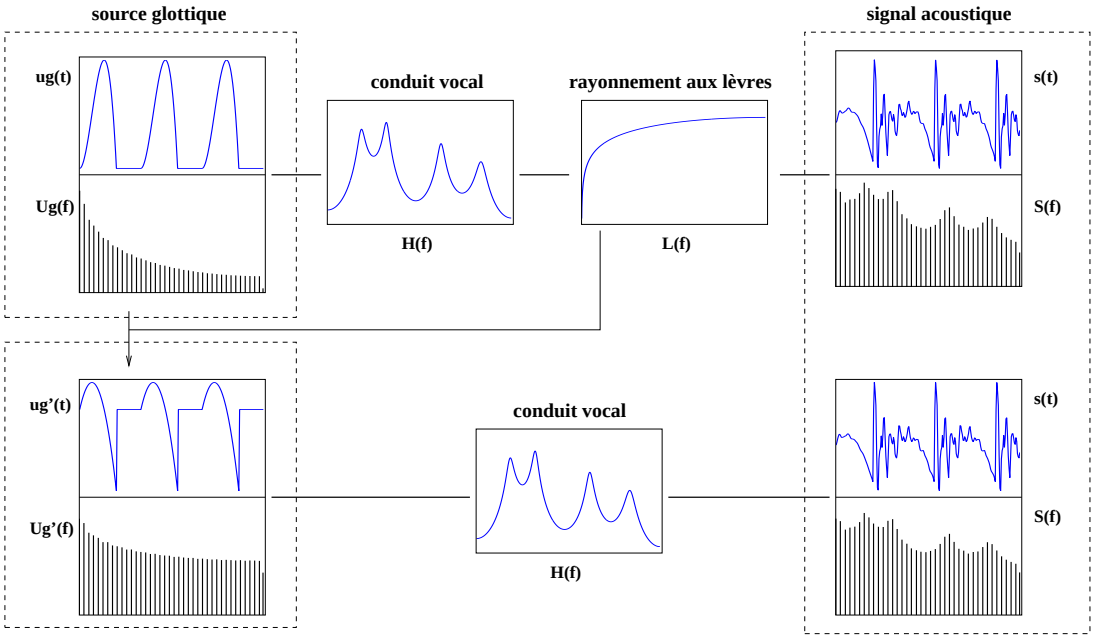


FIG. *2.1 – Illustration du modèle de signal source-filtre : filtrage de la source de voisement par la fonction de transfert du conduit vocal et dérivation par rayonnement aux lèvres.*

dans un tel tube sont données par :

$$F_n = \frac{(2n-1)c}{4L} \qquad pour\ n\ =\ 1, 2, 3, \ldots$$

avec c la vitesse du son dans l'air ($c \simeq 340\ m.s^{-1}$) et L la longueur du tube ($L \simeq 17\ cm$ pour un homme). En remplaçant c et L par leurs valeurs approximatives, on obtient : $F_n \simeq (2n-1) * 500$ Hz. A partir de ce modèle simpliste du conduit vocal, on observe un formant tous les 1000 Hz, à la fréquence F_n (500 Hz, 1500 Hz, 2500 Hz, ...).

le rayonnement aux lèvres $l(t)$ ou $L(f)$: au passage des lèvres, les ondes acoustiques planes se transforment en ondes sphériques et se propagent dans le milieu environnant. Du point de vue du signal, ce rayonnement équivaut à une dérivation du débit (Flanagan [73], 1965).

Le signal acoustique $S(f)$ en sortie des lèvres correspond alors à la convolution temporelle ou multiplication fréquentielle de ces trois composantes :

$$\begin{aligned} s(t) &= u_g(t) * h(t) * l(t) \\ S(f) &= U_g(f).H(f).L(f) \end{aligned} \tag{2.1}$$

Comme le système est linéaire, nous pouvons intervertir les termes. Puisque l'effet du rayonnement aux lèvres correspond à une dérivation, on peut réécrire l'équation 2.1 sous la forme :

$$\begin{aligned} s(t) &= u_g^{'}(t) * h(t) \\ S(f) &= U_g^{'}(f).H(f) \end{aligned} \tag{2.2}$$

avec $u_g^{'}$ la dérivée temporelle de u_g, $U_g^{'}$ et U_g leurs transformées de Fourier respectives.

2.1.2 Les limites

La théorie source-filtre a eu un impact considérable parce qu'appuyée expérimentalement (Fujimura & Lindqvist [82]) et appliquée avec succès en analyse (analyse cepstrale (Noll [158]), en codage par prédiction linéaire (Makhoul [150], Markel & Gray [151])), en synthèse (synthèse par formants (Klatt [133], Rodet [168]), synthèse à partir du texte (Dutoit [58])). Comme toute modélisation, elle a des limites dont il faut être conscient.

En particulier, elle se base sur l'approximation d'une propagation plane des ondes acoustiques dans le conduit vocal, ce qui est vérifié pour des diamètres $d << \lambda$, avec $\lambda = \frac{c}{f}$. Dans le cas d'un diamètre de 6 cm, l'approximation des ondes planes est valide pour des fréquences $f << \frac{c}{d} \simeq 6\ k$ Hz. Or les chanteurs présentent de l'énergie jusqu'à des fréquences très élevées, pouvant largement dépasser 6 kHz. Il existe également des interactions possibles entre la source et le filtre (Fant [68], 1993), comme par exemple, l'influence du conduit supraglottique ou l'influence des ondes acoustiques stationnaires dans le conduit vocal sur la pression transglottique. C'est en particulier le cas quand le premier formant F_1 coïncide avec la fréquence fondamentale f_0 (Schutte & Miller [182], 1986 ; Fant [67], 1986).

2.1.3 Modélisation de la production vocale

L'étude de la production vocale nécessite d'élaborer des modèles qui reproduisent au mieux les sons naturels. Dans le cas des sons voisés, le conduit vocal se modélise de façon très convenable par un filtre tout-pôle, auquel on peut rajouter des zéros en cas de nasalité. La modélisation de la source glottique est beaucoup plus délicate, du fait de la complexité du mouvement vibratoire glottique et d'une compréhension encore réduite de ses caractéristiques. Les approches sont de deux types : modélisation interactive ou non-interactive (Ljungqvist [146]). La modélisation interactive se base sur des aspects physiques du mouvement vibratoire et prend ainsi en compte les interactions avec le conduit vocal. Le débit glottique est alors calculé à partir de l'aire glottique (Guerin & al [91], 1976 ; Ananthapadmanabha & Fant [10], 1982 ; Titze [200], 1984 ; Allen & Strong [8], 1985 ; Titze [201], 1989) ou par modélisation physique des cordes vocales à l'aide de systèmes masse - ressort (Ishizaka & Flanagan [116, 74], 1972 ; Titze [199], 1973 ; Lous & al [147], 1998). Cette approche est limitée par la connaissance que nous avons de la physique du mouvement vibratoire des cordes vocales, ce qui se traduit dans ces modèles par l'utilisation de paramètres (pression sous-glottique, masse des cordes) dont les valeurs peuvent être parfois très éloignées de toute réalité physique. La modélisation non-interactive s'attache à décrire le signal lui-même, en l'approximant par des fonctions paramétriques empiriques qui ne reflètent pas forcément une réalité physique. Cette approche suppose une indépendance entre la source et le filtre. Néanmoins, certains aspects de l'interaction source-filtre sont pris en compte indirectement dans ces modèles. Par exemple, le couplage entre les cavités sus- et sous-glottique se traduit par une accentuation de l'asymétrie de l'impulsion (Ananthapadmanabha & Fant [10, 9]).

Dans cette étude, nous avons choisi l'approche non-interactive car nous nous intéressons particulièrement à la modélisation et à l'estimation du signal associé au débit glottique u_g et à sa dérivée $u_g^{'}$. La fonction qui décrit ce débit glottique est appelée conventionnellement Onde de Débit Glottique (ODG), bien qu'il ne s'agisse pas ici d'une onde au sens physique du terme (mouvement propagatif), mais d'un flux d'air qui traverse la surface glottique au cours du temps.

Nous allons à présent décrire plus en détail les paramètres de source glottique et la terminologie que nous serons amenés à employer au cours de cette étude.

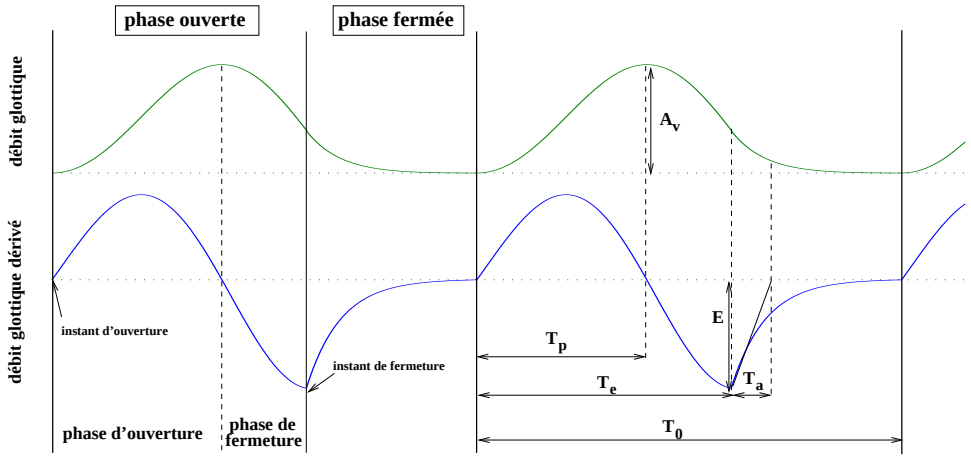


FIG. *2.2 – Illustration de la terminologie utilisée pour décrire la source glottique, visualisée sur le débit glottique et sur sa dérivée.* T_0 *est la période fondamentale d'une impulsion glottique,* T_e *la durée de la phase ouverte,* T_p *la durée de la phase d'ouverture et* T_a *la durée effective de la phase de retour. Le quotient ouvert est défini par* $O_q = \frac{T_e}{T_0}$ *et le quotient de vitesse par* $S_q = \frac{T_p}{T_e - T_p}$. A_v *est l'amplitude de voisement et* E *la vitesse de fermeture.*

2.2 Les paramètres de source glottique

2.2.1 Définition

Les impulsions glottiques sont souvent décrites à l'aide d'une terminologie qui se réfère à des aspects physiologiques de la phonation (i.e. : au mouvement vibratoire des cordes vocales) : ouverture et fermeture glottique, phase d'ouverture, phase ouverte, phase de fermeture, phase fermée, abduction glottique (Ananthapadmanabha [9], 1984). Selon le contexte de l'étude, ces termes ne représenteront pas la même chose, en raison du biais entre des aspects physiologiques (tels la séparation ou l'accolement des cordes vocales, l'amplitude de l'ouverture glottique, ...) et des aspects physiques (variations du débit glottique, de la pression sous-glottique, ...). Ainsi par exemple, l'instant relié au maximum du débit glottique sur une période ne correspond pas à l'instant relié au maximum de l'aire glottique. Il semble donc important de définir précisément les termes que nous serons amenés à employer couramment dans cette étude. Ces termes explicitent des paramètres qui font sens d'un point de vue acoustique. Ils sont illustrés sur la figure 2.2.

Instants d'ouverture et de fermeture glottique :

Quand la glotte est fermée, le débit est soit nul, soit constant (cas d'une fuite d'air permanente au niveau des aryténoïdes). L'instant d'ouverture glottique est défini par l'instant d'augmentation du débit glottique par rapport à sa valeur minimale. Physiologiquement, cela correspond à l'instant où les cordes vocales commencent à se séparer sur leur partie supérieure. L'instant de fermeture glottique est associé au maximum de l'excitation glottique, c'est-à-dire l'instant où la dérivée du débit glottique atteint son minimum. Physiologiquement, cela correspond à l'instant où les cordes vocales s'accolent sur leur partie inférieure. On utilisera le terme de fermeture abrupte si la dérivée du débit glottique est discontinue à l'instant de fermeture glottique.

Phase d'ouverture, de fermeture, phase ouverte, phase fermée, phase de retour :

Ces différentes phases du mouvement glottique sont définies à partir de la connaissance des instants d'ouverture et de fermeture détectés sur le débit glottique, ainsi que l'instant du maximum de débit. La phase d'ouverture est définie entre l'instant d'ouverture glottique et l'instant du maximum de débit. La

phase de fermeture est définie entre l'instant du maximum de débit et l'instant de fermeture glottique. Ces deux phases constituent la phase ouverte, définie entre les instants d'ouverture et de fermeture glottique. A l'inverse, la phase fermée est définie entre les instants de fermeture et d'ouverture glottique.

Il est bon de noter que la phase fermée n'est pas forcément associée avec un arrêt du débit. Ainsi, dans le cas d'une fermeture non abrupte, comme c'est le cas sur la figure 2.2, le débit diminue graduellement après l'instant de fermeture glottique. On définit alors la phase de retour entre l'instant du maximum de l'excitation et l'instant où le débit a atteint sa valeur minimale.

Quotient ouvert, quotient de vitesse, durée effective de la phase de retour, amplitude de voisement, vitesse de fermeture

Il existe une multitude de paramètres indépendants pour décrire la source glottique. Ces paramètres peuvent être dimensionnés (durée d'ouverture, de fermeture, durée effective de la phase de retour, composante continue du flux glottique, amplitude de voisement, ...) ou adimensionnés (quotient ouvert, quotient de vitesse, ...). Dans cette étude, nous privilégierons les paramètres adimensionnés, puisqu'ils permettent une étude de la source glottique indépendamment des variations de période fondamentale. Nous nous intéressons en particulier aux 5 paramètres suivants :

1. **le quotient ouvert** O_q : il est défini comme le rapport entre la durée de la phase ouverte et la période fondamentale. En théorie, il peut prendre des valeurs entre 0 (pas d'ouverture) et 1 (pas de fermeture ou une fermeture incomplète). En pratique, il varie entre 0.3 et 0.98. Dans la littérature, on retrouve également un équivalent, le quotient fermé, défini par $Cq = 1 - O_q$.
2. **le quotient de vitesse** S_q : il est défini comme le rapport entre la durée de la phase d'ouverture et la durée de la phase de fermeture. On peut également retrouver dans la littérature la définition inverse. Il indique une asymétrie plus ou moins marquée de l'impulsion glottique. Comme la phase d'ouverture a une durée égale ou plus longue que la phase de fermeture, S_q est compris entre 1 (impulsion glottique symétrique) et 9 (impulsion glottique très asymétrique). En pratique, les valeurs usuelles sont situées entre 2 et 3.
3. **la durée effective de la phase de retour** T_a : ce paramètre correspond au temps entre l'instant du maximum d'excitation et l'instant où le débit est quasiment revenu à sa valeur minimale. Il s'agit d'un paramètre d'importance puisqu'il a pour effet de modifier la pente du spectre de la source dans les hautes fréquences. Il est nul dans le cas d'une fermeture abrupte.
4. **l'amplitude de voisement** A_v : ce paramètre est défini comme l'amplitude entre la valeur minimale et la valeur maximale du débit glottique. Il s'exprime en $\mathrm{l.s}^{-1}$.
5. **la vitesse de fermeture** E : ce paramètre est défini comme l'amplitude du minimum de la dérivée du débit glottique et correspond à la vitesse du débit glottique à l'instant de fermeture. Il s'exprime en $l.s^{-2}$.

Décomposition périodique / apériodique

Les apériodicités qui se retrouvent dans un signal acoustique contribuent à la qualité vocale (Klatt & Klatt [132], Childers & Lee [35]). Dans le cas des sons voisés, ces apériodicités peuvent provenir d'un bruit ajouté à la source glottique, comme, par exemple, en raison d'une fuite glottique, ou être générées par de petites variations de la fréquence fondamentale (jitter) ou de l'amplitude (shimmer). Un signal acoustique peut se décomposer alors en une partie périodique (due à la source glottique) et une partie apériodique (due au bruit et à la source glottique dans le cas des apériodicités structurelles) (d'Alessandro & al [47, 44, 45], Darsinos & al [48], Yegnanarayana & al [218]). On peut définir le rapport périodique/apériodique comme le rapport entre l'énergie de la partie périodique et celle de la partie apériodique d'un signal acoustique. Dans toute cette étude, nous laisserons ces aspects de côté et nous nous centrerons sur l'étude de la source glottique périodique. Dans le chant classique, la présence d'apériodicité est considérée comme un défaut, ce qui n'est pas le cas dans d'autres styles musicaux.

Nous allons à présent détailler les liens mis en évidence dans la littérature entre les paramètres de source glottique et la qualité vocale.

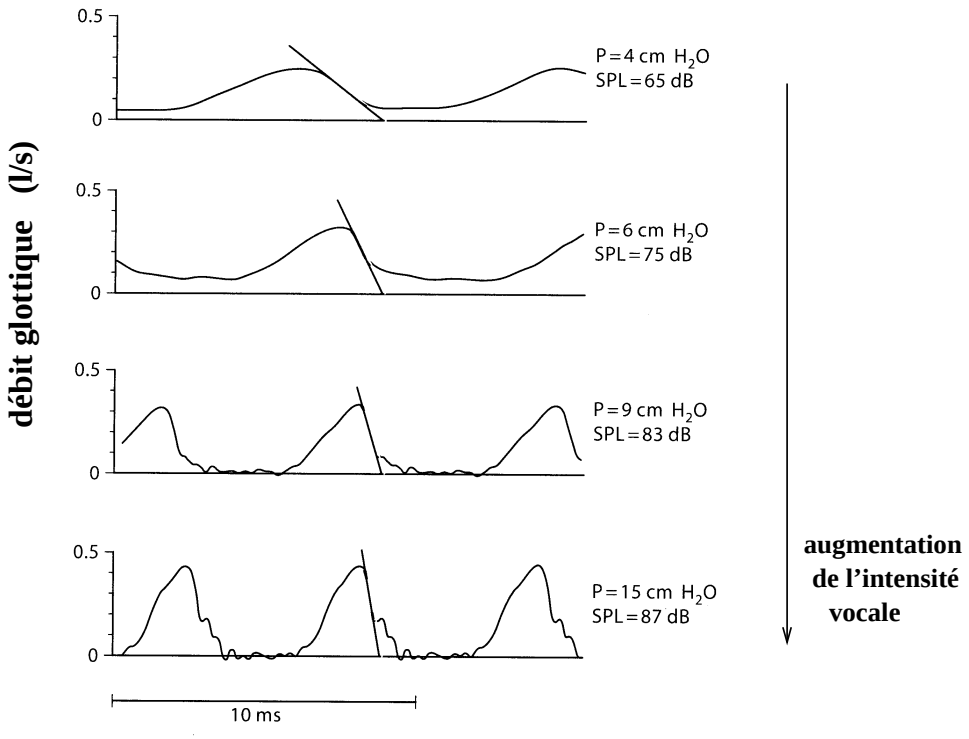


FIG. *2.3 – Illustration des variations de débit glottique en fonction d'une augmentation progressive de l'intensité vocale. D'après Gauffin & Sundberg* [85]

2.2.2 Influence sur la qualité vocale

Intensité vocale

La figure 2.3 illustre les variations du débit glottique estimé par filtrage inverse pour différentes intensités vocales dans le cas d'un locuteur non chanteur. Elle est extraite d'une étude sur les corrélats spectraux des paramètres de source glottique, réalisée par Gauffin & Sundberg [85]. Sur cet exemple, la pression sous-glottique notée P passe de 4 à 15 cm H_2O quand l'intensité vocale notée SPL passe de 65 dB à 87 dB. Cette observation est généralisable : une augmentation d'intensité vocale résulte généralement d'une augmentation de la pression sous-glottique (Ladefoged & McKinney [135], 1963 ; Isshiki [117], 1964 ; Schutte [181], 1981 ; Tanaka & Gould [195], 1983 ; Karlsson [123], 1986 ; Holmberg & al [109], 1988 ; Gauffin & Sundberg [85], 1989 ; Titze & Sundberg [209], 1992 ; Sundberg & al [193], 1993 ; Lecuit & Demolin [140, 141], 1998).

En plus de la pression sous-glottique, les paramètres de source glottique qui varient lors d'un changement d'intensité vocale sont :

1. **La fréquence fondamentale**

   L'intensité varie souvent de pair avec la fréquence fondamentale (Gramming & al [89], 1988 ; Titze [209], 1992 ; Liénard & Di Benedetto [144], 1999). En voix chantée, elle peut néanmoins en être découplée, comme par exemple lors d'un son filé sur une note tenue.

2. **La vitesse de fermeture**
   La vitesse de fermeture E est représentée indirectement sur la figure 2.3 par la pente de la tangente au débit glottique pendant la phase de fermeture. On note que cette pente augmente avec l'intensité. De nombreuses études ont mis en évidence une corrélation forte entre la vitesse de fermeture et l'intensité vocale (Fant [66], 1982 ; Holmberg & al [109], 1988 ; Gauffin & Sundberg [85], 1989 ; Sundberg & al [193], 1993).

3. **L'amplitude de voisement**
   L'amplitude de voisement A_v semble varier conjointement à la vitesse de fermeture dans le cas des chanteurs (Sundberg & al [193], 1993). C'est également ce que l'on observe sur la figure 2.3.

4. **Le quotient ouvert**
   En voix chantée, Sundberg & al [193] notent que, quand l'amplitude de voisement est faible, l'augmentation de l'intensité vocale peut se faire par diminution du quotient ouvert. Une fois que le quotient ouvert a atteint un niveau donné, de l'ordre de 0.4, il se stabilise et l'augmentation de l'intensité vocale se fait alors par augmentation de l'amplitude de voisement. Une forte corrélation entre quotient ouvert et intensité vocale a été observée par Timcke & al [197], 1958 ; Holmberg & al [109], 1988 ; Orlikoff [161], 1991 ; Dromey & al [57], 1992 ; Hacki [93], 1996).

5. **Le quotient de vitesse**
   En voix chantée, Sundberg & al [193] notent également qu'une augmentation d'intensité peut résulter d'une accentuation de l'asymétrie, donc d'une modification du quotient de vitesse. Une augmentation du quotient de vitesse avec l'intensité est également notée par Timcke & al [197], Holmberg & al [109], Dromey & al [57]. D'un autre côté, le quotient de vitesse semble également dépendre de la fréquence fondamentale : il tend à diminuer quand la fréquence fondamentale augmente, ce qui amène à des impulsions glottiques plus symétriques à haute fréquence (Sundberg & al [193]).

Timbre

A côté des variations de qualité induite par les changements d'intensité vocale, on observe également des variations de timbre qui peuvent être le fait de modification des cavités buccales et nasales ou de variations des paramètres de source glottique.

La figure 2.4 présente le débit glottique pour différents types de qualité vocale observés classiquement en production vocale pathologique et non-pathologique. Sundberg [188] (1995) a avancé l'hypothèse que ces types de phonation sont contrôlés par la force de fermeture glottique (compression plus ou moins forte des cordes vocales entre elles). Nous allons détailler les caractéristiques de chacune de ces productions en fonction des paramètres de source.

1. **voix pressée, tendue (*pressed voice*)**
   Ce type de phonation est encore appelé phonation hyperfonctionnelle. Elle traduit une compression accentuée des aryténoïdes l'un contre l'autre (Titze [205], 1995) et une augmentation forte de la pression sous-glottique (Gauffin & Sundberg [85]). On observe effectivement sur la figure 2.4 que la pression sous-glottique est plus élevée que dans les autres types de phonation : $P = 14$ cm H_2O. Comme la pression sous-glottique est importante et que les cordes vocales sont fortement comprimées l'une contre l'autre, l'amplitude de voisement est très réduite, de l'ordre de 0.2 $l.s^{-1}$ sur la figure 2.4. Dans ce type de phonation, le quotient ouvert est généralement très faible puisque la glotte se referme très rapidement. Le quotient de vitesse est par contre généralement peu élevé, traduisant une impulsion relativement symétrique.

2. **voix normale**
   C'est la phonation habituelle dans le cas des voix non-pathologiques. Le quotient ouvert est de

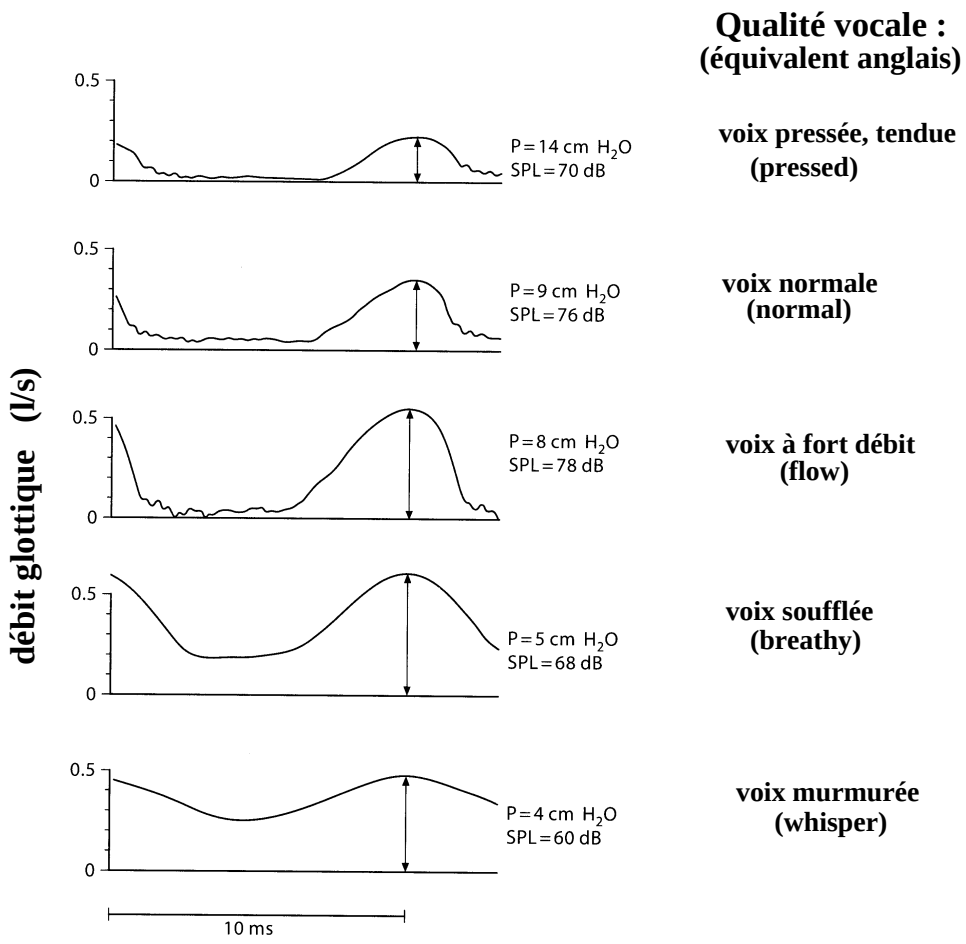


FIG. *2.4 – Illustration du débit glottique pour différentes qualités vocales : voix pressée, normale, à fort débit, soufflée et murmurée. Ces termes sont des traductions des termes anglais courants, précisés entre parenthèses sur la figure. D'après Gauffin & Sundberg* [85]

l'ordre de 0.5, le quotient de vitesse proche de 2.

3. **voix à fort débit** (*flow voice*)
   Cette phonation a été mise en avant par Gauffin & Sundberg [85]. Elle se caractérise par une amplitude de voisement importante, un quotient ouvert plus élevé qu'en phonation normale et un quotient de vitesse comparable à la phonation normale. Selon ces auteurs, cette phonation semble optimale car elle permet de produire des intensités élevées avec un minimum de pression sous-glottique. Ceci se fait bien évidemment au détriment de la consommation d'air, qui s'accroît alors.

4. **voix soufflée** (*breathy voice*)
   Pour ce type de phonation, le débit minimal n'est plus nul. Ceci traduit une fermeture incomplète de la glotte, vraisemblablement au niveau des aryténoïdes. L'amplitude de voisement est comparable à la voix normale, mais le quotient ouvert augmente sensiblement, tandis que le quotient de vitesse diminue, traduisant une impulsion glottique plus symétrique. Dans les exemples que nous avons commentés jusqu'à présent, la phase de fermeture pouvait être considérée comme abrupte. Dans ce cas-ci, la fermeture est beaucoup plus douce, ce qui accentue encore le caractère symétrique de l'impulsion. Le rapport signal/bruit devient très faible.

5. **voix murmurée** (*whisper voice*)
   L'onde acoustique est générée par des turbulences du débit glottique et non plus par un mouvement vibratoire des cordes vocales (Titze [205]). Sur la figure 2.4, la voix murmurée est légèrement voisée (Gauffin & Sundberg [85]), ce qui explique la présence d'une période fondamentale.

6. **voix relâchée** (*relaxed voice*)
   Cette phonation, encore appelée hypofonctionnelle, est le contraire d'une phonation pressée. Elle n'est pas présentée sur la figure 2.4. Dans ce type de phonation, l'accolement des cordes est mou, ce qui se traduit par des valeurs de quotient ouvert élevées et de quotient de vitesse basses. La phase de retour devient importante.

La position du larynx varie selon le type de phonation. Ainsi, une voix pressée sera généralement produite avec un larynx en position haute, tandis que la phonation à fort débit sera produite avec un larynx en position basse (Sundberg [188]).

Les voix relâchées et soufflées se caractérisent spectralement par un premier harmonique (fréquence fondamentale) d'amplitude plus importante que les harmoniques élevés. Au contraire, les voix pressées présentent un premier harmonique fortement atténué par rapport aux harmoniques d'ordre plus élevé (Gauffin & Sundberg [85], Klatt & Klatt [132]).

Paramètre R_d

Fant [69] (1995) a effectué une analyse statistique des covariations des paramètres du modèle LF (Fant & al [72], 1985) à partir de mesures effectuées par Gobl [88] dans le cas de locuteurs et Karlsson [124] dans le cas de locutrices. Les phonations explorées vont des voix très pressées, avec des valeurs de quotient ouvert très basses et des phases de retour très brèves, aux voix soufflées, avec des valeurs de quotient ouvert très élevées et des phases de retour plus longues. Cette analyse a fait ressortir un paramètre unique qui semble permettre d'appréhender à lui seul les variations entre ces diverses productions vocales. Ce paramètre R_d est défini par :

$$R_d = \frac{A_v}{E}\,\frac{f_0}{110} \tag{2.3}$$

Le facteur 110 est une constante sans dimension. A partir de ce paramètre, Fant a alors exprimé les relations entre les différents paramètres de son modèle. En particulier, il en a déduit une relation liant la différence en amplitude entre les deux premiers harmoniques du spectre de source glottique et le quotient ouvert. Nous reviendrons en détail sur cette relation dans le chapitre 4.

2.3 Conclusion

Nous avons présenté ici le modèle source-filtre et défini avec précision la terminologie que nous serons amenés à employer au cours de cette étude. Nous avons caractérisé de manière très générale des variations d'intensité vocale et de timbre à l'aide d'un petit nombre de paramètres de source glottique. La détermination du lien entre les paramètres de source glottique et la qualité du son produit reste néanmoins une question ouverte. Quels sont les paramètres les plus pertinents pour décrire la source glottique ? En particulier, existe-t-il un paramètre unique, comme le suggère Fant [69], qui puisse permettre d'appréhender à lui seul certaines qualités vocales ? Ces paramètres évoluent-ils de façon similaire en parole et en chant ? Observe-t-on des différences marquées selon le mécanisme laryngé utilisé ? Quels sont les corrélats spectraux de ces paramètres ? Comment perçoit-on leurs variations ? Ce sont autant de questions auxquelles nous espérons que cette étude apportera des éléments de réponse.

Première partie

Analyse temporelle et spectrale de la source glottique

La voix est formée uniquement par les compressions et les dilatations périodiques que l'air éprouve lorsque, à sa sortie de la glotte, celle-ci, par une action alternative et régulière, l'arrête et lui livre passage.

(Manuel Garcia, « Traité complet de l'Art du Chant », 1884)

Sommaire de la partie

Cette partie aborde l'étude de la source glottique sous un aspect d'analyse et de synthèse du signal d'Onde de Débit Glottique.

Les principaux paramètres de source glottique qui influencent la qualité vocale ont été évoqués en partie préliminaire. Pour pouvoir comprendre la façon dont ces paramètres peuvent modifier une qualité vocale donnée, il nous importe à présent de déterminer leurs effets spectraux, ce qui sera l'objet du chapitre 3. Nous prendrons appui sur le formalisme mis en place par Doval & d'Alessandro [55] en 1999, qui permet d'exprimer les modèles d'ODG de façon temporelle et spectrale à l'aide d'un jeu unique de paramètres. Les effets spectraux des paramètres de source glottique pourront alors être étudiés de façon analytique et seront illustrés par des exemples de synthèse.

Si l'on connaît les effets spectraux des paramètres de source glottique, ne peut-on pas, par une démarche inverse, estimer ces paramètres à partir de mesures spectrales de la source glottique ? L'idée semble d'autant plus séduisante que des mesures expérimentales ont montré une forte corrélation entre la différence spectrale entre les deux premiers harmoniques du spectre de source glottique, $(H_1 - H_2)$, et le quotient ouvert. Dans le chapitre 4, nous discuterons alors de la possibilité d'estimer spectralement les paramètres de source glottique, en nous centrant sur cette différence spectrale. Nous étudierons le lien entre $(H_1 - H_2)$ et les paramètres de source glottique de façon à la fois théorique et expérimentale.

Cette étude nous amènera à réfléchir à l'application et aux limites des modèles d'ODG, ainsi qu'à la validité de la théorie source-filtre dans le cas du chant.

Chapitre 3

Modélisation de la source glottique

Nous venons de voir l'importance des modèles de signaux pour la compréhension et l'analyse de la production vocale. Dans le cadre de la théorie source-filtre, la source glottique peut être décrite par des modèles paramétriques. Si les modèles proposés dans la littérature génèrent des signaux d'ODG similaires, ils ne partagent ni les mêmes paramètres, ni le même nombre de paramètres. La question se pose alors de savoir si ces modèles ne pourraient pas tous être décrits à l'aide d'un unique jeu de paramètres. Il serait également souhaitable que ces paramètres ne soient pas abstraits mais puissent être reliés à des aspects physiologiques du mouvement vibratoire des cordes vocales. Dans cet esprit, un travail d'unification des modèles d'ODG a été entrepris par Doval & d'Alessandro [54, 56, 55]. Se plaçant dans le cas d'une fermeture abrupte, ils ont établi un formalisme permettant d'exprimer les modèles temporels d'ODG en fonction d'un unique jeu de paramètres glottiques. Ce formalisme, que nous exposerons en partie 3.1, permet d'exprimer de façon simple et unique la variation temporelle de la source glottique indépendamment des modèles considérés. Par application de la transformée de Fourier, il fournit également les équations spectrales associées à l'ODG et à sa dérivée. Une étude analytique des corrélats spectraux des paramètres de source glottique est alors possible, dans le but de comprendre les attributs perceptifs associés à chacun des paramètres de source glottique considérés. Nous présenterons donc en partie 3.2 les effets spectraux de ces paramètres, que nous illustrerons par des exemples de synthèse.

3.1 Unification des modèles temporels de source glottique et formalisme

La majorité des modèles de source glottique présentés dans la littérature et appliqués pour l'analyse et la synthèse de la parole sont des modèles temporels. Des modèles très simples ont été proposés par Rosenberg [169] en 1971. Ils ont ensuite été repris et améliorés par Hedelin [100], Klatt & Klatt [132] et Veldhuis [213]. D'autres modèles ont également été proposés par Fant [65], Liljencrants & Fant [72, 69], Fujisaki & Ljungqvist [146], Milenkovic [153] et Childers [32]. Tous ces modèles présentent une forme similaire :

1. l'ODG est toujours positive ou nulle. En particulier, elle est nulle ou constante quand la glotte est fermée.

2. l'ODG augmente avec l'ouverture glottique, atteint un maximum puis diminue avec la fermeture. Elle présente souvent une forme asymétrique car la vitesse de la fermeture est généralement plus rapide que celle de l'ouverture.

3. l'ODG est une fonction du temps périodique, continue et dérivable, sauf aux instants de fermeture et/ou d'ouverture glottique, où sa dérivée peut présenter parfois des discontinuités.

Néanmoins, les paramètres choisis pour décrire ces modèles diffèrent en nombre et en définition. De plus, ces modèles sont tous exprimés dans le domaine temporel. Notre oreille est très sensible à de petites variations temporelles, mais la forme spectrale d'un signal est souvent plus intéressante à étudier que sa forme temporelle car la description perceptive est plus proche du spectre.

Dans ce contexte, Doval et d'Alessandro [54, 55] (1999) ont effectué un travail d'unification basé sur 4 modèles d'ODG parmi les plus couramment utilisés dans la littérature : un des modèles élaborés par Rosenberg [169] dans les années 70 (modèle Rosenberg C, 1971), le modèle de Liljencrants & Fant [72] (modèle LF, 1985), actuellement le plus utilisé, le modèle de Klatt & Klatt [132] (modèle KLGLOTT88, 1990), récemment amélioré par Veldhuis [213] (modèle R++, 1998). Ils ont ainsi dégagé un jeu de paramètres communs à tous ces modèles et mis en place, dans le cas d'une fermeture abrupte, un formalisme d'écriture des fonctions temporelles et spectrales décrivant chacun de ces 4 modèles. Dans un premier temps, nous allons nous placer dans ce cadre-là, afin de décrire ce formalisme et de l'appliquer à l'étude spectrale de la source glottique. Nous étendrons ensuite cette étude au cas d'une fermeture non-abrupte.

3.1.1 Les paramètres

La fonction génératrice de chacun de ces 4 modèles peut se réécrire à partir d'un unique jeu des 5 paramètres suivants :

* $\mathbf{A_v}$: l'amplitude de voisement,
* $\mathbf{T_0}$: la période fondamentale de l'ODG ($T_0 = \frac{1}{f_0}$),
* $\mathbf{O_q}$: le quotient ouvert,
* $\alpha_\mathbf{m}$: le coefficient d'asymétrie, nombre sans dimension qui définit l'instant du maximum de l'ODG relativement à O_q et T_0. Il varie donc entre 0.5 (forme symétrique) et 1 (cas limite d'une forme très asymétrique). Ce paramètre de source glottique est l'équivalent du quotient de vitesse S_q évoqué en partie 2.2.1 : $\alpha_m = \frac{S_q}{1+S_q}$,
* $\mathbf{Q_a}$: le quotient de phase de retour. Ce paramètre a-dimensionné est l'équivalent de la durée effective de la phase de retour T_a : $Q_a = \frac{T_a}{1-O_q\,T_0}$. Dans le cas d'une fermeture abrupte, $Q_a = 0$.

La figure 3.1 illustre ce jeu de paramètres sur l'ODG et sur sa dérivée (ODGD). Dans le cas de la dérivée, il peut être intéressant de remplacer le paramètre A_v par le paramètre E (vitesse de fermeture). Nous reviendrons sur ce point en partie 3.2.

3.1.2 Expression générique dans le cas d'une fermeture abrupte

Nous ne rentrerons pas ici dans le détail des calculs mais donnerons simplement les résultats essentiels. Les fonctions génériques normalisées pour chaque modèle sont fournies dans l'annexe A. Pour plus de détails, se référer aux travaux de Doval & d'Alessandro [55].

Domaine temporel

Nous nous plaçons dans le cas d'une fermeture abrupte, ce qui suppose que la phase de retour est instantanée, donc $Q_a = 0$. Sous ces hypothèses, chaque modèle peut se réécrire sous la forme :

$$u_g(t) = A_v\, n_g(\frac{t}{O_q T_0}, \alpha_m) \;*\; \perp\!\!\!\perp\!\!\!\perp_{T_0}(t) \tag{3.1}$$

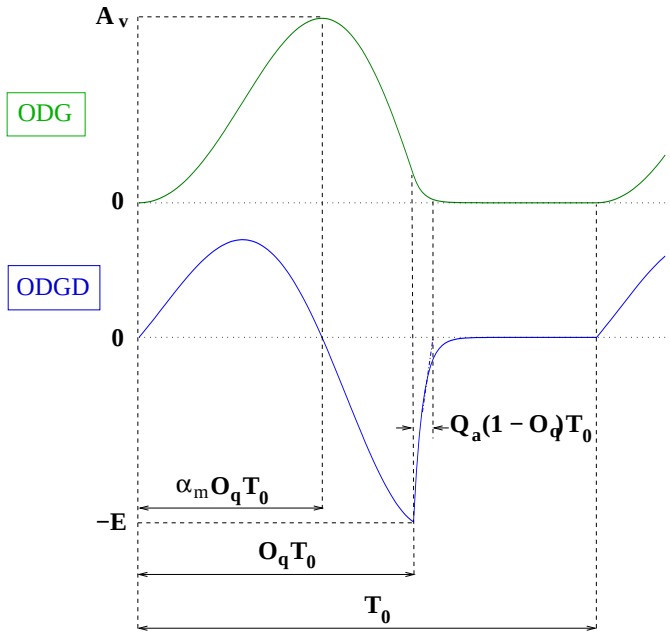


Fig. *3.1 – Représentation d'une période de l'ODG et de sa dérivée et définition des paramètres de source glottique : A_v amplitude de voisement, T_0 période fondamentale, O_q quotient ouvert, α_m coefficient d'asymétrie, Q_a quotient de phase de retour, E vitesse de fermeture.*

$$u_g^{'}(t) = \frac{A_v}{O_q T_0}\, n_g^{'}(\frac{t}{O_q T_0}, \alpha_m) \ * \ \perp\!\!\!\perp\!\!\!\perp_{T_0}(t) \tag{3.2}$$

où $\perp\!\!\!\perp\!\!\!\perp_{T_0}(t)$ représente un peigne de Dirac de période fondamentale T_0 et $n_g(x, \alpha_m)$ une fonction générique du modèle, qui ne dépend plus que du paramètre α_m. La normalisation par rapport à la durée de la phase ouverte $(O_q T_0)$ et à l'amplitude de voisement (A_v) est illustrée sur la figure 3.2. La vitesse de fermeture E se déduit de la vitesse de fermeture normalisée $e_n(\alpha_m)$ par :

$$E = \frac{A_v}{O_q T_0}\, e_n(\alpha_m) \tag{3.3}$$

De même, l'intégrale de l'impulsion glottique J se déduit de l'intégrale de l'impulsion normalisée $j_n(\alpha_m)$ par :

$$J = A_v O_q T_0\, j_n(\alpha_m)$$

Domaine spectral

En appliquant la transformée de Fourier aux équations 3.1 et 3.2, on obtient une expression fréquentielle analytique pour chacun de ces modèles. Les expressions spectrales de l'ODG et de sa dérivée sont données par :

$$U_g(f) = A_v O_q T_0\ N_g(f O_q T_0, \alpha_m)\ (f_0 \perp\!\!\!\perp\!\!\!\perp_{f_0}(f)) \tag{3.4}$$

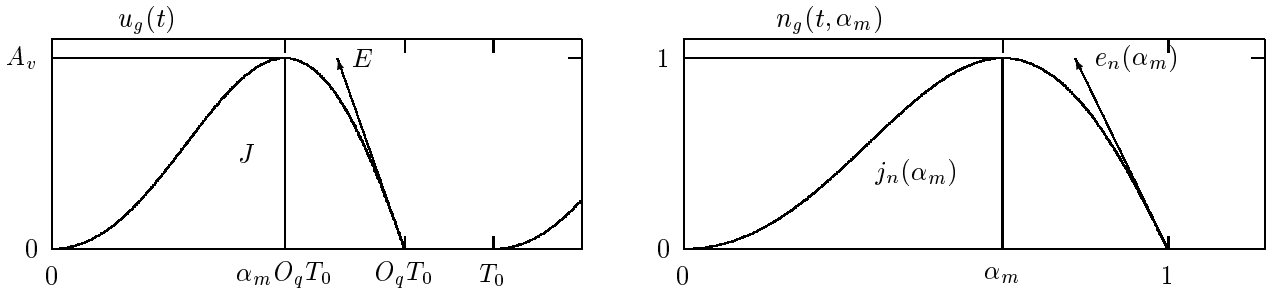


FIG. *3.2 – Forme temporelle générale de l'ODG ($u_g(t)$) et forme normalisée ($n_g(t, \alpha_m)$) par rapport à la période fondamentale T_0 et l'amplitude A_v. E est la vitesse de fermeture, e_n sa valeur normalisée, J l'intégrale de l'impulsion et j_n sa valeur normalisée. La forme normalisée de l'ODG ne dépend plus que du coefficient d'asymétrie α_m. D'après Doval & d'Alessandro* [55].

$$U_g^{'}(f) = A_v \; N_g^{'}(f O_q T_0, \alpha_m) \; (f_0 \text{⊥⊥⊥}_{f_0}(f)) \tag{3.5}$$

où $\text{⊥⊥⊥}_{f_0}(f)$ représente un peigne de Dirac de fréquence fondamentale $f_0 = \frac{1}{T_0}$ et $N_g(f, \alpha_m)$ la transformée de Fourier de $n_g(t, \alpha_m)$.

Comportement asymptotique

Le spectre de l'ODG peut être approximé par un filtre passe-bas d'ordre 2, dont la fréquence de coupure F_g va dépendre du modèle. On montre que :

$$U_g(f) = \int_0^{O_q T_0} u_g(t)\, e^{-j2\pi f t}\, dt \underset{f \to \infty}{\longrightarrow} \frac{E}{(2\pi f)^2}$$

$$et \;\; U_g(f=0) = \int_0^{O_q T_0} u_g(t)\, dt = J$$

On définit alors deux droites asymptotiques au spectre d'ODG : une asymptote horizontale $[(\Delta_1) = J]$ dans les basses fréquences (pour $0 \leq f \leq F_g$), de pente 0 dB/oct, et une asymptote $\left[(\Delta_2) = \frac{E}{(2\pi f)^2}\right]$ dans les hautes fréquences (pour $f \geq F_g$), de pente -12 dB/oct. Dans le cas de l'ODGD, la fréquence de coupure est identique, mais les pentes sont alors de $+6$ dB/oct $\left[(\Delta_1^{'}) = J\, 2\pi f\right]$ et -6 dB/oct $\left[(\Delta_2^{'}) = \frac{E}{2\pi f}\right]$. La figure 3.3 présente les spectres de l'ODG et de l'ODGD (calculés à partir des formules analytiques) dans le cas du modèle KLGLOTT88, ainsi que les asymptotes correspondantes.

L'amplification spectrale visualisée sur le spectre de l'ODGD à la fréquence F_g est appelée « formant glottique » (Fant [65], 1979), en raison des similitudes de forme avec les formants dus au conduit vocal. Ce terme est évidemment discutable et peut prêter à confusion. Il est donc bon de garder à l'esprit que, dans le cas du spectre de source glottique, l'accentuation spectrale n'est évidemment pas le fait de résonances dans des cavités, comme c'est le cas pour le conduit vocal. La fréquence du formant glottique et son amplitude vont dépendre du modèle et des paramètres de source (T_0, O_q et α_m). Dans le cas du spectre normalisé, ils sont donnés par :

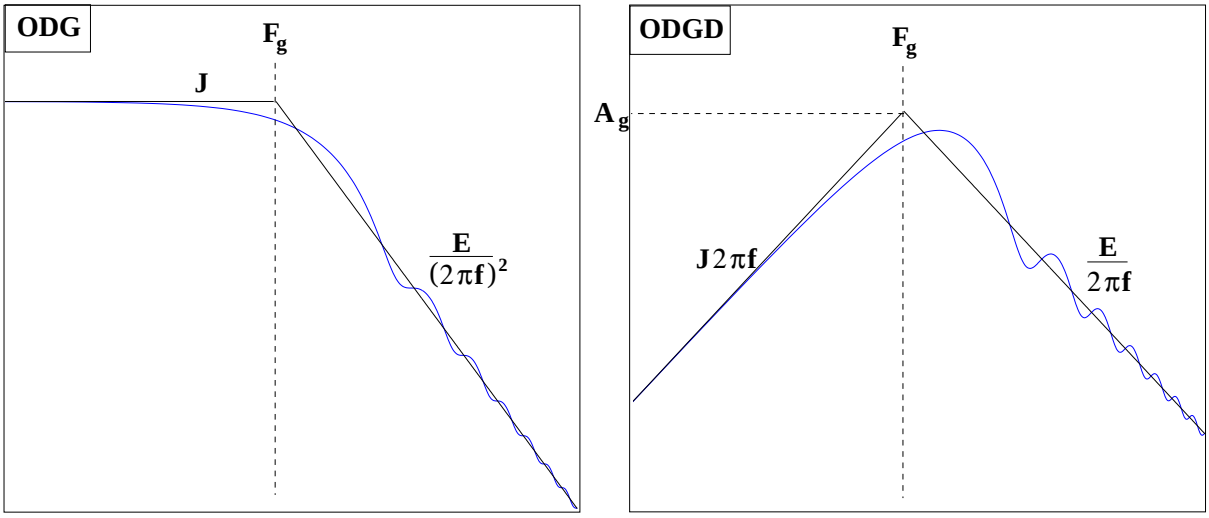


Fig. *3.3 – Spectre de l'ODG et de sa dérivée, et asymptotes correspondantes (modèle KLGLOTT88), dans le cas d'une fermeture abrupte. L'échelle des fréquences (en abscisse) est logarithmique. F_g est la fréquence du formant glottique et A_g l'amplitude associée.*

$$f_g(\alpha_m) = \frac{1}{2\pi}\sqrt{\frac{e_n(\alpha_m)}{j_n(\alpha_m)}}$$

$$a_g(\alpha_m) = \sqrt{e_n(\alpha_m)\, j_n(\alpha_m)}$$

La fréquence et l'amplitude du formant glottique dans le cas général s'en déduisent par :

$$F_g = \frac{f_g(\alpha_m)}{O_q T_0}$$

$$A_g = A_v\, a_g(\alpha_m)$$

Ils peuvent également s'exprimer en fonction des paramètres J et E par :

$$\begin{aligned} F_g &= \frac{1}{2\pi}\sqrt{\frac{E}{J}} \\ A_g &= \sqrt{EJ} \end{aligned} \tag{3.6}$$

Nous nous sommes placés ici dans le cadre d'une fermeture abrupte et le paramètre Q_a n'intervient donc pas. Le cas d'une fermeture non-abrupte se modélise par l'ajout d'un filtre passe-bas d'ordre 1, dont la fréquence de coupure F_a est située dans les hautes fréquences. Cela rajoute alors une asymptote supplémentaire en haute fréquence, $\left[(\Delta_3^{'}) = A_g \frac{F_a F_g}{f^2}\right]$, de pente $-12\ dB/oct$, ainsi que l'illustre la figure 3.4. D'après Fant [72, 69], la fréquence de coupure F_a est reliée au paramètre Q_a par :

$$F_a = \frac{1}{2\pi}\,\frac{1}{Q_a\,(1 - O_q T_0)} \tag{3.7}$$

L'amplitude A_a associée à cette fréquence de coupure est donnée par :

$$A_a = \frac{E}{2\pi F_a} = \frac{F_g}{F_a}\, A_g$$

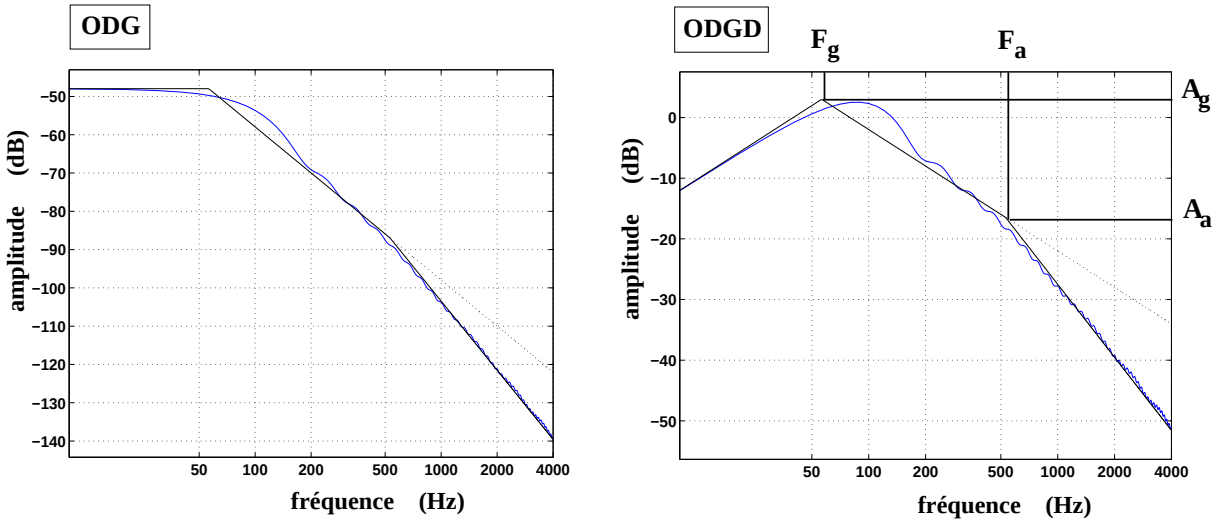


FIG. *3.4 – Spectre de l'ODG, de sa dérivée, et asymptotes correspondantes (modèle LF), dans le cas d'une fermeture non-abrupte. On observe l'atténuation spectrale dans les hautes fréquences, par rapport au cas d'une fermeture abrupte (ligne en pointillée).*

Nous reviendrons plus particulièrement sur l'effet spectral du paramètre Q_a en partie 3.2.4.

3.2 Effet spectral des paramètres de source glottique

Dans cette partie, nous allons nous appuyer sur le modèle LF, car c'est le plus complet des 3 modèles présentés précédemment : il permet en effet une plus grande variation du coefficient d'asymétrie α_m. Il est également le plus utilisé. Néanmoins, les résultats présentés dans cette partie sont généralisables aux trois modèles étudiés, en raison du formalisme mis en place.

3.2.1 Effet spectral du paramètre d'amplitude

Une modification du paramètre d'amplitude A_v, indépendamment des autres paramètres du modèle, a pour effet de translater l'enveloppe spectrale de l'ODG et de l'ODGD sur l'échelle des intensités sans modifier sa forme, ainsi que l'illustre la figure 3.5. Ceci revient à augmenter l'intensité vocale du son produit sans privilégier une zone fréquentielle particulière. Les autres paramètres (T_0, O_q, α_m, Q_a) sont gardés constants (sur cet exemple, $T_0 = \frac{1}{130}$ s, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$ et $Q_a = 0$ (pas de phase de retour)).

Il est intéressant de remarquer qu'une modification du paramètre A_v, indépendamment des autres paramètres, entraîne une modification du paramètre E dans les mêmes proportions. En effet, il existe une relation analytique directe entre A_v et E, donnée par l'équation 3.3. Dans le cas où T_0, O_q et α_m sont gardés constants, A_v et E sont directement proportionnels et ont exactement le même effet spectral sur l'ODG et sa dérivée.

3.2.2 Effet spectral du quotient ouvert et du coefficient d'asymétrie

La figure 3.6 illustre l'effet spectral d'une modification du quotient ouvert, dans le cas où les autres paramètres A_v, T_0, α_m et Q_a sont gardés constants (en particulier $\alpha_m = \frac{2}{3}$ et $Q_a = 0$ (pas de phase de retour)). Les formes temporelles et spectrales de l'ODG et de l'ODGD sont calculées analytiquement à

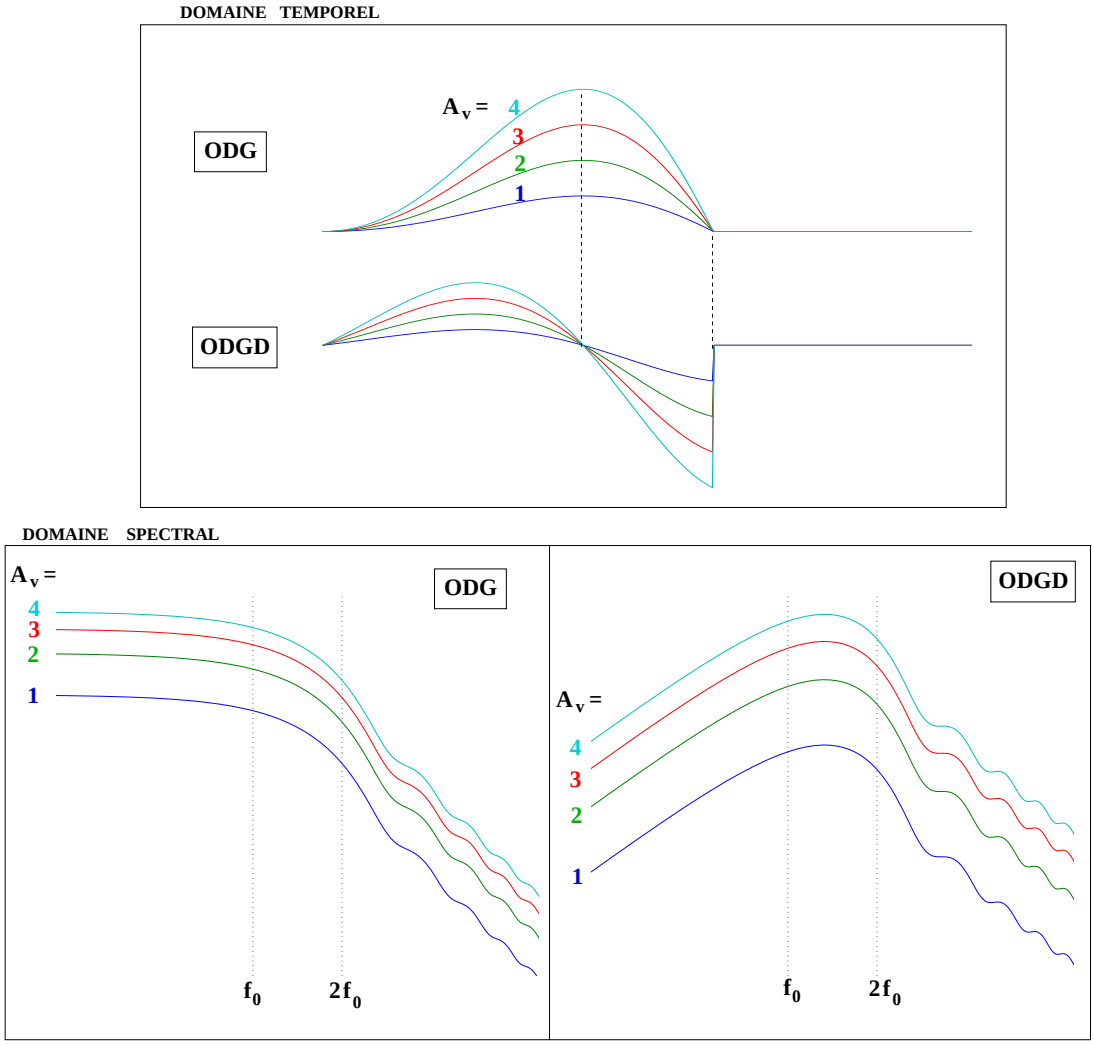


Fig. *3.5 – Variations temporelles et spectrales de l'ODG et de sa dérivée pour* 4 *valeurs de* A_v. *Les autres paramètres sont gardés constants (*$T_0 = \frac{1}{130}$ *s,* $O_q = 0.6$, $\alpha_m = \frac{2}{3}$ *et* $Q_a = 0$*).*

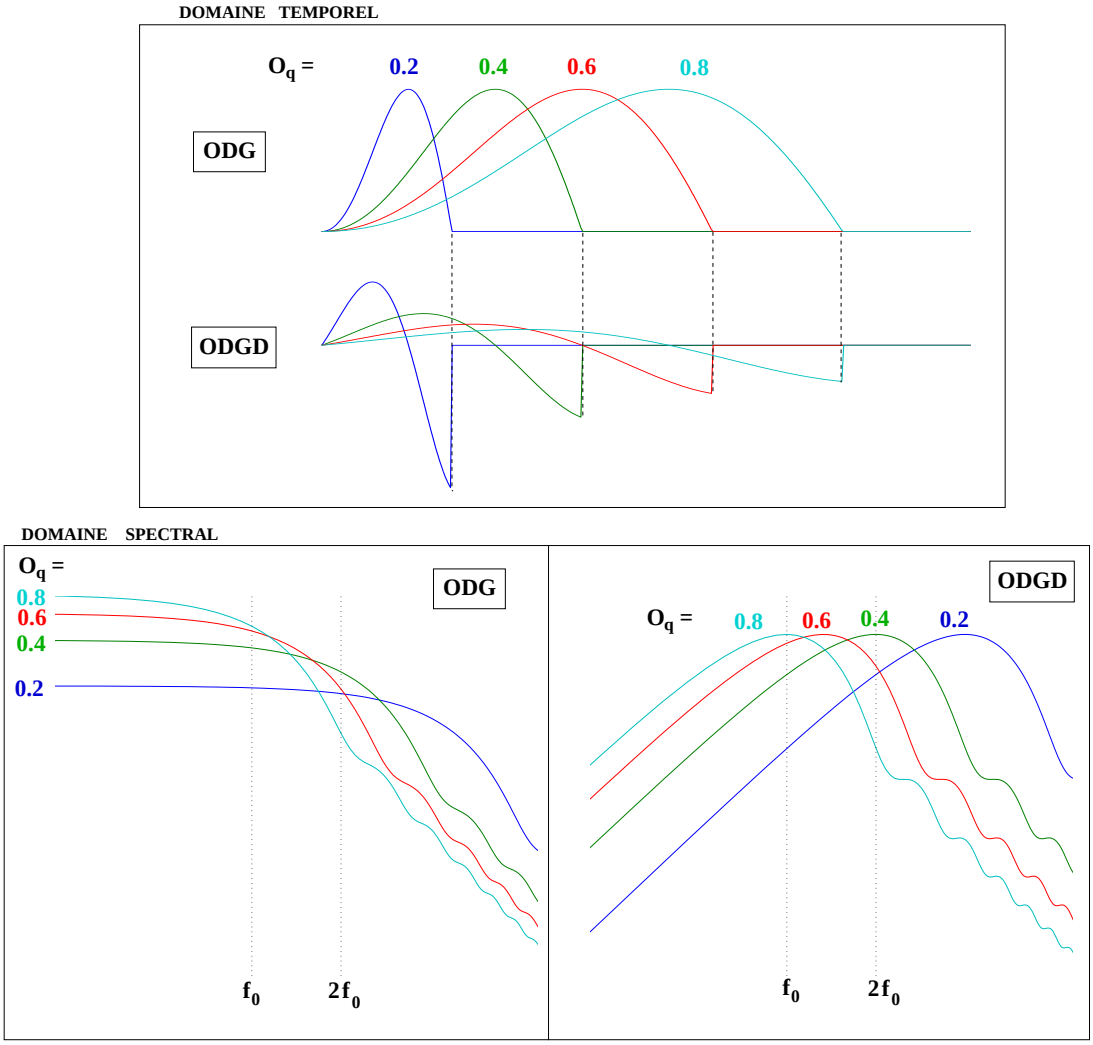


FIG. *3.6 – Variations temporelles et spectrales de l'ODG et de sa dérivée pour* 4 *valeurs de* O_q. *Les autres paramètres sont gardés constants* ($A_v = 1$, $T_0 = \frac{1}{130}$ *s*, $\alpha_m = \frac{2}{3}$ *et* $Q_a = 0$).

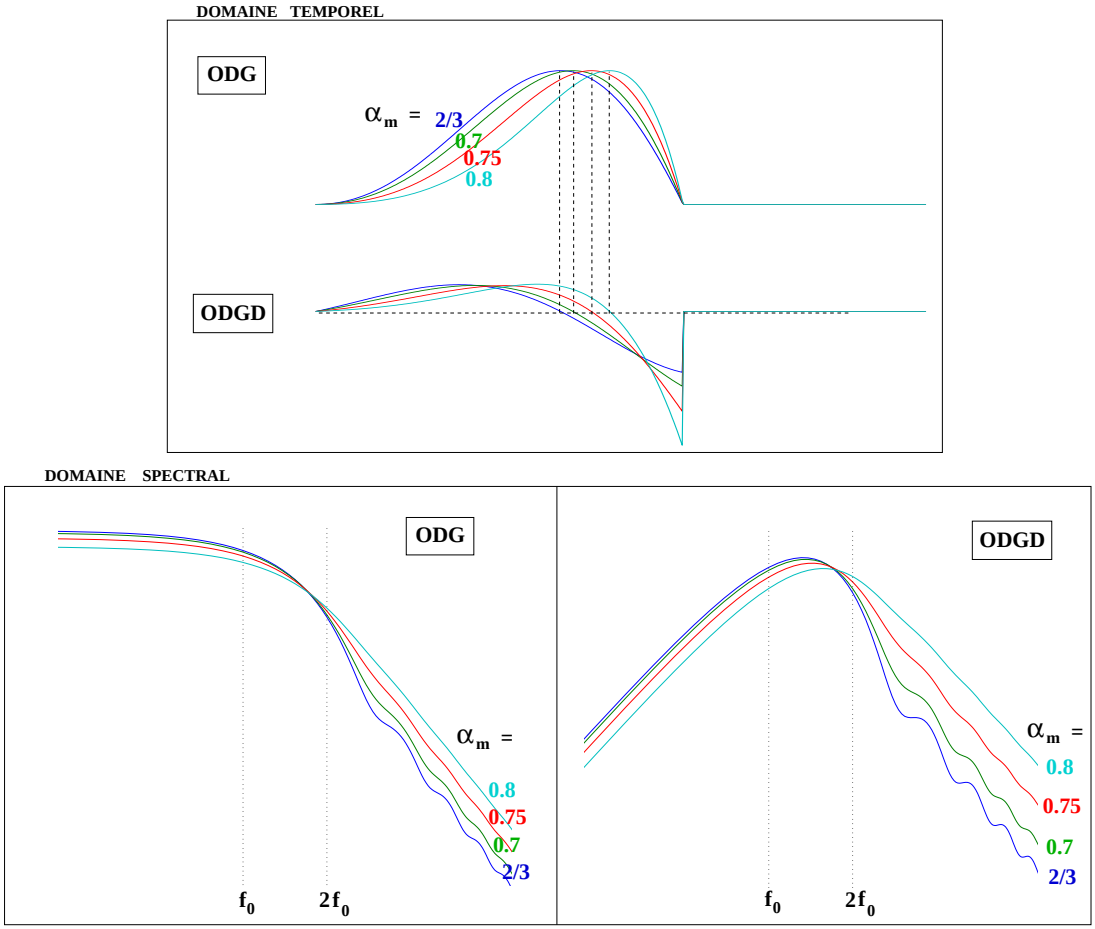


Fig. *3.7 – Variations temporelles et spectrales de l'ODG et de sa dérivée pour* 4 *valeurs de* α_m*. Les autres paramètres sont gardés constants (*$A_v = 1$, $T_0 = \frac{1}{130}$ *s,* $O_q = 0.6$ *et* $Q_a = 0$*)*

partir des équations 3.1, 3.2, 3.4 et 3.5 appliquées au modèle LF, pour 4 valeurs de O_q ($O_q \in [0.2, 0.8]$ par pas de 0.2). On remarque, en particulier sur le spectre de l'ODGD, qu'une diminution du quotient ouvert se traduit par une montée en fréquence du formant glottique et par une augmentation de l'énergie en haute fréquence.

Cet effet spectral se traduit perceptivement par une augmentation nette de l'intensité sonore et une impression de « brillance » du son quand O_q diminue. Il est illustré par les exemples sonores (CD, piste [3]) ($O_q = 0.4$) et (CD, piste [4]) ($O_q = 0.8$), qui correspondent à la synthèse [1] d'une vocalise sur deux hauteurs, où les paramètres de source sont gardés constants. Seule l'information de période fondamentale provient de la détection des instants de fermeture glottique sur le signal EGG d'un chanteur (B2, cf. chapitre 6). La fonction de transfert modélise le conduit vocal dans le cas d'un [a] et est gardée constante tout au long de la synthèse.

De même, la figure 3.7 illustre l'effet spectral d'une modification du coefficient d'asymétrie, dans le cas où les autres paramètres A_v, T_0, O_q et Q_a sont gardés constants (en particulier $O_q = 0.6$ et $Q_a = 0$ (pas de phase de retour)). Une augmentation de α_m entraîne une augmentation de la largeur de bande du formant glottique. Comme pour O_q, on observe également un effet dans les hautes fréquences : une augmentation de α_m entraîne une accentuation des composantes haute-fréquence du spectre de la source

[1] La méthode de synthèse de ces exemples sonores sera détaillée dans la partie 9.3.1 du chapitre 9. Les valeurs par défaut des paramètres de source sont : $A_v = 1$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$ et $Q_a = 0$.

glottique. Ces effets spectraux sont illustrés par les exemples sonores (CD, piste [5]) ($\alpha_m = \frac{2}{3}$) et (CD, piste [6]) ($\alpha_m = 0.8$).

Dans une étude de la voix chantée en mécanisme I, Sundberg & al [193] (1993) ont observé une forte corrélation entre l'intensité vocale et la vitesse de fermeture. Ils ont observé que, selon les chanteurs, 3 comportements différents pouvaient entraîner une augmentation du paramètre E :

1. une augmentation de l'amplitude de voisement,
2. une augmentation du coefficient d'asymétrie,
3. une diminution du quotient ouvert de façon conjointe à une augmentation du coefficient d'asymétrie.

Ces résultats expérimentaux sont tout à fait en accord avec les résultats théoriques que nous obtenons ici. D'après l'équation 3.3 qui lie E à A_v, O_q et α_m, le paramètre E augmente si A_v augmente, si O_q diminue, si f_0 augmente ou si $e_n(\alpha_m)$ augmente. Dans le cas où A_v reste constant, nous observons sur les figures 3.6 et 3.7 que les variations des paramètres O_q et α_m entraînent une modification de E, donc de la pente spectrale dans les hautes fréquences (figure 3.3). Ceci amène à une modification de l'intensité vocale.

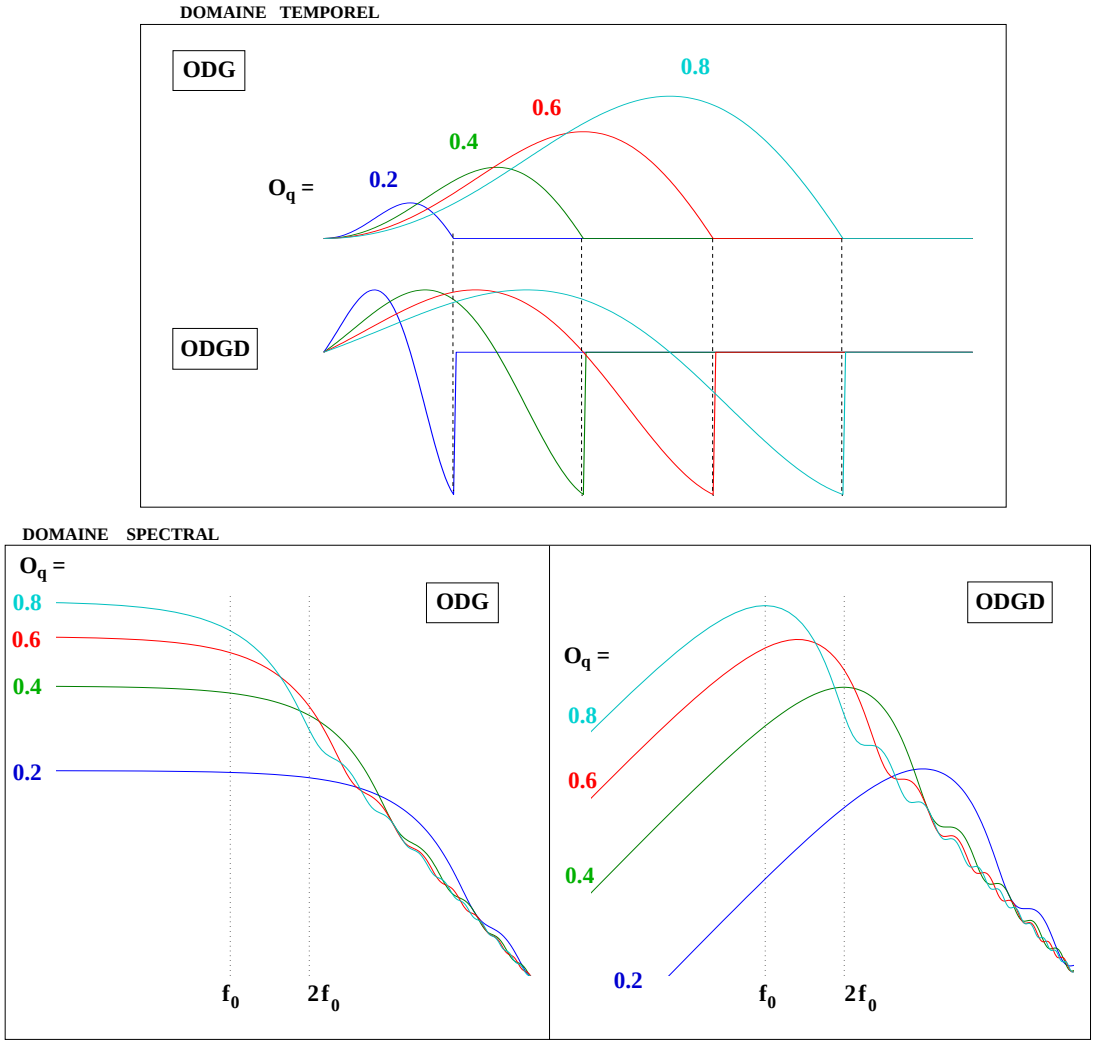


FIG. *3.8 – Variations temporelles et spectrales de l'ODG et de sa dérivée pour 4 valeurs de O_q, dans le cas où le paramètre d'amplitude E est gardé constant.*

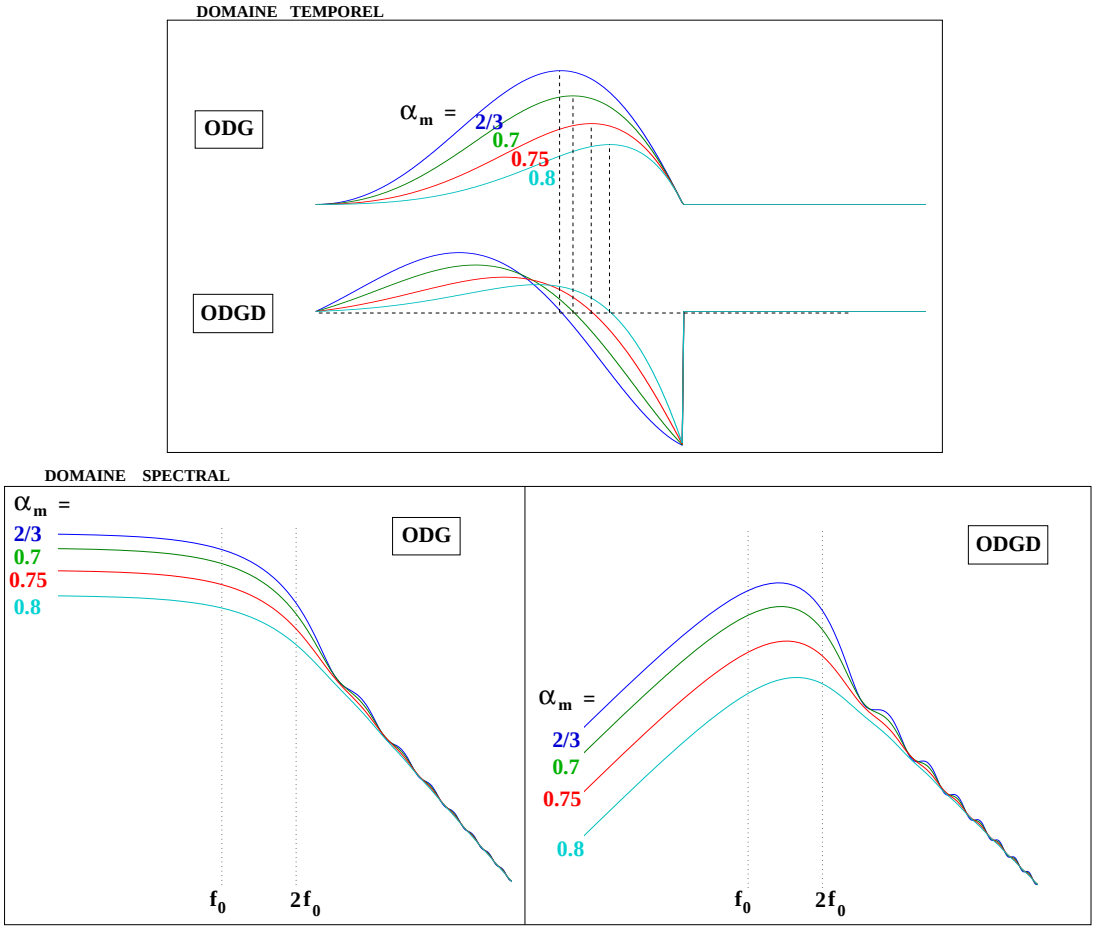


Fig. *3.9 – Variations temporelles et spectrales de l'ODG et de sa dérivée pour* 4 *valeurs de* α_m*, dans le cas où le paramètre d'amplitude* E *est gardé constant.*

Que se passe-t'il dans le cas où E est gardé constant, ce qui implique une variation sous-jacente de A_v ?

Les figures 3.8 et 3.9 présentent les variations temporelles et spectrales du quotient ouvert et du coefficient d'asymétrie, dans le cas du jeu de paramètres (T_0, E, O_q, α_m, Q_a). On remarque que, contrairement à ce qui a été observé précédemment sur les figures 3.6 et 3.7, une modification du quotient ouvert ou du coefficient d'asymétrie n'influence plus le comportement spectral dans les hautes fréquences, mais se traduit uniquement par un déplacement de la position fréquentielle du formant glottique. En effet, la pente spectrale dans les hautes fréquences dépend du paramètre E. Si E est gardé constant, la position en amplitude de cette pente spectrale ne va pas varier. L'effet perceptif d'une modification du quotient ouvert et du coefficient d'asymétrie est alors atténué, car l'oreille humaine est particulièrement sensible aux variations dans les hautes fréquences (approximativement, entre 1000 et 4000 Hz). Cet effet est illustré sur les exemples sonores (CD, piste [3]) ($O_q = 0.4$) et (CD, piste [4]) ($O_q = 0.8$) dans le cas du quotient ouvert, et (CD, piste [5]) ($\alpha_m = \frac{2}{3}$) et (CD, piste [6]) ($\alpha_m = 0.8$) dans le cas du coefficient d'asymétrie. Pour synthétiser ces exemples, le paramètre E a cette fois-ci été gardé constant ($E = 4000$).

Il est intéressant de remarquer que les effets spectraux du quotient ouvert et du coefficient d'asymétrie dépendent du jeu de paramètres que l'on considère. Nous voyons que, dans tous les cas, une modification du quotient ouvert ou du coefficient d'asymétrie se traduit par un déplacement de la position fréquentielle du formant glottique. Nous allons à présent quantifier ce déplacement relativement à la fréquence fondamentale en fonction des paramètres O_q et α_m.

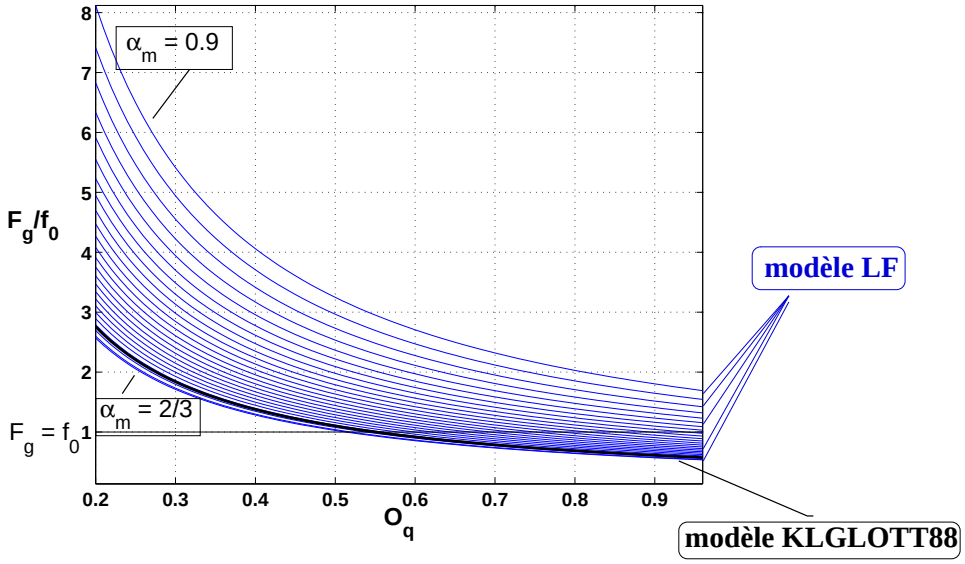


FIG. *3.10 – Position fréquentielle du formant glottique F_g relativement à la fréquence fondamentale, pour différentes valeurs de O_q, dans le cas du modèle KLGLOTT88 ($\alpha_m = \frac{2}{3}$, représenté en trait gras sur la figure) et du modèle LF ($\alpha_m \in [\frac{2}{3}, 0.9]$ par pas de 0.01).*

3.2.3 Position spectrale du formant glottique

Comme nous venons de le voir, la fréquence du formant glottique va dépendre de T_0, O_q et α_m. La figure 3.10 illustre la position de ce formant glottique relativement à f_0, dans le cas du modèle KLGLOTT88 (α_m constant à $\frac{2}{3}$) et dans le cas du modèle LF (où α_m varie entre $\frac{2}{3}$ et 0.9), pour O_q compris entre 0.2 et 0.96. Dans le cas où $F_g = f_0$, le formant glottique est situé sur l'harmonique H_1 (qui correspond à la fréquence fondamentale : $H_1 = f_0$) et la source va donc avoir pour effet spectral d'accentuer fortement cette harmonique. Ceci est en particulier le cas pour des valeurs de O_q situées entre 0.5 et 0.6, avec une asymétrie réduite (α_m proche de $\frac{2}{3}$). Dans le cas du modèle KLGLOTT88, le formant glottique n'accentue que rarement des fréquences supérieures au second harmonique. Par contre, dès que l'on peut modéliser plus finement l'asymétrie de l'impulsion, comme c'est le cas pour le modèle LF, on remarque que, plus cette asymétrie est marquée, plus le formant glottique va avoir un effet dans les hautes fréquences, pouvant même atteindre des valeurs théoriques de 4 à 8 f_0 dans le cas d'un quotient ouvert faible ($O_q \leq 0.4$) et d'un coefficient d'asymétrie élevé ($\alpha_m \geq 0.8$).

3.2.4 Effet spectral du quotient de phase de retour

Le formalisme sur lequel nous nous sommes appuyés jusqu'à présent suppose une fermeture abrupte. Pour étudier l'effet spectral d'une phase de retour non nulle, nous devons compléter les expressions temporelles et spectrales des modèles étudiés en tenant compte du paramètre Q_a. Nous nous intéresserons dans cette partie uniquement au modèle LF, puisqu'il est le plus complet et le plus utilisé des modèles d'ODG, et nous réécrirons les équations de ce modèle en tenant compte de la phase de retour.

Expression temporelle du modèle LF dans le cas d'une fermeture non abrupte

Dans le cas d'une fermeture non abrupte ($Q_a > 0$), le modèle LF se réécrit :

$$u_g(t) = \begin{cases} -\frac{E\, e^{-a O_q T_0}}{\sin(\frac{\pi}{\alpha_m})\,(a^2+(\frac{\pi}{\alpha_m O_q T_0})^2)} \left(\frac{\pi}{\alpha_m O_q T_0} + a\, e^{at} \sin(\frac{\pi}{\alpha_m O_q T_0}\, t) - \frac{\pi}{\alpha_m O_q T_0}\, e^{at} \cos(\frac{\pi}{\alpha_m O_q T_0}\, t)\right) & pour\ 0 \le t \le O_q T_0 \\ -E \left(\frac{1}{\varepsilon\, Q_a\,(1-O_q) T_0} - 1\right) \left(T_0 - t + \frac{1-e^{\varepsilon(T_0-t)}}{\varepsilon}\right) & pour\ O_q T_0 \le t \le T_0 \end{cases}$$

$$u_g^{'}(t) = \begin{cases} -E\, e^{a(t-O_q T_0)}\ \frac{\sin(\frac{\pi}{\alpha_m O_q T_0}\, t)}{\sin(\frac{\pi}{\alpha_m})} & pour\ 0 \le t \le O_q T_0 \\ -\frac{E}{\varepsilon\, Q_a\,(1-O_q) T_0}\, e^{\varepsilon\, O_q T_0} \left(e^{-\varepsilon\, t} - e^{-\varepsilon\, T_0}\right) & pour\ O_q T_0 \le t \le T_0 \end{cases}$$

Le facteur ε est calculé en exprimant la condition de continuité de l'ODGD à l'instant de fermeture glottique :

$$u_g^{'}(t = (O_q T_0)^-) = u_g^{'}(t = (O_q T_0)^+)$$

$$\Rightarrow \quad \frac{e^{\varepsilon O_q T_0}}{\varepsilon\, Q_a\,(1-O_q) T_0} \left(e^{-\varepsilon\, O_q T_0} - e^{-\varepsilon\, T_0}\right) = 1$$

ce qui amène à l'équation implicite suivante :

$$1 + e^{\varepsilon\,(1-O_q) T_0}\ (\varepsilon\, Q_a (1 - O_q) T_0 - 1) = 0 \tag{3.8}$$

$\varepsilon = 0$ est une solution triviale de l'équation 3.8. Pour que cette équation admette une autre solution, il faut que :

$$(\varepsilon\, Q_a (1 - O_q) T_0 - 1) < 0$$

$$\Rightarrow \quad \varepsilon < \frac{1}{Q_a\,(1-O_q) T_0}$$

Ces considérations fixent les bornes de variation du facteur ε : $0 < \varepsilon < \frac{1}{Q_a\,(1-O_q) T_0}$.

Le facteur a se calcule en exprimant la condition de continuité de l'ODG à l'instant de fermeture glottique :

$$u_g(t = (O_q T_0)^-) = u_g(t = (O_q T_0)^+)$$

$$\Rightarrow \quad \frac{e^{-a O_q T_0}}{\sin(\frac{\pi}{\alpha_m})\,(a^2+(\frac{\pi}{\alpha_m O_q T_0})^2)} \left(\frac{\pi}{\alpha_m O_q T_0} + a\, e^{a\, O_q T_0} \sin(\frac{\pi}{\alpha_m}) - \frac{\pi}{\alpha_m O_q T_0}\, e^{a\, O_q T_0} \cos(\frac{\pi}{\alpha_m})\right)$$
$$= \left(\frac{1}{\varepsilon\, Q_a\,(1-O_q) T_0} - 1\right) \left(T_0 (1 - O_q) + \frac{1-e^{\varepsilon\,(1-O_q) T_0}}{\varepsilon}\right)$$

Cette égalité peut se simplifier en tenant compte de l'équation 3.8, qui permet de remplacer le facteur $\left(\frac{1}{\varepsilon\, Q_a\,(1-O_q) T_0} - 1\right)$ par $\left(\frac{1}{e^{\varepsilon\,(1-O_q) T_0} - 1}\right)$. En développant les deux membres, cette équation peut se réécrire sous la forme :

$$\frac{1}{a^2 + (\frac{\pi}{\alpha_m O_q T_0})^2} \left(a + \frac{\pi}{\alpha_m O_q T_0} * \frac{e^{-a\, O_q T_0} - \cos(\frac{\pi}{\alpha_m})}{\sin(\frac{\pi}{\alpha_m})}\right) = \frac{T_0 (1 - O_q)}{e^{\varepsilon\, T_0 (1-O_q)} - 1} - \frac{1}{\varepsilon} \tag{3.9}$$

Dans le cas où $Q_a = 0$, le membre de droite de l'équation 3.9 est nul. Nous retrouvons alors l'équation A.1 implicite sur a_n présentée dans l'annexe A dans le cas du modèle LF normalisé, au facteur $O_q T_0$ multiplicatif près : $a = \frac{a_n}{O_q T_0}$. La normalisation par rapport à $O_q T_0$ revient à remplacer ce facteur par 1.

Expression spectrale du modèle LF

Si on calcule la transformée de Fourier de $u_g^{'}(t)$, on obtient l'équation suivante :

$$U_g^{'}(f) = U_{g1}^{'}(f) + U_{g2}^{'}(f)$$

avec

$$U_{g1}^{'}(f) = -\frac{Ee^{-aO_qT_0}}{\sin(\frac{\pi}{\alpha_m})((a-j2\pi f)^2+(\frac{\pi}{\alpha_m O_q T_0})^2)}\left(\frac{\pi}{\alpha_m O_q T_0} + e^{(a-j2\pi f)O_qT_0}\left((a-j2\pi f)\sin(\frac{\pi}{\alpha_m}) - \frac{\pi}{\alpha_m O_q T_0}\cos(\frac{\pi}{\alpha_m})\right)\right)$$

$$U_{g2}^{'}(f) = \frac{Ee^{-j2\pi f O_q T_0}}{j2\pi f(\varepsilon+j2\pi f)}\left(\frac{e^{-\varepsilon T_0(1-O_q)}(1-e^{-j2\pi f T_0(1-O_q)})}{Q_a(1-O_q)T_0} - j2\pi f\right)$$

On en déduit $U_g(f) = TF(u_g(t))$ par $U_g(f) = \frac{1}{j2\pi f}\,U_g^{'}(f)$.

Etude spectrale des variations du quotient de phase de retour

A l'aide de ces équations, une étude analytique des effets spectraux associés aux variations du quotient de phase de retour Q_a est possible. La figure 3.11 présente les effets temporels et spectraux dans le cas des 4 valeurs de Q_a suivantes : $Q_a = 0$, 0.05 , 0.1 , 0.2. Les valeurs par défaut des autres paramètres sont : $f_0 = 100$ Hz, $E = 500$, $O_q = 0.7$ et $\alpha_m = 0.7$. Pour une meilleure lisibilité de la figure, les asymptotes au spectre ont également été tracées dans chaque cas.

Une augmentation de Q_a entraîne une atténuation importante des hautes fréquences. Cela correspond à l'ajout d'un filtre passe-bas dont la fréquence de coupure F_a est inversement proportionnelle à Q_a, d'après l'équation 3.7. Plus Q_a augmente, plus cette fréquence de coupure se déplace dans les basses fréquences. On observe de ce fait également une petite accentuation des basses fréquences. Par contre, la position du formant glottique reste relativement stable (petite diminution quand Q_a augmente).

La variation de ce paramètre a donc un effet perceptif notable, puisqu'il affecte les hautes fréquences du spectre de source glottique, donc la région fréquentielle à laquelle l'oreille humaine est particulièrement sensible. Cet effet perceptif est illustré sur les exemples sonores (CD, piste [5]) ($Q_a = 0$) et (CD, piste [11]) ($Q_a = 0.2$).

3.3 Conclusion

Nous avons présenté ici un formalisme permettant d'unifier les modèles de source glottique et de les exprimer de façon à la fois temporelle et spectrale à l'aide du jeu unique de paramètres glottiques suivant : la période fondamentale T_0, l'amplitude de voisement A_v, le quotient ouvert O_q, le coefficient d'asymétrie α_m et le quotient de phase de retour Q_a. Nous avons appliqué directement ce formalisme pour étudier les effets spectraux de ces paramètres.

A partir de ce travail théorique d'exploration des paramètres de source glottique, nous voyons que l'intensité vocale du son émis est dépendante de l'ensemble de ces paramètres. Les paramètres d'amplitude (A_v ou E) sont comparables à des gains car ils amplifient de façon globale le spectre de source glottique. A une moindre échelle, les paramètres O_q, α_m et Q_a agissent également sur l'intensité vocale en accentuant ou atténuant les hautes fréquences. O_q et α_m ont peu d'influence sur l'intensité vocale si la vitesse de fermeture E est constante.

Nous avons également discuté des effets basse fréquence du quotient ouvert et du coefficient d'asymétrie. Ces paramètres modifient la fréquence et la largeur de bande du formant glottique. Ils vont

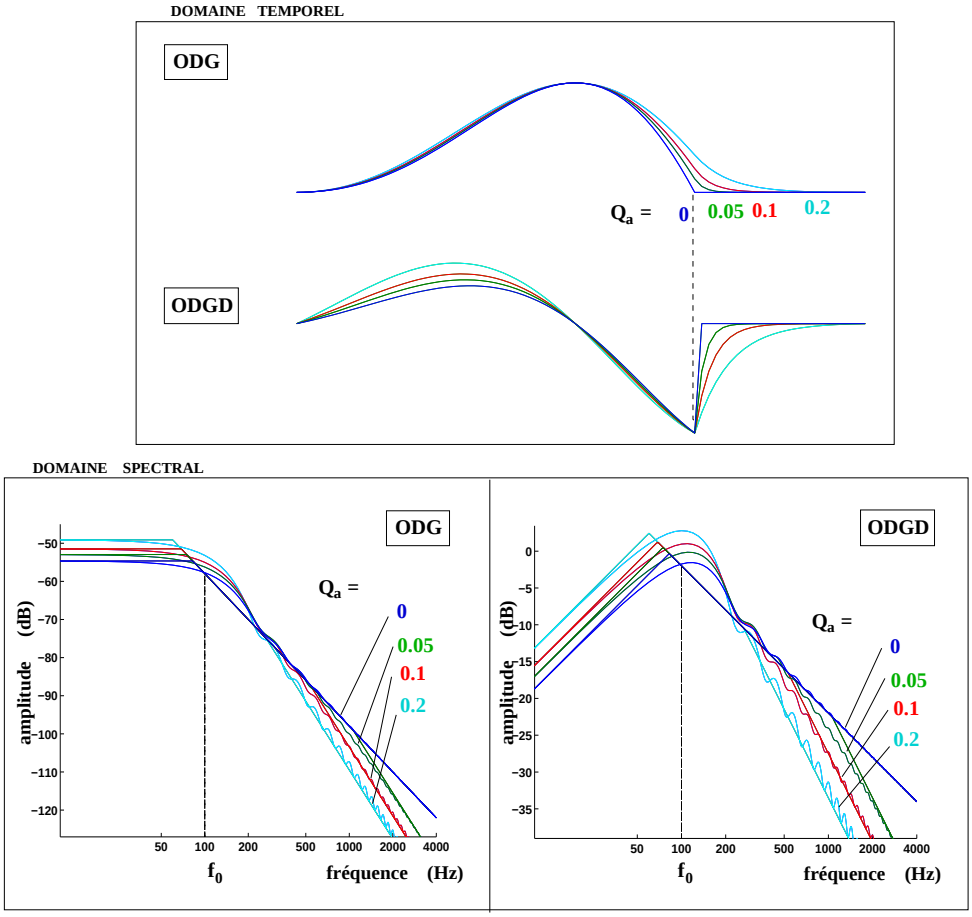


FIG. *3.11 – Variations temporelles et spectrales de l'ODG et de sa dérivée pour 4 valeurs de Q_a. Les autres paramètres sont gardés constants ($A_v = 1$, $T_0 = \frac{1}{100}$ s, $O_q = 0.7$, $\alpha_m = 0.7$).*

influencer l'amplitude spectrale des premiers harmoniques du spectre de source, en particulier H_1 et H_2. Les voix pressées sont caractérisées par des valeurs faibles de quotient ouvert et de coefficient d'asymétrie. Spectralement, on observe pour ces productions une diminution de l'amplitude du premier harmonique par rapport aux harmoniques d'ordre plus élevé (Gauffin & Sundberg [85], Klatt & Klatt [132]). Nous retrouvons ces caractéristiques dans notre étude théorique. Sur la figure 3.6, nous pouvons observer que la fréquence du formant glottique se déplace de f_0 à $2f_0$ quand O_q diminue de 0.8 à 0.4, ce qui entraîne effectivement une diminution de l'amplitude du fondamental par rapport au second harmonique. Ces résultats théoriques sont donc en bon accord avec les résultats expérimentaux présentés dans la littérature. En particulier, plusieurs auteurs montrent des corrélations fortes entre O_q et la différence spectrale entre les deux premiers harmoniques du spectre de source glottique (Holmberg & al [111], Fant [69, 70], Sundberg & al [190]). Il est même suggéré par ces auteurs qu'une mesure spectrale de $(H_1 - H_2)$ permet une mesure indirecte du quotient ouvert. Or nous voyons ici que cette différence spectrale pourrait également dépendre des variations du coefficient d'asymétrie et du quotient de phase de retour. Nous nous proposons alors d'étudier dans le chapitre suivant cette relation entre O_q et $(H_1 - H_2)$ à la lumière du formalisme développé dans le présent chapitre.

En Résumé

Un travail d'unification des modèles d'ODG a été entrepris par Doval et d'Alessandro [55], afin de dégager un jeu de paramètres qui puisse décrire l'ODG et sa dérivée, indépendamment du modèle. Ce jeu de paramètres temporels est : (A_v, T_0, O_q, α_m, Q_a), avec A_v l'amplitude de voisement, T_0 la période fondamentale, O_q le quotient ouvert, α_m le coefficient d'asymétrie et Q_a le quotient de phase de retour.

Un formalisme d'écriture permet d'exprimer de façon très simple les modèles LF (Fant, Liljencrants & Lin [72]), KLGLOTT88 (Klatt & Klatt [132]) et R++ (Veldhuis [213]) dans le domaine temporel et dans le domaine spectral, et d'étudier de façon analytique les propriétés temporelles et spectrales de l'ODG et de l'ODGD lors de variations des paramètres de source glottique.

L'effet spectral du paramètre d'amplitude revient à une amplification globale du signal.

Dans le cas où A_v est constant, une augmentation du quotient ouvert se traduit spectralement par une diminution de la fréquence du formant glottique et une atténuation de l'énergie dans les hautes fréquences. Une augmentation du coefficient d'asymétrie se traduit par une augmentation de la largeur de bande du formant glottique et une petite augmentation de sa fréquence, ainsi qu'une accentuation de l'énergie dans les hautes fréquences. L'effet spectral de ces paramètres dans les hautes fréquences provient de la variation sous-jacente de la vitesse de fermeture E. Cet effet est alors supprimé si l'on garde E constant (ce qui implique une variation de A_v).

La position fréquentielle du formant glottique est fonction de O_q et α_m. Elle est située près du fondamental pour des valeurs moyennes de O_q et α_m, mais peut atteindre des valeurs de 4 à 8 f_0 dans le cas d'un quotient ouvert faible ($O_q \leq 0.4$) et d'un coefficient d'asymétrie élevé ($\alpha_m \geq 0.8$).

Un quotient de phase de retour Q_a non nul entraîne l'ajout en haute fréquence d'un filtre passe-bas dont la fréquence de coupure diminue avec l'augmentation de Q_a. Une augmentation de Q_a entraîne donc une diminution très significative de l'énergie dans les hautes fréquences.

Ce travail illustre de façon théorique des résultats expérimentaux sur la qualité vocale. Dans le cas des voix fortes ou faibles, les paramètres A_v et E sont souvent corrélés à l'intensité vocale. Nous montrons que la variation de ces paramètres équivaut à la modification d'un gain. Dans le cas des voix pressées, le quotient ouvert est faible et on observe spectralement une atténuation du fondamental par rapport aux harmoniques supérieurs. Nous retrouvons ces variations à partir des modèles paramétriques d'ODG.

Chapitre 4

Estimation spectrale des paramètres de source glottique

Dans le chapitre 3, nous avons illustré les effets spectraux des paramètres de source glottique en prenant appui sur le formalisme développé. Nous avons montré en particulier qu'une augmentation du quotient ouvert se traduit par une diminution de la fréquence du formant glottique. Ceci pourrait donc expliquer que les voix relâchées, caractérisées par des quotients ouverts élevés, présentent souvent un fondamental (ou premier harmonique) prédominant par rapport aux harmoniques supérieurs tandis qu'au contraire les voix pressées, pour lesquelles le quotient ouvert est généralement faible, présentent un fondamental fortement atténué. Plusieurs études expérimentales ont même montré une corrélation forte entre le quotient ouvert et la différence spectrale entre les deux premiers harmoniques du spectre de source glottique (Holmberg & al [111], Fant [69, 70], Sundberg & al [190]) et avancé que cette différence spectrale permettrait une mesure indirecte du quotient ouvert (Stevens & Hanson [184], Hanson [97, 98], Fant [69, 70]). Maintenant, nous avons également observé dans le chapitre précédent que la variation du coefficient d'asymétrie ou du quotient de phase de retour pouvait influencer les premiers harmoniques du spectre de source. Nous nous proposons donc d'explorer plus en détail leur influence sur la différence spectrale $(H_1 - H_2)$, afin de déterminer si la relation liant $(H_1 - H_2)$ à O_q peut être dépendante de ces autres paramètres de source glottique. En particulier, nous désirons savoir si l'utilisation de mesures spectrales permet l'estimation du quotient ouvert [1]. Dans la partie 4.1, nous étudierons de façon théorique la relation liant l'amplitude des premiers harmoniques du spectre de source glottique au quotient ouvert, à la lumière du formalisme développé au chapitre 3. Nous validerons ensuite cette étude par l'analyse de signaux réels de parole et de chant. Pour ce faire, nous développerons en partie 4.2 les méthodes classiques d'estimation de la source glottique que nous serons amenées à appliquer dans cette étude et nous présenterons les analyses de $(H_1 - H_2)$ en fonction de O_q pour quelques signaux de parole et de chant dans la partie 4.3.

4.1 Etude théorique de la différence spectrale $(H_1 - H_2)$

4.1.1 Introduction

On note par H_i à la fois l'harmonique i (de fréquence $i * f_0$) et l'amplitude de cet harmonique (en décibel) mesurée sur le spectre de l'ODGD. On rappelle que le premier harmonique H_1 correspond à la fréquence fondamentale f_0 .

[1] Une partie de cette étude a été présentée à l'occasion du congrès Eurospeech 2001 (Henrich & al [102])

La production des sons soufflés ou relâchés est caractérisée par une prédominance du fondamental (premier harmonique), au contraire des productions vocales pressées pour lesquelles le fondamental est souvent fortement attenué par rapport aux harmoniques élevés (Klatt & Klatt [132], 1990). Dans des études portant sur l'analyse acoustique de la source glottique, plusieurs auteurs ont montré des corrélations nettes entre la différence spectrale $(H_1 - H_2)$ et le quotient ouvert (ou fermé). Hanson [97, 98] a étudié cette différence spectrale sur un groupe de 22 locutrices. Elle a observé que les sujets ayant des quotients ouverts élevés présentent généralement aussi une différence spectrale accrue entre les deux premiers harmoniques. Par contre, l'effet d'une modification de cette différence spectrale sur la perception des sons soufflés semble limité (Hanson [97], 1995). De même, Holmberg & al [111] (1995) ont trouvé une corrélation entre $(H_1 - H_2)$ et O_q, mesurés sur l'ODG dans le cas de 20 locutrices. Dans une étude portant sur les effets des variations de la pression sous-glottique dans le cas de chanteurs barytons professionnels, Sundberg & al [190] ont également observé une corrélation nette entre $(H_1 - H_2)_{ODG}$ mesuré sur l'ODG et le quotient fermé C_q, qu'ils quantifient par la relation : $(H_1 - H_2)_{ODG} = 21.5 - 31.1 * C_q$. On peut réécrire cette relation à l'aide de $O_q = 1 - C_q$ par : $(H_1 - H_2)_{ODG} = -9.6 + 31.1 * Oq$.

Fant [69, 70] (1995) s'est attaché à décrire cette corrélation à l'aide du modèle LF. Les valeurs spectrales de H_1 et H_2 sont mesurées sur des signaux synthétiques pour des variations du paramètre R_d [2] entre 0.3 (phonation tendue, avec un quotient ouvert réduit et une fermeture abrupte) et 2.7 (phonation relâchée, avec un quotient ouvert élevé et une longue phase de retour). Une corrélation linéaire quasi-parfaite est trouvée entre $(H_1 - H_2)$ et R_d : $(H_1 - H_2) = -7.6 + 11.1\ R_d$. Le quotient ouvert est également lié de façon directe à $(H_1 - H_2)$ par la relation suivante :

$$(H_1 - H_2) = -6 + 0.27 * e^{5.5\ Oq} \tag{4.1}$$

Nous nous proposons de vérifier sur les 3 modèles présentés au chapitre 3 le lien qui relie $(H_1 - H_2)$ et O_q.

4.1.2 Etude théorique

Influence de l'amplitude de voisement et de la fréquence fondamentale

Dans le cas d'une fermeture abrupte, l'expression analytique de H_1 et H_2 s'exprime à partir de l'équation 3.5 par :

$$\begin{aligned} H_1 &= 20\log_{10}(U_g^{'}(f_0)) = 20\log_{10}(A_v * N_g^{'}(O_q, \alpha_m) * f_0) \\ H_2 &= 20\log_{10}(U_g^{'}(2\,f_0)) = 20\log_{10}(A_v * N_g^{'}(2\,O_q, \alpha_m) * f_0) \end{aligned} \tag{4.2}$$

Ces équations montrent que $(H_1 - H_2)$ est indépendante de l'amplitude de voisement, ainsi que de la fréquence fondamentale :

$$(H_1 - H_2) = 20\log_{10}(\frac{N_g^{'}(O_q, \alpha_m)}{N_g^{'}(2\,O_q, \alpha_m)})$$

$(H_1 - H_2)$ est donc fonction de O_q, mais elle semble également être dépendante du coefficient d'asymétrie α_m.

[2] Le paramètre R_d relatif au modèle LF est décrit en détail au chapitre 3, en partie 2.2.2

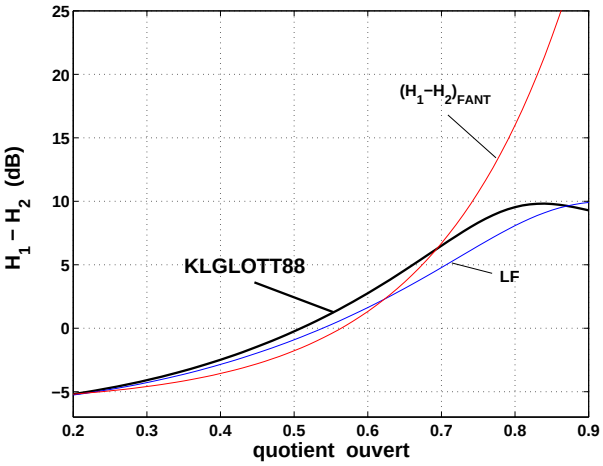


Fig. *4.1 – Différence spectrale $(H_1 - H_2)$ pour O_q variant entre 0.2 et 0.9, dans le cas du modèle KLGLOTT88 (trait épais). Le coefficient d'asymétrie α_m est une constante du modèle fixée à $\frac{2}{3}$. Le cas du modèle LF a également été représenté sur cette figure (trait fin), pour $\alpha_m = \frac{2}{3}$, ainsi que la relation empirique 4.1 proposée par Fant.*

Asymétrie constante : cas du modèle KLGLOTT88

Les travaux de Hanson [97, 98] se basent sur l'utilisation du modèle KLGLOTT88 (Klatt & Klatt [132]). La figure 4.1 présente la différence spectrale $(H_1 - H_2)$ en fonction du quotient ouvert O_q dans le cas de ce modèle (courbe en trait épais). On observe une relation bijective entre $(H_1 - H_2)$ et O_q tant que $O_q \leq 0.78$. Sur la base de ce modèle, on peut donc estimer le quotient ouvert en mesurant l'amplitude spectrale des deux premiers harmoniques. Néanmoins, dans ce modèle, le coefficient d'asymétrie est une constante fixée à $\frac{2}{3}$. Or, nous avons vu en partie 3.2.2 que ce paramètre a une influence importante sur la largeur de bande du formant glottique. On peut donc légitimement penser qu'une variation du coefficient d'asymétrie va avoir une conséquence sur $(H_1 - H_2)$.

Asymétrie variable : cas des modèles R++ et LF

Les modèles R++ et LF permettent de faire varier le coefficient d'asymétrie dans une certaine mesure. La figure 4.2 présente le calcul de $(H_1 - H_2)$ en fonction de O_q dans le cas de ces deux modèles. Les échelles sont identiques pour permettre la comparaison entre les deux modèles. Les différentes courbes correspondent à des valeurs données du coefficient d'asymétrie : α_m varie entre 0.5 et 0.75 dans le cas du modèle R++ et entre 0.52 et 0.98 dans le cas du modèle LF. Le pas de variation de α_m est de 0.01. Sur la figure présentant le modèle R++, la courbe obtenue précédemment dans le cas du modèle KLGLOTT88 est représentée pour permettre la comparaison. Sur la figure présentant le modèle LF, la relation empirique 4.1 proposée par Fant est également représentée.

Nous pouvons observer sur ces figures que la relation entre O_q et $(H_1 - H_2)$ n'est plus bijective : pour une valeur donnée de $(H_1 - H_2)$, il existe un intervalle de variation possible pour O_q. La connaissance de α_m est alors nécessaire si l'on désire déduire une mesure de quotient ouvert de la mesure spectrale $(H_1 - H_2)$. Par exemple, en moyennant ses résultats sur 22 locutrices, Hanson [98] trouve une valeur moyenne $(H_1 - H_2) = 3.4$ dB dans le cas de la voyelle [æ]. Si nous nous reportons alors à la figure 4.2, nous observons que, pour $3.3 \leq H_1 - H_2 \leq 3.5$, les valeurs possibles de quotient ouvert sont comprises entre 0.66 et 0.87. Ainsi, plusieurs couples de valeurs, tels $(O_q = 0.66, \alpha_m = 2/3)$, $(O_q = 0.70, \alpha_m = 0.72)$ ou $(O_q = 0.83, \alpha_m = 0.78)$, amènent à une même valeur de $(H_1 - H_2)$. Afin d'illustrer les plages de variation des couples (O_q, α_m), la figure 4.3 présente le domaine de variation de O_q et α_m pour des valeurs de $(H_1 - H_2)$ comprises dans les intervalles suivants : [-5,-4] dB, [-3,-2] dB, [0,1] dB et [4,5] dB. Quand

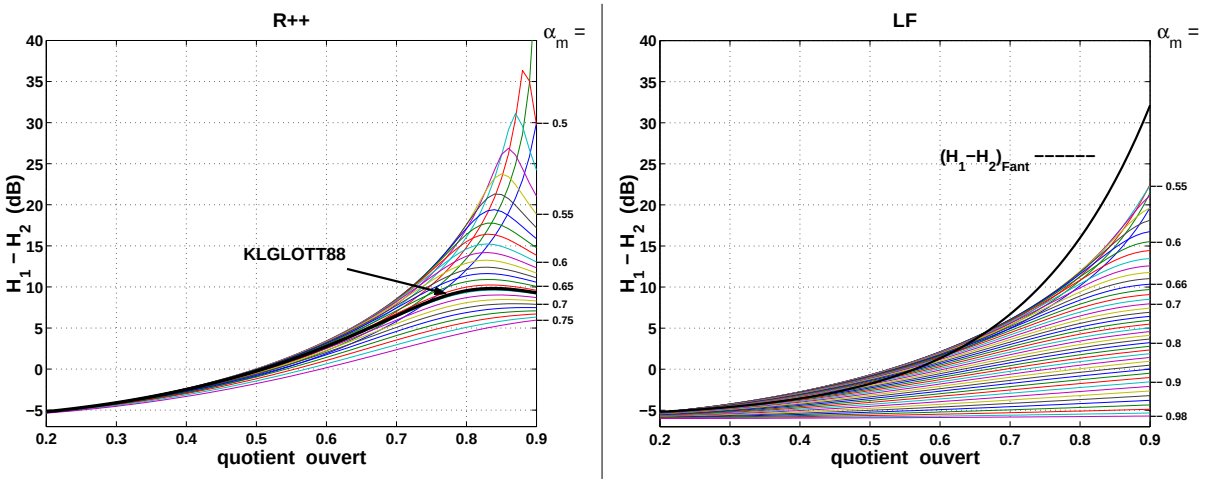


FIG. *4.2 – Différence spectrale* $(H_1 - H_2)$ *pour* O_q *variant entre* 0.2 *et* 0.9. *A gauche : cas du modèle* $R++$ *avec* α_m *variant entre* 0.5 *et* 0.75, *par pas de* 0.01 *; à droite : cas du modèle LF avec* α_m *variant entre* 0.52 *et* 0.98, *par pas de* 0.01.

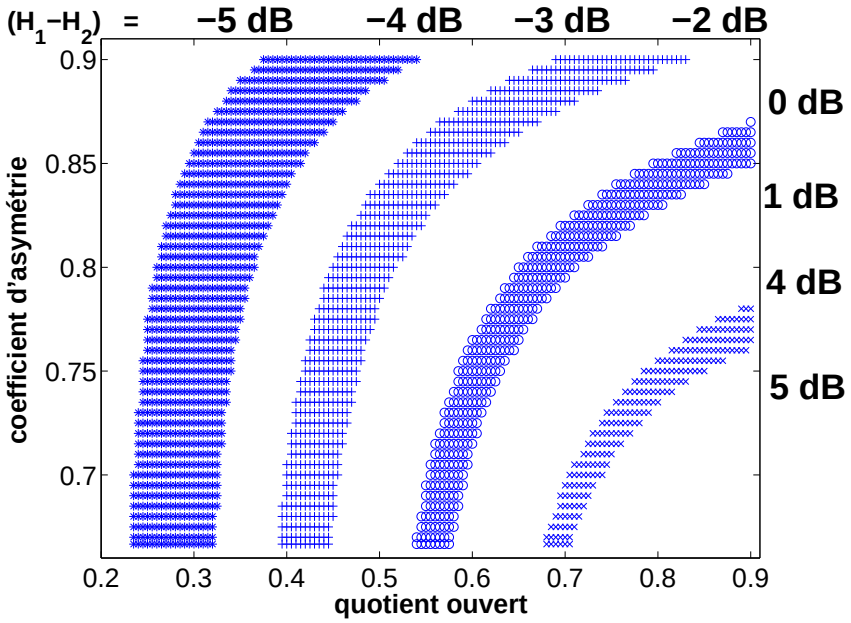


FIG. *4.3 – Domaine de variation de* O_q *et* α_m *pour différentes valeurs de* $(H_1 - H_2)$ *([-5,-4] dB ; [-3,-2] dB ; [0,1] dB ; [4,5]dB), dans le cas du modèle LF.*

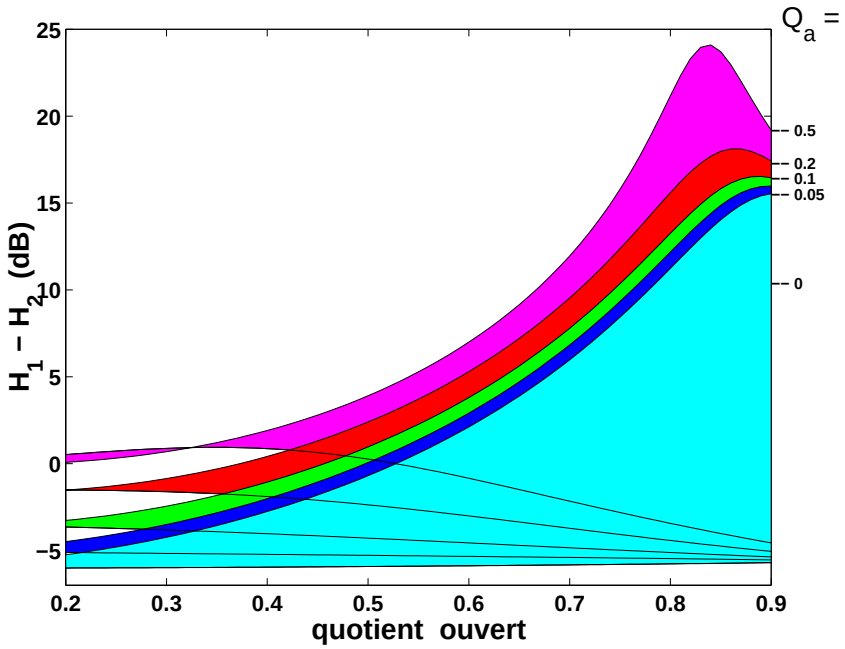


Fig. *4.4 – Domaines de variation de* $(H_1 - H_2)$ *pour* 5 *valeurs données de* Q_a *:* 0 , 0.05 , 0.1 , 0.2 *et* 0.5. *Un domaine donné correspond aux variations de* O_q *entre* 0.2 *et* 0.9 *et* α_m *entre* 0.6 *et* 0.98. *Les domaines se superposent, mais les limites minimales et maximales de chacun des domaines ont été représentées.*

$(H_1 - H_2)$ augmente, la plage de variation du quotient ouvert se déplace vers des valeurs élevées de O_q, ce qui confirme la tendance générale d'une augmentation de O_q avec $(H_1 - H_2)$. En parallèle, la plage de variation de α_m se réduit et se déplace vers des valeurs faibles de α_m. Pour des valeurs faibles de quotient ouvert, la valeur prise par le coefficient d'asymétrie ne modifie que très peu la valeur de $(H_1 - H_2)$, ce qui se traduit sur la figure 4.2 par un resserrement des abaques quand O_q diminue. Pour des valeurs élevées de quotient ouvert, la valeur de $(H_1 - H_2)$ dépend du coefficient d'asymétrie. Pour une valeur donnée de $(H_1 - H_2)$, le coefficient d'asymétrie ne peut alors pas dépasser une valeur limite supérieure : $(\alpha_m)_{max} = 0.78$ pour $(H_1 - H_2) \in [4, 5]$ dB.

Cas d'une fermeture non-abrupte

Jusqu'à présent, nous nous sommes placés dans le cadre d'une fermeture abrupte, ce qui implique la nullité du quotient de phase de retour. Nous avons vu en partie 3.2.4 qu'une modification de Q_a se répercute essentiellement sur les hautes fréquences. Néanmoins, nous avons également noté un petit effet basse fréquence. Cela laisse donc supposer que le quotient de phase de retour peut avoir une influence sur la valeur de $(H_1 - H_2)$ à O_q et α_m constants.

La figure 4.4 présente les domaines de variation de $(H_1 - H_2)$ en fonction de O_q et pour α_m variant entre 0.6 et 0.98. Les domaines correspondant aux 5 valeurs suivantes sont représentés : $Q_a =$ 0 , 0.05 , 0.1 , 0.2 , *et* 0.5. Ces domaines se superposent en partie et pour aider à la lisibilité de la figure, les bornes minimales et maximales de chacun des 5 domaines sont alors représentées. Cette figure illustre l'effet d'une modification du quotient de phase de retour sur les basses fréquences. Moins important qu'une modification du coefficient d'asymétrie, cet effet rajoute à la valeur par défaut de $(H_1 - H_2)$ (cas $Q_a = 0$) de 1 à 2 dB quand Q_a est situé entre 0.05 et 0.1, et jusqu'à 5 dB pour $Q_a = 0.5$. La figure 4.5 complète la figure 4.4 en présentant uniquement deux domaines : $Q_a = 0$ (en plein) et $Q_a = 0.2$ (en trait). La relation 4.1 proposée par Fant est également présentée sur cette figure. On observe donc que cette courbe traduit une augmentation progressive du quotient de phase de retour, quand O_q augmente.

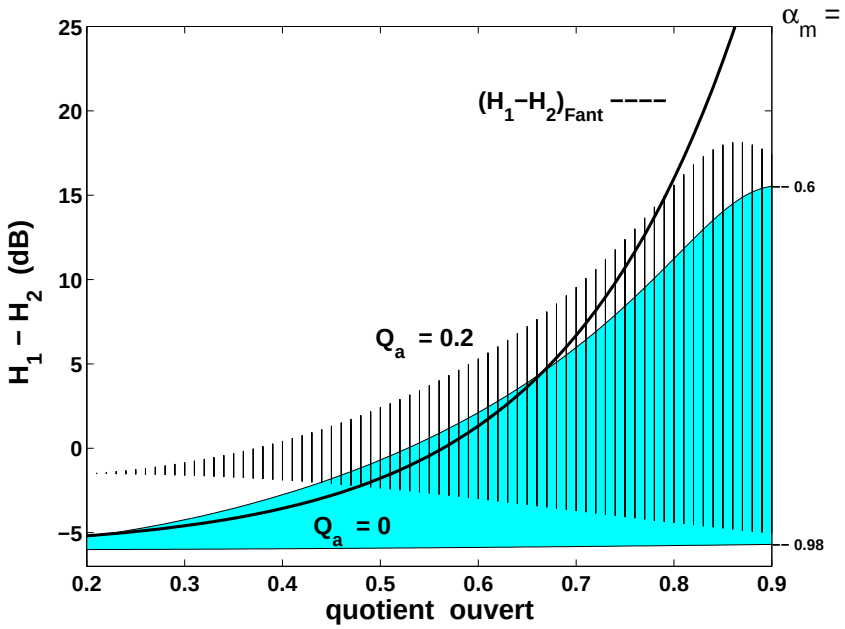


FIG. *4.5 – Domaines de variation de* $(H_1 - H_2)$ *en fonction de* O_q *et* α_m *pour* $Q_a = 0$ *(en plein) et* $Q_a = 0.2$ *(en trait). La relation empirique 4.1 proposée par Fant est également représentée.*

En effet, dans le cas où $Q_a = 0$ et α_m approche 0.6, on sort du domaine de variation de $(H_1 - H_2)$ pour $O_q = 0.66$. Seule une augmentation du quotient de phase de retour permet alors d'expliquer des mesures expérimentales de $(H_1 - H_2)$ situées hors de ce domaine (comme par exemple sur la courbe de Fant).

4.1.3 Discussion

Cette étude théorique de la différence spectrale $(H_1 - H_2)$ amène aux constatations suivantes :

1. Quand le coefficient d'asymétrie et le quotient de phase de retour sont fixés, on observe de façon générale, pour tous les modèles, une augmentation de $(H_1 - H_2)$ quand O_q augmente. Ceci se traduit spectralement par une accentuation progressive du fondamental par rapport au second harmonique. On retrouve ainsi de façon théorique un résultat que l'on observe expérimentalement : l'atténuation sensible du fondamental dans le cas d'une voix pressée, où O_q est faible $(0.4 - 0.6)$ ou, au contraire, la prédominence marquée du fondamental par rapport au second harmonique dans le cas d'une voix relâchée, où O_q est élevé $(0.7 - 0.9)$.

2. $(H_1 - H_2)$ est dépendante du modèle, comme le montre la figure 4.1 qui présente $(H_1 - H_2)$ fonction de O_q pour le modèle KLGLOTT88 (asymétrie constante) et le modèle LF avec $\alpha_m = \frac{2}{3}$ [3]. La différence sur $(H_1 - H_2)$ entre ces deux modèles va en s'accentuant quand O_q augmente. En supposant que l'on puisse déterminer O_q à partir de $(H_1 - H_2)$ à l'aide de ces modèles, les valeurs ainsi obtenues seront donc fonction du modèle choisi.

3. En étudiant $(H_1 - H_2)$ en fonction de O_q dans le cas des modèles R++ et LF, on note que cette différence spectrale devient fortement dépendante du coefficient d'asymétrie quand O_q augmente. Dans l'hypothèse où l'asymétrie de l'impulsion glottique n'est pas constante lors de la phonation, la relation entre $(H_1 - H_2)$ et O_q n'est alors plus bijective. Il existe une plage de variation de O_q

[3] Le modèle R++ n'a pas été représenté puisqu'il est égal au modèle KLGLOTT88 quand $\alpha_m = \frac{2}{3}$

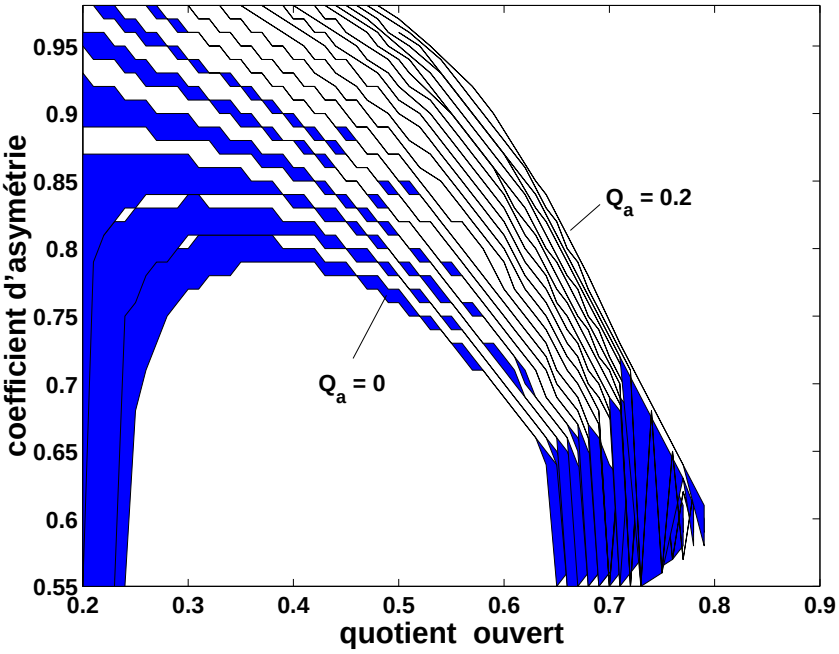


Fig. *4.6 – Calcul des valeurs possibles de α_m en fonction de O_q pour $(H_1 - H_2)$ donnée par la relation empirique de Fant et pour différentes valeurs de Q_a. Le paramètre Q_a varie entre* 0 *et* 0.2 *par pas de* 0.01. *L'incertitude sur $(H_1 - H_2)$ est de* 0.1 dB

qui amène à une même valeur de $(H_1 - H_2)$. Cette plage de variation est d'autant plus importante que O_q est élevé et une mesure spectrale du quotient ouvert devient sujette à caution.

4. Le quotient de phase de retour Q_a a un effet non négligeable sur la différence spectrale $(H_1 - H_2)$. Une augmentation de Q_a revient à translater de quelques décibels vers le haut la limite supérieure du domaine de variation de $(H_1 - H_2)$ en fonction de O_q et α_m, ainsi que l'illustre la figure 4.4. Spectralement, cela correspond à une accentuation du fondamental par rapport au second harmonique.

Dans cette étude théorique, nous n'avons formulé aucune hypothèse sur les possibles corrélations entre paramètres. Ainsi, nous avons donc supposé que les paramètres de source varient indépendamment les uns des autres. Dans la réalité, on s'attend intuitivement à une dépendance entre ces paramètres, d'autant que l'étude des signaux réels montre fréquemment une corrélation importante entre O_q et $(H_1 - H_2)$. En s'appuyant sur des mesures effectuées sur des locuteurs et à partir des équations du modèle LF, Fant en déduit la relation empirique 4.1 liant $(H_1 - H_2)$ et O_q. Une telle relation tente à montrer la dépendance entre O_q, α_m et Q_a, ainsi que l'illustre implicitement la figure 4.5. La question se pose alors de savoir quels sont les paramètres dépendants et de quelle manière ils interagissent. En se basant sur cette relation (4.1), nous calculons, pour un Q_a donné, les valeurs de α_m qui permettraient d'obtenir la différence spectrale associée à une valeur donnée de O_q. La figure 4.6 présente le résultat de ces calculs pour des valeurs de Q_a variant entre 0 et 0.2 par pas de 0.01. Pour chaque cas, nous faisons varier O_q entre 0.2 et 0.9, nous calculons $(H_1 - H_2)$ par la relation 4.1 et nous déterminons la plage de variation de α_m pour laquelle nous obtenons cette valeur de $(H_1 - H_2)$ à 0.1 dB près. Nous observons que, pour $0.25 \leq O_q \leq 0.6$, le coefficient d'asymétrie n'a pas de valeur possible en-dessous de $\alpha_m = 0.7$. Un quotient ouvert faible s'accompagne d'un coefficient d'asymétrie élevé (entre 0.7 et 0.8), ainsi que d'une phase de retour nulle ou peu marquée (Q_a entre 0 et 0.05). Un quotient ouvert plus élevé (entre 0.7 et 0.8) s'accompagne d'un coefficient d'asymétrie plus bas (entre 0.55 et 0.65) et d'un quotient de phase de retour plus élevé. Pour un quotient ouvert donné, des valeurs élevées de α_m sont possibles mais elles impliquent une augmentation parallèle

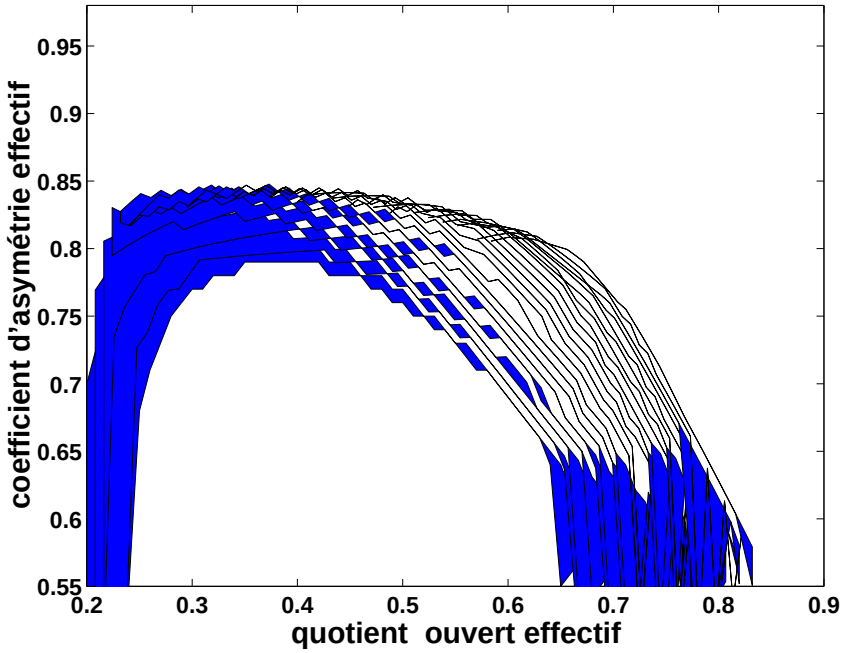


FIG. *4.7 – Calcul des valeurs possibles de* α_{me} *en fonction de* O_{q_e} *pour* $(H_1 - H_2)$ *donnée par la relation empirique de Fant et pour différentes valeurs de* Q_a*. Le paramètre* Q_a *varie entre* 0 *et* 0.2 *par pas de* 0.01*. L'incertitude sur* $(H_1 - H_2)$ *est de* 0.1 dB

de Q_a. Cela peut sembler paradoxal, car on s'attend à ce que l'asymétrie soit moins marquée dans le cas de voix relâchées qui sont caractérisées par des valeurs importantes de Q_a. Cela pourrait s'expliquer par la définition de α_m qui ne prend pas en compte la phase de retour, puisque ce coefficient est défini par rapport au temps ouvert, c'est-à-dire la durée entre l'instant d'initialisation du débit glottique et l'instant du maximum d'excitation. Pour tenir compte de la phase de retour, un terme correctif ajouté à O_q et α_m permet alors calculer les valeurs d'un quotient ouvert « effectif » O_{q_e} et d'un coefficient d'asymétrie effectif α_{me} par :

$$\begin{aligned} & O_{q_e} T_0 = O_q T_0 + Q_a (1 - O_q) T_0 \\ \Rightarrow \quad & O_{q_e} = O_q + Q_a (1 - O_q) \end{aligned}$$

$$et \quad \alpha_{me} = \alpha_m * \frac{O_q}{O_{q_e}} = \alpha_m * \frac{O_q}{O_q + Q_a (1 - O_q)}$$

La figure 4.7 complète la figure 4.6 dans le cas du quotient ouvert effectif et du coefficient d'asymétrie effectif. La courbe pour $Q_a = 0$ reste inchangée, par contre on remarque que le coefficient d'asymétrie effectif plafonne à $\alpha_{me} = 0.85$. On observe toujours que, pour un quotient ouvert donné, des valeurs élevées de α_m sont possibles si elles s'accompagnent d'une augmentation parallèle de Q_a. Néanmoins, à O_{q_e} fixé, la plage de variation de Q_a est alors plus limitée.

Qu'en est-il exactement dans la réalité ? Dans le but de mieux cerner les covariations entre cette mesure spectrale et les paramètres de source glottique dans des situations réels de production vocale, nous nous proposons à présent d'effectuer des mesures de $(H_1 - H_2)$ sur des signaux acoustiques de parole et de chant. Pour ces signaux, les informations portant sur la source glottique sont couplées avec des informations portant sur le système supraglottique. Il devient alors nécessaire de développer au préalable des méthodes d'estimation de la source glottique à partir de tels signaux. Nous allons donc à présent présenter et discuter les méthodes existantes d'estimation de la source glottique que nous appliquerons dans la suite de cette étude.

4.2 Estimation de la source glottique

L'étude de la source glottique indépendamment du conduit vocal nécessiterait de pouvoir mesurer le débit d'air au niveau de la glotte, ce qui est évidemment très difficile à réaliser. Pour contourner cette difficulté, des méthodes indirectes d'estimation du débit glottique par filtrage inverse ont été élaborées. Nous allons présenter ces méthodes, puis nous les évaluerons sur des exemples parlés et chantés.

4.2.1 Les méthodes d'estimation par prédiction linéaire

Introduction

Les méthodes d'estimation de la source glottique s'appuient sur la théorie source-filtre et l'hypothèse d'une non-interaction entre la source et le filtre. Elles tentent d'estimer l'entrée d'un système et le système lui-même, à partir de la connaissance du signal de sortie, en l'occurence le signal acoustique vocal mesuré en sortie des lèvres, et d'une hypothèse très forte sur la nature du signal d'entrée (i.e. à spectre plat). D'après l'équation 2.2, ceci est en théorie possible si l'on arrive à évaluer précisément la fonction de transfert du conduit vocal, $H(z)$ [4]. Dans ce cas, le signal de source glottique dérivé $U_g^{'}$ se déduit du signal acoustique vocal S par filtrage inverse :

$$U_g^{'}(z) = S(z).H^{-1}(z)$$

Miller [157] (1959) a été le premier à estimer la source glottique par filtrage inverse, à l'aide de filtres analogiques. Par la suite, de nombreuses études ont été menées dans le but d'établir le lien entre les signaux obtenus par filtrage inverse et le débit glottique. En particulier, des similitudes ont été montrées entre ces signaux et les signaux électroglottographiques (Rothenberg [171], Fourcin [76], Baer & al [15]) et photoglottographiques (Kitzing [130]). Le filtrage inverse peut s'appliquer au signal de pression mesuré à l'aide d'un microphone placé en sortie des lèvres ou au signal de débit mesuré à l'aide du masque de Rothenberg [170]. Avec l'avènement des techniques numériques, les méthodes d'estimation de la source par filtrage inverse se sont développées et complexifiées (Wong, Markel & Gray [217], 1979 ; Veeneman & BeMent [212], 1985 ; Krishnamurthy & Childers [134], 1986 ; Ljungqvist [146, 83], 1986 ; Liu [145], 1993 ; Alku [1], 1992 ; Alku & Vilkman [4], 1994 ; Oliveira [159, 160] 1993 ; Lu & Smith [148], 1999). La plupart de ces méthodes s'appuient sur l'hypothèse de prédiction linéaire (LP), équivalente à un filtrage linéaire, qui consiste à supposer qu'à un instant donné, le signal pourra être estimé (ou prédit) par une combinaison linéaire des échantillons précédents. Les variantes entre méthodes portent alors principalement sur la détermination d'un critère d'erreur approprié entre le signal estimé et le signal réel.

Méthodes de prédiction linéaire par autocorrélation, par covariance - Analyse DAP

Dans le domaine du traitement de la parole, la prédiction linéaire a été appliquée avec succès à partir du début des années 1970 (Atal & Hanauer [13], 1971 ; Makhoul [149, 150] , 1975 ; Markel & Gray [151]). Simple et pourtant très efficace, elle a suscité un vif intérêt en raison des nombreuses applications en analyse et synthèse de la parole, en particulier pour le codage et la transmission des signaux de parole (Atal & Hanauer [13]) ainsi que pour la synthèse à partir du texte par compression de bases de données de diphones (Dutoit [58]).

Le principe de la prédiction linéaire classiquement adoptée en traitement de la parole est détaillé en partie B de l'annexe B et les méthodes de prédiction linéaire par autocorrélation et de prédiction linéaire

[4] Dans le chapitre précédent, nous avons choisi de présenter la théorie sur les modèles d'ODG dans un cadre continu et nous avons exprimé les signaux dans le domaine fréquentiel à l'aide de la transformée de Fourier. Dans la pratique, les signaux ne sont pas continus mais discrets, échantillonnés à une fréquence Fe. Les méthodes d'estimation s'appliquent sur des signaux échantillonnés et nous exprimerons donc plutôt le spectre fréquentiel d'un signal $h(t)$ à l'aide de sa transformée en Z, $H(z)$

par covariance sont explicitées. Ces méthodes supposent une continuité spectrale, ce qui n'est pas le cas dans la pratique, puisque le signal est échantillonné. L'estimation de l'enveloppe d'un spectre discret par prédiction linéaire ne sera donc jamais exacte, car cette méthode ne prend pas en compte l'effet du repliement causé par l'échantillonnage spectral (El-Jaroudi & Makhoul [59], 1991). Ces auteurs proposent une amélioration de la prédiction linéaire dans le cas des signaux discrets par une modification du critère d'erreur. Cette méthode, appelée « modélisation tout-pôle discrète » ou analyse DAP (*Discrete All-pole Modelling*), est présentée en partie B de l'annexe B. Elle a été appliquée avec succès par Alku dans ses différentes études (Alku [2], 1992 ; Alku & Vilkman [4, 5, 6], 1994, 1996 ; Alku & al [3], 1997 ; Alku & al [7], 1999). En particulier, Alku [2] montre que cette technique permet d'atténuer les « rebonds » observés dans la phase fermée, conséquences d'une mauvaise estimation du premier formant F_1 .

Méthodes de prédiction linéaire avec modélisation de la source

Le plus gros défaut de la prédiction linéaire classique est l'hypothèse très restrictive faite sur le signal de source. En matière de codage et de transmission de données, cette hypothèse n'est pas vraiment limitative, car, même si la source est estimée de façon très simpliste, l'intelligibilité, facteur principal d'intérêt, est conservée. Par contre, la qualité vocale, contenue principalement dans la source, n'est pas prise en compte. Si ce facteur nous intéresse, comme c'est le cas dans cette étude, une modélisation convenable de la source glottique est à ajouter à cette méthode de prédiction. Il est alors nécessaire d'appliquer des procédures d'optimisation pour déterminer les valeurs adéquates des paramètres de source glottique.

Ljungqvist [146, 83] (1986), Isaksson & Millnert [115] (1986), Liu [145] (1993), Lu & Smith [148] (1999) ou très récemment Fröhlich & al [78] (2001) s'appuient sur des algorithmes d'optimisation sous contraintes des paramètres des modèles d'ODG et se rapprochent des méthodes spectrales d'Analyse-par-Synthèse (Stevens [183], Bell & al [19]). La partie B de l'annexe B présente la méthode proposée par Ljungqvist [146], que nous appliquerons par la suite dans cette étude.

4.2.2 Application à l'estimation de la source glottique par filtrage inverse

Nous allons à présent détailler les étapes de l'estimation de la source glottique par filtrage inverse, indépendamment de la méthode d'estimation de la fonction de transfert du conduit vocal considérée.

Méthode générale

La figure 4.8 présente les étapes habituelles de l'estimation de la source glottique, basée sur l'analyse par prédiction linéaire classique. Un signal voisé a été synthétisé à partir d'un filtre $H(z)$ modélisant le conduit vocal dans le cas d'un [a] : les fréquences des formants (en Hz) sont [730, 1281, 2468, 3343] et les largeurs de bande correspondantes (en Hz) sont [140, 120, 140, 200]. La fréquence fondamentale a été fixée à $f_0 = 100\ Hz$ et la fréquence d'échantillonnage à $F_e = 8000\ Hz$. Le modèle de source utilisé est le modèle de Klatt, avec $A_v = 1$ et $O_q = 0.6$. Pour effectuer l'analyse par prédiction linéaire (analyse LP), le signal voisé est au préalable accentué par un filtre tout-zéros de la forme $F(z) = 1 - az^{-1}$, où le paramètre d'accentuation a doit être choisi proche de l'unité (0.98 étant une valeur typique de ce paramètre). Cela revient à effectuer une quasi-dérivation du signal voisé, ce qui supprime une éventuelle composante continue et accentue les hautes fréquences du signal. Le but de cette pré-accentuation est de tenter d'éliminer les effets spectraux dus aux pentes de la source (en -12dB/oct) et du rayonnement aux lèvres (en $+6$dB/oct). Sur l'exemple présenté sur la figure 4.8, on remarque ainsi que les premiers harmoniques ont été fortement atténués. L'analyse LP s'effectue ensuite par application de l'une ou l'autre des deux méthodes présentées précédemment. Elle peut également être remplacée par une analyse DAP, plus lourde en temps de calcul, mais souvent plus précise. Dans le cas de la méthode par autocorrélation, le signal accentué est fenêtré avant analyse par une fenêtre de type Hanning ou Hamming, d'une durée de l'ordre de 20 ms à 40 ms. Dans le cas de la méthode par covariance, le fenêtrage est rectangulaire et,

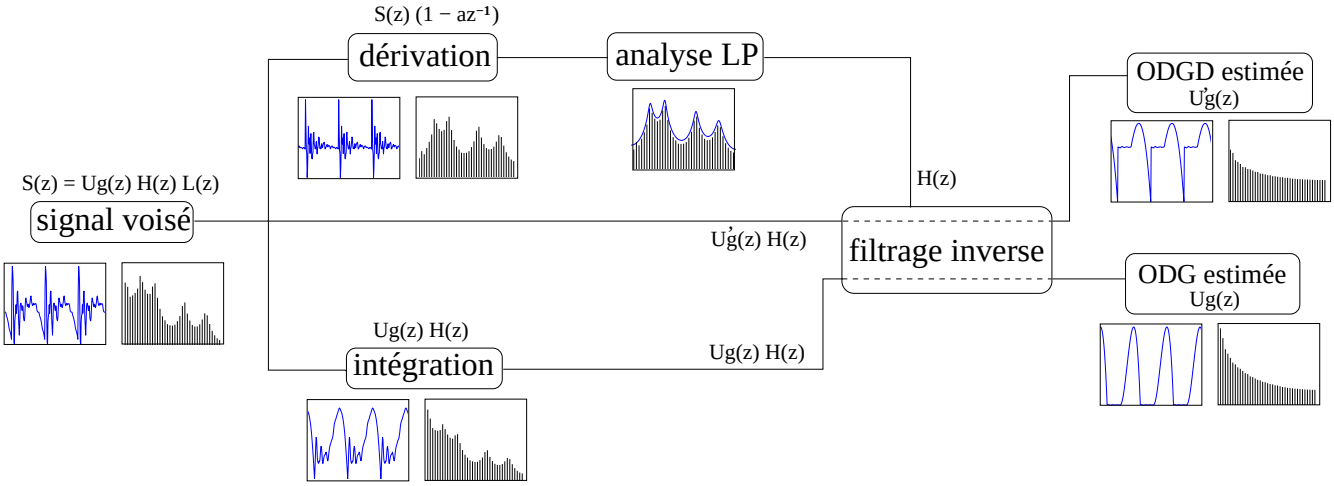


FIG. *4.8 – Etapes de l'estimation du signal de source glottique et de sa dérivée à partir d'une analyse par prédiction linéaire (analyse LP). Pour chaque étape de l'estimation, le signal temporel et le spectre associé sont présentés.*

si l'on connaît au préalable les instants de fermeture glottique, il peut être approprié de se placer dans la partie fermée (débit nul ou constant) pour effectuer l'analyse, car seul le conduit vocal intervient à ces instants. Pour cet exemple, la méthode par autocorrélation a été appliquée sur le signal accentué fenêtré sur 3 périodes. L'effet du filtre estimé par l'analyse LP est ensuite supprimé par filtrage inverse du signal voisé (estimation de l'ODGD) ou du signal voisé désaccentué (intégré), ce qui revient à supprimer l'effet de la dérivation aux lèvres (estimation de l'ODG).

Cette méthode très générale d'estimation de la source glottique peut être raffinée par l'ajout d'un processus itératif, tel que le propose Alku [1] (méthode IAIF, 1992). En effet, la pré-accentuation du signal voisé a pour but de supprimer les effets conjoints de la source et de la dérivation aux lèvres. Une première analyse permet d'estimer plus précisement la pente spectrale moyenne. On peut alors réitérer l'analyse en utilisant le résultat de la première analyse. La figure 4.9 présente un schéma de cette méthode.

Conclusion

La prédiction linéaire classique est une méthode robuste et efficace, qui peut donner de bons résultats dans les cas où les effets de la source glottique et du conduit vocal peuvent être séparés. C'est en particulier le cas des voyelles ayant un premier formant haut en fréquence, dont le [a] est un exemple typique. La méthode fonctionne bien dans le cas où la fréquence fondamentale du signal voisé est basse (de l'ordre de 100 à 200 Hz), ce qui est souvent vérifié en parole pour les hommes, mais pas en voix chantée.

Nous avons vu dans l'étude théorique présentée au chapitre 3 que l'effet de la source glottique, traduit en particulier par la position du formant glottique, peut affecter des fréquences situées bien au-dessus du premier harmonique (f_0). Dans ces cas, même pour des voyelles à premier formant élevé, la prédiction linéaire classique ne sera pas adaptée pour séparer F_1 de F_g, car ceux-ci peuvent alors être situés dans une même zone fréquentielle.

4.2.3 Evaluation des méthodes

Afin de déterminer les méthodes les plus adaptées à une bonne estimation de la source glottique, les différentes méthodes présentées dans la partie précédente ont été évaluées dans le cas idéal de signaux

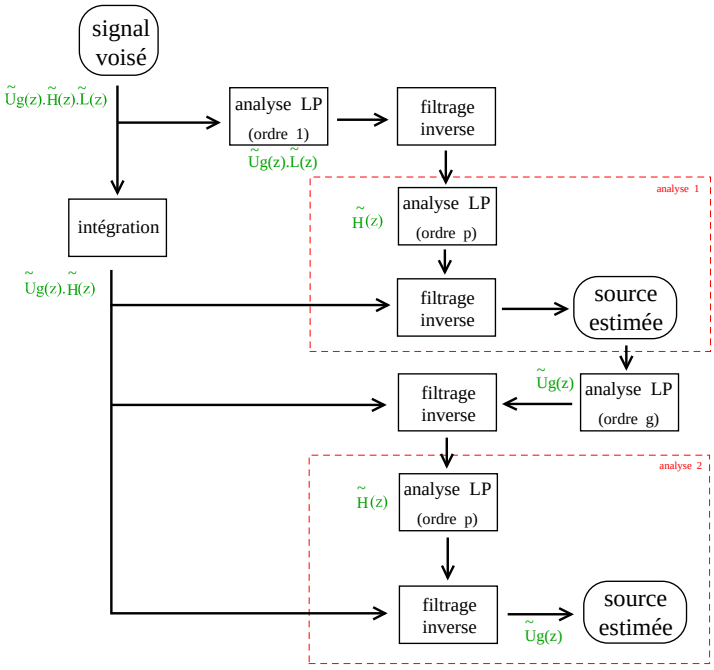


FIG. *4.9 – Estimation de la source glottique par filtrage inverse, à partir d'un processus itératif (Alku [1], 1992)*

synthétiques et sur des exemples de voix parlées et chantées.

Tests sur des signaux synthétiques

Les signaux tests ont été synthétisés à partir du modèle KLGLOTT88 (asymétrie constante). Un signal de source glottique dérivée (ODGD) est synthétisé sur une durée de 5 périodes et filtré par un filtre modélisant une voyelle du conduit vocal. Les trois voyelles explorées sont données dans le tableau 4.1, qui présente les valeurs moyennes des formants et des largeurs de bande (en Hz), mesurées pour des voix d'hommes (Childers [29], 2000, pp. 70).

Les méthodes suivantes ont été testées : prédiction linéaire par autocorrélation avec application ou non de l'algorithme itératif IAIF (Alku [1]), analyse DAP avec application ou non de l'algorithme itératif IAIF, prédiction linéaire par covariance et méthode de Ljungqvist à partir du modèle de Klatt. L'espace

VOYELLE	$\mathbf{F_1}$ (BW_1)	$\mathbf{F_2}$ (BW_2)	$\mathbf{F_3}$ (BW_3)	$\mathbf{F_4}$ (BW_4)
[a]	673 (154)	1097 (154)	2457 (242)	3464 (238)
[i]	303 (139)	2172 (156)	2851 (281)	3572 (227)
[u]	342 (134)	1067 (144)	2219 (191)	3342 (188)

TAB. *4.1 – Valeurs moyennes des fréquences et largeurs de bande des formants dans le cas des 3 voyelles anglaises [a], [i] et [u], déterminées à partir de mesures sur des voix d'hommes (Childers [29], 2000, pp. 70)*

Voyelle		LP cor	LP cor (IAIF)	DAP	DAP (IAIF)	LP cov	Met. LJUNG.
[a]	E_{F_1}	2.8	4.2	2.0	2.6	2.1	2.5
	(E_{BW_1})	(17)	(25)	(37)	(36)	(27)	(14)
	E_{F_2}	1.9	2.4	1.0	1.1	2.2	1.4
	(E_{BW_2})	(14)	(15)	(33)	(33)	(18)	(15)
	E_{F_3}	1.3	1.2	0.8	0.8	0.3	0.9
	(E_{BW_3})	(31)	(38)	(25)	(26)	(6)	(24)
	E_{F_4}	0.2	0.2	0.1	0.1	0.1	0.3
	(E_{BW_4})	(15)	(21)	(18)	(20)	(9)	(8)
[i]	E_{F_1}	9.6	9.6	11.1	10.2	9.3	3.5
	(E_{BW_1})	(30)	(42)	(39)	(69)	(38)	(13)
	E_{F_2}	1.0	1.1	0.3	0.6	0.3	0.5
	(E_{BW_2})	(36)	(58)	(28)	(50)	(8)	(24)
	E_{F_3}	0.4	7.1	0.2	2.1	1.2	2.6
	(E_{BW_3})	(15)	(30)	(16)	(31)	(18)	(14)
	E_{F_4}	0.3	0.1	0.4	0.7	0.3	0.2
	(E_{BW_4})	(13)	(23)	(19)	(35)	(13)	(13)
[u]	E_{F_1}	8.5	9.0	7.2	8.8	7.2	3.2
	(E_{BW_1})	(27)	(37)	(50)	(49)	(50)	(18)
	E_{F_2}	1.9	1.9	1.2	1.0	0.9	1.1
	(E_{BW_2})	(19)	(21)	(26)	(29)	(15)	(14)
	E_{F_3}	1.8	1.9	1.6	1.5	0.2	1.5
	(E_{BW_3})	(27)	(28)	(25)	(23)	(6)	(19)
	E_{F_4}	0.8	0.6	0.3	0.4	0.07	0.6
	(E_{BW_4})	(34)	(27)	(18)	(24)	(5)	(30)

Tab. *4.2 – Erreurs relatives en % sur l'estimation des fréquences et des largeurs de bande des formants, dans le cas des 3 voyelles [a], [i] et [u]. Les méthodes testées sont : la prédiction linéaire classique par autocorrélation (LP cor) avec ou sans l'algorithme itératif IAIF développé par Alku [1], l'analyse DAP (avec ou sans l'algorithme itératif), la prédiction linéaire classique par covariance (LP cov), ainsi que la méthode proposée par Ljungqvist [146]*

des fréquences a été balayé de $f_0 = 100\ Hz$ à $f_0 = 300\ Hz$ par pas de 50 Hz et l'espace des quotients d'ouverture de $O_q = 0.4$ à $O_q = 0.8$ par pas de 0.01. L'amplitude de voisement et le coefficient d'asymétrie sont gardés constants. La fréquence d'échantillonnage vaut $F_e = 8000\ Hz$. Les résultats portant sur l'estimation des formants et largeurs de bande sont donnés dans le tableau 4.2. Les erreurs relatives (en %) sur l'estimation de la fréquence des formants (E_{F_i}) et des largeurs de bande (E_{BW_i}) sont calculées par :

$$E_{F_i} = \frac{|\widetilde{F_i} - F_i|}{F_i} * 100$$

$$E_{BW_i} = \frac{|\widetilde{BW_i} - BW_i|}{BW_i} * 100$$

où $\widetilde{F_i}$ et $\widetilde{BW_i}$ sont les valeurs estimées de la fréquence et de la largeur de bande du formant F_i, et moyennées sur l'ensemble des erreurs obtenues pour les variations de f_0 et O_q.

Nous remarquons que, dans tous les cas, l'estimation des formants élevés est meilleure. Ces méthodes ne fournissent donc pas une bonne estimation du premier formant F_1, en particulier dans le cas des voyelles ([i] et [u]) où ce formant est à une fréquence très basse. Cela traduit le fait que la prédiction linéaire est plus efficace dans les hautes fréquences. Pour ces voyelles, la prédiction linéaire avec modèle de source (méthode développée par Ljungqvist) donne de bons résultats sur l'estimation du premier formant, comparativement aux autres méthodes. Elle semble effectivement plus adaptée que les méthodes classiques pour séparer la source du filtre. Les méthodes par processus itératif (IAIF) n'améliorent pas l'estimation.

La prédiction linéaire par covariance donne des résultats mitigés. En particulier, elle estime mal le premier formant dans le cas des voyelles [i] et [u]. Néanmoins, ces résultats sont des moyennes pour des fréquences variant entre 100 et 300 Hz . La prédiction linéaire par covariance donne d'excellents résultats pour les fréquences basses, mais elle devient moins performante quand f_0 se rapproche de 300 Hz .

Tests sur des signaux réels de parole

Les 6 méthodes testées précédemment sont également appliquées à l'analyse de signaux voisés réels, dans le cas des 3 voyelles retenues précédemment. Les données analysées proviennent des enregistrements audio et EGG effectués par Childers [29] (2000).

La figure 4.10 présente le résultat de l'analyse de quelques périodes extraites d'une voyelle [a] parlée par le locuteur M03. En particulier, cette figure présente le signal voisé original, l'ODGD estimée ($u_n^{'}$), l'ODG estimée (u_n) et l'erreur résiduelle e_n donnée par l'équation B.5. Pour cet exemple, nous observons une grande similarité entre les résultats obtenus à partir des différentes méthodes. Cela s'explique par le fait que les fonctions de transfert estimées sont très semblables.

Des différences notables entre les méthodes apparaissent sur l'analyse des voyelles ayant un premier formant bas. La figure 4.11 présente le résultat de l'analyse dans le cas du [i]. Si nous comparons les signaux u_n, l'analyse DAP semble donner le résultat le plus satisfaisant, puisque nous pouvons observer nettement la phase fermée, caractérisée par un minimum du débit. Néanmoins, en regardant la fonction de transfert estimée par DAP (ou DAP IAIF), nous remarquons que les deux premiers formants sont beaucoup trop résonants. Des différences se retrouvent également dans l'analyse de la voyelle [u], présentée sur la figure 4.12.

Tous ces exemples illustrent la difficulté d'obtenir une ODG ou une ODGD proche des modèles paramétriques actuellement utilisés pour les appréhender (voir chapitre 3). Quelle que soit la méthode, on a du mal à obtenir des signaux montrant une phase fermée nettement marquée. Malgré ses limitations théoriques, la prédiction linéaire semble donner des résultats tout à fait acceptables et même parfois meilleurs que des méthodes plus sophistiquées comme l'analyse DAP itérative ou la prédiction linéaire par covariance.

Nous touchons ici à la difficulté d'appréhender la source glottique par une méthode directe qui ne soit pas biaisée par le système supraglottique. Il nous paraît bien difficile d'effectuer des mesures directes des paramètres de source glottique sur de tels signaux, quand :

1. la phase fermée, à débit minimal, est souvent peu marquée et il est donc difficile de localiser exactement les instants d'ouverture et de fermeture glottique.
2. selon la méthode employée, l'aspect temporel du signal estimé varie considérablement.

Ces difficultés sont-elles liées aux limitations des modèles ou à la méthode de mesure ? Les techniques de filtrage inverse sont appliqués ici sur des signaux de pression mesurés à quelques cm des lèvres par un microphone. Dans la littérature, les signaux ODG estimés par filtrage inverse des signaux de débit mesurés à l'aide du masque de Rothenberg semblent souvent plus proches des formes théoriques attendues (voir par exemple les résultats obtenus par Holmberg & al [109, 110]). Une étude comparative poussée entre ces deux techniques de mesure (la mesure du débit en sortie des lèvres à l'aide du masque de Rothenberg et la mesure de pression à quelques cm des lèvres à l'aide d'un microphone de pression) permettrait d'apporter des éléments de réponse à cette question. Il serait également intéressant de regarder plus en détail les propriétés du rayonnement aux lèvres et de la propagation du son par ondes sphériques dans le milieu ambiant, afin de déterminer, du point de vue du signal, si la modélisation de l'effet du rayonnement par une dérivation, proposée en 1965 par Flanagan [73], ne pourrait pas être affinée, ce qui pourrait peut-être permettre une amélioration du filtrage inverse.

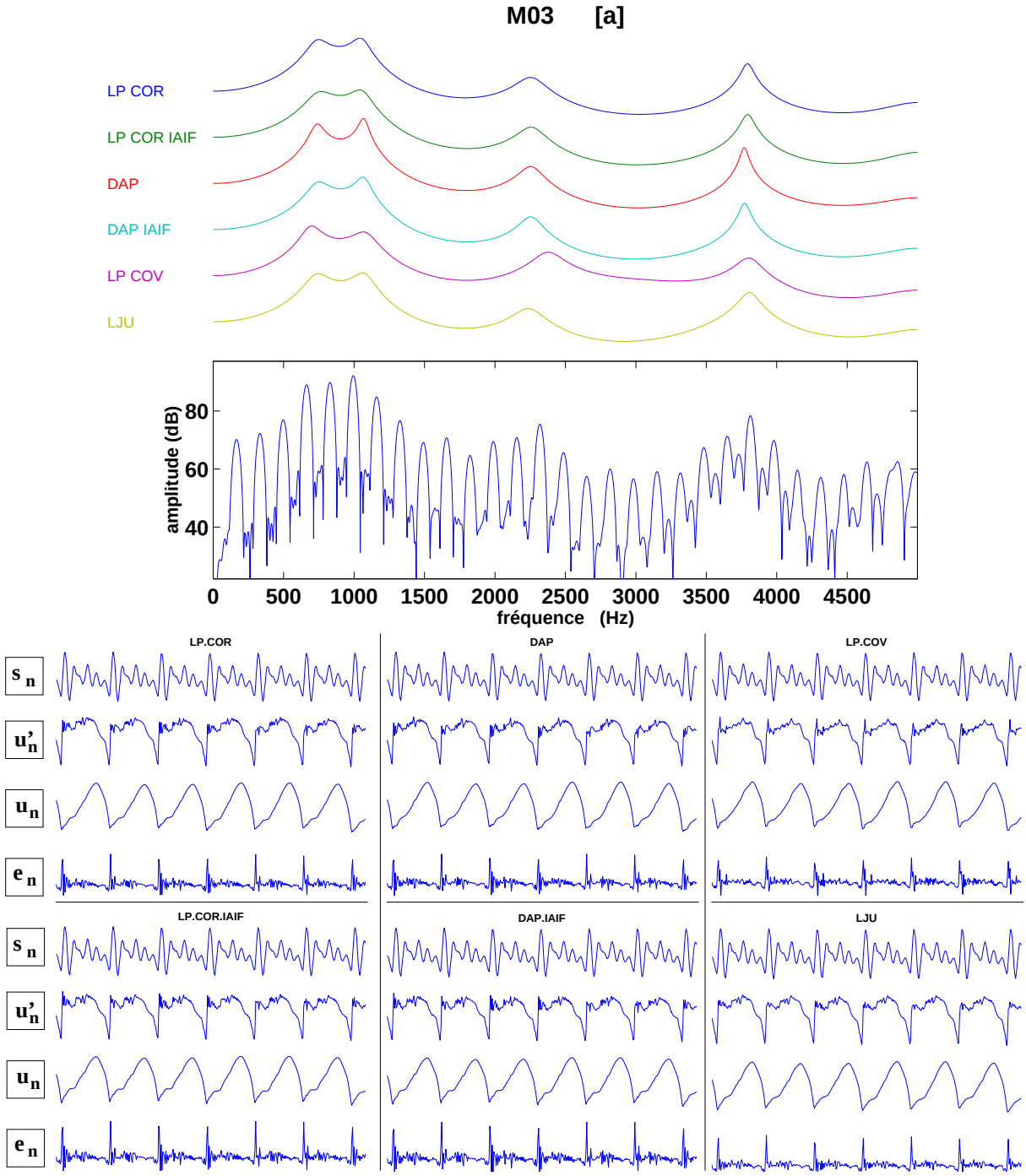


FIG. *4.10 – Analyse de la voyelle [a] parlée par le sujet M03 (cf. Childers [29]). La figure du haut présente les fonctions de transfert estimées dans le cas des 6 méthodes suivantes : prédiction linéaire par autocorrélation (LP COR), avec application de l'algorithme itératif IAIF développé par Alku [1] (LP COR IAIF), analyse DAP, avec application de l'algorithme itératif (DAP IAIF), prédiction linéaire par covariance (LP COV), ainsi que la méthode proposée par Ljungqvist [146] (LJU). La figure du bas présente le signal s_n analysé, les signaux $u_n^{'}$ (ODGD) et u_n estimés par filtrage inverse, ainsi que l'erreur résiduelle e_n.*

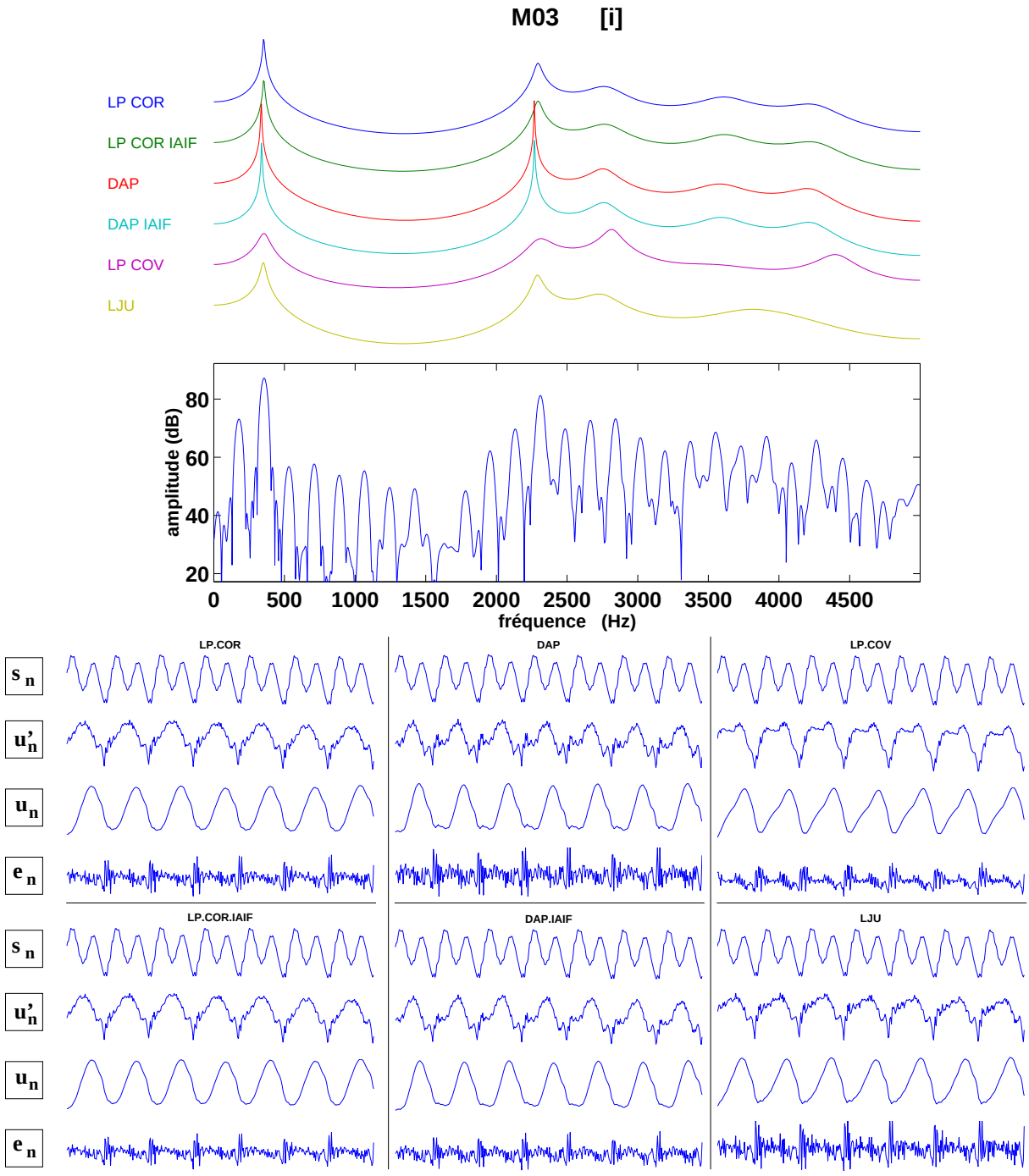


FIG. *4.11 – Analyse de la voyelle [i] parlée par le sujet M03 (cf. Childers [29]). La figure du haut présente les fonctions de transfert estimées dans le cas des 6 méthodes suivantes : prédiction linéaire par autocorrélation (LP COR), avec application de l'algorithme itératif IAIF développé par Alku [1] (LP COR IAIF), analyse DAP, avec application de l'algorithme itératif (DAP IAIF), prédiction linéaire par covariance (LP COV), ainsi que la méthode proposée par Ljungqvist [146] (LJU). La figure du bas présente le signal* s_n *analysé, les signaux* u'_n *(ODGD) et* u_n *estimés par filtrage inverse, ainsi que l'erreur résiduelle* e_n.

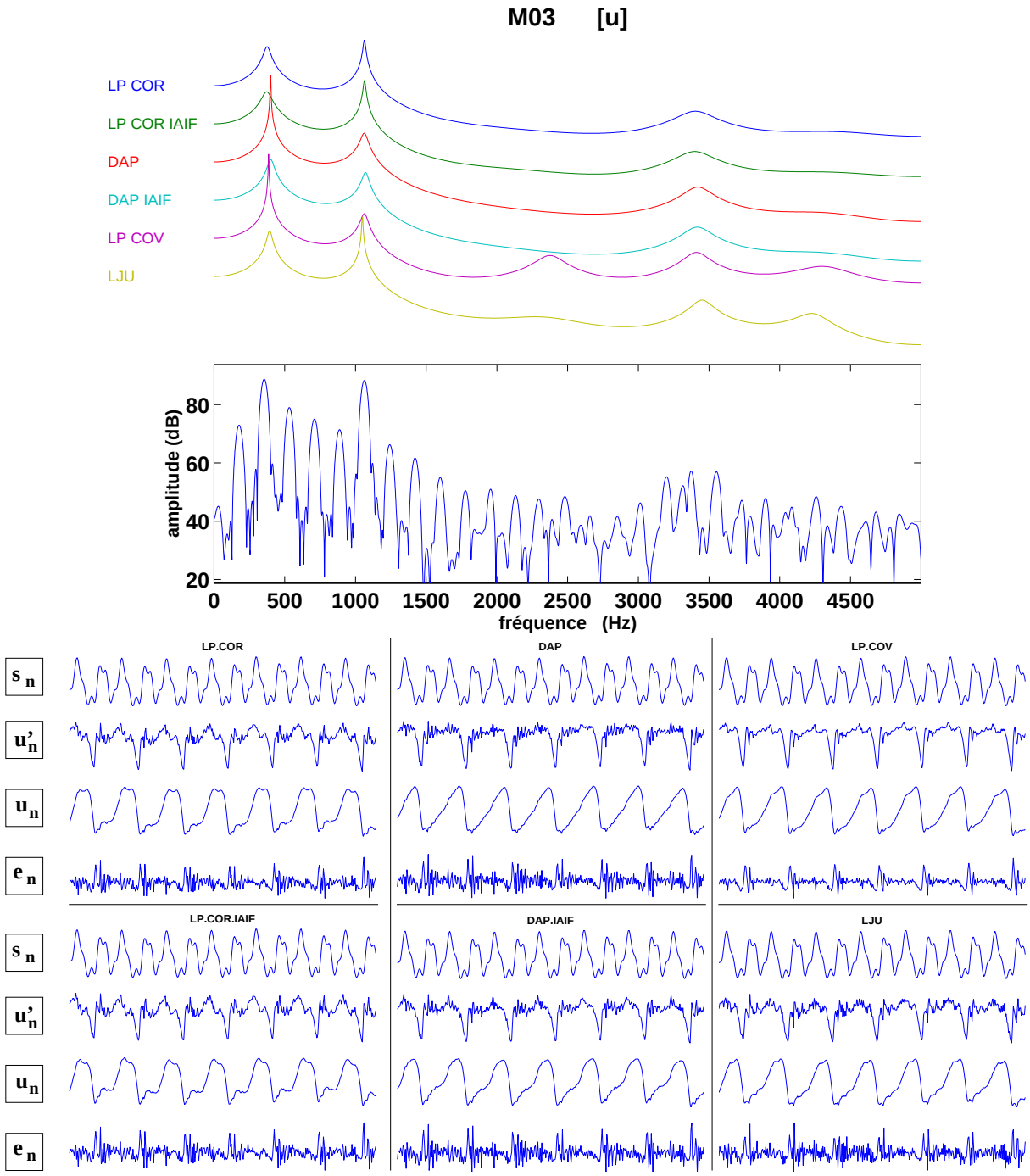


FIG. *4.12 – Analyse de la voyelle [u] parlée par le sujet M03 (cf. Childers [29]). La figure du haut présente les fonctions de transfert estimées dans le cas des 6 méthodes suivantes : prédiction linéaire par autocorrélation (LP COR), avec application de l'algorithme itératif IAIF développé par Alku [1] (LP COR IAIF), analyse DAP, avec application de l'algorithme itératif (DAP IAIF), prédiction linéaire par covariance (LP COV), ainsi que la méthode proposée par Ljungqvist [146] (LJU). La figure du bas présente le signal s_n analysé, les signaux $u_n^{'}$ (ODGD) et u_n estimés par filtrage inverse, ainsi que l'erreur résiduelle e_n.*

Application au chant

La figure 4.13 présente le résultat de l'analyse d'une voyelle [a] chantée par le baryton B1 (cf. chapitre 6). L'instant d'analyse se situe en début d'un crescendo et correspond à une voix mezzo forte. Lors d'une première analyse, l'ODG estimée est comparable à un signal ODGD, ce qui laisserait à penser que dans le chant, la pente spectrale de la source glottique n'est plus en -12dB/oct mais en -6dB/oct. Le signal voisé a donc été au préalable désaccentué pour supprimer cet effet. Nous obtenons ici une forme d'ODG (signal u_n) proche de ce que l'on attendrait en théorie. Néanmoins, nous avons remarqué que l'analyse des signaux chantées était moins aisée que celle des signaux de parole et qu'il était souvent très difficile de supprimer les « rebonds », c'est-à-dire les effets des formants du conduit vocal. Nous remarquons également que le signal d'erreur e_n est plus bruité que dans le cas de la parole, et, en particulier, les impulsions n'apparaissent pas nettement, alors qu'elles sont bien visibles dans l'analyse des voyelles parlées (figures 4.10, 4.11 et 4.12).

Nous allons à présent appliquer ces méthodes d'estimation de la source glottique à l'étude de la différence spectrale $(H_1 - H_2)$ sur des exemples de voix parlée et chantée.

4.3 Mesures expérimentales de $(H_1 - H_2)$

4.3.1 Méthodes de mesure

2 méthodes différentes permettent la mesure de $(H_1 - H_2)$ sur des signaux voisés réels.

La première méthode consiste à estimer l'ODGD par filtrage inverse, à l'aide d'une des méthodes présentées précédemment. Les instants d'ouverture et de fermeture glottique, dont la connaissance est nécessaire pour la prédiction linéaire par covariance et la méthode de Ljungqvist, sont déterminés à partir des signaux électroglottographiques dérivés [5]. Les amplitudes de H_1 et H_2 sont ensuite mesurées sur le spectre de l'ODGD.

La seconde méthode a été proposée par Hanson [97] (1995). Elle consiste à éviter le filtrage inverse en mesurant directement H_1 et H_2 sur le spectre du signal voisé. Ces amplitudes sont accentuées par rapport à celles qui seraient mesurées sur le spectre du signal ODGD, en raison de l'influence spectrale du premier formant sur ces harmoniques. Un terme correctif, appliqué sur H_1 et H_2, compense cet effet. Les amplitudes corrigées des harmoniques H_i^* sont données, pour $i = 1$ ou 2, par :

$$H_i^* = H_i - 20 * \log_{10}(\frac{F1^2}{F1^2 - (i * f_0)^2})$$

Cette correction s'applique sous l'hypothèse que la largeur de bande du premier formant est réduite et est valable tant que la fréquence des deux harmoniques est inférieure de 100 Hz au moins à la fréquence du premier formant F_1.

Nous allons présenter les résultats des analyses sur des exemples de voix parlée et chantée. Ces exemples ont été choisis car ils présentent des variations de quotient ouvert. Dans le cadre de la parole, le quotient ouvert est lié à l'aspect « tendu » d'une voix et nous illustrerons donc le lien entre O_q et $(H_1 - H_2)$ sur l'exemple d'une voyelle émise de façon relâchée à pressée par un locuteur. Dans le cadre du chant, nous avons vu au chapitre 2 de la partie préliminaire que le quotient ouvert peut varier avec l'intensité vocale[6]. Nous avons donc sélectionné comme exemples chantés des crescendos en mécanisme I et en mécanisme II.

[5] Nous reviendrons en détail sur ces signaux et sur leur utilité pour la détection précise des instants d'ouverture et de fermeture glottique au chapitre 7

[6] Nous aborderons ce point très en détail au chapitre 8, dans le cadre de mesures expérimentales de quotient ouvert à partir de signaux électroglottographiques.

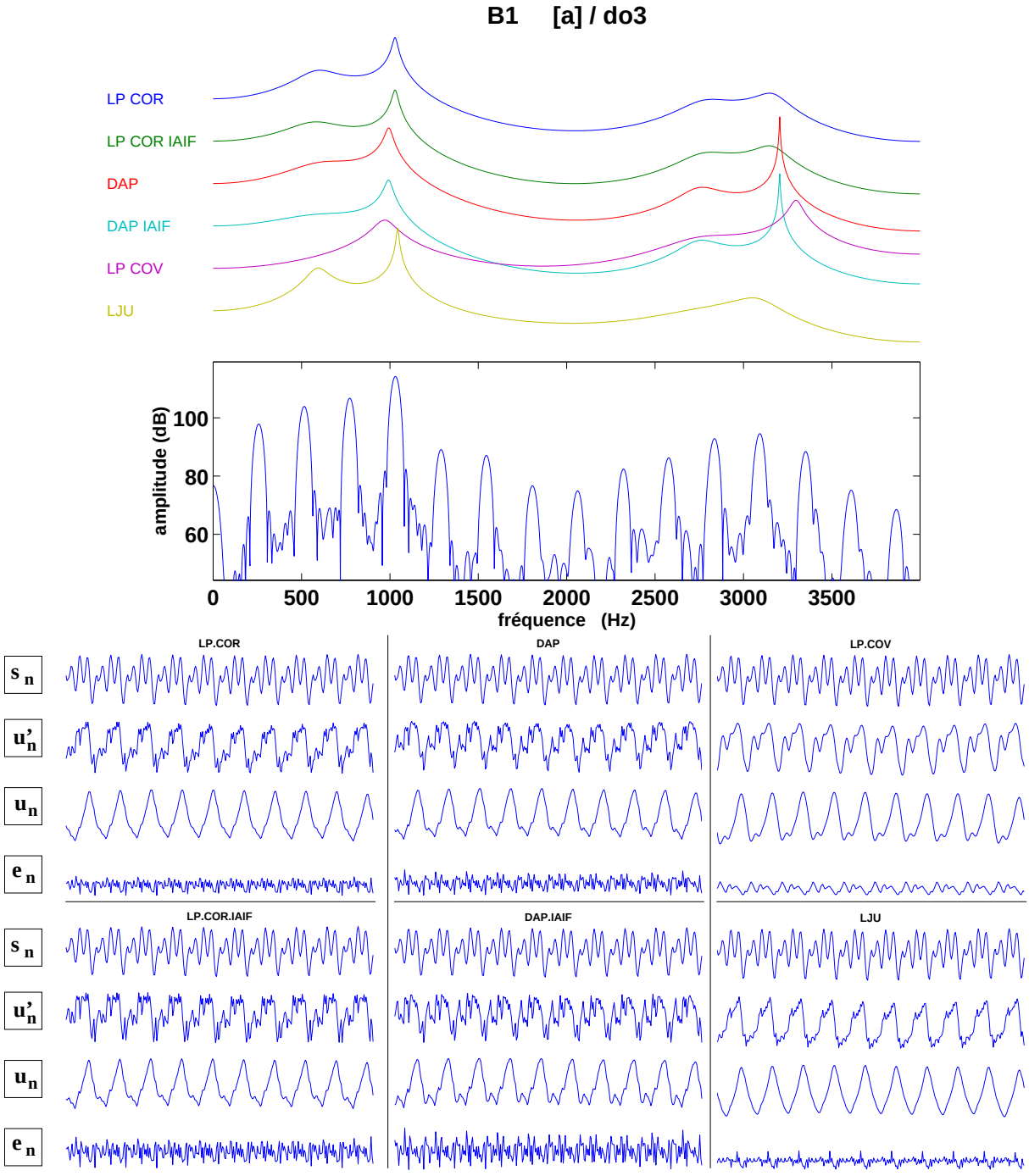


Fig. *4.13 – Analyse d'une voyelle chantée par le baryton B1, extraite du début d'un crescendo chanté sur la voyelle [a]. $f_0 = 260\ Hz$ (do3).*

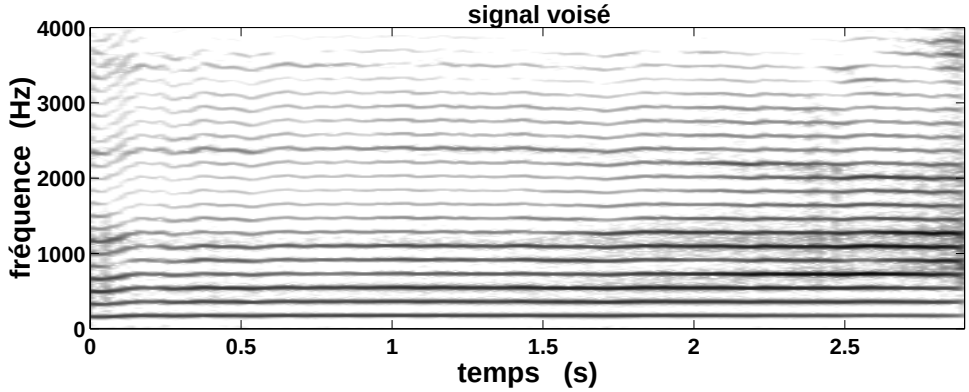


FIG. *4.14 – Voyelle émise de façon relâchée à pressée par un locuteur. (CD, piste [12])*

Nous présenterons $H_1^* - H_2^*$ en fonction du quotient ouvert mesuré sur les signaux DEGG (la méthode appliquée pour les mesures de quotient ouvert est présentée en détail dans la partie 7). A chaque fois, nous tracerons sur ce même graphe les bornes du domaine de variation de $(H_1 - H_2)$ en fonction de O_q et α_m : la borne supérieure en pointillés et la borne inférieure en trait plein correspondront à $Q_a = 0$, la borne supérieure en trait plein correspondra à $Q_a = 0.2$. Nous tracerons également la courbe de variation de $(H_1 - H_2)$ en fonction de O_q donnée par l'équation 4.1. Nous donnerons pour chaque exemple les valeurs de f_0 mesurées spectralement et mesurées sur les signaux DEGG, ainsi que l'évolution sur la phrase analysée de la fréquence F_1 et de la largeur de bande BW_1 du premier formant.

4.3.2 Application à la parole

L'exemple présenté sur la figure 4.14 correspond à une voyelle émise de façon relâchée à pressée par un locuteur.

Le filtrage inverse a été effectué par application de la méthode de Ljungqvist avec modèle de Klatt. Pour ce faire, le signal DEGG enregistré simultanément a été calé temporellement par rapport au signal acoustique afin de compenser le temps de propagation de l'onde acoustique entre la glotte et le microphone de pression. Sur une fenêtre d'analyse (40 ms), les instants d'ouverture glottique ont été détectés et une période commençant à un instant d'ouverture a été sélectionnée. Le signal ODGD a été estimé sur cette fenêtre, puis une mesure de $(H_1 - H_2)$ a été réalisée.

Nous observons tout d'abord que le quotient ouvert diminue effectivement au cours du geste vocal, passant de 0.7 (voix très relâchée, en début de phrase) à 0.4 (voix très tendue, en fin de phrase). En particulier, la figure 4.16 montre que les mesures de quotient ouvert effectuées sur les signaux DEGG sont comparables à celles effectuées à partir de la méthode de Ljungqvist [146, 83] [7]. Cette diminution notable du quotient ouvert semble également très fortement corrélée avec une diminution de la différence spectrale $(H_1 - H_2)$. Ces résultats sont donc en accord avec les résultats obtenus par Holmberg & al [111], Fant [69, 70] et Sundberg & al [190]. Nous observons également que les mesures suivent de façon relativement proche la courbe empirique proposée par Fant [69, 70] et représentée en trait gras sur la figure 4.15. Nous pouvons noter également que les deux méthodes de mesure de $(H_1 - H_2)$ (par filtrage inverse ou par correction formantique) donnent des résultats similaires.

Ces résultats laissent donc à penser que, dans le cadre de la parole, il est possible d'estimer le quotient ouvert à l'aide de la mesure spectrale $(H_1 - H_2)$, en appliquant par exemple la relation proposée par Fant, qui rend visiblement compte des dépendances entre les paramètres.

[7] La correspondance entre les mesures par électroglottographie et par filtrage inverse sera étudiée plus en détail au chapitre 7

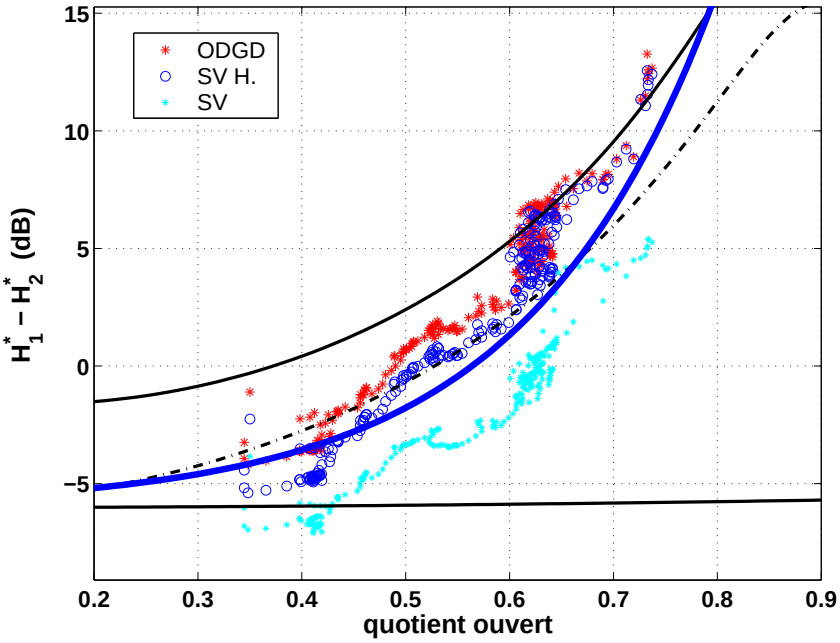


Fig. *4.15 – Mesures de $(H_1 - H_2)$ en fonction de O_q dans le cas d'une voyelle émise de façon relâchée à pressée. Les bornes du domaine théorique de variation de $(H_1 - H_2)$ ont été représentées en trait plein : la limite inférieure correspond à $\alpha_m = 0.98$ et $Q_a = 0$ et la limite supérieure à $\alpha_m = 0.6$ et $Q_a = 0.2$. La courbe en pointillés correspond à la limite du domaine théorique pour $Q_a = 0$ ($\alpha_m = 0.6$). La courbe en trait plein gras correspond à la relation empirique proposée par Fant [69, 70]. ODGD : mesures effectuées sur le signal ODGD obtenu par filtrage inverse, SV H. : mesures effectuées sur le signal voisé avec application d'un terme correctif (méthode de Hanson [97, 98]), SV : mesures effectuées sur le signal voisé sans correction formantique.*

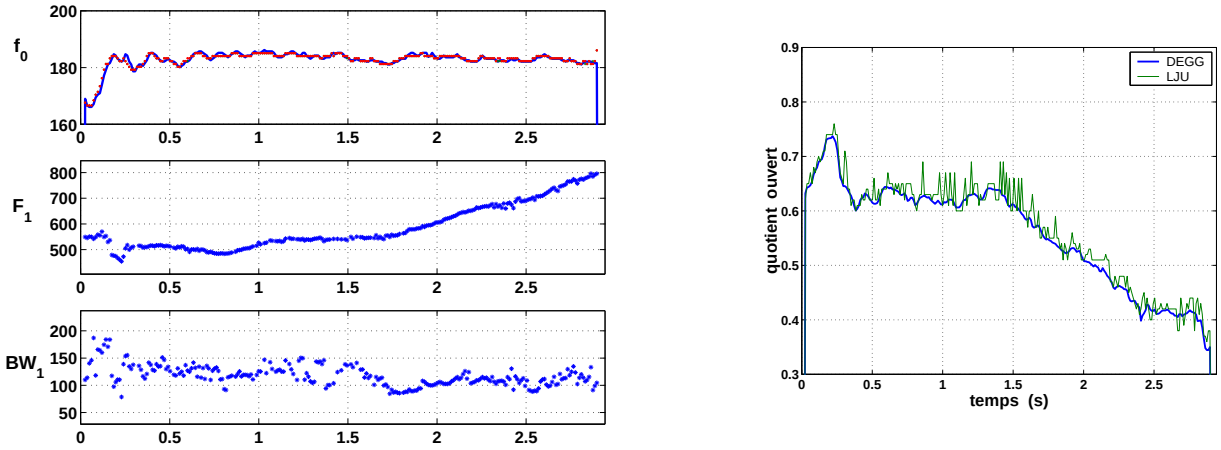


Fig. *4.16 – Analyse d'une voyelle émise de façon relâchée à pressée. Figure de gauche : estimation de la fréquence fondamentale f_0, de la fréquence du 1^{er} formant F_1 et de sa largeur de bande BW_1 sur le signal d'ODGD estimé. Figure de droite : Comparaison entre les mesures de O_q effectuées sur les signaux DEGG et celles obtenues lors de la procédure d'optimisation.*

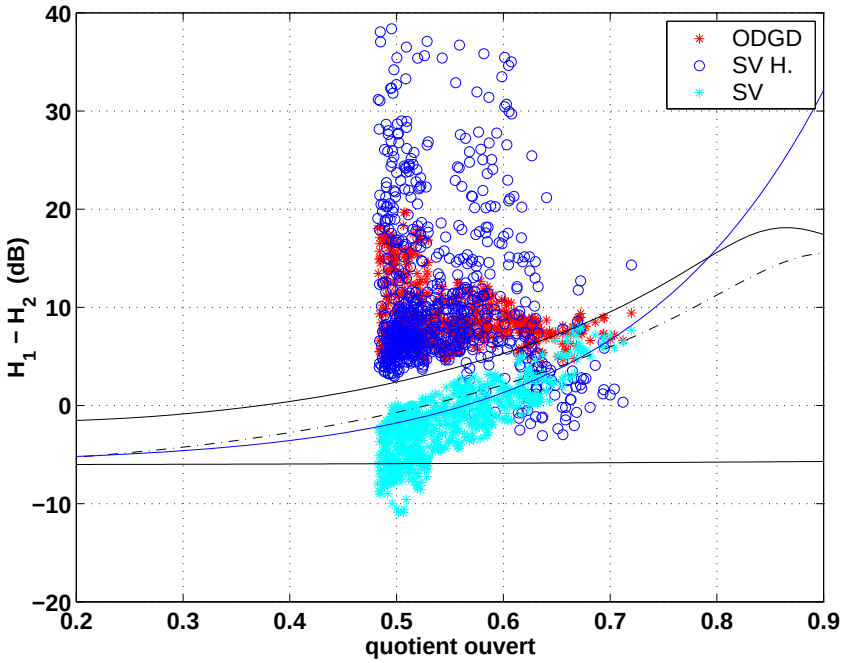


FIG. *4.17 – Analyse de* $(H_1 - H_2)$ *en fonction de* O_q *sur un* **crescendo chanté par le baryton B1, sur un do3 (**$f_0 = 260\ Hz$**) et la voyelle [a], en mécanisme I**. *Les valeurs comprises dans le domaine théorique de variation de* $(H_1 - H_2)$ *correspondent aux mesures non corrigées effectuées sur le signal voisé (SV). Les autres mesures, par la méthode de correction proposée par Hanson* [97, 98] *(SV H.) ou par filtrage inverse (ODGD), se superposent en partie et sont situées à l'extérieur du domaine théorique. Pour le détail des courbes, voir la légende de la figure 4.15.*

4.3.3 Etude de $(H_1 - H_2)$ dans le cas du chant

Trois exemples sont présentés dans cette partie. Le filtrage inverse a été effectué par prédiction linéaire (méthode d'autocorrélation). Les deux premiers exemples correspondent à des productions vocales en mécanisme I et le dernier exemple à une production vocale en mécanisme II. Ces exemples sont représentatifs des résultats obtenus par analyse d'autres productions vocales chantées, qui ne sont pas présentés ici.

Production vocale en mécanisme I

Le premier exemple proposé correspond à l'analyse d'un crescendo chanté par un baryton (B1, voir chapitre 6). Un exemple de l'ODG estimée par filtrage inverse a été présenté en partie 4.2 pour ce même signal. La figure 4.17 montre les valeurs de $(H_1 - H_2)$ en fonction des valeurs de O_q mesurées sur le signal DEGG associé. Nous présentons 3 mesures de la différence spectrale $(H_1 - H_2)$: premièrement, sur le signal d'ODGD obtenu par filtrage inverse (ODGD), deuxièmement sur le signal voisé avec correction (SV. H) et troisièmement sur le signal voisé sans correction (SV). Nous observons sur cet exemple que les valeurs de $(H_1 - H_2)$ mesurées sur le signal obtenu par filtrage inverse ou sur le signal voisé avec correction sont relativement semblables et elles varient dans le même domaine, qui est par ailleurs très éloigné du domaine théorique. Seules les mesures effectuées sur le signal voisé sans correction sont situées dans le domaine théorique. Nous notons également que les valeurs mesurées par la méthode proposée par Hanson [97, 98] sont plus dispersées, tandis que les valeurs mesurées sur l'ODGD ou sans correction sont regroupées pour un quotient ouvert donné dans une bande de variation de 3 à près de 10 dB.

Le second exemple correspond à l'analyse d'un crescendo chanté par le ténor T2, sur la voyelle [a]

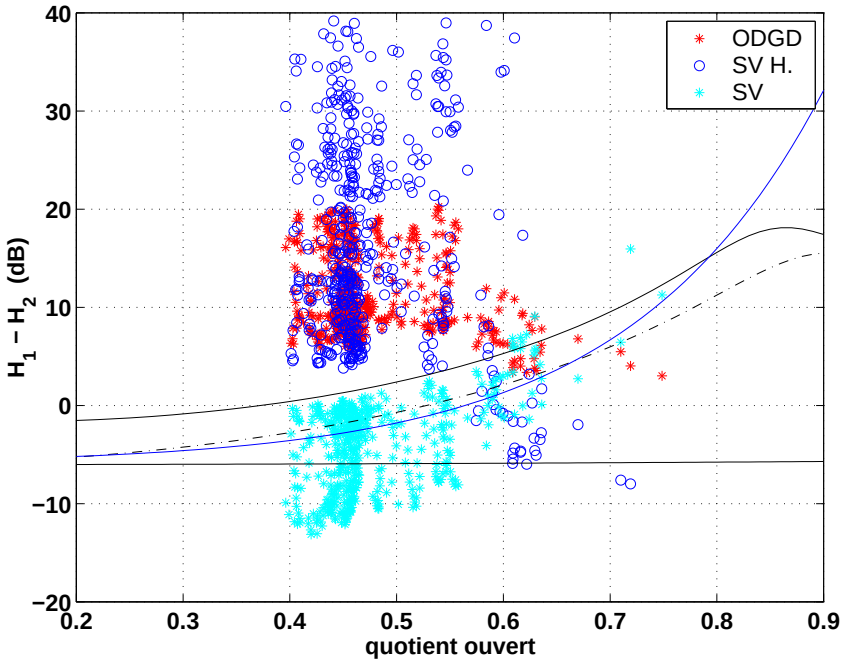


Fig. *4.18 – Mesures de $(H_1 - H_2)$ en fonction de O_q, sur un* **crescendo chanté par le ténor T2, sur un ré3 ($f_0 = 293\ Hz$) et la voyelle [a], en mécanisme I**. *ODGD : mesures effectuées sur le signal ODGD obtenu par filtrage inverse, SV H. : mesures effectuées sur le signal voisé avec application d'un terme correctif (méthode de Hanson [97, 98]), SV : mesures effectuées sur le signal voisé sans correction formantique.*

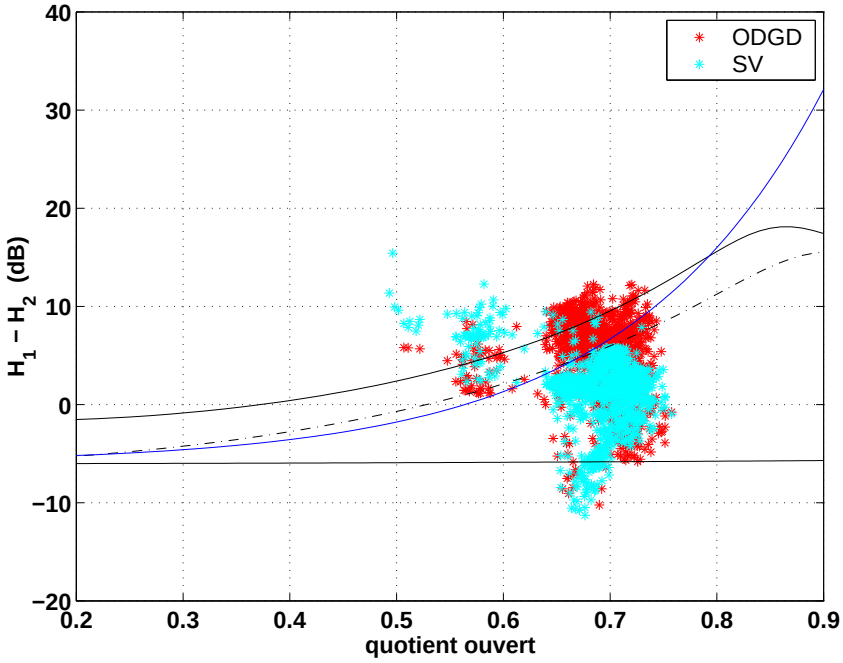


Fig. *4.19 – Mesure de $(H_1 - H_2)$ en fonction de O_q, sur un* **crescendo chanté par la soprano S1 sur un do4 et la voyelle [u], en mécanisme II**. *ODGD : mesures effectuées sur le signal ODGD obtenu par filtrage inverse, SV H. : mesures effectuées sur le signal voisé avec application d'un terme correctif (méthode de Hanson [97, 98]), SV : mesures effectuées sur le signal voisé sans correction formantique.*

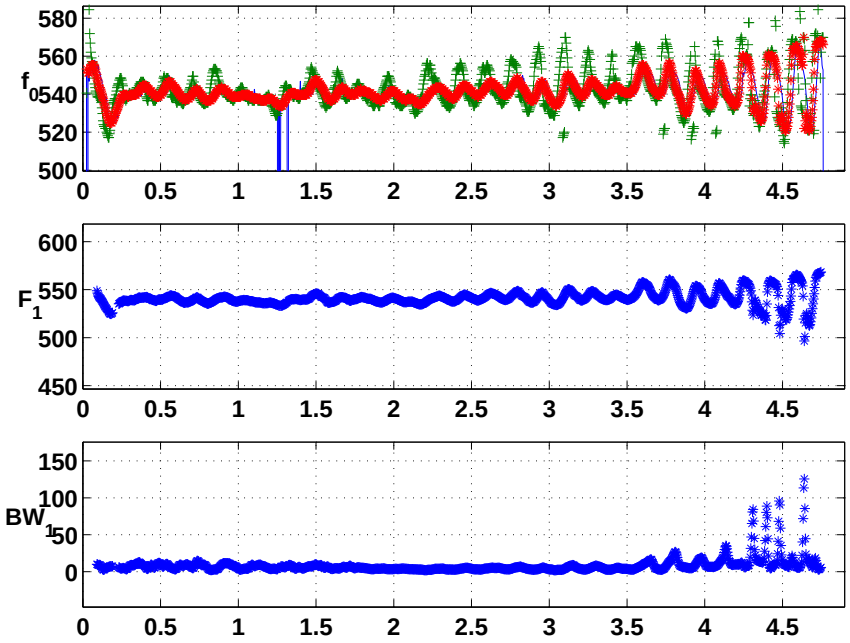


FIG. *4.20 – Estimation de la fréquence fondamentale f_0, de la fréquence du 1^{er} formant F_1 et de sa largeur de bande BW_1 sur le signal d'ODGD estimé, dans le cas d'un crescendo chanté par la soprano S1.*

et sur un ré3 ($f_0 = 293\ Hz$). La figure 4.18 montre les valeurs de $(H_1 - H_2)$ en fonction de valeurs de O_q mesurées sur le signal DEGG associé. De même que pour l'exemple précédent, nous obtenons des valeurs en-dehors des domaines théoriques, et ceci indépendamment du fait que $(H_1 - H_2)$ soit mesuré par la méthode de Hanson ou sur l'ODGD par filtrage inverse. Nous avons présenté ces deux exemples à la même échelle, afin de souligner leurs ressemblances. Dans les deux cas, les valeurs de $(H_1 - H_2)$ sont situées entre 5 et 20 dB. Dans les deux cas, nous remarquons également que la dispersion des mesures est plus importante dans le cas de la méthode de Hanson que dans le cas des mesures sur ODGD estimée par filtrage inverse.

Production vocale en mécanisme II

Le troisième et dernier exemple correspond à l'analyse d'un crescendo chanté par une soprano (S1), sur un do4/do4# (f_0 entre 523 et 554 Hz) et la voyelle [u]. L'analyse de cet exemple est présentée dans le cadre du chapitre 8, en partie 8.4.3. Seules les valeurs de $(H_1 - H_2)$ mesurées sur le signal ODGD obtenu par filtrage inverse sont présentées sur la figure 4.19. En effet, dans ce cas-ci, la correction apportée sur les valeurs mesurées sur le signal voisé n'est plus valable puisque la fréquence fondamentale est très proche de la fréquence du premier formant, ainsi que l'illustre la figure 4.20. Pour cet exemple, les valeurs de $(H_1 - H_2)$ mesurées sont bien comprises dans les limites du domaine théorique et leur variabilité peut s'expliquer par une diminution progressive de Q_a lors de la montée en intensité, tandis que le quotient ouvert reste relativement stable.

4.3.4 Discussion

Dans le cas de la parole, les mesures réalisées sur des signaux réels peuvent correspondre à des cas modélisés par la théorie. Dans le cas du chant, il en va tout autrement. Nous avons observé sur les deux exemples en mécanisme I que les valeurs de $(H_1 - H_2)$ estimées étaient situées bien en dehors des limites

théoriques et qu'elles étaient très élevées : H_1 est dans les deux cas d'une amplitude d'au moins 5 dB supérieure à H_2. Pour tenter d'expliquer ces observations, nous avançons trois hypothèses, qui peuvent se compléter :

1. les perturbations induites sur la mesure par l'environnement acoustique et les mouvements du chanteur ne sont pas négligeables ;
2. les méthodes de filtrage inverse ne sont pas adaptées dans le cas du chant et elles ne permettent pas d'appréhender la source glottique ;
3. les modèles d'ODG ne permettent pas de reproduire les caractéristiques vocales du chant.

Mesure du signal acoustique

Les exemples analysés correspondent à des enregistrements du signal acoustique à 50 cm de la bouche du chanteur, dans une chambre insonorisée mais non anéchoïque [8]. La distance au microphone, ainsi que les possibles phénomènes de réverbération, pourraient avoir une influence sur le signal acoustique mesuré. De plus, bien qu'il lui soit demandé de rester statique, le chanteur ne peut s'empêcher d'accompagner sa production vocale de petits mouvements lents de la tête et du haut du corps. Ces mouvements pourraient également perturber la mesure de pression.

Dans cette étude, ces perturbations ont été supposées négligeables. Il serait intéressant d'effectuer une étude comparative avec la mesure du débit à l'aide d'un masque de Rothenberg, méthode plus invasive mais pour laquelle le signal est enregistré directement en sortie des lèvres et n'est pas perturbé par les mouvements du chanteur.

Limitations du filtrage inverse

Les figures 4.21 et 4.22 complètent les analyses en mécanisme I, en présentant les variations mesurées de la fréquence fondamentale, de la fréquence du 1^{er} formant F_1 et de sa largeur de bande BW_1. Nous observons que la largeur de bande BW_1 varie avec le vibrato de fréquence fondamentale. La fréquence du premier formant F_1 est également variable, surtout dans le cas du baryton. Ces variations pourraient s'expliquer comme un artéfact de la méthode d'estimation par prédiction linéaire. En effet, la prédiction linéaire est sensible à la structure harmonique du spectre, elle a tendance à « s'accrocher » aux harmoniques ayant une amplitude très élevée et la fréquence du premier formant estimé serait alors proche de la fréquence d'un harmonique particulièrement accentué dans le spectre d'ODGD. Cela expliquerait cette variation en fonction de la fréquence fondamentale. Nous observons en effet que la fréquence F_1 estimée est légèrement supérieure à la fréquence de l'harmonique H_2 dans les deux cas. Ces observations vont dans le sens d'une mauvaise estimation du filtre modélisant le conduit vocal, en raison d'un second harmonique d'amplitude élevée. En supprimant l'effet des formants par filtrage inverse, l'amplitude de H_1 serait alors trop accentuée.

Ces problèmes de filtrage inverse peuvent résulter d'une interaction forte entre la source et le filtre dans le cas des productions chantées. Pour produire des sons de forte intensité, les chanteurs ajustent volontairement les cavités du conduit vocal, en soulevant le voile du palais, en modifiant la position de la langue, en abaissant ou en élevant le larynx, ...D'un point de vue acoustique, ces ajustements favorisent la correspondance entre les fréquences des harmoniques et les fréquences des formants. On parle alors d'accord formantique (*formant tuning*). En abaissant la mâchoire, la fréquence du formant F_1 peut augmenter jusqu'à coïncider avec la fréquence fondamentale dans le cas de notes aiguës (Titze [204]). Ainsi, les sopranos peuvent faire coïncider f_0 avec F_1 et accentuer l'intensité vocale du son émis (Sundberg [185, 187]). De même, il est envisageable que les barytons ou les ténors, qui chantent à des fréquences fondamentales plus basses, puissent ajuster leurs cavités pour obtenir un accord formantique entre F_1 et H_2 ou H_3. Dans ce cas, l'estimation exacte des formants sera faussée par cette accentuation.

Il est particulièrement intéressant de noter que cette recherche d'un couplage entre la source et le filtre, très marquée dans le chant, va à l'encontre de la théorie source-filtre et constitue un paradoxe entre

[8] Le dispositif expérimental d'enregistrement est décrit en partie 6.1 du chapitre 6.

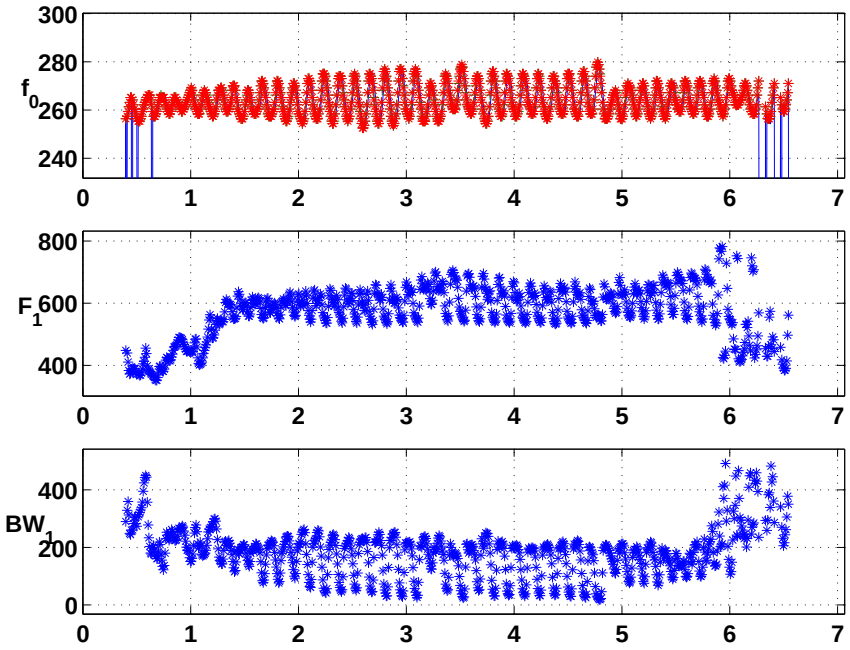


FIG. *4.21 – Estimation de la fréquence fondamentale f_0, de la fréquence du 1^{er} formant F_1 et de sa largeur de bande BW_1 sur le signal d'ODGD estimé, dans le cas d'un crescendo chanté par le baryton B1.*

les hypothèses faites par la théorie et le but recherché pour l'efficacité vocale.

Une limitation des modèles d'ODG ?

On peut également avancer une dernière hypothèse pour expliquer ces résultats. L'excitation acoustique est supposée être générée par la discontinuité de pression à l'instant de fermeture glottique. Il n'est jamais envisagé de discontinuité à l'ouverture. Or Hanson & Chuang [99] ont observé dans le cas de productions en mécanisme I qu'une excitation pouvait apparaître sur le signal d'erreur, vraisemblablement à l'instant d'ouverture. La plupart des modèles d'ODG ne prennent pas en compte cette possibilité et seul un modèle proposé dans la littérature par Fujisaki & Ljungqvist [146, 83] possède une discontinuité du débit à l'instant d'ouverture glottique. Il serait très intéressant d'explorer plus en avant les effets spectraux d'une discontinuité à l'ouverture.

4.4 Conclusion

Nous venons d'étudier de façon à la fois théorique et expérimentale le lien entre le quotient ouvert et la différence en amplitude entre les deux premiers harmoniques du spectre de source glottique. En théorie, la relation qui lie O_q et $(H_1 - H_2)$ n'est pas bijective dans le cas général mais elle dépend des valeurs prises par α_m et Q_a. Si dans la réalité une corrélation forte est trouvée entre $(H_1 - H_2)$ et O_q, cela semble indiquer que α_m et Q_a varient conjointement à O_q. Si nous pouvions quantifier cette dépendance entre ces 3 paramètres de source glottique, nous pourrions simplifier l'approche théorique et n'utiliser dans les modèles qu'un unique paramètre pertinent pour appréhender les qualités vocales pressées, relâchées et soufflées.

Dans cet esprit, Fant [69] a proposé une relation liant $(H_1 - H_2)$ à O_q. Nous avons observé la pertinence de cette relation sur l'analyse d'une production relâchée à pressée. En revenant à la théorie, nous avons tenté de visualiser la dépendance qu'une telle relation entraîne entre O_q et α_m, selon les valeurs prises

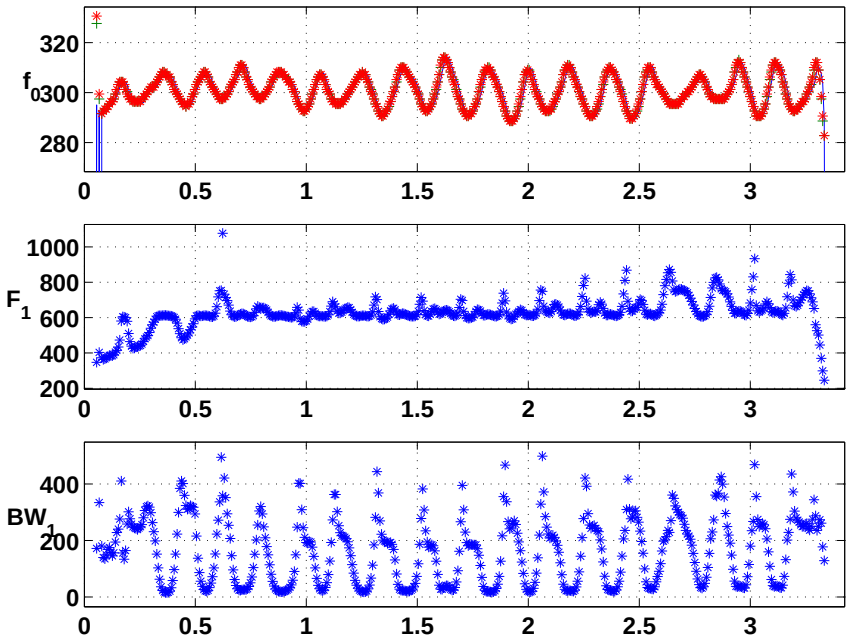


Fig. *4.22 – Estimation de la fréquence fondamentale f_0, de la fréquence du 1^{er} formant F_1 et de sa largeur de bande BW_1 sur le signal d'ODGD estimé, dans le cas d'un crescendo chanté par le ténor T2.*

par Q_a. Il en ressort que pour les phonations à faible quotient ouvert, α_m a des valeurs élevées (entre 0.75 et 0.85). Quand O_q augmente ($O_q \geq 0.65$), α_m reste confiné entre 0.55 et 0.75 et Q_a augmente alors progressivement.

Dans le cas du chant, par contre, les valeurs de $(H_1 - H_2)$ mesurées sont situées bien en dehors des limites théoriques et ne peuvent plus s'expliquer à l'aide des modèles classiques. Nous touchons peut-être là aux limites des méthodes classiques de filtrage inverse par prédiction linéaire. En effet, il y a très certainement dans ces productions vocales une forte interaction entre la source et le filtre. Il nous paraît alors nécessaire, pour appréhender la relation entre les paramètres de source glottique et la qualité vocale, de pouvoir effectuer des mesures de ces paramètres de façon indépendante du système supraglottique. Ce sera l'objet de la partie II que nous allons à présent aborder. Nous y développerons une méthode de mesure de la fréquence fondamentale et du quotient ouvert par utilisation des signaux électroglottographiques dérivés et nous l'appliquerons à l'étude du quotient ouvert en fonction des mécanismes laryngés, de la fréquence fondamentale et de l'intensité, dans le cas de productions vocales parlées et chantées.

En Résumé

Nous avons présenté dans ce chapitre une étude théorique et expérimentale de la différence spectrale entre les deux premiers harmoniques du spectre de source glottique $(H_1 - H_2)$, qui semble permettre, d'après la littérature, l'estimation spectrale du quotient ouvert.

A partir du formalisme mis en place au chapitre 3, nous avons étudié de façon analytique la relation entre $(H_1 - H_2)$ et O_q et montré que cette relation est dépendante du modèle paramétrique d'ODG : elle n'est effectivement bijective que dans le cas du modèle KLGLOTT88, où α_m est constant. Dans les autres cas (modèle LF, R++), elle va dépendre fortement de la valeur prise par les autres paramètres de source.
La relation entre $(H_1 - H_2)$ et O_q est donc fonction du coefficient d'asymétrie α_m et du quotient de phase de retour Q_a :

- $(H_1 - H_2)$ augmente si α_m diminue,
- $(H_1 - H_2)$ augmente si Q_a augmente.

Ces résultats théoriques sont en accord avec les résultats expérimentaux qui montrent une augmentation de l'amplitude du premier harmonique dans le cas de productions vocales relâchées (quotient ouvert élevé) ou une diminution de l'amplitude du premier harmonique dans le cas de productions vocales pressées (quotient ouvert faible).

Dans le cas d'une production réelle, les paramètres de source glottique varient conjointement. En introduisant la relation obtenue par Fant entre $(H_1 - H_2)$ et O_q dans les équations analytiques, nous avons simulé une dépendance entre ces paramètres :
- $0.25 < O_q < 0.6$ s'accompagne de $\alpha_m > 0.7$,
- $O_q > 0.6$ s'accompagne de $0.55 < \alpha_m < 0.65$.

Quand le quotient ouvert augmente, le coefficient d'asymétrie tend à diminuer et le quotient de phase de retour à augmenter.

Afin d'étudier cette différence spectrale en fonction du quotient ouvert sur des signaux réels de parole et de chant, 6 méthodes d'estimation de la fonction de transfert du conduit vocal et de la source glottique par filtrage inverse ont été testées sur des signaux voisés synthétiques, dans le cas des 3 voyelles [a], [i] et [u]. Ces méthodes estiment avec précision les formants élevés, mais que l'estimation du premier formant est moins précise, surtout dans le cas où sa fréquence est basse (cas du [i] et du [u]). Dans ce cas, la méthode par optimisation, qui prend en compte le signal de source glottique à partir d'un modèle, donne de bien meilleurs résultats que les autres méthodes pour estimer correctement la fréquence et la largeur de bande de ce formant.

Ces méthodes ont également été appliquées pour l'estimation de la source glottique dans le cas de signaux réels de parole et de chant. Nous observons qu'il est difficile d'obtenir une phase fermée bien marquée. Les ODG estimées avec les différentes méthodes présentent des formes très semblables dans le cas de la voyelle [a], mais des différences dans le cas des voyelles [i] et [u] à premier formant bas. Dans le cas du chant, une désaccentuation préalable est nécessaire avant l'analyse.

Nous avons enfin effectué des mesures expérimentales de $(H_1 - H_2)$ sur deux types de signaux voisés réels :

1. une voyelle émise de façon relâchée à pressée par un locuteur,
2. 3 crescendos chantés (1 baryton et 1 ténor, en mécanisme I, et une soprano, en mécanisme II)

Dans le cas d'une production vocale relâchée à pressée, une forte corrélation est apparue entre quotient ouvert et $(H_1 - H_2)$, en accord avec la relation proposée par Fant. Dans le cas de signaux de voix chantée, nous avons observé que, dans le plan O_q - $(H_1 - H_2)$, les mesures sont situées très en-dehors des domaines théoriques pour les deux productions vocales en mécanisme I présentées. Par contre, dans le cas d'une production vocale en mécanisme II, les valeurs expérimentales coïncident bien avec les domaines théoriques. Ces résultats pourraient s'expliquer par une limitation des méthodes de filtrage inverse dans le cas du chant, où l'interaction entre la source et le filtre n'est vraisemblablement plus négligeable.

Deuxième partie

Mesures acoustiques et électroglottographiques

Au cours de l'émission d'un son, voyelle chantée par exemple, les cordes vocales exécutent des oscillations qui assurent alternativement la fermeture et l'ouverture de la glotte. La fermeture ne se réalise pas instantanément par tout ou rien, sur toute la longueur de l'orifice, mais d'avant en arrière à partir de la commissure antérieure, quelquefois aussi en même temps d'arrière en avant, à partir des aryténoïdes rapprochés.

(Philippe Fabre, Académie Nationale de Médecine, 1957)

Sommaire de la partie

Dans la partie I, nous avons montré les limites des modèles d'ODG pour la mesure des paramètres de source glottique dans le cas du chant, quand l'hypothèse d'un découplage entre source et filtre devient discutable. Comment étudier la source glottique sans faire de telles hypothèses ? L'électroglottographie, que nous avons évoquée au chapitre 1 de la partie préliminaire, est une méthode non-invasive de mesure de l'activité glottique qui semble répondre à ce besoin. Quels paramètres peut-on mesurer avec cette technique et comment ? Peut-on alors caractériser la variation de ces paramètres pour des productions vocales parlées et chantées ?

L'électroglottographie, que nous présenterons au chapitre 5, renseigne sur le contact entre les cordes vocales, sans que ces mesures soient biaisées par un couplage entre la source glottique et le conduit vocal. Nous nous intéresserons en particulier à la dérivée du signal EGG, qui apporte des informations sur les instants d'ouverture et de fermeture glottique. L'analyse du signal DEGG permet ainsi une mesure précise de la fréquence fondamentale et du quotient ouvert.

Afin d'effectuer des analyses par électroglottographie de la variation du quotient ouvert pour diverses situations parlées et chantées, une base de données comprenant les signaux acoustiques et électroglottographiques de 18 chanteurs entraînés a été enregistrée. Elle sera détaillée au chapitre 6.

Les signaux DEGG de cette base de données seront ensuite étudiés au chapitre 7. Dans la littérature, ces signaux sont très peu exploités. Quelles sont donc leurs possibilités et limites d'utilisation ? Le quotient ouvert mesuré par électroglottographie correspont-il à ce que l'on mesure sur un signal d'ODG estimé par filtrage inverse du signal acoustique ? Nous illustrerons sur des exemples concrets l'intérêt d'utiliser ces signaux et nous proposerons une méthode de mesure du quotient ouvert.

Une fois le contexte expérimental mis en place et validé, nous présenterons au chapitre 8 les mesures de quotient ouvert pour des productions vocales parlées et chantées. Le quotient ouvert est-il dépendant du mécanisme laryngé utilisé ? Si c'est le cas, peut-on déterminer la nature d'un mécanisme laryngé par la seule mesure du quotient ouvert ? Observe-t-on des variations conjointes des paramètres de source glottique (quotient ouvert, fréquence fondamentale) et de l'intensité vocale émise ? La voyelle a-t-elle une influence sur le quotient ouvert ? Observe-t-on des différences sur la mesure du quotient ouvert entre parole et chant ? Voilà autant de questions auxquelles nous tenterons alors de répondre.

Ainsi, cette partie propose une méthode de mesure fiable du quotient ouvert, en parole ou en chant, et l'exploite dans le cas de productions vocales parlées et chantées pour étudier principalement le lien entre quotient ouvert, fréquence fondamentale et intensité vocale, selon le mécanisme laryngé considéré.

Chapitre 5

Principe et application de l'électroglottographie

Au chapitre 4, nous avons touché aux limites des méthodes de filtrage inverse pour l'estimation des paramètres de source glottique en voix chantée. Ces méthodes s'appuient toutes sur la théorie source-filtre (Fant [64], 1960) et font l'hypothèse d'un découplage entre la source et le filtre. Pour étudier les paramètres de source glottique en voix chantée, nous désirons effectuer des mesures indépendamment de cette possible interaction acoustique entre les cordes vocales et le système supraglottique. L'électroglottographie, qui permet d'effectuer des mesures directement au niveau de la glotte, nous semble la méthode exploratoire la plus adaptée. Elle permet l'investigation du mouvement vibratoire des cordes vocales en cours de phonation et son principal intérêt, qui explique aussi son utilisation massive depuis le début des années 1980, provient du fait qu'elle est totalement non-invasive, très simple d'application et qu'elle permet une mesure de l'activité glottique, indépendamment du système supraglottique.

Mise au point par Fabre en 1956 [60, 61, 62, 63], cette méthode s'est répandue avec l'apparition des premiers modèles commercialisés au début des années 1970 (Frokjær-Jensen [81, 79, 80], Fourcin [77, 75]). Les études ont alors fleuri, aussi bien dans le domaine scientifique que dans le domaine médical, où l'utilisation d'un tel dispositif pouvait apporter de précieux renseignements pour la détection et l'analyse des pathologies vocales (Kitzing [129]). Cette utilisation massive s'est accompagnée d'un souci de compréhension de la nature exacte des signaux mesurés et de leur adéquation avec l'objet d'étude (i.e. : la caractérisation du mouvement vibratoire des cordes vocales au cours de la production vocale). Childers & Krishnamurthy [33] présentent une revue très détaillée des études menées sur l'électroglottographie jusqu'au milieu des années 80. Colton & Conture [40], ainsi que récemment Orlikoff [162], détaillent l'utilisation et les limitations du dispositif.

Le but de ce chapitre est de caractériser les signaux EGG, de comprendre quelles informations ils apportent et de quels paramètres de source glottique ils permettent la mesure. Après avoir décrit le dispositif de mesure en partie 5.1, nous présenterons dans la partie 5.2 une analyse de la forme des signaux électroglottographiques et leur caractérisation. Enfin, nous discuterons en partie 5.3 de l'utilisation des signaux électroglottographiques pour la mesure des paramètres de source glottique.

5.1 Description du dispositif

L'électroglottographe est représenté schématiquement sur la figure 5.1. Il est constitué d'un générateur qui délivre un courant alternatif haute fréquence (porteuse à une fréquence de l'ordre de 1 MHz), de deux électrodes qui sont appliquées sur le cou du sujet, au niveau du cartilage thyroïdien, et d'un circuit

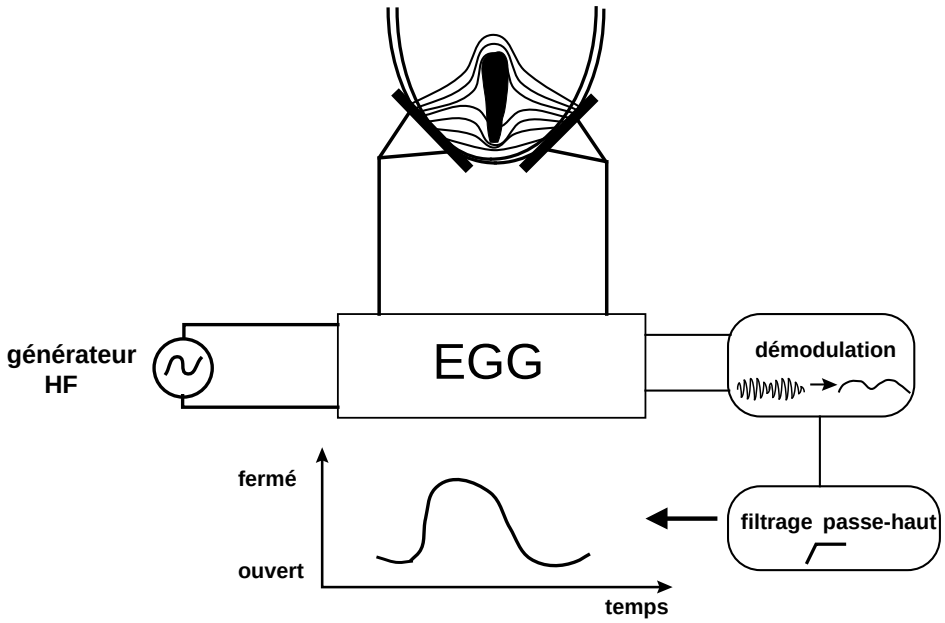


FIG. *5.1 – Principe d'un électroglottographe.*

électronique qui agit principalement en démodulateur de fréquence. Un filtrage passe-haut (de fréquence de coupure comprise entre 5 et 40 Hz) permet d'éliminer les composantes parasites basse fréquence, dues aux mouvements du larynx en cours de phonation, au débit sanguin dans les artères et les veines du cou, ainsi qu'aux contractions des muscles laryngés extrinsèques (Colton & Conture [40]). Comparable à un ohmmètre, l'électroglottographe mesure la différence de potentiel u entre les deux électrodes. Celle-ci est reliée à l'impédance électrique R du cou, qui est la résistance au passage du courant à travers la peau, le cartilage thyroïdien, les tissus, les muscles et la glotte. La relation entre u et R est donnée par la loi d'Ohm $u = Ri$, la proportionnalité étant évidente si le courant i est gardé constant. Quand la glotte est fermée, la résistance au courant est moindre, et cette résistance augmente avec l'ouverture de la glotte, l'air étant un bien moins bon conducteur que les tissus humains. Le signal électrique délivré par le générateur est donc modulé en fréquence par le mouvement vibratoire des cordes vocales. La période de ce signal correspond donc à la fréquence fondamentale du son émis. Le spectre d'un signal EGG est également très riche en harmoniques, ainsi que l'illustre la figure 5.2. On observe ainsi de l'énergie jusqu'à des fréquences de l'ordre de 20 kHz.

5.2 Caractérisation des signaux électroglottographiques

5.2.1 Une représentation du contact entre les cordes vocales

Les études comparatives avec d'autres méthodes exploratoires comme la photographie stroboscopique (Fourcin [75], Lecluse [139], Perdersen [166], Teaney & Fourcin [196]), la vidéostroboscopie (Anastaplo [11] et Karnell [125]), la cinématographie ultra-rapide (Childers & al [37, 38, 34, 33, 31], Baer & al [14]), la photoglottographie (Kitzing [126, 130, 128], Dejonckere [52], Baer [14, 15] and Titze [206], Berke & al [21], Gerratt & al [86]), les mesures de pression sous-glottiques (Kitzing [130]) et le filtrage inverse (Fourcin [76], Rothenberg [171, 173], Childers & al [37]) ont permis de caractériser les signaux électriques délivrés par l'électroglottographe.

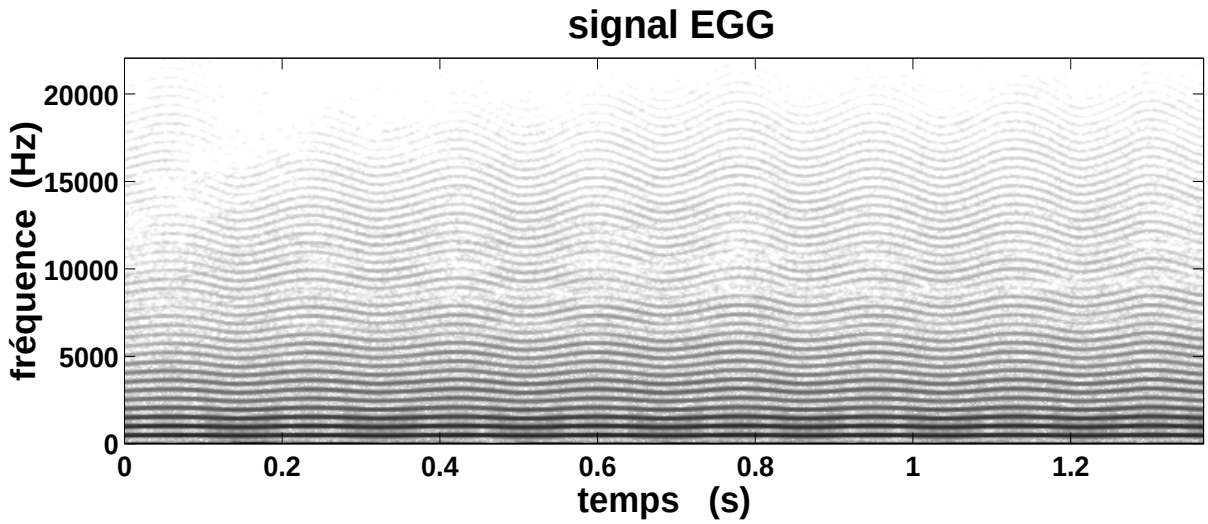


Fig. *5.2 – Sonagramme d'un signal électroglottographique.*

Ces études s'accordent à montrer que le signal électroglottographique est modulé en fonction du contact entre les cordes vocales : plus ce contact est important, plus l'impédance mesurée est faible. La fermeture glottique, qui se caractérise par une augmentation du contact entre les cordes vocales, se traduira donc sur le signal EGG par une augmentation de l'admittance mesurée (diminution de l'impédance, donc de la tension). L'ouverture se traduira de façon inverse par une diminution de l'admittance, puisque caractérisée par une diminution du contact entre les cordes.

5.2.2 Discussion sur la représentation d'un signal EGG

Deux conventions opposées sont utilisées pour présenter les signaux EGG, selon que l'on considère ces signaux comme une mesure de l'impédance du cou ou comme une mesure de l'aire de contact. Dans le premier cas (convention *fermé* → *ouvert*), le signal augmente avec l'impédance, ce qui implique que le maximum correspond à la phase ouverte (impédance élevée) et le minimum à la phase fermée (impédance faible). Dans le second cas (convention *ouvert* → *fermé*), le signal augmente avec le contact entre les cordes vocales, ce qui implique une représentation inverse de la précédente : en phase ouverte, le contact est réduit et il augmente avec la fermeture. Dans le milieu médical, cette seconde convention est la plus répandue, du fait de l'utilisation massive du laryngographe de Fourcin[76]. Par contre, les avis sont plus partagés chez les chercheurs (convention *fermé* → *ouvert* : Fabre [60], Pedersen [166], Kitzing [126], Rothenberg [171], Childers [33] ; convention *ouvert* → *fermé* : Lecluse [139], Fourcin [76], Baer [14], Titze [200, 206], Gilbert [87], Haji [94]). En 1992, Baken [16] propose comme norme, au nom du *Voice Committee of the IALP* [1], la convention *ouvert* → *fermé*, puisque le signal électroglottographique est avant tout représentatif du contact glottique. Cette représentation permet également d'éviter de possibles confusions entre le flux glottique et le signal électroglottographique. En effet, si l'on se place dans le cadre de la convention *fermé* → *ouvert*, les signaux correspondant à l'ODG et à l'EGG varient de pair (augmentation du signal avec l'ouverture et diminution avec la fermeture). Mais mise à part cette similitude de forme, il n'existe aucun lien physique démontré entre ces deux signaux : l'ODG représente le débit d'air passant à travers la glotte au cours du temps et l'EGG l'influence du mouvement glottique sur un courant électrique traversant le cou. Ce sont donc des signaux complémentaires mais non directement comparables, ce qui se traduit de façon plus explicite dans la convention *ouvert* → *fermé*.

Dans cette étude, nous avons fait le choix de respecter cette tentative de cohésion entre les diverses

[1] IALP : International Association of Logopedics and Phoniatrics

représentations et nous présenterons donc les signaux EGG dans la convention *ouvert* → *fermé* : le signal croît lorsque la glotte se ferme et décroît lorsqu'elle s'ouvre.

5.2.3 Description d'une période de signal EGG

La visualisation d'un signal EGG renseigne sur le degré de contact entre les cordes vocales. Lecluse & Brocaar [138] ainsi que Childers & Krishnamurthy [33] ont proposé une description schématique du signal EGG lors d'une période du mouvement glottique. Cette schématisation est fondée sur une caractérisation de la forme du signal en comparaison avec l'observation par cinématographie ultra-rapide, dans le cas du mécanisme I. Nous nous appuyons ici sur leurs observations pour décrire de façon schématique les différentes phases du mouvement glottique sur une période du signal EGG, dans le cas du mécanisme I et du mécanisme II. Pour ce faire, deux signaux EGG représentatifs des deux mécanismes ont été choisis dans la base de données décrite au chapitre 6 et sont présentés sur la figure 5.3 (signal réel et forme schématique). Le signal DEGG, qui permet de visualiser les changements de pente du signal EGG, ainsi qu'une schématisation de l'ODG correspondant à ces signaux, viennent compléter cette description.

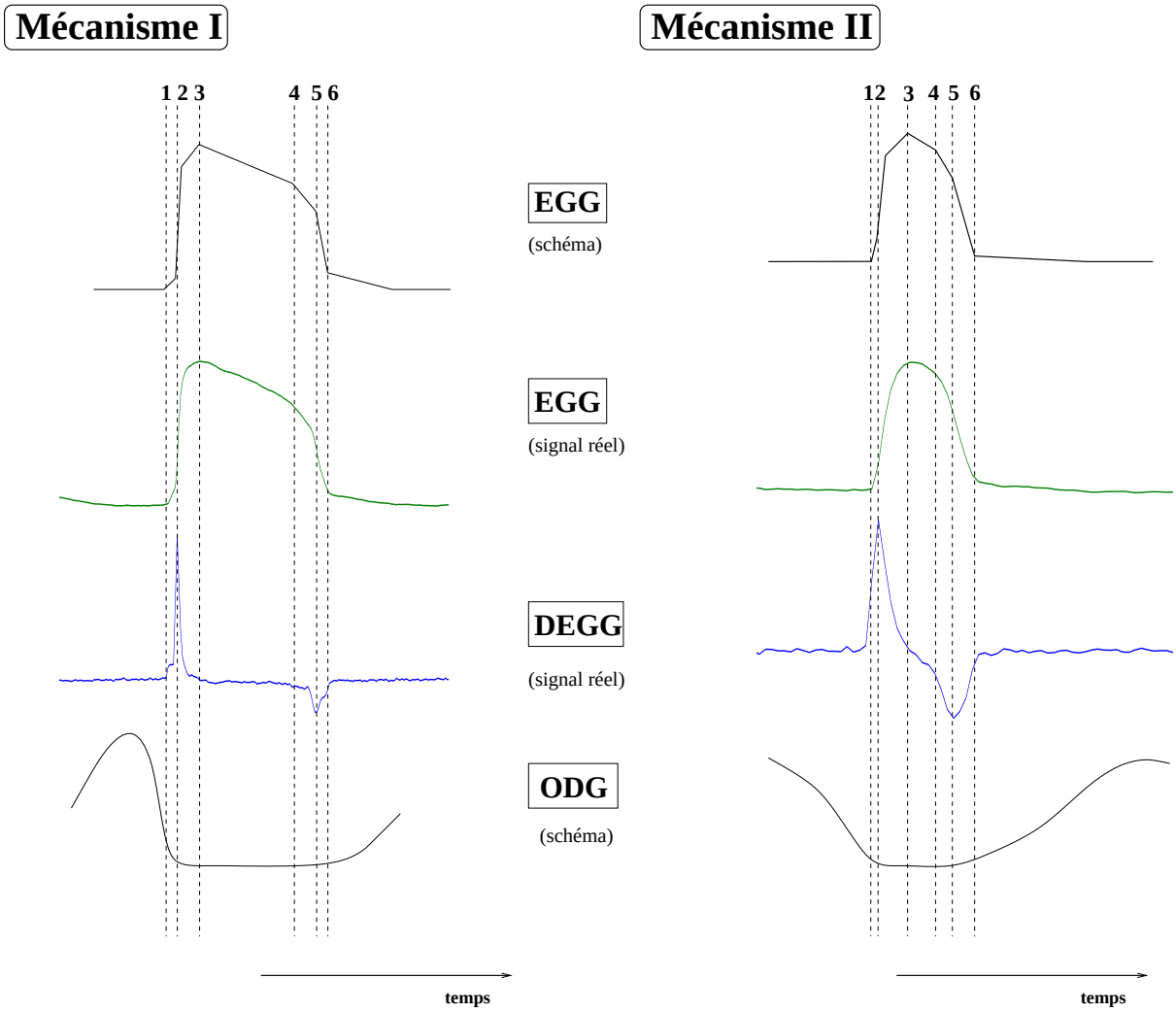


FIG. *5.3 – Illustration des étapes du mouvement glottique visualisées sur un signal EGG et sur sa dérivée. (1)-(3) : phase de fermeture, (3)-(4) : phase fermée, (4)-(6) : phase d'ouverture, (6)-(1) : phase ouverte. Pour comparaison, un schéma du débit glottique correspondant (ODG) a été réprésenté.*

Quatre phases principales, illustrées sur cette figure, permettent de décrire un signal EGG.

① - ③ **: phase de fermeture** . Les cordes vocales se referment, d'abord sur leur partie inférieure (① à ②), puis sur leur partie supérieure (② à ③). La fermeture étant généralement plus rapide que l'ouverture, cette phase est marquée par une pente importante sur le signal EGG. En ② , la dérivée de l'EGG présente un pic très marqué, dont le maximum local est associé à l'instant de fermeture glottique. Physiologiquement, cela correspond à l'instant où les cordes vocales se touchent sur la totalité du bord inférieur. Une visualisation détaillée de cette phase par cinématographie ultra-rapide et électroglottographie simultanées est présentée sur la figure 5.4.

③ - ④ **: phase fermée** . Les cordes vocales sont accolées sur toute leur longueur. On peut néanmoins observer une variation du signal EGG, qui correspond au degré d'accolement des cordes vocales.

④ - ⑥ **: phase d'ouverture** . Les cordes vocales, accolées sur toute leur longueur (en ④), commencent à se séparer sur leur partie inférieure (④ à ⑤), puis sur leur partie supérieure (⑤ à ⑥). En ⑤ , la dérivée de l'EGG présente un pic négatif plus ou moins marqué, dont le minimum local est associé à l'instant d'ouverture glottique. Physiologiquement, cela correspond à l'instant où les cordes vocales commencent à se séparer sur leur bord supérieur. Une visualisation détaillée de cette phase par cinématographie ultra-rapide et électroglottographie simultanées est présentée sur la figure 5.5.

⑥ - ① **: phase ouverte** . Les cordes vocales sont séparées. L'impédance varie alors très peu et on observe donc un signal relativement plat. L'électroglottographe ne peut fournir aucun renseignement relatif à cette phase, puisque le degré d'ouverture de la glotte influence peu le signal électrique : l'air étant un mauvais conducteur, les lignes de courant sont déviées autour de la glotte.

Les instants de fermeture (en ②) et d'ouverture (en ⑤) correspondent à des points d'inflexion ou à des changements de pente discontinus du signal EGG. Nous appellerons par la suite *pic d'ouverture* le pic présent sur le signal DEGG pendant la phase d'ouverture et *pic de fermeture* celui présent pendant la phase de fermeture. Une équipe du *Department of Phoniatrics and Pediatric Audiology* de l' *University of Erlangen-Nuremberg* (Allemagne), spécialisée dans l'étude des pathologies vocales, a très gracieusement mis à disposition pour nos recherches l'enregistrement par cinématographie ultra-rapide et électroglottographie simultanées d'un locuteur en phonation normale (non pathologique, en mécanisme I et à $f_0 = 110 Hz$). Cet enregistrement a été effectué par J. Neubauer, M. Tigges, T. Wittenberg et U. Eysholdt. Avec ces données et grâce à leur aimable autorisation, nous pouvons illustrer ici les phases d'ouverture et de fermeture glottique par une visualisation simultanée des images obtenues par cinématographie ultra-rapide et des instants correspondants sur le signal électroglottographique et sa dérivée.

Dans une étude détaillée sur la forme des signaux EGG, Roubeau [174] a observé une grande variabilité morphologique intra et inter-individuelle, ce qui limite la détermination de la nature du mécanisme vibratoire mis en jeu par la seule observation de ces signaux. Des tendances générales ont néanmoins été observées, portant sur la vitesse de fermeture (plus grande en mécanisme I qu'en mécanisme II) et sur la durée d'ouverture (plus longue en mI qu'en mII). On retrouve ces tendances sur les deux exemples caractéristiques présentés sur la figure 5.3. En mécanisme I, la phase de fermeture est brève (① à ③), se distinguant sur le signal EGG par une pente très raide. La phase d'ouverture est beaucoup plus graduelle (④ à ⑥). Cette différence de vitesse est très nette sur le signal DEGG, où le pic de fermeture a une amplitude bien plus importante que celle du pic d'ouverture. La durée de la phase fermée est souvent longue par rapport à la période, comme nous le verrons en étudiant le quotient ouvert (chapitre 8). Dans le cas du mécanisme II, les phases d'ouverture et de fermeture ont des durées comparables et se traduisent par des pics d'ouverture et de fermeture d'amplitudes presque égales. La durée de la phase fermée est réduite, ce qui se traduit également sur les mesures de quotient ouvert.

Ces différences observées sur les signaux EGG et DEGG reflètent les différences mises en évidence sur l'ODG : ouverture graduelle et fermeture rapide dans le cas du mécanisme I, onde beaucoup plus symétrique dans le cas du mécanisme II (Childers & Lee [35]). En se basant sur un modèle mathématique de l'aire de contact entre les cordes vocales [200], Titze a décrit l'effet de différentes configurations glottiques

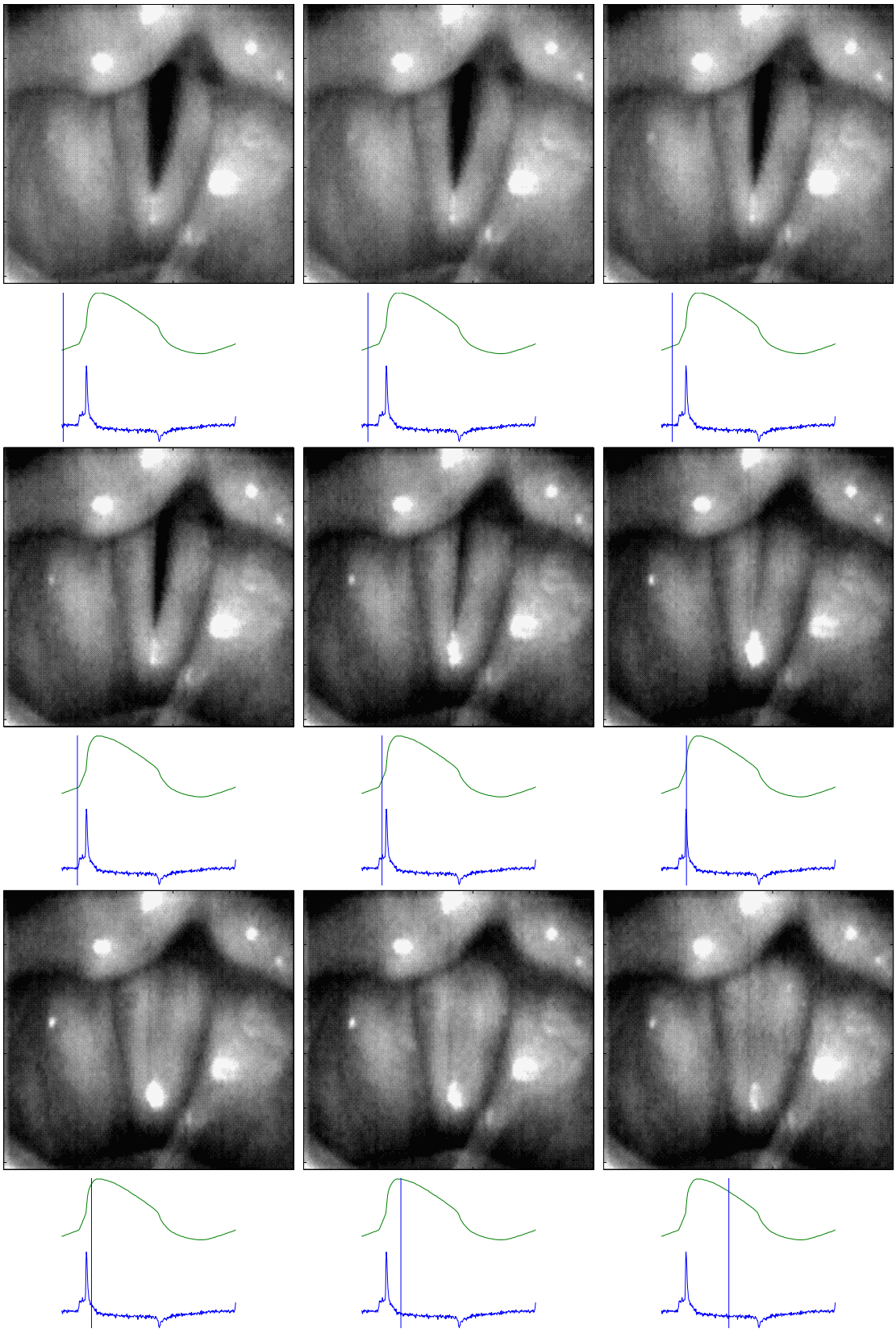


FIG. *5.4* – **Visualisation de la fermeture** *par cinématographie ultra-rapide et électroglottographie simultanées (locuteur en phonation normale,* $f_0 = 110\,Hz$*, mécanisme I).*

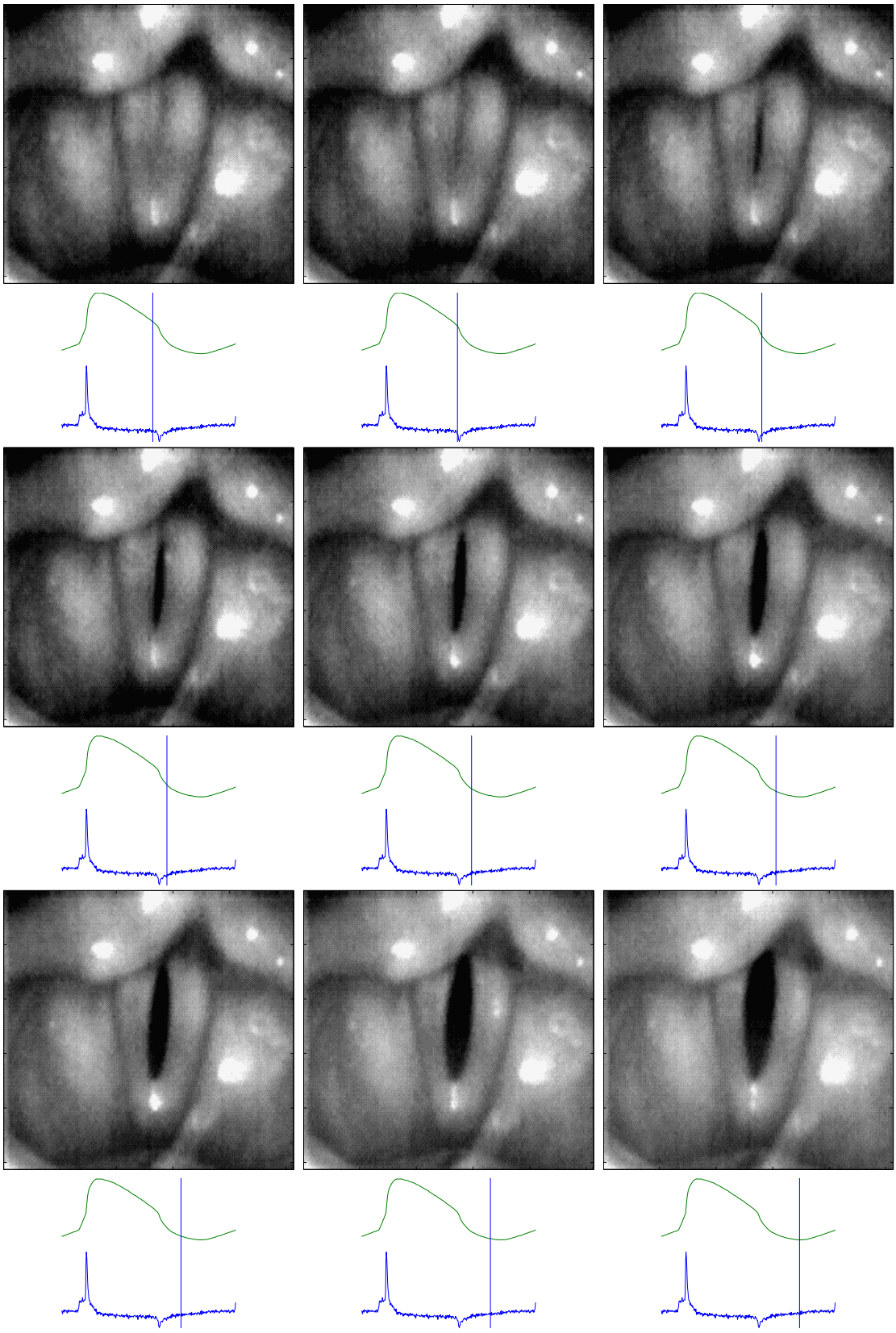


Fig. 5.5 – **Visualisation de l'ouverture** *par cinématographie ultra-rapide et électroglottographie simultanées. (locuteur en phonation normale,* $f_0 = 110\,Hz$*, mécanisme I).*

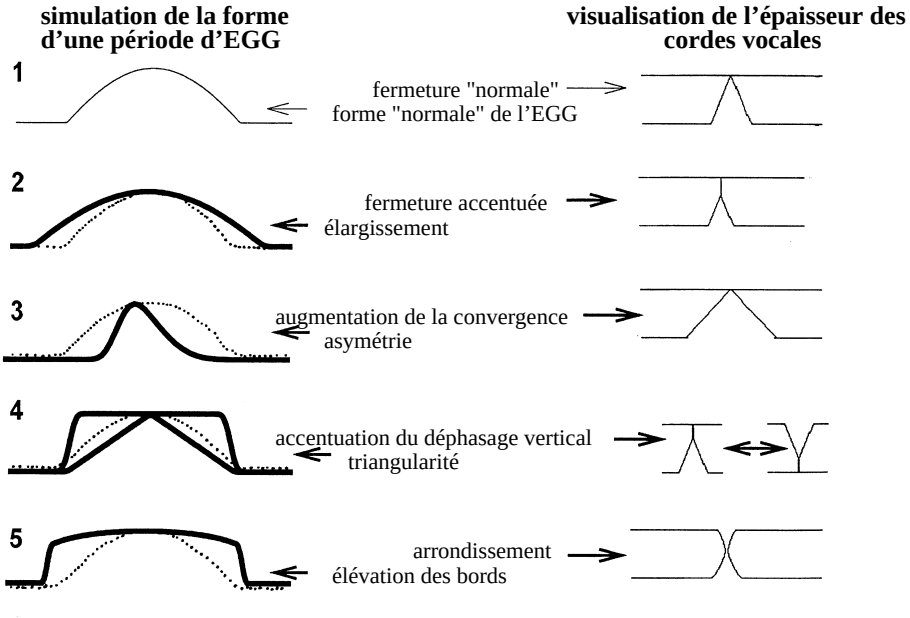


FIG. *5.6 – Simulation de la forme d'une période d'un signal EGG à partir d'un modèle mathématique de l'aire de contact entre les cordes vocales (Titze [200]) pour différentes configurations glottiques, visualisées sur l'épaisseur des cordes vocales (d'après Baken & Orlikoff [17]).*

sur la forme du signal EGG [198], ainsi que l'illustre la figure 5.6. En particulier, il ressort de son étude que l'asymétrie du signal augmente avec la convergence du profil vertical de la corde vocale. La convergence est accrue quand les bords supérieurs des cordes sont plus proches et les bords inférieurs plus éloignés (cas 3.). L'asymétrie augmente également avec un déphasage accentué entre le mouvement du bord supérieur de la corde et celui du bord inférieur. La pente du signal à la fermeture comme à l'ouverture s'accentue si le profil vertical de la corde s'arrondit (cas 5.). On peut donc penser que la différence générale de forme du signal EGG observée entre le mécanisme I et le mécanisme II est principalement le reflet de la configuration glottique. Les cordes vocales sont épaisses et peu tendues en mécanisme I, ce qui accentue la convergence, le déphasage et amène à un arrondissement possible du bord des cordes. En mécanisme II, les cordes vocales sont fines et tendues : il n'y a donc pas (ou peu) de convergence et pas de déphasage. Les différences de forme plus localisées du signal EGG, comme la présence ou non du point d'inflexion ou une pente plus ou moins raide, seraient alors l'effet de modification du mouvement glottique. Elles traduiraient des variations de vitesse de fermeture ou d'amplitude du mouvement vibratoire lors d'un changement de qualité vocale par modification de la pression sous-glottique et des tensions musculaires laryngées.

5.3 Apport des signaux électroglottographiques

Nous avons représenté simultanément sur la figure 5.3 des signaux EGG et ODG, ceci dans le but d'illustrer la distinction fondamentale qui existe entre ces deux signaux. L'ODG, qui correspond au flux d'air traversant la glotte au cours du temps, est un signal continûment nul (fermeture complète) ou minimal (fuite glottique) lorsque la glotte est fermée. Les informations que ce signal apporte concernent donc la phase ouverte. Au contraire, le signal EGG, qui correspond à l'aire de contact entre les cordes vocales, ne varie que très peu lors de la phase ouverte. Les informations qu'il apporte concernent la phase fermée, qui peut alors être caractérisée par le calcul de divers quotients (quotient de fermeture, quotient fermé, quotient d'ouverture, quotient ouvert (Roubeau [174], 1993)). Les signaux EGG et ODG apportent des informations complémentaires, mais ne permettent en aucun cas de déduire la forme d'un signal par

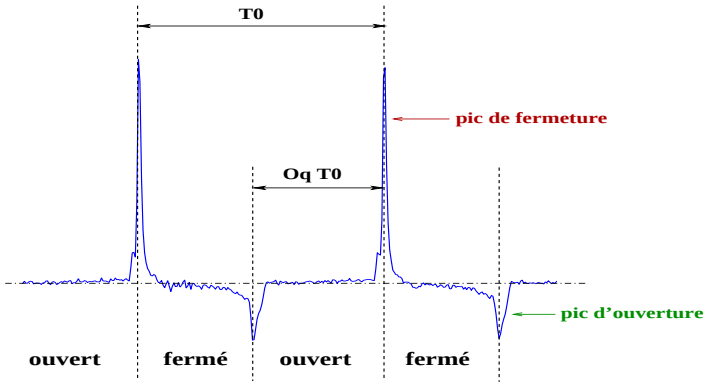


Fig. *5.7 – Signal DEGG présentant des pics marqués à l'ouverture et à la fermeture de la glotte, ce qui permet la mesure de la période fondamentale T_0 et du quotient ouvert O_q.*

la connaissance de l'autre.

Dans le cadre de notre étude sur la source glottique, l'enregistrement du signal EGG ne nous permet donc pas d'appréhender la forme de l'ODG, mais il nous apporte des renseignements précieux concernant les instants de fermeture et d'ouverture glottique. Ces instants, qui correspondent à des points d'inflexion ou des discontinuités de pente du signal EGG, sont facilement mesurables sur la dérivée de ce signal. Plutôt que de s'attacher au signal EGG lui-même, nous sommes donc amenés à utiliser le signal DEGG pour la détermination de ces instants. Dans la littérature, les études portant sur les signaux EGG ont généralement omis de s'intéresser à la dérivée. Nous allons donc discuter dans le détail de la caractérisation et de l'utilisation possible de ces signaux DEGG au chapitre 7. Le signal DEGG véhicule en effet des informations essentielles pour l'analyse des signaux voisés, en particulier par la connaissance de l'instant de fermeture glottique (utile pour le filtrage inverse, les transformations « pitch-synchrones »). La figure 5.7 présente deux périodes d'un signal DEGG. Puisque l'on peut déterminer avec précision sur ce signal la durée séparant deux instants de fermeture glottique successifs, on obtient ainsi une mesure de la période fondamentale T_0 et on en déduit la fréquence fondamentale par $f_0 = 1/T_0$. La détermination de la durée entre un instant d'ouverture glottique et l'instant de fermeture glottique successif permet une mesure du quotient ouvert O_q (ou fermé $Q_c = 1 - O_q$) découplée du système supraglottique.

La fréquence fondamentale et le quotient ouvert sont donc les paramètres de source que l'on peut envisager de mesurer à partir d'un signal DEGG. Dromey & al [57] se sont penchés sur la mesure du quotient de vitesse. Ils ont conclu que le quotient mesuré par EGG n'était pas comparable au quotient de vitesse mesuré sur l'onde de débit glottique, s'agissant plus, dans le cas de l'EGG, d'une mesure des variations rapides d'aire de contact que d'un rapport entre temps de fermeture et d'ouverture. Il serait néanmoins intéressant d'explorer la signification et les corrélats acoustiques et perceptifs d'une telle mesure, qui semble reliée à la pente spectrale d'après les résultats obtenus par Chuberre [39], donc à la vitesse de fermeture E. Ce serait une étude en soi et nous avons donc fait le choix de nous limiter ici à f_0 et O_q.

5.4 Conclusion

Nous venons de présenter et de caractériser la forme des signaux électroglottographiques, en distinguant les mécanismes I et II. Cette méthode ne va pas nous permettre de mesurer tous les paramètres de source glottique que nous avons étudiés de façon théorique en partie I de cette étude. Nous allons donc nous limiter à présent à la fréquence fondamentale et au quotient ouvert, qui peuvent être mesurés sur ces signaux ou sur leur dérivée.

Pour mener cette étude, nous avons besoin de données électroglottographiques et acoustiques enregistrées simultanément pour des productions vocales parlées et chantées mettant en oeuvre différentes qualités vocales. Comme nous n'avions aucune base de données à disposition, nous avons élaboré un protocole d'enregistrement de ces données. 18 chanteurs entraînés ont participé de façon bénévole à ces enregistrements, que nous allons à présent détailler.

En Résumé

Le signal EGG caractérise le degré d'accolement des cordes vocales au cours de la phonation. Il renseigne donc sur les phases de fermeture, fermée et d'ouverture, mais n'apporte aucune information exploitable lors de la phase ouverte.

Selon le mécanisme laryngé (mI ou mII), la forme du signal EGG diffère généralement : durée de la phase fermée plus longue en mI qu'en mII, accentuation de la symétrie du signal en mII. Néanmoins, la seule visualisation du signal EGG ne renseigne pas sur la nature du mécanisme laryngé utilisé.

Ce signal n'est pas comparable à l'ODG et il ne permet pas de l'estimer mais il apporte des informations complémentaires. Ces informations portent sur la détermination des instants d'ouverture et de fermeture glottique. Ces instants correspondent à des points d'inflexion ou des discontinuités de pente du signal et se caractérisent par des pics marqués sur le signal DEGG (dérivée).

Une mesure précise de la fréquence fondamentale et du quotient ouvert est possible à partir d'un signal DEGG, par détection des instants d'ouverture et de fermeture glottique.

Chapitre 6

Constitution d'une base de données

Dans cette étude, nous cherchons à comprendre les changements de qualité vocale en voix chantée (intensité, timbre), en termes de modification des paramètres de source glottique. Ainsi, nous nous proposons d'étudier la façon dont le chanteur va utiliser l'instrument vocal pour produire des sons et moduler leur intensité, varier leur hauteur et changer leur timbre. Nous désirons donc nous placer dans le cadre d'une production vocale naturelle, non forcée, et ceci n'est envisageable que si les appareillages utilisés pour les mesures ne perturbent pas l'émission vocale du chanteur.

Dans cette optique, la mesure la plus simple envisageable est l'enregistrement du signal acoustique émis par le chanteur. Dans la partie 4.2 du chapitre 4, nous avons donc développé des méthodes d'estimation de la source glottique. Ces méthodes s'appliquent sur un signal de pression enregistré à l'aide d'un microphone placé en sortie des lèvres du chanteur (la mesure du débit aux lèvres à l'aide du masque de Rothenberg étant plutôt invasive). Plus généralement, l'enregistrement du signal acoustique permet la mesure de l'intensité sonore produite par le chanteur et apporte des renseignements sur l'évolution spectrale (harmonique et formantique) du signal vocal au cours du temps. Néanmoins, cette mesure n'est pas sans inconvénients et le problème majeur rencontré porte sur la difficulté d'effectuer un filtrage inverse convenable. Comme nous venons de le voir, une telle mesure peut alors être avantageusement complétée par l'électroglottographie, qui permet de contourner ce problème par des mesures découplées de l'interaction avec le système supraglottique. Qui plus est, par son caractère non-invasif, l'électroglottographie ne va pas perturber le geste vocal du chanteur. Néanmoins, seuls la fréquence fondamentale et le quotient ouvert peuvent ainsi être mesurés.

Il vient donc tout naturellement que cette étude doit s'appuyer sur des enregistrements simultanés des signaux acoustiques et électroglottographiques pour les productions vocales chantées. Comme il n'existait pas de base de données accessible qui satisfasse à ces besoins, nous avons donc été amenés à la construire. Nous allons tout d'abord décrire le dispositif expérimental mis en place pour cette étude (partie 6.1), puis nous détaillerons les caractéristiques vocales des chanteurs enregistrés en partie 6.2 et nous examinerons le protocole proposé en partie 6.3.

6.1 Le dispositif expérimental

L'enregistrement simultané des signaux acoustiques et électroglottographiques a été effectué dans une chambre fortement insonorisée mais non anéchoïque, ce qui a permis aux chanteurs de se sentir à l'aise, même dans un petit espace [1]. La figure 6.1 présente une photographie de la chambre et du dispositif de mesure (illustré par le schéma 6.2).

[1] Les dimensions de la chambre sont les suivantes : longueur 1.8 m, largeur 1.8 m, hauteur 2.1 m.



FIG. *6.1 – Chambre insonorisée utilisée pour les enregistrements et dispositif de mesure (enregistrement sur DAT et visualisation des signaux à l'oscilloscope).*

Le signal acoustique est enregistré à l'aide d'un microphone de pression 1/2" (Brüel & Kjær 4165) placé à 50 cm du chanteur, un préamplificateur (Brüel & Kjær 2669) et un amplificateur de mesure (Brüel & Kjær NEXUS 2690). Le signal est numérisé directement sur un DAT (PORTADAT PDR1000) à une fréquence d'échantillonnage de 44,1 kHz ou 48 kHz sur 16 bits.

Le signal électroglottographique est enregistré simultanément sur la seconde piste du DAT. L'électroglottographe utilisé dans cette étude, mis au point par Rothenberg [172], se compose de deux paires d'électrodes, ce qui permet d'ajuster au mieux le placement des électrodes sur le cou du sujet et de prendre en compte les déplacements verticaux du larynx au cours de la phonation. Il délivre un courant de 10 mA , modulé en fréquence à 2 MHz. La tension délivrée en sortie de l'appareil est de quelques centaines de millivolts et ne dépasse pas 1.5 V. Les électrodes sont circulaires, d'un diamètre de 34 mm. Un gel appliqué sur les électrodes au moment du placement permet un meilleur contact avec la peau.

Deux microphones de prises de sons (Neuman) ont également été placés dans la chambre, afin de permettre un enregistrement stéréophonique de qualité sur un second DAT (DA30). Ces enregistrements n'ont pas été exploités au cours de la thèse.

Le déroulement du protocole est suivi au casque par l'expérimentateur et un oscilloscope analogique-numérique permet un retour visuel sur les signaux acoustiques et électroglottographiques. Les hauteurs sont données à l'aide d'un diapason ou d'un synthétiseur placé en dehors de la chambre, selon le choix du chanteur. Lors d'une calibration préalable, un signal sinusoïdal de référence est envoyé par l'amplificateur NEXUS dans la chaîne de mesure (amplitude 1 $VRMS$, fréquence 1 kHz). Par la connaissance des caractéristiques du microphone et du pré-amplificateur, cette mesure permet de déterminer le gain de la chaîne de mesure et de déterminer la pression en Pa enregistrée au niveau du microphone. L'intensité vocale en décibel, I, est alors calculée à partir du signal de pression au microphone par :

$$I = 10\log_{10}[\frac{1}{T}\int_0^T (\frac{p_c(t)}{p_0})^2\, dt]$$

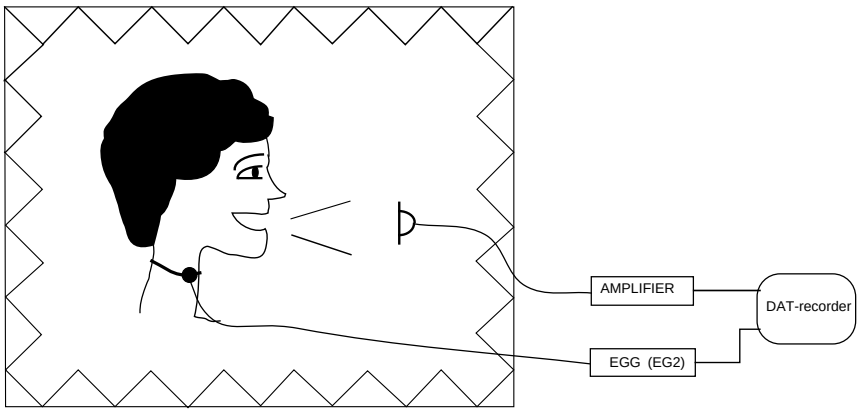


Fig. *6.2 – Schéma du dispositif de mesure : les signaux audio et électroglottographiques sont enregistrés simultanément sur les deux pistes d'un DAT*

avec $p_c(t)$ la pression acoustique au niveau du capteur à l'instant t, $p_0 = 2.10^{-5}\,Pa$ la pression acoustique de référence et T l'intervalle temporel de mesure. D'après la norme IEC 651, $T = 1s$ pour la calibration « Slow » des sonomètres et $T = 125\,ms$ pour la calibration « Fast », valeur que nous avons choisie pour cette étude. Pour vérification, le chanteur émet, au début de l'enregistrement, un son tenu dont l'intensité est mesurée à l'aide d'un sonomètre analogique, placé au niveau du microphone.

6.2 Les chanteurs

Les sujets enregistrés pour cette étude sont des chanteurs semi-professionnels ou professionnels, entraînés, tous de formation classique. Ils sont issus pour la plupart du Conservatoire National Supérieur de Musique et de Danse de Paris (CNSMDP). Le tableau 6.1 détaille les caractéristiques de ces chanteurs. Ces données ont été recueillies auprès des chanteurs et les terminologies utilisées (pour la tessiture par exemple) sont celles exprimées par le chanteur.

La première colonne fournit l'identifiant du chanteur, qui sera utilisé dans la suite de l'étude. Les chanteurs professionnels sont marqués en gras. La seconde colonne donne la tessiture du chanteur. La troisième et la quatrième colonne fournissent des renseignements sur l'étendue vocale du chanteur en mécanisme I et II, telle que le chanteur la perçoit. Les notes extrêmes sont données, avec, entre parenthèses, les notes où le chanteur se sent à l'aise. La cinquième colonne renseigne sur la formation vocale du chanteur en nombre d'années d'entraînement vocal, ainsi que sur le type de formation (c : conservatoire, p : cours particuliers, ch : choeurs). Sont également fournis des indications sur l'âge du chanteur, le répertoire général d'étude, et d'autres aspects tels que l'enseignement du chant (pédagogie vocale), une formation instrumentale notable en parallèle ou une profession en rapport avec la voix (huitième colonne).

La figure 6.3 complète ce tableau en présentant sur une même portée l'étendue vocale de chaque chanteur en mécanisme I et en mécanisme II (notes extrêmes dont les valeurs sont données dans les colonnes « mécanisme I » et « mécanisme II » du tableau 6.1). Nous notons en particulier sur cette figure la zone de recouvrement entre les deux mécanismes, zone fréquentielle dans laquelle le chanteur va pouvoir utiliser l'un ou l'autre des deux mécanismes.

TAB. *6.1 – Tableau présentant les caractéristiques vocales des chanteurs enregistrés pour cette étude. Les chanteurs professionnels sont marqués en gras. L'étendue vocale est donnée pour les mécanismes I et II (les notes où le chanteur se sent à l'aise sont entre parenthèses) . La colonne formation indique le nombre d'années de travail vocal, ainsi que le type de formation (c : conservatoire, p : cours particuliers, ch : choeurs).*

Sujet	Tessiture	mécanisme I	mécanisme II	Formation	âge	Répertoire	remarques
B1	baryton	fa1 - la3 (sib1 - sol3)	sol2 - sol4	7 (p)	44	mélodie, oratorio, musique baroque	orthophoniste
B2	baryton lyrique	sol1 - sib3 (sib1 - sol3)	sol2 - réb4	8 (c)	26	romantique, baroque, XXième	
B3	baryton basse	ré1 - sib3 (fa#1 - sol3)	fa2 - fa4	10 (c+p)	30	baroque, lyrique	phoniatre
B4	baryton dramatique	mib1 - la3 (fa1 - sol#3)	mi2 - do4	5 + 3(c)	26	grand opéra	pédagogie vocale
B5	baryton basse	mi1 - fa3 (la1 - mib3)	non connu	10 (p)	32	baroque	organiste
B6	baryton	mib1 - sib3 (sol1 - sol3)	mi2 - mi4	10 (c+p)	35	romantique, classique, baroque allemand	pédagogie vocale
B7	baryton basse	ré1 - fa#3 (mi1 - fa3)	mi3 - sib3	10 (c)	29	grégorien, baroque	choriste
T1	ténor léger	sol1 - fa#3 (sol2 - fa#3)	fa#3 - la4	4 (p)	30	baroque français, classique	voix très mixée haute-contre
T2	ténor lyrique	mi1 - fa#3 (la1 - fa#3)	ré3 - lab4	13 (p)	32	romantique allemand, 18ième, 19ième	pédagogie vocale
CT1	contre-ténor (alto masculin)	fa#1 - la3 (fa#1 - fa#3)	fa#2 - la4 (la2 - fa#4)	5 (c+p)	23	baroque, contemporain	pédagogie vocale
CT2	contre-ténor (alto masculin)	ré1 - la3 (la1 - mi3)	mi2 - la4 (sol2 - fa#4)	12 (c)	36	baroque, romantisme, contemporain	biologiste, pédagogie vocale
CT3	contre-ténor (sopraniste)	sol1 - fa3 (si1 - ré3)	fa2 - si4 (sol2 - sol4)	5 (c+p)	22	baroque italien, castrats, Mozart	violoniste et pianiste
MS1	mezzo-soprano	fa2 - si3 (sol2 - fa3)	do3 - sib4 (do3 - sol4)	5 (c+ch)	25	baroque, lyrique	choriste, harmonie
MS2	mezzo-soprano	fa2 - sol3 (sol2 - do3)	si2 - sol#4 (do3 - sol4)	3 (c)	23	baroque, classique	choriste flûtiste
MS3	mezzo-soprano léger	fa#2 - sib3 (réb3 - la3)	do3 - do5 (la3 - sol4)	2(p) 5(ch)	22	lyrique, chants traditionnels (algérien, bulgare)	contre-bassiste
S1	soprano colorature	do2 - mib3	sib2 - do6 (do3 - fa5)	22 (c+p)	38	opéra lyrique colorature	voice coach pédagogie vocale
S2	soprano lyrique	mi2 - la3 (mi2 - fa3)	la2 - do#5 (la2 - si4)	7 (c+p)	31	lyrique 19ième	pédagogie vocale
S3	soprano léger	ré2 - ré3 (fa2 - ré3)	ré3 - sol5 (ré3 - fa5)	10 (c+p)	29	lyrique léger	

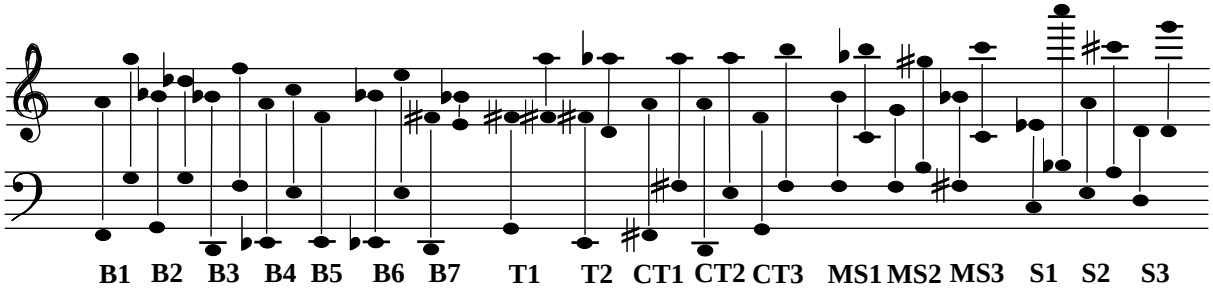


Fig. *6.3 – Visualisation de l'étendue vocale de chaque chanteur, en mécanisme I et II. Les notes présentées sur la figure proviennent des colonnes « mécanisme I » et « mécanisme II » du tableau 6.1 et correspondent aux notes extrêmes que le chanteur se sent capable d'émettre dans chaque mécanisme.*

6.3 Le protocole d'enregistrement

Une scéance d'enregistrement dure environ 20 *mn* en continu. Il est demandé au chanteur de suivre un protocole que nous allons détailler, en limitant au possible son vibrato et en caractérisant au besoin oralement sa production à la demande de l'expérimentateur, comme par exemple dans le cas d'un changement de mécanisme laryngé. Les chanteurs sont généralement à même de reconnaître avec quel mécanisme ils produisent un son donné et ces informations orales, recueillies en cours d'enregistrement, peuvent donc être utiles par la suite, en particulier dans le cas d'une indétermination.

Les paramètres mesurables lors de cette étude expérimentale sont : l'intensité acoustique I au niveau du microphone, la fréquence fondamentale f_0 et le quotient ouvert O_q . Nous nous intéressons aux variations du quotient ouvert dans des productions vocales parlées et chantées diverses, en tenant compte des deux autres paramètres. Le protocole a été élaboré pour permettre l'étude de ces variations lors de productions vocales isolées (voyelles tenues, sons filés) ou en contexte musical (phrase chantée). Voici en détail les points d'étude retenus et la tâche demandée en conséquence :

1. **variations de quotient ouvert au cours d'une phrase**, en fonction du type d'émission (parlé, chanté, crié) et, selon les possibilités du chanteur, en mécanisme I ou II.

   On demande au chanteur de choisir une phrase en français de son choix, souvent extraite de son répertoire, puis de l'énoncer et de la chanter. Il a été demandé aux premiers chanteurs passant le protocole de la crier également, mais cette tâche a été retirée par la suite, car ne présentant qu'un intérêt relatif pour cette étude et pouvant entraîner une fatigue vocale chez le chanteur dès le début du protocole, déjà relativement long.

   Afin qu'une comparaison puisse se faire entre chanteurs, une seconde phrase commune a été demandée, uniquement en production chantée (premières mesures de l' *Ave Maria* de Gounod).

2. **variations de quotient ouvert en fonction de l'intensité**.

   Il a été demandé au sujet de chanter des voyelles tenues sur trois intensités subjectives (piano, médium, forte) et des sons filés (sons crescendos et decrescendos) sur les mêmes voyelles. Les trois voyelles retenues sont : [a], [e] et [u], ce qui permet une exploration de trois plages distinctes du triangle vocalique.

   Afin de replacer cela dans un contexte dynamique et musical, il est également demandé de chanter les premières mesures de l' *Ave Maria* de Gounod avec ces trois intensités subjectives.

3. **variations de quotient ouvert en fonction de la hauteur**.

   Les voyelles tenues et les sons filés sont chantés sur différentes hauteurs sélectionées en fonction de la tessiture du chanteur. Dans le cas des contre-ténors et des altos, ces changements de hauteur peuvent s'accompagner d'un changement de mécanisme laryngé, que l'on s'attache alors à identifier avec le chanteur.

4. **variations de quotient ouvert lors des changements spontanés de mécanisme laryngé**.

   Les glissandos permettent de balayer toute la tessiture du chanteur, avec la possibilité d'un changement de mécanisme laryngé. La transition entre mécanismes I et II est un point de technique très travaillé par les voix d'alto homme et femme et le glissando est donc fait spontanément sans rupture perceptible au moment de la transition. Par contre, les barytons utilisent principalement le mécanisme I et les sopranos le mécanisme II. Dans ces cas, s'il y a, au cours du glissando, incursion dans le mécanisme peu utilisé par ces chanteurs, la transition est souvent très nette, marquée par un saut en fréquence.
   Il a donc été demandé aux chanteurs d'effectuer plusieurs glissandos lents et rapides avec utilisation (si possible) des deux mécanismes laryngés, de façon naturelle dans un premier temps, puis en ne cherchant pas à masquer le passage d'un mécanisme à un autre dans un second temps. Cette tâche a été placée à la fin du protocole, en raison de la fatigue vocale que peut entraîner la production des glissandos.

Ce protocole portait également sur une étude très générale de la qualité vocale liée à des changements de timbre et différentes tâches relatives à ce point ont été demandées en cours de protocole. Nous les exposerons très rapidement ici, bien que cette partie du protocole n'ait pas pu être analysée.

1. **modifications de timbre**.

   La qualité vocale est un terme un peu vague et nous avons voulu, dans un premier temps, laisser le chanteur libre de s'exprimer. Il a donc été demandé de reprendre la phrase chantée commune (*Ave Maria* de Gounod) et de la chanter en utilisant différentes qualités vocales, le choix de ces qualités étant laissé à l'appréciation du chanteur. Pour chaque exemple chanté, le chanteur devait exprimer par un qualificatif la qualité émise. Les qualificatifs le plus souvent cités par les chanteurs sont : voix sombre, voix claire, voix blanche.

2. **comparaison voix naturelle / voix lyrique**.

   Il nous a paru intéressant d'explorer la différence entre une production dite « naturelle » (i .e . sans technique vocale) et une production dite « lyrique » (i .e . avec technique vocale). Nous avons donc demandé au sujet de chanter un arpège très classique (*le petit Rossini*) de façon naturelle puis de façon lyrique. Dans le même esprit et comme les chanteurs éprouvent de grandes difficultés à oublier leur technique, nous avons sélectionné un chant populaire (air du « Joyeux anniversaire ») que nous leur avons demandé de chanter de façon « populaire » puis de façon lyrique.

Le tableau 6.2 récapitule les différentes étapes du protocole, dans l'ordre dans lequel elles ont été demandées. Cet ordre a été élaboré dans une étude préalable, sur les conseils de Marie-Agnès Faure (phoniatre), afin d'éviter une fatigue vocale du chanteur en début de protocole et pour contrebalancer des tâches longues et répétitives (sons filés) par des tâches plus agréables (arpège, air). Ce tableau rappelle également les points d'étude associés à chaque tâche demandée. Un exemple de protocole est fourni en Annexe C.

Une fois la segmentation en fichiers de données effectuée, une scéance d'enregistrement fournit en moyenne un total de 10 mn de signaux effectivement analysables par chanteur. Il y a en moyenne 80

Tab. *6.2 – Tableau récapitulatif des différentes étapes du protocole et de leur application.*

étape	tâche demandée	point d'étude
1	phrase en français parlée, chantée et criée	comparaison voix parlée / chantée variations de O_q sur une phrase
2	premières mesures de l'*Ave Maria* de Gounod	variations de O_q sur une phrase chantée : en fonction de l'intensité en fonction des mécanismes en fonction du timbre
3	arpège du « petit Rossini » sur différentes qualités	modification du timbre vocal comparaison voix naturelle / voix lyrique
4	voyelles tenues et sons filés	variations de O_q en fonction : de l'intensité de la hauteur de la voyelle
5	air du *Joyeux Anniversaire*	comparaison voix naturelle / voix lyrique
6	glissandos lent et rapide	visualisation des mécanismes laryngés détection des passages

paires de fichiers par chanteur (fichiers audio et EGG). La durée totale de cette base de données est de $3h\ 4mn\ 40s$.

Une fois mise en forme, cette base de données a pour vocation d'être largement diffusée, afin de permettre son utilisation par toute personne qui le jugerait utile pour ses travaux de recherche. En effet, de nombreuses recherches nécessitent de s'appuyer sur un enregistrement simultané des signaux acoustiques et électroglottographiques. Une seconde base de données (enregistrement de la lecture de 3 articles d'un journal par 4 locuteurs, avec utilisation des 3 premiers mécanismes laryngés) a également été enregistrée en cours de thèse et utilisée pour les travaux de Vu Ngoc et d'Alessandro [216]. Elle est actuellement en diffusion libre et a déjà été exploitée par de Cheveigné & Kawahara [50, 51] pour leurs travaux portant sur l'estimation de la fréquence fondamentale.

6.4 Conclusion

Nous avons élaboré et enregistré une base de données que nous pouvons exploiter pour analyser les liens entre fréquence fondamentale, quotient ouvert et les modifications de qualité liées à l'intensité vocale ou à un changement de timbre. Cette base est très riche et nous n'en exploiterons qu'une partie dans cette étude.

Avant de passer aux résultats des analyses, nous allons détailler les méthodes de mesure de la fréquence fondamentale et du quotient ouvert. Nous désirons nous appuyer sur les signaux DEGG qui nous semblent très adaptés pour ces mesures. La caractérisation et l'utilisation de ces signaux est l'objet du chapitre suivant.

En Résumé

Une base de données a été élaborée pour cette étude. Elle est constituée de l'enregistrement simultané des signaux acoustiques et électroglottographiques de 18 chanteurs : 7 barytons, 2 ténors, 3 contre-ténors, 3 mezzo-sopranos et 3 sopranos.

Dans le contexte de cette étude, les paramètres mesurés à partir de cette base de données sont :

l'intensité acoustique I, sur le signal acoustique calibré ;

la fréquence fondamentale f_0, sur la dérivée du signal EGG ;

le quotient ouvert O_q, sur la dérivée du signal EGG ;

Le protocole choisi pour ces enregistrements permet de couvrir des variations de fréquence fondamentale et d'intensité, sous forme de vocalises ou dans un contexte musical, d'appréhender les changements de mécanismes laryngés (par des glissandos ou sur une même phrase), de comparer différents types de production vocale (parlée, chantée, criée), ainsi que d'autres qualités vocales.

Chapitre 7

Etude des signaux DEGG

Les signaux électroglottographiques ont été caractérisés au chapitre 5. En particulier, nous avons noté la présence de pics marqués sur la dérivée de ces signaux, indiquant les instants de fermeture et d'ouverture glottique. Les signaux DEGG semblent donc très adaptés pour la mesure de la fréquence fondamentale et du quotient ouvert. Pourtant, ils ne sont que très peu exploités dans la littérature. Est-ce par manque d'une étude précise permettant de les caractériser ? Le but de ce chapitre est donc de discuter de l'utilisation des signaux DEGG pour la caractérisation du mouvement glottique et la mesure des paramètres de source glottique. Ces signaux présentent-ils des formes distinctes selon le mécanisme laryngé, la fréquence fondamentale ou l'intensité vocale ? Observe-t-on toujours des pics bien marqués à l'ouverture ou à la fermeture glottique ? Quelle est la validité des mesures de quotient ouvert effectuées sur ces signaux ? Le quotient ouvert mesuré sur un signal DEGG est-il comparable au quotient ouvert mesuré sur l'ODG estimée par filtrage inverse ?

Une bibliographie détaillée de l'utilisation des signaux DEGG sera donc présentée en partie 7.1. Nous approfondirons ensuite l'étude en partie 7.2, en analysant les signaux de la base de données présentée dans le précédent chapitre. Deux méthodes de mesure de la fréquence fondamentale et du quotient ouvert à partir de ces signaux seront proposées en partie 7.3. Les correspondances entre le quotient ouvert estimé sur un signal acoustique et mesuré par électroglottographie seront enfin discuté en partie 7.3.3, dans le cadre de la parole.

7.1 Des signaux électroglottographiques à leurs dérivées

Dans la littérature, la majorité des études utilisant l'électroglottographie comme moyen de mesure de f_0 et O_q s'appuie sur les signaux électroglottographiques eux-mêmes et non sur leurs dérivées. Les paramètres de source sont alors déterminés par une méthode de niveau (Rothenberg & Mahshie [173]) : une ligne de niveau est fixée, soit arbitrairement, soit en fonction d'un pourcentage sur l'ambitus en amplitude du signal EGG (50% dans le cas d'une voix normale à tendue et 35% dans le cas d'une voix relâchée), soit en égalisant l'aire calculée en-dessous et au-dessus d'un niveau choisi. Les points d'intersection entre cette ligne de niveau et le signal EGG sont utilisés comme valeurs approximatives des instants d'ouverture et de fermeture glottique, ainsi que l'illustre la figure de gauche 7.1. De la connaissance de ces instants, on peut alors déduire la période fondamentale (donc f_0) ainsi que le quotient ouvert ou fermé. Le principal avantage de cette méthode est qu'elle s'applique de façon robuste à tous les signaux EGG, même ceux de faible amplitude, ce qui est très appréciable en milieu médical. Son inconvénient majeur est qu'une telle mesure est par nature imprécise et ne correspond pas toujours à ce que l'on pourrait mesurer sur une fonction d'aire ou de débit glottique [173]. Ce point est illustré sur la figure de droite 7.1 dans le cas d'une voyelle émise de façon relâchée à pressée (voir chapitre 4, figure 4.16). Dans le cas des 2 méthodes de seuil (à 35 % et 50 % de l'amplitude du signal EGG), les instants de fermeture et d'ouverture sont associés aux

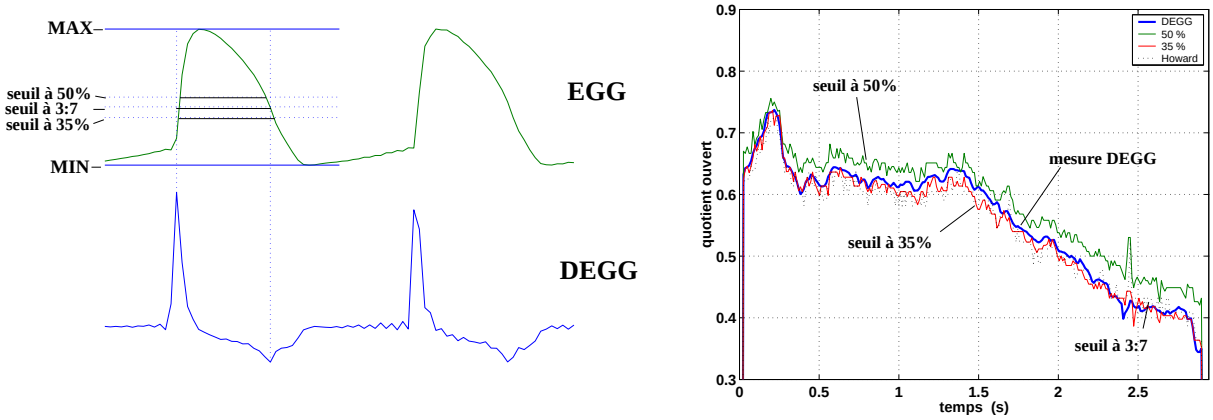


FIG. *7.1 – Principe de 4 méthodes de mesure du quotient ouvert par électroglottographie : sur le signal EGG par détection de seuil (35 %, 50 %), sur le signal DEGG par détection des pics d'ouverture et de fermeture (méthode DEGG) ou en combinant les deux signaux : l'instant de fermeture est détecté par le pic de fermeture sur le signal DEGG et l'instant d'ouverture par un seuil à 3/7 sur le signal EGG. La figure de droite illustre le résultat de ces mesures dans le cas d'une voyelle émise de façon relâchée à pressée. Nous quantifierons les différences entre ces méthodes en partie 7.3.3.*

instants où le signal dépasse ces seuils. Howard [113, 112] applique une méthode par seuil pour la détection de l'instant d'ouverture (seuil à 3/7), mais détecte de manière plus précise les instants de fermeture sur le signal DEGG. Nous présentons également une méthode basée uniquement sur les signaux DEGG que nous décrirons dans ce chapitre. Nous remarquons sur cette figure 7.1 une différence marquée entre les deux méthodes de seuil à 35 et 50 %, ainsi que de petites différences entre les autres méthodes. Nous effectuerons une comparaison quantitative entre ces 4 méthodes en partie 7.3.3.

Ces différences illustrent les imprécisions de la mesure dans le cas d'une méthode par seuil appliquée sur le signal EGG. Dans une étude comparative entre signaux EGG et ODG obtenue par filtrage inverse du débit aux lèvres, Rothenberg [171] (1981) a mis en évidence une bonne concordance entre les accentuations (changements de pente) observées sur le signal EGG à l'ouverture et à la fermeture et les instants d'ouverture et de fermeture visualisés sur l'ODG. Childers a été le premier à se pencher en 1983 [37] sur la caractérisation de ces points d'inflexion du signal EGG par utilisation de la dérivée. Il a entrepris des travaux comparatifs entre signaux EGG et DEGG, onde de débit glottique estimée par filtrage inverse du signal de parole et aire glottique mesurée à partir d'images obtenues par cinématographie ultra-rapide [36, 37, 34, 33, 31] (de 1982 à 90). Ses recherches expérimentales, couplées avec une étude théorique comparant le mouvement vibratoire des cordes vocales et l'aire de contact calculée par modèle physique [30], ont montré que les instants de fermeture et d'ouverture glottique coïncidaient avec les pics de fermeture et d'ouverture observés sur le signal DEGG, dans le cas de productions vocales non pathologiques. Néanmoins, tous les sujets étudiés étaient de sexe masculin et seul le mécanisme I a été exploré. Il est à noter également qu'un cas très particulier fait exception : celui où un lien constitué de mucus subsiste entre les cordes vocales à l'ouverture de la glotte, réalisant ainsi un pont électrique entre les deux cordes en train de se séparer. Le pic d'ouverture observé sur le signal DEGG correspond alors à l'instant de rupture du filet de mucus, et non à l'instant de séparation des cordes vocales et succède donc à l'instant d'ouverture mesuré sur la fonction d'aire glottique. Néanmoins, en terme d'onde de débit, l'effet de ce filet de mucus équivaut à un arrêt du débit sur toute la surface glottique couverte. Si cette surface est importante comparativement à l'aire glottique, on peut raisonnablement penser que l'instant d'ouverture mesuré sur le signal DEGG à la rupture du filet n'est pas significativement différent de celui que l'on pourrait mesurer sur un signal de débit glottique. En effet, Colton & Conture [40] ont observé que la présence d'un tel filet provoquait une petite variation sur le tracé du signal EGG, de même que sur celui de l'ODG obtenue par filtrage inverse du débit d'air aux lèvres.

D'autres études ont appuyé les résultats obtenus par Childers. Une étude comparative entre signaux électroglottographiques, photoglottographiques, de pression sous-glottique et des photographies stroboscopiques a été menée *in vivo* sur des chiens par Berke & al [21] (1987), les signaux glottographiques enregistrés sur les canins étant comparables à ceux enregistrés sur les humains [21]. Ils ont observé que les changements de pente du signal EGG étaient associés à l'initialisation de l'ouverture sur la partie supérieure des cordes vocales et à la fermeture complète de la partie inférieure des cordes vocales. Anastaplo & Karnell [11] (1988) se sont intéressés tout particulièrement à l'instant d'ouverture glottique, qu'ils ont étudié à l'aide de la vidéostroboscopie synchronisée avec le signal électroglottographique. Leurs résultats sur 4 sujets (2 hommes, 2 femmes) ont également montré que le changement de pente observé sur le signal EGG marquait le début de l'ouverture glottique le long de la partie supérieure des cordes vocales. Cranen [42](1991) a effectué des comparaisons entre signaux électroglottographiques, photoglottographiques et d'onde de débit estimée par filtrage inverse, sur des segments voyelle-consomne-voyelle. Les pics observés sur le signal DEGG coïncidaient parfaitement avec ceux de la dérivée du signal photoglottographique, ainsi qu'avec les instants d'ouverture et de fermeture discriminables sur l'ODG et sa dérivée.

Néanmoins, une étude très récente menée par Hess & Ludwigs [106] (2000) a avancé que l'inflexion observée sur le signal EGG pourrait être plutôt reliée à l'ouverture de la partie inférieure des cordes vocales. La méthode de visualisation utilisée, appelée transillumination strobophotoglottographique, consiste à éclairer la glotte simultanément par le dessous (sur le principe de la photoglottographie, une source lumineuse est appliquée sur le cou du sujet, face au ligament crico-thyroïdien) et par le dessus (l'utilisation d'une lumière bleue pour l'éclairage supérieur permet de différencier les deux sources lumineuses à l'aide de filtre, la lumière traversant les tissus du cou paraissant rouge). Cette méthode de visualisation très astucieuse et prometteuse n'en est qu'à ses débuts. Pour l'instant, les comparaisons entre le signal EGG et l'image obtenue ont été approximatives, en ce sens que la position de l'inflexion du signal EGG à l'ouverture a été déterminée arbitrairement et visuellement sur un moniteur de contrôle, ce qui limite la précision temporelle. Dans les exemples de signaux EGG et DEGG que nous présentons par la suite, on peut remarquer que l'inflexion observée sur le signal EGG n'est pas forcément un phénomène instantané, surtout dans le cas de l'ouverture, et qu'il est donc nécesssaire d'avoir recours à la dérivée des signaux EGG pour déterminer les instants exacts des changements de pente du signal EGG. Les images présentées par Hess & Ludwigs au niveau de l'inflexion (diffusion de la lumière par le dessous des cordes) et très légèrement après (ouverture médiane déjà importante) donnent même à penser que cette méthode nouvelle de visualisation serait tout à fait appropriée pour accréditer (s'il le fallait encore) le fait que le pic d'ouverture du signal DEGG représente l'instant d'initialisation du flux glottique. Une collaboration avec cette équipe est en cours.

7.1.1 Discussion sur la validité des mesures

Au regard de ces études, la détection des instants de fermeture et d'ouverture glottique sur le signal DEGG peut être considérée comme précise et appropriée. La fermeture étant généralement brutale, le pic de fermeture est souvent très marqué. L'ouverture peut être plus molle, et le pic qui lui correspond dans le signal DEGG est bien moins marqué que celui associé à la fermeture, ce qui explique les réserves exprimées par Baken [16] (1992) quant à l'utilisation de ce pic comme mesure de l'instant d'ouverture glottique. En raison de ces incertitudes, il nous a paru important de visualiser au préalable les signaux DEGG de la base de données, afin de déterminer les signaux exploitables et d'établir un critère de sélection.

7.2 Analyse des signaux DEGG de la base de données

7.2.1 Pics bien marqués

Sur les figures 5.4 et 5.5 présentées en 5.2.3, nous avons pu voir la bonne concordance entre les instants de fermeture et d'ouverture glottique et les pics marqués présents sur le signal DEGG. La figure

7.2 présente des exemples caractéristiques extraits de la base de données pour lesquels les pics de fermeture et d'ouverture observés sur les signaux DEGG sont bien marqués.

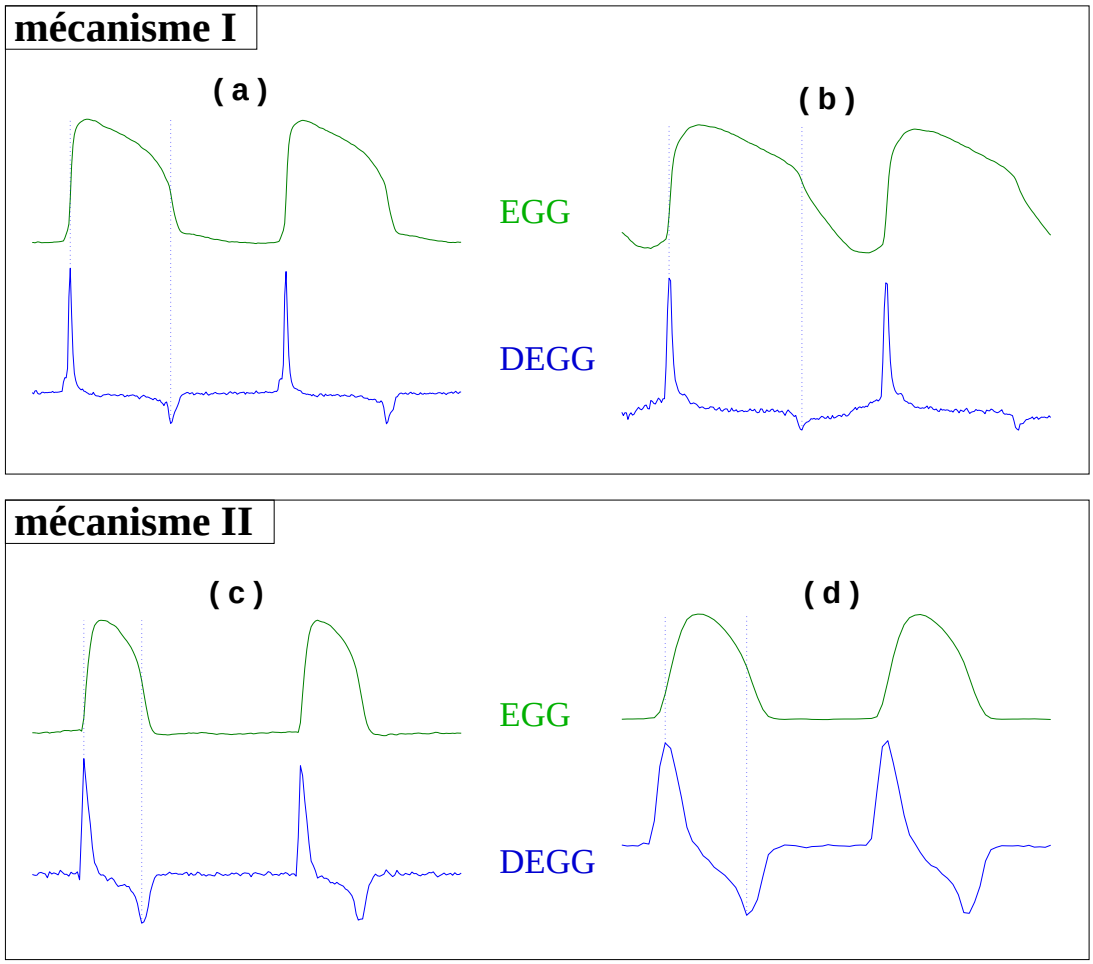


FIG. *7.2 – Exemples caractéristiques de signaux EGG et DEGG, extraits de la base de données, pour lesquels les pics correspondant à l'ouverture et à la fermeture sont bien marqués. (a) sujet B2, voyelle [u], do3 (*260 Hz*) ; (b) sujet B4, voyelle [e], mi3 (*330 Hz*) ; (c) sujet S2, voyelle [a], la3 (*440 Hz*) ; (d) sujet S3, voyelle [a], do5 (*1040 Hz*).*

Les exemples (a) et (b) correspondent à des productions en mécanisme I, provenant de crescendos chantés par deux barytons (sujets B2 et B4). Ces productions sont caractérisées par un temps fermé notable : celui-ci est précisé sur la figure 7.2 par la distance entre les deux lignes verticales successives. Les exemples (c) et (d) correspondent à des productions en mécanisme II, provenant de crescendos chantés par deux sopranos (sujets S2 et S3). L'exemple (d) correspond à un contre-ut chanté par le sujet S3 (do5, 1040 Hz). A cette hauteur, l'ouverture et la fermeture sont comparables. Cette tendance se dessine déjà sur l'exemple (c), qui correspond pourtant à une note beaucoup plus basse (la3, 440 Hz). La forme de ces signaux peut s'expliquer par le mécanisme laryngé utilisé, comme nous l'avons vu en 5.2.3.

7.2.2 Pics dédoublés ou peu marqués

Imprécision sur le pic d'ouverture

On observe également des cas où le pic d'ouverture est peu marqué, ce qui le rend indéterminable. La figure 7.3 illustre ces cas sur deux exemples typiques ((e) et (f)).

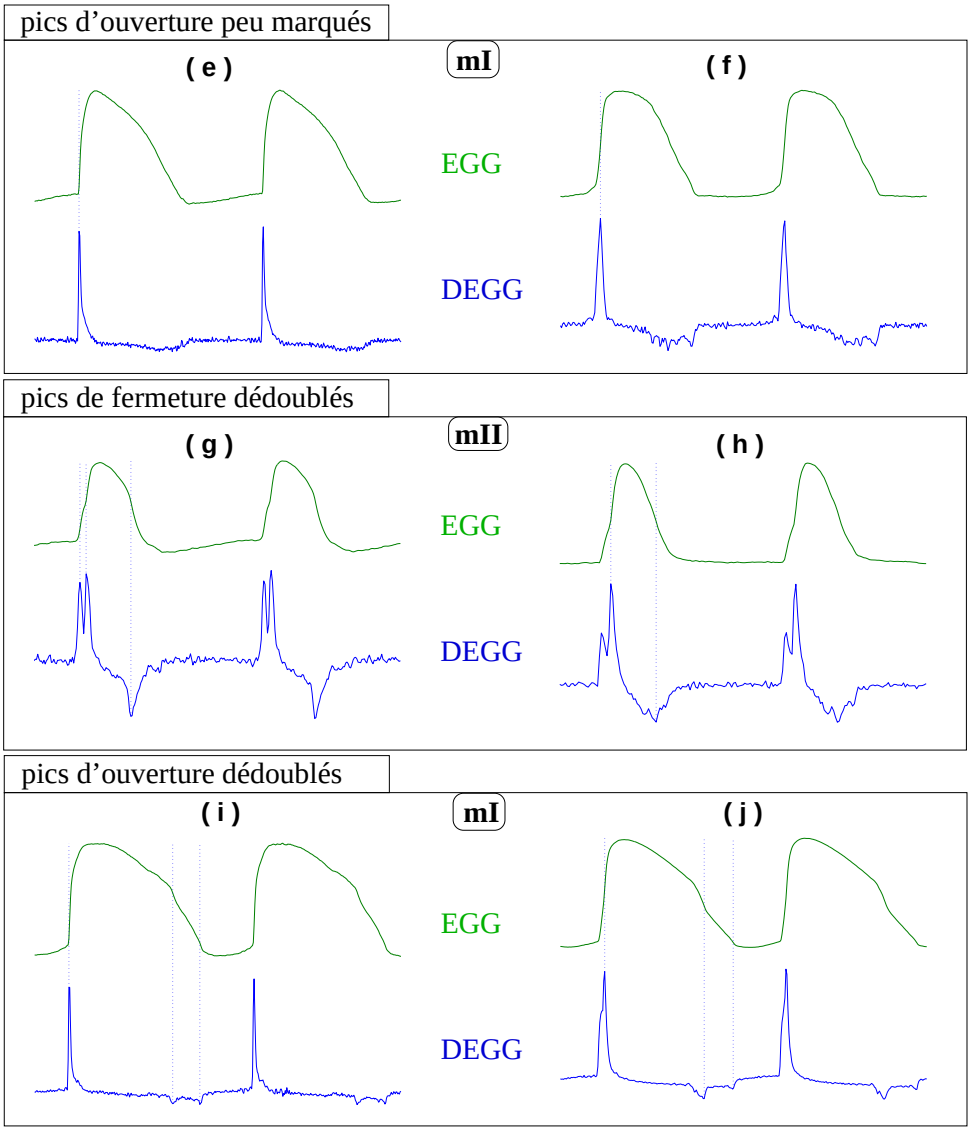


Fig. *7.3 – Exemples caractéristiques de signaux EGG et DEGG, extraits de la base de données, pour lesquels le pic d'ouverture est peu marqué ou pour lesquels on observe un dédoublement du pic d'ouverture ou de fermeture. (e) sujet MS2, voyelle [a], fa2 # (*185 Hz*) ; (f) sujet B2, voyelle [u], do3 (*260 Hz*) ; (g) sujet CT3, voyelle [a], mi3 (*330 Hz*) ; (h) sujet S2, voyelle [a], la3 (*440 Hz*) ; (i) sujet B4, voyelle [a], sol2 (*196 Hz*) ; (j) sujet B6, voyelle [a], si2 (*245 Hz*).*

Dans le premier cas (e), il n'y a quasiment pas de pic à l'ouverture, si ce n'est une légère bosse très étalée. Dans le second cas (f), il y a plusieurs petits pics successifs, ce qui rend la mesure du quotient ouvert inadaptée. Ces observations peuvent s'expliquer par une ouverture progressive de la glotte, qui rend l'instant d'ouverture indéfinissable. Cette hypothèse est illustrée par la figure 7.4, qui présente une

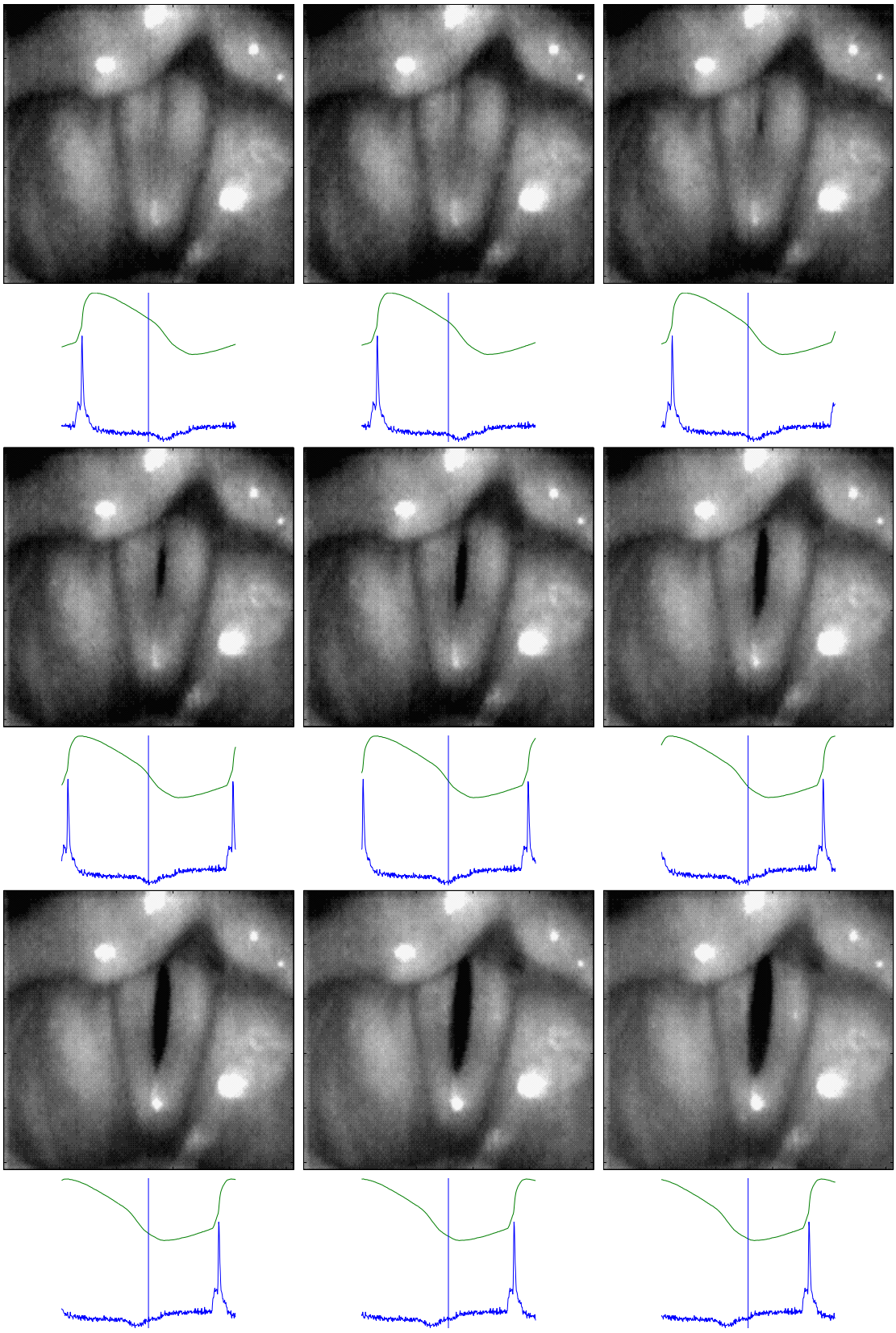


FIG. 7.4 – *Visualisation d'une ouverture peu marquée sur le signal DEGG par cinématographie ultra-rapide et électroglottographie simultanées (locuteur en phonation normale, $f_0 = 110$ Hz).*

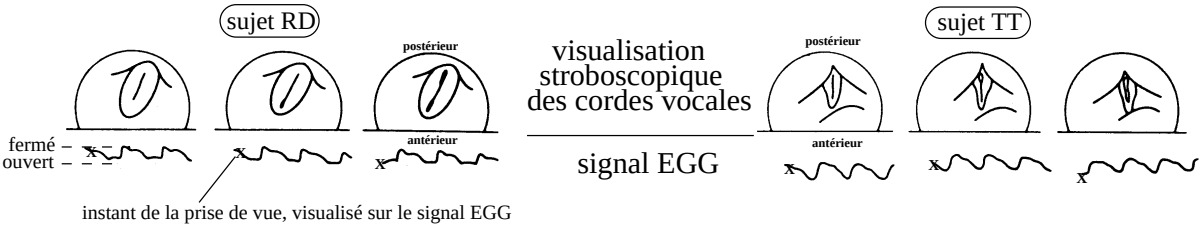


Fig. *7.5 – Visualisation d'images de la glotte obtenues par vidéostroboscopie synchronisée à l'électroglottographie, pour deux sujets présentant une ouverture sur deux parties distinctes de la glotte (d'après Karnell [125]). Pour le sujet RD, l'ouverture sur la partie antérieure des cordes vocales précède l'ouverture sur la partie postérieure. Pour le sujet TT, l'ouverture a d'abord lieu au niveau des aryténoïdes (partie postérieure des cordes vocales).*

visualisation simultanée de la glotte et des signaux EGG et DEGG dans le cas d'une ouverture peu marquée. Le signal DEGG a été amplifié afin de permettre la visualisation de la courbure à l'ouverture. L'ouverture se fait de façon médiane (i. e. par le milieu de la glotte) et très graduelle.

Dédoublement marqué des pics de fermeture et d'ouverture

Cette imprécision sur le pic d'ouverture ne se retrouve jamais dans le cas du pic de fermeture. La seule indétermination sur l'instant de fermeture provient de cas où le pic de fermeture est très nettement dédoublé (exemples (g) et (h) de la figure 7.3). Ce phénomène s'observe également à l'ouverture (exemples (i) et (j) de la figure 7.3), bien que les pics soient alors moins marqués et beaucoup moins rapprochés que dans le cas de la fermeture. Dans les cas (g) et (i), les deux pics ont même amplitude, tandis que dans les cas (h) et (j), un des deux pics est plus marqué. Il est à noter que l'on n'observe que très rarement un triplement ou un quadruplement marqué de ces pics, tant à la fermeture qu'à l'ouverture.

Discussion

Ces observations pourraient s'expliquer par une ouverture (ou fermeture) sur deux parties distinctes de la glotte, de façon quasiment simultanée ou légèrement différée, ou par une ouverture rapide des cordes sur l'épaisseur, suivie d'une ouverture sur la longueur. A l'appui de la première hypothèse (ouverture déphasée sur la longueur des cordes), de tels signaux ont été observés par Karnell [125] lors d'une étude alliant vidéostroboscopie synchronisée à l'électroglottographie. La visualisation des cordes vocales a permis de montrer une ouverture au niveau antérieur de la glotte précédant l'ouverture de la partie postérieure [125], comme le montre la figure 7.5.

Ce type de comportement glottique n'est en fait pas rare, puisqu'il avait déjà été observé dans une étude antérieure réalisée par Anastaplo & Karnell [11] et qu'il a également été observé très récemment par Hess & Ludwigs [106]. Ces derniers ont d'ailleurs noté une correspondance entre la qualité vocale et le type d'ouverture glottique, schématisé sur la figure 7.6 : ouverture postérieure à antérieure dans le cas d'une voix relâchée, médiane dans le cas d'une voix normale et antérieure à postérieure dans le cas d'une voix tendue.

Ces observations correspondent tout à fait à ce que l'on attendrait intuitivement. En effet, dans le cas d'une production vocale tendue, les aryténoïdes exercent une pression importante au niveau postérieur et l'ouverture se fait alors là où la résistance au passage de l'air est moindre, c'est-à-dire dans la partie antérieure de la glotte. Dans le cas d'une production vocale relâchée, la compression au niveau des aryténoïdes est moindre et l'air peut s'échapper à partir de la partie postérieure de la glotte. Souvent même, les aryténoïdes sont complètement séparés, laissant une petite ouverture triangulaire, comme dans le cas des fuites glottiques.

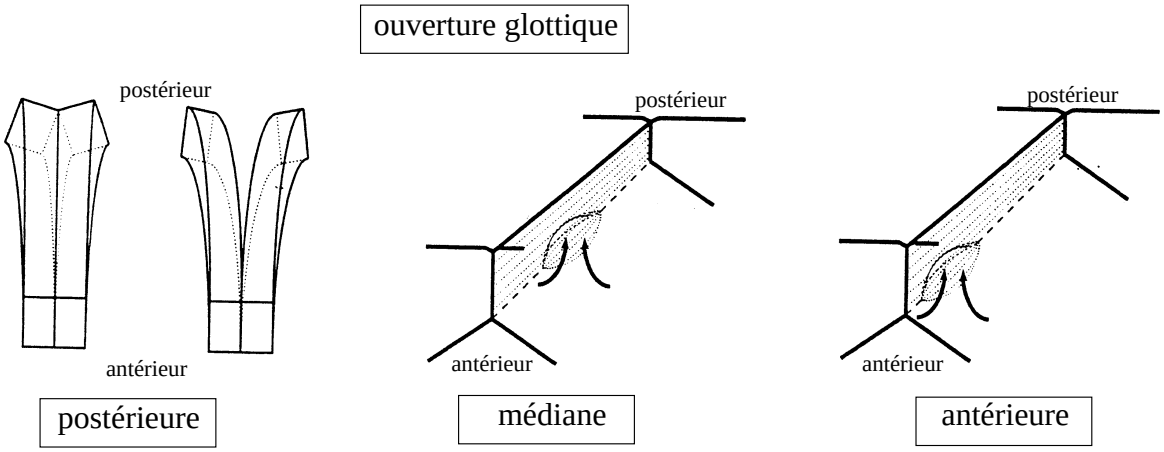


FIG. *7.6 – Types d'ouverture glottique observés par Hess & Ludwigs* [106] *: ouverture postérieure (par les aryténoïdes), médiane (par le milieu de la glotte) ou antérieure (par l'avant de la glotte) (d'après Hess & Ludwigs* [106]*).*

Visualisation

Sans prétendre à expliquer ces phénomènes de façon générale, la figure 7.7 présente la visualisation simultanée des images cinématographiques et des signaux EGG et DEGG dans le cas d'un dédoublement du pic de fermeture. Cet extrait provient d'un début de phonation. Sur la figure présentant le signal EGG et sa dérivée sur quelques périodes, on observe que l'initialisation de la phonation se caractérise ici par un dédoublement du pic de fermeture, phénomène qui s'estompe au bout de 4 périodes. Les images montrent que ce dédoublement correspond à la fermeture antérieure de la glotte (premier pic) suivie d'une fermeture postérieure au niveau des aryténoïdes (second pic), ce qui confirme les remarques précédentes, car on peut raisonnablement penser que la compression au niveau des aryténoïdes est moindre en début de phonation (phase relâchée).

La figure 7.8 illustre le cas d'un dédoublement du pic d'ouverture. Le premier pic, qui est également le pic le plus ample, correspond à l'ouverture postérieure de la glotte. On peut remarquer que l'ouverture se fait par le tiers postérieur de la corde, mais n'est pas localisée au niveau des aryténoïdes. Le second pic correspond à l'accentuation de l'ouverture sur la partie antérieure.

Les exemples (a) et (f) (figures 7.2 et 7.3) sont extraits du même signal, un crescendo chanté par le baryton B2 sur une note tenue (do3). Dans ce cas-ci, le début du crescendo est marqué par une imprécision de l'ouverture, qui s'efface quand le chanteur monte en intensité. De même, les exemples (c) et (h) proviennent d'un crescendo chanté par la soprano S2. Dans ce cas-là par contre, la montée en intensité s'accompagne d'un dédoublement du pic de fermeture. Le tableau 7.1 résume les observations faites sur la base de données, pour tous les chanteurs. Seuls les voyelles tenues et les crescendos ont été visualisés. On observe une grande variabilité entre chanteurs et aucune tendance marquée ne ressort de ce tableau. Ainsi, certains chanteurs peuvent présenter un dédoublement du pic d'ouverture quand ils montent en intensité (B3, B4, T2), alors que pour d'autres chanteurs, on observe exactement l'inverse (B2, CT2, S1).

Dans ses conclusions, Karnell [125] fait remarquer que, sans l'appui visuel des images de la glotte, ces irrégularités observées sur le signal EGG auraient été mal interprétées et considérées comme des artefacts de mesure. Dans notre cas, nous n'avons pas à disposition de visualisations directes de la glotte qui permettent de lever ces doutes. Néanmoins, nous pouvons noter que ce phénomène ne présente pas les caractéristiques propres à un artefact de mesure, puisqu'il n'est pas régulier et que, selon le chanteur, il peut dépendre des caractéristiques de la phonation, telle l'intensité ou la hauteur du son produit. Il serait donc très intéressant de pouvoir relier précisément ces phénomènes observés sur le signal

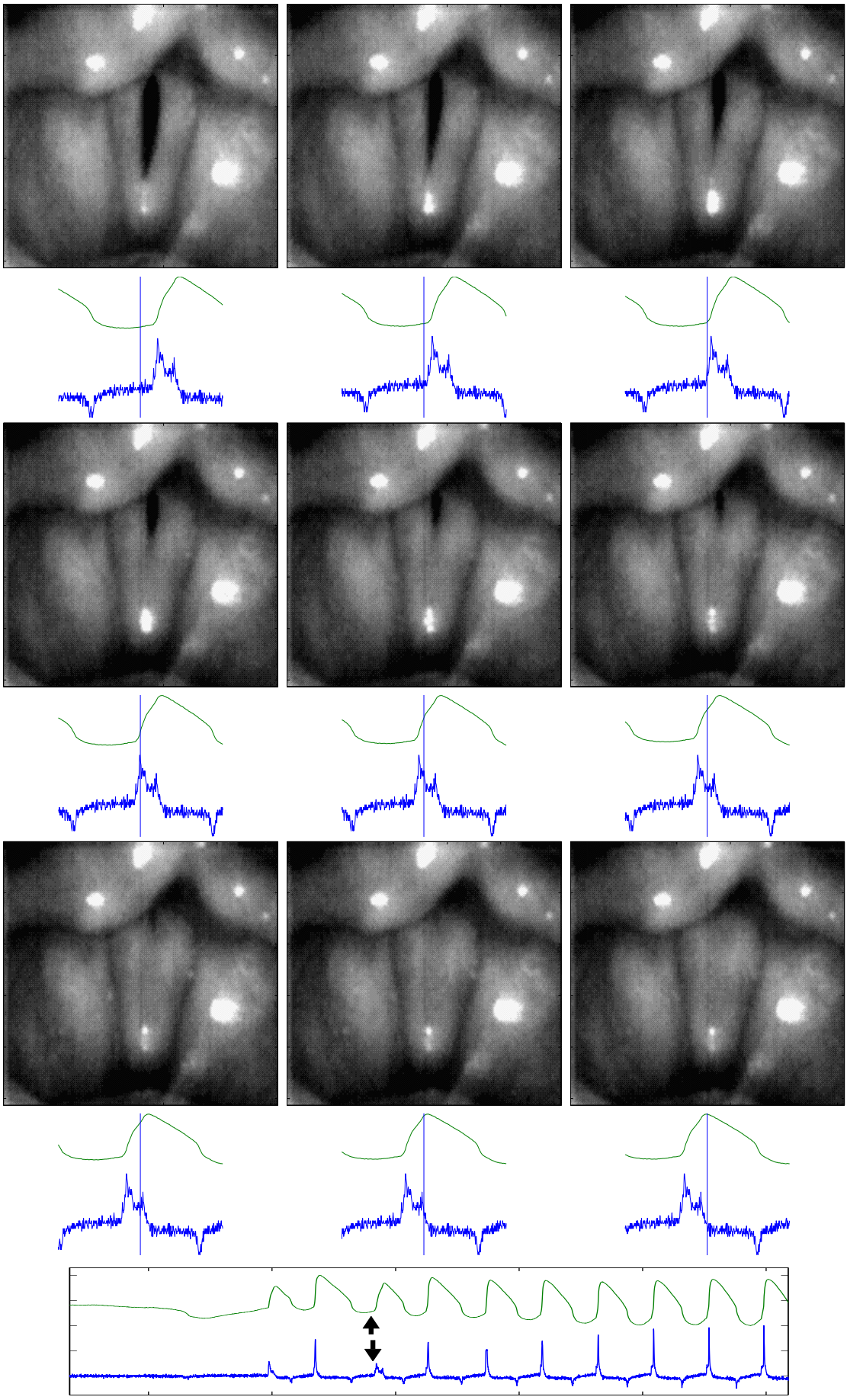


Fig. *7.7 – Visualisation d'un dédoublement du pic de fermeture par cinématographie ultra-rapide et électroglottographie simultanées (locuteur en phonation normale, $f_0 = 110$ Hz).*

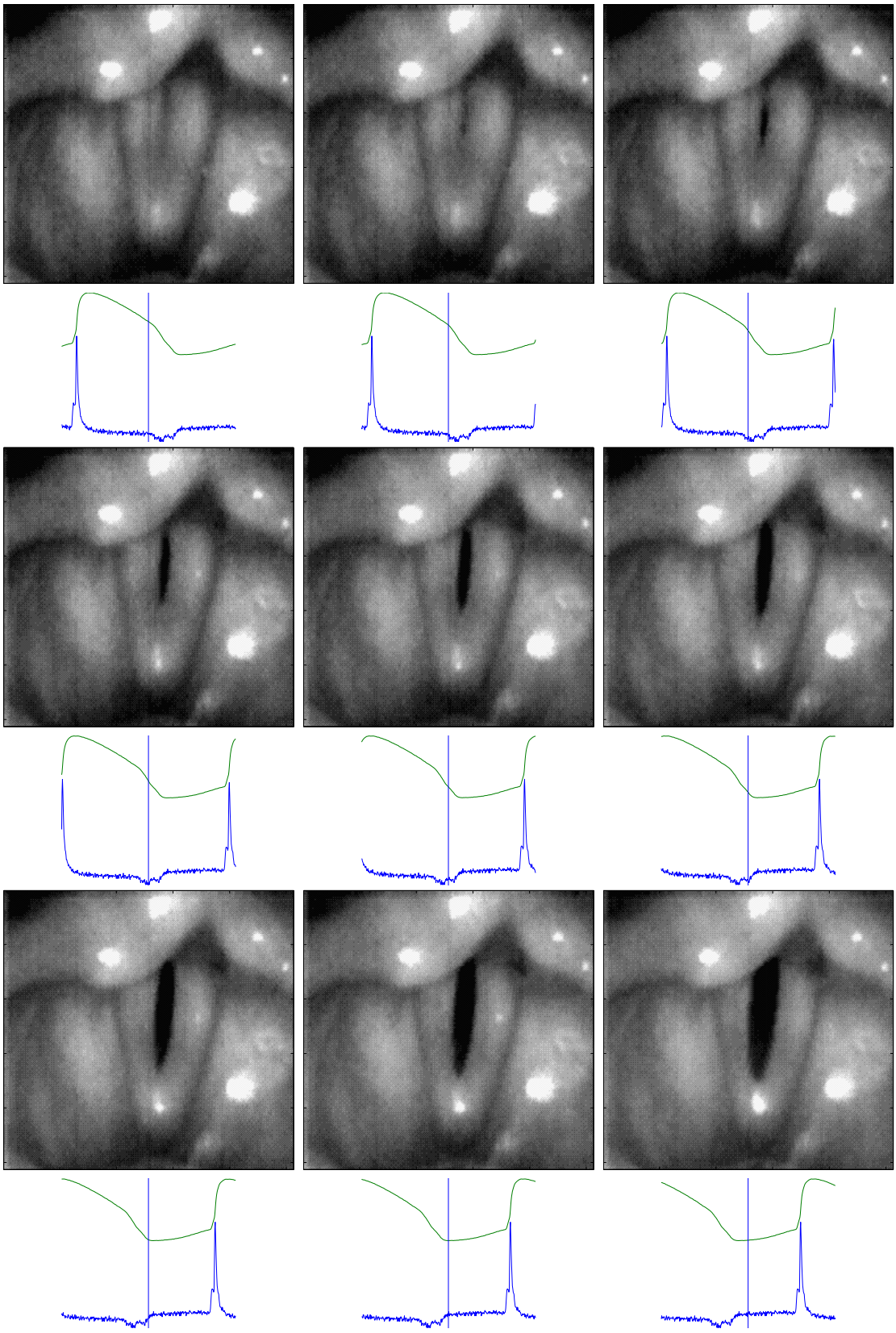


FIG. 7.8 – *Visualisation d'un dédoublement du pic d'ouverture par cinématographie ultra-rapide et électroglottographie simultanées (locuteur en phonation normale, $f_0 = 110$ Hz).*

Tab. *7.1 – Observations des pics de fermeture et d'ouverture sur les signaux DEGG de la base de données, dans le cas des voyelles tenues et des sons filés. « X » signifie « observé fréquemment ». Quand les observations dépendent de la hauteur et/ou de l'intensité, elles sont précisées par : HF (« observé à haute fréquence »), BF (« observé à basse fréquence »), p (« observé pour les sons piano »), m (« observé pour les sons mezzo-forte »), f (« observé pour les sons forte »). Pour chaque catégorie, on rappelle les exemples types : les exemples (a),(b),(c) et (d) se rapportent à la figure 7.2 et les exemples (e),(g),(h),(i) et (j) à la figure 7.3.*

chanteur	**pics marqués mI** (ex. (a) (b))	**pics marqués mII** (ex. (c) (d))	**pics peu marqués** (ex. (e))	**pics de fermeture dédoublé** (ex. (g) (h))	**pics d'ouverture dédoublé** (ex. (i) (j))
B1	m-f		BF	HF	
B2	X		X		p
B3	p-m				f
B4	X				f
B5	X			HF	HF
B6	X			X	X
B7	BF		p	X	X
T1	BF		X	BF	HF
T2	X				f
CT1	X		X	X	X
CT2	X				HF / p
CT3		HF		X	
MS1	X		BF		
MS2	BF		X		X
MS3	HF / m-f				
S1	X	X			p
S2	X	X		X	X
S3	X	X		p	

DEGG des chanteurs à des visualisations de la glotte, tel que nous avons essayé de le faire à partir des données portant sur la phonation d'un locuteur, aimablement prêtées par l'équipe du *Department of Phoniatrics and Pediatric Audiology* de l' *University of Erlangen-Nuremberg* (Allemagne). Une telle étude n'a jamais été réalisée et pourtant son apport serait fondamental pour la caractérisation et l'utilisation des signaux DEGG. Devant l'importance de détecter précisement les instants d'ouverture et de fermeture glottique pour nombre de recherches sur la voix (analyse de la source glottique, détection de la fréquence fondamentale, ...), les signaux DEGG, très simples à enregistrer, présentent des informations essentielles que l'on ne peut, à l'heure actuelle, que partiellement exploiter. Nous allons illustrer ce point sur un exemple concret.

7.2.3 Visualisation des passages entre mécanismes laryngés sur les signaux DEGG

Les études menées par Lecluse [137, 138], Askenfelt et al. [12], Kitzing [127], Roubeau [177, 174, 175] ont montré l'influence des mécanismes laryngés sur les signaux EGG, tant sur la forme que sur l'amplitude de ces signaux. La transition entre mécanismes se caractérise par une modification brutale de la forme du signal sur environ deux périodes de voisement (Roubeau [177], Svec [194]). En parallèle, un changement plus progressif s'opère au niveau de l'amplitude (sur une durée de l'ordre de 130-150 ms d'après Roubeau [177]), suivi d'une perte de contrôle en fréquence (Roubeau [177], Miller [154], Svec [194]).

Ces variations se retrouvent sur les signaux DEGG, comme l'illustre la figure 7.9 qui présente les signaux EGG et DEGG lors d'un glissando chanté par le ténor T2 (glissando présenté sur la figure 8.1 (CD, piste [40])). La visualisation des signaux DEGG peut compléter celle des signaux EGG pour la caractérisation des passages entre mécanismes. La figure 7.10 présente les signaux EGG et DEGG provenant du glissando chanté par le baryton B6 (glissando présenté dans le recueil de figures, (CD, piste [38])). Peu avant le passage (mII-mI), les signaux EGG et DEGG présentent un dédoublement marqué des pics d'ouverture et de fermeture. Pendant le passage, l'amplitude marquée d'un des deux pics diminue progressivement tandis que celle de l'autre pic augmente, que ce soit à l'ouverture ou à la fermeture. On aboutit après le passage à une unicité des pics à l'ouverture et à la fermeture. Le passage (mII-mI) présente les mêmes caractéristiques : unicité des pics d'ouverture et de fermeture en mécanisme II, dédoublement progressif du pic de fermeture et du pic d'ouverture. Les signaux DEGG montrent des détails qui passeraient inaperçus à la seule visualisation des signaux EGG. On remarque par exemple que, si le passage d'un mécanisme à l'autre semble très brusque (quelques périodes) à la vue du changement d'amplitude sur l'enveloppe des signaux EGG ou DEGG, des « ajustements » s'opèrent pendant quelques dizaines de périodes. Malheureusement, il est difficile d'expliquer ces observations sans avoir recours à une visualisation directe des cordes vocales.

En conclusion, les signaux DEGG permettent de montrer des détails du mouvement d'ouverture et de fermeture glottique, bien que ces détails ne puissent pas toujours être interprétés sans l'apport de la visualisation directe ou de signaux complémentaires (par exemple, l'ODG). Pour effectuer des mesures quantitatives à l'aide de ces signaux, une détection automatique des pics a été établie. Dans le cas où les pics de fermeture et d'ouverture observés sur ces signaux sont précis et non dédoublés, cette détection permet une mesure de la fréquence fondamentale et du quotient ouvert, comme nous l'avons illustré dans le chapitre 5 sur la figure 5.7. Nous allons maintenant détailler l'algorithme mis en place pour l'analyse des signaux DEGG.

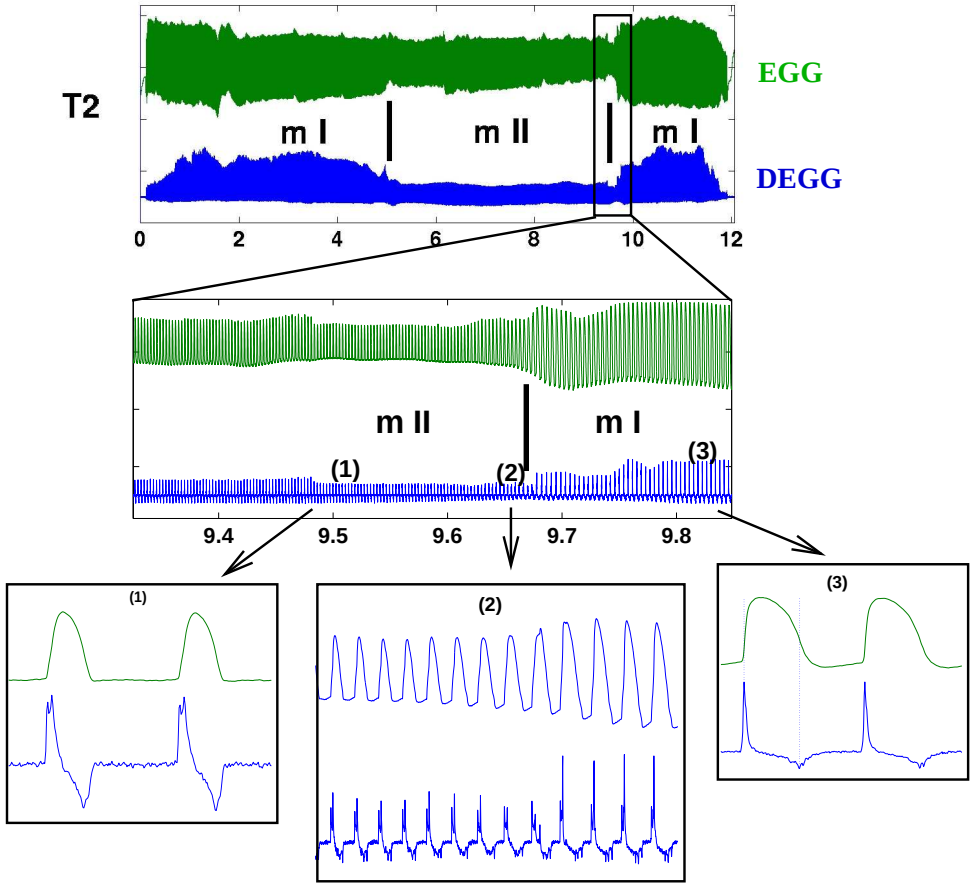


Fig. *7.9 – Visualisation des signaux EGG et DEGG lors d'un glissando effectué par le ténor T2. La figure du dessus présente la forme générale des signaux sur toute la phrase. Les passages entre mécanismes laryngés se traduisent par un changement marqué d'amplitude. La figure du milieu focalise sur le passage (mII - mI) (un peu avant* $t = 9.7$ s*) et les 3 figures du bas présentent quelques périodes des signaux EGG et DEGG juste avant (1), pendant (2) et après (3) le passage. (CD, piste [40])*

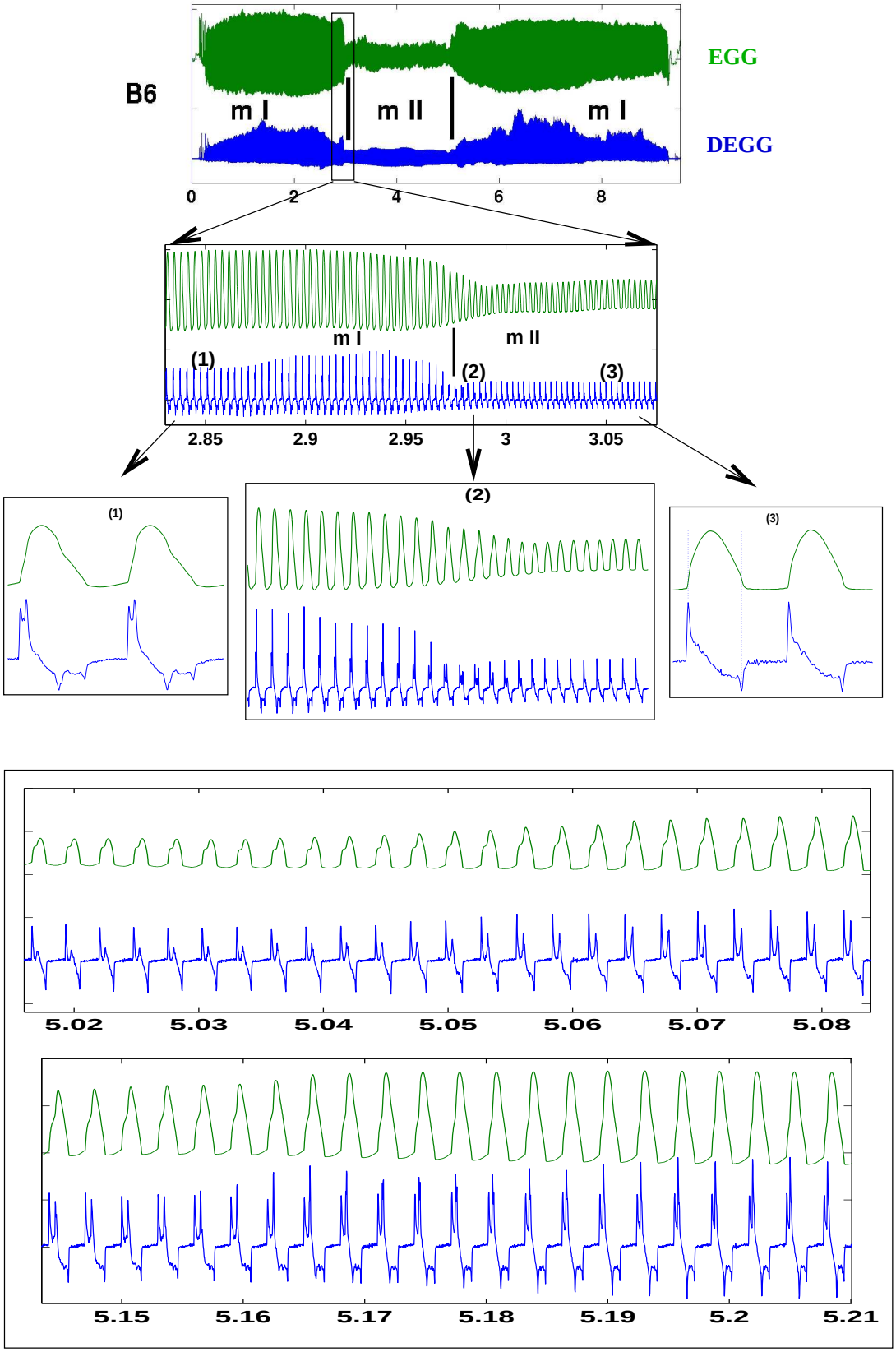


FIG. *7.10 – Visualisation des signaux EGG et DEGG lors d'un glissando effectué par le baryton B6. La figure du dessus présente la forme générale des signaux sur toute la phrase. Comme sur l'exemple précédent, les passages entre mécanismes laryngés se traduisent par un changement marqué d'amplitude. Les figures du milieu focalisent sur le passage (mI - mII) (un peu avant* $t = 3$ s*) et présentent quelques périodes des signaux EGG et DEGG juste avant (1), pendant (2) et après (3) le passage. La figure du bas fournit une visualisation détaillée du passage (mII - mI). (CD, piste [38])*

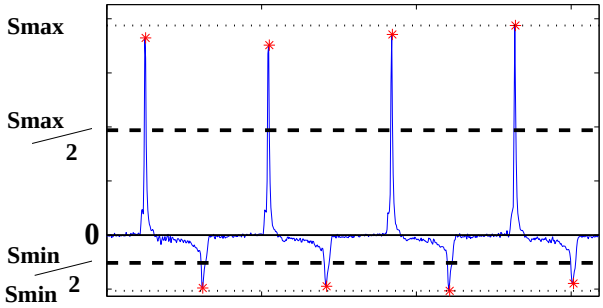


Fig. *7.11 – Détection des pics de fermeture et d'ouverture par seuillage : les seuils sont placés ici à* 50% *du maximum (détection des pics de fermeture) et du minimum (détection des pics d'ouverture).*

7.3 Mesure de la fréquence fondamentale et du quotient ouvert

7.3.1 Détection directe par seuillage

Une méthode très simple de mesure de la fréquence fondamentale et du quotient ouvert sur un signal DEGG revient à déterminer la position des maxima situés au-dessus d'un certain seuil, ainsi que l'illustre la figure 7.11.

Cette méthode est simple à appliquer, et elle donne de bons résultats quand les pics sont, comme sur cet exemple, très marqués. Quand le pic d'ouverture est étalé ou peu marqué, elle ne convient plus. Nous avons donc choisi d'appliquer ici une méthode plus générale et plus robuste, basée non pas sur une détection directe des pics, mais sur la détermination par corrélation de l'écart entre les pics.

7.3.2 Mesure par corrélation

Description de l'algorithme d'analyse

L'algorithme d'analyse du signal DEGG est représenté sur la figure 7.12 et se décompose comme suit :

1. **sélection des trames d'analyse**

   Les trames d'analyse ont une longueur de 4 périodes, ce qui suppose donc de connaître, au moins de façon approximative, la période fondamentale. En cours d'analyse, celle-ci peut être approchée par la valeur calculée sur la trame précédente. Si ce n'est pas le cas, comme lors d'un début d'analyse ou dans le cas de non-voisement pour la trame précédente, un fenêtrage de taille fixe est appliqué : 40 ms pour une voix d'homme et 20 ms pour une voix de femme. Le pas entre deux trames successives est de 2 périodes ou 5 ms si la valeur de la période n'est pas connue.

2. **séparation du signal en partie positive (pics de fermeture) et en partie négative (pics d'ouverture)**

   On génère ainsi deux signaux S_f (f pour « fermeture ») et S_o (o pour « ouverture »). Le signal S_f est composé de la partie positive du signal d'analyse, la partie négative étant mise à 0. Le signal S_o est composé de l'opposée de la partie négative du signal d'analyse, la partie positive étant mise à 0. S_o, tout comme S_f, est alors un signal positif ou nul. Dans la suite, l'analyse se fera à partir

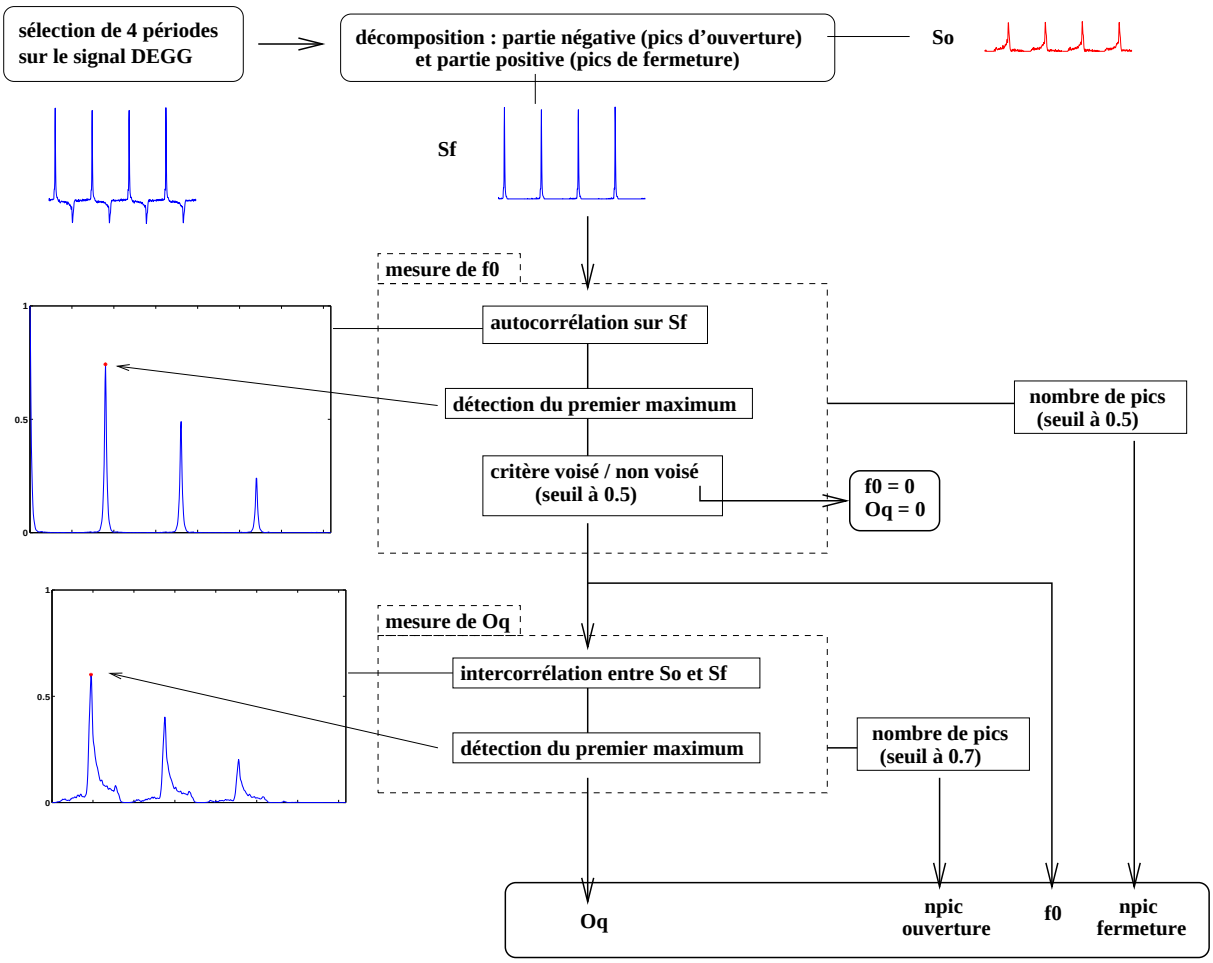


FIG. *7.12 – Schéma présentant les différentes étapes de l'analyse des signaux DEGG pour la mesure de* f_0 *et* O_q*. L'algorithme fournit également le nombre de pics détectés à la fermeture (npic fermeture) et à l'ouverture (npic ouverture).*

de ces deux signaux.

3. **détermination de la fréquence fondamentale**

   Elle se fait par calcul de la fonction d'autocorrélation normalisée et biaisée appliquée à S_f (les pics de fermeture étant de manière générale plus marqués que les pics d'ouverture). La détection du premier maximum de cette fonction donne la valeur de l'écart entre deux pics de fermeture successifs, donc de la période fondamentale T_0. Il est à noter qu'étant donnée la forme « quasi impulsive » du signal S_f, nous ne rencontrons pas ici de problème d'atténuation du fondamental par rapport aux harmoniques supérieurs, comme cela peut arriver sur un signal voisé.

   La détection du voisement se fait à cette étape de l'algorithme, par comparaison entre l'amplitude du pic détecté sur la fonction d'autocorrélation et une valeur seuil, choisie empiriquement à 0.5. Dans le cas d'un non-voisement, on pose $f_0 = 0$, $O_q = 0$ et l'algorithme s'arrête.

   Dans le cas contraire, on précise la valeur de f_0 mesurée, en effectuant une interpolation cubique sur 5 points centrés autour du maximum détecté. L'incertitude sur f_0 est dépendante de l'incertitude sur T_0 par : $\frac{\Delta f_0}{f_0} = \frac{\Delta T_0}{T_0}$. Pour obtenir une précision Δf_0 sur la mesure de f_0, il nous faut alors une précision sur la mesure de T_0 de $\Delta T_0 = \frac{\Delta f_0}{f_0^2}$. Le pas de l'interpolation effectuée est donc calculé à partir de la valeur approximative mesurée pour f_0, en posant $\Delta f_0 = 0.5$.

4. **détermination du quotient ouvert**

   Elle se fait par calcul de la fonction d'intercorrélation normalisée et biaisée entre S_o et S_f. Le premier maximum de cette fonction correspond à l'écart entre un pic d'ouverture et le pic de fermeture suivant. Si on note T_1 le temps correspondant, on en déduit le quotient ouvert par le rapport : $O_q = \frac{T_1}{T_0}$. L'incertitude sur O_q est donc fonction de celle sur T_1 et T_0 par : $\frac{\Delta O_q}{O_q} = \frac{\Delta T_1}{T_1} + \frac{\Delta T_0}{T_0}$. Comme précédemment, on détermine la position exacte du maximum par interpolation cubique entre 5 points centrés autour du maximum détecté, avec un pas identique à celui choisi pour la détermination de f_0, ce qui implique $\Delta T_1 = \Delta T_0$. Comme $T_1 = O_q T_0$, $\Delta O_q = \frac{\Delta T_0}{T_0}(1 + O_q) < 2\frac{\Delta T_0}{T_0} = 2\frac{\Delta f_0}{f_0}$. Pour la gamme de fréquences mesurées ($f_0 \geq 100$ Hz), l'erreur est inférieure à 0.01 .

5. **détermination du nombre de pics à la fermeture et à l'ouverture**

   Cette dernière partie de l'algorithme d'analyse permet de détecter automatiquement les cas où les pics sont dédoublés ou peu marqués. Elle est illustrée par la figure 7.13. On calcule la fonction d'intercorrélation entre le signal S_f (respectivement S_o) et un signal synthétique de même longueur, de même période et composé de gaussiennes très étroites (largeur de bande de 2% de la période T_0). Ce signal synthétique est très semblable à un signal S_f où les pics seraient très marqués. La détection se fait dans une fenêtre de taille T_0 centrée autour des premiers maxima détectés et seuls les maxima situés au-dessus d'un certain seuil sont pris en compte, afin de ne pas tenir compte des cas où il existe une différence importante en amplitude entre les pics à discriminer. Ce seuil, choisi empiriquement, est placé à 50% du maximum dans le cas de la fermeture et 70% dans le cas de l'ouverture.

En sortie de l'algorithme sont donc donnés : les valeurs de f_0 et O_q mesurées pour la trame, ainsi que le nombre de pics détectés à la fermeture (npic fermeture) et à l'ouverture (npic ouverture). Cette dernière information est ensuite prise en compte dans les mesures expérimentales, où l'on ne gardera que les valeurs de f_0 et O_q pour lesquelles les pics de fermeture et d'ouverture sont marqués et uniques.

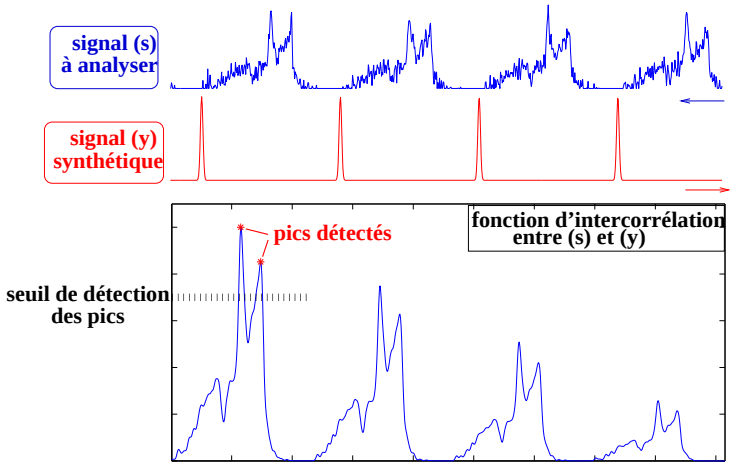


FIG. *7.13 – Illustration de la méthode de sélection des pics du signal DEGG, par intercorrélation avec un peigne de gaussienne.*

Illustration sur deux exemples

Nous avons émis précédemment deux hypothèses pour expliquer un dédoublement du pic d'ouverture : ouverture déphasée sur deux parties distinctes de la glotte (hypothèse A) ou vitesse d'ouverture sur l'épaisseur puis sur la longueur des cordes vocales (hypothèse B). Dans le cas de l'hypothèse A, l'initialisation du débit glottique serait reliée au premier des deux pics et nous pourrions alors choisir comme critère de sélectionner le premier des deux pics d'ouverture rencontrés comme marquant l'instant d'initialisation du flux glottique. Dans le cas de l'hypothèse B, c'est tout le contraire, car, lors de l'ouverture sur l'épaisseur, il n'y a pas de débit sortant de la glotte. C'est alors le second pic qui marquerait l'instant d'initialisation du flux glottique, et les valeurs de O_q ainsi mesurées seraient plus faibles car le temps relatif entre l'ouverture et la fermeture est réduit.

Nous nous sommes donc intéressés à tester ces hypothèses sur deux exemples de production vocale en mécanisme I. Nous avons analysé deux fois chacun de ces exemples. Pour la première analyse, nous sélectionnons, lors d'un dédoublement du pic d'ouverture, le premier des deux maxima détectés sur la fonction d'intercorrélation entre S_o et S_f. Pour la seconde analyse, nous sélectionnons le second pic, ce qui amène à une valeur plus élevée de O_q.

Le premier exemple présenté sur la figure 7.14 illustre le résultat de l'analyse d'un crescendo chanté par le baryton B2. Les mesures pour lesquelles l'algorithme a détecté un dédoublement du pic d'ouverture sont indiquées par « o » et celles pour lesquelles le pic d'ouverture est unique sont indiquées par « * ». Le début du crescendo (de $t = 1.4$ à 2.6 s) est ainsi marqué par un dédoublement marqué du pic d'ouverture. Ce dédoublement disparaît quand le chanteur monte en intensité. La courbe en pointillé présente les mesures où l'on a choisi de sélectionner le premier des deux pics détectés sur la fonction d'intercorrélation (valeurs plus faibles de O_q mesuré). Comme nous le voyons, dans ce cas, ce choix amène à des valeurs de quotient ouvert qui paraissent aberrantes ($O_q = 0.5$ pour un son piano) et la courbe de mesure présente des discontinuités marquées (à $t = 0.5$ et 2.4 s). Quand on choisit de sélectionner le second des deux pics détectés sur la fonction d'intercorrélation, on obtient une courbe de mesure continue et qui varie en fonction de l'intensité tel qu'on pourrait l'attendre en mécanisme I (nous reviendrons en détail sur le lien entre intensité vocale et quotient ouvert au chapitre 8). Cet exemple-ci va donc en faveur de l'hypothèse A, puisqu'on semble obtenir des mesures correctes quand on prend en compte le premier des pics d'ouverture (c'est-à-dire le second maximum de la fonction d'intercorrélation). On remarque également que, selon le choix effectué, l'écart sur les valeurs de O_q mesuré est considérable, allant jusqu'à $\Delta O_q = 0.2$.

Le contre-exemple est fourni par l'analyse suivante. La figure 7.15 présente le résultat de l'analyse

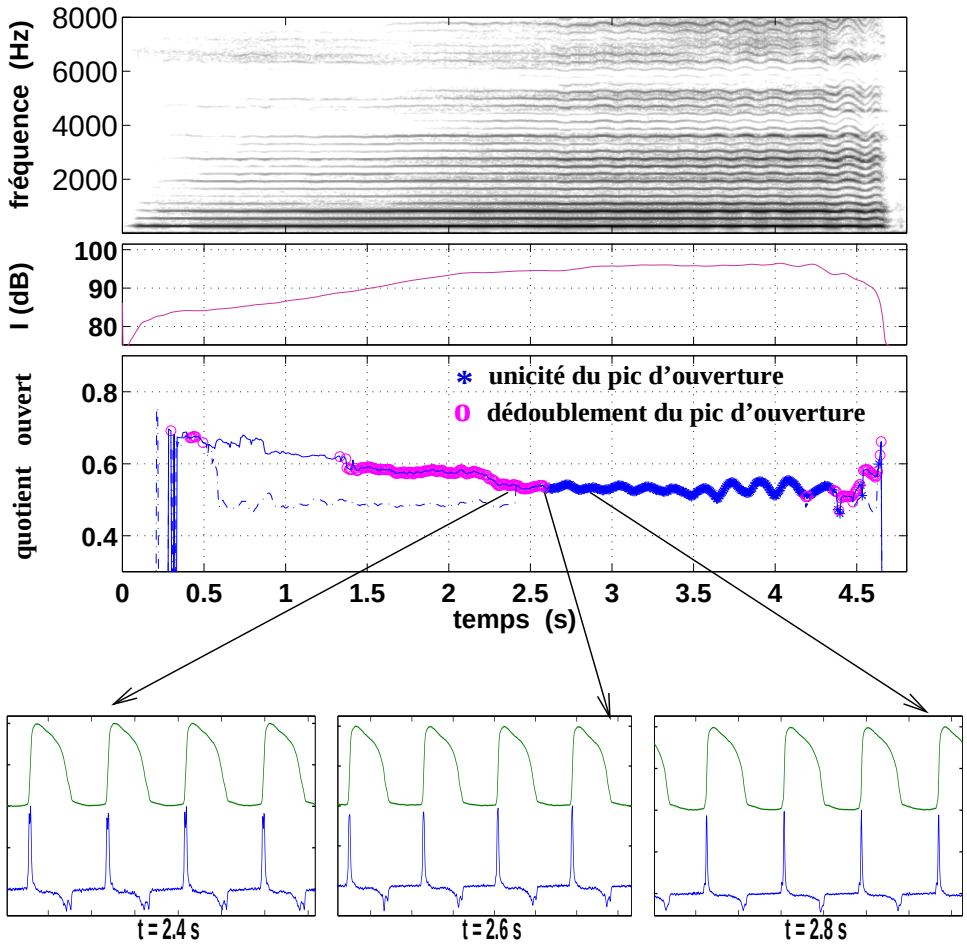


FIG. *7.14 – Illustration du choix du premier pic à l'ouverture lors d'un dédoublement, dans le cas d'un crescendo chanté par le baryton B2 en mécanisme I, sur la voyelle [u] et la hauteur do3 ($f_0 = 260$ Hz). Sur la figure présentant O_q mesuré en fonction du temps, les mesures où l'unicité du pic d'ouverture a été détectée sont indiquées par « * » et celles où un dédoublement du pic d'ouverture a été détecté sont indiqués par « o ». En trait plein, sélection du second maximum de la fonction d'intercorrélation, correspondant au premier pic d'ouverture ; en pointillés, sélection du premier maximum, correspondant au second pic d'ouverture. (CD, piste [13])*

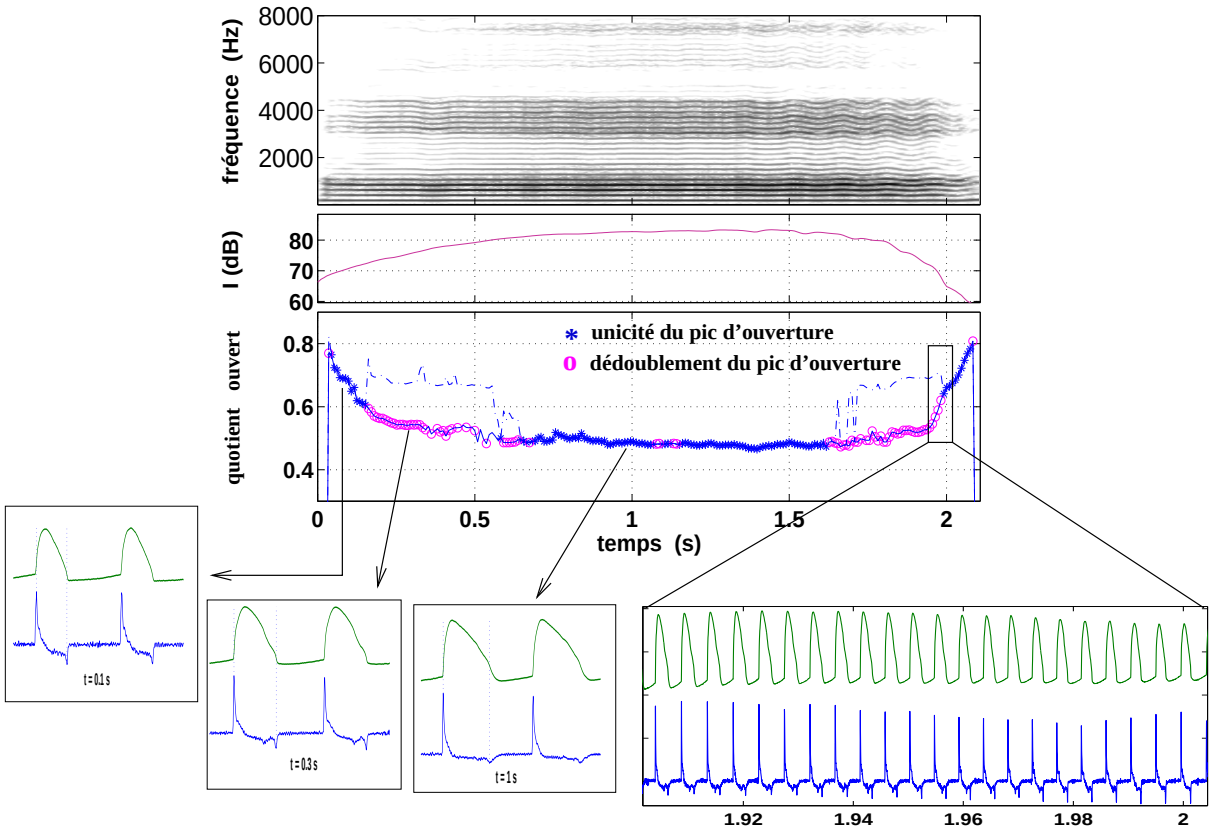


FIG. *7.15 – Illustration du choix du second pic à l'ouverture lors d'un dédoublement, dans le cas d'une voyelle [a] chantée par la mezzo-soprano MS2 en mécanisme I, sur un la2 ($f_0 = 220$ Hz). Sur la figure présentant O_q mesuré en fonction du temps, les mesures où l'unicité du pic d'ouverture a été détectée sont indiquées par « * » et celles où un dédoublement du pic d'ouverture a été détecté sont indiquées par « o ». En trait plein, sélection du premier maximum de la fonction d'intercorrélation, correspondant au second pic d'ouverture ; en pointillés, sélection du second maximum, correspondant au premier pic d'ouverture. (CD, piste [14])*

sujet :	B1	B2	B3	B4	B5	B6	B7	T1	T2	CT1	CT2	CT3	MS1	MS2	MS3	S1	S2	S3
$\mathbf{N_f = 1}$	51	85	66	75	76	68	16	81	65	81	82	35	86	59	61	68	89	90
$\mathbf{N_o = 1}$	34	49	21	56	41	50	19	64	31	49	53	64	63	44	46	52	82	80
$\left\{ \begin{array}{l} \mathbf{N_f = 1} \\ \mathbf{N_o = 1} \end{array} \right.$	31	48	19	56	40	43	8	61	25	46	46	30	62	44	45	35	81	78

Tab. *7.2 – Analyse des voyelles tenues et des sons filés de la base de données des chanteurs, en terme d'unicité des pics de fermeture ($N_f = 1$) et d'ouverture ($N_o = 1$), exprimée en pourcentage par rapport à la totalité des mesures voisées effectuées.*

d'une voyelle tenue chantée par la mezzo-soprano MS2 sur un la2 ($f_0 = 220$ Hz) (CD, piste [14]). Cette chanteuse présente un dédoublement du pic d'ouverture entre $t = 0.2$ et $t = 0.4$ s (valeurs marquées par des « o »). La courbe en pointillés correspond cette fois-ci à la sélection du second maximum de la fonction d'intercorrélation et la courbe en trait plein à celle du premier maximum. Comme précédemment, l'écart sur les valeurs de O_q ainsi mesurées est considérable, allant sur cet exemple de 0.1 à 0.2. Cet exemple-ci va plutôt en faveur de la seconde hypothèse.

En conlusion, nous voyons sur ces exemples qu'une mauvaise détection de l'instant d'ouverture, du fait d'un dédoublement des pics d'ouverture, amène à des valeurs erronées de quotient ouvert. Il est donc important de faire preuve de prudence quant aux mesures réalisées. On ne peut pas simplement choisir l'un ou l'autre des deux pics. Dans la suite, nous mesurerons O_q en choisissant le pic maximum de la fonction d'intercorrélation dans le cas d'une multiplicité du pic d'ouverture (ou de fermeture). Ces mesures n'auront qu'une valeur indicative, et nous n'utiliserons pour l'étude statistique de O_q en fonction de l'intensité vocale et de la fréquence fondamentale que les valeurs pour lesquelles les pics d'ouverture et de fermeture ont été détectés par l'algorithme comme étant uniques. Afin de donner une idée du nombre de données écartées par cette sélection, le tableau 7.3.2 fournit pour chaque chanteur les pourcentages de détection des cas suivants par rapport à la totalité des mesures voisées effectuées :

$N_f = 1$ unicité du pic de fermeture,
$N_o = 1$ unicité du pic d'ouverture,
$N_f = 1$ et $N_o = 1$ unicité des pics de fermeture et d'ouverture.

Ces données montrent que les situations pour lesquelles il y a multiplicité des pics d'ouverture et/ou de fermeture sont courantes : dans le cas du chanteur B7, en particulier, seules 8 % des mesures seront effectivement exploitables. Nous n'observons pas de tendances marquées au sein d'une même catégorie vocale. Dans le cas des sopranos, par exemple, seules 35 % des mesures effectuées sur S1 seront exploitables, alors que 81 % et 78 % des mesures le seront pour les sopranos S2 et S3.

Cela explique peut-être pourquoi ces signaux ne sont pas utilisés de façon courante dans la littérature. Pourtant, quand les pics d'ouverture et de fermeture sont bien marqués, la mesure de quotient ouvert est vraisemblablement plus correcte que si elle était effectuée sur les signaux EGG. Quelle différence y a-t-il alors entre ces mesures ? Nous nous proposons de répondre à cette question dans la partie suivante.

7.3.3 Comparaison entre les méthodes de mesure de quotient ouvert

En début de ce chapitre, nous avons illustré sur la figure 7.1 différentes méthodes de mesure de quotient ouvert. Nous avions également présenté au chapitre 4 sur une même figure (figure 4.16) les mesures de quotient ouvert effectuées sur les signaux DEGG et sur le signal acoustique par la méthode de Ljungqvist [146, 83] présentée en partie B de l'annexe B. Nous nous proposons d'effectuer une comparaison quantitative entre ces différentes méthodes de mesure du quotient ouvert. Cette comparaison est effectuée dans le cadre de la parole.

Nous avons analysé les signaux acoustiques et électroglottographiques de 5 locuteurs (M01 à M05), extraits de la base de données de Childers [29] (2000). Nous avons mesuré le quotient ouvert sur le signal

Mesures de quotient ouvert moyen

voyelle [a]

sujet	50 %	35 %	3/7	**DEGG**	**LJU**
M01	0.62	0.58	0.59	0.60	0.59
M02	0.53	0.47	0.50	0.47	0.46
M03	0.55	0.50	0.51	0.52	0.50
M04	0.54	0.46	0.53	0.55	0.57
M05	0.44	0.36	0.40	0.37	0.37

voyelle [i]

sujet	50 %	35 %	3/7	**DEGG**	**LJU**
M01	0.61	0.58	0.59	0.59	0.58
M02	0.53	0.47	0.50	0.51	0.52
M03	0.53	0.46	0.48	0.49	0.47
M04	0.50	0.45	0.49	0.48	0.49
M05	0.43	0.38	0.39	0.40	0.41

voyelle [u]

sujet	50 %	35 %	3/7	**DEGG**	**LJU**
M01	0.66	0.62	0.63	0.64	0.63
M02	0.51	0.47	0.49	0.48	0.49
M03	0.51	0.46	0.47	0.48	0.46
M04	0.50	0.43	0.49	0.50	0.52
M05	0.39	0.34	0.35	0.33	0.33

différences moyennes (en %) $|O_q - O_{q\,LJU}|$

voyelle [a]

sujet	50 %	35 %	3/7	**DEGG**
M01	2.99	1.25	0.79	0.99
M02	7.37	2.02	3.75	1.56
M03	4.71	1.08	1.39	2.26
M04	3.30	10.89	3.81	2.39
M05	6.03	1.45	2.44	1.25

voyelle [i]

sujet	50 %	35 %	3/7	**DEGG**
M01	3.26	1.10	1.12	0.93
M02	1.45	4.86	1.84	0.96
M03	6.47	2.08	3.06	2.98
M04	2.00	4.03	1.59	1.45
M05	4.84	4.70	4.13	2.47

voyelle [u]

sujet	50 %	35 %	3/7	**DEGG**
M01	3.52	1.01	0.83	1.29
M02	2.37	1.97	0.97	1.58
M03	5.22	1.70	2.39	2.82
M04	2.60	8.55	3.00	2.60
M05	5.39	1.21	1.89	1.10

TAB. *7.3 – Comparaison, pour* 5 *locuteurs énonçant les* 3 *voyelles [a], [i] et [u], des mesures de quotient ouvert : sur les signaux EGG (méthodes par seuil à* 35 *%,* 50 *% et* 3/7*), sur les signaux DEGG (par détection des pics d'ouverture et de fermeture) et sur les signaux acoustiques (méthode de Ljungqvist* [146, 83]

acoustique en appliquant la méthode de Ljungqvist [146, 83]. Nous l'avons également mesuré sur le signal électroglottographique par une méthode de seuil (à 35 ou 50 %), par la méthode proposée par Howard [113, 112] (détection de l'instant d'ouverture sur le signal EGG par un seuil à 3/7 et détection de l'instant de fermeture sur le signal DEGG) et par la méthode que nous venons de présenter dans ce chapitre (détection des instants d'ouverture et de fermeture sur le signal DEGG).

Le tableau 7.3 fournit les résultats des mesures de O_q pour chacune des 3 voyelles [a], [i] et [u]. Le tableau de gauche donne les valeurs moyennes de O_q mesurées pour chaque voyelle et chaque locuteur et le tableau de droite donne l'erreur absolue moyenne entre les valeurs de O_q mesuré par électroglottographie et sur le signal acoustique. Cette erreur est exprimée en %. Nous remarquons que dans le cas des méthodes de seuil ne prenant pas en compte le signal DEGG (seuil à 50 et 35 %), l'erreur obtenue par rapport à la mesure effectuée sur le signal acoustique est variable : elle peut être réduite, de l'ordre de 1 % ou très élevée, montant jusqu'à près de 10 %. La méthode par seuil à 35 % donne en général des valeurs de quotient ouvert plus proches que la méthode par seuil à 50 %. Dès que l'on prend en compte le signal électroglottographique, les valeurs se rapprochent des valeurs de quotient ouvert mesurées sur le signal acoustique. Les erreurs moyennes sont situées entre 0.79 % et 4.13 % dans le cas de la méthode proposée

par Howard et entre 0.96 % et 2.98 % dans le cas de la méthode présentée ici.

Ces résultats montrent que, dans l'ensemble, il y a une bonne correspondance entre les valeurs de quotient ouvert mesurées par électroglottographie et par filtrage inverse du signal acoustique. Les valeurs sont plus proches quand les méthodes prennent en compte le signal DEGG. La méthode proposée par Howard, qui donne ici de bons résultats, pourraient compléter la méthode que nous proposons, dans le cas où le pic d'ouverture est dédoublé ou peu marqué.

7.4 Conclusion

Nous avons discuté dans ce chapitre de l'intérêt des signaux DEGG. Nous avons également présenté une méthode de mesure de la fréquence fondamentale et du quotient ouvert à partir de ces signaux. Les mesures sur les signaux DEGG sont exploitables si les pics d'ouverture et de fermeture sont bien marqués. Dans ce cas, nous montrons que ces mesures sont plus précises qu'avec des méthodes basées sur le signal EGG, par rapport aux mesures effectuées sur le signal acoustique dans le cas de la parole. Les mesures de quotient ouvert sur les signaux DEGG peuvent même suppléer de façon très fiable aux mesures acoustiques lorsque celles-ci sont déficientes, comme dans le cas du chant.

Dans le cas où les pics d'ouverture et/ou de fermeture ne sont pas bien marqués, les mesures sur signaux DEGG sont sujettes à caution. Nous pourrions alors envisager d'utiliser dans ces cas la méthode proposée par Howard [113, 112], qui nous semble un bon compromis entre les méthodes de seuil appliquées sur les signaux EGG et les méthodes de détection des pics des signaux DEGG. En combinant la méthode proposée par Howard avec la méthode que nous développons ici, une détection systématique du quotient ouvert à partir des signaux DEGG, qui ne soit pas limitée aux cas où les pics de fermeture et d'ouverture sont bien marqués, est alors envisageable.

De manière très générale, la différence observée entre les méthodes est-elle significative du point de vue de la perception des variations de quotient ouvert ? Nous aborderons cette question dans la partie III, par la mesure de seuils de perception associés au quotient ouvert.

Indépendamment de ces mesures, il serait très intéressant d'effectuer une étude plus précise des signaux DEGG à l'aide de la visualisation par cinématographie ultra-rapide. Nous avons observé des comportements particuliers au moment des transitions entre mécanismes laryngés. Nous avons formulé des hypothèses pour expliquer les dédoublements observés sur les signaux DEGG, mais seule une comparaison avec la visualisation du mouvement glottique nous permettrait de comprendre exactement ces phénomènes. En particulier, la technique de transillumination strobophotoglottographique nous semble particulièrement adaptée pour cette caractérisation, puisqu'elle permet d'obtenir des informations sur l'ouverture par le dessous des cordes vocales alors que la glotte est fermée. Une collaboration avec le Department of Phoniatrics and Pedaudiology de l'University Hospital Eppendorf à Hamburg est en cours pour étendre cette technique à la cinématographie ultra-rapide et l'appliquer à la caractérisation des signaux DEGG et à l'étude de l'ouverture et de la fermeture sur l'épaisseur des cordes vocales (collabaration avec M. M. Hess, G. Schade, C. Mantay et J. Neubauer).

En Résumé

Relativement peu exploités dans la littérature, les signaux DEGG informent sur les instants d'ouverture et de fermeture glottique.

L'analyse des signaux DEGG de la base de données de chanteurs permet de montrer que les instants de fermeture et d'ouverture ne sont pas toujours facilement détectables sur ces signaux, dans les cas où l'on observe un dédoublement des pics de fermeture et/ou d'ouverture. Dans ces cas, les mesures peuvent aboutir à des valeurs aberrantes de quotient ouvert.

Deux hypothèses ont été avancées pour expliquer ces observations :

1. ouverture (fermeture) déphasée de deux parties distinctes de la glotte.
   Une comparaison entre la visualisation de la fermeture glottique par cinématographie ultra-rapide et les signaux DEGG appuie cette hypothèse, qui semble également confirmée par l'analyse du crescendo effectué par un chanteur.
2. différence de vitesse entre l'ouverture sur l'épaisseur et sur la longueur.
   L'analyse d'une voyelle tenue semble accréditer cette seconde hypothèse.

Un algorithme d'analyse des signaux DEGG a été mis en place. A partir de fonctions de corrélation, il mesure la fréquence fondamentale, le quotient ouvert, et détermine également la multiplicité des pics de fermeture et d'ouverture.

Les méthodes de mesure du quotient ouvert par électroglottographie et par estimation sur le signal acoustique ont été comparées dans le cadre de la parole. Les méthodes par seuil sur le signal EGG sont moins précises que celles utilisant tout ou partie du signal DEGG. Les mesures de quotient ouvert obtenues sur les signaux DEGG sont très proches de celles effectuées sur le signal acoustique : le quotient ouvert défini par rapport à l'ODG est donc égal à celui défini sur le signal DEGG.

Chapitre 8

Mesures du quotient ouvert en voix parlée et chantée

Nous avons développé dans le chapitre précédent une méthode de mesure de la fréquence fondamentale et du quotient ouvert sur les signaux DEGG. Cette méthode permet d'effectuer des mesures de façon découplée du système supraglottique. Nous avons montré sur des signaux de parole que les paramètres ainsi mesurés étaient bien ceux définis au chapitre 2 à partir du débit glottique. Nous pouvons donc à présent appliquer cette méthode à l'analyse du quotient ouvert dans le cas de productions vocales parlées et chantées. Les mesures de quotient ouvert vont-elles être dépendantes du mécanisme laryngé considéré ? Quelles sont les corrélations entre quotient ouvert, fréquence fondamentale et intensité vocale ? Observe-t-on des tendances bien marquées au sein d'une même catégorie vocale, entre catégories vocales ? Y a-t-il une variation de quotient ouvert selon la voyelle produite ? Des différences entre la parole et le chant ?

Voilà autant de questions auxquelles nous nous proposons de répondre. Après une revue bibliographique (partie 8.1), nous étudierons donc l'effet des mécanismes laryngés sur les valeurs de quotient ouvert (partie 8.2), puis en effectuant la distinction entre mécanismes, nous étudierons les corrélats entre quotient ouvert et intensité vocale en partie 8.4, ainsi qu'entre quotient ouvert et fréquence fondamentale en partie 8.5 [1]. Nous évoquerons également de façon qualitative un effet possible de la voyelle demandée au chanteur sur le quotient ouvert en partie 8.6 et nous terminerons en illustrant dans la partie 8.7 les productions d'une même phrase en voix parlée, chantée et criée, sur les exemples d'un baryton et d'une soprano.

8.1 Revue bibliographique

Dans la littérature, le paramètre de source mesuré est soit le quotient ouvert (mesures sur l'ODG obtenue par filtrage inverse ou par photoglottographie), soit le quotient fermé (mesures électroglottographiques). Dans cette étude, nous nous intéressons à l'ODG, l'électroglottographie n'étant qu'un moyen expérimental d'appréhender les variations de ce paramètre de source glottique. Nous avons donc fait le choix de parler en terme de quotient ouvert et, afin de garder une homogénéité dans cette revue bibliographique, nous présenterons tous les résultats en ces termes et convertirons au besoin quotient fermé en quotient ouvert ($O_q = 1 - C_q$).

[1] Les résultats préliminaires de cette étude ont été présentés à l'occasion de deux congrès en 2000 (Henrich & al [104, 101])

8.1.1 Mesures par glottographie (électro-, photo- et visualisation cinématographique)

Les premières mesures de quotient ouvert remontent à l'étude du larynx par cinématographie rapide de Timcke & al [197] en 1958. Elles montrent en particulier que le quotient ouvert est lié à l'intensité vocale : quand l'intensité augmente, le quotient ouvert diminue (la phase fermée est allongée). L'influence de la hauteur semble, quant à elle, minime. En 1974, dans une étude photoglottographique menée sur 20 sujets féminins, Kitzing & Sonesson [131] trouvent également une diminution du quotient ouvert lors d'une augmentation d'intensité vocale. Par contre, ils observent une augmentation du quotient ouvert quand la fréquence fondamentale augmente. Il est à noter que le mécanisme laryngé n'est pas pris en compte dans leur étude. En parole normale, ils mesurent une valeur moyenne de quotient ouvert de 0.65. En 1977, Lecluse & Brocaar [138] effectuent les premières mesures de quotient ouvert basées sur le signal électroglottographique, dans le cas de 6 sujets masculins non-entrainés. Ils étudient trois intensités (faible, confortable et élevée), plusieurs hauteurs et distinguent les deux mécanismes laryngés. Leur définition des quotients ouverts et fermés est légèrement différente de celle que nous avons présentée précédemment, car le temps d'ouverture précède le temps ouvert et n'est donc pas inclu dans ce dernier. Ceci se traduit par des valeurs plus faibles de O_q : elles sont comprises entre 0.2 et 0.45 dans le cas du mécanisme I, entre 0.45 et 0.7 dans le cas du mécanisme II. Si leurs résultats ne montrent pas de grande différence ni avec l'intensité, ni avec la fréquence fondamentale, ils montrent en revanche une différence très importante en fonction du mécanisme laryngé (valeurs moyennes de 0.33 pour le mécanisme I et 0.54 pour le mécanisme II, soit une différence moyenne de 0.2 entre les deux mécanismes).

En 1982, Kitzing & al [130] effectuent des mesures de quotient ouvert par photoglottographie combinée à l'électroglottographie, sur un sujet masculin, pour différentes qualités vocales (« breathy voice », $O_q = 0.79$ et « creaky voice », $O_q = 0.35$), différentes attaques vocales (« breathy », $O_q = 0.67$ et « hard », $O_q = 0.46$), en tenant compte des mécanismes laryngés (O_q passant de 0.65 à 0.83 au moment du passage mI $\rightarrow$ mII). Une étude centrée sur les mécanismes [127], portant sur un chanteur non-entraîné et un ténor, montre également de grandes différences de quotient ouvert selon le mécanisme laryngé. Elle met aussi en avant des différences entre registres résonantiels ($O_q = 0.6$ en « voix de poitrine » et 0.5 en « voix de tête », tous deux vraisemblablement produit en mécanisme I (Roubeau [174, 175], Chuberre [39])).

Mesurant le quotient ouvert O_q et le quotient de vitesse S_q par photoglottographie sur 12 sujets masculins, Hanson & al [96] ne mettent pas en évidence de liens significatifs entre l'intensité vocale et ces deux paramètres de source, seulement une tendance : diminution de O_q et augmentation de S_q quand l'intensité augmente. Ils observent une augmentation significative de O_q avec la fréquence fondamentale. L'étude menée par Orlikoff [161] montre, quant à elle, une corrélation très forte entre intensité vocale et quotient ouvert (mesure par seuil à 25 % sur 10 sujets masculins).

Howard s'intéresse à l'évolution du quotient fermé mesuré par électroglottographie en fonction de la fréquence fondamentale et des années d'entraînement vocal. Il étudie un groupe de chanteurs masculins entraînés ou non-entraînés [113], pour lequel il reporte une absence de lien significatif entre O_q et f_0 et une diminution progressive de O_q avec l'entraînement vocal. Dans le cas d'un groupe de sujets féminins (chanteuses entraînées ou non) [112], il observe que le quotient ouvert diminue de façon significative quand f_0 augmente, montrant même une relation linéaire entre $\log(f_0)$ et O_q. Cette relation dépend de l'entraînement vocal, étant très marquée chez les chanteuses entraînées et pas ou peu chez les chanteuses non-entraînées.

Dans une étude sur l'effet de l'ouverture de la bouche et de la pression mandibulaire sur le quotient fermé, Cookman & al [41] observent également une corrélation significative entre quotient ouvert et fréquence fondamentale. Par contre, ils montrent une augmentation du quotient ouvert avec la fréquence (résultat à l'inverse de celui obtenu par Howard), mais il est à noter que seules deux valeurs de f_0 ont été étudiées (fréquence fondamentale basse / haute). De plus, les résultats des hommes et des femmes ont été associés et moyennés et les deux valeurs obtenues (0.48 pour f_0 bas et 0.53 pour f_0 élevé) sont très proches. Les résultats qu'ils présentent ne sont donc pas forcément très représentatifs de la réalité. Cookman & al [41] montrent également que l'ouverture de la bouche en parole entraîne une diminution

du quotient ouvert, aussi bien chez les femmes que chez les hommes.

Childers & al [31] et Dromey & al [57] comparent le quotient ouvert mesuré sur les signaux DEGG et celui mesuré sur la fonction d'aire glottique obtenue par cinématographie ultra-rapide [31] ou sur le débit obtenu par filtrage inverse [57]. Pour 4 sujets masculins non-entraînés, Childers & al trouvent un bon accord général entre les deux types de mesure du quotient ouvert (des différences apparaissent lorsque la fréquence augmente), ainsi qu'une augmentation de O_q avec l'intensité vocale et avec la fréquence fondamentale. Pour 10 sujets féminins non-entraînés, Dromey & al trouvent des différences notables entre les deux types de mesure, et des tendances différentes également (diminution de O_q avec l'intensité dans le cas des mesures sur l'ODG et peu de variations dans le cas des mesures sur DEGG).

8.1.2 Mesures par filtrage inverse du signal de pression

Holmberg & al [109, 110, 111] effectuent des mesures des paramètres temporels de source glottique (entre autres, quotient ouvert et quotient de vitesse) sur 25 locuteurs et 20 locutrices, par estimation sur l'onde de débit obtenue par filtrage inverse du débit aux lèvres (masque de Rothenberg). Leur étude comparative de production syllabique pour des voix de faible, normale et forte intensité [109] montre des différences significatives de quotient ouvert en fonction de l'intensité. Dans le cas des sujets masculins, le quotient ouvert diminue quand l'intensité augmente. Dans le cas des voix de femmes, cette tendance se retrouve entre intensité faible et intensité normale. Par contre, ils n'observent pas de différences significatives entre intensité moyenne et forte. Les valeurs moyennes de O_q mesurées pour le groupe de sujets masculins (M) et le groupe de sujets féminins (F) sont : 0.74 (M) et 0.78 (F) pour une intensité faible, 0.60 (M) et 0.73 (F) pour une intensité normale, 0.56 (M) et 0.68 (F) pour une intensité forte. Les valeurs obtenues par filtrage inverse ont été comparées aux valeurs mesurées sur le signal EGG (par critère de niveau à 35%). Dans le cas de signaux de forte amplitude et non-bruités, ces deux mesures sont fortement corrélées [111]. Leur étude comparative de production syllabique pour des hauteurs graves, moyennes et aigües ne montre pas de corrélation forte entre O_q et f_0. La tendance observée est une augmentation de O_q avec f_0 dans le cas des hommes, comme dans le cas des femmes. Pour toutes ces études, le mécanisme laryngé utilisé est supposé être le mécanisme I, bien que ce point n'ait pas été expérimentalement vérifié.

Dans plusieurs études portant sur des productions vocales chantées, Sundberg & al [193, 188, 190, 192, 191] ont effectué des mesures de quotient fermé sur l'ODG obtenue par filtrage inverse du débit aux lèvres (méthode similaire à celle utilisée par Holmberg). Dans le cas de 6 chanteurs de « country music » [191] (phonation en mécanisme I et qualité de voix généralement « pressée »), ils mesurent des valeurs de quotient ouvert entre 0.4 et 0.7 et observent des variations selon l'intensité : les valeurs de O_q plus faibles pour des intensités élevées, descendant très souvent jusqu'à 0.4 et atteignant parfois 0.25. Ils n'observent pas d'effet significatif de la hauteur. Ils étudient également les effets des variations de pression sous-glottique sur 5 barytons [190], tous chanteurs d'opéra professionnels. Ils observent une corrélation entre O_q et la pression sous-glottique : le quotient ouvert diminue progressivement quand la pression sous-glottique augmente, puis plafonne entre 0.6 et 0.5. Dans une seconde étude portant sur 4 barytons, 5 ténors et 4 contre-ténors [192], ils observent l'effet des mécanismes laryngés, le quotient ouvert étant généralement plus élevé en mII qu'en mI à intensité vocale et hauteur équivalentes. Les différences sur O_q entre mécanismes sont plus marquées dans le cas des barytons que dans le cas des ténors et contre-ténors où, pour 3 d'entre eux, les valeurs sont même comparables. Des différences marquées sont également observées sur les sons pianos, où les barytons ne présentent que rarement une phase fermée ($O_q = 1$) alors que les ténors et contre-ténors en présentent fréquemment une ($O_q < 1$).

8.1.3 Résumé

Il ressort de ces études que :

le quotient ouvert dépend des mécanismes laryngés de façon directe et univoque : il est plus élevé en mécanisme II qu'en mécanisme I où la phase fermée est plus longue.

le quotient ouvert est influencé par l'intensité vocale, sans que le lien entre ces deux paramètres soit précisément établi. La tendance générale observée est une diminution du quotient ouvert quand l'intensité augmente.

le quotient ouvert est influencé par la pression sous-glottique : une augmentation de pression sous-glottique entraîne une diminution du quotient ouvert jusqu'à un certain seuil, après quoi le quotient ouvert reste stable.

le quotient ouvert est influencé par la fréquence fondamentale : il diminue ou il augmente avec f_0, selon les études.

Dans toute étude du quotient ouvert (ou fermé), il est donc primordial de distinguer les mécanismes laryngés, ce que nous illustrerons et quantifierons dans la partie 8.2. L'intensité vocale, la pression sous-glottique et la fréquence fondamentale semblent influencer le quotient ouvert. Dans cette étude, seules l'intensité vocale et la fréquence fondamentale seront mesurées et l'étude de l'influence de ces paramètres dans le cadre de la voix chantée sera l'objet des parties 8.4 et 8.5.

8.2 Effet des mécanismes laryngés

8.2.1 Illustration

Dans l'esthétique lyrique occidentale, les basses, les barytons et les ténors utilisent principalement le mécanisme I et ne travaillent donc pas ou peu le second mécanisme (les ténors légers peuvent être amenés à l'utiliser pour produire avantageusement certains aigus). La transition entre mécanisme I et mécanisme II se fait donc dans l'aigu de leur voix, alors que la tension musculaire des cordes et la pression sous-glottique sont très élevées (Miller [154]). De ce fait, elle peut s'accompagner d'un saut marqué en fréquence (Roubeau [174], Svec [194], Miller [155]) et souvent d'une baisse d'intensité vocale. La figure 8.1, qui présente un glissando chanté par un ténor, illustre ces phénomènes.

Le sonagramme du signal acoustique permet de visualiser les variations temporelles de la structure harmonique et formantique du glissando. Les évolutions temporelles du quotient ouvert et de la fréquence fondamentale sont présentées sur la même figure. Les mesures de f_0 et O_q pour lesquelles il y a unicité du pic de fermeture et du pic d'ouverture (détection automatique par analyse des signaux DEGG) sont présentées sous forme de points épais. On observe souvent une grande continuité entre ces valeurs et celles obtenues avec détection simultanée de pics peu marqués ou doubles, ce qui laisse à penser que l'erreur de mesure n'est pas forcément très importante dans ces cas. Cela valide également le choix de la méthode de mesure par intercorrélation plutôt que par détection directe, qui, dans le cas de signaux avec pics peu marqués, permet une mesure fiable et proche de celle que l'on obtiendrait avec des pics marqués. De ce fait, et afin d'aider à la lisibilité de la figure 8.1, l'ensemble des valeurs mesurées est superposé à la figure sous forme d'un trait fin discontinu. Les deux barres verticales distinguent les parties chantées en mécanisme I de celles chantées en mécanisme II et sont placées subjectivement au moment du passage entre mécanismes. Dans ce cas-ci, ces transitions sont facilement identifiables par le saut en fréquence. Nous verrons sur l'exemple suivant (figure 8.2) que ce n'est pas toujours aussi évident et qu'une analyse du sonagramme peut s'avérer utile pour tenter de localiser les transitions (sans s'aider de la courbe de O_q, qui, comme nous allons le voir, est également un bon indicateur).

Le passage (mI $\rightarrow$ mII) est marqué par un saut en fréquence rapide (quelques périodes), passant de $f_0 = 370$ Hz (fa3 #) à $f_0 = 444$ Hz (la3), ce qui correspond à une tierce mineure (3 demi-tons). Le passage (mII $\rightarrow$ mI) est également marqué par un saut en fréquence rapide, passant de $f_0 = 357$ Hz (entre fa3 et fa3 #) à $f_0 = 263$ Hz (do3), ce qui correspond à une quarte juste à augmentée (5 à 6 demi-tons). Ces résultats sont tout à fait en accord avec ceux trouvés pour des ténors par Miller [155], lors d'une étude sur l'intervalle caractéristique du saut en fréquence au moment du passage entre « voix de poitrine »

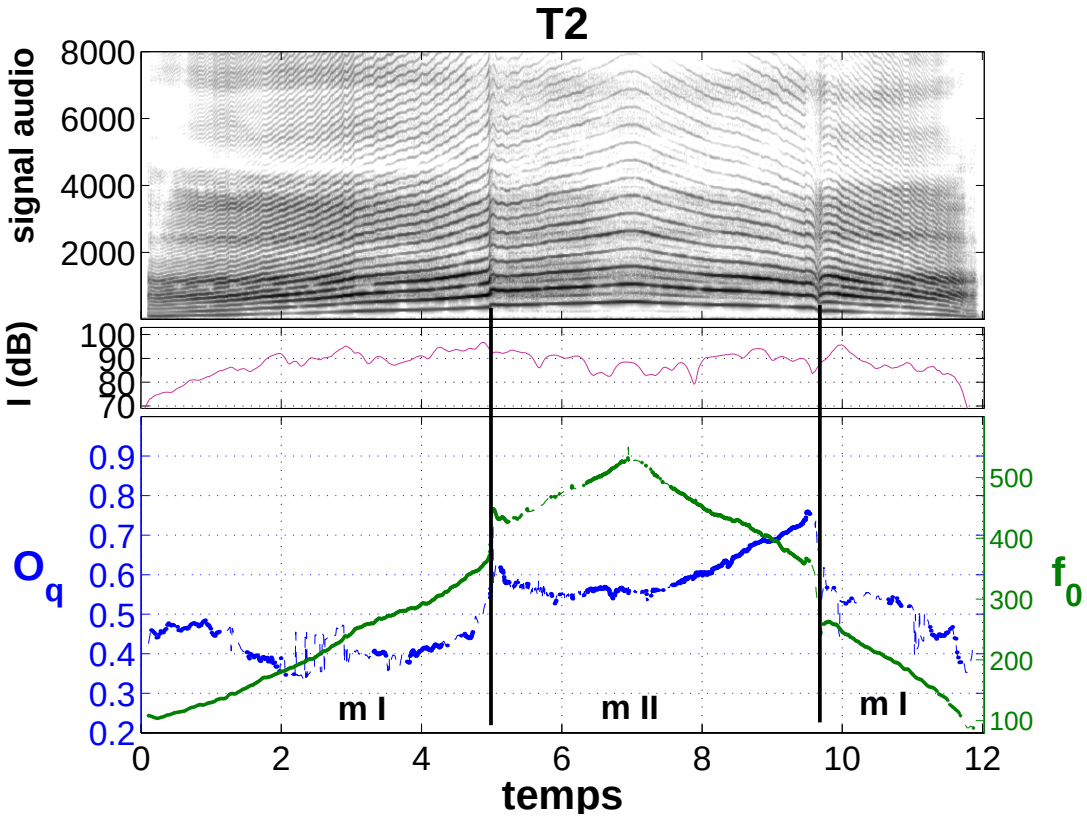


Fig. *8.1 – Glissando chanté par le ténor T2. Le passage mécanisme I → mécanisme II se traduit par un saut en fréquence de* 74 Hz *(fa3 → la3), précédé par une augmentation progressive du quotient ouvert (de* 0.4 *à* 0.62*). Le passage mécanisme II → mécanisme I se traduit par un saut en fréquence de* 94 Hz *(fa3 → do3)et une diminution brutale du quotient ouvert (de* 0.75 *à* 0.53*). (CD, piste [40])*

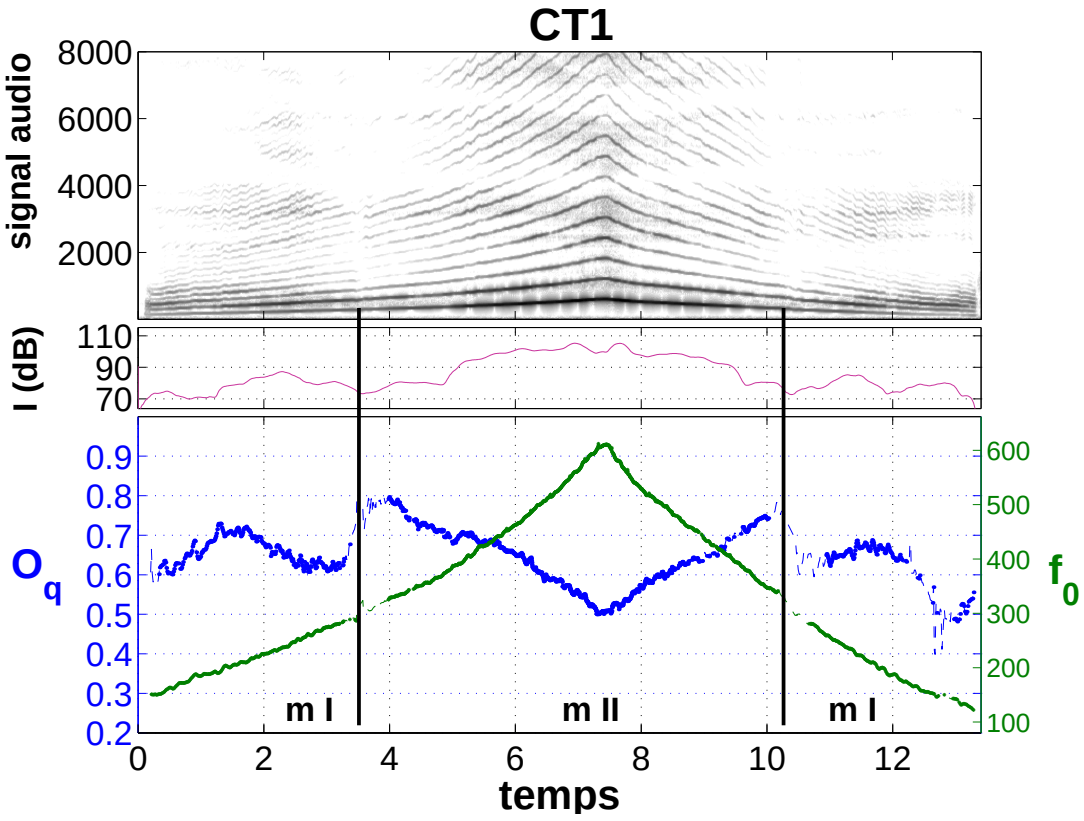


FIG. *8.2 – Glissando effectué par le contre-ténor CT1. Le passage entre les mécanismes I et II est bien géré et quasiment inaudible. Malgré tout, on retrouve comme dans le cas du ténor un saut de quotient ouvert qui précède la transition. Une atténuation des hautes fréquences est visible sur le spectre au moment du passage. On notera également que la transition se fait à faible intensité. (CD, piste [45])*

(mI) et « falsetto » (mII) réalisée sur des chanteurs entraînés (une basse, un baryton basse, 3 ténors, 4 mezzo-sopranos et 3 sopranos). Ces sauts en fréquence s'accompagnent de sauts également marqués du quotient ouvert : O_q passe de 0.4 à 0.62 lors de la montée en fréquence et de 0.75 à 0.53 lors de la descente. Dans le cas du passage (mI → mII), cette variation est progressive ($\sim 300\,ms$), plus lente que celle de la fréquence et elle précède le passage. Un résultat équivalent avait été mentionné par Miller [155], bien que le temps de variation de O_q dans son étude était plutôt de l'ordre de $100\,ms$. Dans le cas du passage (mII → mI), la modification de quotient ouvert est aussi ample, mais plus rapide (comparable au saut en fréquence).

En chant classique, les chanteurs qui développent l'étendue de leur voix sur les 2 mécanismes (comme, par exemple, les altos hommes ou femmes) doivent apprendre à « lisser » le passage d'un mécanisme à un autre, afin d'éviter toute rupture de l'homogénéité du timbre dans la ligne mélodique. La figure 8.2 présente le glissando réalisé par un contre-ténor (CD, piste [45]).

Dans ce cas-ci, le saut en fréquence à la transition est imperceptible. On peut néanmoins observer sur le sonagramme et sur la courbe d'intensité une diminution significative de l'amplitude au moment des transitions, phénomène déjà remarqué par Roubeau [174]. Ces passages ne sont quasiment pas perceptibles à première écoute. On remarque ici aussi une très nette modification du quotient ouvert, qui précède également le passage. Cette modification est moins importante que pour le ténor, puisque l'on passe de 0.62 à 0.78 au moment du passage (mI → mII) et 0.76 à 0.6 au moment du passage (mII → mI).

On note une corrélation visuelle très marquée entre le quotient ouvert et la fréquence fondamentale

en mécanisme II. Cette corrélation se retrouve sur tous les glissandos effectués par ce chanteur (exemples présentés en Annexe F, sur les figures F.15, F.16 et F.18). Cette tendance se retrouve également sur les glissandos chantés par les autres contre-ténors, les 2 ténors (voir exemple précédent), et dans une moindre mesure les 3 sopranos. Le lien possible entre quotient ouvert et fréquence fondamentale sera étudié plus en détail en 8.5.

Pour les chanteurs masculins utilisant principalement le mécanisme I, la production vocale en mécanisme II se caractérise souvent par une diminution sensible de l'intensité vocale par rapport au mécanisme I, alors pourtant qu'ils montent en fréquence. Ce contre-ténor fait exactement l'inverse : il est en piano dans le mécanisme I et il laisse par contre augmenter son intensité vocale avec la fréquence fondamentale en mécanisme II. Il baisse à nouveau son intensité vocale au moment d'aborder la transition (mII $\rightarrow$ mI). Visiblement, ce chanteur gère son passage en l'abordant à faible intensité vocale. Si l'on compare sa production avec celle du ténor précédent, on observe que :

en mécanisme I : le contre-ténor a déjà des valeurs élevées de O_q (entre 0.6 et 0.7), alors que celles du ténor sont plutôt situées entre 0.4 et 0.5. De ce fait, le saut de quotient ouvert est plus ample dans le cas du ténor que dans celui du contre-ténor.

au moment de la transition entre mécanismes : le contre-ténor chante à une intensité vocale bien plus faible ($I \simeq 72$ dB) que celle du ténor ($I \simeq 95$ dB).

Nous verrons en partie 8.4 que l'intensité vocale est souvent fortement corrélée avec le quotient ouvert en mécanisme I : à faible intensité, le quotient ouvert a des valeurs plus élevées qu'à forte intensité. Pour éviter, ou du moins limiter en amplitude le saut de quotient ouvert, une solution est donc de baisser l'intensité vocale en mécanisme I pour que le quotient ouvert ait des valeurs proches de celles qu'il aurait une fois passé en mécanisme II. Cette solution est celle vraisemblablement choisie par ce contre-ténor. Il limite la variation de O_q au moment du passage, ce qui se répercute sur la qualité vocale du son produit et permet une homogénéité perceptive du timbre.

Cette technique de « lissage » vocal du changement de mécanisme se retrouve bien évidemment dans d'autres catégories vocales, en particulier chez les ténors, les altos et les mezzo-sopranos. Le ténor précédent (T2) peut, par exemple, chanter à la demande des glissandos sans rupture de fréquence. Ces exemples sont présentés en Annexe F sur les figures F.13 et F.14.

Le contre-ténor CT1 a également chanté des voyelles tenues en mécanisme I (CD, piste [15]) puis en II (CD, piste [16]) sur une note de la zone commune aux 2 mécanismes (ré3, 293 Hz). La figure 8.3 illustre le résultat des mesures, dans le cas de la voyelle [a], tandis que le tableau 8.1 donne les moyennes et écart-types pour les trois voyelles [a], [e] et [u]. Cet exemple montre également la différence de quotient ouvert qui caractérise les mécanismes I et II. Dans le cas de ce contre-ténor, le mécanisme I se caractérise en moyenne par $O_q = 0.64$ et le mécanisme II par $O_q = 0.77$. L'intensité vocale moyenne pour cet exemple est de 88 dB en mI et 79 dB en mII et les sonagrammes présentés sur la figure 8.3 montrent une baisse de l'énergie dans les hautes fréquences ($f \geq 4000$ Hz) dans le cas du mécanisme II. Perceptivement, les deux productions vocales sont très semblables.

mécanisme laryngé	[a]		[e]		[u]	
	O_q	I	O_q	I	O_q	I
mécanisme I	0.64 (0.02)	88 (3)	0.64 (0.01)	83 (3)	0.65 (0.02)	82 (2)
mécanisme II	0.77 (0.02)	79 (3)	0.77 (0.04)	77 (3)	0.77 (0.03)	79 (3)

Tab. *8.1 – Mesure du quotient ouvert et de l'intensité vocale [moyenne (écart-type)] sur des voyelles tenues chantées par le contre-ténor CT1 en mécanisme I ou II, à même hauteur.*

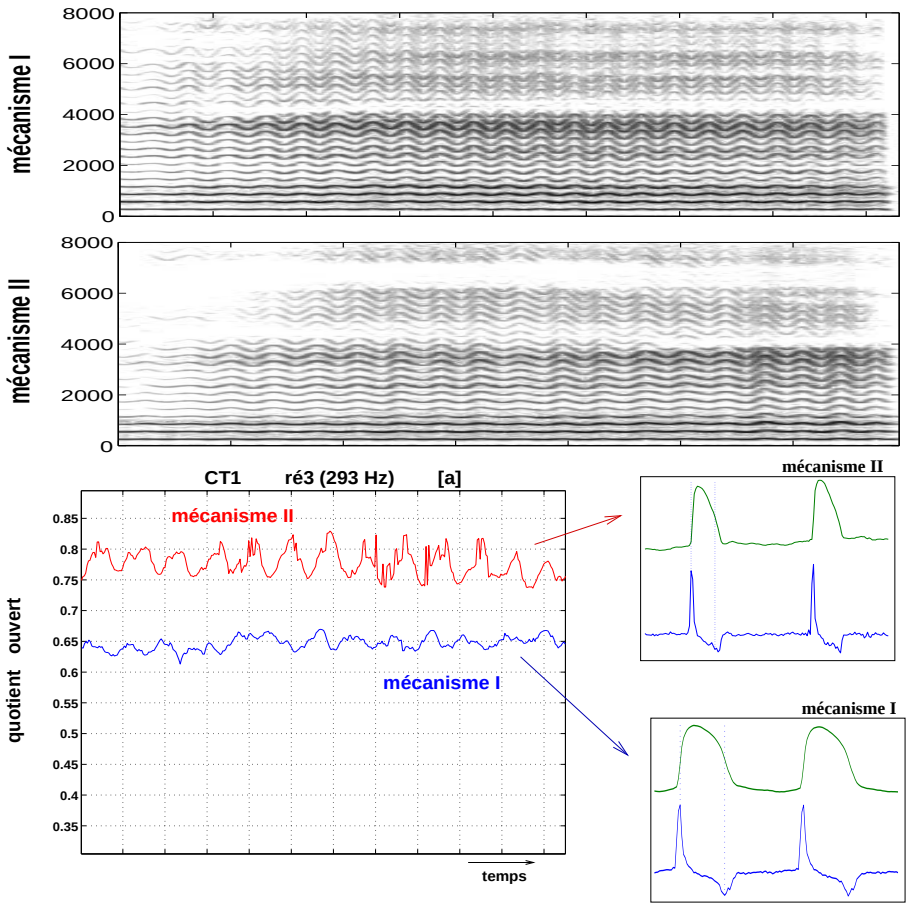


FIG. *8.3 – Voyelle [a] chantée par le contre-ténor CT1 sur une même note (ré3) en mécanisme I (CD, piste [15]) puis en II (CD, piste [16]). Les signaux EGG et DEGG correspondant à ces deux productions sont également représentés.*

8.2.2 Analyse en contexte (parlé, chanté)

Il a été demandé aux chanteurs de choisir une phrase en français de leur choix, de la parler, de la chanter et de la crier. Dans certains cas, les chanteurs ont pu effectuer cette tâche en utilisant l'un ou l'autre des mécanismes I et II. Nous reviendrons sur la différence voix parlée / voix chantée dans la partie 8.7. Pour l'instant, nous nous intéressons aux différences entre mécanismes dans ces productions et nous présenterons des valeurs moyennées sur une phrase vocale. La figure 8.4 illustre les résultats obtenus pour des chanteurs ayant réussi à effectuer l'une ou l'autre de ces tâches dans les deux mécanismes. On note qu'un même sujet n'a pas forcément effectué les 3 tâches dans les deux mécanismes. Ainsi, le contre-ténor CT1 a chanté et crié la phrase dans deux mécanismes différents, tandis que le contre-ténor CT2 a chanté et énoncé la phrase dans deux mécanismes différents.

On note la variabilité inter-individuelle. Indépendamment de cette variabilité, on remarque la très nette différence de quotient ouvert entre mécanisme I et II (entre 0.4 et 0.65 pour le mécanisme I, 0.65 et 0.8 pour le mécanisme II), quelle que soit le type de production (parlée, chantée ou criée). Dans le cas du contre-ténor CT2, il ne semble pas y avoir de différence au niveau du quotient ouvert entre voix parlée et voix chantée dans un mécanisme donné. On observe une différence entre voix chantée et criée pour le mécanisme I (exemple du contre-ténor CT1). Cette différence s'accompagne d'une augmentation de l'intensité acoustique moyenne de 10 dB entre la voix chantée et la voix criée, ce qui traduit un effort vocal accentué.

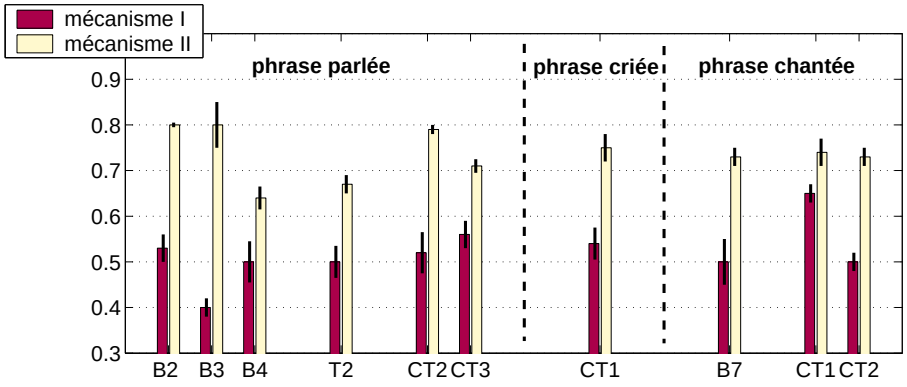


Fig. *8.4 – Moyenne et écart-type de quotient ouvert mesurés sur des phrases parlées, chantées et criées en mécanisme I et II, dans le cas de chanteurs ayant effectués une ou plusieurs de ces tâches en utilisant successivement les deux mécanismes.*

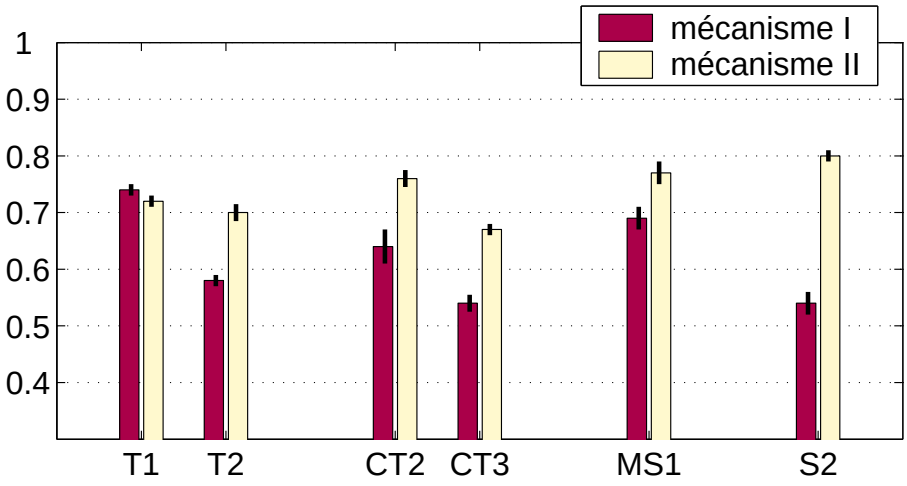


Fig. *8.5 – Moyenne et écart-type de quotient ouvert mesurés sur les premières mesures de l'Ave Maria de Gounod, chantées en mécanisme I et II.*

Dans le cas de la phrase chantée commune (les premières mesures de l'*Ave Maria* de Gounod), certains chanteurs ont également pu effectuer la tâche en utilisant les deux mécanismes. La figure 8.5 présente les résultats de deux barytons, deux contre-ténors, une mezzo et une soprano. L'intensité vocale est comparable entre les deux productions, par contre la fréquence fondamentale moyenne est très différente, afin de permettre aux chanteurs de chanter l'intégralité de la phrase musicale dans un mécanisme donné. Seul le contre-ténor CT2 a chanté la phrase à même hauteur en utilisant l'un ou l'autre des deux mécanismes. On remarque l'écart très important entre mécanisme I et mécanisme II dans le cas de la soprano S2 (0.54 et 0.8). Comme les sopranos n'utilisent que très peu le mécanisme I, on peut penser qu'un effort vocal important leur est peut-être nécessaire pour effectuer cette production vocale dans un mécanisme inhabituel.

On remarque également la production du ténor T1 qui se distingue des autres par le fait que le quotient ouvert est le même en mécanisme I et II. On peut se demander si ce chanteur, un ténor léger à la voix jugée « très mixée » par ses professeurs, n'utilise pas principalement le mécanisme II dans ses productions vocales. Si l'on regarde le glissando effectué par ce chanteur (figure F.11, (CD, piste [39])), on remarque également que le quotient ouvert reste toujours à des valeurs élevées (≥ 0.7), que le chanteur soit en mécanisme I ou en mécanisme II. La détermination des mécanismes se base alors sur les transitions, qui

sont ici perceptibles grâce au saut en fréquence qui les caractérise, mais qui ne sont pas très marquées sur la courbe de quotient ouvert (petite modification de O_q au passage mI $\rightarrow$ mII). On peut donc en conclure que ce chanteur utilise bien les deux mécanismes dans ses productions vocales. Il chante avec une production vocale reconnue par ses professeurs comme très mixée, à la façon des « haute-contres », qui sont des contre-ténors « à la française » (Jacobs [119]), chantant avec un timbre propre aux contre-ténors, mais avec utilisation du mécanisme I. De ce fait, on peut penser que la production vocale des « haute-contres » se caractériserait par un quotient ouvert élevé, même en mécanisme I (ce qui ferait exception à la tendance générale observée chez les sujets masculins). Cette hypothèse mériterait d'être étudiée sur d'autres sujets utilisant ce type de production vocale, avant que des conclusions puissent être avancées sur ce point.

8.2.3 Conclusion

Cette étude confirme donc l'influence très forte du mécanisme laryngé sur la valeur du quotient ouvert, indépendamment de la production vocale considérée. Les valeurs de quotient ouvert sont très souvent plus faibles en mécanisme I qu'en mécanisme II. Ce résultat sans surprise reflète simplement les différences de configuration glottique entre les deux mécanismes, que nous avons détaillées en partie 1.2 du chapitre 1 et il confirme donc la réalité physiologique de ces mécanismes. Dans toute étude portant sur le quotient ouvert, il est donc fondamental de préciser le mécanisme laryngé dans lequel a lieu la production vocale étudiée.

Il est par contre surprenant de remarquer à quel point certains chanteurs peuvent reproduire les qualités acoustiques d'un mécanisme en restant vraisemblablement dans l'autre mécanisme. Cette caractéristique vocale particulière est appelée « voix mixte ». La voix mixte ne constitue pas un mécanisme en soi, mais pourrait être considérée comme un registre résonantiel (Chuberre [39], 2000). Le ténor T1 présente un exemple typique de voix mixée, à tel point que l'on peut se demander s'il n'est pas toujours en mécanisme II, à la vue de ses valeurs de quotient ouvert élevées. Nous pensons que ce n'est pas le cas, mais qu'en mécanisme I, ce chanteur joue inconsciemment sur son quotient ouvert pour produire des sons dont les qualités acoustiques sont identiques à celles des sons produits en mécanisme II.

Nous aurions pu conclure que le quotient ouvert est un indicateur fiable et robuste du mécanisme laryngé, à la vue des résultats généraux. Nous voyons par ces contre-exemples que ce n'est pas toujours le cas. De même, l'oreille humaine n'est pas toujours à même de déterminer, à l'écoute d'une production vocale donnée, la nature du mécanisme laryngé sous-jacent, puisqu'elle peut avoir l'illusion d'un mécanisme donné. La combinaison de l'écoute et de la visualisation des signaux acoustique, EGG et DEGG, des courbes de quotient ouvert et d'intensité vocale, peut parfois lever ces doutes.

8.3 Fréquence fondamentale et intensité vocale

Avant de présenter les mesures de quotient ouvert en fonction de l'intensité vocale et de la fréquence fondamentale, nous allons brièvement discuter du lien entre ces deux dernières variables.

En effet, nous avons noté au chapitre 2 de la partie préliminaire que la fréquence fondamentale et l'intensité vocale sont souvent fortement corrélées. Comparativement à la parole, une des caractéristiques du chant est pourtant de pouvoir chanter fort tout en restant à la même hauteur et de monter en fréquence en limitant la montée en intensité. Un chanteur apprend à contrôler au mieux la dépendance naturelle entre fréquence fondamentale et intensité vocale pour une production vocale donnée. Qu'en est-il globalement sur l'ensemble d'une tessiture ? Le tableau 8.2 présente le coefficient de corrélation totale entre fréquence fondamentale et intensité vocale mesuré sur les voyelles tenues et les sons filés, dans le cas des deux mécanismes. Nous observons qu'en analysant de façon globale l'ensemble de la tessiture d'un chanteur, la fréquence fondamentale et l'intensité vocale semblent souvent fortement corrélées. Le baryton B6 fait exception, ainsi que le baryton B7 sur le [e] et le [u] et le contre-ténor CT2 en mécanisme

mécanisme I

chanteur	[a]	[e]	[u]
B1	**0.73**	0.47	**0.75**
B2	0.68	0.69	**0.84**
B3	0.52	0.55	0.47
B4	**0.81**	**0.75**	0.69
B5	0.50	0.62	**0.73**
B6	0.11	0.18	n.s.
B7	**0.72**	0.22	0.23
T1	0.69	**0.88**	**0.94**
T2	0.45	0.60	0.59
CT1	0.64	0.49	0.66
CT2	0.28	0.18	0.47
MS1	0.53	**0.71**	**0.84**
MS2	**0.80**	**0.73**	0.64
MS3	0.55	**0.79**	0.70

mécanisme II

chanteur	[a]	[e]	[u]
CT1	**0.90**	**0.83**	0.63
CT2	0.60	0.68	0.64
CT3	**0.75**	**0.86**	**0.84**
MS1	**0.75**	**0.79**	**0.75**
MS2	0.70	**0.93**	**0.92**
MS3	**0.86**	**0.80**	**0.86**
S1	0.65	**0.83**	**0.74**
S2	**0.78**	**0.85**	**0.87**
S3	**0.83**	**0.78**	**0.78**

Tab. *8.2 – Coefficient de corrélation totale entre* **fréquence fondamentale et intensité vocale**, *dans le cas des productions vocales en mécanisme I et II, pour les 3 voyelles [a], [e] et [u]. Les coefficients en gras présentent une corrélation forte ($r > 0.70$). Une corrélation non significative ($p \geq 0.001$) est indiquée par n.s.*

I.

En comparant le quotient ouvert avec l'intensité vocale, il nous faudra donc garder à l'esprit qu'une variation sous-jacente de la fréquence fondamentale peut également entraîner une variation de l'intensité vocale. Pour pouvoir en tenir compte lors des calculs de corrélation, nous introduirons le coefficient de corrélation partielle, détaillé dans l'annexe D, qui permet de vérifier si la dépendance entre deux variables n'est pas le fait des variations sous-jacentes d'une troisième variable.

8.4 Quotient ouvert et intensité vocale

8.4.1 Illustration sur un son filé

La figure 8.6 présente un son filé chanté par le baryton B1 sur la voyelle [a] et à une hauteur de 260 Hz (do3). Nous nous plaçons ici dans un cas où la fréquence fondamentale est gardée constante, ce qui supprime la possible dépendance entre fréquence fondamentale et intensité vocale. Nous observons que la variation du quotient ouvert est fonction de l'intensité : quand l'intensité augmente, le quotient ouvert diminue, et vice versa. Le quotient ouvert passe ainsi de 0.7 à 0.5 alors que l'intensité vocale augmente de 20 dB. L'augmentation d'intensité vocale, accompagnée par une diminution du quotient ouvert, se traduit sur le sonagramme par une augmentation de la richesse spectrale.

Cette corrélation entre quotient ouvert et intensité vocale est-elle généralisable ? Est-elle dépendante du mécanisme laryngé utilisé ? De la catégorie vocale du chanteur ? Nous allons étudier ce lien possible entre intensité vocale et quotient ouvert, en ayant soin de distinguer les mécanismes laryngés, puisque nous venons de voir que, selon le mécanisme considéré, les valeurs de quotient ouvert pouvaient très différentes,

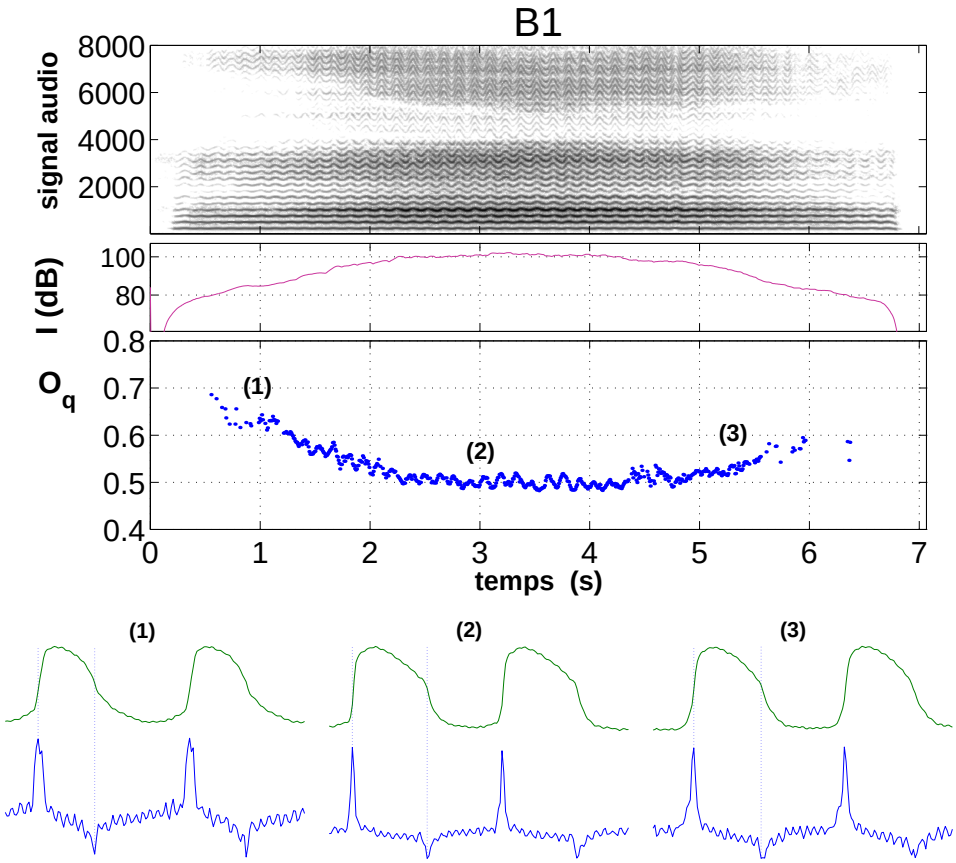


FIG. *8.6 – Son filé chanté par le baryton B1 sur un do3 (*260 Hz*) et la voyelle [a].*

indépendamment de l'intensité vocale. Nous prendrons également en compte les possibles variations sous-jacentes de la fréquence fondamentale qui pourraient impliquer des variations de l'intensité vocale.

8.4.2 Production vocale en mécanisme I

Les figures E.1, E.2, E.3 et E.4 présentées dans l'annexe E montrent les variations d'intensité vocale en fonction du quotient ouvert en mécanisme I pour les 7 barytons, les 2 ténors, 2 des contre-ténors et les 3 mezzo-sopranos. Seules ont été prises en compte les valeurs mesurées pour les voyelles tenues et les sons filés. Les échelles sont maintenues constantes, afin de pouvoir visualiser les différences entre chanteurs. En complément de ces graphes, le tableau 8.3 donne les valeurs des coefficients de corrélation totale et partielle entre quotient ouvert et intensité vocale, dans le cas des productions vocales en mécanisme I (voir Annexe D pour des précisions sur le calcul de ces coefficients).

Les résultats sont présentés toutes hauteurs confondues et pour chacune des 3 voyelles [a], [e] et [u]. Il est important de noter que ces voyelles sont celles demandées au chanteur et que nous n'avons pas étudié la correspondance avec les fréquences des formants effectivement mesurées sur les signaux enregistrés. Si en parole il est relativement simple d'associer à une voyelle donnée des positions fréquentielles précises pour les formants, dans le chant, il en va tout autrement. Ainsi, en étudiant la position fréquentielle des formants pour plusieurs voyelles dans le cas d'une chanteuse soprano, Sundberg [187] a montré que cette position évolue avec la fréquence fondamentale. Comme l'objet de cette partie n'est pas l'étude des variations de la position fréquentielle des formants, nous avons fait le choix de sélectionner les voyelles en fonction de ce que le chanteur a voulu émettre.

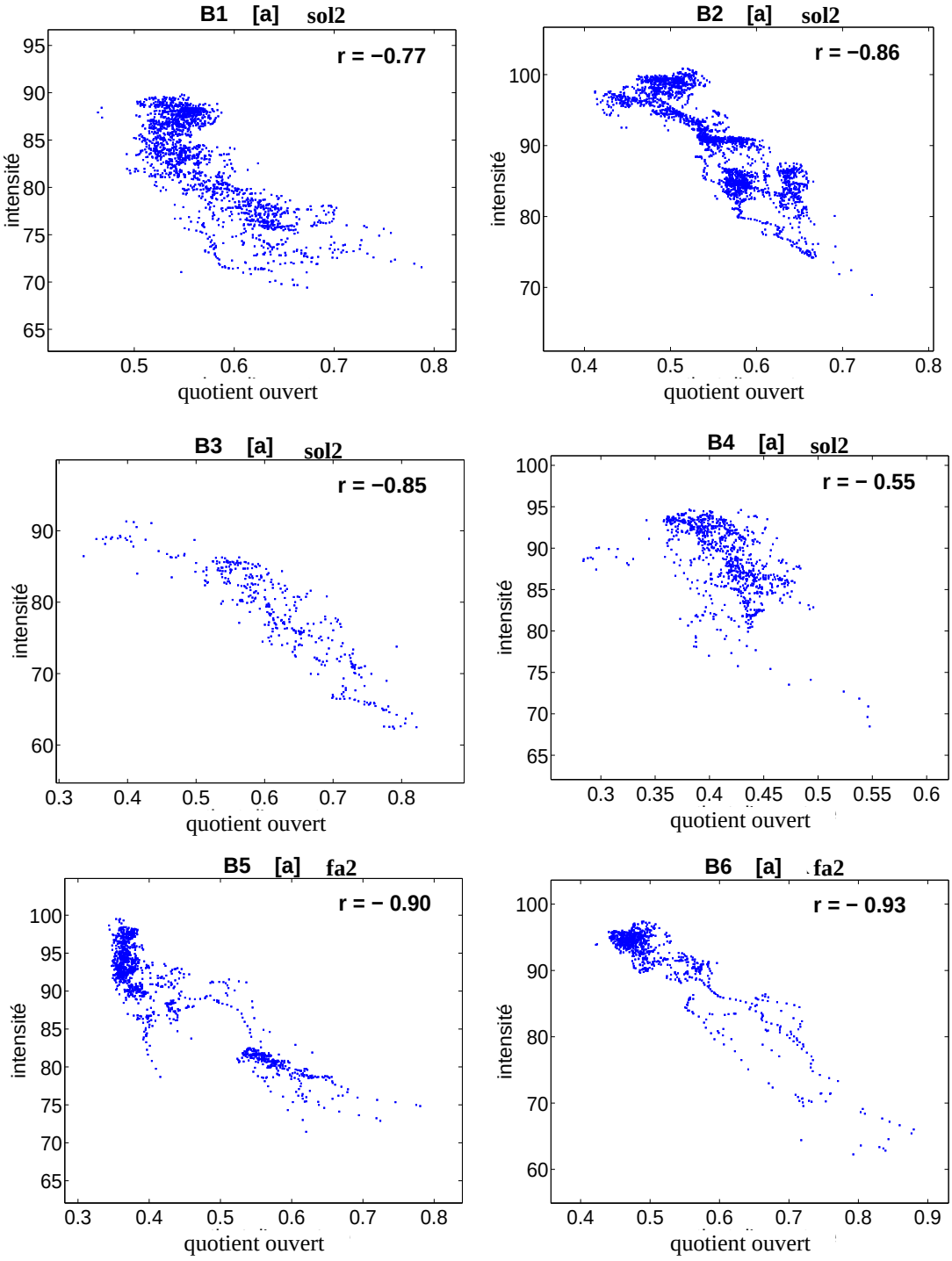


Fig. *8.7 – Intensité vocale en fonction du quotient ouvert à une hauteur donnée dans le cas des barytons présentant une corrélation forte (indiquée par r) en mécanisme I. Seule la voyelle [a] est présentée.*

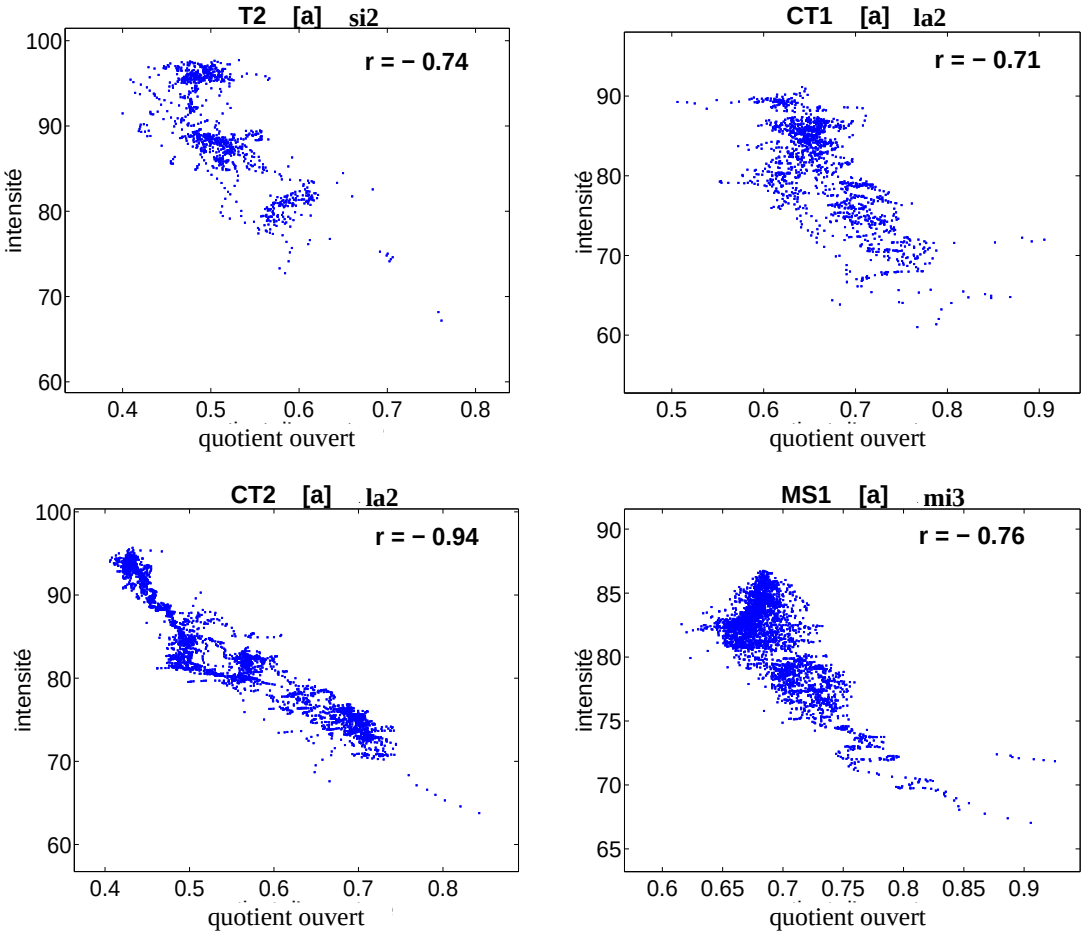


FIG. *8.8 – Intensité vocale en fonction du quotient ouvert à une hauteur donnée pour le ténor, les 2 contre-ténors et la mezzo-soprano présentant une corrélation forte (indiquée par r) en mécanisme I. Seule la voyelle [a] est présentée.*

chanteur	[a]	[e]	[u]
B1	$-\mathbf{0.78}$ ($-\mathbf{0.72}$)	-0.43 (-0.48)	-0.25 (-0.20)
B2	$-\mathbf{0.76}$ ($-\mathbf{0.79}$)	$-\mathbf{0.83}$ ($-\mathbf{0.76}$)	-0.68 (-0.46)
B3	-0.60 ($-\mathbf{0.78}$)	-0.62 ($-\mathbf{0.83}$)	-0.57 ($-\mathbf{0.76}$)
B4	-0.44 (-0.30)	-0.49 (-0.41)	-0.43 (-0.24)
B5	$-\mathbf{0.84}$ ($-\mathbf{0.90}$)	$-\mathbf{0.76}$ (-0.65)	-0.66 (-0.45)
B6	$-\mathbf{0.87}$ ($-\mathbf{0.86}$)	$-\mathbf{0.84}$ ($-\mathbf{0.83}$)	$-\mathbf{0.85}$ ($-\mathbf{0.85}$)
B7	n.s. (n.s.)	-0.21 (-0.43)	-0.09 (-0.34)
T1	-0.30 (0.07)	-0.43 (n.s.)	-0.08 (-0.12)
T2	-0.60 ($-\mathbf{0.71}$)	-0.69 (-0.55)	-0.49 (-0.31)
CT1	-0.39 (-0.55)	-0.33 (-0.26)	-0.39 (-0.37)
CT2	$-\mathbf{0.75}$ ($-\mathbf{0.92}$)	$-\mathbf{0.75}$ ($-\mathbf{0.94}$)	-0.57 ($-\mathbf{0.77}$)
MS1	-0.33 (-0.53)	-0.41 (-0.51)	-0.06 (-0.32)
MS2	0.33 ($-\mathbf{0.71}$)	0.37 (-0.34)	0.32 (-0.32)
MS3	-0.20 (-0.41)	0.11 (-0.35)	-0.12 (-0.35)

Tab. *8.3 – Coefficients de corrélation totale (partielle) entre* **intensité vocale et quotient ouvert** *pour les chanteurs utilisant* **le mécanisme I** *et les 3 voyelles [a], [e] et [u]. La corrélation partielle permet de tenir compte de la variation sous-jacente de fréquence fondamentale. Les coefficients en gras présentent une corrélation forte ($r > 0.70$). Une corrélation non significative ($p \geq 0.001$) est indiquée par n.s.*

Discussion

L'intensité vocale varie de $60 - 70$ dB pour les sons pianos à $95 - 105$ dB pour les sons forte. Le quotient ouvert varie entre 0.3 et 0.9. Il descend rarement en-dessous de 0.5 pour les chanteurs T1, CT1 et les 3 mezzo-sopranos (MS1, MS2, MS3) et il est rarement supérieur à 0.7 dans le cas des barytons. Les 6 barytons (B1, B2, B3, B4, B5 et B6), le ténor T2 et les 2 contre-ténors (CT1, CT2) présentent des comportements semblables : ils tendent à diminuer le quotient ouvert quand ils montent en intensité. Les figures 8.7 et 8.8 présentent les résultats obtenus sur la voyelle [a] pour une hauteur donnée, dans le cas des chanteurs présentant une forte corrélation. Cette tendance est particulièrement marquée (corrélation partielle forte, ($r > 0.70$)) dans le cas de la voyelle [a] pour les chanteurs B1, B2, B3, B5, B6, T2 et CT2, dans le cas de la voyelle [e] pour les chanteurs B2, B3, B6 et CT2, dans le cas de la voyelle [u] pour les chanteurs B3, B6 et CT2. La voyelle semble donc avoir un effet sur le degré d'ouverture glottique. Dans le cas des voyelles [e] et [u] où le premier formant est bas (F_1 en-dessous de 400 Hz), l'interaction de la source avec le conduit vocal peut influencer le mouvement glottique. Cet effet de la voyelle sera analysé et commenté en 8.6.

Les barytons B4 et B7 se distinguent des autres barytons. B4 ne présente pas de variations marquées du quotient ouvert, ayant toujours des valeurs comprises entre 0.35 et 0.5 quelles que soient la hauteur et l'intensité vocale. Dans le cas du baryton B7, les valeurs de quotient ouvert sont comprises entre 0.4 et 0.55, allant très rarement au-dessus de 0.6. A intensité vocale équivalente, ses valeurs correspondent à celle du chanteur B4. Si l'on regarde le tableau 6.1, on remarque que ces deux chanteurs se distinguent des autres barytons par leur répertoire : B4 travaille essentiellement le grand opéra et B7 pratique couramment le chant grégorien. Ces différences marquées avec les autres barytons pourraient donc s'expliquer par une technique vocale différente.

Le ténor T1 est également à part et se distingue nettement du ténor T2. Il ne présente que très peu de variations du quotient ouvert (valeurs comprises entre 0.65 et 0.85). Cette plage de variation de O_q est caractéristique du mécanisme II mais nous avions déjà remarqué précédemment (cf. figure 8.5) que ce chanteur ne présentait curieusement pas de valeurs différentes de quotient ouvert en mécanisme I ou II.

Si l'on s'intéresse maintenant aux résultats des mezzo-sopranos, on remarque qu'il y a également une tendance à la diminution du quotient ouvert quand l'intensité augmente, bien que peu marquée. Les coefficients de corrélation partielle varient entre -0.32 et -0.53, excepté pour MS2 sur la voyelle [a] où la corrélation est alors forte ($r = -0.71$). L'influence sous-jacente de la fréquence fondamentale est ici très importante : sur l'exemple de la chanteuse MS2, on note que le signe du coefficient de corrélation change, selon que l'on regarde la corrélation totale ou partielle. Cela signifie que, si l'on s'intéresse à la tendance générale (toutes intensités et hauteurs confondues), le quotient ouvert tend à augmenter quand l'intensité augmente. Par contre, si l'on regarde plus spécifiquement l'influence de l'intensité vocale à hauteur constante, on observe la même tendance que dans le cas des chanteurs masculins. On peut également noter que, là-aussi, les valeurs de quotient ouvert sont très élevées bien que l'on soit en mécanisme I.

8.4.3 Production vocale en mécanisme II

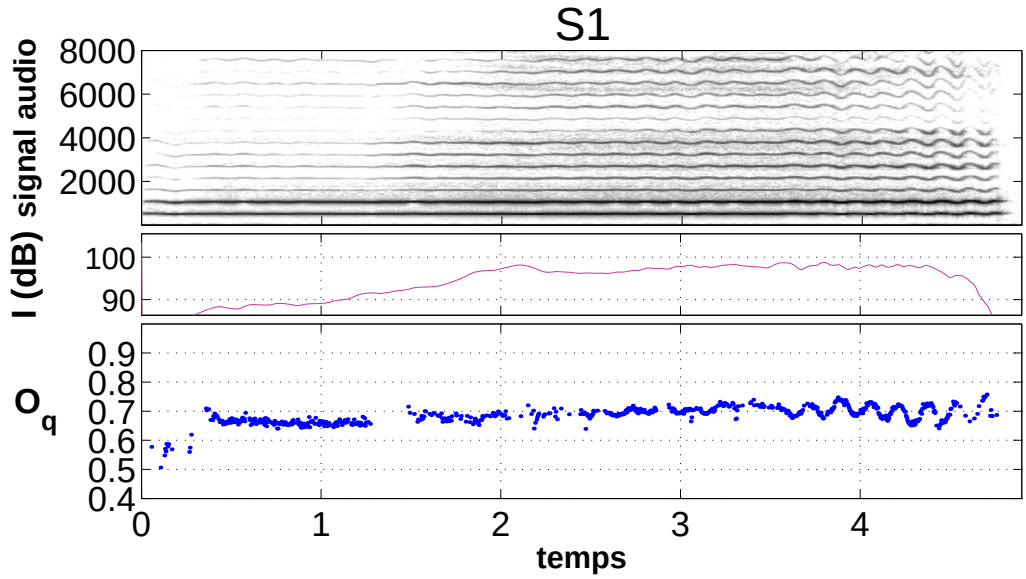


FIG. *8.9 – Son filé chanté par la soprano S1 sur un do4 (*520 Hz*) et la voyelle [u]. (CD, piste [17])*

La figure 8.9 illustre le crescendo réalisé par la soprano S1 (CD, piste [17]). L'intensité vocale augmente d'environ 10 dB. On ne retrouve pas dans ce cas la tendance observée précédemment, car le quotient ouvert reste à des valeurs constantes alors que l'intensité augmente (on observe même une très légère augmentation de O_q en début de phrase). Nous allons voir que nous pouvons généraliser cette observation dans le cas du mécanisme II.

Les figures E.5, E.6 et E.7 présentées dans l'annexe E montrent les variations d'intensité vocale en fonction du quotient ouvert en mécanisme II pour les 3 contre-ténors, les 3 mezzo-sopranos et les 3 sopranos. De même que précédemment, seuls les voyelles tenues et les sons filés ont été pris en compte. En complément, le tableau 8.4 présente les résultats de la corrélation totale et partielle entre quotient ouvert et intensité vocale, dans le cas des productions vocales en mécanisme II.

Contrairement à ce qui a été observé précédemment dans le cas du mécanisme I, aucune corrélation forte n'émerge ici. On observe même un changement de signe constant entre coefficient de corrélation totale et partielle dans le cas des 3 contre-ténors, de la mezzo-soprano MS1 et de la soprano S1, qui peut s'expliquer par le lien très fort entre intensité et fréquence fondamentale pour ces chanteurs (voir le tableau 8.2). Dans le cas des trois contre-ténors, dont les résultats sont relativement semblables, la tendance observée est une diminution générale du quotient ouvert avec l'intensité vocale, si l'on ne regarde que la corrélation totale. Par contre, nous observons la tendance inverse (i. e. une augmentation du

chanteur	[a]	[e]	[u]
CT1	−0.63 (0.42)	−0.55 (0.27)	−0.26 (0.40)
CT2	−0.44 (0.07)	−0.33 (0.33)	−0.29 (0.29)
CT3	−0.56 (0.21)	−0.50 (0.44)	−0.61 (−0.09)
MS1	n.s. (−0.54)	0.33 (−0.31)	0.24 (−0.46)
MS2	0.44 (0.22)	0.47 (−0.09)	0.49 (0.23)
MS3	n.s. (−0.41)	n.s. (−0.30)	n.s. (−0.29)
S1	0.06 (0.68)	−0.26 (0.53)	−0.10 (0.64)
S2	−0.42 (0.07)	−0.60 (−0.15)	−0.64 (n.s.)
S3	−0.54 (−0.14)	−0.39 (−0.37)	−0.53 (−0.41)

Tab. *8.4 – Coefficients de corrélation totale (partielle) entre* **intensité vocale et quotient ouvert** *pour les chanteurs utilisant* **le mécanisme II** *et les 3 voyelles [a], [e] et [u]. La corrélation partielle permet de tenir compte de la variation sous-jacente de fréquence fondamentale. Une corrélation non significative ($p \geq 0.001$) est indiquée par n.s.*

quotient ouvert avec l'intensité vocale) si l'on prend en compte la variation sous-jacente très significative de hauteur. Cette tendance se retrouve également chez la soprano S1. Les mezzo-sopranos MS1 et MS3, ainsi que les deux sopranos S2 et S3, montrent quant à elles une tendance à diminuer le quotient ouvert quand l'intensité augmente.

8.4.4 Analyse en contexte chanté

Il a été demandé aux sujets de chanter les premières mesures de l'*Ave Maria* de Gounod sur trois nuances d'intensité (piano, médium, forte). Les figures 8.10, 8.11 et 8.12 présentent les résultats des mesures du quotient ouvert moyennées sur toute la phrase musicale, selon la catégorie vocale (barytons, ténors et contre-ténors, mezzo-sopranos et sopranos). Le tableau 8.5 donne les valeurs (moyenne et écart-type) de O_q, I et f_0 pour chaque chanteur et pour les trois nuances. Les courbes d'analyse de ces phrases sont présentées par chanteur en Annexe G.

En mécanisme I, dans un contexte chanté, les tendances sont similaires à celles observées chez les chanteurs sur des voyelles isolées de tout contexte musical : le quotient ouvert diminue quand l'intensité augmente. C'est en particulier le cas pour les chanteurs B1, B2, B6, T2 et CT2 en mI. Cette tendance, bien que moins marquée, est également présente chez les chanteurs B3 et B5. Les chanteurs B4, B7 et T1 ne semblent pas montrer cette tendance. L'application du test statistique sur observations appariées (détaillé en Annexe D) montre que la diminution du quotient ouvert est fortement significative entre les productions piano et médium ($p < 0.001$) et entre les productions médium et forte ($p < 0.001$), si l'on excepte les résultats du ténor T1, dont le comportement est atypique. Si l'on inclut ce chanteur dans la statistique, la diminution de quotient ouvert avec l'augmentation de l'intensité reste malgré tout très significative ($p < 0.01$).

En mécanisme II, les tendances sont partagées. On observe une diminution de O_q avec l'intensité vocale pour les chanteurs MS1, MS3 et S3. Par contre, la tendance inverse, c'est-à-dire une augmentation de O_q avec l'intensité vocale, est observée chez les chanteurs CT3 et S1. Les chanteurs CT2 en mII, MS2 et S2 ne semblent pas montrer de tendance particulière. L'application du test statistique sur observations appariées ne montre aucune différence significative sur le quotient ouvert entre les 3 productions, alors que l'intensité varie significativement ($p < 0.01$ entre piano et médium et $p < 0.001$ entre médium et forte).

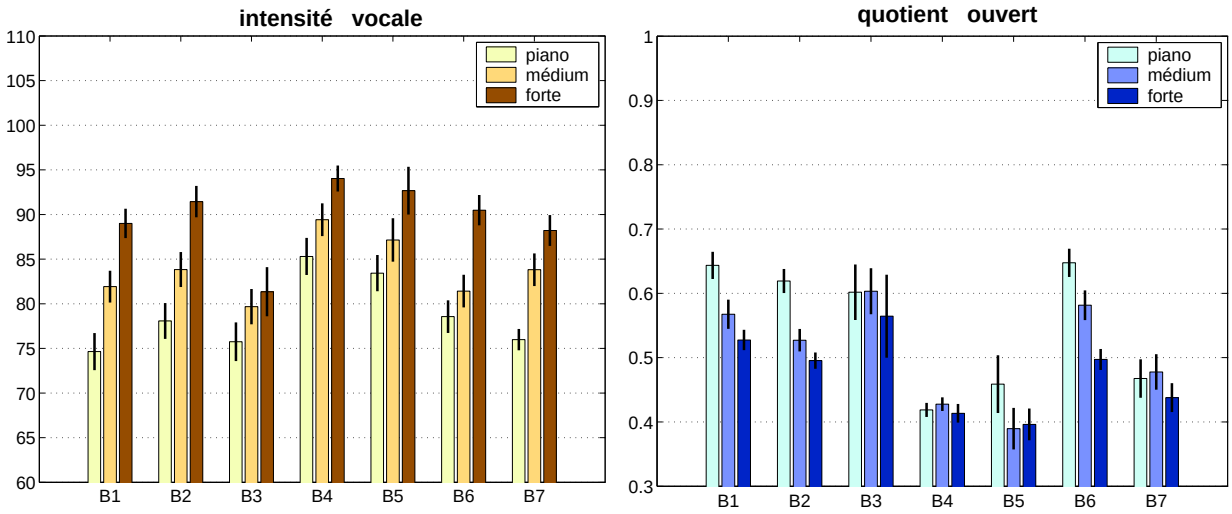


FIG. *8.10* – **Cas des barytons** *: moyenne et écart-type des valeurs de quotient ouvert mesurées sur les premières mesures de l'Ave Maria de Gounod pour trois nuances d'intensité (piano, médium, forte).*

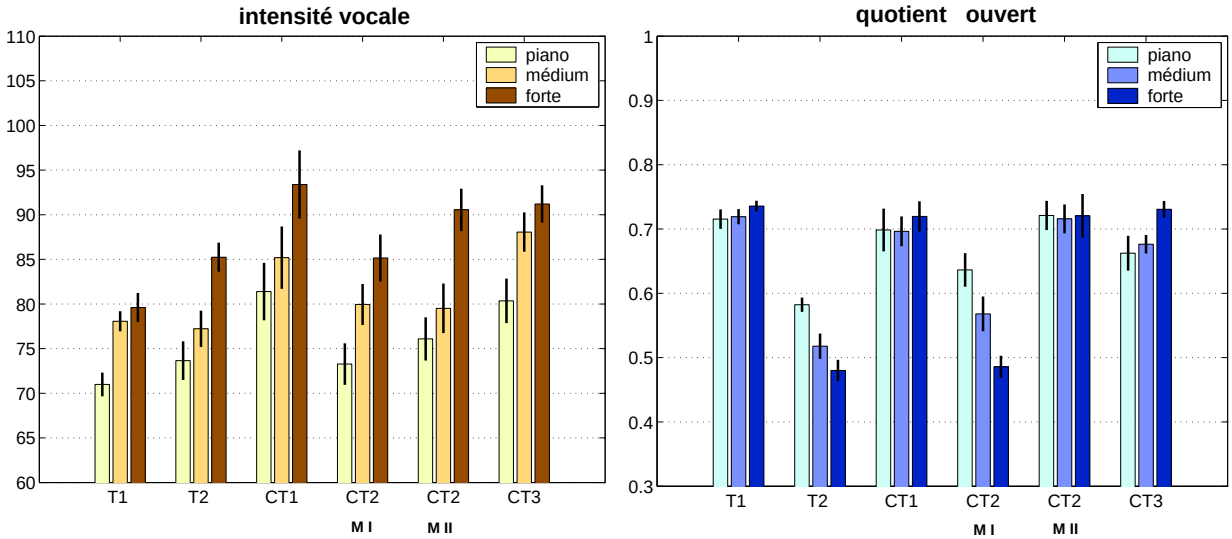


FIG. *8.11* – **Cas des ténors et des contre-ténors** *: moyenne et écart-type des valeurs de quotient ouvert mesurées sur les premières mesures de l'Ave Maria de Gounod pour trois nuances d'intensité (piano, médium, forte). A noter que le contre-ténor CT2 a effectué cette tâche dans les deux mécanismes, les 2 autres contre-ténors étant en mII.*

	piano			médium			forte		
	I_m	O_{qm}	f_{0m}	I_m	O_{qm}	f_{0m}	I_m	O_{qm}	f_{0m}
B1	74 (4)	0.64 (0.04)	231 (24)	81 (3)	0.57 (0.05)	231 (21)	89 (3)	0.53 (0.03)	227 (17)
B2	78 (4)	0.62 (0.04)	237 (23)	83 (3)	0.53 (0.04)	236 (15)	91 (3)	0.50 (0.03)	241 (20)
B3	75 (4)	0.60 (0.09)	239 (24)	79 (3)	0.60 (0.07)	230 (17)	81 (5)	0.56 (0.13)	233 (20)
B4	85 (4)	0.42 (0.02)	229 (17)	89 (3)	0.43 (0.02)	231 (18)	94 (2)	0.41 (0.03)	236 (18)
B5	83 (4)	0.46 (0.09)	238 (21)	87 (4)	0.39 (0.06)	246 (21)	92 (5)	0.40 (0.05)	234 (22)
B6	78 (3)	0.65 (0.04)	234 (20)	81 (3)	0.58 (0.05)	248 (19)	90 (3)	0.50 (0.03)	237 (22)
B7	75 (2)	0.47 (0.06)	214 (9)	83 (3)	0.48 (0.06)	211 (12)	88 (3)	0.44 (0.04)	210 (9)
T1	70 (2)	0.72 (0.03)	228 (16)	78 (2)	0.72 (0.02)	230 (15)	79 (3)	0.74 (0.02)	228 (16)
T2	73 (4)	0.58 (0.02)	256 (23)	77 (4)	0.52 (0.04)	236 (22)	85 (3)	0.48 (0.03)	234 (18)
CT1 (mII)	81 (6)	0.70 (0.07)	391 (34)	85 (6)	0.70 (0.05)	388 (28)	93 (7)	0.72 (0.05)	404 (34)
CT1 (mI)	73 (4)	0.64 (0.05)	208 (23)	79 (4)	0.57 (0.05)	208 (18)	85 (5)	0.49 (0.03)	205 (18)
CT1 (mII)	76 (4)	0.72 (0.05)	393 (26)	79 (5)	0.72 (0.04)	401 (26)	90 (4)	0.72 (0.07)	402 (41)
CT3 (mII)	80 (4)	0.66 (0.05)	394 (27)	88 (4)	0.68 (0.03)	382 (20)	91 (4)	0.73 (0.03)	383 (19)
MS1	78 (8)	0.74 (0.04)	487 (37)	82 (6)	0.69 (0.04)	468 (39)	83 (6)	0.64 (0.03)	463 (37)
MS2	71 (3)	0.82 (0.03)	323 (24)	76 (3)	0.74 (0.05)	326 (25)	82 (4)	0.80 (0.04)	333 (32)
MS3	76 (5)	0.86 (0.06)	449 (47)	78 (3)	0.81 (0.05)	443 (26)	81 (3)	0.78 (0.04)	452 (36)
S1	79 (2)	0.57 (0.04)	477 (42)	88 (2)	0.70 (0.03)	463 (34)	93 (3)	0.71 (0.04)	471 (44)
S2	88 (4)	0.81 (0.03)	455 (43)	91 (4)	0.80 (0.02)	448 (40)	94 (4)	0.80 (0.03)	458 (43)
S3	77 (5)	0.78 (0.04)	553 (44)	84 (3)	0.72 (0.04)	555 (42)	87 (3)	0.63 (0.03)	552 (39)

Tab. *8.5 – Moyenne et écart-type (entre parenthèses) des variations de O_q sur les premières mesures de l'Ave Maria de Gounod. Les moyennes et écart-types correspondant de I (en dB) et f_0 (en Hz) sont également indiqués.*

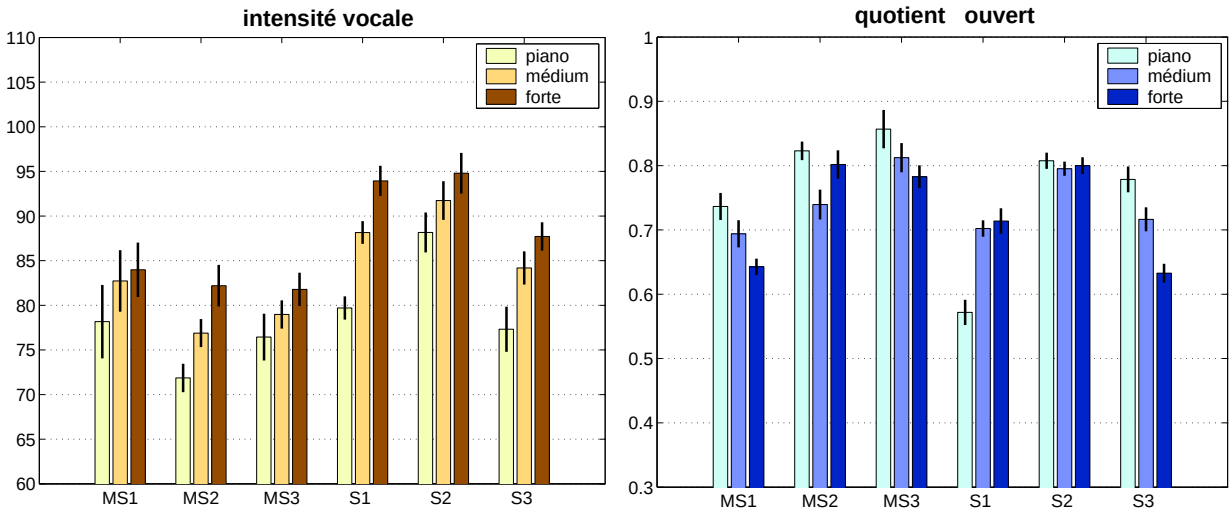


FIG. *8.12* – **Cas des mezzo-sopranos et des sopranos** *: moyenne et écart-type des valeurs de quotient ouvert mesurées sur les premières mesures de l'Ave Maria de Gounod pour trois nuances d'intensité (piano, médium, forte).*

8.4.5 Conclusion

En conclusion, nous observons en mécanisme I une diminution marquée du quotient ouvert quand l'intensité augmente, que ce soit pour des productions vocales hors d'un contexte musical (cas des voyelles tenues, des sons filés) ou dans un contexte musical donné. Par contre, cette tendance n'est plus vérifiée dans le cas du mécanisme II, où la tendance inverse peut même être observée chez certains chanteurs.

Ces résultats sont en accord avec les observations de Timcke & al [197], Kitzing & Sonesson [131], Hanson & al [96], Orlikoff [161], Dromey & al [57], Holmberg & al [109], Sundberg & al [193, 188, 190, 192, 191]. Ils confirment donc que la méthode de mesure sur signaux DEGG semble adaptée à la mesure du quotient ouvert, puisque nous retrouvons les tendances attendues entre quotient ouvert et intensité vocale.

A la vue de ces résultats, il semble qu'une augmentation de l'intensité vocale ne soit pas le fait des mêmes facteurs, en mécanisme I ou en mécanisme II. Comme nous l'avons noté en partie préliminaire, une augmentation de l'intensité vocale résulte généralement d'une augmentation de la pression sous-glottique, ceci indépendamment du mécanisme laryngé considéré. Par contre, la participation du muscle vocal au mouvement glottique n'est pas la même selon les mécanismes. En mécanisme I, la contraction du vocalis a une incidence directe sur le mouvement vibratoire glottique. Dans l'aigu du mécanisme I, ce muscle atteint sa limite physiologique de contraction. En mécanisme II, on observe alors un relâchement relatif de ce muscle au profit des muscles crico-thyroïdiens. La diminution du quotient ouvert observée en mécanisme I pourrait donc être le fait d'une contraction du muscle vocal lors d'une augmentation d'intensité. Cela expliquerait qu'en mécanisme II, où son action se trouve plus limitée, nous n'observons plus de variation du quotient ouvert avec l'intensité vocale.

Bien que la validité des modèles d'ODG dans le cadre du chant soit discutable, il nous paraît néanmoins intéressant de noter qu'une augmentation de l'intensité vocale en relation avec une diminution du quotient ouvert est en accord avec ce que prédisent ces modèles. En effet, nous avons vu au chapitre 3 que la vitesse de fermeture E s'exprime, dans le cas d'une fermeture abrupte, sous la forme $E = \frac{A_v}{O_q T_0} e_n(\alpha_m)$ (équation 3.3). Cette relation traduit qu'en théorie, la vitesse de fermeture peut augmenter avec une augmentation de A_v, de f_0 ou de α_m ou avec une diminution de O_q. En pratique, cette vitesse de fermeture semble fortement corrélée à l'intensité vocale, d'après les résultats expérimentaux obtenus par Fant [66], Holmberg & al [109], Gauffin & Sundberg [85], Sundberg & al [193]. Une diminution de O_q se traduit donc par une

chanteur	[a]	[e]	[u]
B1	−0.47 (0.22)	n.s. (0.24)	−0.15 (n.s.)
B2	−0.29 (0.45)	−0.51 (0.15)	−0.57 (n.s.)
B3	0.13 (0.64)	0.12 (**0.71**)	0.19 (0.63)
B4	−0.34 (n.s.)	−0.30 (0.11)	−0.39 (−0.15)
B5	−0.12 (0.62)	−0.53 (−0.11)	−0.56 (−0.15)
B6	−0.11 (n.s.)	−0.15 (n.s.)	−0.11 (−0.13)
B7	n.s. (n.s.)	0.60 (0.68)	0.67 (**0.71**)
T1	−0.50 (−0.42)	−0.49 (−0.27)	n.s. (0.10)
T2	n.s. (0.49)	−0.51 (−0.17)	−0.44 (−0.22)
CT1	n.s. (0.42)	−0.22 (n.s.)	−0.18 (n.s.)
CT2	0.32 (**0.83**)	0.46 (**0.90**)	0.20 (0.65)
MS1	0.21 (0.48)	−0.07 (0.34)	0.14 (0.34)
MS2	**0.75** (**0.87**)	**0.73** (**0.72**)	**0.75** (**0.75**)
MS3	0.24 (0.43)	0.40 (0.50)	0.18 (0.37)

Tab. *8.6 – Coefficients de corrélation totale (partielle) entre* **fréquence fondamentale et quotient ouvert** *pour les chanteurs utilisant* **le mécanisme I** *et les 3 voyelles [a], [e] et [u]. La corrélation partielle permet de tenir compte de la variation sous-jacente d'intensité vocale. Les coefficients en gras présentent une corrélation forte ($r > 0.70$). Une corrélation non significative ($p \geq 0.001$) est indiquée par n.s.*

augmentation de E, donc de l'intensité vocale. Dans le cas du mécanisme II, si le quotient ouvert reste constant, seules une augmentation de l'amplitude de voisement A_v ou de l'asymétrie de l'impulsion α_m pourraient alors entraîner une augmentation de la vitesse de fermeture, donc de l'intensité vocale, à fréquence fondamentale constante.

8.5 Quotient ouvert et fréquence fondamentale

Lors de la visualisation des glissandos, nous avons noté une corrélation visuelle très nette entre fréquence fondamentale et quotient ouvert. Cette corrélation semble être caractéristique du mécanisme II et ne se retrouve pas lors des passages en mécanisme I. Nous allons maintenant explorer ce lien possible entre quotient ouvert et fréquence fondamentale dans le cas des 2 mécanismes, en nous appuyant sur les mesures effectuées sur les voyelles tenues et les sons filés chantés sur plusieurs hauteurs.

8.5.1 Production vocale en mécanisme I

Le tableau 8.6 présente les résultats de la corrélation entre quotient ouvert et fréquence fondamentale, dans le cas des productions vocales en mécanisme I. De manière générale, il ne semble pas y avoir de corrélation marquée entre O_q et f_0. Néanmoins, des individualités se distinguent. Les barytons B3 et B7, le contre-ténor CT2 et la mezzo-soprano MS2 présentent une corrélation positive importante voire forte entre O_q et f_0 : dans le cas de ces chanteurs, le quotient ouvert tend à augmenter avec la fréquence fondamentale. Cet effet est contrebalancé par l'effet de l'intensité chez les chanteurs B3 et CT2, ce qui explique que la corrélation totale soit faible, alors que la corrélation partielle est importante.

Une augmentation de O_q lors d'une augmentation de f_0 a également été observée par Kitzing & Sonesson [131], Holmberg & al [110], Hanson & al [96], Childers & al [31], Cookman & al [41]. Néanmoins, la fréquence fondamentale est souvent fortement corrélée à l'intensité vocale. Comme le quotient ouvert diminue avec une augmentation de l'intensité vocale en mécanisme I, cet effet peut compenser une

chanteur	[a]	[e]	[u]
CT1	**−0.82** (**−0.74**)	**−0.78** (−0.69)	**−0.74** (**−0.77**)
CT2	**−0.79** (**−0.73**)	**−0.73** (**−0.73**)	**−0.71** (**−0.71**)
CT3	**−0.85** (**−0.78**)	**−0.75** (**−0.73**)	−0.69 (−0.41)
MS1	0.36 (0.62)	0.61 (0.61)	0.63 (**0.71**)
MS2	0.42 (0.18)	0.53 (0.31)	0.44 (n.s.)
MS3	0.26 (0.47)	0.37 (0.45)	0.22 (0.35)
S1	−0.57 (**−0.80**)	−0.60 (**−0.71**)	−0.60 (**−0.79**)
S2	−0.59 (−0.46)	−0.64 (−0.30)	**−0.73** (−0.46)
S3	−0.57 (−0.26)	−0.20 (0.17)	−0.37 (0.08)

TAB. *8.7 – Coefficients de corrélation totale (partielle) entre* **fréquence fondamentale et quotient ouvert** *pour les chanteurs utilisant* **le mécanisme II** *et les 3 voyelles [a], [e] et [u]. La corrélation partielle permet de tenir compte de la variation sous-jacente d'intensité vocale. Les coefficients en gras présentent une corrélation forte ($r > 0.70$). Une corrélation non significative ($p \geq 0.001$) est indiquée par n.s.*

augmentation due à la fréquence fondamentale et limiter, dans la plupart des cas, le lien entre O_q et f_0.

8.5.2 Production vocale en mécanisme II

Le tableau 8.7 présente les résultats de la corrélation entre quotient ouvert et fréquence fondamentale, dans le cas des productions vocales en mécanisme II. La corrélation entre O_q et f_0 est très forte chez les trois contre-ténors, ce que nous avions déjà pu observer sur les glissandos, ainsi que chez les deux sopranos S1 et S2. Une augmentation de fréquence fondamentale se traduit ici par une diminution sensible du quotient ouvert. La troisième soprano (S3) ne semble pas présenter de tendance particulière. Les mezzo-sopranos présentent plutôt une corrélation inverse (effet comparable à celui observé précédemment dans le cas du mécanisme I), bien que celle-ci ne soit pas très marquée (sauf chez le sujet MS1).

Ces résultats sont particulièrement en accord avec ceux obtenus par Howard [113, 112], qui note que le lien entre O_q et f_0 dépend de l'entraînement vocal. Ce lien est donc peu marqué chez les chanteuses pas ou peu entraînées. Cela pourrait expliquer les résultats inverses obtenus dans le cas des 3 mezzo-sopranos, de jeunes chanteuses non professionnelles, chantant principalement en chorale, et qui ont moins d'années d'entraînement vocal que les autres chanteurs enregistrés pour cette étude.

8.5.3 Conclusion

En mécanisme I, le quotient ouvert ne semble pas corrélé à la fréquence fondamentale. Nous observons néanmoins pour certains chanteurs une augmentation de O_q quand f_0 augmente. Dans le cas du mécanisme II, par contre, la corrélation est très nette entre O_q et f_0 chez les contre-ténors et les sopranos : une augmentation de fréquence fondamentale s'accompagne d'une diminution du quotient ouvert.

Ces résultats semblent en accord avec les observations faîtes par Howard [113, 112] : diminution de O_q quand f_0 augmente chez des sujets féminins et pas de relation entre ces deux paramètres chez des sujets masculins. Cette concordance suggère également que les différences entre sexes présentées dans ces études sont vraisemblablement dues à une différence de mécanismes (sujets masculins $\leftrightarrow$ mécanisme I, sujets féminins $\leftrightarrow$ mécanisme II).

En mécanisme I, la fréquence fondamentale augmente principalement par l'augmentation de raideur

chanteur	comparaisons voyelles [a] - [e]	comparaisons voyelles [a] - [u]	comparaisons voyelles [e] - [u]
mécanisme I			
B1	3.77 ***	6.03 ***	2.27 *
B2	4.42 ***	9.11 ***	4.69 *
B3	8.45 ***	10.31 ***	1.86 *
B4	−0.06	−1.46	−1.41
B5	8.03 ***	8.60 ***	0.57
B6	6.39 ***	8.56 ***	2.17
B7	0.58	−3.09 **	−3.67
T1	3.14 *	2.90	−0.24
T2	6.96 ***	4.16 *	−2.80 *
CT1	−1.54	0.04	1.59
CT2	4.20 **	6.51 ***	2.31 *
MS1	3.59 ***	2.29 ***	−1.30 *
MS2	0.42	−2.89 *	−3.31 *
MS3	5.33 **	2.41	−2.92
mécanisme II			
CT1	2.51	8.15 **	5.63 *
CT2	−0.13	0.56	0.69
CT3	1.66	−1.97	−3.63 **
MS1	−3.20 *	−3.62	−0.42
MS2	5.02	2.34	−2.68
MS3	9.11 ***	5.27 ***	−3.84 **
S1	0.50	1.38	0.88
S2	0.48	0.37	−0.11
S3	2.37 **	−1.09 *	−3.46 ***

Tab. *8.8 – Différence moyenne de quotient ouvert (en %) entre deux voyelles. Les différences significatives ($p < 0.05$) sont marquées par *, les différences très significatives ($p < 0.01$) par ** et les différences fortement significatives ($p < 0.001$) par ***.*

des cordes vocales, du fait de la contraction du vocalis. Le rôle de la pression sous-glottique dans une modification de la fréquence fondamentale reste limité. En mécanisme II, l'augmentation de fréquence fondamentale est plus le fait d'un étirement des cordes vocales sous l'effet de la contraction des muscles crico-thyroïdiens et d'une augmentation de la pression sous-glottique. On peut alors se demander si une diminution du quotient ouvert ne serait pas le reflet d'une augmentation de la pression sous-glottique, indépendamment du mécanisme laryngé. Ceci irait dans le sens des observations effectuées par Sundberg & al [190] entre quotient ouvert et pression sous-glottique.

8.6 Effet de la voyelle

Sans rentrer dans une étude exhaustive de la variance possible du quotient ouvert en fonction de la voyelle, nous désirons néanmoins illustrer qualitativement l'effet de la voyelle sur le quotient ouvert. Nous précisons qu'il s'agit ici de la voyelle que le chanteur a voulu émettre, mais que, en particulier selon la hauteur du son, la notion même de voyelle peut perdre de sa consistance. Cette étude pourrait se compléter par une étude détaillée de la position fréquentielle de formants associés à ces voyelles, que nous n'avons pas entreprise ici.

La comparaison s'effectue à l'aide des graphes présentés sur les figures 8.13, 8.14 et 8.15, où les

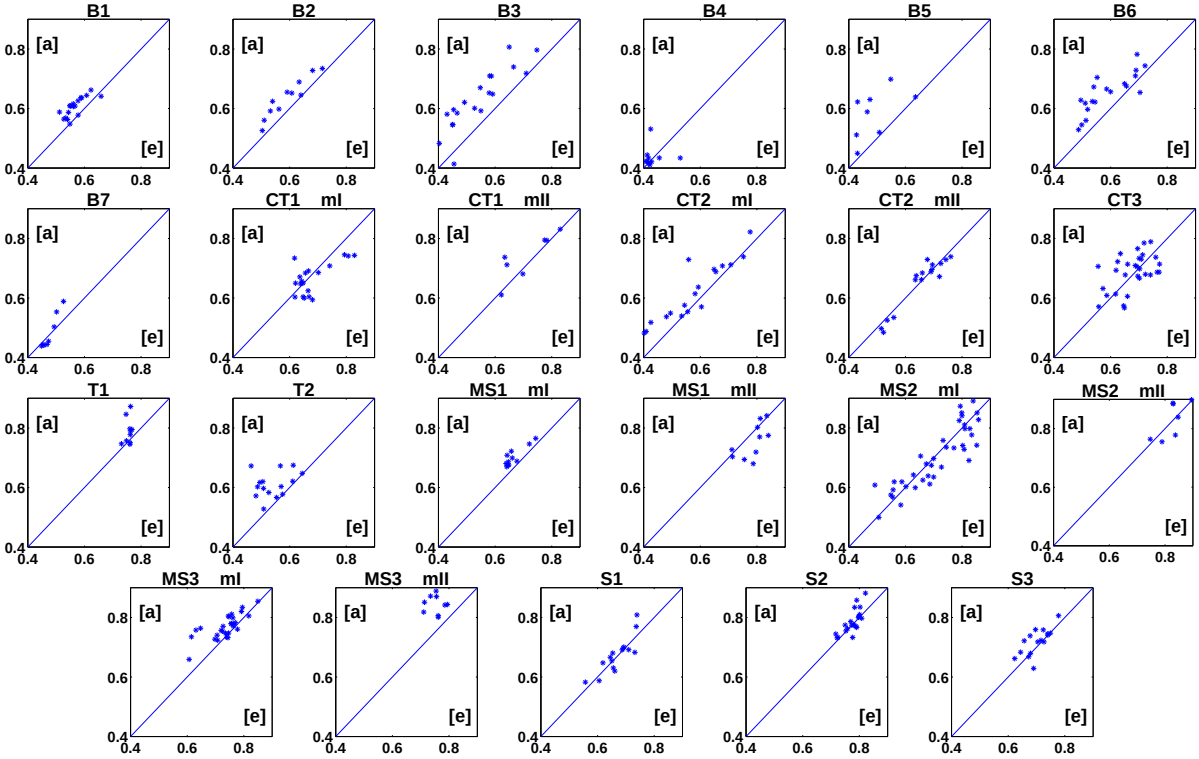


FIG. *8.13* – **Comparaison sur O_q entre la voyelle [e] et la voyelle [a]**. *Les points représentent des valeurs identiques de f_0 et I pour les deux voyelles. La droite représente l'équivalence des deux voyelles.*

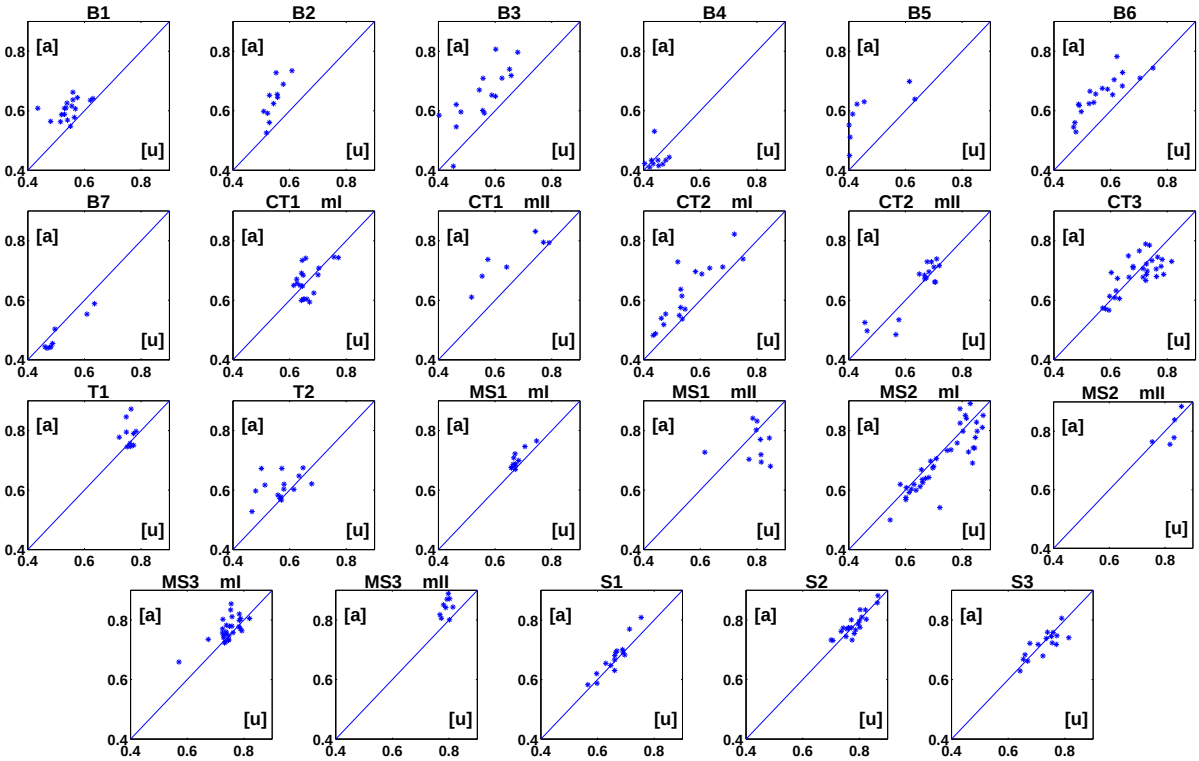


FIG. *8.14* – **Comparaison sur O_q entre la voyelle [u] et la voyelle [a]**. *Les points représentent des valeurs identiques de f_0 et I pour les deux voyelles. La droite représente l'équivalence des deux voyelles.*

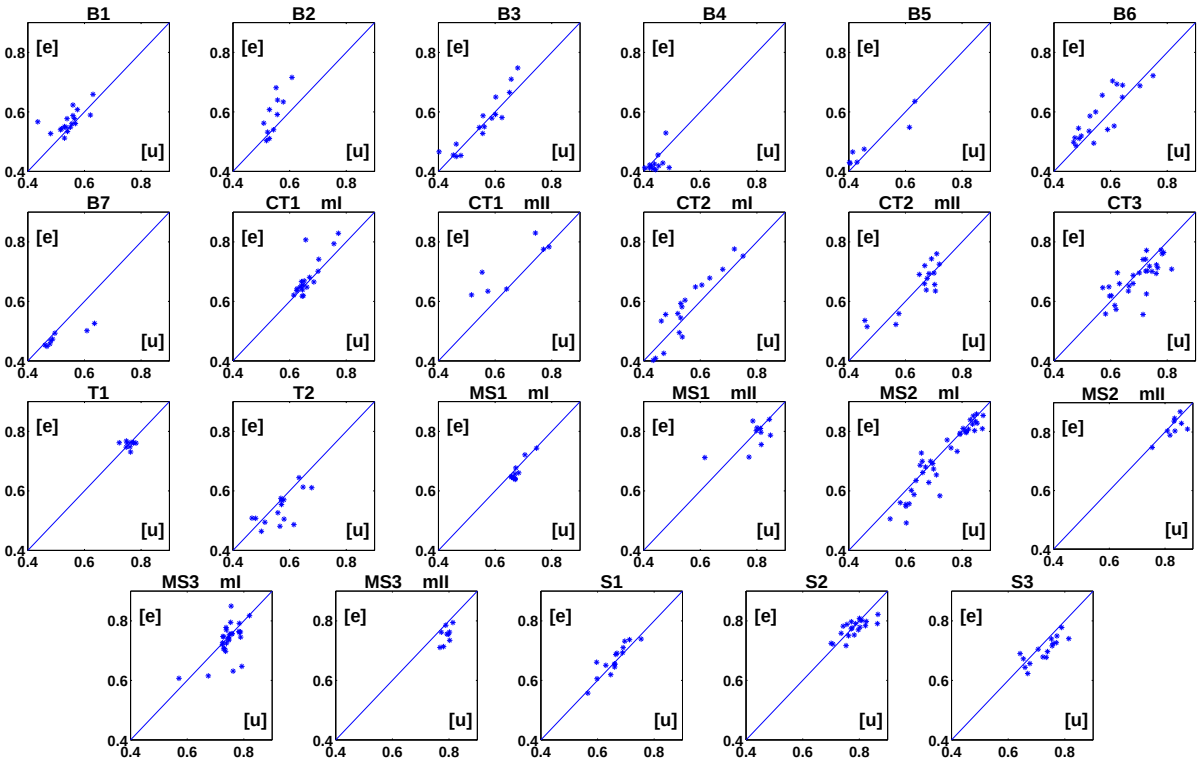


Fig. *8.15* – **Comparaison sur O_q entre la voyelle [u] et la voyelle [e]**. *Les points représentent des valeurs identiques de f_0 et I pour les deux voyelles. La droite représente l'équivalence des deux voyelles.*

valeurs de quotient ouvert sont comparées entre deux voyelles, à même intensité vocale et même fréquence fondamentale. La voyelle dont le premier formant a la fréquence la plus basse est placée en abscisse. La droite représentée sur la figure traduit l'équivalence de quotient ouvert entre les deux voyelles. Pour un couple intensité - fréquence donné (I , f_0), la sélection se fait dans des intervalles fixes d'intensité et de fréquence (I ± 1 dB , f_0 ± 5 Hz). Le tableau 8.8 donne les valeurs des différences moyennes calculées pour chaque chanteur entre les voyelles [a] et [e], les voyelles [a] et [u] et les voyelles [e] et [u], ainsi que la signification de la différence observée.

Les quotients ouverts mesurés dans le cas des voyelles [e] et [u] ne sont pas significativement très différents, à l'exception des cas du contre-ténor CT3, de la mezzo-soprano MS3 et de la soprano S3, tous les 3 en mécanisme II. Par contre, on peut observer une différence très significative entre la voyelle [a] et les voyelles [e] et [u], essentiellement en mécanisme I. Dans ces cas, les valeurs de quotient ouvert sont plus faibles pour les voyelles ayant un premier formant bas ([e] et [u]). Ceci est vérifié de façon très significative chez les chanteurs B1, B2, B3, B5, B6, T2, CT2 en mI, T2, CT2 et MS1 en mI. Pour les autres chanteurs, la différence est moins marquée ou inexistante, ce qui se traduit visuellement par des valeurs plus ou moins réparties de part et d'autre de la droite d'équivalence. En mécanisme II, seules les chanteuses MS3 et S3 présentent des différences très significatives entre les voyelles.

Nous n'avons pas prétention à tirer des conclusions générales sur l'effet de la voyelle par rapport aux valeurs de quotient ouvert, à partir de cette étude préliminaire. Il faudrait pour cela compléter ces résultats par une analyse de la position fréquentielle des formants associés à ces voyelles, dans le cas de chaque chanteur. Nous avons simplement observé ici que le quotient ouvert semble être influencé par la voyelle demandée au chanteur dans le cas du mécanisme I, mais peu influencé dans le cas du mécanisme II. Dans le cas où une différence très significative est rapportée, on observe que la voyelle [a] est généralement produite avec des valeurs de quotient ouvert plus élevées que les voyelles [e] et [u]. Les valeurs de quotient ouvert associées aux voyelles [e] et [u] sont en général comparables. Une grande variabilité inter-sujets au sein d'un même mécanisme est à noter, certains sujets comme les chanteurs $B4$, $B7$ et $T1$ présentant des comportements particuliers.

La voyelle [a] est une voyelle ouverte, tandis que les voyelles [e] et [u] sont des voyelles fermées. On peut alors se demander si l'ouverture de la bouche pourrait expliquer la différence de quotient ouvert entre ces voyelles. La tendance que nous observons est pourtant en contradiction avec les résultats de Cookman & al[41], qui observent une diminution du quotient ouvert quand l'ouverture de la bouche augmente. Néanmoins, ces mêmes auteurs notent que leurs résultats vont à l'encontre des techniques enseignés dans le chant ou lors d'une thérapie vocale, où justement des exercices d'ouverture buccale sont appliqués pour tenter de supprimer une possible tension laryngée, par exemple dans le cas de voix hyperfonctionnelles. Nos résultats seraient donc en accord avec ces techniques, puisque nous observons une augmentation du quotient ouvert dans le cas de la voyelle ouverte [a] comparativement aux voyelles fermées [e] et [u], ce qui laissent supposer un relâchement de la compression au niveau laryngé.

Dans une étude portant sur la mesure de pression sous-glottique dans le cas des 3 voyelles [a], [i] et [u] en français, les études présentées par Demolin & al [53] et Bucella & la [24] (2000) montrent que la pression sous-glottique est plus faible dans le cas de la voyelle [a], comparativement aux 2 autres voyelles. Par contre, aucune différence significative ne se dégage des mesures de pression sous-glottique entre les voyelles [i] et [u]. Notre étude semble en accord avec ces résultats, puisque nous observons que le quotient ouvert est plus élevé dans le cas du [a] relativement aux cas du [e] et du [u]. L'augmentation du quotient ouvert serait alors lié à une diminution de la pression sous-glottique. Comme le notent ces auteurs, une bascule vers l'avant du cartilage thyroïdien a lieu lors de la production des voyelles [i] et [u], ce qui accroît la tension des cordes vocales. L'augmentation de la pression sous-glottique et la diminution du quotient ouvert lors de la production des voyelles [i] et [u] seraient alors induites par ce basculement. Une autre hypothèse avancée par ces auteurs porte sur le contrôle respiratoire. Dans ce cas, la variation de la pression sous-glottique serait alors la cause directe d'une modification du quotient ouvert.

8.7 Comparaison parlé - chanté

Au début du protocole, il était demandé aux chanteurs de choisir une phrase en français de leur choix, de l'énoncer et de la chanter. Nous présentons ici l'exemple d'un homme utilisant principalement le mécanisme I en voix chantée, le baryton B2, et l'exemple d'une femme utilisant principalement le mécanisme II en voix chantée, la soprano S2. Ces deux chanteurs ont également crié la phrase choisie.

8.7.1 Exemple du baryton B2

La phrase choisie par le baryton B2 est « *Oh sainte médaille* ». La figure 8.16 présente la production parlée, la figure 8.17 la production chantée et la figure 8.18 la production criée. Seule l'échelle de quotient ouvert a été maintenue fixe entre ces trois figures pour permettre la comparaison visuelle. L'échelle temporelle est dilatée dans le chant : ce chanteur a pris plus de 6 s pour chanter l'extrait contre près de 3 s pour l'énoncer ou le crier. On note également que le chant est émis à une intensité vocale plus forte, de 5 à 10 dB, que la parole, mais à une intensité vocale plus faible, de quelques dB, que le cri. La fréquence fondamentale varie également selon la production. En particulier, nous pouvons remarquer que le cri se traduit par une augmentation de la tessiture d'environ d'environ 50 Hz (de l'ordre d'une quinte). Si l'on écoute attentivement l'exemple sonore associé (CD, piste [20]), on remarque également que ce chanteur élève la voix, mais ne crie pas comme le fait la soprano S2 que nous étudierons ensuite (CD, piste [24]).

Entre la parole et le chant, le quotient ouvert semble varier de façon comparable. Il faut noter que la phrase a été énoncée de manière théâtrale, ce qui pourrait expliquer des similitudes entre expression parlée et chantée dans ce cas, similitudes qui ne se retrouveraient pas forcément sur des phrases dites de façon plus naturelle. On note qu'il y a plus de mesures incertaines en parole, qui ne sont donc pas présentées sur cette figure, et que les parties voisées sont plus hachées. Néanmoins, une évolution similaire se remarque : la phrase débute et se termine à des valeurs de O_q proche de 0.7 dans les deux cas. L'ambitus des valeurs de O_q en parole semble un peu plus étendue puisque les valeurs extrêmes sont $O_{q\,min} = 0.42$ et $O_{q\,max} = 0.75$ pour la phrase parlée, tandis que pour la phrase chantée, $O_{q\,min} = 0.46$ et $O_{q\,max} = 0.69$.

Les transitions entre syllabes ainsi que les débuts et fins de phrase sont marquées par une augmentation sensible et rapide du quotient ouvert. Ces variations rapides de quotient ouvert sont comparables à ce que l'on peut observer dans le cas d'un changement de mécanisme, bien qu'ici la production soit réalisée en mécanisme I tout du long. Elles traduisent vraisemblablement un relâchement dans le mouvement vibratoire des cordes vocales. Cette caractéristique du quotient ouvert se retrouve sur tous les exemples de production vocale en contexte présenté dans cette étude, ainsi que nous pouvons le voir en particulier sur les figures présentées dans l'Annexe G qui montrent les variations de O_q sur les premières mesures de l'*Ave Maria* de Gounod.

Une autre caractéristique du quotient ouvert dont nous n'avons pas parlé jusqu'à présent s'observe sur ces productions vocales chantées : la variation conjointe du quotient ouvert et du vibrato. La production vocale chantée se distingue bien souvent de la production vocale parlée par l'ajout d'un vibrato. Ainsi la différence principale sur O_q entre la production parlée du chanteur B2 sur la figure 8.16 et sa production chantée sur la figure 8.17 porte sur la présence marquée d'un « vibrato de quotient ouvert » dans le cas du chant. Ce vibrato de quotient ouvert est exactement en opposition de phase avec celui de la fréquence fondamentale. Ce point n'est pas le fait unique de ce chanteur, mais se retrouve pour toutes les productions vocales chantées présentant un vibrato. Il demanderait une étude en soi.

La production vocale criée se distingue des deux autres par des variations rapides de O_q, en dehors des changements syllabiques et des attaques. Le quotient ouvert a également des valeurs plus basses que pour les deux autres productions : la plupart des mesures sont situées entre $O_q = 0.4$ et $O_q = 0.5$.

Ce chanteur a également énoncé la phrase en mécanisme II. L'analyse de cette production est présentée sur la figure 8.19. L'échelle de quotient ouvert a été modifiée pour permettre de visualiser les détails des variations de O_q. L'amplitude de ces variations est très restreinte : $O_{q\,min} = 0.77$ et $O_{q\,max} = 0.89$.

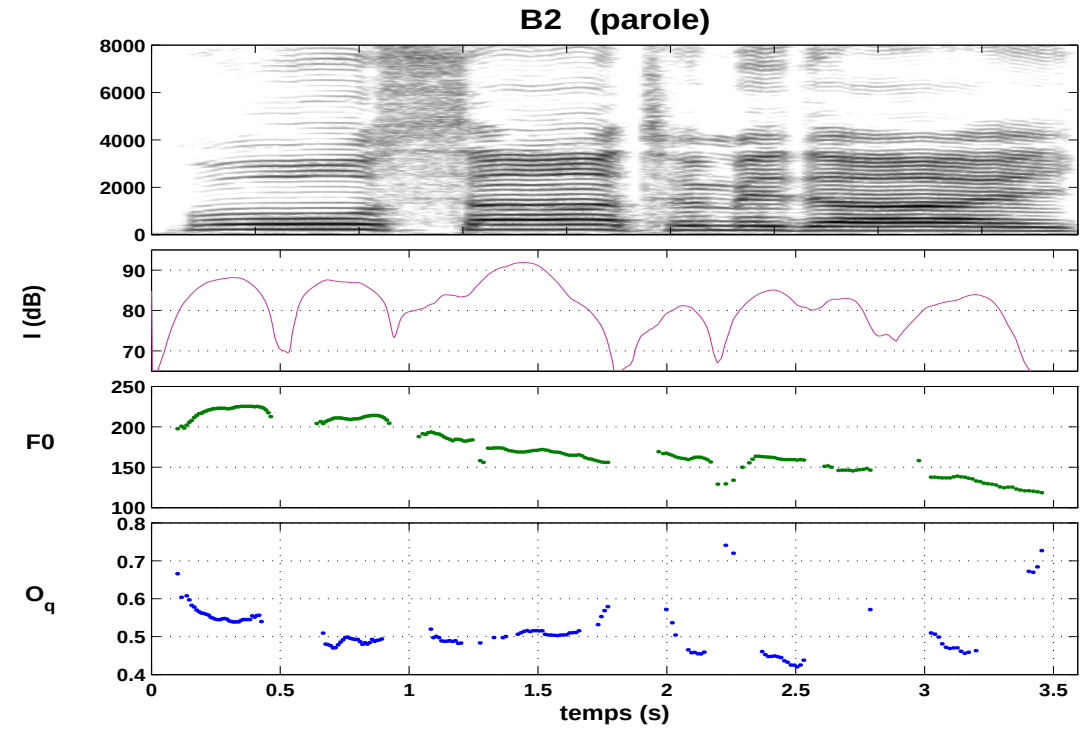


FIG. *8.16 – Mesure de I_{db}, f_0 et O_q pour la phrase énoncée par le baryton B2 : « Oh sainte médaille ». (CD, piste [18])*

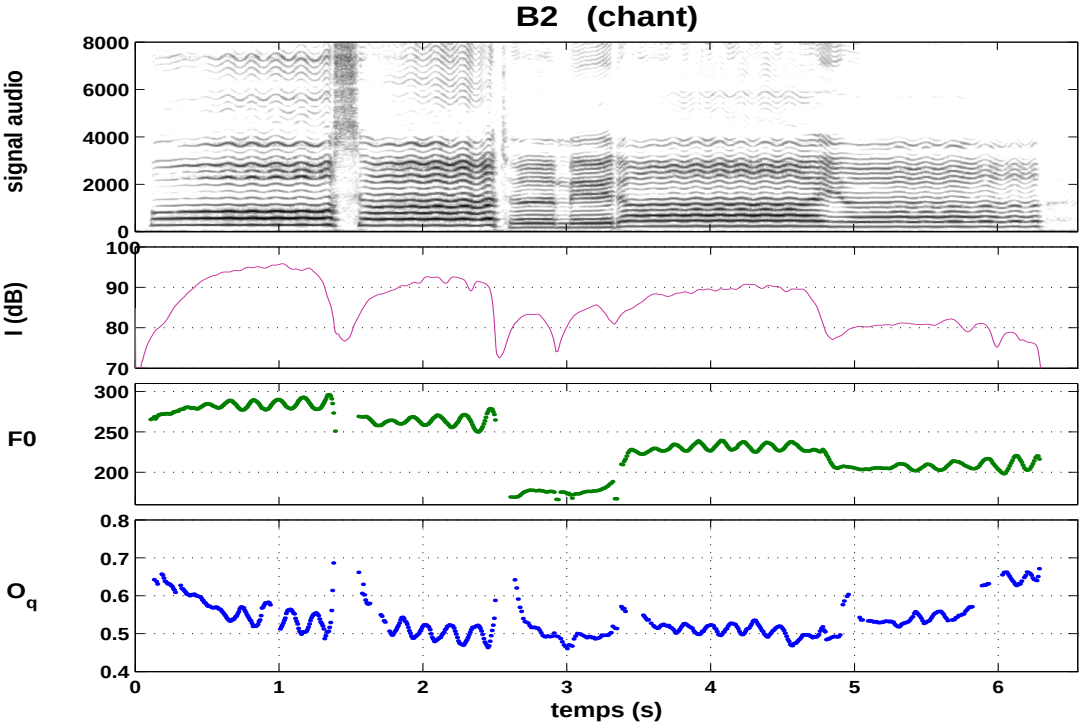


FIG. *8.17 – Mesure de I_{db}, f_0 et O_q pour la phrase chantée par le baryton B2 : « Oh sainte médaille ». (CD, piste [19])*

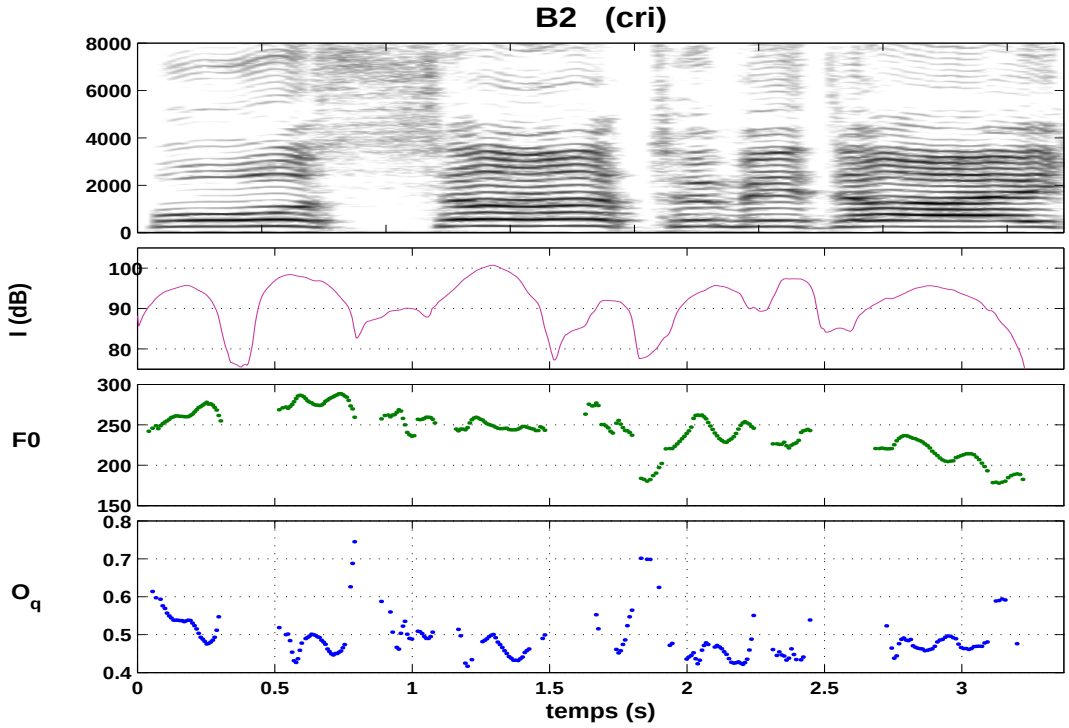


Fig. *8.18 – Mesure de I_{db}, f_0 et O_q pour la phrase criée par le baryton B2 : « Oh sainte médaille ». (CD, piste [20])*

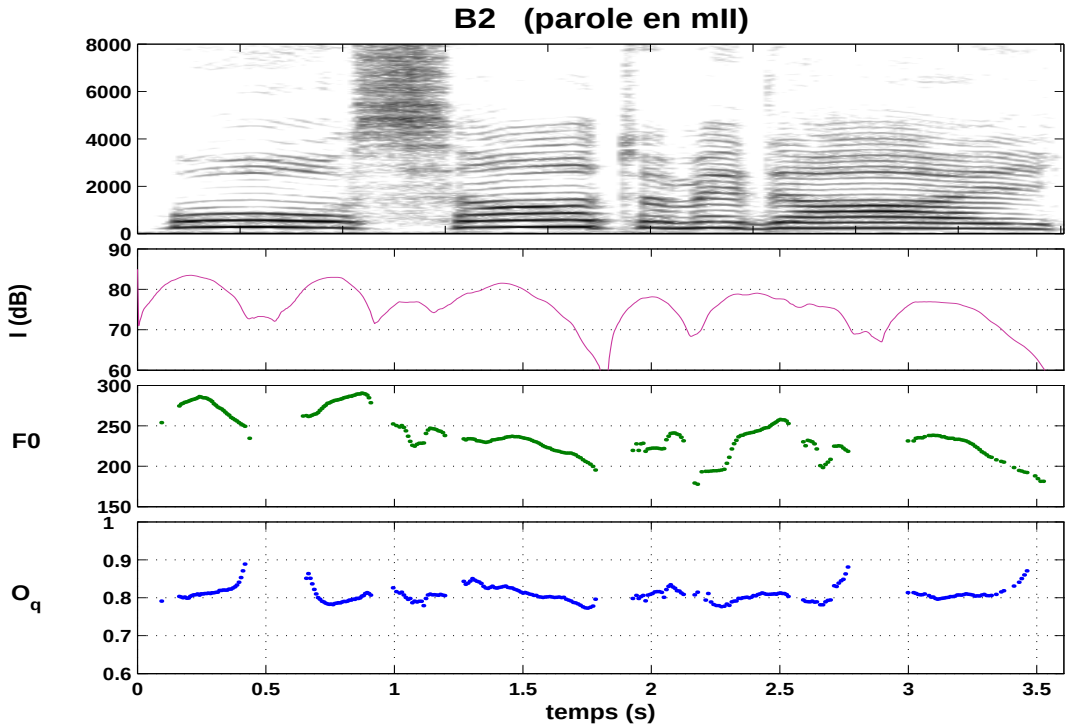


Fig. *8.19 – phrase « Oh sainte médaille » énoncée en mécanisme II par le baryton B2. (CD, piste [21])*

La plupart des valeurs de O_q sont ainsi situées entre 0.78 et 0.84, dans une plage de quotient ouvert totalement distincte de celle mesurée pour les productions en mécanisme I. La zone fréquentielle couverte dans cet exemple en mécanisme II est comparable à celle utilisée pour la production chantée. Ceci illustre le fait qu'il existe une zone fréquentielle de recouvrement, où les deux mécanismes peuvent être utilisés au choix par le chanteur. L'étendue de cette zone de recouvrement est variable selon les chanteurs. Le tableau 6.1 présenté au chapitre 6 donne des indications sur les zones de recouvrement des chanteurs ayant participé à cette étude, par comparaison entre leur étendue vocale en mécanisme I et en mécanisme II.

8.7.2 Exemple de la soprano S2

La phrase choisie par la soprano S2 est « *Celui dont la parole efface toute peine* ». La figure 8.20 présente la production vocale parlée, la figure 8.21 la production vocale chantée et la figure 8.22 la production vocale criée. De même que pour l'exemple précédent, l'échelle de quotient ouvert a été maintenue fixe pour aider à la comparaison entre les exemples, tandis que les échelles d'intensité vocale et de fréquence fondamentale sont choisies pour rendre au mieux les détails de la variation de ces paramètres. Dans cet exemple également, le chant se caractérise par une augmentation sensible de l'intensité vocale (près de 20 dB en moyenne) par rapport à la parole et une petite diminution (4 dB en moyenne) par rapport à la voix criée. Comme précédemment, la production vocale criée est effectuée avec une augmentation de la fréquence fondamentale par rapport à la production vocale parlée, qui pourrait être induite par l'augmentation sensible de l'intensité vocale entre ces deux types de production. Comme précédemment, nous remarquons qu'une différence majeure entre phrase parlée et phrase chantée est l'apparition d'un vibrato de quotient ouvert, en opposition de phase avec celui de fréquence.

Une autre différence majeure entre voix parlée et chantée porte sur les plages de variations du quotient ouvert. Celui-ci varie entre 0.58 et 0.74 en voix parlée et entre 0.72 et 0.9 en voix chantée. Cet écart très marqué s'explique par le fait que, si la soprano s'exprime en mécanisme I, elle chante par contre en utilisant le mécanisme II. Nous pouvons aussi noter, au sein d'un même mécanisme, une différence homme/femme : les valeurs de quotient ouvert sont plus élevées chez la femme que chez l'homme.

La voix criée, en mécanisme I, se caractérise par des variations de quotient ouvert plus rapides et plus fréquentes que pour les deux autres productions. Le quotient ouvert atteint également des valeurs plus faibles, allant jusqu'à $O_q = 0.4$. Comme dans le cas précédent du baryton, sa tessiture monte d'une quinte entre voix parlée et criée.

Cette chanteuse a tenté par deux fois de chanter cette phrase en utilisant le mécanisme I. Il nous a paru intéressant de présenter également ces deux exemples. Lors de la première tentative, présentée sur la figure 8.23, cette soprano, qui n'utilise pas principalement le mécanisme I, n'a pas réussi à s'y maintenir lors de la montée mélodique. A ce moment-là, elle a alors basculé dans le second mécanisme. Elle a réussi à revenir en mécanisme I à la fin de la phrase musicale, lors d'une descente mélodique. Si nous observons la courbe de quotient ouvert, nous notons très distinctement le début du chant en mécanisme I, caractérisé par des valeurs faibles de O_q (entre 0.6 et 0.7) puis un brusque saut de O_q qui accompagne la montée en fréquence. La chanteuse est alors en mécanisme II et son vibrato s'installe. Elle tente néanmoins de revenir en mécanisme I dès que la fréquence redevient suffisamment basse. Pour son second essai, représenté sur la figure 8.24, la chanteuse a donc commencé intentionnellement son chant à une fréquence fondamentale plus basse, ce qui lui permet de rester en mécanisme I tout au long de la phrase musicale. Le quotient ouvert reste alors à des valeurs proches de 0.6. Il diminue même légèrement au cours de la phrase, ce qui pourrait traduire un petit effort de la part de cette chanteuse pour rester dans ce mécanisme.

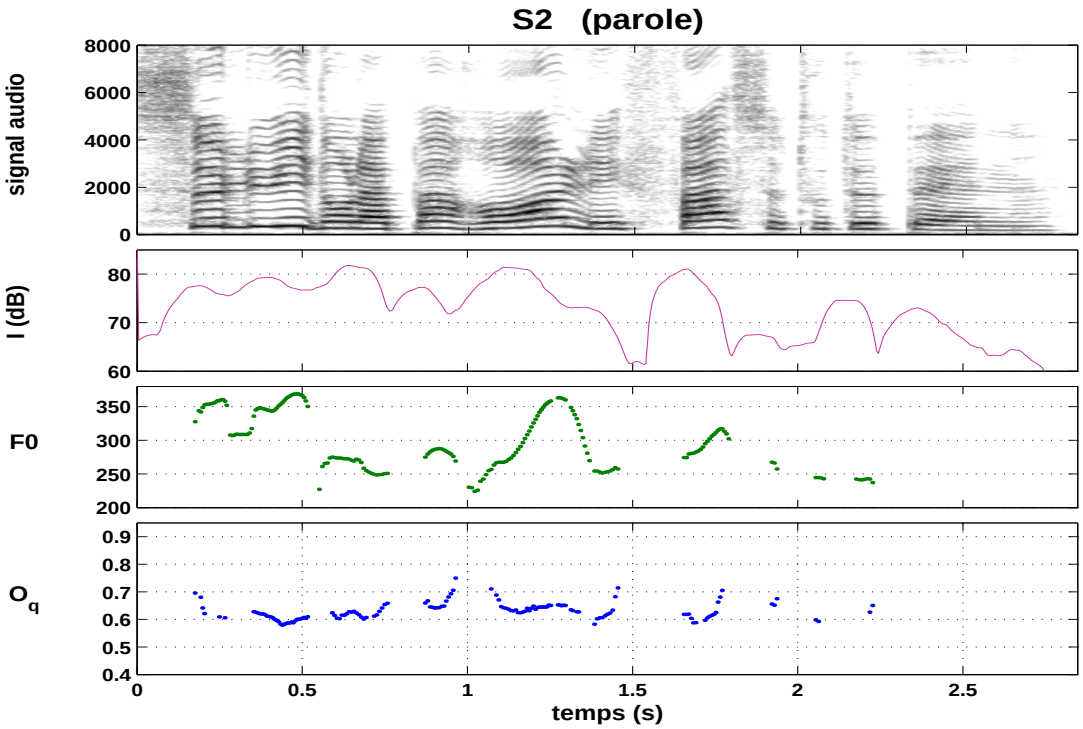


Fig. *8.20 – Mesure de I_{db}, f_0 et O_q pour la phrase énoncée par la soprano S2, vraisemblablement* **en mécanisme I** *: « Celui dont la parole efface toute peine ». (CD, piste [22])*

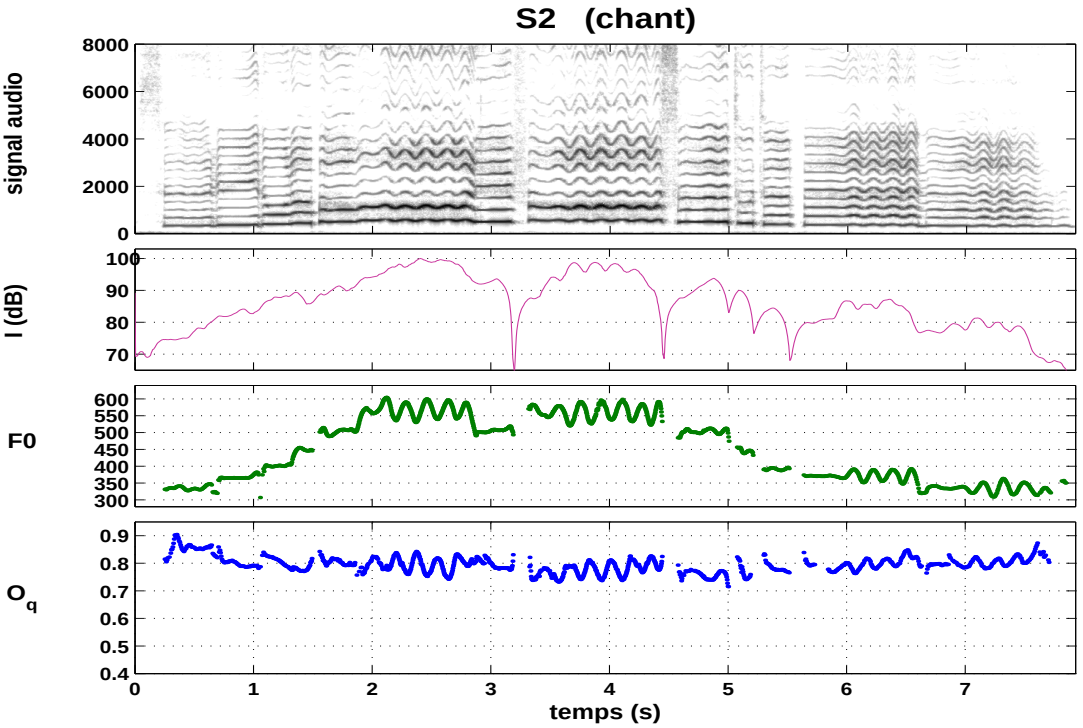


Fig. *8.21 – Mesure de I_{db}, f_0 et O_q pour la phrase chantée par la soprano S2* **en mécanisme II** *: « Celui dont la parole efface toute peine ». (CD, piste [23])*

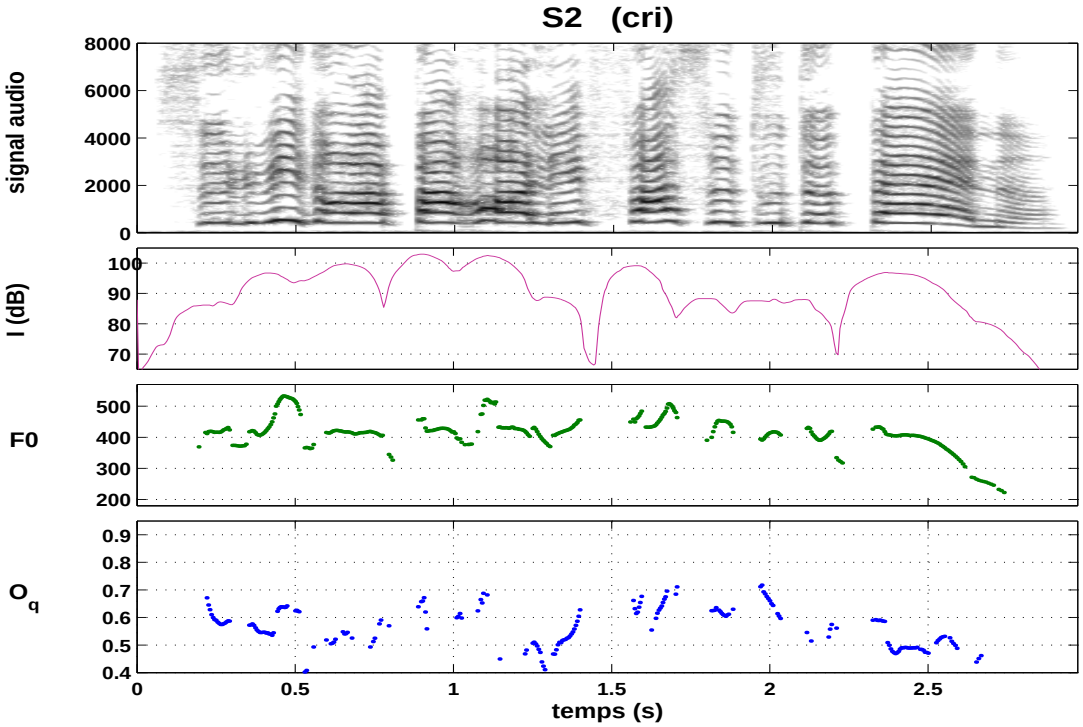


FIG. *8.22 – Mesure de I_{db}, f_0 et O_q pour la phrase criée par la soprano S2 (***mécanisme I***) : « Celui dont la parole efface toute peine ». (CD, piste [24])*

8.8 Conclusion

Si nous faisons le bilan de cette étude, nous notons tout d'abord la grande variabilité entre les sujets, un trait caractéristique déjà mis en avant par Schutte [181] dans ses travaux sur l'efficience vocale. Nous observons donc de grandes différences de comportements selon les chanteurs, au sein d'une même catégorie vocale ou entre les différentes catégories vocales. En particulier, nous n'avons enregistré que deux ténors et ceux-ci présentent des comportements très différents. Il serait donc intéressant d'enregistrer plus de chanteurs au sein d'une même catégorie vocale, pour pouvoir déterminer s'il existe ou non une tendance marquée à laquelle seuls quelques chanteurs font exception.

Cette étude montre que le quotient ouvert permet une indication du mécanisme laryngé dans lequel se situe la production vocale considérée. Le mécanisme II se caractérise par des valeurs de quotient ouvert plus élevées qu'en mécanisme I, généralement situées entre 0.5 et 0.95. En mécanisme I, la marge de variation du quotient ouvert est plus étendue et nous mesurons des valeurs de O_q entre 0.3 et 0.8. La seule mesure de O_q ne va donc pas permettre de déterminer de façon sûre la nature du mécanisme laryngé, puisque le quotient ouvert peut prendre des valeurs élevées en mécanisme I, comparables à celles que l'on mesurerait en mécanisme II. Nous avons vu qu'il existe de nombreux cas où le doute est permis, par exemple dans le cas de sons chantés à faible intensité. La valeur de quotient ouvert mesurée donne donc une indication et non une certitude sur le mécanisme laryngé utilisé. Une écoute attentive des exemples sonores peut permettre, dans certains cas, de résoudre cette indétermination mais notre perception peut être habilement trompée par le chanteur. Néanmoins, la perception auditive et la mesure de quotient ouvert se complètent et permettent, dans la majorité des cas, de déterminer la nature du mécanisme laryngé utilisé.

Cette étude donne également des indications sur la façon dont un chanteur va augmenter l'intensité vocale. Cette augmentation est généralement le fait d'une augmentation de la pression sous-glottique,

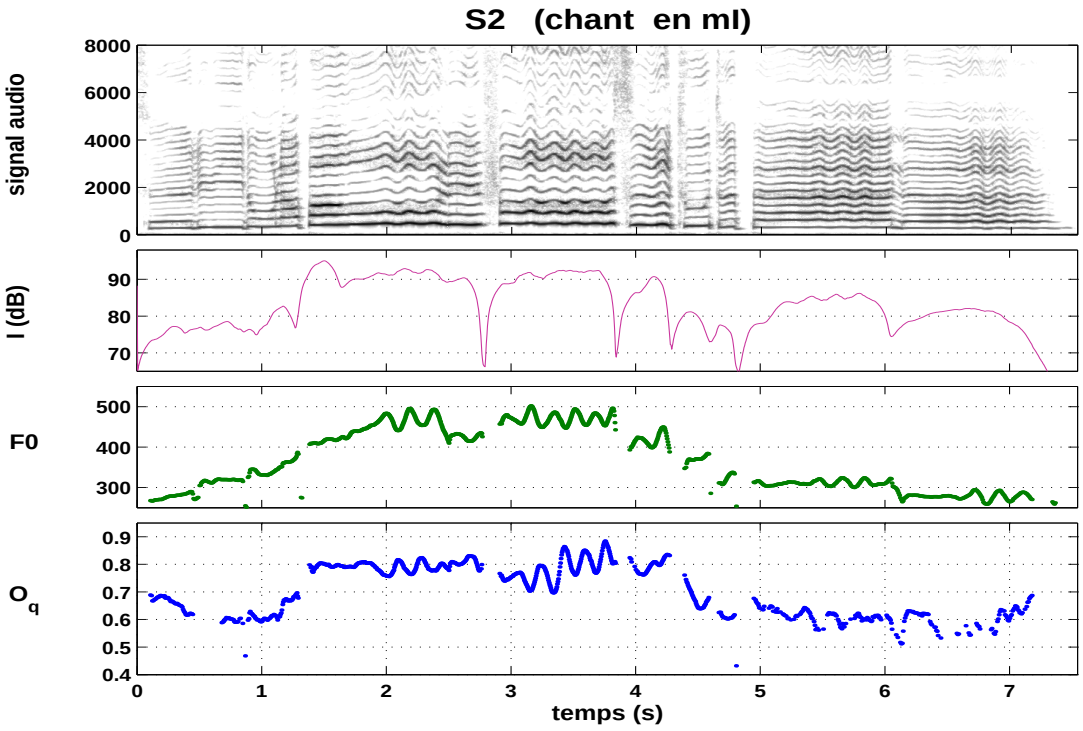


Fig. *8.23 – Phrase « Celui dont la parole efface toute peine » chantée par la soprano S2 : tentative de chanter la phrase* **en mécanisme I** *mais passage* **en mécanisme II** *entre* $t = 1.2$ *et* 4.5 s. *(CD, piste [25])*

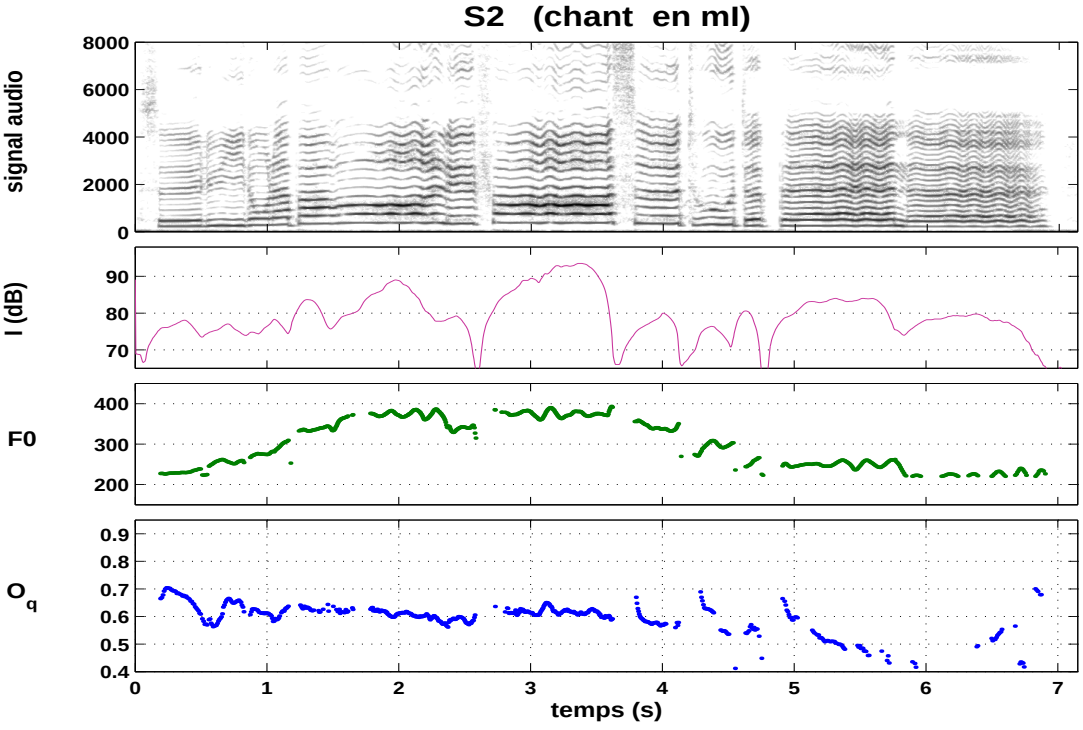


Fig. *8.24 – Phrase « Celui dont la parole efface toute peine » chantée par la soprano S2* **en mécanisme I**. *(CD, piste [26])*

indépendamment du mécanisme laryngé considéré. Qu'en est-il de l'action musculaire qui l'accompagne ? En mécanisme I, nous avons observé une forte corrélation entre le quotient ouvert et l'intensité vocale. Comme le vocalis est principalement en action dans ce mécanisme, nous pouvons avancer l'hypothèse que le chanteur va agir de façon plus ou moins consciente sur la contraction de ce muscle pour augmenter l'intensité du son émis. Cette augmentation de la contraction se reflètera alors directement sur les mesures de quotient ouvert : plus le vocalis sera contracté, plus le quotient ouvert sera faible. En mécanisme II, nous n'observons plus de corrélation marquée entre le quotient ouvert et l'intensité vocale, ce qui pourrait alors s'expliquer par une diminution d'activité du vocalis.

Nous avons observé en mécanisme II une forte corrélation entre le quotient ouvert et la fréquence fondamentale, qui ne se retrouve pas en mécanisme I. Cette corrélation est particulièrement marquée chez les 3 contre-ténors. Elle pourrait s'expliquer par le rôle de la pression sous-glottique sur le contrôle de la fréquence fondamentale, plus marqué en mécanisme II qu'en mécanisme I. La diminution du quotient ouvert serait liée à une augmentation de la pression sous-glottique. Cela permettrait alors également d'expliquer les différences de quotient ouvert observées entre la voyelle [a] et les voyelles [e] et [u].

La grande variabilité observée sur les résultats nous amène à nous demander dans quelle mesure une différence de quotient ouvert est-elle significative du point de vue de la perception auditive. De même, jusqu'à quelle limite le saut de quotient ouvert qui accompagne un changement de mécanisme laryngé peut-il être perçu ? Nous allons tenter d'apporter des réponses à ces questions dans le chapitre suivant.

En Résumé

Dans ce chapitre, nous avons présenté les résultats des mesures de quotient ouvert effectuées sur une base de donnée de 18 chanteurs. En particulier, nous nous sommes intéressés aux corrélations possibles entre quotient ouvert et intensité vocale ou fréquence fondamentale. Nous avons également étudié de façon qualitative l'effet possible des voyelles demandées sur les valeurs de quotient ouvert mesurées et illustré la différence entre production vocale parlée, chantée et criée sur les exemples d'un baryton et d'une soprano. Les points principaux qui ressortent de cette étude sont l'importance des mécanismes laryngés pour l'étude du quotient ouvert et une grande variabilité entre les chanteurs, au sein d'une même catégorie vocale ou entre catégories vocales. Néanmoins, des tendances générales se dégagent au sein d'un même mécanisme laryngé.

Importance des mécanismes laryngés

Nous avons montré sur de nombreux exemples de voix parlée et chantée, hors contexte ou en contexte musical, l'importance des mécanismes laryngés sur les valeurs de quotient ouvert. De façon générale, le quotient ouvert a des valeurs plus élevées en mécanisme II qu'en mécanisme I. Les transitions entre mécanismes peuvent alors se caractériser, lors d'un passage d'un mécanisme à l'autre, par des sauts marqués de quotient ouvert. L'augmentation de quotient ouvert est de l'ordre de 0.1 à 0.2 entre mécanisme I et mécanisme II. Ces sauts peuvent accompagner un saut en fréquence, caractéristique d'un passage non-contrôlé entre mécanismes, mais ils se retrouvent également dans le cas où l'habileté du chanteur lui permet de masquer perceptivement ce passage.
Nous avons montré également qu'il existe des exceptions à cette tendance générale. Certains chanteurs montrent des valeurs de quotient ouvert élevées alors qu'ils sont vraisemblablement en mécanisme I. Il devient alors difficile de déterminer avec assurance dans quel mécanisme a lieu leur production vocale.
Devant l'importance du mécanisme laryngé pour l'étude du quotient ouvert, il nous apparaît nécessaire de préciser dans toute étude portant sur le quotient ouvert le mécanisme dans lequel a lieu la production vocale. Même en parole, il est possible que le locuteur utilise l'un ou l'autre des deux mécanismes.

Production vocale en mécanisme I

Les caractéristiques générales du quotient ouvert en mécanisme I sont :

- des valeurs de O_q entre 0.3 et 0.8 (barytons, ténors, contre-ténors),
- une diminution de O_q quand l'intensité vocale augmente,
- une absence de lien significatif entre O_q et la fréquence fondamentale,
- à intensité vocale et fréquence fondamentale constantes, des valeurs de O_q légèrement supérieures dans le cas de la voyelle [a] comparativement aux voyelles [e] et [u].

Production vocale en mécanisme II

Les caractéristiques générales du quotient ouvert en mécanisme II sont :

- des valeurs de O_q entre 0.5 et 0.95 (contre-ténors, mezzo-sopranos, sopranos),
- une absence de lien significatif entre O_q et l'intensité vocale,
- une diminution de O_q quand la fréquence fondamentale augmente,
- pas d'effet significatif de la voyelle.

Les variations du quotient ouvert en fonction de l'intensité pourraient dépendre de l'activité du vocalis, importante en mécanisme I et réduite en mécanisme II. Le lien entre O_q et f_0 en mécanisme II, ainsi que la différence de quotient ouvert entre la voyelle [a] et les voyelles [e] et [u] pourraient dépendre des variations de la pression sous-glottique.

Troisième partie

Etude perceptive des paramètres de source glottique

Enigme pour la pensée qui vient toujours trop tard pour la saisir et qu'en outre elle menace, la voix est énigme pour le corps qui ne sait s'il la crée, la reçoit ou s'en délivre.

(Marie-France Castarède, « La voix et ses sortilèges », 1987)

Sommaire de la partie

Dans la partie I, les modèles d'ODG ont été examinés pour déterminer les corrélats spectraux des paramètres de source glottique A_v, E, O_q, α_m et Q_a. Des mesures de quotient ouvert ont été effectuées à l'aide des signaux DEGG, dans la partie II. Comment les variations de ces paramètres de source glottique sont-elles perçues ?

Une expérience d'appréciation globale de la qualité, proche de ce que font les chanteurs ou les professeurs de chant, est difficile à mettre en oeuvre. Nous limiterons donc notre exploration des changements de qualité à l'évaluation de seuils de perception des paramètres de source glottique et seules les variations de quotient ouvert ou de coefficient d'asymétrie seront étudiées.

Cette évaluation perceptive sera l'objet du chapitre 9. Elle s'appuie sur une étude réalisée en 1998 par Scherer & al [180], qu'elle se propose de compléter et d'étendre pour déterminer le rôle du contexte (la voyelle, la hauteur, le vibrato ou le paramètre d'amplitude) dans la détermination de ces seuils.

La liste détaillée des exemples sonores qui ont illustrés cette thèse sera fournie à la fin de cette partie.

Chapitre 9

Seuils de perception des paramètres de source glottique

9.1 Problématique

Dans le chapitre 8, nous avons mesuré le quotient ouvert par électroglottographie pour des productions vocales parlées et chantées. Un des résultats de cette étude porte sur la grande variabilité observée dans les valeurs obtenues. Se pose alors la question de savoir dans quelle mesure des différences de quotient ouvert sont significatives du point de vue de la perception. Peut-on déterminer des seuils, fonctions des paramètres de source glottique, en-dessous desquels les différences entre deux sons voisés ne seront pas perceptibles ? La connaissance de tels seuils permettrait également d'apporter des éléments de comparaison entre les méthodes de mesure de quotient ouvert par électroglottographie évoquées au chapitre 7. En effet, si les différences de mesure entre ces méthodes ne sont pas significativement perceptibles, ces méthodes peuvent alors être considérées comme équivalentes. Un autre résultat obtenu dans le chapitre 8 porte sur la différence marquée de quotient ouvert selon le mécanisme laryngé dans lequel a lieu la production vocale. Cette différence a-t-elle un sens du point de vue perceptif ? Là encore, la connaissance de tels seuils permettrait d'apporter une réponse à cette question.

9.1.1 Le seuil différentiel

La mesure du seuil différentiel nous a paru la mesure la plus adaptée à cette étude, car cette expérience psychophysique renseigne sur les capacités de discrimination auditive d'un sujet dans une situation donnée. Le seuil différentiel [1] est défini comme la limite des différences perçues et des différences non-perçues entre un stimulus dit standard (S), gardé fixe au cours de l'expérience, et un stimulus dit variable (V) (Bonnet [22]). Sa détermination est généralement unidimensionnelle, car il est aisé de faire varier la valeur d'un paramètre choisi (comme, par exemple, l'intensité, la fréquence fondamentale ou la durée) tout en gardant les autres paramètres constants. Dans notre cas, la situation est plus complexe, en raison de l'interdépendance des paramètres de source glottique que nous avons soulignée en partie I. Il nous faudra donc garder ce point à l'esprit tout au long de cette étude perceptive.

Dans la littérature, nous n'avons trouvé qu'une unique étude, menée par Scherer & al [180] (1998), qui se propose de mesurer des seuils différentiels en fonction des paramètres de source glottique. Cette étude constitue donc le point de départ de notre exploration et nous allons à présent la détailler.

[1] Le seuil différentiel se traduit en anglais par le terme *Just Noticeable Difference* ou JND

9.1.2 Expérience de Scherer & al [180]

Scherer & al [180] effectuent des mesures de seuil différentiel pour le quotient ouvert O_q et le quotient de vitesse S_q dans le cas d'un signal de source ou d'une voyelle synthétique ([a]), à partir d'une étude perceptive menée sur des sujets entraînés, 6 hommes et 5 femmes, tous professeurs de chant. Les valeurs de référence choisies pour cette étude sont : $f_0 = 130.39\ Hz$, $O_q = 0.6$ et $S_q = 2$ (équivalent à $\alpha_m = 2/3$). Le modèle d'ODG utilisé pour la synthèse [2] est défini à l'aide des 4 paramètres de source glottique (T_0, A_v, O_q, S_q) et la nature de la phase de retour (abrupte ou molle) dépend de O_q et S_q . La méthode expérimentale utilisée pour la détermination des seuils est une méthode adaptative à choix forcé « 1-up 2-down » [3] .

Les seuils différentiels obtenus par Scherer sont les suivants (moyennes sur les 11 sujets) :

1. **à partir d'un signal de source glottique**, $\Delta O_q = 0.0264$ pour $O_q = 0.6$ et $\Delta S_q = 0.154$ pour $S_q = 2$ (ce qui correspond à $\Delta\alpha_m = 0.0163$ pour $\alpha_m = \frac{2}{3}$) ;
2. **à partir d'un signal voisé**, $\Delta O_q = 0.0344$ pour $O_q = 0.6$ et $\Delta S_q = 0.319$ pour $S_q = 2$ (ce qui correspond à $\Delta\alpha_m = 0.0328$ pour $\alpha_m = \frac{2}{3}$).

La figure 9.1 présente les seuils obtenus par chaque sujet, pour chacune des quatre tâches de discrimination (signal de source glottique avec $O_q = 0.6$ ou $S_q = 2$ et signal voisé avec $O_q = 0.6$ ou $S_q = 2$). Pour pouvoir effectuer facilement des comparaisons avec l'étude qui va suivre, nous avons préféré présenter les résultats concernant le quotient de vitesse en terme de coefficient d'asymétrie et nous avons donc converti S_q en α_m par la relation $\alpha_m = \frac{S_q}{1+S_q}$.

Une étude statistique montre qu'il n'a pas été observé de différences sur le seuil selon le sexe du sujet. Des différences significatives sur le seuil ont été obtenues entre la tâche portant sur le signal de source glottique et celle portant sur le signal voisé dans le cas du quotient de vitesse, mais pas dans le cas du quotient ouvert. Enfin, la variance sur les mesures du seuil est plus marquée pour le signal voisé que pour le signal de source glottique.

Les résultats obtenus par Scherer & al [180], avec des sujets entraînés et dans des conditions rigoureuses, sont toutefois limités. En effet :

- Ces résultats ne sont valables que pour une seule valeur des paramètres de source glottique ($O_q = 0.6$ et $S_q = 2$), une voyelle ([a]) et une fréquence fondamentale ($f_0 = 130.39$Hz).
- Les signaux synthétisés pour l'expérience ne sont pas du tout représentatifs d'une situation vocale réelle, la fréquence fondamentale étant gardée constante au cours du son.

Nous nous proposons donc d'explorer les seuils différentiels de quotient ouvert et de coefficient d'asymétrie pour différentes valeurs de ces paramètres, en nous plaçant dans une situation proche de la réalité, afin de tenter de répondre aux questions suivantes : ces seuils dépendent-ils de la valeur cible du quotient ouvert (ou du coefficient d'asymétrie) choisie ? De la hauteur ? De la voyelle ? Le paradigme de test, reflet de celui choisi par Scherer, sera l'objet de la partie 9.2. Le protocole du test perceptif et la synthèse des stimuli seront décrits en partie 9.3. Les résultats seront détaillés dans la partie 9.4.

Une partie du travail a été effectuée par Gunilla Sundin (auditrice libre du DEA ATIAM [4]), dans le cadre d'un stage de 3 mois.

---

[2] La synthèse a été réalisée à l'aide d'un programme de synthèse « SPEAK », d'après Titze & al [208]

[3] Nous reviendrons en détail sur cette méthode dans la partie 9.2. En effet, ce type de méthode nous a paru très approprié pour l'étude de seuils différentiels et dans un souci de continuité par rapport aux travaux menés par Scherer, nous l'avons appliquée telle qu'elle est présentée dans cet article.

[4] Diplôme d'Etudes Approfondies en Acoustique, Traitement de signal et Informatique Appliqués à la Musique

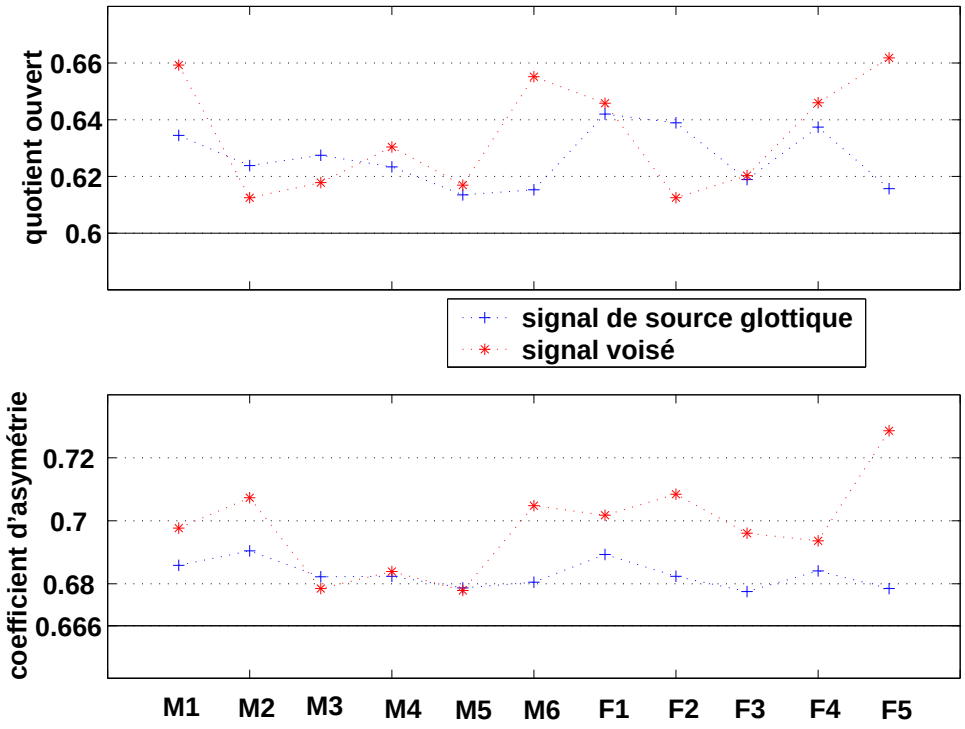


FIG. *9.1 – Présentation des résultats obtenus par Scherer* [180] *pour 6 sujets masculins (M1 à 6) et 5 sujets féminins (F1 à 5), tous entraînés. Dans le cas du coefficient d'asymétrie, les résultats présentés sont calculés à partir de ceux obtenus pour le quotient de vitesse* S_q*, par la relation d'équivalence* $\alpha_m = \frac{S_q}{1+S_q}$*. Les valeurs choisies pour le stimulus standard sont* $0_q = 0.6$ *et* $\alpha_m = 2/3$*.*

9.2 Méthode adaptative à choix forcé

9.2.1 Quelle méthode pour estimer un seuil différentiel ?

Dans cette étude, une réflexion a été menée pour choisir la méthode de mesure psychophysique qui nous paraissait la plus adaptée. En effet, de nombreuses méthodes permettent la détermination des seuils différentiels (Bonnet [22], p.37) :

* **la méthode d'ajustement :** le sujet est amené à ajuster lui-même le seuil à la valeur qui lui semble appropriée ;
* **la méthode des limites :** sur présentation d'une série ordonnée de stimuli (croissante ou décroissante), le sujet signale l'instant où le seuil lui paraît être atteint ;
* **la méthode constante :** les stimuli sont présentés au sujet de façon aléatoire et celui-ci répond soit à un paradigme « Oui/Non » (différences perçues/non-perçues), soit à un paradigme de choix forcé (sélection forcée d'un stimulus parmi ceux proposés) ;
* **les procédures adaptatives :** elle sont dérivées de la méthode des limites. La valeur du stimulus V est modifiée en fonction des réponses du sujet aux essais précédents. Comme pour la méthode constante, les procédures adaptatives peuvent s'appliquer avec un paradigme « Oui/Non » ou de choix forcé.

La précision de la mesure (réduction de la variance des estimations du seuil) et son exactitude (estimation valide, qui ne soit pas faussée par des variations sur les critères de décision des sujets) sont des caractéristiques souhaitées. La méthode d'ajustement et la méthode des limites sont des méthodes

précises, généralement rapides, mais peu exactes. La méthode constante est une méthode exacte mais peu précise en raison du compromis à trouver entre un pas fin et un nombre important d'essais. Elle est facilement implémentable, mais elle nécessite une hypothèse sur la valeur du seuil et sa stabilité en cours d'expérience.

Les procédures adaptatives proposent un bon compromis entre précision et exactitude. Elles permettent de cibler rapidement le seuil du sujet sans en connaître la valeur à priori et d'obtenir un maximum d'informations concernant ce seuil en un minimum d'essais. Néanmoins, elles peuvent entraîner une lassitude chez le sujet, car les différences perçues sont plus minimes à mesure que l'on se rapproche du seuil. Contrairement à la méthode constante, elles ne supposent pas une stabilité du sujet au cours de l'expérience et elles peuvent même rendre compte d'une dérive du seuil en cours de session (due à l'inattention, la fatigue ou, au contraire, un effet d'apprentissage). Elles sont plus complexes à implémenter que les autres méthodes, puisqu'elles tiennent compte des réponses du sujet.

Le principe d'une procédure adaptative a été retenue en raison des avantages majeurs qu'elle propose. Ce choix permet également de placer cette étude dans la continuité directe des travaux de Scherer & al [180]. Dans le cadre des méthodes adaptatives, de nombreuses variantes sont possibles (selon le paradigme utilisé et la règle de progression). En ce sens, les choix effectués par Scherer & al nous ont paru très adaptés à cette étude et nous les avons également adoptés ici. Nous allons donc maintenant décrire en détail le principe et l'application de la méthode retenue.

9.2.2 Description de la méthode adaptative à choix forcé 1-up 2-down

Le principe de la méthode adaptative à choix forcé mise en place pour cette étude est de partir d'un seuil très élevé (différence nettement perceptible entre les sons S (stimulus standard) et V (stimulus variable)), de diminuer progressivement ce seuil en cas de détection du sujet ou de l'augmenter en cas de non-détection, jusqu'à converger vers une valeur qui représente la limite entre les différences perçues et non-perçues. Le nombre d'essais (ou écoutes) réalisés par le sujet dépend de la rapidité et de la stabilité de la convergence lors de la session. En particulier, il est fonction du nombre maximal d'inversions et du pas (valeur initiale, valeur minimale) choisis par l'expérimentateur. Une inversion correspond à un changement de sens de la progression (diminution ou augmentation du seuil).

Afin d'optimiser au mieux la procédure, plusieurs paramètres doivent être judicieusement choisis (Bonnet [22], p. 48) :

* **le niveau de départ (stimulus variable initial)**. Celui-ci doit être suffisamment élevé par rapport au niveau du seuil différentiel que l'on cherche à approcher. Néanmoins, si ce niveau est placé trop haut, cela va alors rallonger d'autant la phase de convergence et donc le nombre d'essais nécessaire.

* **le pas.** Il ne doit être ni trop élevé (manque de précision), ni trop faible (augmentation importante du nombre d'essais sans apport significatif sur la précision). Il peut varier au cours du test. Dans cette étude, nous avons choisi de diviser le pas par deux à chaque fois que le nombre d'inversions prend une valeur impaire, jusqu'à se stabiliser à une valeur limite qui dépend de la précision que nous désirons obtenir pour la mesure.

* **la règle d'inversion de la progression.** L'inversion dépend des réponses fournies par le sujet. Une réponse sera dite « correcte » si le sujet détecte le stimulus variable et « incorrecte » s'il se trompe (non-détection). La règle la plus simple (« 1-up 1-down ») consiste à diminuer le seuil après chaque réponse correcte du sujet et à l'augmenter après chaque réponse incorrecte. La convergence a lieu pour une probabilité de réponses correctes de 0.5 (50% de réponses correctes et 50% de réponses incorrectes). La règle choisie dans cette étude, « 1-up 2-down », améliore l'exactitude de la mesure, en ramenant la convergence à une probabilité de réponses correctes de 0.707 (Levitt [143]). Elle consiste à diminuer le seuil après deux réponses correctes successives et à l'augmenter après

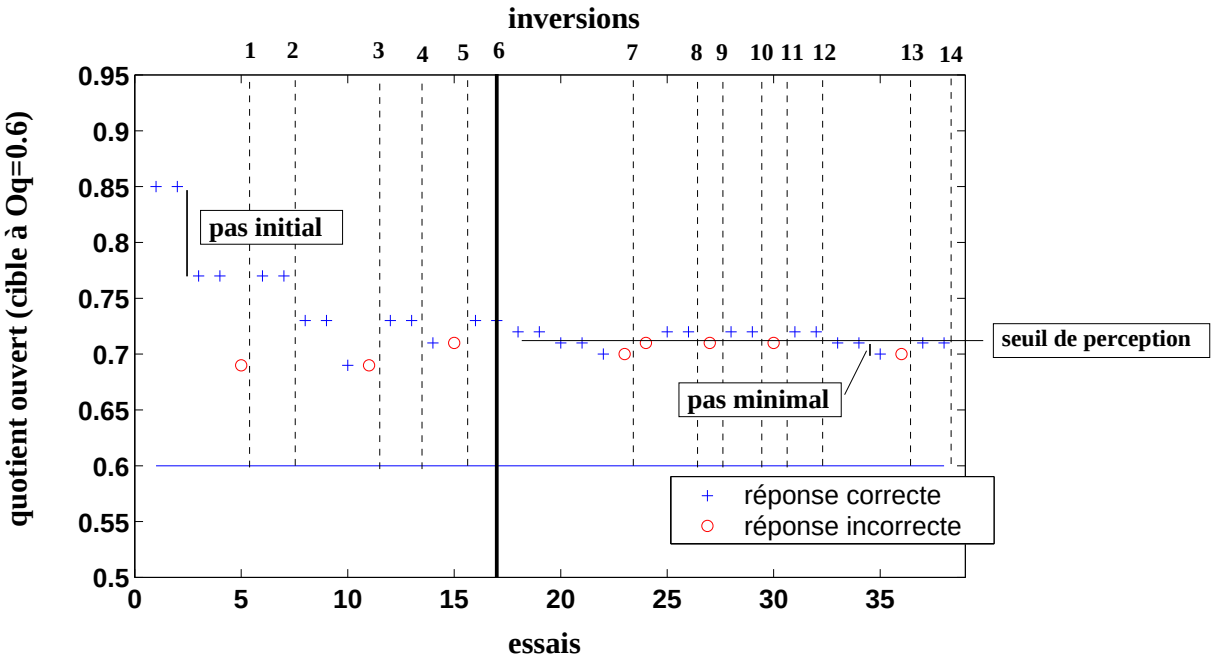


Fig. *9.2 – Exemple du déroulement d'une session : le sujet a effectué* 38 *écoutes (essais) et le test a pris fin au bout de la* 14^{ieme} *inversion. Le pas initial* (0.08) *et le pas minimal* (0.01) *sont également représentés. Le pas est divisé par* 2 *après chaque inversion paire* (2, 4, 6) *jusqu'à atteindre sa valeur minimale. Le seuil différentiel pour cette session est déterminé comme la moyenne des réponses du sujet lors des* 8 *dernières inversions (seule la valeur du seuil au moment de l'inversion est prise en compte).*

une réponse incorrecte.

* **la règle d'arrêt.** La session prend fin au bout d'un nombre donné d'inversions. La valeur habituellement choisie est de 10 inversions. Lors du calcul du seuil, les 2 premières inversions ne sont alors pas prises en compte.

A la fin d'une session, le seuil est calculé à partir des inversions retenues, par une moyenne arithmétique ou géométrique entre les valeurs à chaque inversion.

La figure 9.2 illustre ces paramètres sur un exemple concret. La session consiste à mesurer le seuil différentiel par rapport au quotient ouvert ($O_q = 0.6$). Le premier stimulus variable proposé au sujet est synthétisé en prenant $O_q = 0.85$, ce qui revient à un seuil différentiel initial de $\Delta O_q = 0.25$. Ce seuil est diminué par un pas, dont la valeur initiale de 0.08 est divisée par 2 après chaque inversion paire, jusqu'à atteindre sa valeur minimale de 0.01. La session prend fin à la 14^{ieme} inversion, ce qui correspond à 38 écoutes successives dans le cas de ce sujet. A partir de ces résultats, le seuil différentiel est calculé en moyennant les valeurs des seuils obtenus au moment des 8 dernières inversions.

9.2.3 Implémentation de la méthode

La méthode a été implémentée en Matlab comme un module de l'environnement de tests perceptifs *LISE* développé par Rioux [167]. Ce module, appelé « Adaptative », s'appuie sur les structures de gestion des données et des résultats de *LISE*. Il possède une interface qui lui est propre, présentée sur la figure 9.3 et fait appel à un programme de présentation des sons et de gestion des réponses du sujet qui est présenté schématiquement sur la figure 9.4.

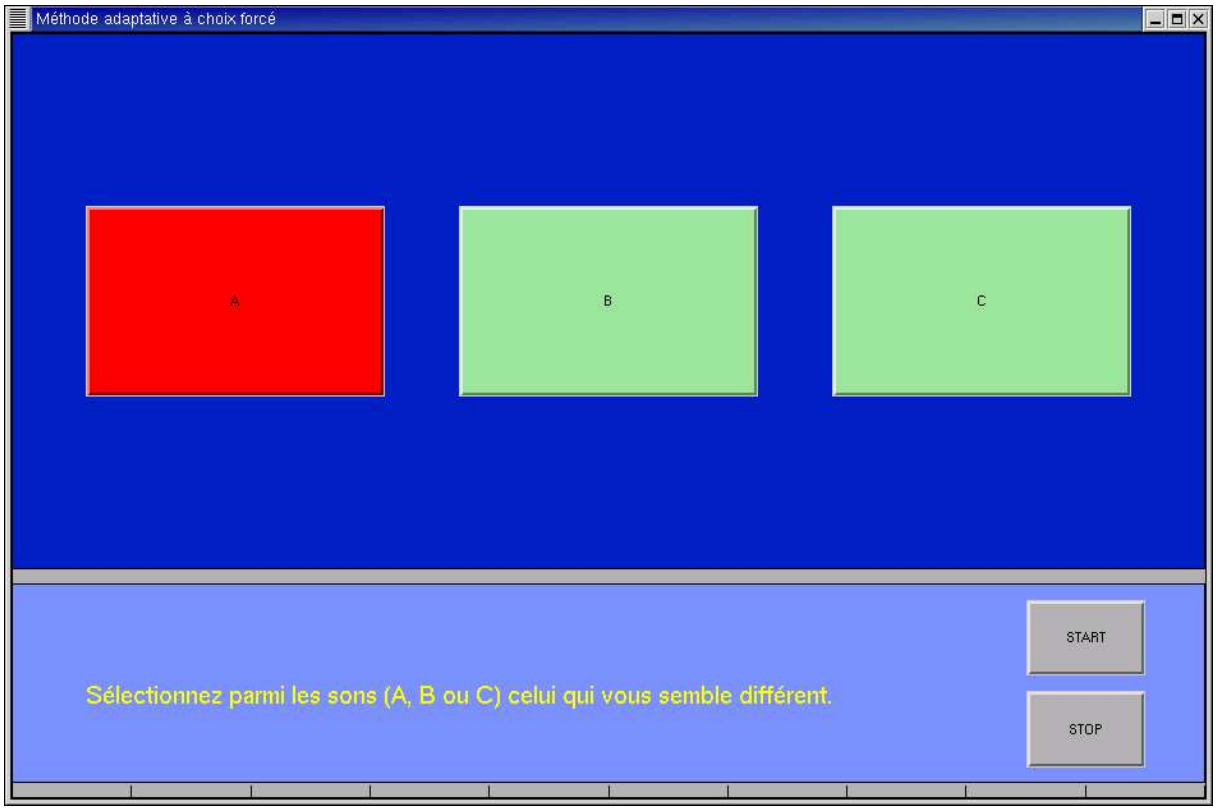


FIG. *9.3 – Interface du module « Adaptative » développé sous Matlab. Sur cet exemple, le son associé à la case A est joué, ce qui se traduit par une coloration rouge de la case correspondante. La consigne demandée au sujet est : « Sélectionnez parmi les sons (A, B ou C) celui qui vous semble différent ».*

Présentation des stimuli et sélection

Une session correspond à la détermination d'un seuil différentiel dans une situation donnée. Elle est composée d'un nombre d'essais compris en moyenne entre 30 et 90. Ce nombre varie avec la rapidité de convergence et la stabilité des réponses du sujet. Lors d'un essai, 3 sons sont présentés successivement au sujet, séparés par une pause de 0.5 s. Ces 3 sons correspondent à une répartition aléatoire du stimulus standard S (constant au cours de la session) répété deux fois et d'un stimulus variable V (fonction des réponses du sujet). Les cases A, B ou C, de couleur verte par défaut, se colorent successivement en rouge quand le son qui leur est associé est joué, de manière à donner un retour visuel au sujet. A la fin de la série, la main est donnée au sujet, qui peut aller sélectionner, à l'aide de la souris, la case correspondant au son qui lui a semblé différent des deux autres. La case sélectionnée par le sujet se colore en rouge, puis, après une pause de 0.5 s, un nouvel essai est lancé.

Programme adaptatif de présentation des stimuli

Un schéma du programme qui gère la sélection des stimuli de façon adaptative en fonction des réponses des sujets est présenté sur la figure 9.4. Le programme fait appel à une base de sons synthétiques, dont l'élaboration sera détaillée dans la partie 9.3.1. Les données fournies par l'expérimentateur au programme sont : le code du chanteur, la voyelle choisie, les paramètres de source glottique associés à la synthèse du stimulus standard (en particulier la fréquence fondamentale, le quotient ouvert et le coefficient d'asymétrie), le paramètre du stimulus variable initial (valeur ($Param$), type (O_q ou α_m)), ainsi que des données sur le pas (pas initial, pas limite) et le nombre maximal d'inversions.

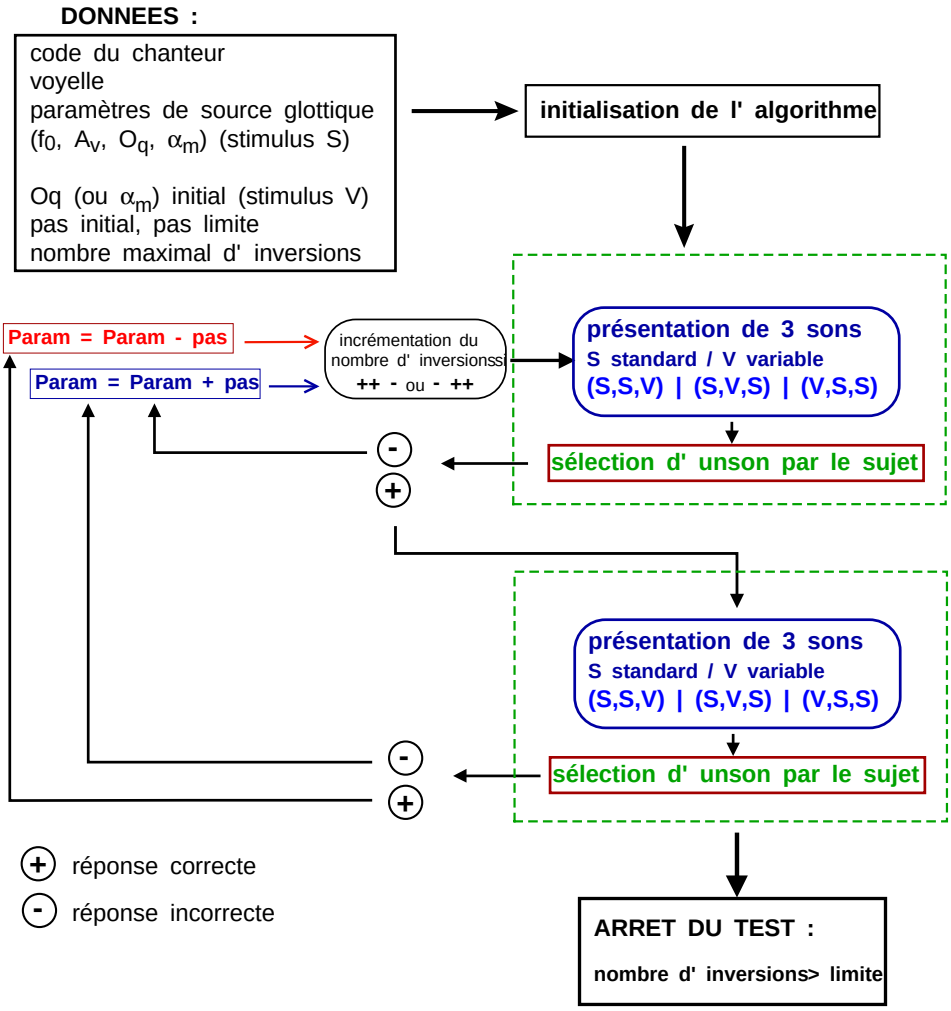


FIG. *9.4 – Schéma de l'algorithme mis en place pour la méthode adaptative à choix forcé 1-up 2-down.*

Lors d'une phase d'initialisation, le programme détermine les caractéristiques du stimulus standard et du stimulus variable initial en fonction des données, en déduit le signe du pas (si $(Param_V - Param_S) \geq 0$, le signe du pas est positif, sinon négatif). Le nombre d'inversions et le nombre d'essais sont initialisés à 1.

Le programme boucle sur deux écoutes successives où le paramètre $Param$ garde la même valeur. Lors d'une écoute, le programme sélectionne de façon aléatoire un triplet parmi les 3 possibles ((S,S,V) ou (S,V,S) ou (V,S,S)) et le sujet choisit le son qu'il pense correspondre à « V ». Le programme détermine si la réponse du sujet est correcte ($\oplus$) ou incorrecte ($\ominus$) et applique les règles suivantes :

- dans le cas d'une réponse incorrecte dès la première écoute, le paramètre est modifié par $Param = Param + pas$ (on s'éloigne de la cible) et il n'y a pas de seconde écoute ;
- dans le cas d'une réponse correcte suivie d'une réponse incorrecte, le paramètre est modifié par $Param = Param + pas$ (on s'éloigne de la cible) ;
- dans le cas de deux réponses correctes succesives, le paramètre est modifié par $Param = Param - Pas$ (on se rapproche de la cible).

Une variable, incrémentée à chaque écoute, comptabilise le nombre d'essais. Le nombre d'inversions est incrémenté quand 3 réponses successives sont de la forme $++-$ (fin d'une phase de rapprochement)

ou $-++$ (fin d'une phase d'éloignement). La valeur du pas est modifiée en fonction de la parité (changement à chaque valeur impaire) du nombre d'inversions, jusqu'à se stabiliser à la valeur limite fixée par l'expérimentateur.

Le test prend fin lorsque le nombre d'inversions atteint le nombre limite fixé par l'expérimentateur. Ce nombre d'inversions limite est toujours un nombre pair, afin que l'arrêt ait lieu à la fin d'une phase d'éloignement. Les résultats (données initiales, évolution du paramètre $Param$, réponses du sujet à chaque essai) sont alors sauvegardés sous deux formats : écriture dans un fichier texte et dans un fichier de données Matlab.

La position du stimulus à chaque test (case A, B ou C) n'a pas été enregistrée par le programme. A postériori, cet oubli est une erreur, car il aurait été très intéressant de comparer les réponses obtenues en fonction de la place du stimulus, pour déterminer si cette place pouvait avoir une importance dans la facilité (ou difficulté) à entendre une différence. Dans le cadre de la Théorie de la Détection du Signal, la position de présentation du stimulus variable ne devrait pas introduire de biais conséquent sur le choix du sujet (Green & Swets [90], p. 408). Cependant, dans une étude portant sur l'effet de la position du stimulus pour différents types de tests à choix forcé sur 3 intervalles (présentations successives de 3 stimuli), Johnson & al [120] ont mis en évidence des différences de performance sur les réponses du sujet selon la place de présentation du stimulus standard. Ainsi, le pourcentage de réponses correctes est plus élevé quand le stimulus variable est présenté en 3^{ieme} position(cas (S,S,V)) que dans le 1^{e} (cas (V,S,S)). Cette observation est néanmoins variable selon les sujets. En particulier, cet effet d'intervalle est nettement moins marqué chez les sujets ayant les seuils les plus bas.

9.3 Mise en place du test

Les points que nous désirons étudier sont les suivants :

1. **influence de la valeur de O_q cible sur le seuil différentiel.** Va-t-on obtenir un seuil identique si on modifie la valeur de quotient ouvert cible ?
2. **influence de la valeur de α_m cible sur le seuil différentiel.** De même que pour le quotient ouvert, va-t-on obtenir un seuil identique si on modifie la valeur du coefficient d'asymétrie cible ?
3. **influence de la voyelle.** Le choix d'une voyelle donnée décide de la position des formants sur le spectre du signal voisé synthétique. Dans le cas de la voyelle [a], la fréquence de F_1 est élevée (entre 600 et 800 Hz). Ce n'est pas le cas d'une voyelle comme le [e], où F_1 avoisine plutôt les 300 Hz. Cette accentuation de l'une ou l'autre des zones fréquentielles va-t-elle avoir une influence sur le seuil ?
4. **influence de la hauteur.** Le seuil différentiel va-t-il varier selon la hauteur à laquelle on se place ?
5. **influence du vibrato.** L'étude de Scherer & al [180] a porté sur un son où la fréquence fondamentale est gardée constante. La perception du seuil est-elle modifiée si l'on présente des exemples vocaux plus proches de la réalité, où la fréquence fondamentale n'est pas constante ? Le vibrato est un exemple typique de variation locale de fréquence fondamentale au cours de la production vocale. Dans quelle mesure cet ornement, qui rajoute toujours un côté plus « réel » à un son synthétique (en particulier en voix chantée), va-t-il influencer la perception du seuil ?
6. **influence du paramètre d'amplitude.** On peut fixer soit l'amplitude de l'ODG (en fixant le paramètre A_v), soit celle de l'ODGD (en fixant le paramètre E), les deux paramètres ne pouvant être gardés constants simultanément (la relation entre les deux dépend de O_q et de α_m). Va-t-on alors retrouver dans les deux cas des valeurs de seuils comparables ?

Nous allons tout d'abord décrire la méthode de synthèse des stimuli, réfléchie pour rendre l'impression d'un son naturel, puis nous donnerons le détail des sessions qui composent le test.

9.3.1 Synthèse des stimuli

Une des étapes majeures dans la mise en place de ce test perceptif a été de réussir à synthétiser des stimuli qui s'approchent au plus près de sons vocaux réels, pour une voyelle tenue sur une hauteur donnée, tout en gardant constants tout à la fois les paramètres de source glottique (hormis la fréquence fondamentale) et la fonction de transfert du conduit vocal. La première approche a donc été de rajouter un vibrato de fréquence à une fréquence fondamentale fixe, ce qui permet déjà de donner à un son purement synthétique une petite impression de naturel. Pour améliorer néanmoins de façon plus conséquente le naturel de la synthèse, l'idée est venue ensuite de ne pas synthétiser le vibrato mais de déduire de l'analyse d'un son réel l'évolution de la fréquence fondamentale au cours de l'émission sonore. La synthèse des stimuli a alors été effectuée en 3 étapes successives que nous allons à présent détailler.

1. **Détection des instants de fermeture glottique**

   Des exemples vocaux (voyelles tenues ou sons crescendos) ont été extraits de la base de données des chanteurs présentée dans le chapitre 6. Ces exemples ont été analysés afin de déterminer les instants de fermeture glottique. Cette analyse a été effectuée en utilisant le programme en C développé par Vu Ngoc Tuan qui permet la détection des instants de fermeture glottique (GCI ou « Glottal Closure Instant ») sur un signal de parole (Vu Ngoc & d'Alessandro [215, 216]). Le signal à analyser est décomposé en ondelettes par un banc de filtres. Les signaux ainsi obtenus pour différentes fréquences présentent des maxima locaux, qui sont reliés à des minima locaux du signal de parole (en raison du signe de la fonction de transfert des filtres). En parcourant sur une période l'espace temps-échelle depuis les hautes fréquences jusqu'aux basses fréquences, on observe un regroupement progressif des multiples maxima présents dans les hautes fréquences, jusqu'à ne former plus qu'un unique maximum (moins résonant, plus étalé) dans les basses fréquences. Un suivi des lignes de maxima depuis les basses fréquences jusqu'aux hautes fréquences est effectué par programmation dynamique. Ce suivi permet de déterminer le maximum présent dans les hautes fréquences qui est corrélé au maximum basse fréquence. Sa position temporelle donne une mesure précise de l'instant de fermeture glottique.

   Cette méthode a vocation à s'appliquer à un signal voisé, où la détection du minimum correspondant à la fermeture peut s'avérer problématique. Elle détecte également parfaitement bien les instants de fermeture glottique sur un signal DEGG inverse (pic de fermeture vers le bas), puisque les minima correspondants sont alors très marqués. Comme nous avions à disposition les signaux DEGG, nous avons utilisé ces derniers pour la détection des GCI. Une détection directe des pics sur le signal DEGG était également envisageable.

2. **Synthèse d'un signal de source glottique dérivé**

   A partir des GCI détectés précédemment, un signal de source dérivé a été reconstruit période par période en utilisant comme modèle d'ODGD le modèle LF, à partir d'un jeu de paramètres (T_0, A_v, O_q, α_m, Q_a) donné (voir le chapitre 3 et l'annexe A pour plus de détails sur ce modèle). La période fondamentale est donnée par les GCI, tandis que les autres paramètres de source glottique sont gardés constants. La variation de fréquence fondamentale est donc « naturelle », car elle est reproduite fidèlement.

3. **Détermination de la fonction de transfert du conduit vocal**

   Le signal de source glottique dérivé ainsi synthétisé est ensuite filtré par la fonction de transfert du conduit vocal. Dans le but de synthétiser un signal vocal qui se rapproche au mieux du signal dont sont issus les GCI, une analyse par prédiction linéaire (méthode par autocorrélation) a été effectuée sur une sélection du signal voisé original (sélection manuelle de l'instant d'analyse sur le sonagramme et fenêtrage par une fenêtre de Hanning de 40 ms). La fréquence d'échantillonnage choisie pour cette analyse était volontairement élevée (F_e = 44.1 kHz), afin de rendre au mieux compte de l'énergie présente dans les hautes fréquences. Le filtre auto-régressif ainsi obtenu a été validé perceptivement par la synthèse, afin que la voyelle synthétique ait une couleur proche de celle

du son original.

Illustration de la synthèse

La figure 9.5 montre le sonagramme d'une voyelle [a] chantée par le baryton B3 sur un sol2 (f_0 = 196 Hz). La figure 9.6 présente le résultat de la synthèse. Les paramètres de source glottique sont gardés constants : $A_v = 1$, $O_q = 0.5$, $\alpha_m = 0.75$ et $Q_a = 0$. L'intensité de l'exemple de synthèse n'a pas été ajustée par rapport à l'exemple original et elle est donc plus élevée.

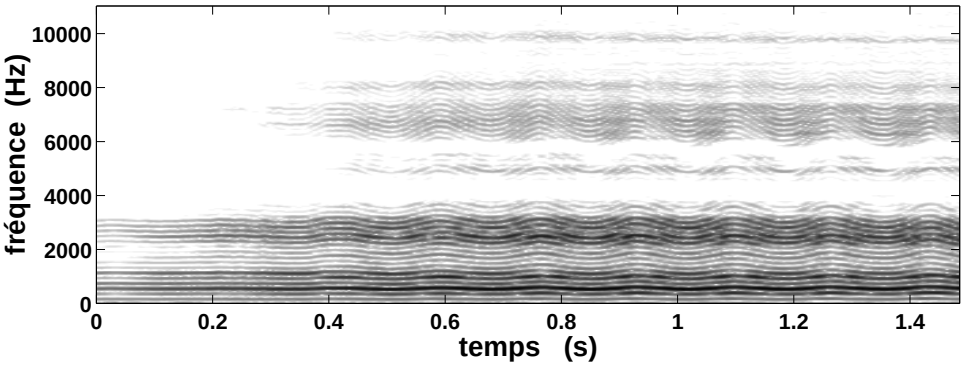


FIG. *9.5 – Sonagramme du signal original utilisé pour la détection des GCI (f_0 = 196 Hz, voir figure 9.7) et la voyelle [a] (voir figure 9.8). (CD, piste [27])*

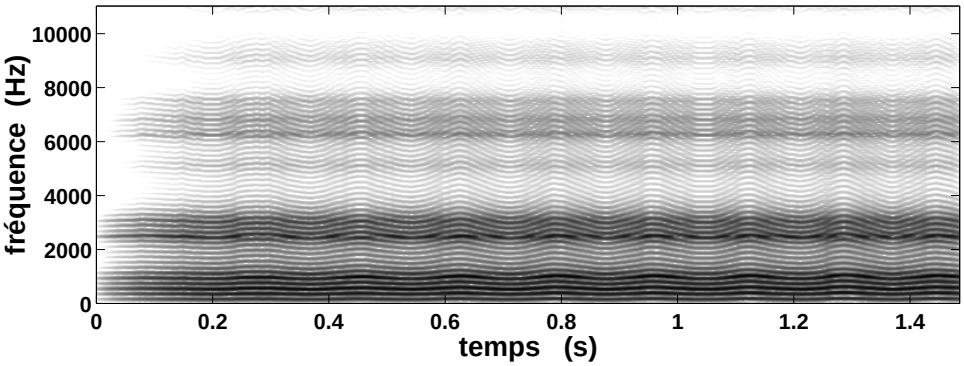


FIG. *9.6 – Sonagramme d'un exemple de synthèse à partir de la connaissance des GCI du signal original ci-dessus et du filtre modélisant un [a] présenté sur la figure 9.8. Les paramètres de source glottique ont été fixés à : $A_v = 1$, $O_q = 0.5$, $\alpha_m = 0.75$ et $Q_a = 0$. (CD, piste [28])*

9.3.2 Description des sessions

Chaque point que nous désirons aborder dans cette expérience perceptive et que nous avons énuméré précédemment en introduction de cette partie nécessite la mise en place de sessions comparatives. Afin de ne pas alourdir excessivement le test, nous avons choisi de n'effectuer la comparaison que sur deux sessions, ce qui est limitatif mais permet d'aborder les diverses questions posées. Le test perceptif mis en place comprend donc une session d'entraînement et 9 sessions pour lesquelles soit le quotient ouvert soit le coefficient d'asymétrie varient dans une situation donnée. Le tableau 9.1 présente les caractéristiques de chaque session, dans l'ordre d'apparition dans le test. Les valeurs par défaut sont : $f_0 = 196\,Hz$, $A_v = 1$, $O_q = 0.6$, $\alpha_m = 2/3$, voyelle [a]. Dans tous les cas, le pas limite a été choisi à 0.01 après des tests préliminaires. Le pas de départ a été fixé à 0.04. Le nombre maximal d'inversions est de 14 pour les

session	variable	O_q	α_m	f_0	voyelle	autre paramètre
1	O_q	**0.4**	2/3	196 Hz	[a]	
2	O_q	0.6	2/3	196 Hz	[a]	**sans vibrato**
3	O_q	**0.6**	2/3	196 Hz	[a]	
4	O_q	0.6	2/3	196 Hz	[a]	**E constant**
5	O_q	0.6	2/3	**130 Hz**	[a]	
6	O_q	**0.8**	2/3	196 Hz	[a]	
7	α_m	0.6	**2/3**	196 Hz	[a]	
8	α_m	0.6	**0.8**	196 Hz	[a]	
9	O_q	0.6	2/3	196 Hz	**[e]**	

Tab. *9.1 – Caractéristiques des 9 sessions composant le test perceptif, présentées dans leur ordre d'apparition. Pour chaque session, le paramètre en gras précise le paramètre modifié pour la synthèse du stimulus standard S, par rapport aux valeurs par défaut ($f_0 = 196\ Hz$, $A_v = 1$, $O_q = 0.6$, $\alpha_m = 2/3$, voyelle [a]). La colonne « variable » précise le paramètre qui varie lors de la synthèse du stimulus variable V.*

sessions où O_q varie et 12 pour les sessions où α_m varie (en raison d'une convergence souvent plus rapide dans ces cas-là). Dans tous les cas, le calcul du seuil se fait sur les 8 dernières inversions.

Pour étudier l'influence de la hauteur, les deux fréquences fondamentales moyennes choisies sont $f_0 = 130\ Hz$ et $f_0 = 196\ Hz$, ce qui nécessite les données GCI sur deux exemples. Les deux fichiers GCI retenus pour la synthèse proviennent de l'analyse d'une voyelle [a] chantée par le chanteur B3 sur deux hauteurs séparées par une quinte : do2 ($f_0 = 130\ Hz$) et sol2 ($f_0 = 196\ Hz$). Ils sont présentés sur la figure 9.7, sous la forme d'une courbe de fréquence fondamentale, obtenue par : $f_0(i) = \frac{1}{T_{GCI}(i+1) - T_{GCI}(i)}$ pour $1 \leq i < N$ avec N la longueur du vecteur T_{GCI} composé des instants de fermeture glottique détectés sur le signal analysé.

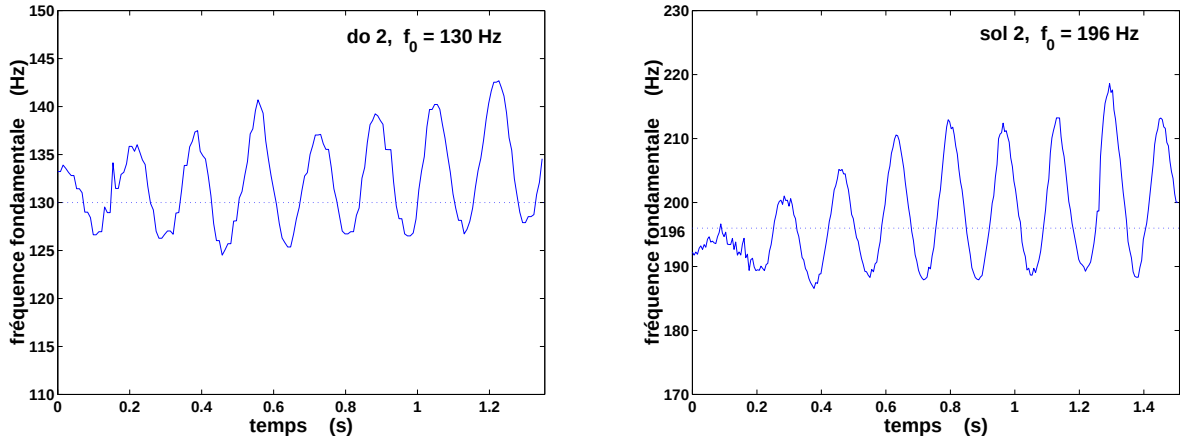


Fig. *9.7 – Courbe de fréquence fondamentale provenant de la détection des GCI sur une voyelle chantée par le chanteur B3 sur les deux hauteurs do2 et sol2.*

Pour étudier l'influence de la voyelle, les voyelles retenues (extraites de la base de données des chanteurs) sont le [a] et le [e]. Les deux filtres modélisant le conduit vocal proviennent de l'analyse des voyelles [a] et [e] chantées par le même chanteur B3. Les figures 9.8 et 9.9 présentent les fonctions de transfert

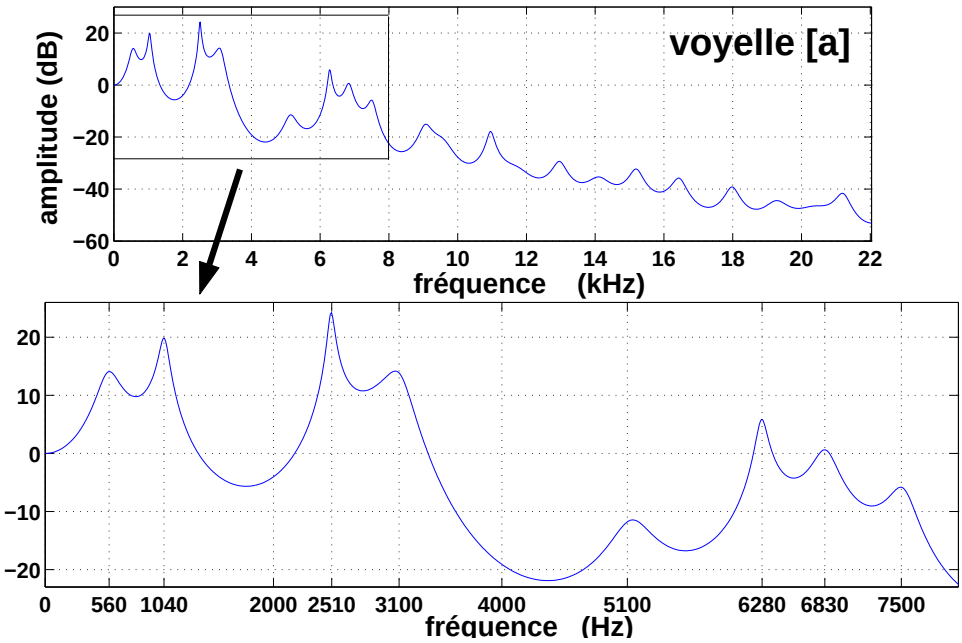


FIG. *9.8 – Fonction de tranfert associée à la synthèse de la voyelle [a].*

utilisées pour synthétiser les deux voyelles [a] et [e].

Les filtres choisis modélisent les pics formantiques présents dans la zone fréquentielle $0-8000\ Hz$, mais rendent également compte de l'énergie présente dans les hautes fréquences (ici, jusqu'à 22050 Hz). Il est intéressant de remarquer que, contrairement au cas de la parole, les signaux de voix chantée possèdent de l'énergie jusqu'à des fréquences très élevées. En rendre compte convenablement permet alors d'améliorer la qualité générale de la synthèse.

9.3.3 Déroulement du test

Le test se déroule dans une chambre insonorisée, précédemment utilisée pour l'enregistrement des chanteurs. L'écoute se fait au casque, de façon binaurale, à un niveau confortable d'écoute. Ce niveau est maintenu constant et identique pour tous les sujets.

Par ce test perceptif, nous explorons les capacités de discrimination sensorielle (plutôt que cognitive) des sujets et il est donc nécessaire de connaître l'état de l'organe auditif. Les sujets passent donc au préalable un test auditif (« screening » à 20 dB), qui consiste à vérifier qu'ils perçoivent correctement, par l'oreille droite comme par l'oreille gauche, des sons purs d'une intensité de 20 dB et de fréquence fondamentale $f_0 = 250$, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000 et 8000 Hz.

Le test commence par une session d'entraînement, qui a pour but de familiariser le sujet avec la tâche demandée. Cette session d'entraînement, qui comprend 3 inversions, dure environ 5 mn. Le test se poursuit par 9 sessions présentée dans l'ordre précisé dans le tableau 9.1. La durée d'une session varie entre 10 et 20 mn, selon la rapidité de convergence du sujet. Entre deux sessions, le sujet est autorisé à faire une pause s'il le juge nécessaire et des pauses après la session 3 et la session 6 lui sont même conseillées.

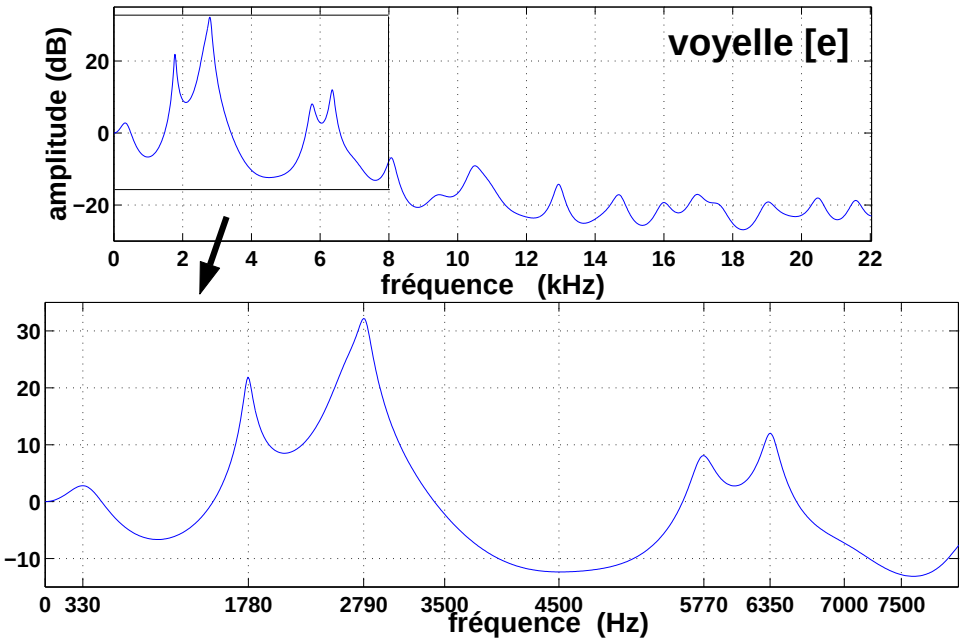


Fig. *9.9 – Fonction de tranfert associée à la synthèse de la voyelle [e].*

9.3.4 Les sujets

27 sujets ont participé à cette expérience perceptive, tous de façon bénévole. Lors du test auditif préalable, 7 ont révélé une faiblesse auditive, généralement dans les hautes fréquences (sur 4000, 6000 et 8000 Hz). Leurs résultats ne sont donc pas directement exploitables dans le cadre de cette étude, mais nous ne les écarterons pas complètement, car, comme nous le verrons par la suite, ils sont intéressants comparativement aux résultats du groupe des 20 sujets ayant passé le test auditif de façon positive.

Le tableau 9.2 présente les caractéristiques du groupe des 20 sujets (screening normal) et du groupe des 7 sujets (faiblesse auditive). L'âge des sujets varie entre 23 et 58 ans, avec un âge moyen (valeur médiane) de 27 ans. Le sexe des sujets n'a pas été pris en compte puisque Scherer & al [180] ont montré que cela n'entrainait aucune différence sur les seuils mesurés. Dans le groupe S1 à S20, les sujets sont tous soit acousticiens, soit musiciens, à l'exception du sujet S16. Il s'agit donc d'un groupe de sujets « experts », en ce sens qu'ils ont tous une oreille formée à l'écoute. Parmi ces sujets, 10 ont passé le test 3 fois (S11 à S20), à quelques jours ou quelques semaines d'intervalle. Seul le sujet S11 a effectué son 3^{e} passage à plus de 3 mois d'intervalle du 2^{nd}. Les sujets S15 et S17 sont les expérimentatrices de l'étude.

Nous allons considérer dans cette étude 3 groupes de sujets :

1. le groupe des 20 sujets, pour lesquels seul le premier passage du test a été pris en compte. Il constitue un groupe de 20 sujets non-entraînés ;

2. le groupe des 10 sujets qui ont passé le test 3 fois. Comme nous allons le voir, ce groupe peut être considéré comme un groupe de 10 sujets entraînés.

3. le groupe des 7 sujets qui ont présenté des faiblesses auditives.

Sujets non-entraînés

sujets	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
passage	1	1	1	1	1	1	1	1	1	1
âge	45	29	36	39	23	28	25	24	27	24
acousticien	X	X	X	X	X	X				X
musicien		X	X	X	X	X	X	X	X	X

Sujets entraînés

sujets	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
passage	3	3	3	3	3	3	3	3	3	3
âge	28	32	23	23	26	26	35	25	24	36
acousticien		X		X	X		X	X	X	X
musicien	X		X	X	X		X	X	X	X
Intervalles entre passage (en jours)										
2^{nd} **passage**	7	36	3	13	11	3	15	16	21	14
3^{ie} **passage**	111	30	61	47	18	9	15	15	3	6

Sujets présentant une faiblesse auditive

sujets	A1	A2	A3	A4	A5	A6	A7
âge	58	37	23	55	24	23	44
acousticien				X		X	
musicien	X	X			X	X	
screening	OG/OD $2-8\ kHz$ $60\ dB$	OG $4\ kHz$ $25\ dB$	OG $6\ kHz$ $30\ dB$	OG/OD $6-8\ kHz$ $30/55\ dB$	OG $4/6\ kHz$ $25/30\ dB$	OG $6\ kHz$ $30\ dB$	OG/OD $3-8\ kHz$ $45/50\ dB$

TAB. *9.2 – Caractéristiques des sujets. La case « passage » indique le nombre de fois que le sujet a passé le test complet. On précise également si le sujet est acousticien et/ou musicien (repérés par X) . Les sujets A1 à A7, qui présentent une faiblesse auditive détaillée dans la case « screening », ont tous passé le test une seule fois. Dans le cas des sujets ayant passé 3 fois le test (sujets S11 à S20), l'intervalle temporel (en jours) entre le premier et le 2^{nd} passage est précisé dans la case « 2^{nd} passage » et celui entre le 2^{nd} et le 3^{ie} dans la case « 3^{ie} passage ».*

Seuil	Sessions								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	$O_q = 0.4$	sans vib.	$O_q = 0.6$	E const.	$f_0 = 130$	$O_q = 0.8$	$\alpha_m = 2/3$	$\alpha_m = 0.8$	voyelle [e]
Groupe de 20 sujets non-entraînés									
seuil moyen	0.058	0.068	0.087	0.173	0.092	0.106	0.033	0.027	0.080
écart-type	0.019	0.032	0.027	0.077	0.036	0.039	0.009	0.009	0.029
seuil min	0.022	0.009	0.043	0.034	0.049	0.049	0.019	0.012	0.042
seuil max	0.092	0.161	0.153	0.298	0.190	0.190	0.052	0.049	0.144
Groupe de 10 sujets entraînés									
seuil moyen	0.037	0.041	0.063	0.099	0.071	0.079	0.027	0.022	0.053
écart-type	0.010	0.008	0.015	0.047	0.018	0.014	0.010	0.008	0.013
seuil min	0.024	0.027	0.043	0.044	0.051	0.061	0.013	0.012	0.033
seuil max	0.053	0.052	0.090	0.178	0.093	0.101	0.048	0.037	0.067
Groupe de 7 sujets présentant des faiblesses auditives									
seuil moyen	0.063	0.075	0.133	0.235	0.149	0.156	0.041	0.032	0.142
écart-type	0.012	0.019	0.029	0.087	0.052	0.017	0.008	0.012	0.031
seuil min	0.048	0.051	0.099	0.118	0.088	0.128	0.030	0.010	0.100
seuil max	0.081	0.106	0.180	0.356	0.224	0.173	0.049	0.044	0.198

Tab. *9.3 – Valeurs moyennes, écart-types et valeurs extrêmes du seuil mesuré pour un groupe de* 20 *sujets non-entraînés, de* 10 *sujets entraînés et de* 7 *sujets présentant des faiblesses auditives.*

9.4 Résultats

9.4.1 Résultats généraux

Seuil moyen, valeurs extrêmes

Pour chaque sujet et pour chaque session, une valeur de seuil différentiel a été mesurée. Pour une session et un groupe donnés, la moyenne des seuils différentiels mesurés, que nous appellerons seuil moyen, renseigne sur le comportement global de l'ensemble des sujets du groupe. Les valeurs extrêmes apportent des informations sur les plus petites et les plus grandes valeurs rencontrées dans la session. Plus intéressantes que l'écart-type dans ce cas, elles indiquent clairement en-dessous de quelle valeur les sujets n'ont plus perçu de différence, à l'unanimité, et au-dessus de quelle valeur les sujets ont tous perçu une différence, sans exception. Ces trois informations (valeur moyenne et valeurs extrêmes de seuil différentiel) sont donc complémentaires et sont présentées dans le tableau 9.3 par groupe de sujets et par session. Les valeurs des écart-types sont également fournies.

En considérant les résultats du groupe des sujets entraînés, nous pouvons noter que les seuils obtenus sont en accord avec ceux présentés par Scherer & al [180]. Ils mesurent un seuil moyen de quotient ouvert de $\Delta O_q = 0.0344$ pour une fréquence fondamentale constante $f_0 = 130.39\ Hz$ et un coefficient d'asymétrie $\alpha_m = \frac{2}{3}$. Si nous comparons cette valeur avec le seuil moyen $\Delta O_q = 0.041$ obtenu par le groupe des sujets entraînés en session 2 (fréquence fondamentale constante), nous notons que cette dernière valeur est légèrement supérieure.

Nous observons de nettes différences entre les 3 groupes. Quelle que soit la session considérée, les 7 sujets ayant une faiblesse auditive ont toujours les seuils significativement les plus élevés, que ce soit au niveau du seuil moyen ($p < 0.01$) ou du seuil minimal ($p < 0.01$). A l'opposé, les sujets entraînés présentent les meilleurs résultats au niveau du seuil moyen ($p < 0.01$) et du seuil maximal ($p < 0.01$) par rapport aux deux autres groupes. Par contre, il n'y a pas de différence significative sur le seuil minimal entre les groupes des sujets entraînés et ceux non-entraînés. Il n'y a pas non plus de différence significative sur le seuil maximal entre les groupes des sujets non-entraînés et ceux présentant une faiblesse auditive.

comparaisons entre passages	SESSIONS 1	2	3	4	5	6	7	8	9
$1 \leftrightarrow 2$	2.57 **	1.66 *	1.08	6.88 *	0.86	0.86	0.61	0.48 *	2.19 *
$1 \leftrightarrow 3$	1.83 **	2.60 **	1.12	4.77	1.14	1.62	0.19	0.28	2.19 **
$2 \leftrightarrow 3$	−0.73	0.94	0.04	−2.11	0.29	0.76	−0.42	−0.20	0.00

TAB. *9.4 – Comparaisons entre les* 3 *passages effectués par les* 10 *sujets entraînés. Pour chaque session, les différences de seuil entre deux passages sont moyennées sur les* 10 *sujets. Cette différence moyenne entre* 2 *passages donnés est exprimée en %. Un effet d'apprentissage significatif est marqué par* * *et correspond à* $p < 0.05$ *(*$t > t_{(9\,;\,0.05)} = 2.26$*). Un effet d'apprentissage très significatif est marqué par* ** *et correspond à* $p < 0.01$ *(*$t > t_{(9\,;\,0.01)} = 3.25$*).*

Les différences marquées entre le groupe des sujets présentant une faiblesse auditive et les autres sujets montrent la nécessité de définir soigneusement le cadre de l'expérience en testant au préalable la capacité auditive des sujets. Celle-ci semble avoir un impact important sur la détermination de ces seuils. La différence entre sujets entraînés et non-entraînés pourrait s'expliquer par un effet d'apprentissage par répétition du test. Avant de détailler les résultats par session, nous allons donc étudier cet effet sur le groupe des 10 sujets ayant passés le test perceptif 3 fois.

Effet de l'apprentissage

L'effet de l'apprentissage a été étudié sur le groupe de 10 sujets ayant passé 3 fois le test perceptif. Des tests statistiques sur observations appariées (effectués d'après Daudin & al [49], voir Annexe D) ont été mis en place pour comparer les différents passages ($1 \leftrightarrow 2$, $2 \leftrightarrow 3$, $1 \leftrightarrow 3$), pour chacune des 9 sessions, afin de déterminer un éventuel effet d'apprentissage entre les passages. Pour chacune des 3 comparaisons et pour chaque session, le tableau 9.4 fournit la moyenne des différences de seuil entre 2 passages donnés, calculée sur les 10 sujets entraînés. Cette différence moyenne est exprimée en % pour plus de lisibilité. L'Annexe D précise les calculs effectués.

Nous observons une différence très significative entre le premier et le second passage, ainsi qu'entre le premier et le dernier passage, pour les deux premières sessions. Cette différence n'est plus significative pour la plupart des sessions suivantes. On peut noter une différence significative entre $1 \leftrightarrow 2$ et $1 \leftrightarrow 3$ au niveau de la dernière session (session 9) qui correspond au changement de voyelle. Nous n'observons aucune différence significative entre le second et le dernier passage. Le signe de la différence moyenne fournit une indication sur l'apprentissage entre les passages P_i et P_j : si elle est positive, cela signifie qu'en moyenne les seuils obtenus lors du passage P_j pour la session correspondante sont plus bas que ceux obtenus lors du passage P_i. Nous remarquons que, pour la comparaison $1 \leftrightarrow 2$ et $1 \leftrightarrow 3$, les différences moyennes sont toutes positives, ce qui indique une amélioration de la détection du seuil dans toutes les sessions. Par contre, entre les passages 2 et 3, cette différence peut être positive (sessions 2 , 3 , 5 , 6 , 9) ou négative (sessions 1 , 4 , 7 , 8), ce qui confirme qu'il n'y a plus d'effet d'apprentissage entre ces deux passages.

En conclusion, l'effet d'apprentissage est marqué entre le premier et le second passage du test, principalement sur les premières sessions. Le fait que cet effet d'apprentissage ne soit plus marqué sur les sessions suivantes indique qu'au sein du premier test, les sujets ont en général dépassé très rapidement le stade de l'apprentissage et les valeurs de seuils obtenus sont alors comparables d'un test à l'autre. Nous observons néanmoins une diminution du seuil détecté entre les deux premiers passages.

Les second et troisième passages sont comparables et amènent à des seuils très proches. Selon les

comparaisons	quotient ouvert			coefficient d'asymétrie
	$Oq = 0.4 \leftrightarrow 0.6$ sessions $1 \leftrightarrow 3$	$Oq = 0.6 \leftrightarrow 0.8$ sessions $3 \leftrightarrow 6$	$Oq = 0.4 \leftrightarrow 0.8$ sessions $1 \leftrightarrow 6$	$\alpha_m = 2/3 \leftrightarrow 0.8$ sessions $7 \leftrightarrow 8$
sujets non-entraînés	-2.96 ***	-1.86 *	-4.81 ***	0.56 *
sujets entraînés	-2.59 ***	-1.59 **	-4.18 ***	0.58 *

Tab. *9.5 – Comparaisons entre deux sessions données portant sur O_q ou α_m, dans le cas des groupes de sujets non-entraînés et entraînés. Pour chaque groupe, les différences de seuil entre deux sessions sont moyennées sur l'ensemble des sujets. Cette différence moyenne entre 2 sessions données est exprimée en %. Une différence significative entre les deux sessions est marquée par * et correspond à $p < 0.05$. Une différence très significative est marquée par ** ($p < 0.01$) et fortement significative par *** ($p < 0.001$).*

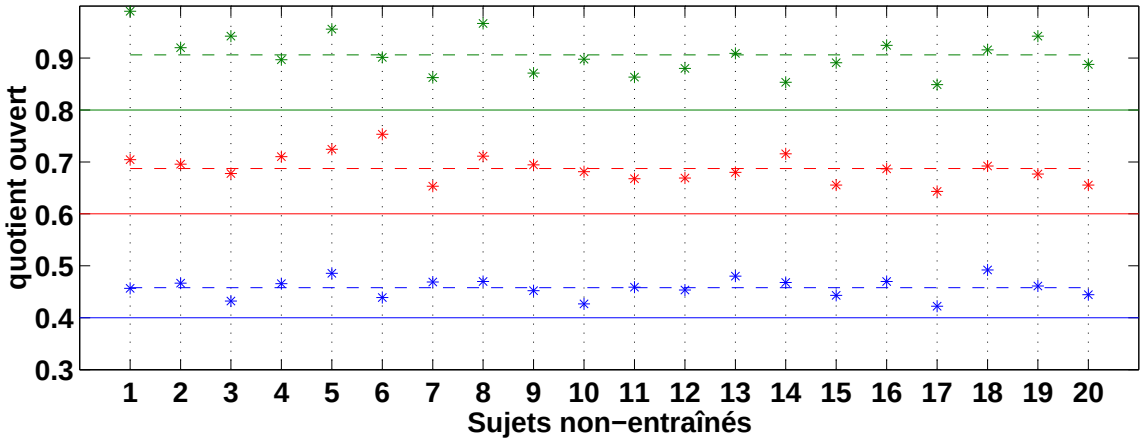


Fig. *9.10 – Valeurs du quotient ouvert associées aux seuils mesurés sur les 20 sujets non-entraînés pour 3 valeurs cibles de O_q (0.4, 0.6, et 0.8, ce qui correspond respectivement aux sessions 1, 3 et 6). Les cibles sont indiquées par des lignes pleines et les valeurs moyennes pour l'ensemble du groupe par des lignes en pointillés.*

sessions, nous observons soit une augmentation, soit une diminution du seuil détecté.

Pour ces raisons, nous avons choisi comme valeur de seuil dans le cas des 10 sujets entraînés la moyenne des seuils détectés lors du 2^{nd} et du 3^{ieme} passage.

9.4.2 Mesure du seuil pour 3 valeurs de quotient ouvert

Résultats

Les 3 valeurs de quotient ouvert explorées sont : $O_q = 0.4$, $O_q = 0.6$ et $O_q = 0.8$. Les résultats pour chacun des 20 sujets non-entraînés sont illustrés sur la figure 9.10 qui présente les valeurs de quotient ouvert associées aux seuils différentiels mesurés ($O_q + \Delta O_q$) pour chacun de ces sujets.

Les lignes pleines correspondent aux 3 valeurs cibles et les lignes en pointillés aux valeurs de quotient ouvert associées aux seuils moyens calculés sur l'ensemble des premiers (ou uniques) passages effectués par les 20 sujets.

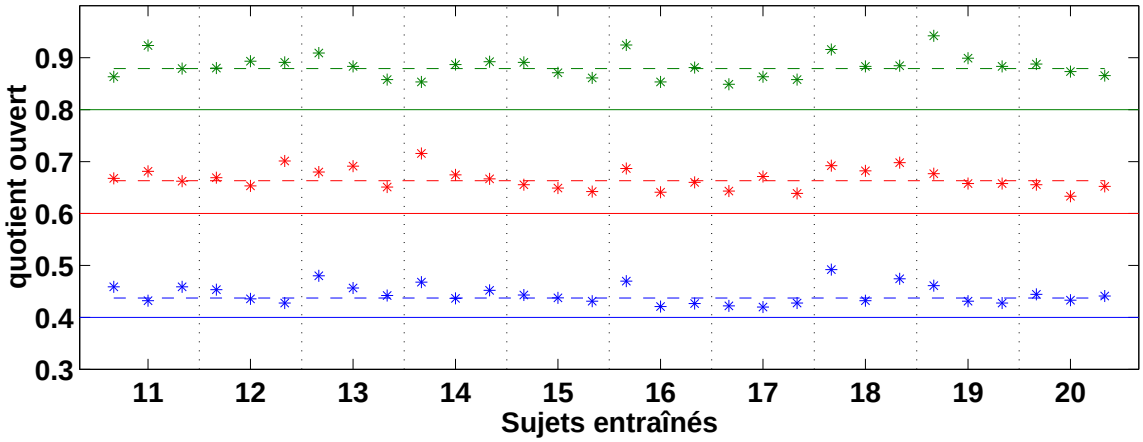


FIG. *9.11 – Valeurs du quotient ouvert associées aux seuils mesurés sur les* 10 *sujets entraînés pour* 3 *valeurs de* O_q *(*0.4*,* 0.6*, et* 0.8*, ce qui correspond respectivement aux sessions* 1*,* 3 *et* 6*). Les cibles sont indiquées par des lignes pleines et les valeurs moyennes pour l'ensemble du groupe par des lignes en pointillés.*

Pour $O_q = 0.4$, le seuil moyen est $\Delta O_q = 0.058$. Il augmente jusqu'à 0.087 pour $O_q = 0.6$ et 0.106 pour $O_q = 0.8$.

Les résultats obtenus par les 10 sujets entraînés sont présentés sur la figure 9.11. Nous avons associé à chaque sujet 3 points, qui correspondent aux seuils obtenus lors des 3 passages et que nous faisons figurer dans l'ordre chronologique de passage. Pour $O_q = 0.4$, le seuil moyen est $\Delta O_q = 0.037$. Il augmente jusqu'à 0.063 pour $O_q = 0.6$ et 0.079 pour $O_q = 0.8$. Nous remarquons l'effet d'apprentissage entre le premier et le second passage, très net pour les sujets S16, S18 et S19. Le sujet S17 est l'expérimentateur et il présente donc dès le départ des valeurs de seuil très faibles. Il est également très stable d'un passage à l'autre.

Enfin, la partie gauche de la figure 9.12 présente les résultats obtenus par les 7 sujets présentant une faiblesse auditive. Ils ont obtenu les seuils moyens les plus élevés : $\Delta O_q = 0.063$ pour $0_q = 0.4$, 0.133 pour $O_q = 0.6$ et 0.156 pour $O_q = 0.8$.

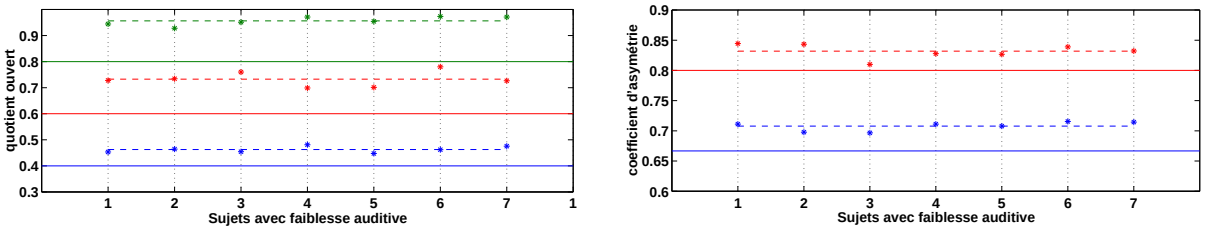


FIG. *9.12 – Figure de gauche : valeurs du quotient ouvert associées aux seuils mesurés sur les* 7 *sujets présentant une faiblesse auditive pour* 3 *valeurs de* O_q *(*0.4*,* 0.6*, et* 0.8*). Figure de droite : valeurs du coefficient d'asymétrie associées aux seuils mesurés sur ces* 7 *sujets pour* 2 *valeurs de* α_m *(*$\frac{2}{3}$*,* 0.8*). Les cibles sont indiquées par des lignes pleines et les valeurs moyennes pour l'ensemble du groupe par des lignes en pointillés.*

$\frac{\Delta O_q}{O_q}$ (en %)	$O_q = 0.4$	$O_q = 0.6$	$O_q = 0.8$
sujets non-entraînés	14.5	14.5	13.2
sujets entraînés	9.2	10.5	9.9

TAB. *9.6 – Seuil différentiel relatif moyen ($\frac{\Delta O_q}{O_q}$) exprimé en % pour $O_q = 0.4$ (session 1), 0.6 (session 3) et 0.8 (session 6).*

$\frac{\Delta \alpha_m}{\alpha_m}$ (en %)	$\alpha_m = 2/3$	$\alpha_m = 0.8$
sujets non-entraînés	4.9	3.3
sujets entraînés	4.0	2.7

TAB. *9.7 – Seuil différentiel relatif moyen ($\frac{\Delta \alpha_m}{\alpha_m}$) exprimé en % pour $\alpha_m = 2/3$ (session 7) et 0.8 (session 8).*

Discussion

Quel que soit le groupe de sujets concerné, plus O_q est élevé, moins ses variations sont perceptibles. Cet effet est fortement significatif ($p < 0.001$) ainsi que l'indique le tableau 9.5, que ce soit pour les sujets non-entraînés ou pour les sujets entraînés. Par contre, si l'on s'intéresse au seuil différentiel relatif, $\frac{\Delta O_q}{O_q}$, on note que ce rapport reste relativement constant, ainsi que l'indique le tableau 9.6 qui présente le seuil différentiel relatif moyen pour les 3 valeurs de O_q, dans le cas des groupes de sujets non-entraînés et entraînés. La différence sur le seuil différentiel relatif entre les 3 sessions (1, 3 et 6) n'est alors plus significative. Ces résultats sont en accord avec la loi de Bouguer-Weber, qui affirme que le seuil différentiel (ici ΔO_q) est proportionnel à l'amplitude du stimulus (O_q) et que donc le seuil différentiel relatif est une constante (Bonnet [22], Botte [23]). Bien qu'il soit difficile de généraliser à partir de 3 mesures seulement, nous pouvons avancer les deux relations suivantes liant le seuil différentiel aux valeurs de quotient ouvert (pour $f_0 = 196$ Hz, $A_v = 1$, $\alpha_m = 2/3$ et $Q_a = 0$) :

1. dans le cas des sujets non-entraînés, $\Delta O_q \simeq 0.14\ O_q$.
2. dans le cas des sujets entraînés, $\Delta O_q \simeq 0.1\ O_q$.

Cet accroissement du seuil différentiel en fonction du quotient ouvert peut-il s'interpréter à l'aide de variations temporelles ou spectrales de la source glottique ? En gardant tous les paramètres constants et en faisant varier O_q, on va indirectement faire varier la vitesse de fermeture E, paramètre lié à l'intensité vocale. Les variations de O_q vont donc s'accompagner d'une modification de l'intensité sonore perçue. La partie gauche de la figure 9.13 présente l'évolution de E en fonction de O_q, quand les autres paramètres de source glottique sont fixés aux valeurs suivantes : $A_v = 1$, $f_0 = 196$ Hz et $\alpha_m = 2/3$. Les tangentes à la courbe pour $O_q = 0.4$, 0.6 et 0.8 sont représentées. Nous voyons que leurs pentes diminuent quand O_q augmente, ce qui signifie que les variations de E induites par des variations de O_q seront moins amples pour des valeurs élevées de O_q que pour des valeurs basses. Les modifications de l'intensité sonore seront donc plus marquées pour de faibles valeurs de O_q et donc plus perceptibles.

Peut-il y avoir une influence du formant glottique ? Nous avons vu au chapitre 3 que sa fréquence est proportionnelle à l'inverse de O_q pour f_0 et α_m constants. Quand le quotient ouvert augmente, la fréquence du formant glottique diminue. Pour $f_0 = 196$ Hz et $\alpha_m = \frac{2}{3}$, il est situé à $F_g = 250$ Hz pour $O_q = 0.4$, donc un peu au-dessus de H_1, et il descend à $F_g = 125$ Hz pour $O_q = 0.8$, en-dessous de H_1. Dans le domaine des basses fréquences, la sensibilité de l'oreille humaine à des sons purs est d'autant plus faible que la fréquence fondamentale de ces sons est basse. De plus, l'oreille est particulièrement sensible à de petites variations de la fréquence fondamentale (i.e. de H_1). Pour ces raisons, de petites variations autour des valeurs faibles de O_q pourraient être mieux perçues qu'autour de valeurs élevées. Néanmoins, les résultats obtenus par le groupe des 7 sujets présentant une faiblesse auditive qui affecte, pour tous, les hautes fréquences, laissent à penser que, si l'effet du formant glottique pouvait être important, les seuils différentiels obtenus par ces sujets devraient être comparables à ceux des autres sujets puisque cet effet est basse fréquence. Or leurs seuils sont nettement supérieurs, ce qui nous amène à penser que les variations perçues par les sujets sont principalement des variations de la richesse spectrale et que l'influence du

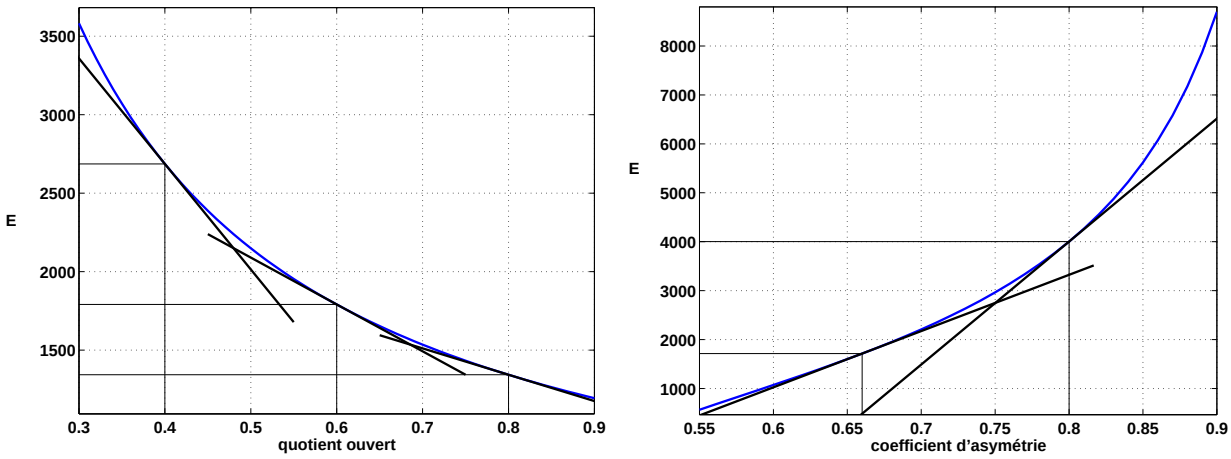


FIG. *9.13 – Etude des variations de E en fonction du quotient ouvert (figure de gauche) et du coefficient d'asymétrie (figure de droite). Les valeurs par défaut sont celles de la session 3 : $A_v = 1$, $f_0 = 196$ Hz, $O_q = 0.6$ et $\alpha_m = 2/3$. Les tangentes à la courbe sont tracées pour $O_q = 0.4$, 0.6 et 0.8, ainsi que pour $\alpha_m = 2/3$ et 0.8.*

formant glottique, situé dans les basses fréquences, reste limitée.

9.4.3 Mesure du seuil pour 2 valeurs de coefficient d'asymétrie

Résultats

Les 2 valeurs de coefficient d'asymétrie explorées sont : $\alpha_m = 2/3$ et $\alpha_m = 0.8$.

Les résultats pour chacun des 20 sujets non-entraînés sont présentés sur la figure 9.14. Le seuil moyen est $\Delta\alpha_m = 0.033$ pour $\alpha_m = 2/3$ et $\Delta\alpha_m = 0.027$ pour $\alpha_m = 0.8$. Les résultats obtenus par les 10 sujets entraînés sont présentés sur la figure 9.15. Le seuil moyen est de 0.027 pour $\alpha_m = 2/3$ et de 0.022 pour $\alpha_m = 0.8$. Enfin, la partie de droite de la figure 9.12 présente les résultats obtenus par les 7 sujets ayant une faiblesse auditive.

Discussion

Inversement à O_q, nous observons une diminution du seuil quand α_m augmente. Cet effet est significatif ($p < 0.05$) ainsi que l'indique le tableau 9.5, mais il est moins marqué que dans le cas du quotient ouvert. Le tableau 9.7 fournit les valeurs du seuil différentiel relatif pour les deux sessions considérées : $\alpha_m = 2/3$ (session 7) et $\alpha_m = 0.8$ (session 8). Il est difficile de discuter de la constance de ce rapport sur deux valeurs uniquement. Par contre, la différence sur le seuil différentiel relatif entre les 2 sessions devient alors très significative, ce qui laisse à penser que, contrairement au quotient ouvert, le seuil différentiel relatif de coefficient d'asymétrie n'est pas une constante.

Comme dans le cas du quotient ouvert, une modification de α_m entraîne une modification sous-jacente de la vitesse de fermeture E et donc de l'intensité sonore. D'après la courbe présentée sur la figure de droite 9.13, la pente de la tangente à la courbe $E = f(\alpha_m)$ augmente quand α_m augmente. Les variations d'intensité sonore seront donc d'autant plus amples que α_m sera élevé, ce qui peut expliquer qu'on en perçoive alors plus facilement les modifications. La fréquence du formant glottique est également modifiée par une variation de α_m : pour $O_q = 0.6$ et $f_0 = 196$ Hz, $F_g = 168$ Hz quand $\alpha_m = 2/3$ et 280 Hz quand $\alpha_m = 0.8$.

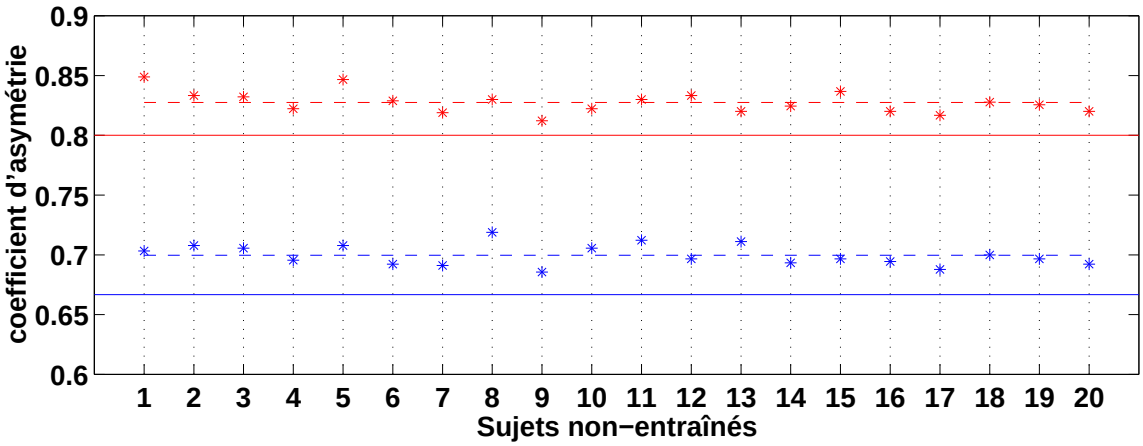


FIG. *9.14 – Valeurs du coefficients d'asymétrie associées aux seuils mesurés sur les* 20 *sujets non-entraînés pour* 2 *valeurs de* α_m *(*2/3 *et* 0.8*, ce qui correspond aux sessions* 7 *et* 8*). Les cibles sont indiquées par des lignes pleines et les valeurs moyennes pour l'ensemble du groupe par des lignes en pointillés.*

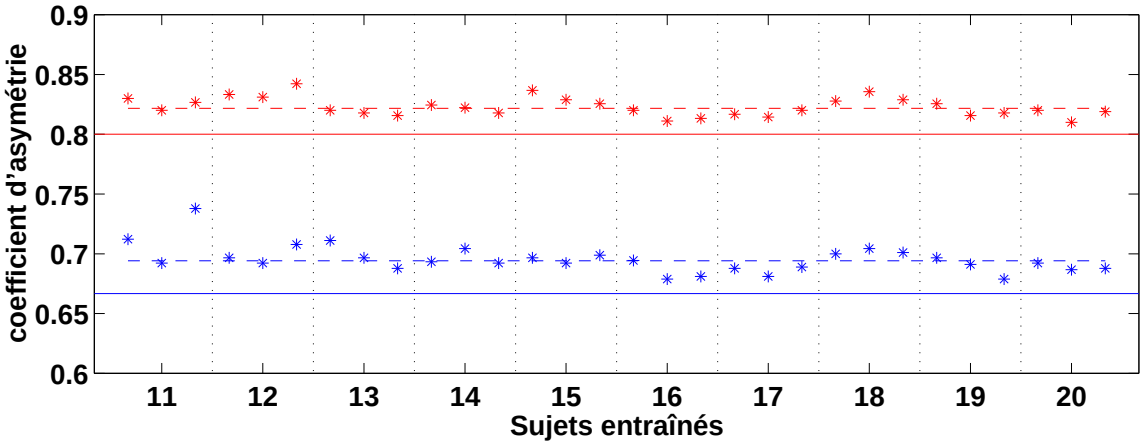


FIG. *9.15 – Valeurs du coefficients d'asymétrie associées aux seuils mesurés sur les* 10 *sujets entraînés pour* 2 *valeurs de* α_m *(*2/3 *et* 0.8*, ce qui correspond aux sessions* 7 *et* 8*). Les cibles sont indiquées par des lignes pleines et les valeurs moyennes pour l'ensemble du groupe par des lignes en pointillés.*

comparaisons	sujets non-entraînés	sujets entraînés
effet de la voyelle sessions $3 \leftrightarrow 9$	0.75	1.01 *
effet de la hauteur sessions $3 \leftrightarrow 5$	-0.42	-0.80
effet du vibrato sessions $3 \leftrightarrow 2$	1.91 *	2.18 **
comparaison A_v / E sessions $3 \leftrightarrow 4$	-8.55 ***	-3.57 *

TAB. *9.8 – Comparaisons entre deux sessions données, dans le cas des groupes de sujets non-entraînés et entraînés. Pour chaque groupe, les différences de seuil entre deux sessions sont moyennées sur l'ensemble des sujets. Cette différence moyenne entre 2 sessions données est exprimée en %. Une différence significative entre les deux sessions est marquée par * et correspond à* $p < 0.05$*. Une différence très significative est marquée par ** (*$p < 0.01$*) et fortement significative par *** (*$p < 0.001$*).*

9.4.4 Variations du seuil différentiel de quotient ouvert en fonction du contexte

Effet de la voyelle

Les sessions 3 et 9 sont identiques du point de vue des paramètres de source, mais la voyelle a été modifiée. La figure 9.16 présente les résultats dans le cas des sujets non-entraînés et des sujets entraînés. Nous remarquons que, selon les sujets, le seuil obtenu est plus petit pour l'une ou l'autre des deux voyelles sans qu'une tendance bien marquée ne se dégage. Ceci est confirmé par le tableau 9.8 qui ne montre pas de différence significative dans le cas des sujets non-entraînés, mais une différence significative ($p < 0.05$) dans le cas des sujets entraînés. Il semble donc y avoir un léger effet de la voyelle. Dans les deux cas, le seuil moyen est légèrement plus petit dans le cas de la voyelle [e] que dans le cas de la voyelle [a]. Ceci pourrait s'expliquer par le fait que, pour la voyelle [e], la fréquence du premier formant est proche de celle du formant glottique, ce qui n'est pas le cas pour la voyelle [a].

Effet de la hauteur

Deux hauteurs séparées par une quinte ont été étudiées : $f_0 = 196$ Hz dans la session 3 et $f_0 = 130$ Hz dans la session 5. La figure 9.17 présente les résultats obtenus par les sujets non-entraînés et les sujets entraînés dans ces deux sessions. Comme précédemment, les seuils moyens obtenus dans chacune des deux sessions ont des valeurs très proches et aucune tendance ne se dégage. Ceci est confirmé par le tableau 9.8 qui ne montre pas de différence significative dans le cas des sujets non-entraînés ou entraînés.

Effet du vibrato

L'ajout d'un vibrato à la synthèse d'une voyelle chantée rajoute toujours une impression de naturel. Les sessions avec vibrato et sans vibrato (session 2) sont comparées sur la figure 9.18 dans le cas des sujets non-entraînés et des sujets entraînés. Ces courbes reflètent les résultats montrés par le tableau 9.8 : nous observons une différence significative pour les sujets non-entraînés ($p < 0.05$) et très significative pour les sujets entraînés ($p < 0.01$). Les seuils obtenus pour la session sans vibrato sont nettement inférieurs aux seuils obtenus si l'on rajoute un vibrato.

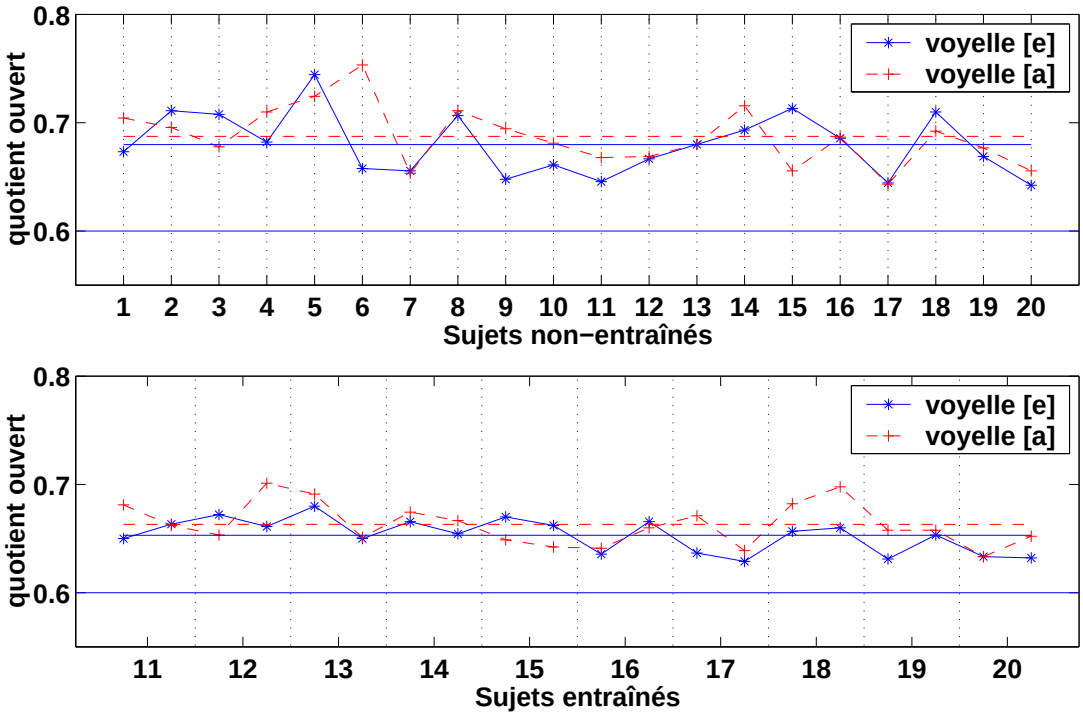


Fig. *9.16 – Comparaisons des seuils obtenus pour la voyelle [a] et la voyelle [e] par les sujets non-entraînés et les sujets entraînés.*

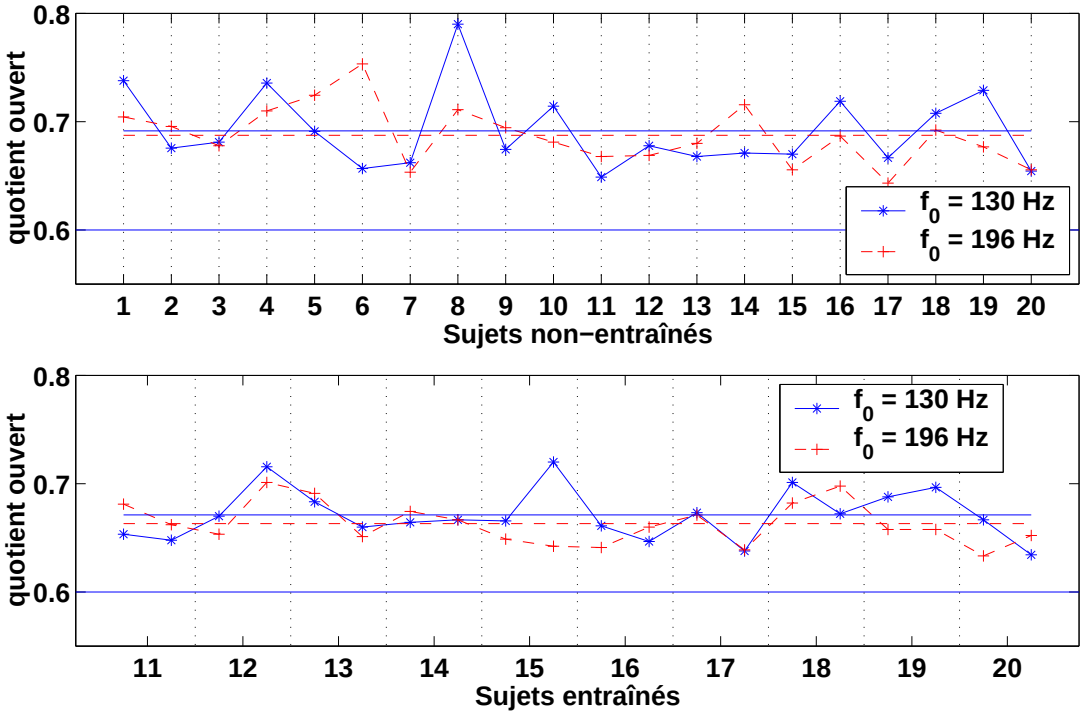


Fig. *9.17 – Comparaisons des seuils obtenus pour* $f_0 = 196$ Hz *(session 3) et* $f_0 = 130$ Hz *(session 5) par les sujets non-entraînés et les sujets entraînés.*

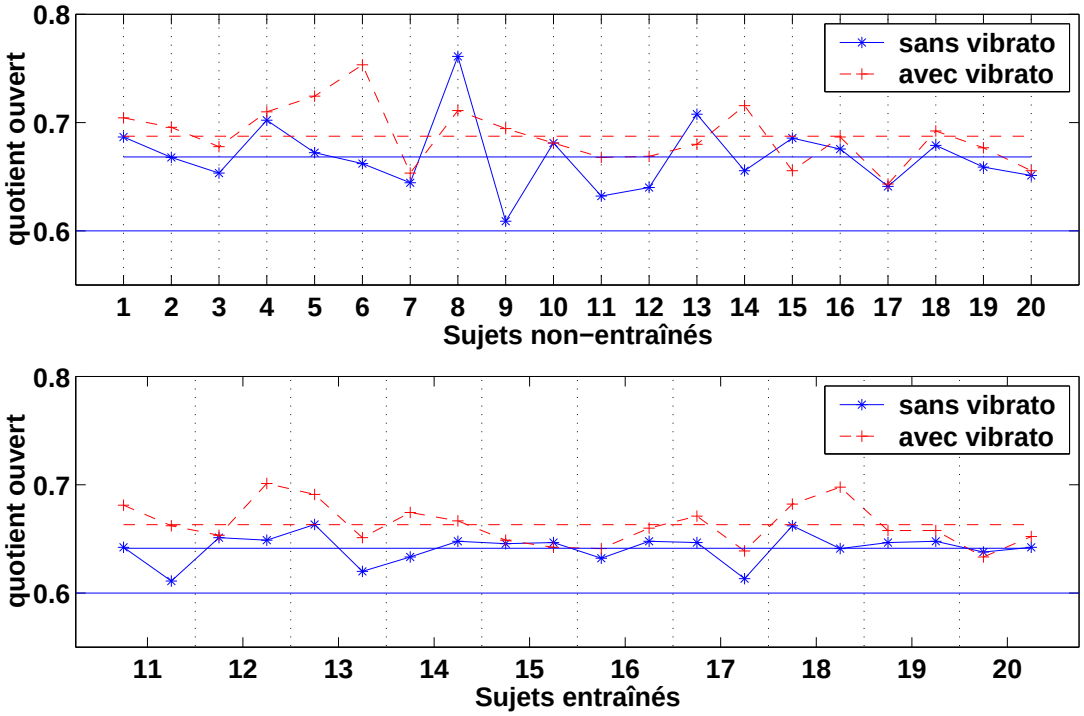


FIG. *9.18 – Comparaisons des seuils obtenus pour une fréquence fondamentale gardée constante (sans vibrato, session 2) ou variable (avec vibrato, session 3) par les sujets non-entraînés et les sujets entraînés.*

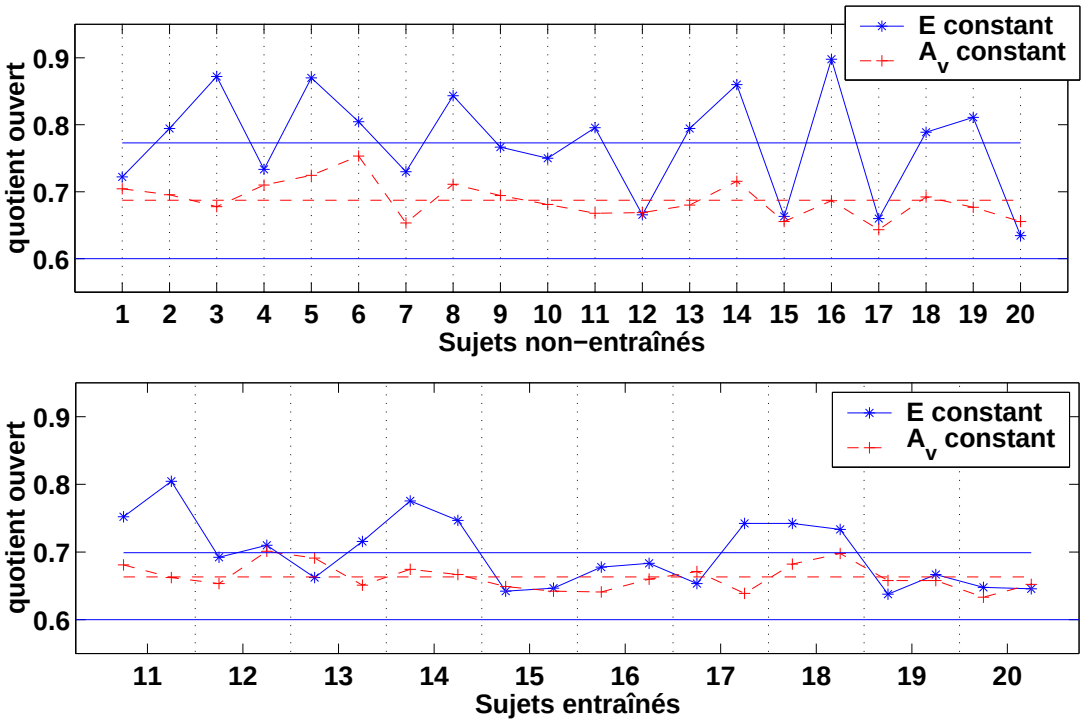


FIG. *9.19 – Comparaisons des seuils obtenus avec soit l'amplitude de voisement A_v constante (session 3), soit la vitesse de fermeture E constante (session 4) par les sujets non-entraînés et les sujets entraînés.*

Effet du paramètre d'amplitude

Nous avons vu au chapitre 3 en partie 3.2.2 que les effets spectraux du quotient ouvert et du coefficient d'asymétrie étaient différents selon que l'amplitude de voisement A_v ou que la vitesse de fermeture E sont gardées constantes : ces effets spectraux dépendent du jeu de paramètres considéré. Dans ce test, nous avons choisi par défaut de fixer A_v. Seule la session 4 fixe le paramètre E, ce qui implique une variation implicite de A_v. La figure 9.19 présente les résultats des seuils obtenus par les sujets non-entraînés et entraînés pour ces deux sessions. Nous observons de façon très nette que le seuil est fortement augmenté quand E est maintenu constant. Ce résultat est confirmé par le tableau 9.8 : la différence entre les deux sessions est fortement significative pour les sujets non-entraînés ($p < 0.001$) et significative pour les sujets entraînés ($p < 0.05$). Une telle différence était attendue puisque maintenir E constant revient à supprimer l'effet haute fréquence du quotient ouvert, et les différences entre stimuli ne sont alors marquées que dans les basses fréquences.

9.5 Conclusion

Nous avons prolongé par cette étude les travaux réalisés par Scherer & al [180] en 1998, en étudiant un plus grand nombre de cas et en nous plaçant dans une situation plus proche de la réalité. Nous observons que le seuil différentiel dépend du paramètre considéré et de sa valeur cible. Il peut dépendre également du contexte : nous avons noté un effet significatif du vibrato et du paramètre d'amplitude gardé constant (soit A_v, soit E). Par contre, ce seuil n'est pas modifié par la voyelle choisie ou par la fréquence fondamentale, pour $O_q = 0.6$ et $\alpha_m = 2/3$.

Nous pouvons maintenant apporter des éléments de réponse aux questions posées en introduction de ce chapitre.

- **Les variabilités observées sur les mesures de quotient ouvert dans le chapitre 8 sont-elles perceptibles ?**
  A la vue des seuils obtenus ici, cette variabilité sera effectivement perceptible dès qu'elle dépasse 14 % (en moyenne), ce qui est bien souvent le cas.

- **Les méthodes présentées au chapitre 7 sont-elles comparables ?**
  La méthode proposée par Howard [113, 112] et la méthode basée uniquement sur les signaux DEGG semblent comparables. Par contre, il peut exister de grandes différences entre ces méthodes et les méthodes de seuil sur signal EGG.

- **La variation de quotient ouvert au moment du passage entre mécanismes a-t-elle un sens du point de vue perceptif ?**
  Cela semble le cas, puisque cette différence est de l'ordre de 0.1 à 0.2, ce qui est souvent au-dessus des valeurs maximales des seuils détectés sur les sujets non-entraînés, quelle que soit la session. Néanmoins, il nous faut rester prudent car ces seuils ont été mesurés sur des sons fixes, alors que le passage entre mécanismes a souvent lieu dans une phrase musicale. D'autre part, lorsqu'un chanteur change de mécanisme, il se produit d'autres phénomènes plus importants perceptivement, comme par exemple un saut de fréquence, une chute d'intensité ou un défaut sur le vibrato.

Cette étude pourrait se compléter de diverses façons. Nous avons comparé les seuils différentiels de quotient ouvert sur deux voyelles et deux fréquences pour un jeu donné de paramètres de source glottique. Il serait intéressant d'explorer d'autres voyelles et d'autres hauteurs par rapport à ce même jeu ou un autre jeu de paramètres. En effet, l'analyse de l'influence du quotient ouvert et du coefficient d'asymétrie sur le formant glottique, effectuée en partie 3.2.3, montre que la fréquence de ce formant relativement à la fréquence fondamentale dépend en particulier de la valeur de ces deux paramètres. Plus le quotient ouvert sera faible et le coefficient d'asymétrie élevé, plus cette fréquence sera élevée et à même de venir interférer avec les fréquences des formants du conduit vocal. Pour certains jeux de paramètres, des différences

pourraient alors apparaître selon la voyelle ou la hauteur.

Nous avons noté la constance du seuil différentiel relatif de quotient ouvert, par rapport aux 3 valeurs cibles de O_q. Il serait intéressant de vérifier si la loi de Bouguer-Weber est également vérifiée dans d'autres cas (une autre voyelle, une autre hauteur, ...). En particulier, une extension importante de ce travail serait d'explorer les mêmes variations en fixant la vitesse de fermeture E. L'écart considérable sur le seuil différentiel entre A_v constant et E constant (8.55 % dans le cas des sujets non-entraînés) renforce l'idée que l'essentiel de l'effet d'un changement de O_q est le changement concomitant de E.

Cette étude a été effectuée sur des voyelles isolées de tout contexte. On peut s'attendre à ce que l'effet du contexte augmente encore la valeur de ces seuils puisque l'attention de l'auditeur ne serait alors plus focalisée sur une voyelle. Un prolongement de cette étude serait de reprendre ces expériences en plaçant le stimulus en contexte, c'est-à-dire en ne modifiant qu'une partie donnée d'une phrase parlée ou chantée et en reprenant le reste de la phrase à l'identique.

En Résumé

Un test perceptif a été mis en place dans le but de déterminer les seuils différentiels associés au quotient ouvert et au coefficient d'asymétrie. Cette étude se place dans la continuité d'une étude réalisée par Scherer & al [180] en 1998, où les seuils différentiels ont été explorés pour une valeur donnée de O_q (0.6) et une valeur donnée de α_m (2/3). Dans l'étude de Scherer, les stimuli sont soit des signaux de source glottique, soit des sons voisés avec $f_0 = 130.39$ Hz constant. Notre étude se propose de la compléter :

1. en s'appuyant sur des stimuli dont la qualité perceptive doit être proche de sons réels. Ceci implique en particulier que la fréquence fondamentale n'est pas une constante de la synthèse ;
2. en appliquant le test perceptif à plusieurs valeurs de quotient ouvert et de coefficient d'asymétrie ;
3. en testant les effets possibles d'autres facteurs entrant en jeu, comme le vibrato, la hauteur, la voyelle ou le paramètre d'amplitude.

Les paramètres de source glottique (A_v, O_q, α_m et Q_a) ainsi que le filtre modélisant le conduit vocal sont gardés constant pour la synthèse d'un stimulus donné. Pour améliorer autant que possible la qualité de la synthèse sonore, les instants de fermeture glottique sont détectés sur le signal voisé réel d'un chanteur et permettent une synthèse de qualité.

La méthode psychoacoustique choisie pour ce test est une méthode adaptative à choix forcé « 1-up 2-down », identique à celle choisie par Scherer & al [180] pour son étude. Elle a été implémentée en Matlab et ajoutée comme un module à l'environnement de tests perceptifs *LISE* (Rioux [167]). 9 sessions ont été mises en place, pour lesquelles soit le quotient ouvert, soit le coefficient d'asymétrie varie dans un cadre donné. 27 sujets ont passé le test. Parmi ces 27 sujets, 7 présentaient une faiblesse auditive dans les hautes fréquences. 10 sujets ont repassé le test perceptif 3 fois et constituent alors un groupe de sujets entraînés.

Les résultats obtenus dans cette étude sont les suivants :

1. Le seuil différentiel associé au quotient ouvert ou au coefficient d'asymétrie dépend de la valeur cible de ce paramètre. Ainsi, plus le quotient ouvert sera élevé, plus le seuil différentiel sera élevé. Plus le coefficient d'asymétrie sera élevé, plus le seuil différentiel sera bas.

   Ces observations s'expliquent par la variation sous-jacente de la vitesse de fermeture E (effet haute fréquence) et, dans une moindre mesure, par la modification de la position fréquentielle du formant glottique (effet basse fréquence).

   Le seuil différentiel relatif de quotient ouvert reste constant, ce qui est en accord avec la loi de Bouguer-Weber.
2. Dans le cas de sujets non-entraînés, le seuil sur O_q est de $\Delta O_q = 0.058$ pour $O_q = 0.4$, 0.087 pour $O_q = 0.6$ et 0.106 pour $O_q = 0.8$, ce qui amène à la relation : $\Delta O_q \simeq 0.14\, O_q$. Le seuil sur α_m est de 0.033 pour $\alpha_m = 2/3$ et 0.027 pour $\alpha_m = 0.8$.

   Dans le cas des sujets entraînés, les valeurs de seuil sont réduites de façon très significative. Le seuil sur O_q est de 0.037 pour $O_q = 0.4$, O.063 pour $O_q = 0.6$ et 0.079 pour $O_q = 0.8$, ce qui amène à la relation : $\Delta O_q \simeq 0.1\, O_q$. Le seuil sur α_m est de 0.027 pour $\alpha_m = 2/3$ et 0.022 pour $\alpha_m = 0.8$.
3. Dans cette étude, nous n'avons pas observé de différence significative sur le seuil différentiel à $O_q = 0.6$ dans le cas d'un changement de voyelle ou de hauteur. Par contre, nous avons observé une différence significative sur le seuil différentiel par ajout ou non du vibrato : $p < 0.05$ pour les sujets non-entraînés et $p < 0.01$ pour les sujets entraînés. Une différence significative apparaît également quand on fixe soit l'amplitude de voisement, soit la vitesse de fermeture : $p < 0.001$ pour les sujets non-entraînés et $p < 0.05$ pour les sujets entraînés.

Ces résultats permettent d'estimer dans quelle mesure la variation du quotient ouvert est perceptible. Nous pouvons alors utiliser ces résultats pour déterminer, par exemple, s'il existe des différences perceptibles dans le cas d'un changement de mécanisme laryngé, si des méthodes de mesure différentes sont comparables ou si la variabilité intra-individuelle et inter-individuelle observée dans la mesure du quotient ouvert a un sens du point de vue perceptif.

Exemples sonores

Le CD contenant ces exemples sonores est disponible sur demande au :
Laboratoire d'Acoustique Musicale (Université Paris 6 / case 161 / 4 place Jussieu / 75252 Paris cedex 05)

[1] crescendo sur un [e], contre-ténor CT2 (figure 1)

[2] glissando, soprano S3 (figure 1.6)

[3] arpège, voix synthétique avec $\mathbf{O_q = 0.4}$ ($A_v = 1$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$).

[4] arpège, voix synthétique avec $\mathbf{O_q = 0.8}$ ($A_v = 1$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$).

[5] arpège, voix synthétique avec $\alpha_\mathbf{m} = \mathbf{\frac{2}{3}}$ ($A_v = 1$, $O_q = 0.6$, $Q_a = 0$).

[6] arpège, voix synthétique avec $\alpha_\mathbf{m} = \mathbf{0.8}$ ($A_v = 1$, $O_q = 0.6$, $Q_a = 0$).

[7] arpège, voix synthétique avec $\mathbf{O_q = 0.4}$ ($E = 4000$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$).

[8] arpège, voix synthétique avec $\mathbf{O_q = 0.8}$ ($E = 4000$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$).

[9] arpège, voix synthétique avec $\alpha_\mathbf{m} = \mathbf{\frac{2}{3}}$ ($E = 4000$, $O_q = 0.6$, $Q_a = 0$).

[10] arpège, voix synthétique avec $\alpha_\mathbf{m} = \mathbf{0.8}$ ($E = 4000$, $O_q = 0.6$, $Q_a = 0$).

[11] arpège, voix synthétique avec $\mathbf{Q_a = 0.2}$ ($A_v = 1$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$).

[12] voyelle émise de façon relâchée à pressée par un locuteur (figure 4.14)

[13] crescendo, [u], do3, baryton B2 (figure 7.14)

[14] voyelle tenue, [a], la2, mezzo-soprano MS2 (figure 7.15)

[15] voyelle [a] chantée par le contre-ténor CT1 en mécanisme I (figure 8.3)

[16] même voyelle que l'exemple précédent, chantée sur une même hauteur par le contre-ténor CT1, mais en mécanisme II (figure 8.3)

[17] crescendo, [u], do4, soprano S1 (figure 8.9)

[18] phrase énoncée par le baryton B2 (figure 8.16)

[19] phrase chantée par le baryton B2 (figure 8.17)

[20] phrase criée par le baryton B2 (figure 8.18)

[21] phrase énoncée par le baryton B2 en mécanisme II (figure 8.19)

[22] phrase énoncée par la soprano S2 (figure 8.20)

[23] phrase chantée par la soprano S2 (figure 8.21)

[24] phrase criée par la soprano S2 (figure 8.22)

[25] phrase chantée par la soprano S2 avec tentative d'utilisation du mécanisme I (figure 8.23)

[26] phrase chantée par la soprano S2 en mécanisme I (figure 8.23)

[27] voyelle [a] chantée par le baryton B3 sur un sol2 ($f_0 = 196\ Hz$) (figure 9.5)

[28] exemple de synthèse avec $A_v = 1$, $O_q = 0.5$, $\alpha_m = 0.75$ et $Q_a = 0$ (figure 9.6)

[29] glissando chanté par le baryton B1 (figure F.1)

[30] glissando chanté par le baryton B1 (figure F.2)
[31] glissando chanté par le baryton B2 (figure F.3)
[32] glissando chanté par le baryton B2 (figure F.4)
[33] glissando chanté par le baryton B3 (figure F.5)
[34] glissando chanté par le baryton B4 (figure F.7)
[35] glissando chanté par le baryton B4 (figure F.8)
[36] glissando chanté par le baryton B5 (figure F.9)
[37] glissando chanté par le baryton B5 (figure F.10)
[38] glissando chanté par le baryton B6 (figures 7.10 et F.6)
[39] glissando chanté par le ténor T1 (figure F.11)
[40] glissando chanté par le ténor T2 (figures 7.9, 8.1 et F.12)
[41] glissando chanté par le ténor T2 (figure F.13)
[42] glissando chanté par le ténor T2 (figure F.14)
[43] glissando chanté par le contre-ténor CT1 (figure F.15)
[44] glissando chanté par le contre-ténor CT1 (figure F.16)
[45] glissando chanté par le contre-ténor CT1 (figures 8.2 et F.17)
[46] glissando chanté par le contre-ténor CT1 (figure F.18)
[47] glissando chanté par le contre-ténor C21 (figure F.19)
[48] glissando chanté par le contre-ténor CT3 (figure F.20)
[49] glissando chanté par le contre-ténor CT3 (figure F.21)
[50] glissando chanté par la mezzo-soprano MS1 (figure F.22)
[51] glissando chanté par la mezzo-soprano MS2 (figure F.23)
[52] glissando chanté par la mezzo-soprano MS2 (figure F.24)
[53] glissando chanté par la mezzo-soprano MS3 (figure F.25)
[54] glissando chanté par la soprano S1 (figure F.26)
[55] glissando chanté par la soprano S2 (figure F.27)
[56] glissando chanté par la soprano S2 (figure F.28)
[57] glissando chanté par la soprano S3 (figure F.29)
[58] glissando chanté par la soprano S3 (figure F.30)
[59] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B1 (figure G.1)
[60] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B2 (figure G.2)
[61] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B3 (figure G.3)
[62] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B4 (figure G.4)
[63] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B5 (figure G.5)
[64] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B6 (figure G.6)
[65] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le baryton B7 (figure G.7)
[66] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le ténor T1 (figure G.8)
[67] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le ténor T2 (figure G.9)
[68] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le contre-ténor CT1 (figure G.10)
[69] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le contre-ténor CT2 en mécanisme I (figure G.11)
[70] Premières mesures de l'*Ave Maria* de Gounod, interprétées par le contre-ténor CT2 en mécanisme II (figure G.12)

[71] Premières mesures de l'*Ave Maria* de Gounod, interprétées par la mezzo-soprano MS1 (figure G.13)
[72] Premières mesures de l'*Ave Maria* de Gounod, interprétées par la mezzo-soprano MS2 (figure G.14)
[73] Premières mesures de l'*Ave Maria* de Gounod, interprétées par la mezzo-soprano MS3 (figure G.15)
[74] Premières mesures de l'*Ave Maria* de Gounod, interprétées par la soprano S1 (figure G.16)
[75] Premières mesures de l'*Ave Maria* de Gounod, interprétées par la soprano S2 (figure G.17)
[76] Premières mesures de l'*Ave Maria* de Gounod, interprétées par la soprano S3 (figure G.18)

Conclusion générale de l'étude

Bilan de l'étude

Dans cette étude, nous nous sommes attachés à caractériser et décrire les paramètres de source glottique en voix parlée et chantée, en particulier le quotient ouvert et le coefficient d'asymétrie. Pour aborder cette étude sous plusieurs angles, nous avons utilisé 3 approches : une approche théorique, une approche expérimentale et une approche perceptive. Ces 3 approches sont complémentaires.

Approche théorique

Nous avons présenté au chapitre 3 un formalisme qui permet d'exprimer un modèle d'ODG à l'aide de 5 paramètres de source : la période fondamentale T_0, l'amplitude de voisement A_v ou la vitesse de fermeture E, le quotient ouvert O_q, le coefficient d'asymétrie α_m et le quotient de phase de retour Q_a. Dans le cadre de ce formalisme, nous avons formulé des expressions temporelles et spectrales, sur lesquelles nous nous sommes appuyées tout au long de cette étude pour étudier de façon analytique les effets temporels et spectraux de ces paramètres.

L'effet spectral de l'amplitude de voisement ou de la vitesse de fermeture a été souligné : une augmentation de ce paramètre de source glottique amène une amplification générale de l'intensité sur l'ensemble du spectre de source glottique. Les effets spectraux du quotient ouvert et du coefficient d'asymétrie sont dépendants du fait que l'on fixe l'un ou l'autre de ces deux paramètres. Dans le cas où A_v reste constant, une augmentation du quotient ouvert entraîne une atténuation spectrale des hautes fréquences et une translation de la position fréquentielle du formant glottique vers les basses fréquences. Une augmentation du coefficient d'asymétrie entraîne une accentuation spectrale dans les hautes fréquences, une augmentation de la largeur de bande du formant glottique et une légère augmentation de sa position fréquentielle. Dans le cas où E reste constant, les effets haute fréquence de ces deux paramètres de source glottique sont supprimés. Enfin, une augmentation du quotient de phase de retour entraîne une atténuation spectrale des hautes fréquences. Un quotient de phase de retour non nul correspond à l'ajout d'un filtre passe-bas dans les hautes fréquences, dont la fréquence de coupure diminue quand Q_a augmente.

La position fréquentielle du formant glottique dépend essentiellement des paramètres O_q et α_m mais peut également être modifiée légèrement par une variation de Q_a. Cette position est de l'ordre de la fréquence fondamentale, pour des conditions standard de phonation, c'est-à-dire O_q proche de 0.6, α_m proche de $2/3$ et Q_a nul. La fréquence du formant glottique peut atteindre 4 à 8 fois la fréquence fondamentale quand O_q diminue et α_m augmente.

Nous avons voulu étudier plus en détail les effets basse fréquence des paramètres de source glottique en nous focalisant sur la différence spectrale $(H_1 - H_2)$ entre les 2 premiers harmoniques, différence qui est souvent associée à une mesure du quotient ouvert (chapitre 4). Nous avons étudié cette différence spectrale sur un plan théorique et montré que, selon le modèle d'ODG employé, la relation entre O_q et $(H_1 - H_2)$ n'était pas forcément bijective. Pour une valeur donnée de $(H_1 - H_2)$, il peut exister une plage

de variation du quotient ouvert, en fonction des valeurs prises par le coefficient d'asymétrie. Si α_m reste constant, comme c'est le cas dans le modèle KLGLOTT88, la relation entre O_q et $(H_1 - H_2)$ devient bijective. Dans la réalité, le coefficient d'asymétrie n'est pas une constante de la phonation, mais, par contre, les paramètres de source peuvent varier conjointement et ne prendre que des valeurs situées dans certaines plages précises en fonction du contexte. Ainsi, Fant a pu proposer une relation bijective liant O_q et $(H_1 - H_2)$ sur la base de mesures effectuées dans le cadre de la parole. Cette relation semble indiquer que, lorsque le quotient ouvert a des valeurs basses, le coefficient d'asymétrie a des valeurs élevées (entre 0.7 et 0.8) et le quotient de phase de retour a des valeurs proches de 0. Quand le quotient ouvert augmente, les valeurs possibles de α_m diminuent et le quotient de phase de retour augmente progressivement.

Des mesures spectrales de $(H_1 - H_2)$ ont été effectuées sur des signaux réels de parole et de chant, afin de déterminer des plages de variation de cette différence spectrale et de discuter les possibilités et limites des modèles d'ODG. Nous nous sommes appuyés sur une étude du filtrage inverse présentée en partie 4.2 du chapitre 4. Les méthodes de prédiction linéaire pour l'estimation de la source glottique par filtrage inverse ont été décrites. Des tests comparatifs ont été effectués sur des signaux synthétiques dans le cas des 3 voyelles [a], [i] et [u]. Ils ont montré que les méthodes par prédiction linéaire permettent d'estimer de façon très précise les fréquences et largeurs de bande des formants dans les hautes fréquences, mais que le premier formant F_1 est souvent mal estimé par ces méthodes. La prédiction linéaire avec modèle de source peut alors donner de meilleurs résultats qu'une prédiction linéaire classique. Néanmoins, cette méthode nécessite la connaissance précise des instants de fermeture ou d'ouverture glottique. Nous avons comparé les signaux de source glottique obtenus par application de ces méthodes de filtrage inverse dans le cas de signaux de parole et de chant. Nous avons observé que, de manière générale, il est difficile d'obtenir une phase fermée bien marquée sur des tels signaux. Des différences marquées apparaissent également sur la forme de l'ODG estimée selon la méthode employée, surtout dans le cas des voyelles ayant un premier formant bas comme le [i] et le [u], puisque le premier formant est souvent mal estimé. Dans le cadre du chant, nous avons observé qu'une première atténuation du signal pouvait être nécessaire pour amener à des ODG estimées plus proche de ce que l'on attendrait d'après la théorie. Cela laisse supposer que les modèles d'ODG ne seraient alors plus valables dans le cas du chant.

Les mesures spectrales de $(H_1 - H_2)$ effectuées avec l'une ou l'autre de ces méthodes montrent que les valeurs mesurées ne sont pas toujours contenues dans les domaines de variation théoriques calculés à partir des modèles d'ODG. Ceci est surtout marquant pour le chant, où l'amplitude de variation de $(H_1 - H_2)$ peut être considérable, dépassant plus que largement les bornes des domaines théoriques dans le cas du mécanisme I. Les limitations du filtrage inverse pourraient expliquer en partie ces résultats. Néanmoins, cela renforce également l'impression que les modèles d'ODG appliqués dans le cadre de la parole ne sont pas à même de reproduire les productions vocales chantées. En particulier, ces modèles s'appuient sur la théorie source-filtre, qui suppose la source sans interaction avec le filtre, alors que les chanteurs cherchent justement à ajuster leurs cavités résonantielles pour amplifier l'intensité vocale perçue, en faisant coïncider les harmoniques du son produit avec la fréquence des formants.

Approche expérimentale

Les résultats des mesures de $(H_1 - H_2)$ dans le cadre du chant nous rappellent que les modèles d'ODG ne sont que des modèles, qu'ils ont donc leurs limitations et qu'ils ne sont pas toujours à même de traduire toutes les possibilités vocales d'une production réelle. En particulier, dans une approche théorique, nous pouvons étudier les variations des paramètres de source en les supposant indépendants les uns des autres, ce qui ne doit pas faire oublier que ces paramètres ne sont vraisemblablement pas indépendants dans la réalité, mais qu'ils varient conjointement. Nous avons donc effectué une étude expérimentale du quotient ouvert en fonction de l'intensité vocale et de la fréquence fondamentale. Ces mesures ont été effectuées à partir de l'électroglottographie, une méthode d'investigation du mouvement glottique simple et non-invasive dont l'avantage majeur est qu'elle permet la caractérisation de certains aspects du mouvement glottique sans avoir recours au filtrage inverse.

N'ayant pas à disposition de base de données exploitable dans le cadre du chant, un protocole d'une

durée de 20 mn a été élaboré, dans le but d'amener le chanteur à interpréter, dans sa production vocale naturelle et non-contrainte, des vocalises, des voyelles tenues, des sons filés et des phrases musicales sur différentes intensités, différentes hauteurs et différents timbres. Une attention toute particulière a été portée sur la détermination du mécanisme laryngé dans lequel avait lieu la production vocale, par écoute et discussion avec le chanteur. Les signaux acoustiques et électroglottographiques de 18 chanteurs ont ainsi été enregistrés : 7 barytons, 2 ténors, 3 contre-ténors, 3 mezzo-sopranos et 3 sopranos.

A l'aide de ces enregistrements, la variation du quotient ouvert a pu être étudiée en fonction de la fréquence fondamentale et de l'intensité vocale (qui est fortement corrélée à la vitesse de fermeture E). La mesure du quotient ouvert nécessite la connaissance précise des instants d'ouverture et de fermeture glottique. Nous nous sommes intéressés à la dérivée des signaux électroglottographiques, car elle peut présenter des pics marqués qui coïncident alors avec ces instants. La détection des pics d'ouverture et fermeture glottique sur les signaux DEGG permet une détermination très fiable de la fréquence fondamentale et du quotient ouvert dans le cas où ces pics sont bien marqués. Dans un premier temps, nous avons exploré la base de données relativement à la qualité de ces pics et nous avons noté qu'il existe de nombreux cas pour lesquels ces pics sont dédoublés soit à l'ouverture, soit à la fermeture. Quelques cas typiques ont été décrits qualitativement, comme par exemple le passage entre mécanismes. Nous avons alors mis en place une procédure de sélection des signaux exploitables pour la mesure de fréquence fondamentale et de quotient ouvert, basée sur la détection de la multiplicité du pic d'ouverture et du pic de fermeture. Une méthode indirecte de mesure du quotient ouvert par intercorrélation entre les parties positives et négatives du signal DEGG a également été présentée. La comparaison des méthodes par électroglottographie et par filtrage inverse du signal acoustique dans le cas de la parole montre que la mesure de quotient ouvert est plus fiable si elle prend en compte les signaux DEGG. Les méthodes par détection de seuil sur les signaux EGG sont néanmoins plus robustes et nous pouvons envisager de compléter la méthode basée sur les signaux DEGG par une méthode de seuil pour la détection de l'instant d'ouverture quand le pic d'ouverture est peu marqué ou dédoublé.

Nous avons enfin présenté les résultats des mesures de quotient ouvert effectuées sur une partie des signaux fournies par la base de données. Nous avons tout d'abord montré l'importance des mécanismes laryngés et leur influence très marquée sur la valeur de quotient ouvert mesurée. Ainsi, le mécanisme I se traduit de façon générale par des quotients ouverts plus faibles qu'en mécanisme II. L'observation des valeurs de O_q peut alors apporter des indications concernant le mécanisme utilisé pour une production vocale donnée, mais ne permet pas une détermination sûre de la nature de ce mécanisme. Sur les glissandos effectués par les chanteurs, la transition entre mécanismes, qui peut ou non s'accompagner d'un saut brusque de la fréquence fondamentale, est généralement marquée par un saut de quotient ouvert. Cette variation du quotient ouvert en dehors d'une variation marquée d'intensité vocale ou de fréquence fondamentale peut alors donner une indication sur le mécanisme laryngé et elle confirme la réalité physiologique des mécanismes laryngés. Il nous paraît donc nécessaire de spécifier dans toute étude portant sur le quotient ouvert le mécanisme dans lequel a lieu la production.

Nous avons noté une très grande variabilité de comportement entre les chanteurs, que ce soit au sein d'une même catégorie vocale ou entre catégories vocales. Au-delà de cette variabilité, des tendances générales se dégagent néanmoins. La corrélation possible entre quotient ouvert et intensité vocale a été étudiée. En particulier, nous avons identifié une corrélation forte dans le cas du mécanisme I, mais aucune corrélation dans le cas du mécanisme II. A l'inverse, nous avons mis en évidence une corrélation forte entre quotient ouvert et fréquence fondamentale dans le cas du mécanisme II, qui ne se trouve pas dans le cas du mécanisme I. Nous avons également visualisé les variations possibles du quotient ouvert en fonction de la voyelle demandée au chanteur et observé que le quotient ouvert avait généralement des valeurs plus élevées pour une voyelle [a] que pour les voyelles [e] et [u], dans le cas du mécanisme I. Dans le cas du mécanisme II, il ne semble pas y avoir d'effet par rapport à la voyelle demandée. Les 3 productions vocales parlée, chantée et criée ont été illustrées sur l'exemple d'un baryton et d'une soprano. Nous avons observé des similitudes sur le quotient ouvert entre voix parlée et voix chantée, bien qu'en voix parlée, la dynamique de variation du quotient ouvert soit accentuée puisque le débit est plus rapide. De grandes variations de quotient ouvert ont été observées dans le cas de la voix criée. Quelle que soit la production, nous avons noté que les attaques, les fins de phrase et les transitions syllabiques se traduisent par une augmentation

ou une diminution rapide du quotient ouvert. Enfin, nous avons observé un vibrato de quotient ouvert qui se met en place en même temps que le vibrato de fréquence fondamentale, indépendamment du mécanisme laryngé et du chanteur. Ce vibrato de quotient ouvert semble comparable en amplitude relative au vibrato de fréquence fondamentale et il semble exactement en opposition de phase.

Approche perceptive

Nous avons enfin procédé à une étude perceptive du seuil minimal de détection des variations du quotient ouvert et du coefficient d'asymétrie afin de déterminer dans quelle mesure la variation de l'un ou l'autre de ces paramètres pouvait amener à une différence effectivement perceptible sur un son voisé donné. Le test perceptif a été élaboré dans la continuité d'une étude réalisée par Scherer & al [180] (1998) pour $O_q = 0.6$ et $\alpha_m = \frac{2}{3}$. Nous avons repris la même méthode adaptative à choix forcé « 1-up 2-down », que nous avons implémentée sous forme d'un module Matlab. Ce module a été ajouté à la structure de l'environnement de tests perceptifs LISE. Les stimuli ont été synthétisés en mettant l'accent sur le côté naturel du son. Pour ce faire, nous avons utilisé pour la synthèse la connaissance des instants de fermeture glottique détectés sur un signal de voix chantée réel, les autres paramètres de source glottique et la fonction de transfert modélisant le conduit vocal étant gardés constants.

Le seuil différentiel a été déterminé pour 3 valeurs de O_q (0.4, 0.6 et 0.8) et 2 valeurs de α_m (2/3 et 0.8). Il dépend de la valeur cible du paramètre de source glottique. Il augmente quand O_q augmente ou quand α_m diminue. Par comparaison des seuils déterminés dans deux situations différentes, les effets possibles d'un changement de voyelle, de hauteur, d'une suppression du vibrato ou du choix du paramètre d'amplitude gardé constant ont été étudiés. Il n'a pas été observé d'effet très significatif de la voyelle (comparaison [a] / [e]) ni de la fréquence fondamentale (comparaison $f_0 = 130$ Hz / 196 Hz). Par contre, un effet très significatif du vibrato et du paramètre d'amplitude a été observé . Quand la fréquence fondamentale est gardée constante (suppression du vibrato), le seuil différentiel diminue. Quand la vitesse de fermeture E est gardée constante, le seuil différentiel augmente par rapport à la situation où l'amplitude de voisement A_v est gardée constante.

Discussion

A la lumière de ces résultats, nous nous proposons d'apporter à présent des réponses aux questions que nous avons soulevées en introduction de ce travail.

- **Peut-on caractériser spectralement les effets perceptifs de la source glottique ?**

  Il est possible de caractériser ces effets par une analyse théorique basée sur des modèles d'ODG essayant de reproduire autant que possible des réalités physiques. Une analyse perceptive conjointe est alors nécessaire pour déterminer les facteurs pertinents. Il faut être conscient des limites, à savoir que :
  - les effets spectraux d'un paramètre donné peuvent dépendre du jeu de paramètres considéré ;
  - dans la réalité, les paramètres vont varier conjointement, alors qu'ils peuvent être considérés comme indépendants dans une étude théorique. L'effet spectral suscité par la modification d'un paramètre peut très bien être compensé par la modification parallèle d'un autre paramètre ;
  - les modèles ne sont pas toujours à même de représenter toute la variabilité possible de la production vocale humaine.

  Les effets perceptifs de la source glottique peuvent aussi se déduire de mesures spectrales effectuées sur des signaux voisés réels. Cette approche a également ses limites. L'appareil vocal est un instrument difficilement accessible et seules des méthodes indirectes sont applicables pour appréhender une production vocale naturelle. En premier plan de ces méthodes se place le filtrage inverse, qui est lui-même basé sur une théorie linéaire de la production vocale. Appliquer un filtrage inverse pour estimer la source glottique implique donc déjà d'avoir formulé certaines hypothèses sur son

comportement. L'électroglottographie se place alors en avant plan des méthodes non-invasives qui permettent d'explorer le mouvement vibratoire glottique de façon décorrélée du système supra-glottique. Cette méthode nécessite d'être utilisée avec précaution et de savoir exactement ce qui est mesuré et quel est l'intervalle de confiance associé à ces mesures. Enfin, il faut être très prudent dans ses conclusions. Si l'on observe la variation d'un paramètre de source glottique et conjointement un effet spectral sur la source, rien ne prouve qu'il s'agisse d'une relation de cause à effet, pour peu qu'un paramètre non pris en compte varie de façon sous-jacente.

- **Peut-on estimer la source glottique à partir des théories et des modèles actuels ?**

  Dans le cadre de la parole, nos résultats semblent confirmer que les modèles actuels permettent d'estimer la source glottique de façon tout à fait convenable. Il en va tout autrement pour le chant, pour lequel on touche très sûrement aux limites de la théorie source-filtre. En parole, l'énergie est concentrée dans des fréquences situées en-dessous de 4000 Hz alors que, dans le chant, on retrouve de l'énergie jusqu'à près de 16000 Hz. En particulier, l'hypothèse d'onde plane n'est plus valide pour des fréquences supérieures à quelques 6000 Hz. Une meilleure compréhension des phénomènes d'interaction entre les cavités sous- et sus-glottiques et de l'interaction de l'onde acoustique avec les parois semble alors nécessaire pour pouvoir améliorer notre compréhension et notre modélisation de la source glottique dans le cas de ces productions vocales.

- **Quels sont les paramètres les plus pertinents pour décrire la source glottique ? Quelles sont leurs variations dans un contexte vocal ?**

  La fréquence fondamentale est évidemment un paramètre de premier plan. Nous avons observé l'importance des petites variations de la fréquence fondamentale sur la qualité de la synthèse. Quand on ajoute un vibrato constant sur un signal voisé synthétique, on augmente le naturel perçu de façon très nette. Quand, de plus, on utilise comme fréquence fondamentale celle mesurée sur un signal réel par détection des instants de fermeture glottique, la synthèse paraît alors très naturelle, même si les autres paramètres de source glottique et le filtre modélisant le conduit vocal sont gardés constants. Il est difficile de réaliser une synthèse de voix chantée qui semble naturelle. Des règles précises sur les variations de f_0 par ajout d'un vibrato mais également par ajout de petites variations aléatoires devraient permettre d'apporter des améliorations notables aux systèmes actuels.
  Nous pensons que la pertinence des autres paramètres de source dépend du mécanisme laryngé. Dans la littérature, la pertinence des paramètres de source glottique tels que la vitesse de fermeture et le quotient de phase de retour a été soulignée, car ces paramètres ont une influence dans les hautes fréquences auxquelles l'oreille humaine est particulièrement sensible. La vitesse de fermeture est vraisemblablement corrélée à l'intensité vocale et ses valeurs peuvent être ajustées par modification de l'amplitude de voisement, du quotient ouvert, du coefficient d'asymétrie ou de la fréquence fondamentale. Existe-t'il une relation qui lie ces paramètres entre eux ?
  Dans cette étude, nous n'avons étudié le lien qu'entre trois de ces paramètres : la fréquence fondamentale, le quotient ouvert et l'intensité vocale (donc indirectement, la vitesse de fermeture). Ainsi le quotient ouvert varie en fonction de l'intensité vocale en mécanisme I, alors qu'il reste relativement stable en mécanisme II. Malgré ces observations, nous ne sommes pas en mesure d'affirmer que le quotient ouvert soit un paramètre pertinent en mécanisme I, puisque ses variations peuvent être le fait des variations d'un ou plusieurs autres paramètres que nous n'avons pas mesurés. Si l'on se réfère aux travaux de Fant, le paramètre $R_d = \frac{A_v}{E}\frac{f_0}{110}$ semble être le paramètre de source glottique pertinent, à l'aide duquel les variations de qualité vocale pourraient être modélisées. Ce paramètre relie les variations de O_q, α_m et Q_a et mériterait une étude approfondie dans le cadre de la voix chantée.

Ce travail s'inscrit dans le cadre d'une étude très générale de la qualité vocale et des paramètres de source qui permettent de la décrire. La dernière question posée en introduction était de savoir s'il était ou non possible de caractériser une qualité donnée à l'aide d'un jeu unique de paramètres. Cela nous semble

effectivement possible. Les aspects de modification de la qualité par changement de l'intensité vocale ont été abordés et des paramètres de source glottique comme la fréquence fondamentale, le quotient ouvert et indirectement la vitesse de fermeture par l'intensité vocale, ont été mesurés et comparés. Les changements de qualité vocale par modification du timbre n'ont été abordés que sous un aspect théorique. Nous avons développé dans cette étude tout un contexte théorique et expérimental qui peut à présent s'appliquer à la caractérisation de la source glottique pour des productions vocales de toutes sortes.

Perspectives

De nombreux points n'ont pu être abordés dans cette étude et mériteraient d'être développés dans une étude ultérieure.

Analyse des extraits de la base de données portant sur les changements de timbre

Ainsi que nous venons de le mentionner, ce travail peut s'appliquer à la caractérisation de la qualité vocale en termes de variation des paramètres de source glottique. En particulier, les enregistrements portant sur des changements de qualité vocale liés au timbre effectués par les 18 chanteurs n'ont pas été analysés. Dans un premier temps, les chanteurs avaient la liberté de choisir les qualités vocales qui les inspiraient, de les nommer et de les illustrer. Perceptivement, d'un chanteur à l'autre, des types de qualité se retrouvent et ils sont souvent décrits par les différents chanteurs à l'aide d'un vocabulaire identique et précis (voix blanche, voix sombre, ...). L'étude de ces exemples permettrait de déterminer quels sont les paramètres sur lesquels ces chanteurs peuvent avoir une action, de façon directe ou indirecte, pour modifier leur production vocale et de quelle façon ces paramètres varient pour une qualité donnée. Dans un second temps, les deux qualités vocales « voix naturelle » / « voix lyrique » étaient imposées au chanteur. L'étude de ces exemples permettrait peut-être de déterminer les paramètres qui différencient une voix travaillée d'une voix non-travaillée. D'autres types de productions vocales ont également été enregistrés en cours de thèse : des exemples de chant diphonique, de chant grégorien et de chant bulgare. L'analyse de ces exemples permettrait d'apporter des indications quand à ces productions vocales particulières. Le mécanisme I chez la femme mériterait également une étude poussée, comme dans le cas des voix d'altos féminins et des voix de variété, qui sont toutes deux des productions vocales féminines utilisant le mécanisme I. L'intérêt d'une telle analyse serait de savoir s'il existe des différences homme/femme au niveau du mécanisme I chez des chanteurs habitués à l'utiliser. Dans le cas du mécanisme II, nous avons vu des similitudes entre la production vocale des contre-ténors, celle des mezzo-sopranos en mII et celle des sopranos.

Analyse spectrale et modification de la qualité vocale par filtrage

Au vue de l'analyse des effets spectraux des paramètres de source effectuée sur les modèles d'ODG, l'effet de la source glottique sur le spectre d'un signal voisé semble être comparable à celui d'un filtre passe-bas du second ou du troisième ordre (selon que l'on suppose une fermeture abrupte ou non) dont la fréquence de coupure et la largeur de bande sont ajustables. Si la source glottique pouvait effectivement être modélisée par un tel filtre, cela permettrait d'envisager de nombreuses applications dans le domaine de la modification vocale. Il serait en effet très simple et peu coûteux en temps de calcul de filtrer simplement un signal voisé donné par un filtre bien choisi pour permettre la modification de qualité vocale à partir de variations des paramètres de source glottique. Un travail d'analyse spectrale par application d'un modèle linéaire de source glottique a été entrepris au cours de cette thèse (Henrich & al [103], 1999) sur la base des travaux développés par Doval & d'Alessandro doval97a,doval97b (1997) mais n'a pas été présenté ici faute de résultats pertinents. Il serait néanmoins intéressant de poursuivre dans cette voie, en simulant et analysant des modifications vocales sur des signaux voisés synthétiques, et en testant de façon perceptive des modifications par filtrage de signaux réels.

Etude des apériodicités

Nous n'avons pas abordé ici l'étude des apériodicités qui peuvent influencer la qualité vocale. Pourtant, la partie apériodique d'un signal acoustique y contribue, en particulier dans les voix de faible intensité, les voix soufflées ou murmurées (d'Alessandro & Doval [46]). Nous avons noté au chapitre 9 que l'on pouvait améliorer le naturel d'une synthèse vocale si l'on gardait comme information pour la synthèse les instants de fermeture glottique détectés sur un signal réel. Cela implique de prendre en compte les petites apériodicités du mouvement vibratoire glottique. Il serait donc important de prolonger cette étude par une caractérisation de l'apériodicité dans les signaux exploités.

De l'utilisation des signaux DEGG

L'utilisation des signaux DEGG est appréciable pour la visualisation et la caractérisation du mouvement glottique : ils apportent souvent des informations complémentaires à celles obtenues sur les signaux EGG eux-mêmes. Pourtant, ils sont très peu utilisés dans la littérature, vraisemblablement par manque d'une étude poussée. Nous avons proposé dans ce travail une description des formes caractéristiques de ces signaux, d'après ce que nous avons observé dans la base de données. Nous avons formulé des hypothèses pour tenter d'expliquer les dédoublements observés des pics d'ouverture et de fermeture. Ce travail doit à présent être complété par la visualisation directe du mouvement glottique et une collaboration est en cours pour appliquer la technique récente de transillumination strobophotoglottographique à la détermination précise des instants d'ouverture et de fermeture glottique par comparaison avec les signaux DEGG. Nous envisageons également d'appliquer cette technique de visualisation à la reconstruction partielle du mouvement vertical des cordes vocales avant l'ouverture et après la fermeture, par analyse de la densité lumineuse diffusée par la glotte fermée.

De nombreux aspects de la production vocale humaine n'ont été que peu ou pas abordés dans cette étude. Nous nous sommes principalement attachés à décrire la source glottique en dehors d'un contexte articulatoire et phonétique. Néanmoins, il faut garder à l'esprit que la communication vocale s'appuie sur une covariation de la source glottique et de la fonction de transfert du conduit vocal (Fant [69], 1995). La source glottique peut être influencée par des constrictions au niveau du système sus-glottique (Bickley & Stevens, 1986), par un couplage entre les systèmes sus- et sous-glottiques. Elle peut avoir une influence sur la largeur de bande et la fréquence du premier formant. Les bruits d'aspiration et le rapport signal/bruit sont souvent des données d'importance (Fant [71], 2000). Dans le chant classique, nous pouvions les négliger, mais dans l'étude d'autres productions vocales, ces paramètres sont assurément à prendre en compte.

En conclusion, ce travail contribue à une meilleure connaissance des paramètres de source glottique qui gouvernent la qualité vocale. Dans sa pratique vocale, le chanteur ne se préoccuppe pas de quotient ouvert ou de coefficient d'asymétrie mais il règle d'autres paramètres, tels l'intensité, la hauteur ou la richesse spectrale. Par notre approche, nous avons donc tenté de relier ces aspects perceptifs à des paramètres physiques.

Annexe A

Expressions temporelles et spectrales des modèles d'ODG

Doval & d'Alessandro [54, 55] ont établi un formalisme permettant d'exprimer de façon générique les formes temporelles et spectrales des modèles d'ODG et d'ODGD. D'après cette étude, nous avons retenu les 3 modèles suivants : le modèle KLGLOTT88 proposé par Klatt & Klatt [132] (1990), référencé couramment dans la littérature, le modèle R++ proposé par Veldhuis [213] (1998), qui est une version étendue du modèle KLGLOTT88, ainsi que le modèle LF proposé par Liljencrants, Fant & Lin [72] (1985), qui est actuellement le modèle le plus utilisé et, comparativement aux deux précédents, le plus complet. Nous nous proposons d'expliciter les expressions normalisées, temporelles et spectrales, de ces 3 modèles.

Modèle KLGLOTT88

Ce modèle a été élaboré à partir du modèle Rosenberg B (Rosenberg [169], 1971). Il a été utilisé dans le synthétiseur de Klatt ([133], [132]), ainsi que dans de nombreuses études par la suite (Hanson [97, 98, 99]). Les fonctions génératrices de l'ODG et de l'ODGD sont données par :

domaine temporel	**domaine spectral**
$n_g(t) = \begin{cases} \frac{27}{4}(t^2 - t^3) & pour\ 0 \leq t \leq 1 \\ 0 & sinon \end{cases}$	$N_g(f) = \frac{27}{4} \frac{1}{(j2\pi f)^2} \left(e^{-j2\pi f} + \frac{2(1+2e^{-j2\pi f})}{j2\pi f} - \frac{6(1-e^{-j2\pi f})}{(j2\pi f)^2} \right)$
$n_g'(t) = \begin{cases} \frac{27}{4}(2t - 3t^2) & pour\ 0 \leq t \leq 1 \\ 0 & sinon \end{cases}$	$N_g'(f) = j2\pi f\ N_g(f)$

On remarque que le paramètre α_m n'intervient pas dans ces formules : en effet, ce paramètre est une constante du modèle. Sa valeur est fixée à $\alpha_m = \frac{2}{3}$. Cela signifie que l'asymétrie de l'impulsion ne peut pas être modifiée à partir de ce modèle.

La vitesse de fermeture normalisée e_n et l 'intégrale de l'impulsion normalisée j_n et sont données par :

$$e_n = \frac{27}{4} \quad , \quad j_n = \frac{9}{16}$$

La fréquence normalisée f_g et l'amplitude normalisée a_g du formant glottique sont données par :

$$f_g = \frac{\sqrt{3}}{\pi} \quad , \quad a_g = \frac{9\sqrt{3}}{8}$$

Modèle R++

Le modèle R++ (Rosenberg ++, Veldhuis [213] 1998) s'appuie également sur le modèle Rosenberg B, mais il diffère du modèle KLGLOTT88 par la prise en compte de l'asymétrie de l'impulsion, qui peut être modifiée. L'objectif de ce modèle est de présenter une alternative au modèle LF que nous décrirons juste après. Il est comparable sur le plan perceptif mais plus facilement implémentable (gain en terme de temps de calcul). Les fonctions génératrices de l'ODG et de l'ODGD sont données par :

domaine temporel

$$n_g(t,\alpha_m) = \begin{cases} \frac{1}{\alpha_m^3(1-\alpha_m)^2}\, t^2(t-1)\left((2-3\alpha_m)t - \alpha_m(3-4\alpha_m)\right) & pour\ 0 \le t \le 1 \\ 0 & sinon \end{cases}$$

$$n_g'(t,\alpha_m) = \begin{cases} \frac{1}{\alpha_m^3(1-\alpha_m)^2}\, 4t(t-\alpha_m)\left((2-3\alpha_m)t - \frac{3-4\alpha_m}{2}\right) & pour\ 0 \le t \le 1 \\ 0 & sinon \end{cases}$$

domaine spectral

$$N_g(f) = \frac{2}{\alpha_m^3(1-\alpha_m)^2}\frac{1}{(j2\pi f)^2}\left[e^{-j2\pi f}(1-\alpha_m)(2\alpha_m-1) + \frac{1}{j2\pi f}\left(\alpha_m(3-4\alpha_m) - e^{-j2\pi f}(8\alpha_m^2 - 15\alpha_m + 6)\right) - \frac{6}{j2\pi f^2}\left(1 - 2\alpha_m^2 + e^{-j2\pi f}(2\alpha_m^2 - 6\alpha_m + 3)\right) + \frac{12}{j2\pi f^3}(1-e^{-j2\pi f})(2-3\alpha_m)\right]$$

$$N_g'(f) = j2\pi f\ N_g(f)$$

Pour respecter la propriété de positivité de l'ODG ($n_g(t) \ge 0$), ce modèle impose que $\alpha_m \in [0.5\ ,\ 0.75]$.

La vitesse de fermeture normalisée e_n et l 'intégrale de l'impulsion normalisée j_n et sont données par :

$$e_n = \frac{2(2\alpha_m - 1)}{\alpha_m^3(1-\alpha_m)} \quad , \quad j_n = \frac{-3+12\alpha_m - 10\alpha_m^2}{30\alpha_m^3(1-\alpha_m)^2}$$

La fréquence normalisée f_g et l'amplitude normalisée a_g du formant glottique sont données par :

$$f_g = \frac{1}{2\pi}\sqrt{\frac{60(1-\alpha_m)(2\alpha_m-1)}{-3+12\alpha_m-10\alpha_m^2}} \quad , \quad a_g = \frac{\sqrt{(2\alpha_m-1)(-3+12\alpha_m-10\alpha_m^2)}}{\sqrt{15}\ \alpha_m^3(1-\alpha_m)^{\frac{3}{2}}}$$

Modèle LF

Le modèle LF, développé par Fant, Liljencrants & Lin [72] (1985), s'attache à décrire l'ODGD, plutôt que l'ODG. Les fonctions génératrices de l'ODG et de l'ODGD sont données par :

domaine temporel

$$n_g(t,\alpha_m) = \frac{1+e^{a_n t}(a_n \frac{\alpha_m}{\pi}\sin(\frac{\pi}{\alpha_m}t)-\cos(\frac{\pi}{\alpha_m}t))}{1+e^{a_n \alpha_m}}$$

$$n_g^{'}(t,\alpha_m) = \frac{e^{a_n t}\ (a_n^2+(\frac{\pi}{\alpha_m})^2)\ \sin(\frac{\pi}{\alpha_m}t)}{\frac{\pi}{\alpha_m}\ (1+e^{a_n \alpha_m})}$$

domaine spectral

$$N_g(f,\alpha_m) = \frac{1}{1+e^{a_n \alpha_m}} \frac{\pi^2+\alpha_m^2 a_n^2}{\pi^2+\alpha_m^2(a_n-j2\pi f)^2}\left(\frac{1-e^{-(j2\pi f)}}{j2\pi f} - e^{(a_n-j2\pi f)}\frac{\sin(\frac{\pi}{\alpha_m})}{\frac{\pi}{\alpha_m}}\right)$$

$$N_g^{'}(f,\alpha_m) = j2\pi f\ N_g(f)$$

La valeur du paramètre a_n qui apparaît dans ces équations se calcule à partir de la condition de nullité de l'ODG à l'instant de fermeture (pas de phase de retour) : $n_g(1,\alpha_m) = 0$. a_n est alors solution de l'équation implicite :

$$e^{-a_n} + a_n \frac{\alpha_m}{\pi}\sin(\frac{\pi}{\alpha_m}) - \cos(\frac{\pi}{\alpha_m}) = 0 \qquad \text{(A.1)}$$

$n_g^{'}(t,\alpha_m)$ n'est définie que si $\sin(\frac{\pi}{\alpha_m}) \neq 0$, ce qui implique que $0.5 < \alpha_m < 1$.

La vitesse de fermeture normalisée e_n et l 'intégrale de l'impulsion normalisée j_n et sont données par :

$$e_n = \frac{-e^{a_n}\sin(\frac{\pi}{\alpha_m})(a_n^2+(\frac{\pi}{\alpha_m})^2)}{\frac{\pi}{\alpha_m}(1+e^{a_n \alpha_m})} \quad , \quad j_n = \frac{1-e^{a_n}\frac{\sin(\frac{\pi}{\alpha_m})}{\frac{\pi}{\alpha_m}}}{1+e^{a_n \alpha_m}}$$

La fréquence normalisée f_g et l'amplitude normalisée a_g du formant glottique sont données par :

$$f_g = \frac{1}{2\pi}\sqrt{\frac{a_n^2+(\frac{\pi}{\alpha_m})^2}{\frac{\pi}{\alpha_m}\ 1-\frac{\pi}{\alpha_m}\ e^{a_n}\ \sin(\frac{\pi}{\alpha_m})}} \quad , \quad a_g = \frac{1}{1+e^{a_n \alpha_m}}\sqrt{(a_n^2+(\frac{\pi}{\alpha_m})^2)e^{a_n}\ \frac{\sin(\frac{\pi}{\alpha_m})}{\frac{\pi}{\alpha_m}}(1-e^{a_n}\frac{\sin(\frac{\pi}{\alpha_m})}{\frac{\pi}{\alpha_m}})}$$

Annexe B

Analyse par prédiction linéaire

Nous présentons en partie B de cette annexe les méthodes d'analyse par prédiction linéaire classiques : la prédiction linéaire par autocorrélation et la prédiction linaire par covariance. Nous détaillons également en partie B la modélisation discrète tout-pôle (analyse DAP pour *Discrete All-pole modelling*) qui est une généralisation de la prédiction linéaire dans le cas des signaux à temps discret.

La prédiction linéaire classique

Dans sa version première, l'analyse par prédiction linéaire est appliquée dans le but de déterminer les paramètres du filtre modélisant le conduit vocal $H(z)$, que l'on peut réécrire sous la forme :

$$H(z) = \frac{S(z)}{U'_g(z)} = G\,\frac{1 + \sum_{j=1}^{q} b_j z^{-j}}{1 + \sum_{i=1}^{p} a_i z^{-i}} \tag{B.1}$$

où $(a_i)_{i=[1,p]}$, $(b_j)_{j=[1,q]}$ et G sont les paramètres du système à estimer. Dans le domaine temporel, cette équation exprime l'hypothèse de prédiction linéaire, par :

$$s_n = -\sum_{i=1}^{p} a_i s_{n-i} + G\,(u_n + \sum_{j=1}^{q} b_j u_{n-j}) \tag{B.2}$$

Des hypothèses sur le conduit vocal et la source permettent alors d'apporter des simplifications à l'équation B.2. Hormis les cas de nasalité, le filtre du conduit vocal peut être approché par un filtre tout-pôle (filtre autorégressif (AR)), ce qui implique que : $b_j = 0$ pour $1 \leq j \leq q$. L'équation B.2 prend alors la forme très classique suivante :

$$s_n = -\sum_{i=1}^{p} a_i s_{n-i} + G\,u_n \tag{B.3}$$

L'influence de la source est supposée restreinte aux basses fréquences et, en première approximation, négligeable. Le signal de source u_n est alors modélisé par un signal impulsionnel ou du bruit blanc. Comme il n'est pas connu, on estime s_n uniquement à partir des p échantillons précédents :

$$\widetilde{s}_n = -\sum_{i=1}^{p} a_i s_{n-i} \tag{B.4}$$

L'erreur e_n entre le signal réel s_n et le signal estimé $\widetilde{s}_n$ est alors donnée par :

$$e_n = s_n - \widetilde{s}_n = \sum_{i=0}^{p} a_i s_{n-i} = G\ u_n \tag{B.5}$$

avec $a_0 = 1$. Les $(a_i)_{i=1,p}$ sont déterminés par minimisation d'un critère d'erreur bien choisi. Classiquement, le critère retenu est la minimisation de l'erreur quadratique moyenne $E = \sum_n e_n^2$ par rapport aux paramètres a_i : $\frac{\partial E}{\partial a_i} = 0$ pour $i = 1, p$ Par calcul direct de ces dérivées partielles ou par application du principe d'orthogonalité (Papoulis [164], pp. 407) qui montre que l'erreur e_n est perpendiculaire aux échantillons $s_{n-1} \ldots s_{n-p}$ ($\sum_n s_{n-i} e_n = 0$ pour $i = 1, \ldots n$), on obtient les p équations de Yule-Walker suivantes, encore appelées équations normales :

$$\sum_{k=1}^{p} a_k \ \sum_n s_{n-i} s_{n-k} = -\sum_n s_{n-i} s_n \qquad pour\ i \in [1, p] \tag{B.6}$$

L'erreur minimale est donnée par :

$$E_{min} = \sum_{k=0}^{p} a_k \ \sum_n s_n s_{n-k} = G^2 \tag{B.7}$$

Le choix des valeurs prises par n a son importance dans la résolution de ce système de $p+1$ équations à $p+1$ inconnues (les $(a_i)_{i=[1,p]}$ et G).

1. **sommation sur une durée infinie** $(-\infty < n < \infty)$: **prédiction linéaire par autocorrélation**

   Les $p+1$ équations du système B.6 et B.7 deviennent alors :

$$\begin{aligned} &\sum_{k=1}^{p} a_k r_{0,k} = G^2 - r_{0,0} \\ &\sum_{k=1}^{p} a_k r_{i,k} = -r_{i,0} \qquad\qquad pour\ i \in [1, p] \end{aligned} \tag{B.8}$$

   avec $r_{i,k} = \sum_{n=-\infty}^{+\infty} s_{n-i} s_{n-k}$ les coefficients de la matrice d'autocorrélation du signal s_n. En pratique, le signal s_n est borné en temps et cette sommation ne peut s'effectuer que sur un nombre fini d'échantillons. Un fenêtrage approprié du signal, par une fenêtre de type Hanning ou Hamming, permet de contourner ce problème (Makhoul [149], 1975). Si nous considérons le signal sur une fenêtre de longueur N, nous obtenons :

$$r_{i,k} = \sum_{n=max(i,k)}^{N-1-min(i,k)} s_{n-i} s_{n-k}$$

   La matrice d'autocorrélation est une matrice symétrique de Toëplitz (tous les éléments d'une même diagonale sont égaux), et le système B.8 peut se résoudre de façon rapide par un algorithme récursif,

par exemple celui proposé par Levinson[142] en 1947 ou Le Roux & Gueguen [178] (1977). Le filtre AR estimé par cette méthode sera toujours stable (pôles à l'intérieur du cercle unité).

2. **sommation sur une durée finie** N $(0 \leq n \leq N-1)$: **prédiction linéaire par covariance**

   Il n'est pas toujours possible d'avoir accès à un nombre important d'échantillons pour effectuer la prédiction linéaire. Dans ce cas, la minimisation de E se fait sur un intervalle temporel borné (entre 0 et $N-1$) et les coefficients $r_{i,k}$ prennent la forme suivante :

$$r_{i,k} = \sum_{n=0}^{N-1} s_{n-i} s_{n-k}$$

   Ces coefficients correspondent aux éléments de la matrice de covariance du signal s_n. Cette matrice reste toujours symétrique, mais les éléments d'une diagonale ne sont plus nécessairement égaux et la résolution du système d'équations de Yule-Walker est alors plus complexe, se faisant généralement par application de l'algorithme de décomposition de Choleski. La stabilité du filtre AR estimé par cette méthode n'est pas assurée.

La prédiction linéaire par autocorrélation semble mieux adaptée à des signaux stationnaires pour lesquels l'analyse peut s'effectuer sur un grand nombre d'échantillons, tandis que la méthode par covariance s'applique mieux à des signaux non-stationnaires (Parsons [165]). En traitement de la parole, les fricatives seront analysées de façon plus précise par autocorrélation et les voyelles par covariance (Parsons [165]).

Expression spectrale et modélisation tout-pôle discrète

Reformulation spectrale de la prédiction linéaire classique (Makhoul [149, 150], 1975)

Par application du théorème de Parseval, l'erreur quadratique moyenne E, calculée à partir de l'équation B.5 dans le cas de la méthode par autocorrélation, peut se réécrire sous forme spectrale (avec $\omega = 2\pi f$) par :

$$E = \sum_{n=-\infty}^{+\infty} e_n^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S(\omega)\ A(\omega)|^2\ d\omega \tag{B.9}$$

$|S(\omega)|^2 = P(\omega)$ est le spectre de puissance du signal s_n . L'estimation $\widetilde{P}(\omega)$ de ce spectre de puissance est donnée par :

$$\widetilde{P}(\omega) = |H(\omega)|^2 = \frac{G^2}{|A(\omega)|^2} \tag{B.10}$$

L' équation B.9 peut alors se réécrire sous la forme :

$$E = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{\widetilde{P}(\omega)}\ d\omega \tag{B.11}$$

La minimisation de cette erreur par rapport aux coefficients $(a_i)_{i=[1,p]}$ amène au système d'équations B.8, où les coefficients d'autocorrélation $r_{i,j}$ s'écrivent sous la forme :

$$r_{i,j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) cos((i-j)\,\omega)\; d\omega$$

Modélisation tout-pôle discrète (DAP)

Les équations précédentes supposent une continuité spectrale, ce qui n'est pas le cas dans la pratique, puisque le signal est échantillonné et périodique. Dans le cas d'un spectre discret, l'erreur E donnée par l'équation B.11 s'écrit (Makhoul [150], 1975) :

$$E = \frac{G^2}{N} \sum_{m=1}^{N} \frac{P(\omega_m)}{\widetilde{P}(\omega_m)}$$

Il est alors possible de montrer que la prédiction linéaire n'est pas l'approche la plus adaptée pour estimer l'enveloppe d'un spectre discret, car elle ne prend pas en compte l'effet du repliement causé par l'échantillonnage spectral (El-Jaroudi & Makhoul [59], 1991). Une amélioration de la prédiction linéaire, basée sur une modification du critère d'erreur, est proposée par El-Jaroudi & Makhoul [59] et appliquée avec succès par Alku dans ses différentes études (Alku [2], 1992 ; Alku & Vilkman [4, 5], 1994, 1996 ; Alku & al [3], 1997 ; Alku & al [7], 1999). En particulier, Alku [2] montre que cette technique permet d'atténuer les « rebonds » observés dans la phase fermée, conséquences d'une mauvaise estimation du premier formant F_1 .

Le critère d'erreur appliqué par El-Jaroudi & Makhoul [59] s'inspire du critère d'erreur développé par Itakura & Saito [118] (1970) dans le cas de spectres continus et modifié par McAulay [152] (1984) dans le cas de spectres discrets. Il est donné par :

$$E = \frac{1}{N} \sum_{m=1}^{N} \left[\frac{P(\omega_m)}{\widetilde{P}(\omega_m)} - \ln\left(\frac{P(\omega_m)}{\widetilde{P}(\omega_m)} \right) - 1 \right] \tag{B.12}$$

Cette erreur E est toujours positive ou nulle, et elle s'annule uniquement quand $\widetilde{P}(\omega_m) = P(\omega_m)$ pour tout $m \in [1, N]$. Si l'on intègre le gain G aux coefficients $(a_i)_{i=[0,N]}$ (en particulier, $a_0 = \frac{1}{G}$), la minimisation de E par rapport à ces coefficients amène au système d'équations suivant :

$$\sum_{j=0}^{p} a_k \left[r_{i,j} - \widetilde{r}_{i,j} \right] = 0 \qquad pour\ 0 \leq i \leq p \tag{B.13}$$

avec

$$\begin{aligned} r_{i,j} &= \tfrac{1}{N} \sum_{m=1}^{N} P(\omega_m) cos(\omega_m (i-j)) \\ \widetilde{r}_{i,j} &= \tfrac{1}{N} \sum_{m=1}^{N} \widetilde{P}(\omega_m) cos(\omega_m (i-j)) \end{aligned} \tag{B.14}$$

Le système d'équations B.13 peut se réécrire sous une forme matricielle :

$$\left(R - \widetilde{R} \right)\, A = 0 \tag{B.15}$$

avec $R = [r_{i,j}]_{i=[0,p],\ j=[0,p]}$, $\widetilde{R} = [\widetilde{r}_{i,j}]_{i=[0,p],\ j=[0,p]}$ et $A = [a_i]_{i=[0,p]}$. Les matrices R et $\widetilde{R}$ sont des matrices de Toëplitz symétriques, de dimension $(p+1,\ p+1)$. Ce système est non-linéaire, car la matrice $\widetilde{R}$ dépend des coefficients $(a_i)_{i=[0,N]}$ (d'après l'équation B.10).

Les solutions de l'équation B.15 sont :

1. $R = \widetilde{R}$,
2. A est vecteur propre de la matrice $\left(R - \widetilde{R}\right)$ de valeur propre associée nulle.

Le premier cas correspond à la solution qui nous intéresse ($\widetilde{P}(\omega_m) = P(\omega_m)$ pour tout $m \in [1, N]$) et l'unicité de cette solution est vérifiée à condition que le nombre N de points spectraux d'estimation soit suffisament élevé (en pratique, $N \geq 2p+1$). Le second cas amène une solution inappropriée mais qui peut être mise en évidence car le filtre AR trouvé possède alors des pôles sur le cercle unité.

Quand $R = \widetilde{R}$, on a alors :

$$\frac{1}{N} \sum_{m=1}^{N} \left[\frac{P(\omega_m)}{\widetilde{P}(\omega_m)} \right] = 1 \tag{B.16}$$

L'erreur minimale est donnée par :

$$E_{min} = \frac{1}{N} \sum_{m=1}^{N} \left[-\ln \left(\frac{P(\omega_m)}{\widetilde{P}(\omega_m)} \right) \right] \tag{B.17}$$

Etapes de l'analyse par modélisation tout-pôle discrète (analyse DAP) :

Comme l'équation B.15 est non-linéaire par rapport aux coefficients $(a_i)_{i=[0,p]}$, sa résolution s'effectue de façon itérative. Une simplification est apportée si on remarque que :

$$\sum_{j=0}^{p} a_j \, \widetilde{r}_{i,j} = \widetilde{h}(-i) \qquad \textit{pour tout } i \in [0, p] \tag{B.18}$$

où $\widetilde{h}(-i)$ est la réponse impulsionnelle anticausale du filtre AR aux fréquences discrètes sélectionnées. Elle est donnée par :

$$\widetilde{h}(-i) = \frac{1}{N} \sum_{m=1}^{N} \frac{e^{-j\omega_m i}}{A(\omega_m)} \tag{B.19}$$

En appliquant l'équation B.15, on a alors également :

$$\sum_{j=0}^{p} a_j \, r_{i,j} = \widetilde{h}(-i) \qquad \textit{pour tout } i \in [0, p] \tag{B.20}$$

La résolution de l'équation B.15 se fait selon les étapes suivantes :

1. calcul du spectre de puissance du signal s_n et sélection du nombre N de points fréquentiels d'analyse, des pulsations ω_m et de l'amplitude $P(\omega_m)$ correspondantes. Les fréquences sélectionnées n'ont pas besoin d'être des multiples de la fréquence fondamentale ;
2. calcul de $r_{i,j}$ d'après la relation B.14 ;

3. estimation initiale des coefficients $(a_i)_{i=[0,p]}$ par application de la prédiction linéaire par autocorrélation ;

   début du processus itératif

4. calcul de $A(\omega_m) = \sum_{i=0}^{p} a_i \, e^{-j\omega_m i}$ pour $1 \leq m \leq N$;
5. évaluation de $\widetilde{h}(-i)$ pour $1 \leq i \leq p$, à partir de l'équation B.19 ;
6. résolution de l'équation B.20 ($R\ A = \widetilde{h}$, avec $\widetilde{h}$ le vecteur $(\widetilde{h}(-i))_{i=[0,p]}$) et estimation des coefficients $(a_i)_{i=[0,p]}$;
7. évaluation de l'erreur E_{min} à partir de l'équation B.12, en estimant $\widetilde{P}(\omega_m)$ par $\widetilde{P}(\omega_m) = \frac{1}{|A(\omega_m)|^2}$;
8. comparaison de l'erreur avec une valeur seuil. Si E_{min} est inférieure à ce seuil, le processus itératif s'arrête, sinon on recommence à l'étape 4. ;

   fin du processus itératif

9. normalisation des coefficients $(a_i)_{i=[0,p]}$ à partir de l'équation B.16 et fin de l'analyse .

La convergence de l'algorithme d'analyse va dépendre principalement de N et elle sera beaucoup plus rapide pour de grandes valeurs de N. Sur le principe des méthodes de convergence de type quasi-Newton, il est également possible d'accélérer la convergence en introduisant un facteur de pondération α. Le calcul de A_k à l'itération k, donné par $A_k = R^{-1}\widetilde{h}$ (réécriture de l'équation B.20), est modifié en tenant compte de la valeur précédente estimée, par :

$$A_k = (1-\alpha)\, A_{k-1} + \alpha\, R^{-1}\widetilde{h}$$

El-Jaroudi & Makhoul [59] ont montré que les valeurs adéquates de ce facteur de pondération sont comprises entre 0.4 et 0.8 . Dans nos expériences, nous utiliserons $\alpha = 0.7$.

Conclusion

La méthode proposée par El-Jaroudi & Makhoul [59] en 1991, basée sur le critère d'erreur d'Itakura-Saito, semble mieux adaptée que la prédiction linéaire pour l'estimation de spectres discrets. Il est également intéressant de remarquer que, pour de très grandes valeurs de N (approximation du cas continu), l'analyse DAP revient à effectuer une prédiction linéaire, ce qui amène El-Jaroudi & Makhoul à poser que la prédiction linéaire est un cas particulier de l'analyse DAP pour lequel $N \to \infty$.

La prédiction linéaire avec modélisation de la source

Le plus gros défaut de la prédiction linéaire est l'hypothèse très restrictive faite sur le signal de source. Si ce facteur nous intéresse, comme c'est le cas dans cette étude, une modélisation convenable de la source glottique est à ajouter à cette méthode de prédiction.

Principe

Les équations B.2 et B.3 (cas du modèle AR) restent inchangées, mais l'expression de l'erreur entre signal réel et signal estimé est modifiée, en tenant compte cette fois-ci de la source, par :

$$\begin{aligned} \widetilde{s}_n &= -\sum_{i=1}^{p} a_i s_{n-i} + G\,(u_n + \sum_{j=1}^{q} b_j u_{n-j}) \\ e_n &= s_n - \widetilde{s}_n = \sum_{i=0}^{p} a_i s_{n-i} - G\,(u_n + \textstyle\sum_{j=1}^{q} b_j u_{n-j}) \end{aligned} \tag{B.21}$$

Comme précédemment, les coefficients $(a_i)_{i=[1,p]}$, $(b_j)_{j=[1,q]}$ et G s'obtiennent par minimisation de l'erreur quadratique moyenne E. Cela suppose de connaître le signal u_{n-j} pour $j \in [0,q]$, qui est justement le signal que l'on cherche à estimer ! Les solutions proposées par Ljungqvist [146, 83] (1986), Isaksson & Millnert [115] (1986), Liu [145] (1993), Lu & Smith [148] (1999) ou très récemment Fröhlich & al [78] (2001) s'appuient sur des algorithmes d'optimisation sous contraintes des paramètres des modèles d'ODG et se rapprochent des méthodes spectrales d'Analyse-par-Synthèse (Stevens [183], Bell & al [19]). Un modèle de source est choisi et le signal $u_{n-j\ (j=[0,q])}$ est synthétisé à partir d'un jeu connu de paramètres. L'espace des paramètres décrivant ce modèle est balayé, à la recherche du jeu de paramètres pour lequel l'erreur quadratique moyenne E est minimale. Ces méthodes sont appliquées de façon pitch-synchrone et sont donc comparables à la prédiction linéaire par covariance. Nous allons détailler ici la méthode proposée par Ljungqvist [146], qui sera utilisée par la suite dans cette étude.

Méthode de Ljungqvist [146, 83]

La transformée en z de l'erreur quadratique moyenne E s'écrit : $E(z) = S(z) - H(z)\,U_g^{'}(z)$ avec $H(z)$ donné par l'équation B.1. Cette formulation amène à un problème de minimisation non-linéaire. Le critère d'erreur peut alors être modifié d'après Kalman [122] (Ljungqvist [146]) par :

$$E(z) = A(z)S(z) \;-\; B(z)U_g^{'}(z)$$

avec $A(z) = 1 + \sum_{i=1}^{p} a_i z^{-i}$ et $B(z) = \sum_{j=0}^{q} b_j z^{-j}$) (par simplification d'écriture, le paramètre de gain G a été introduit dans les paramètres b_j comme commun diviseur). La minimisation de E par rapport aux paramètres $(a_i)_{i=[1,p]}$ et $(b_j)_{j=[0,q]}$ amène à un système de $p+q+1$ équations, que l'on peut écrire sous forme matricielle :

$$\begin{bmatrix} S_{i_{[1,p]},j_{[1,p]}} & -X_{i_{[1,p]},j_{[0,q]}} \\ -X_{i_{[0,q]},j_{[1,p]}} & U_{i_{[0,q]},j_{[0,q]}} \end{bmatrix} \begin{bmatrix} a_{i_{[1,p]}} \\ b_{i_{[0,q]}} \end{bmatrix} = \begin{bmatrix} -S_{0,j_{[1,p]}} \\ X_{0,j_{[0,q]}} \end{bmatrix} \tag{B.22}$$

avec

$$\begin{aligned} S_{i,j} &= \sum_{n=0}^{N-1} s_{n-i} s_{n-j} \\ X_{i,j} &= \sum_{n=0}^{N-1} s_{n-i} u_{n-j}^{'} \\ U_{i,j} &= \sum_{n=0}^{N-1} u_{n-i}^{'} u_{n-j}^{'} \end{aligned}$$

On remarque que ce système très général peut se ramener à deux cas particuliers :

1. $\mathbf{u_{n-j}^{'} = 0}$ pour $j = [0,q]$: on retrouve alors le cas de la prédiction linéaire classique.

2. $\mathbf{b_j = 0}$ pour $j = [1, q]$: le filtre modélisant le conduit vocal est supposé tout-pôle (filtre AR).

Comme la matrice du système d'équations B.22 est symétrique, la résolution de ce système se fait à l'aide des mêmes méthodes utilisées dans le cas de la prédiction linéaire par covariance classique. Comme le signal u_n n'est pas connu, un modèle de source glottique est choisi et l'espace des paramètres est balayé pour trouver le jeu de paramètres qui minimise l'erreur E.

Cette méthode est appliquée de façon synchronisée à la période fondamentale de voisement. D'après Ljungqvist [146], la position et la taille de la fenêtre d'analyse influencent de façon conséquente l'erreur de prédiction. Cette erreur est minimale quand la fenêtre d'analyse est placée au moment de l'ouverture glottique et est de la taille d'une période. L'accentuation du signal voisé avant analyse a également une influence importante sur l'erreur.

Les résultats présentés par Ljungqvist sont tout à fait encourageants, car ils montrent une amélioration sensible de la décomposition source-filtre par rapport aux méthodes de prédiction traditionnelles, se traduisant sur l'erreur de prédiction par une réduction de 3 à 5 dB. Une étude comparative sur les modèles utilisés pour synthétiser le signal de source montre également que le modèle d'ODGD proposé dans le cadre de cette étude et le modèle LF semblent les mieux adaptés à représenter des ODGD dans le cas de signaux réels, par rapport aux modèles de Rosenberg [169] et Hedelin [100].

Avantages et inconvénients d'une telle méthode

Ces méthodes par optimisation tiennent compte du signal de source glottique et semblent donc plus appropriées qu'une prédiction linéaire classique. En même temps que l'estimation du signal de source, elles permettent d'estimer les paramètres de source glottique d'un modèle donné. Leur inconvénient majeur est qu'elles s'appliquent de façon synchrone à la période fondamentale et nécessitent donc une bonne connaissance des instants de fermeture ou d'ouverture glottique. Dans cette étude, nous avons à disposition les signaux électroglottographiques enregistrés de façon simultanée aux signaux acoustiques. Au chapitre 7, l'utilisation des signaux EGG dérivés pour la détection des instants de fermeture et d'ouverture glottique est développée très en détail. Lors de l'application de la méthode de Ljungqvist pour le filtrage inverse, ces signaux sont alors utilisés pour détecter avec précision les instants d'ouverture glottique . Une autre limitation des méthodes d'optimisation est qu'elles sont souvent très coûteuses en temps de calcul. C'est un point que nous n'avons pas pris en compte dans cette étude exploratoire.

Erreur résiduelle, erreur moyenne normalisée

Lors de l'estimation d'un signal par un modèle, l'observation du signal d'erreur e_n entre le signal réel s_n et le signal estimé $\widetilde{s_n}$ donne une indication sur la pertinence de l'estimation. Pour chaque échantillon n, l'erreur e_n, encore appelée erreur résiduelle, est donnée par l'équation B.5. Dans le cas de la prédiction linéaire, cette erreur se calcule par rapport au signal voisé accentué, où l'on suppose que les effets de la source glottique ont été fortement atténués. Le signal correspondant est composé d'impulsions situées aux instants de fermeture glottique et de bruit. De ce fait, ce signal peut être utilisé pour la détection des instants de fermeture glottique (Atal & Hanauer [13]). L'erreur moyenne normalisée est donnée par :

$$E_{norm} = \frac{\sum_n e_n^2}{\sum_n s_n^2}$$

Il est intéressant de remarquer que l'ODGD estimée n'est autre que l'erreur résiduelle calculée par rapport au signal voisé. De même, l'ODG estimée correspond à l'erreur résiduelle calculée par rapport au signal voisé désaccentué.

Annexe C

Protocole d'enregistrement

Laboratoire d'Acoustique Musicale

Contre-ténor

Notre étude porte sur la qualité de la voix et ces enregistrements ont donc pour but de permettre une analyse de la source (les cordes vocales) en voix chantée pour différentes qualités d' émission vocale.

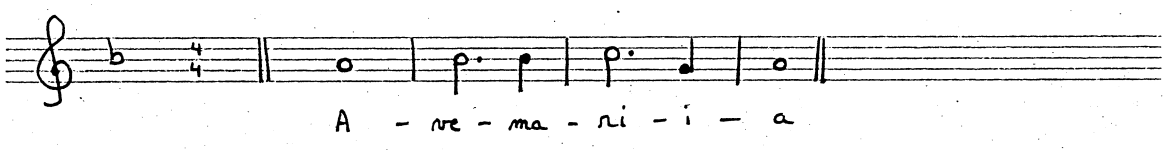
Remarque : pour ces exemples vocaux, il est préférable d'avoir **le moins de vibrato possible**.

1. Comparaison voix parlée / voix chantée

En français et sur une phrase de votre choix :
énoncer, chanter, puis énoncer à nouveau cette même phrase, en utilisant la voix de poitrine puis la voix de tête. La reprendre en criant.

2. Une phrase musicale classique

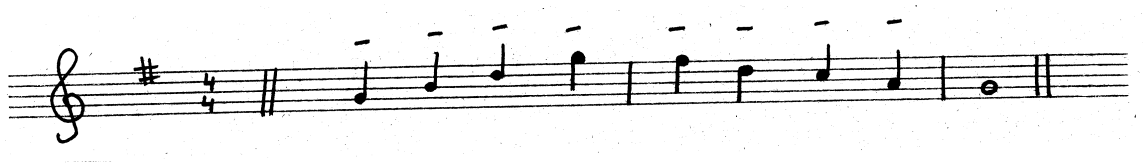
Interpréter les premières mesures de l' *Ave Maria* de Gounod ci-dessous en transposant de telle sorte que vous puissiez interpréter la phrase dans la même tonalité, d'abord en voix de poitrine puis en voix de tête, sur trois intensités différentes : piano, médium et forte.



Reprendre cette phrase musicale, en essayant de modifier **la qualité vocale** (*à votre appréciation*). *Préciser oralement les différentes qualités que vous produirez.*

3. Arpège : *petit Rossini*

En gardant la même intensité, en limitant le vibrato si possible et en tenant chaque note au moins 1 seconde (vous pouvez reprendre votre souffle en cours d'arpège au besoin), chanter un arpège (*le petit Rossini*) commençant sur le sol2, d'abord de façon naturelle, puis de façon lyrique.

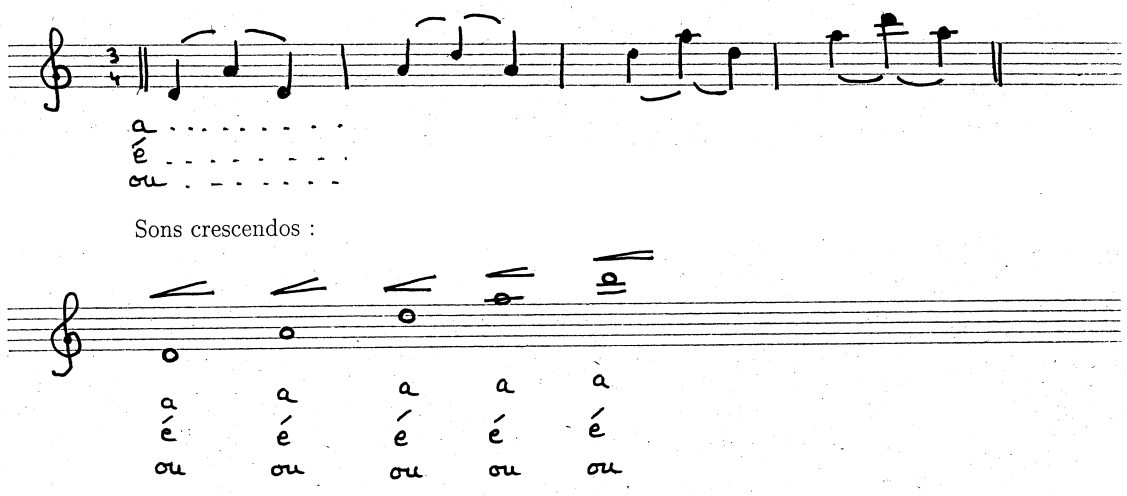


Selon votre inspiration, vous pourrez ensuite reprendre l'arpège, en essayant de modifier la qualité vocale, comme précédemment.

4. Voyelles tenues et sons crescendo

Il s'agit de chanter des sons tenus et des sons crescendos, en portant son attention sur le maintien de la couleur de la voyelle et de la hauteur. Les sons crescendos auront une durée variant entre 4 à 8 secondes et chaque son tenu environ 1 seconde. Lorsque vous le pourrez, vous utiliserez la voix de poitrine puis la voix de tête, en le précisant oralement.

Les trois voyelles choisies sont le "a", le "é" et le "ou". Pour chacune d'entre elle, et pour trois intensités différentes (piano, médium,forte), chanter les mesures suivantes :



5. Expression vocale dans un autre style musical

Interpréter l'air "*Joyeux Anniversaire*" à la manière d'une chanson populaire, puis de façon lyrique, mais sans trop de vibrato.

6. Un glissando

Chanter un **glissando ascendant et descendant**, mezzo-forte (plutôt vers le forte), continu (image de la sirène) et de préférence avec le moins de vibrato possible, en passant de la voix de poitrine à la voix de tête.

Le reprendre plus lentement, puis uniquement en voix de poitrine, et enfin uniquement en voix de tête.

Annexe D

Un peu de statistique ...

Cette annexe présente les calculs statistiques effectués pour déterminer les coefficients de corrélation totale et partielle, leurs significations statistiques et l'axe majeur de régression, dans le cadre des mesures de quotient ouvert effectuées au chapitre 8, ainsi que les tests sur observations appariées effectuées au chapitre 9.

Corrélation entre plusieurs variables

Coefficient de corrélation totale

Le coefficent de corrélation totale, dit de « Bravais-Pearson » (Sapora [179]), mesure le caractère linéaire d'une liaison entre deux variables. Il se calcule en fonction des valeurs prises par les variables x et y sur n échantillons et de leurs valeurs moyennes $\bar{x}$ et $\bar{y}$, par la formule :

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)(\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2)}$$

Le dénominateur représente les produits des écart-types de x et y et le numérateur la covariance observée entre ces deux variables.

Le test de signification du coefficient de corrélation ainsi calculé compare la valeur de $t_{obs} = \frac{|r_{xy}|\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$ à celle donnée dans les tables de distribution de T (loi de Student) (Dagnelie [43]). Si $t_{obs} \geq t_{1-\alpha/2}$, l'hypothèse nulle est rejetée pour un niveau de signification α et avec $n-2$ degrés de liberté. Pour les grandes valeurs de n ($n > 200$), $t_{0.9995} = 3.3$ ($p < 0.001$). Ce sera souvent le cas dans cette étude.

Coefficient de corrélation partielle

Il peut arriver que la dépendance apparente entre deux variables soit en réalité due aux variations d'une troisième variable (Sapora [179]). Dans ce cas, le coefficient de corrélation partielle constitue le moyen d'éliminer l'influence de cette variable. Dans le cadre de cette étude, nous avons vu que l'intensité vocale et la fréquence fondamentale peuvent être fortement corrélées (voir Tableau 8.2). Le calcul des coefficients de corrélation totale entre quotient ouvert et intensité vocale ou fréquence fondamentale peut

donc être biaisé par une corrélation sous-jacente entre intensité et fréquence, dont il faut tenir compte. Le coefficient de corrélation partielle entre x et y en tenant des variations de la variable z se calcule par la formule (Dagnelie [43], Jolicoeur [121]) :

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

Ce résultat est même généralisable à plus de trois variables. Le caractère significatif d'une corrélation partielle se teste de la même façon que dans le cas d'une corrélation totale, puisque les lois décrivant les coefficients de corrélation totale et partielle sont les mêmes. Seul le degré de liberté $(n-2)$ est diminué du nombre de variables fixées (1 dans le cas de trois variables). On calculera donc $t_{obs} = \frac{|r_{xy.z}|\sqrt{n-3}}{\sqrt{1-r_{xy.z}^2}}$ et on le comparera aux valeurs données dans les tables.

Axe majeur

Le calcul de l'axe majeur entre deux variables x et y revient à estimer une équation de régression linéaire entre ces variables, sans pour autant privilégier une direction donnée (ce qui est le cas dans le calcul d'une droite de régression classique). Son estimation est basée sur le principe des moindres carrés (Jolicoeur [121]). On obtient alors le coefficient directeur de la droite $y = a + b * x$ par l'équation :

$$b = \frac{(S_Y^2 - S_X^2) + \sqrt{(S_Y^2 - S_X^2)^2 + 4S_{XY}^2}}{2S_{XY}}$$

avec

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

$S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

et on déduit a de l'équation : $a = \bar{y} - b\bar{x}$, où $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ et $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Tests statistiques sur observations appariées

Le test sur observations appariées, encore appelé test des couples, permet d'effectuer des comparaisons statistiques entre deux moyennes (Daudin & al [49], 1999). Il est plus puissant et plus précis qu'un test classique de comparaisons entre deux populations, car il n'inclut pas dans la variabilité de la statistique du test celle induite par le changement d'individu.

Soit $D_i = (x_i - y_i)$ pour tout $i \in [1, n]$ la distance entre les mesures $(x_i)_{i=[1,n]}$ et $(y_i)_{i=[1,n]}$ effectuées sur deux unités statistiquement appariées. Dans le cas du test perceptif présenté au chapitre 9 par exemple, les mesures sont les seuils obtenus pour chaque sujet S_i et les unités statistiques sont les passages (dans le cas où 3 passages du test ont été effectués) ou les sessions (pour des comparaisons entre sessions). La distance moyenne $\bar{D}$ et l'écart-type S_D sont donnés par :

$$\bar{D} = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)$$

$$S_D = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2}$$

L'hypothèse d'égalité entre les deux moyennes est rejetée pour un niveau de signification α et un degré de liberté $n-1$ si :

$$|\bar{D}| > t_{n-1\,;\,1-\alpha/2} \frac{S_D}{\sqrt{n}}$$

$$\Rightarrow \quad t_{obs} = \sqrt{n}\, \frac{|\bar{D}|}{S_D} > t_{n-1\,;\,1-\alpha/2}$$

où, comme précédemment, $t_{n-1\,;\,1-\alpha/2}$ correspond à la distribution donnée par la loi de Student. Dans notre cas, $n = 20$ pour le groupe des 20 sujets non-entraînés et $n = 10$ pour le groupe des 10 sujets entraînés.

Par exemple, si l'on désire tester l'effet d'apprentissage entre deux tests, pour une session donnée, on calcule t_{obs} et on le compare aux valeurs tabulées. Il y a un effet d'apprentissage significatif entre deux passages si $t_{obs} > t_{n-1\,;\,0.05} = 2.26$ ($p < 0.05$; $n = 10$ correspond au nombre de sujets). Un effet très significatif, pour $p < 0.01$, est observé si $t_{obs} > 3.25$.

Annexe E

Variations de l'intensité vocale en fonction du quotient ouvert

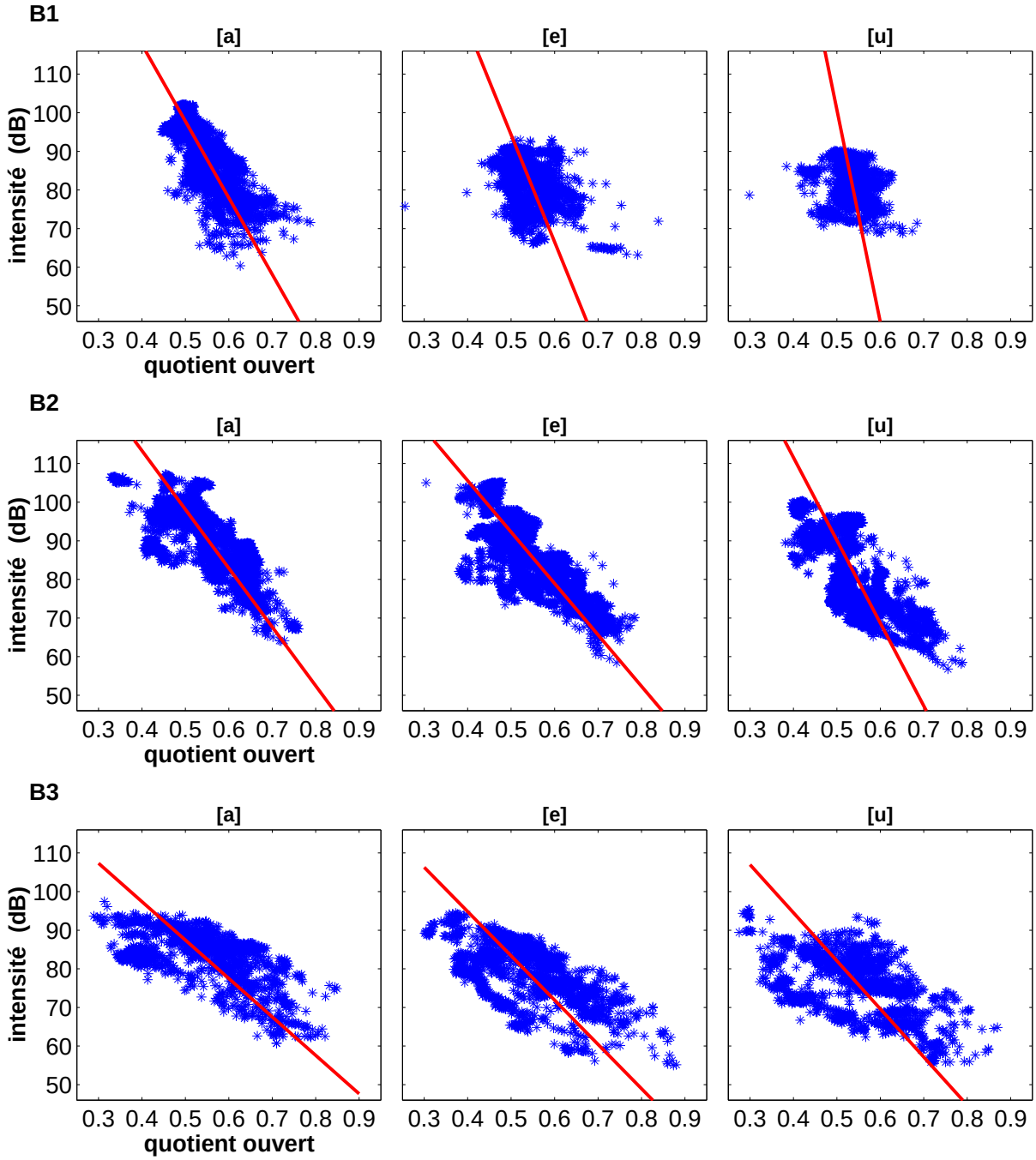


FIG. *E.1 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : barytons B1, B2 et B3. La droite représentée sur chaque figure correspond à l'axe majeur.*

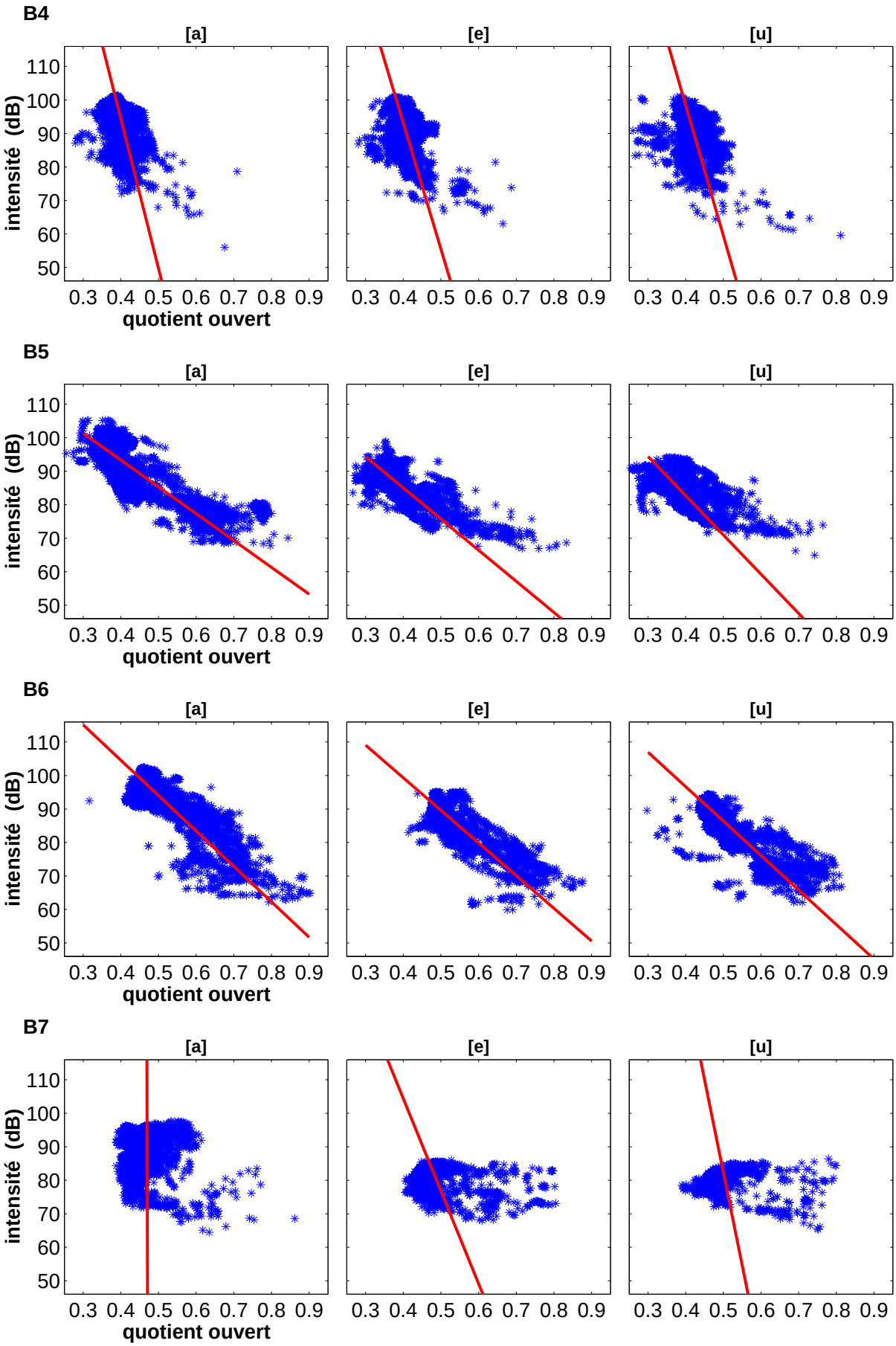


FIG. *E.2 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : barytons B4, B5, B6 et B7. La droite représentée sur chaque figure correspond à l'axe majeur.*

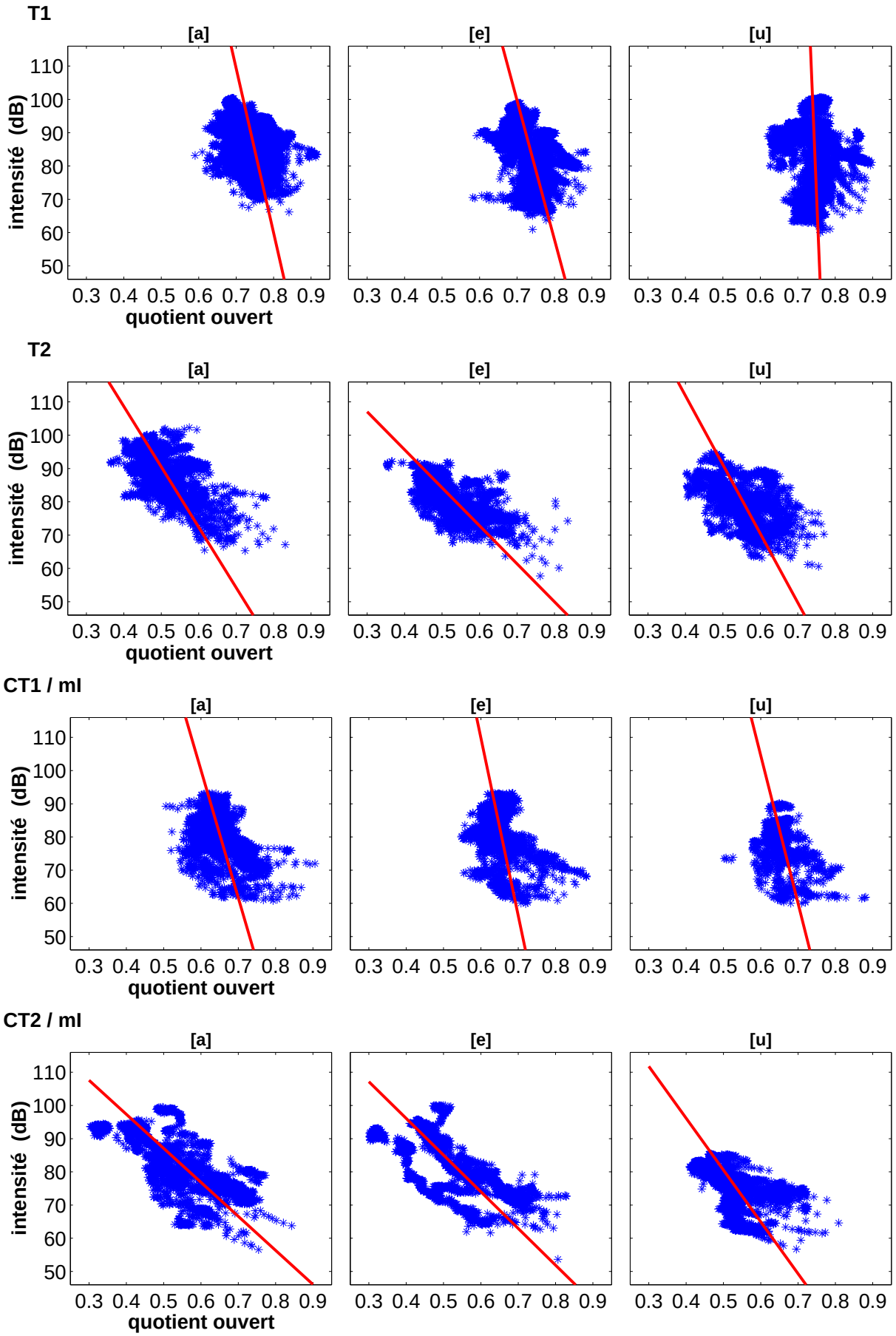


FIG. *E.3 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : ténors T1 et T2, contre-ténors CT1 et CT2 en mécanisme I. La droite représentée sur chaque figure correspond à l'axe majeur.*

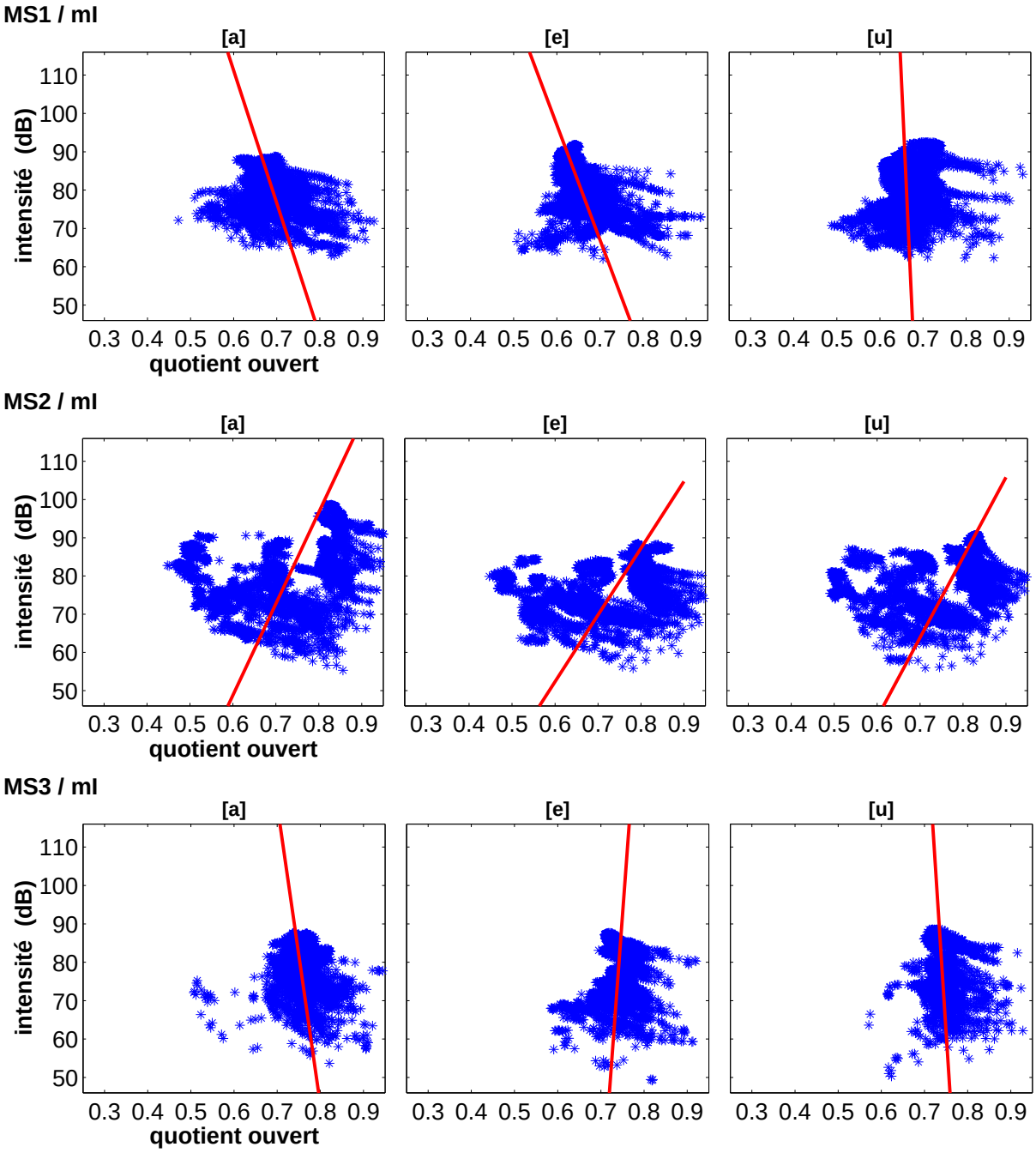


FIG. *E.4 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : mezzo-sopranos MS1, MS2 et MS3. La droite représentée sur chaque figure correspond à l'axe majeur.*

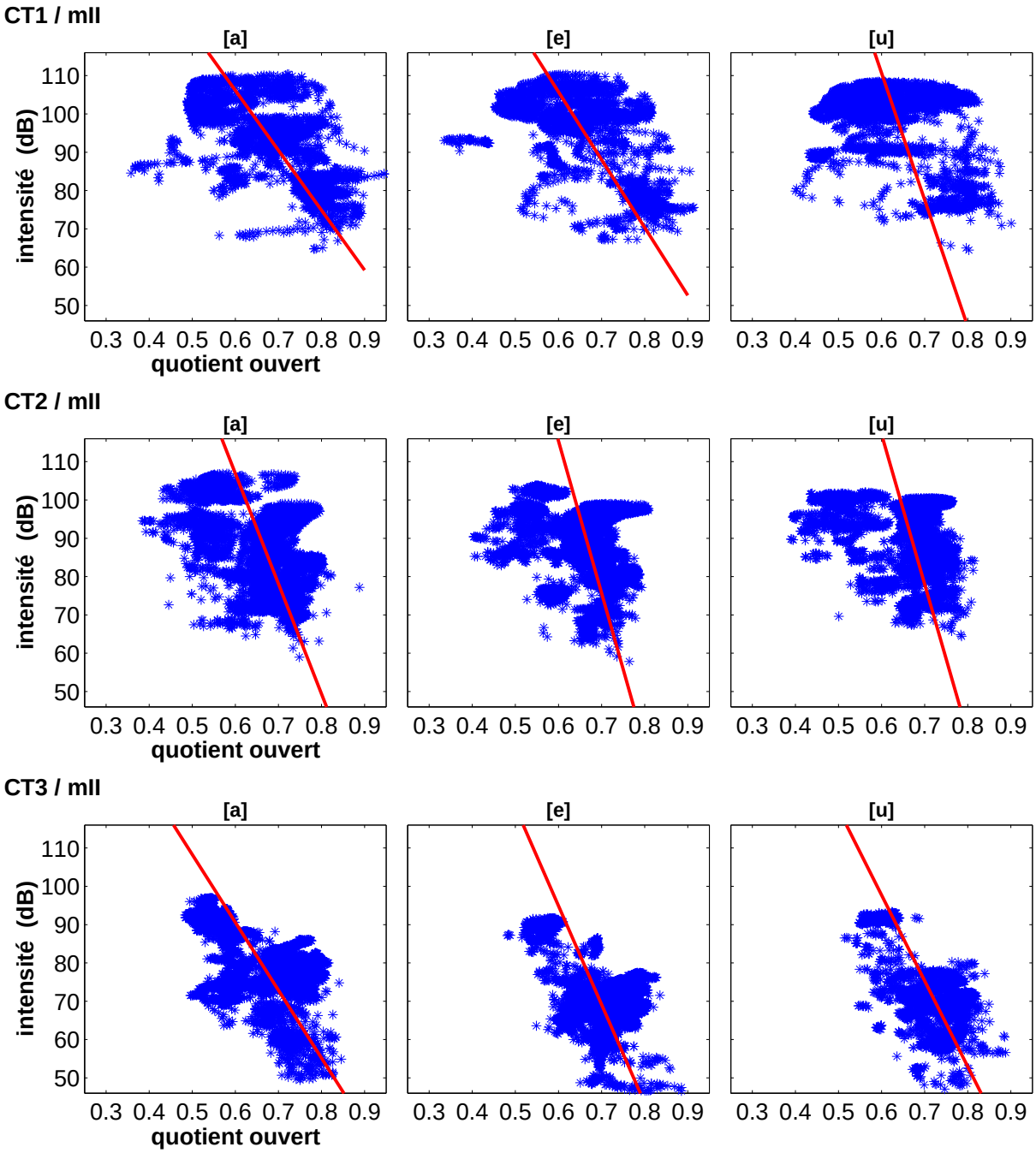


FIG. *E.5 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : contre-ténors CT1, CT2 et CT3 en mécanisme II. La droite représentée sur chaque figure correspond à l'axe majeur.*

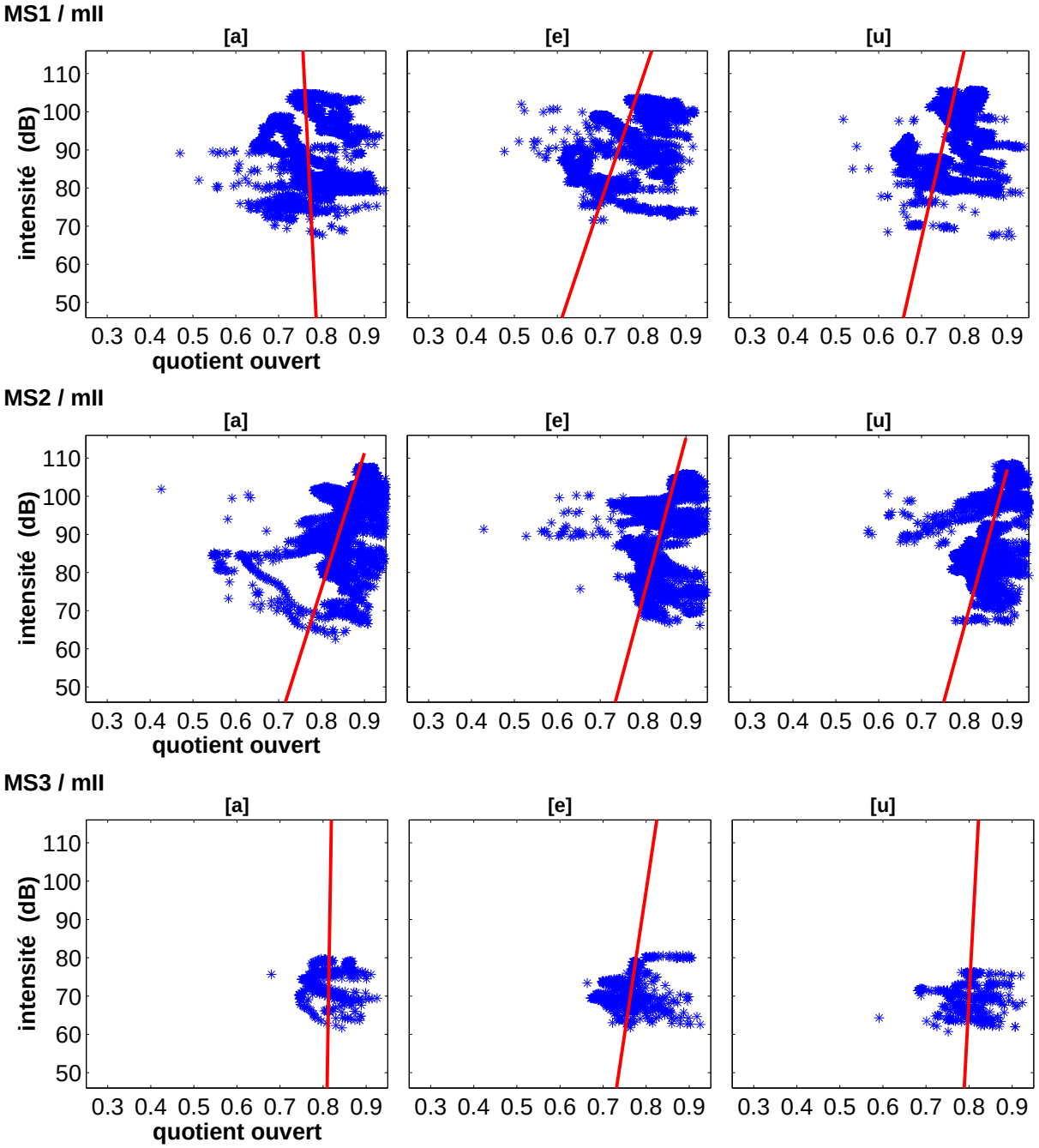


FIG. *E.6 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : mezzo-sopranos MS1, MS2 et MS3 en mécanisme II. La droite représentée sur chaque figure correspond à l'axe majeur.*

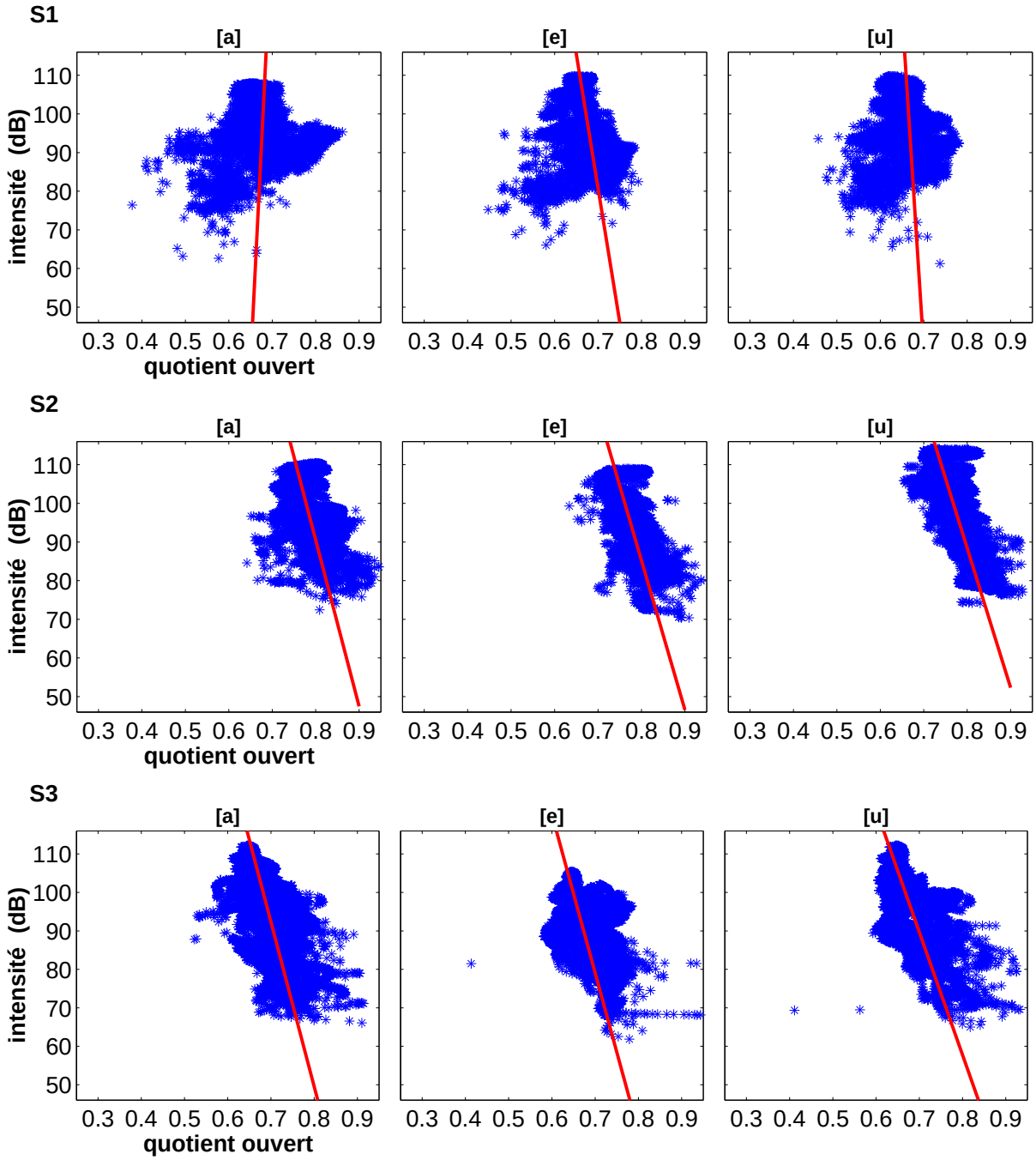


FIG. *E.7 – Intensité vocale en fonction du quotient ouvert pour les 3 voyelles [a], [e] et [u] : sopranos S1, S2 et S3 en mécanisme II. La droite représentée sur chaque figure correspond à l'axe majeur.*

Annexe F

Visualisation des glissandos

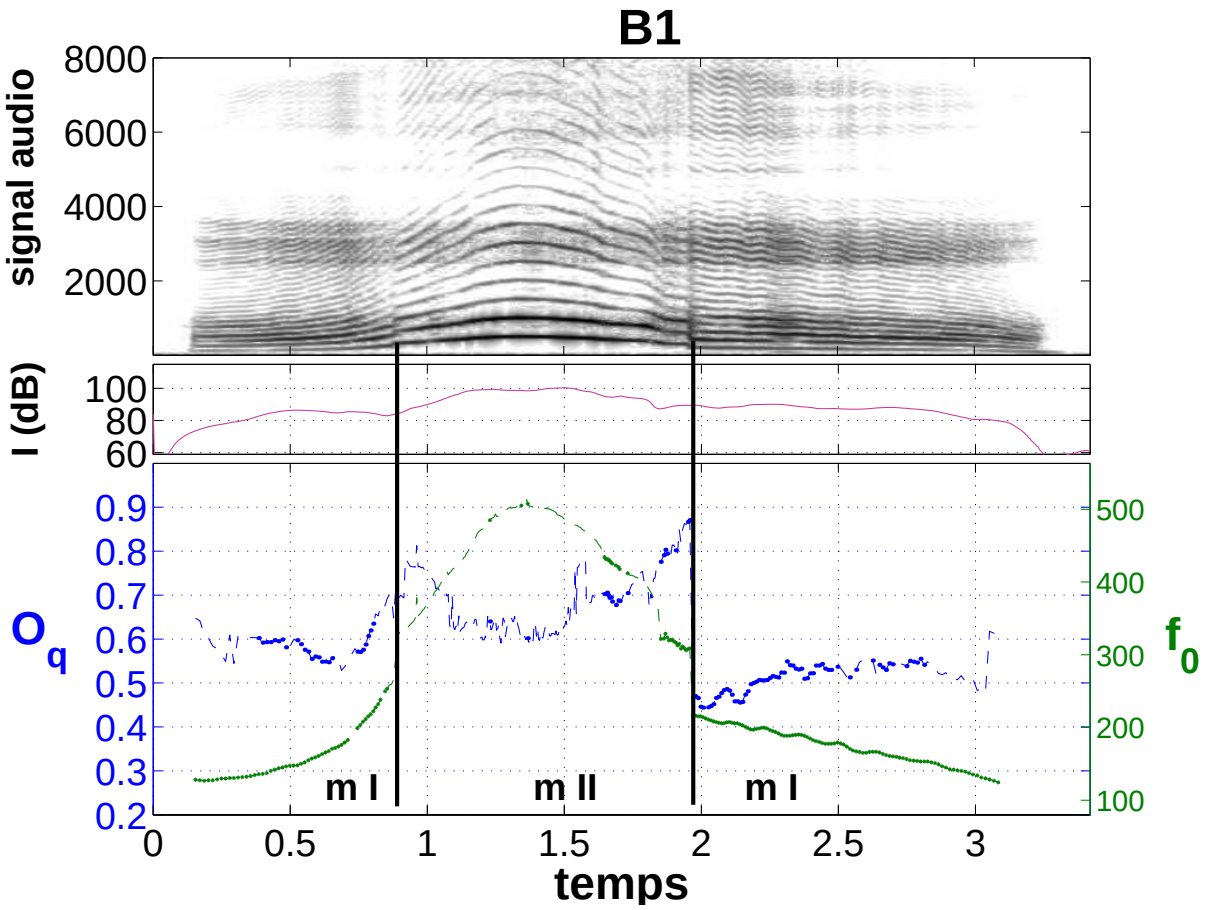


FIG. *F.1 – baryton B1, (CD, piste [29])*

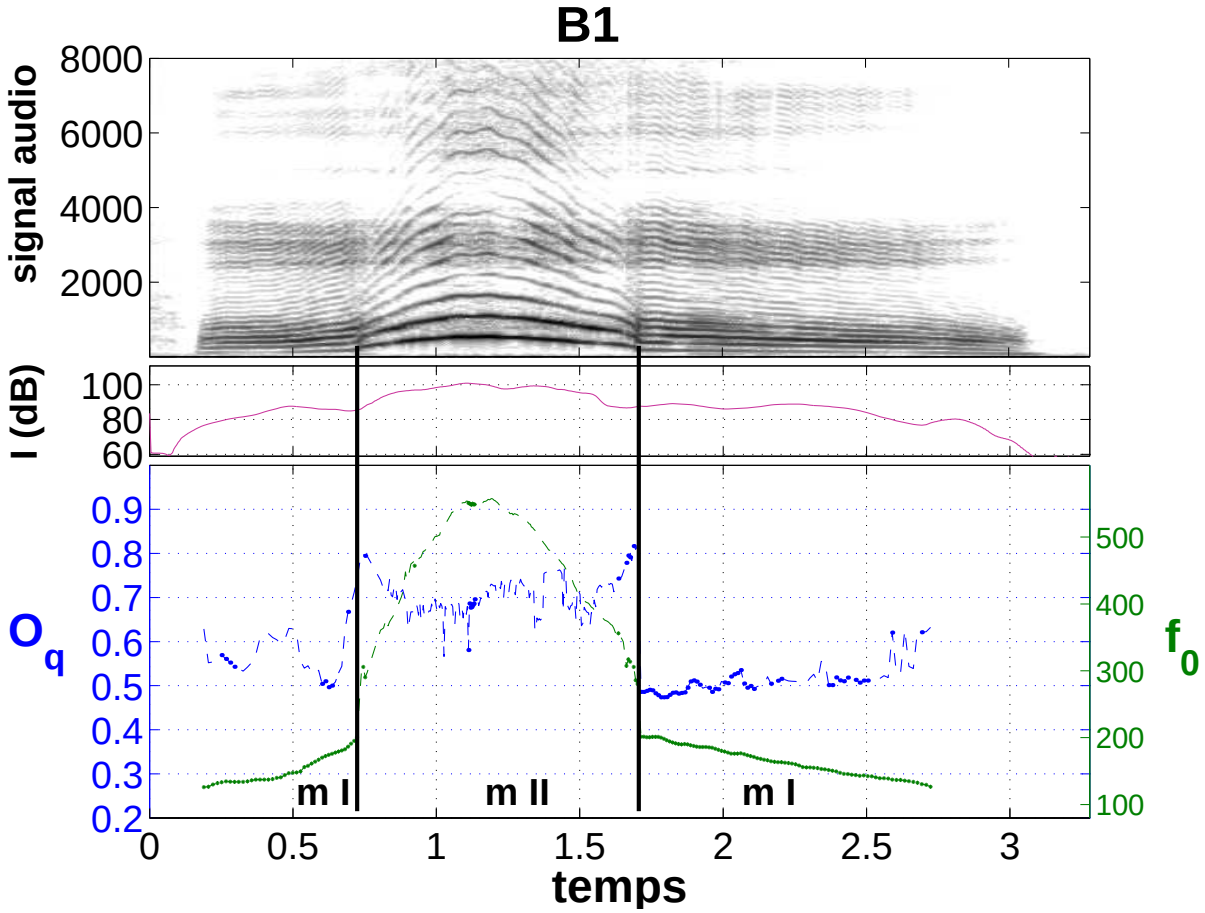


FIG. *F.2 – baryton B1, (CD, piste [30])*

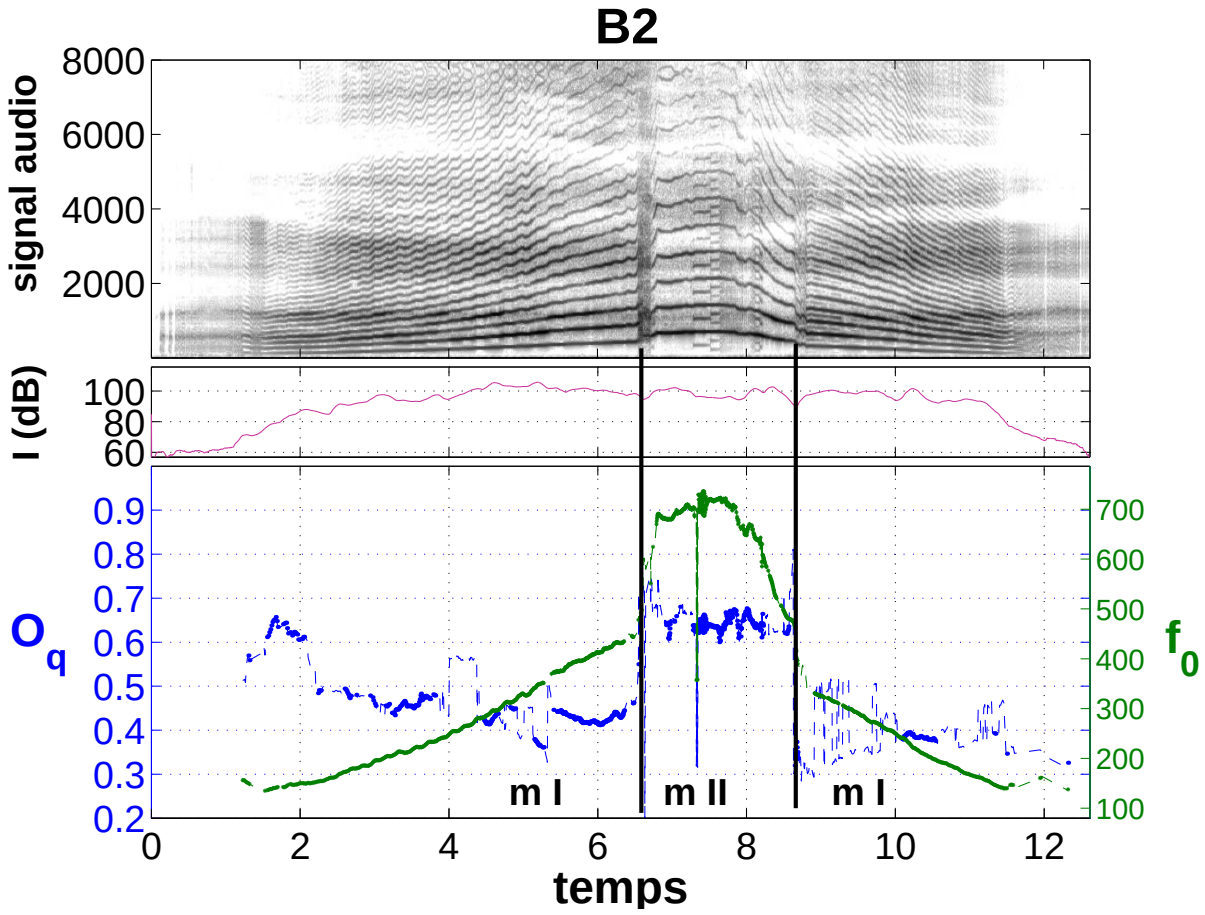


FIG. *F.3 – baryton B2, (CD, piste [31])*

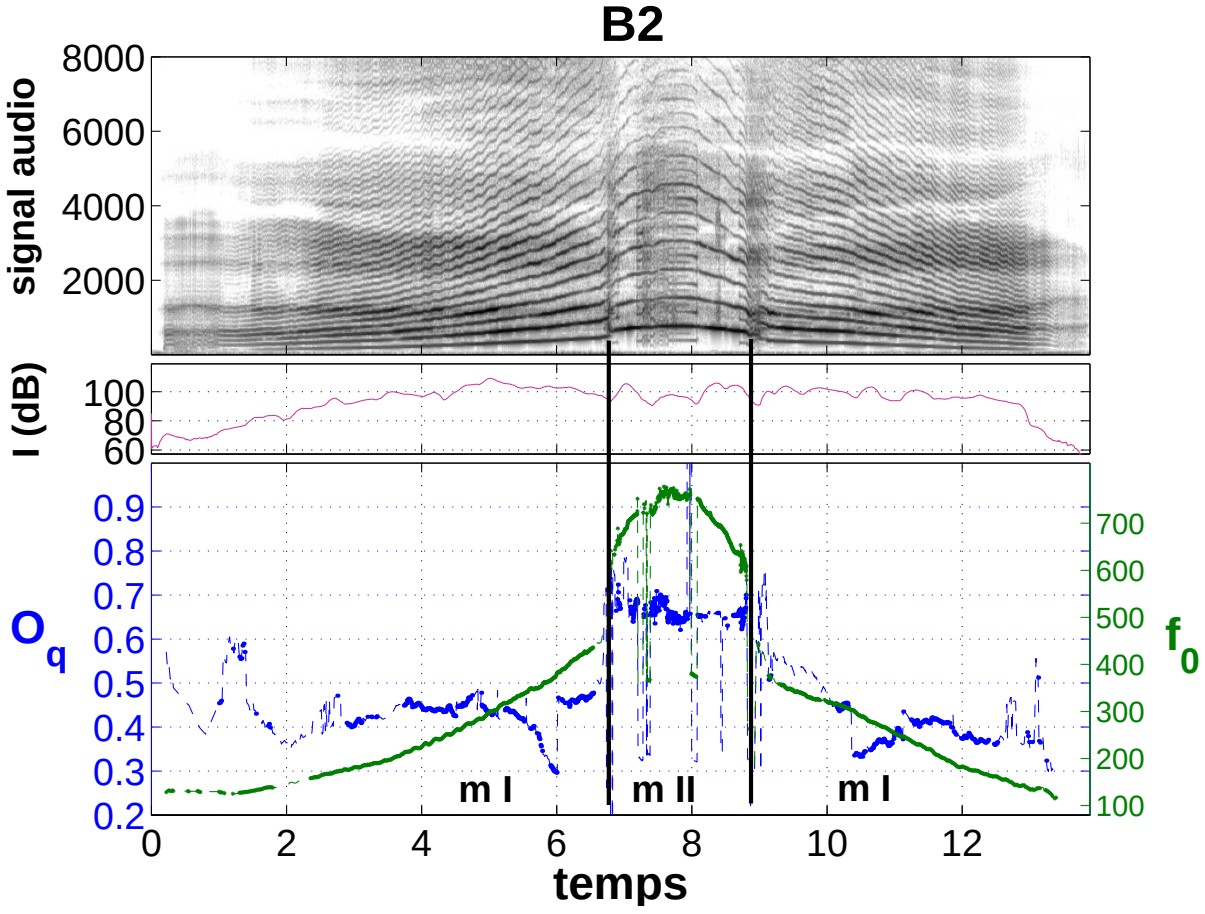


FIG. *F.4 – baryton B2, (CD, piste [32])*

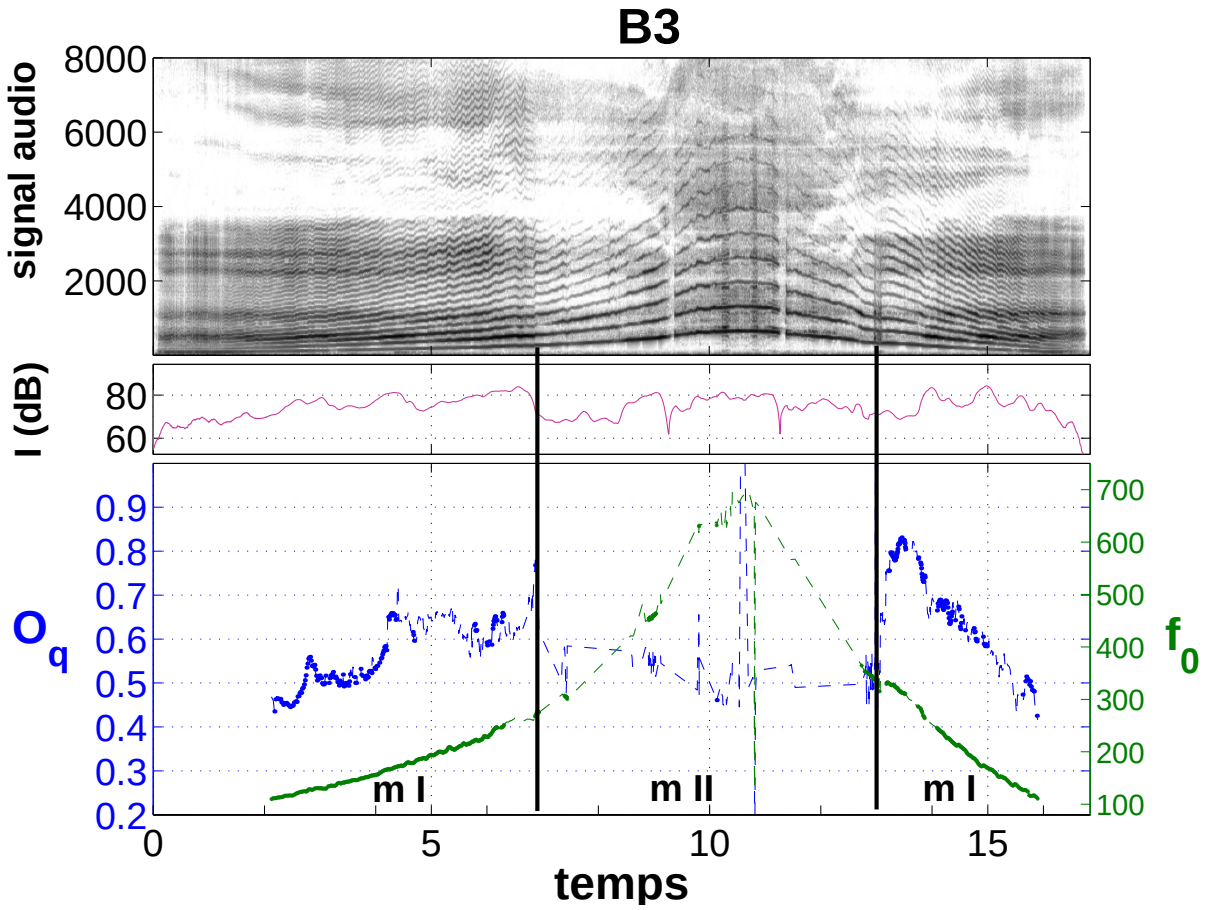


FIG. *F.5 – baryton B3, (CD, piste [33])*

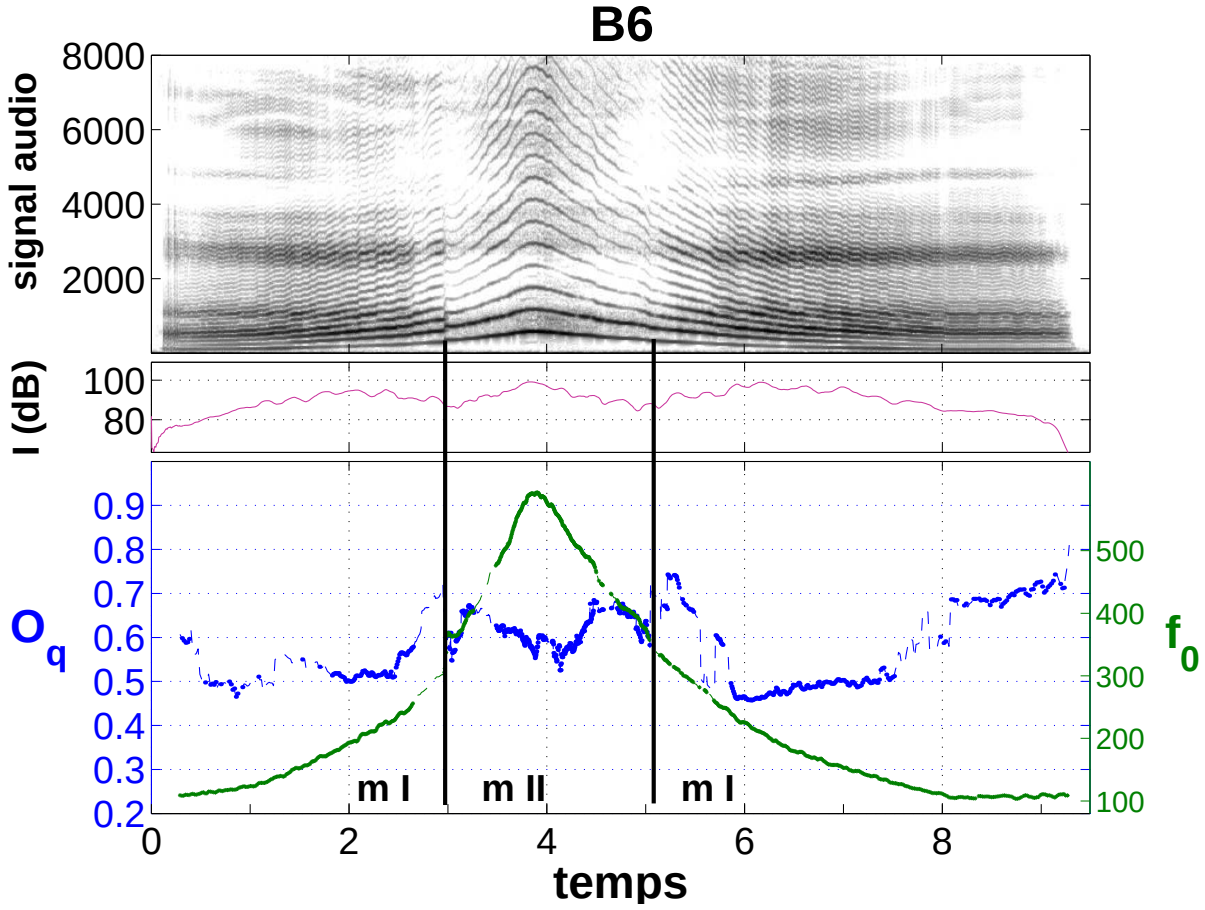


FIG. *F.6 – baryton B6, (CD, piste [38])*

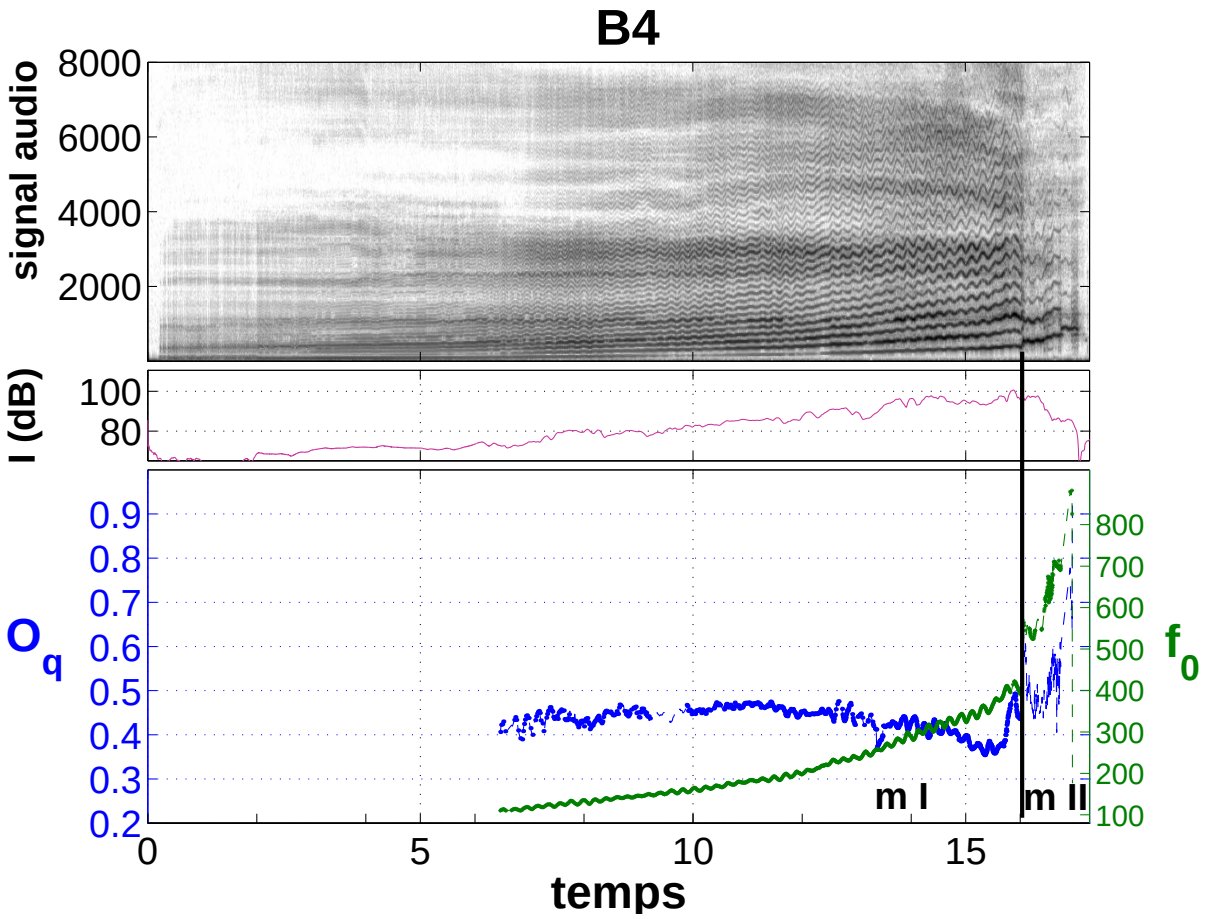


FIG. *F.7 – baryton B4, (CD, piste [34])*

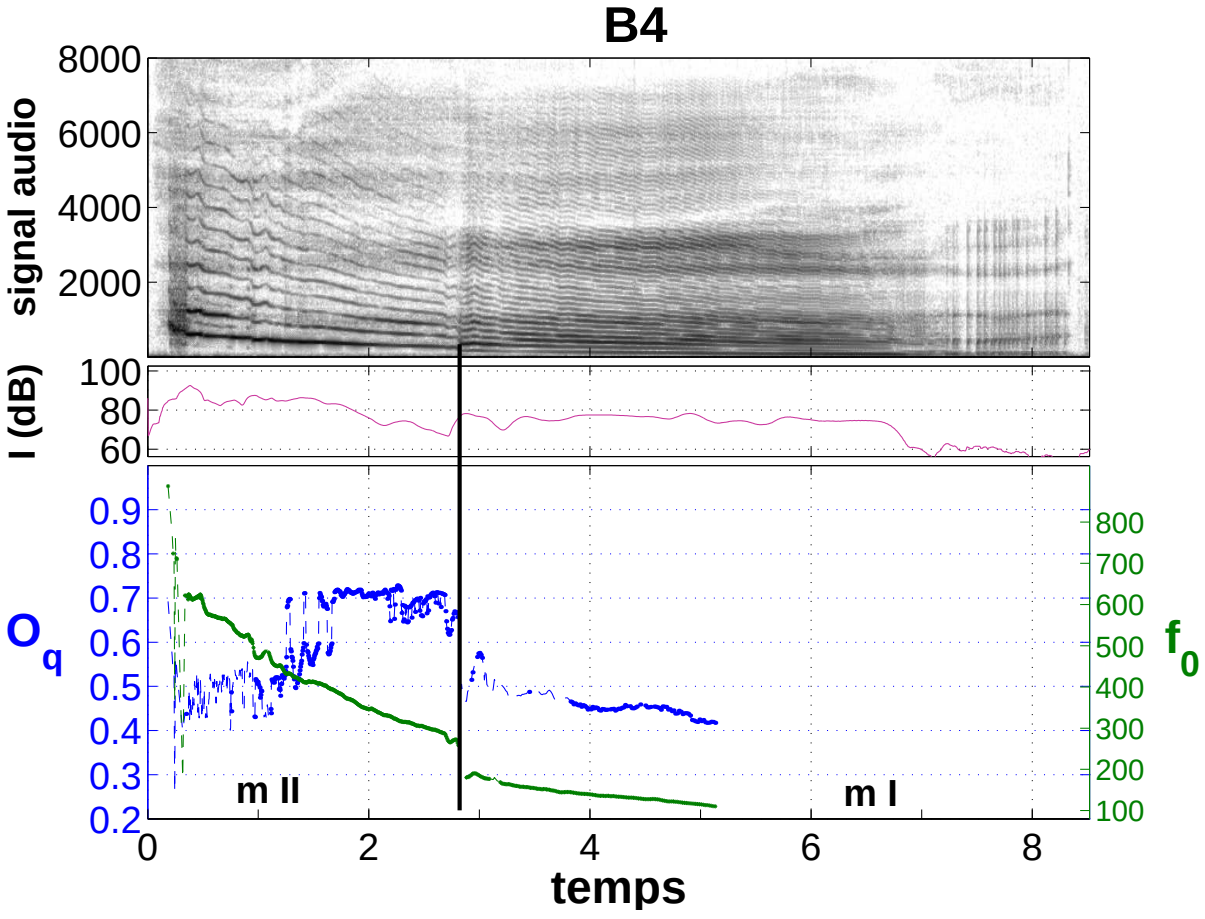


FIG. *F.8 – baryton B4, (CD, piste [35])*

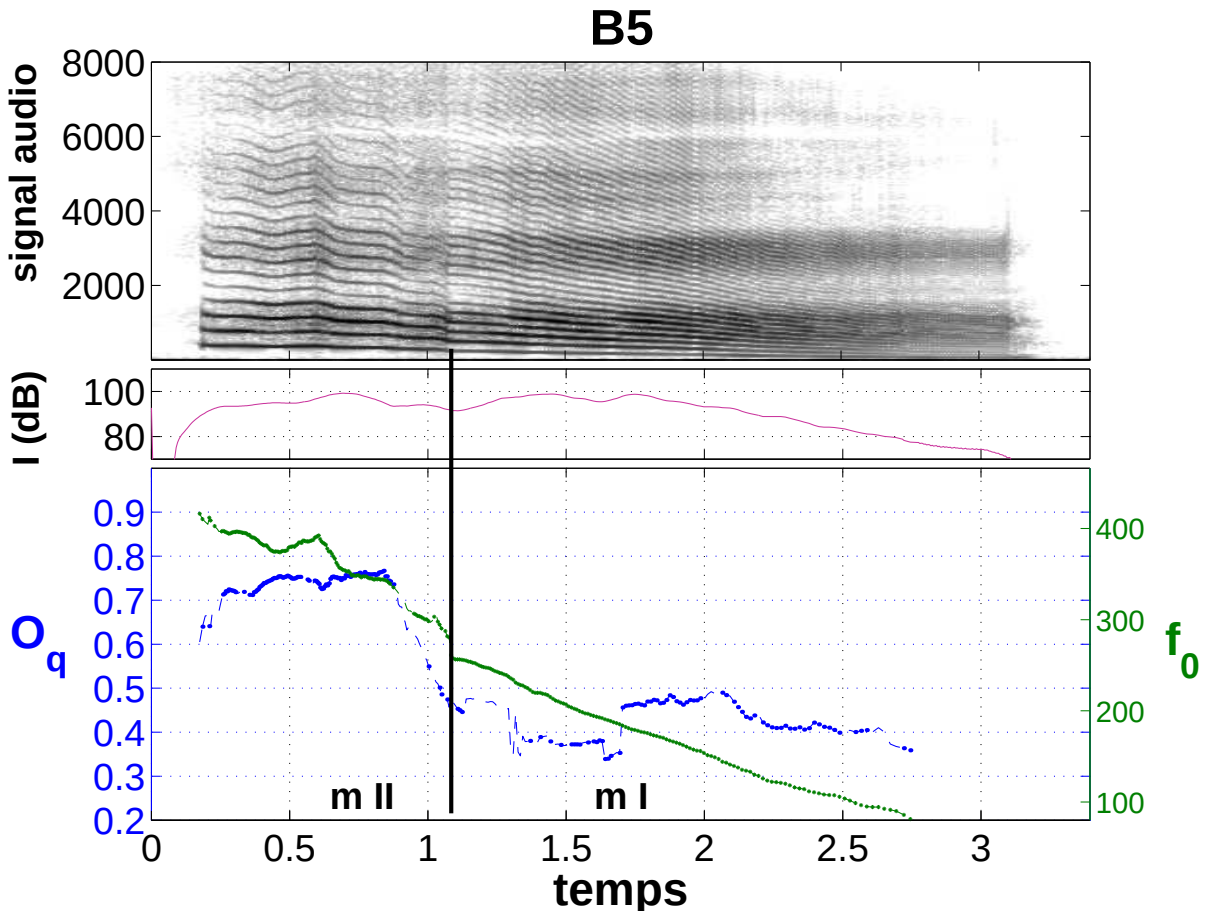


FIG. *F.9 – baryton B5, (CD, piste [36])*

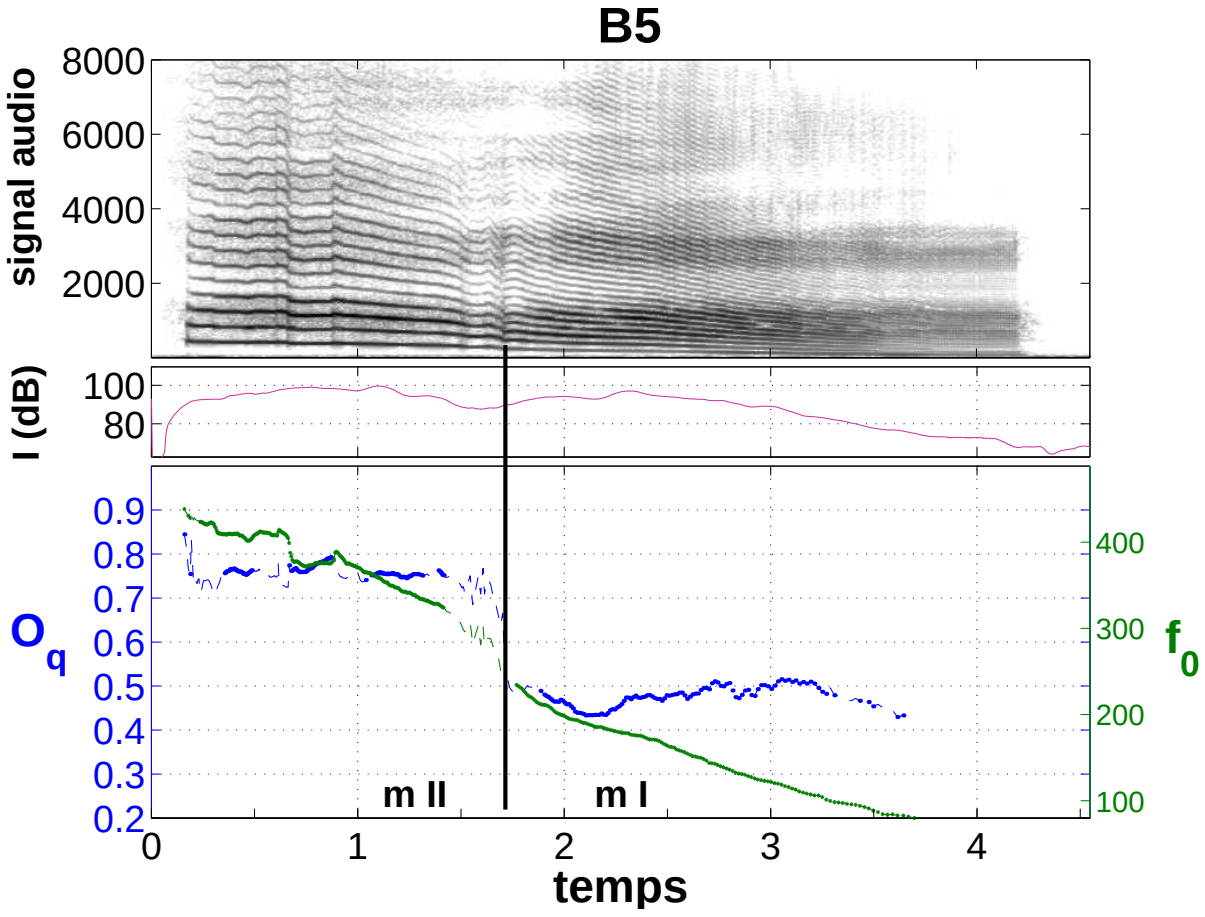


FIG. *F.10 – baryton B5, (CD, piste [37])*

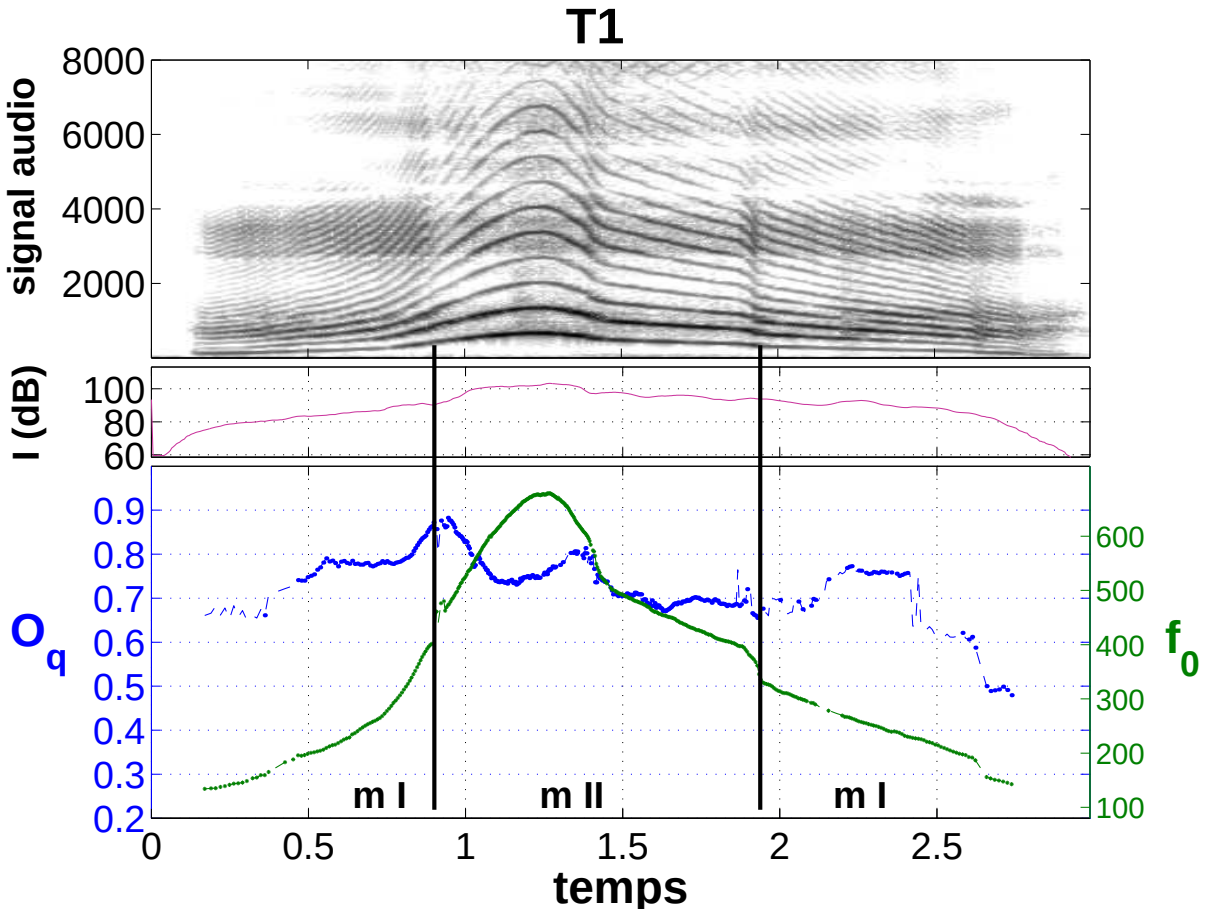


FIG. *F.11 – ténor T1, (CD, piste [39])*

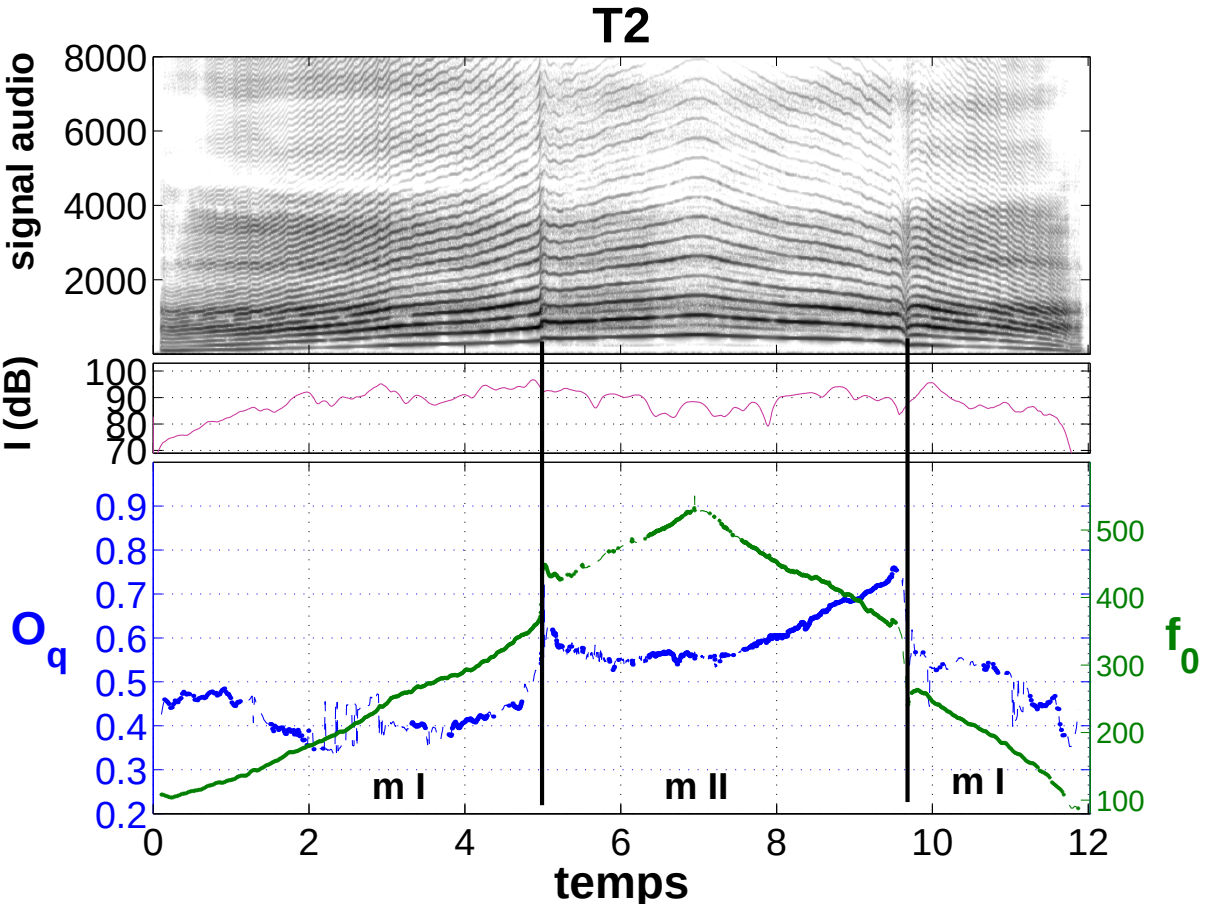


FIG. *F.12 – ténor T2, (CD, piste [40])*

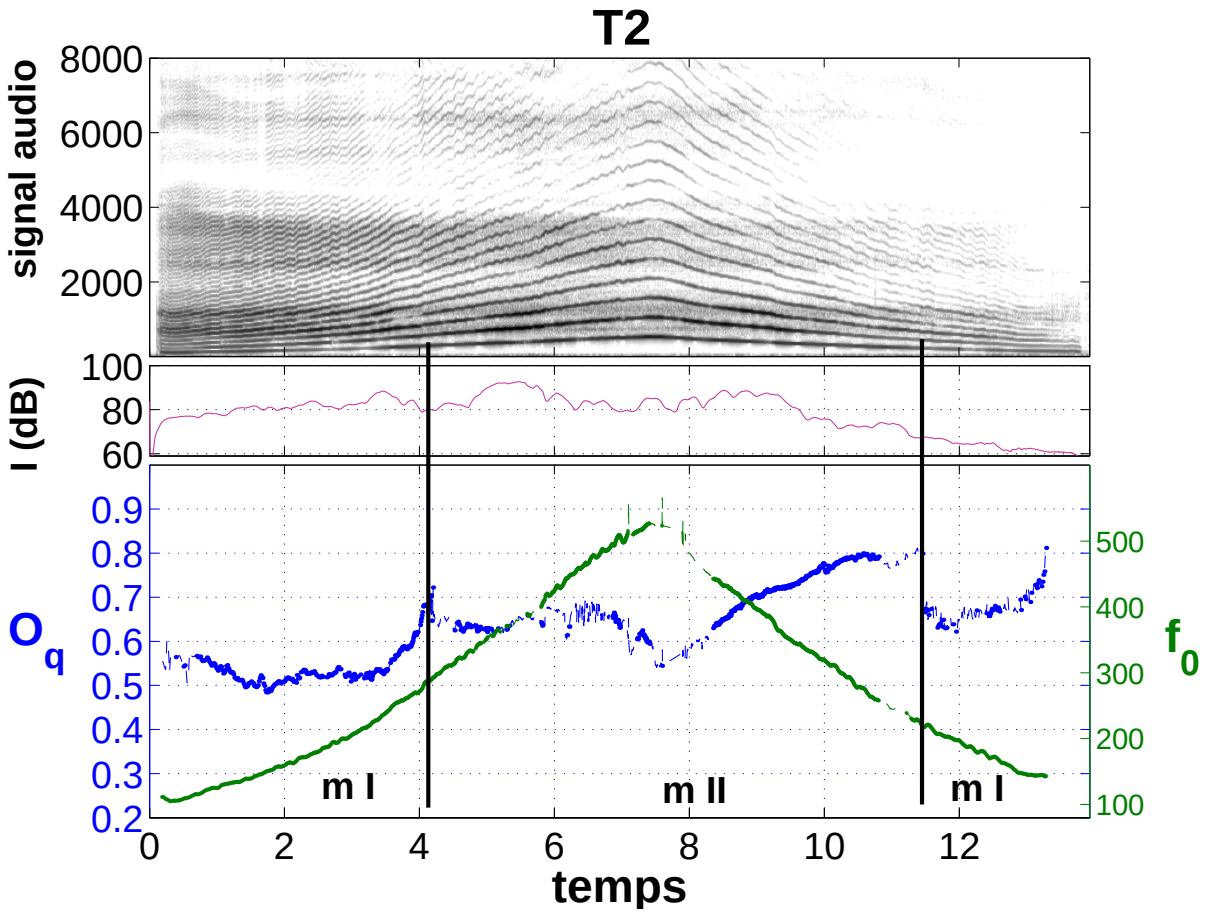


FIG. *F.13 – ténor T2, (CD, piste [41])*

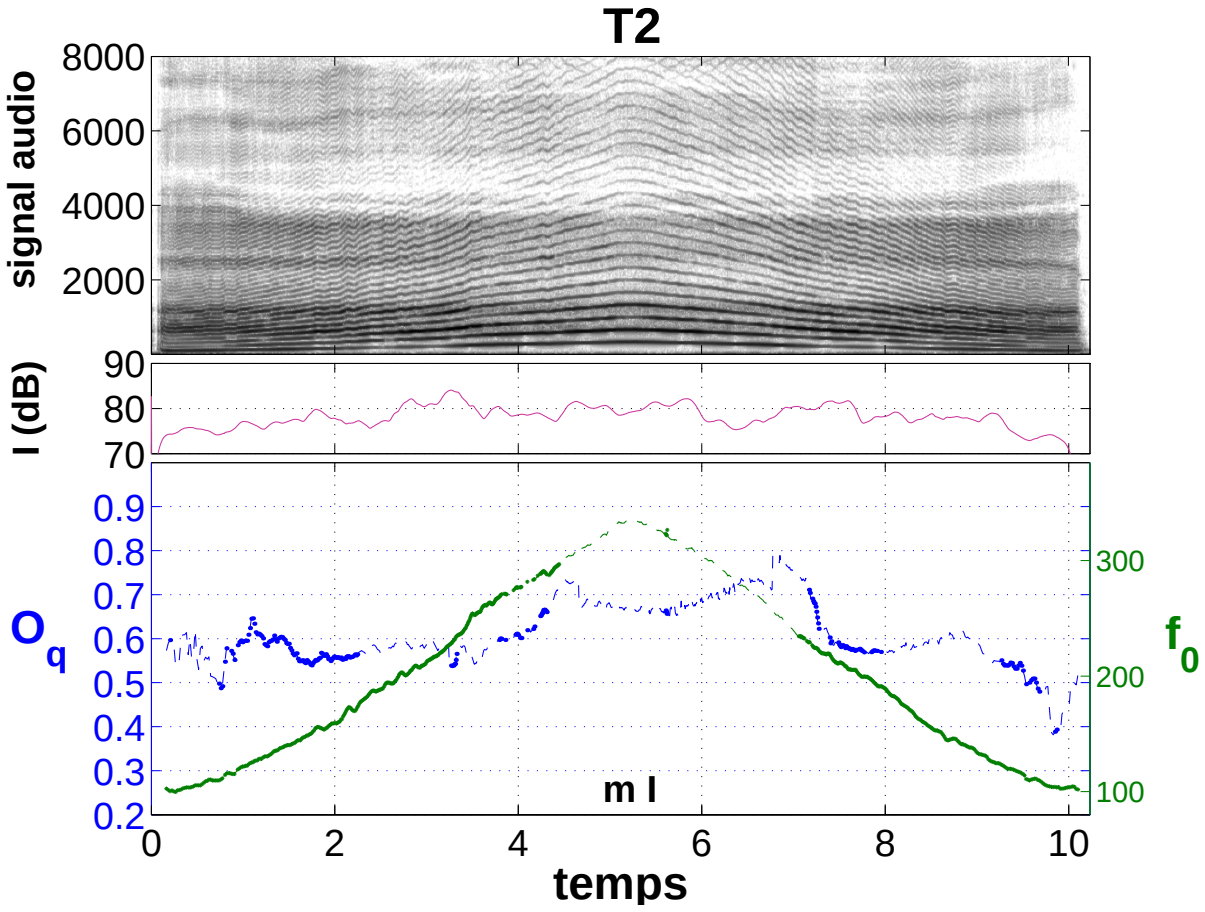


FIG. *F.14 – ténor T2, (CD, piste [42])*

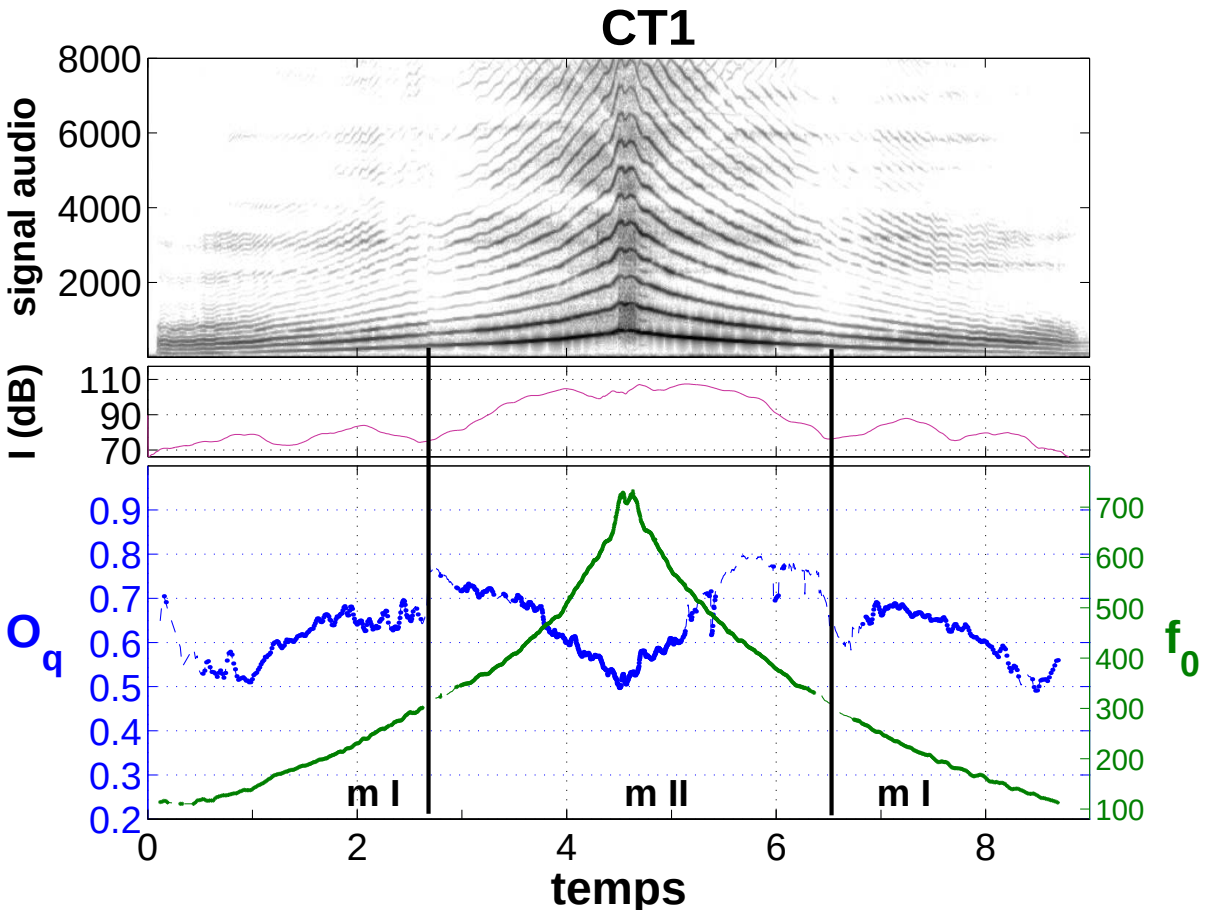


FIG. *F.15 – contre-ténor CT1, (CD, piste [43])*

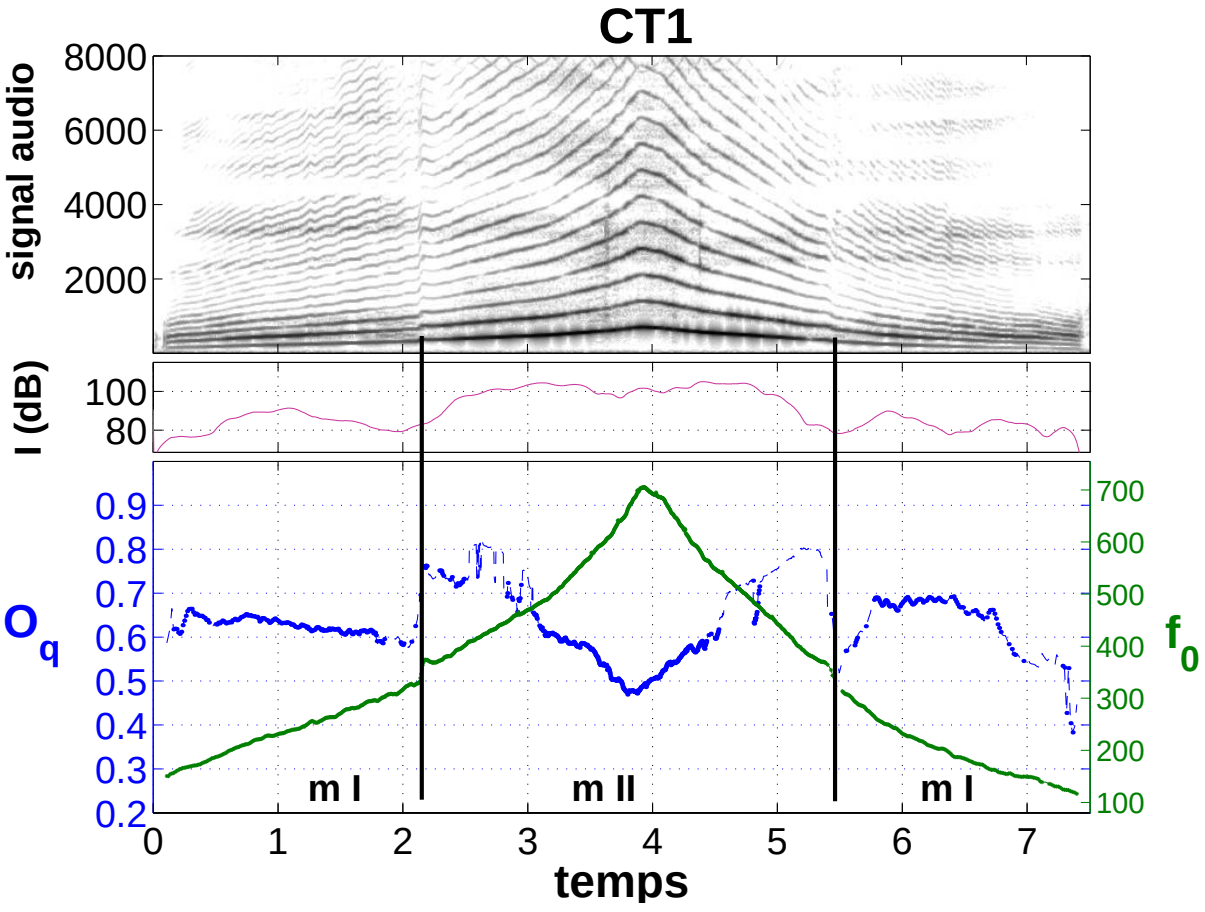


FIG. *F.16 – contre-ténor CT1, (CD, piste [44])*

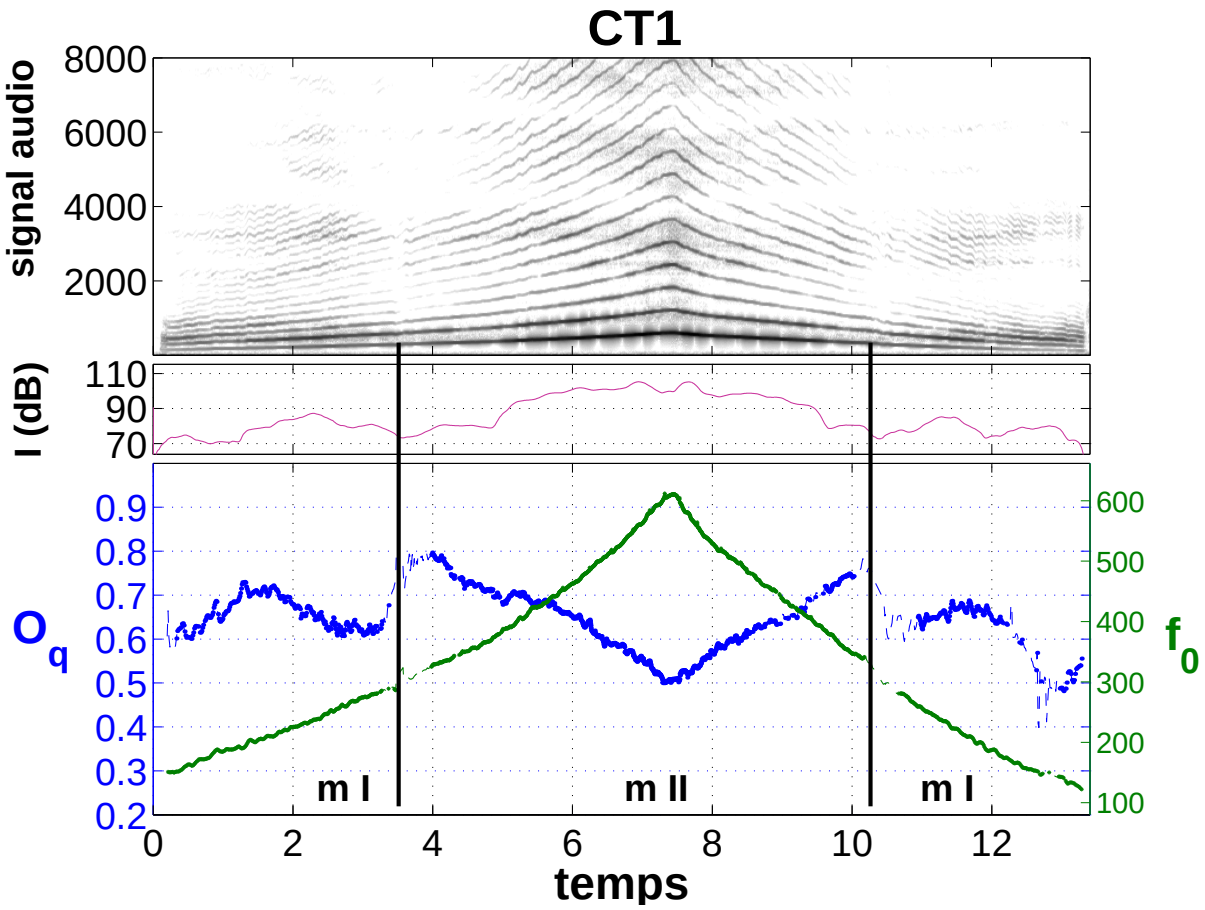


FIG. *F.17 – contre-ténor CT1, (CD, piste [45])*

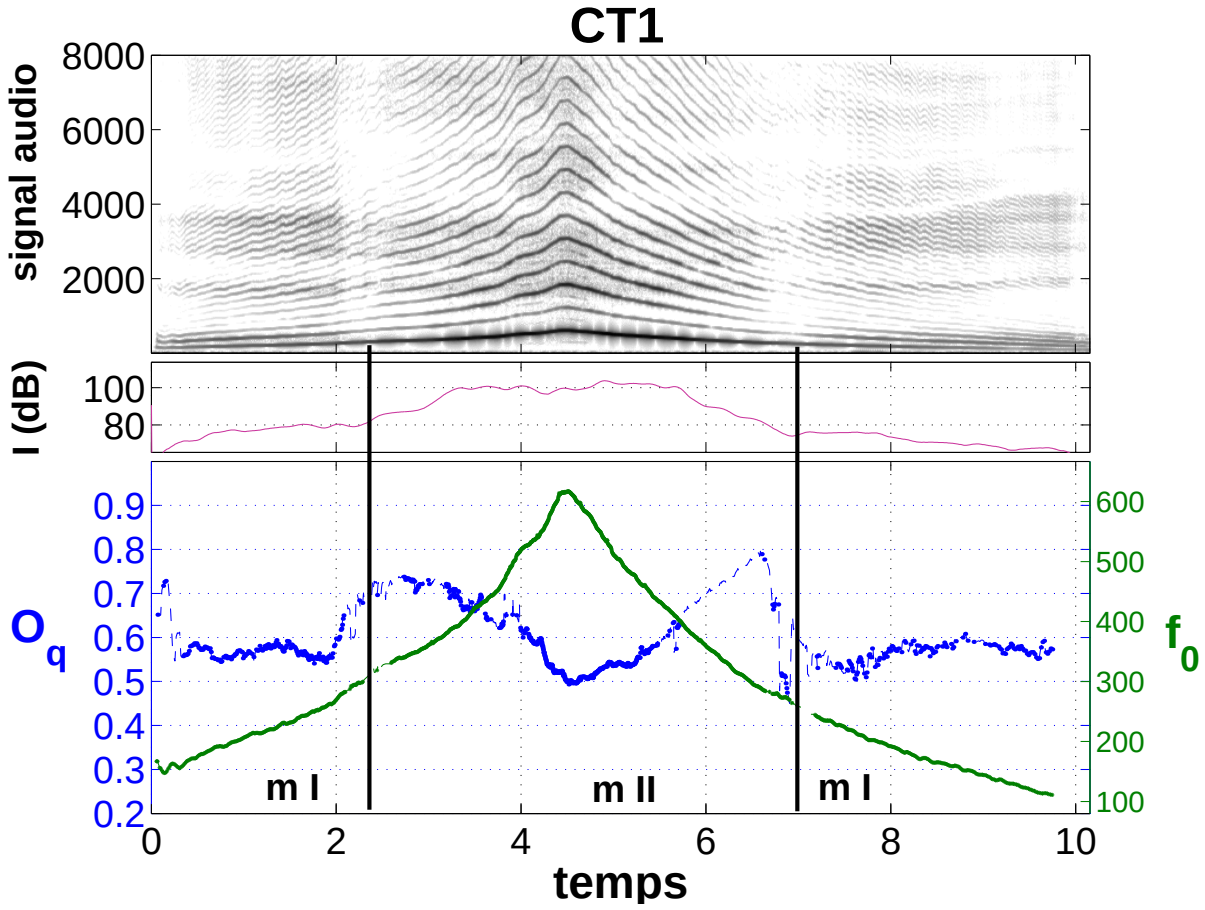


FIG. *F.18 – contre-ténor CT1, (CD, piste [46])*

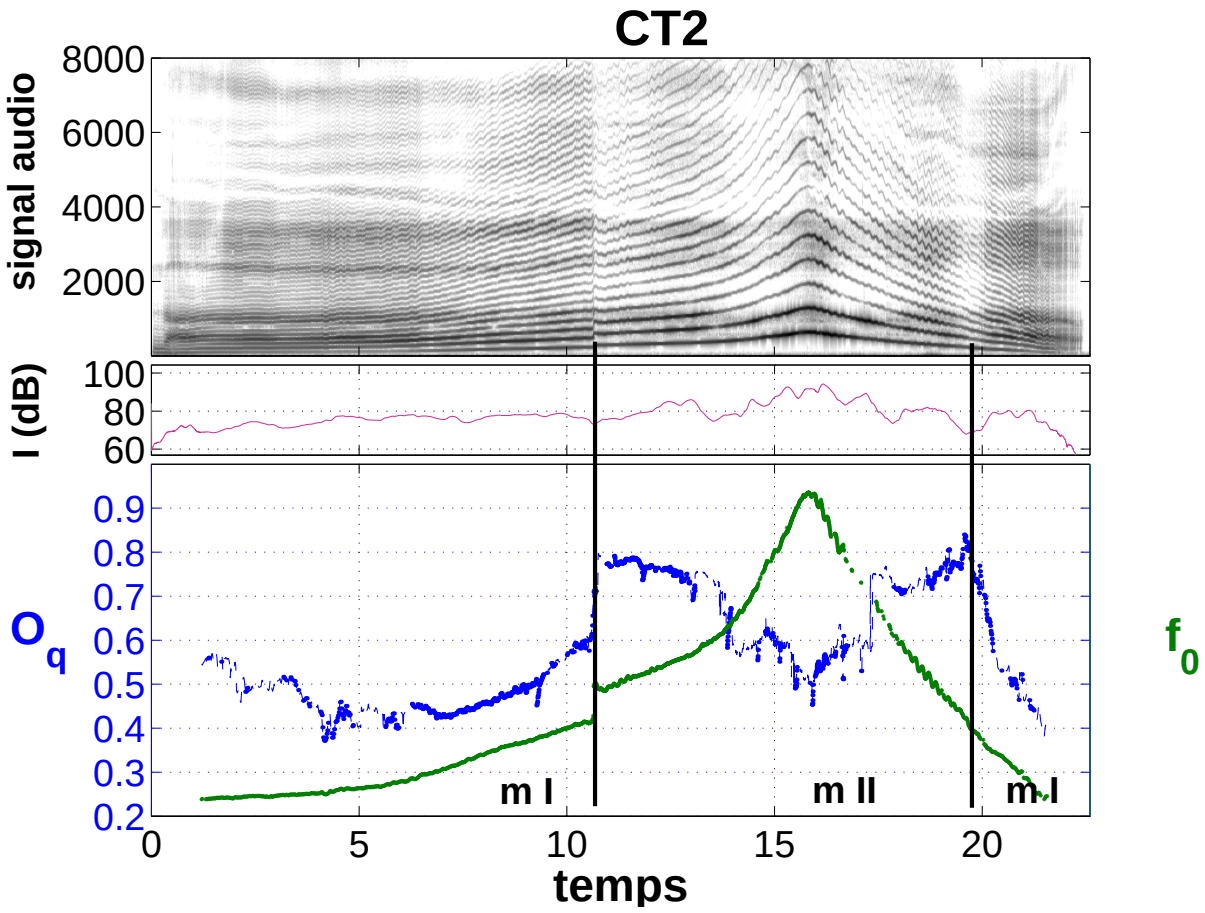


FIG. *F.19 – contre-ténor CT2, (CD, piste [47])*

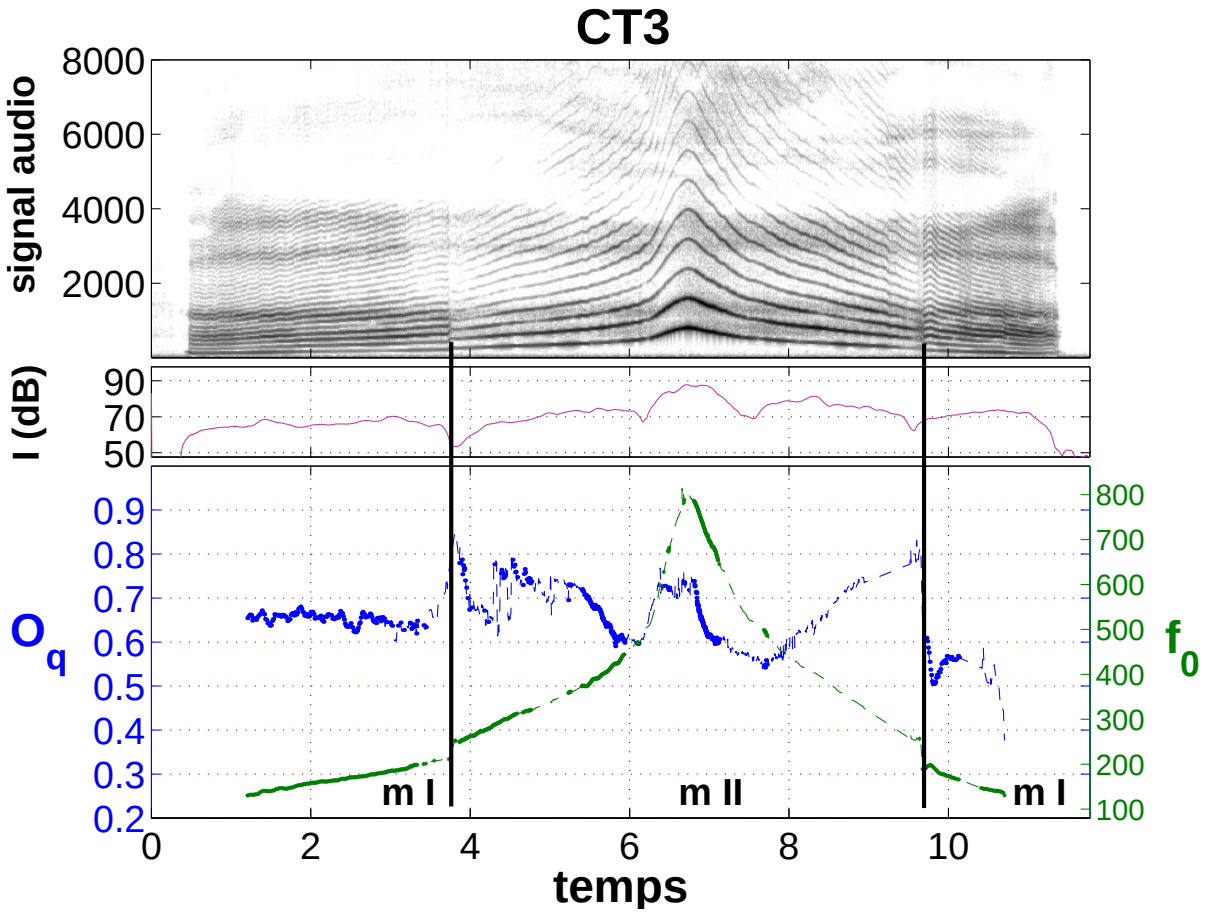


FIG. *F.20 – contre-ténor CT3, (CD, piste [48])*

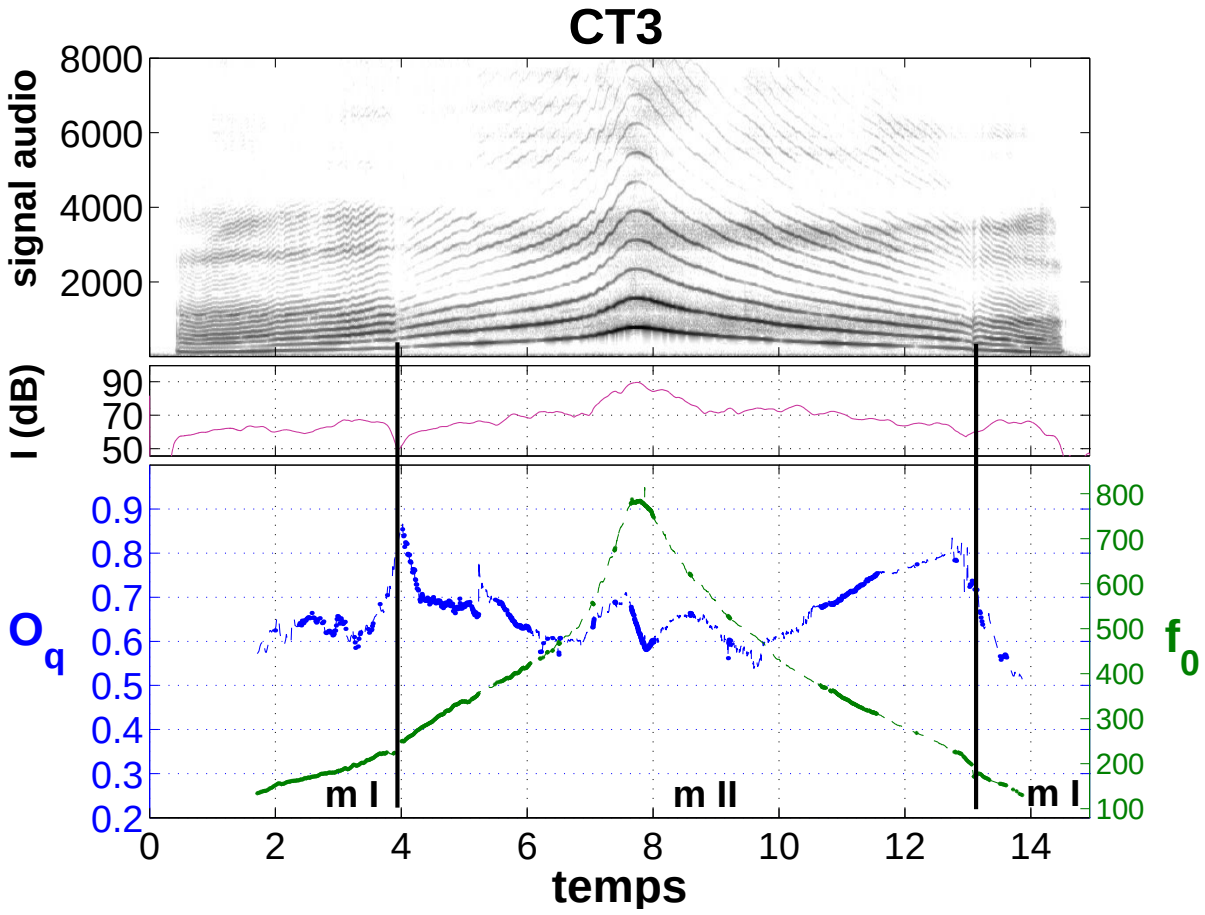


FIG. *F.21 – contre-ténor CT3, (CD, piste [49])*

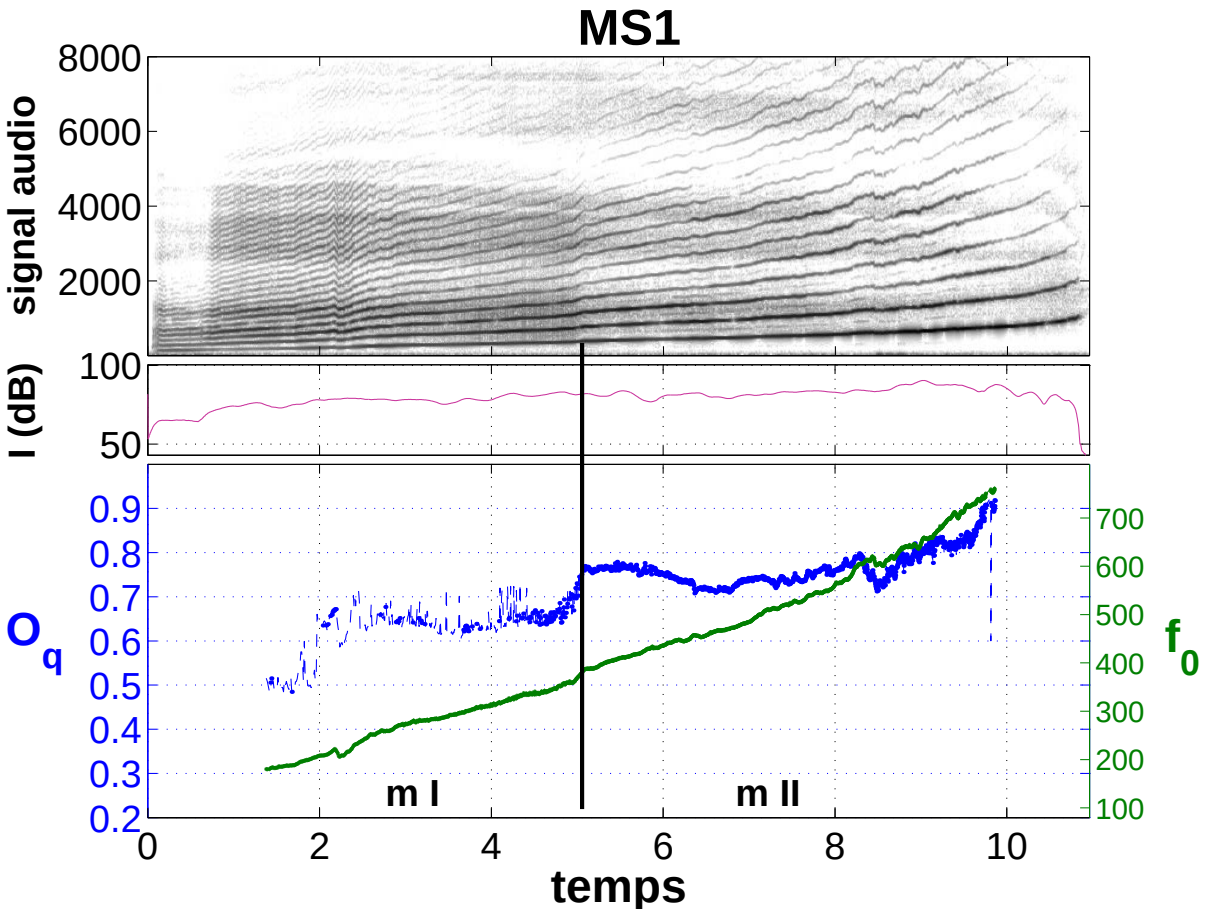


FIG. *F.22 – mezzo-soprano MS1, (CD, piste [50])*

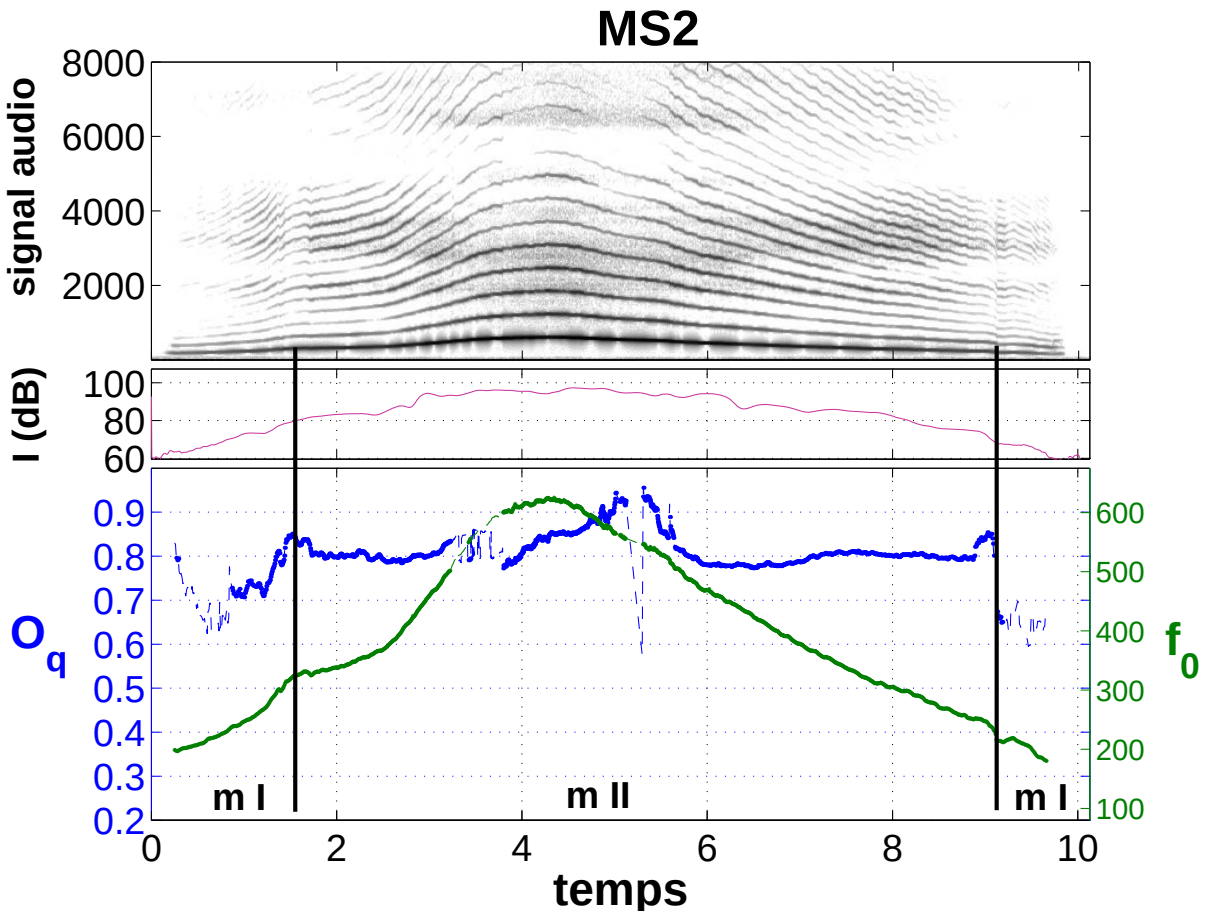


FIG. *F.23 – mezzo-soprano MS2, (CD, piste [51])*

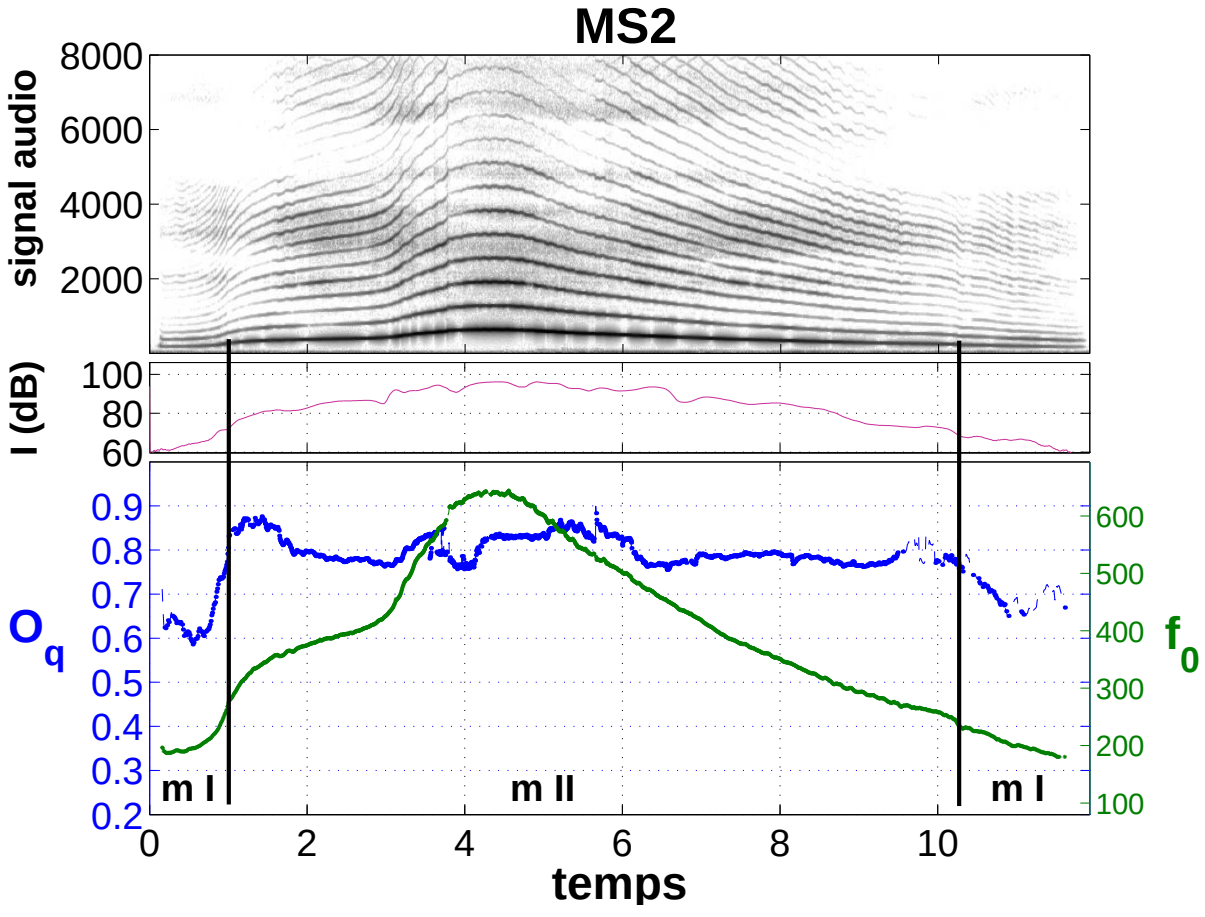


FIG. *F.24 – mezzo-soprano MS2, (CD, piste [52])*

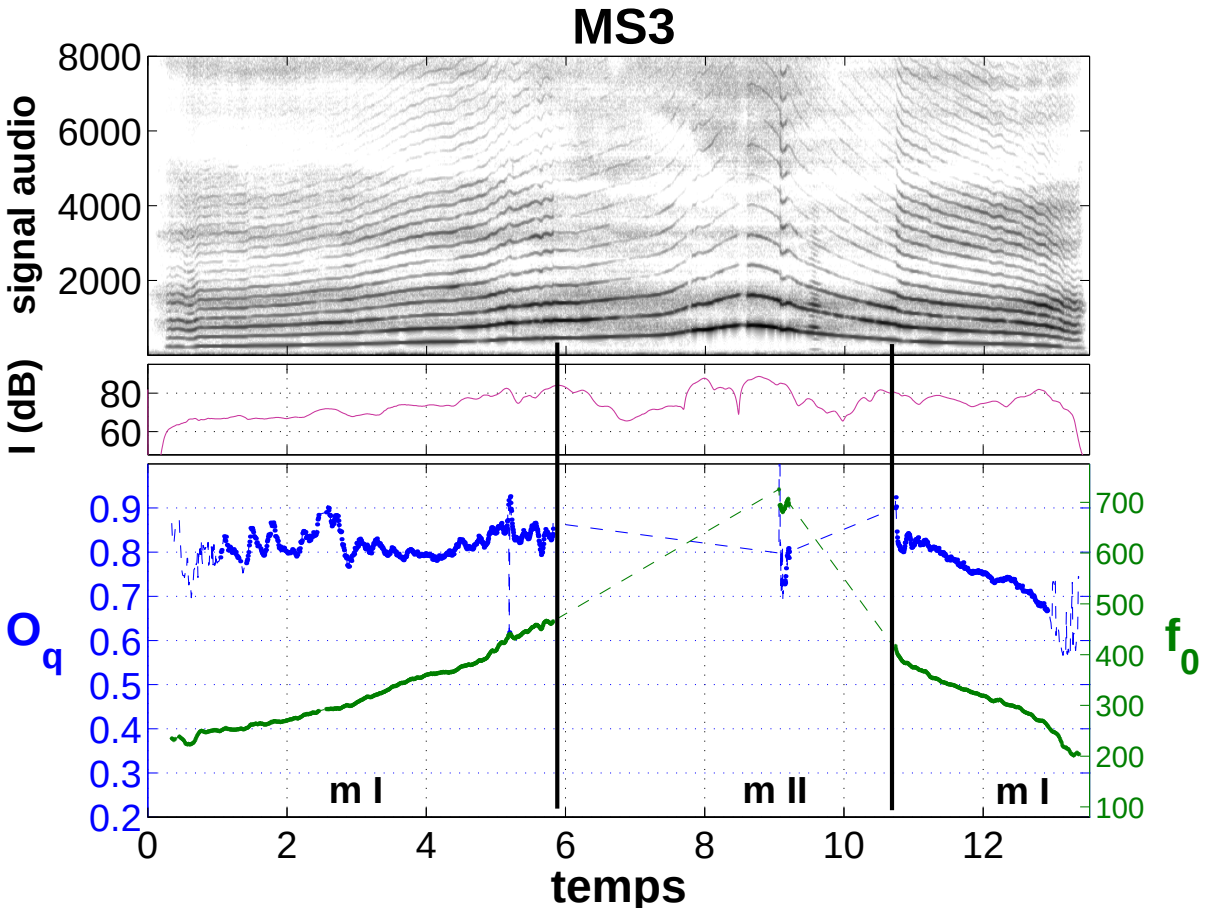


FIG. *F.25 – mezzo-soprano MS3, (CD, piste [53])*

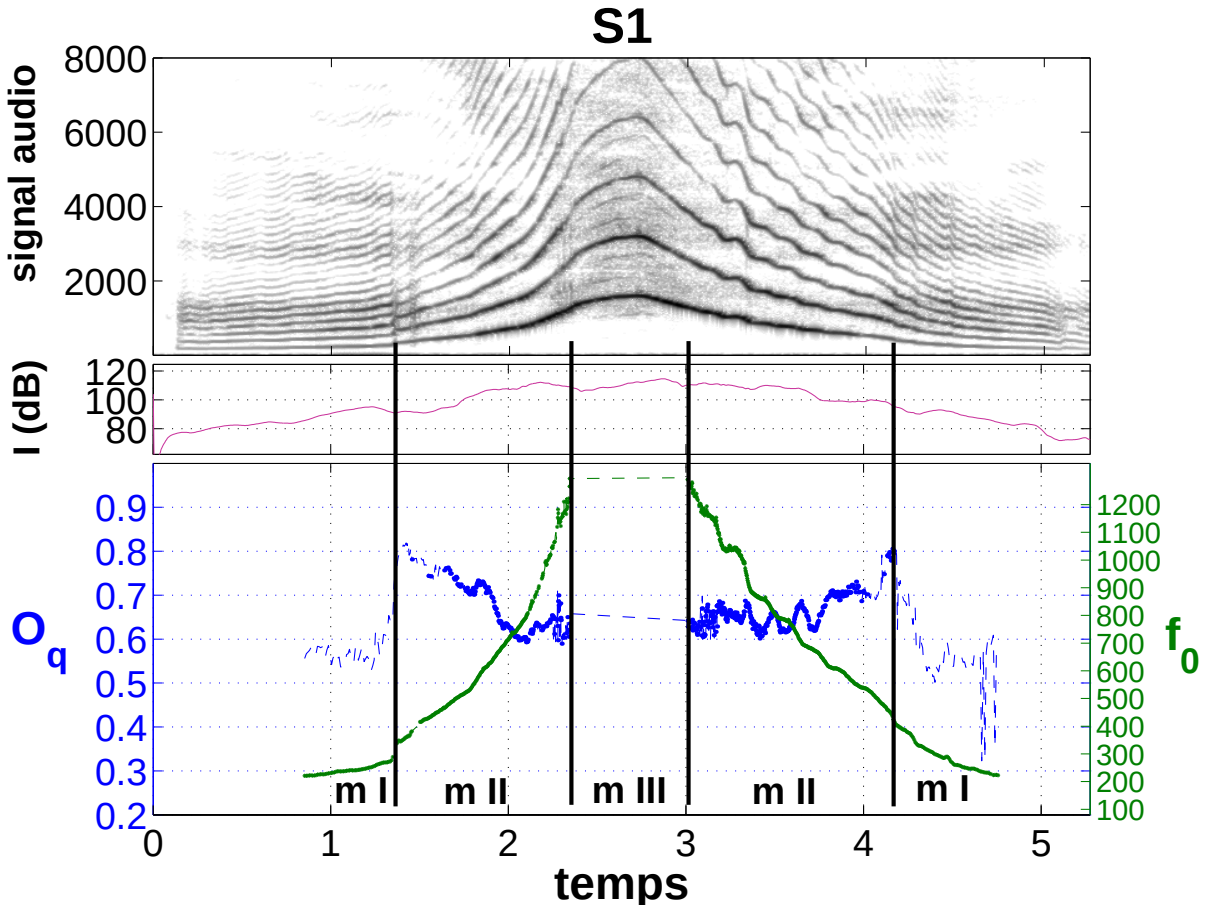


FIG. *F.26 – soprano S1, (CD, piste [54])*

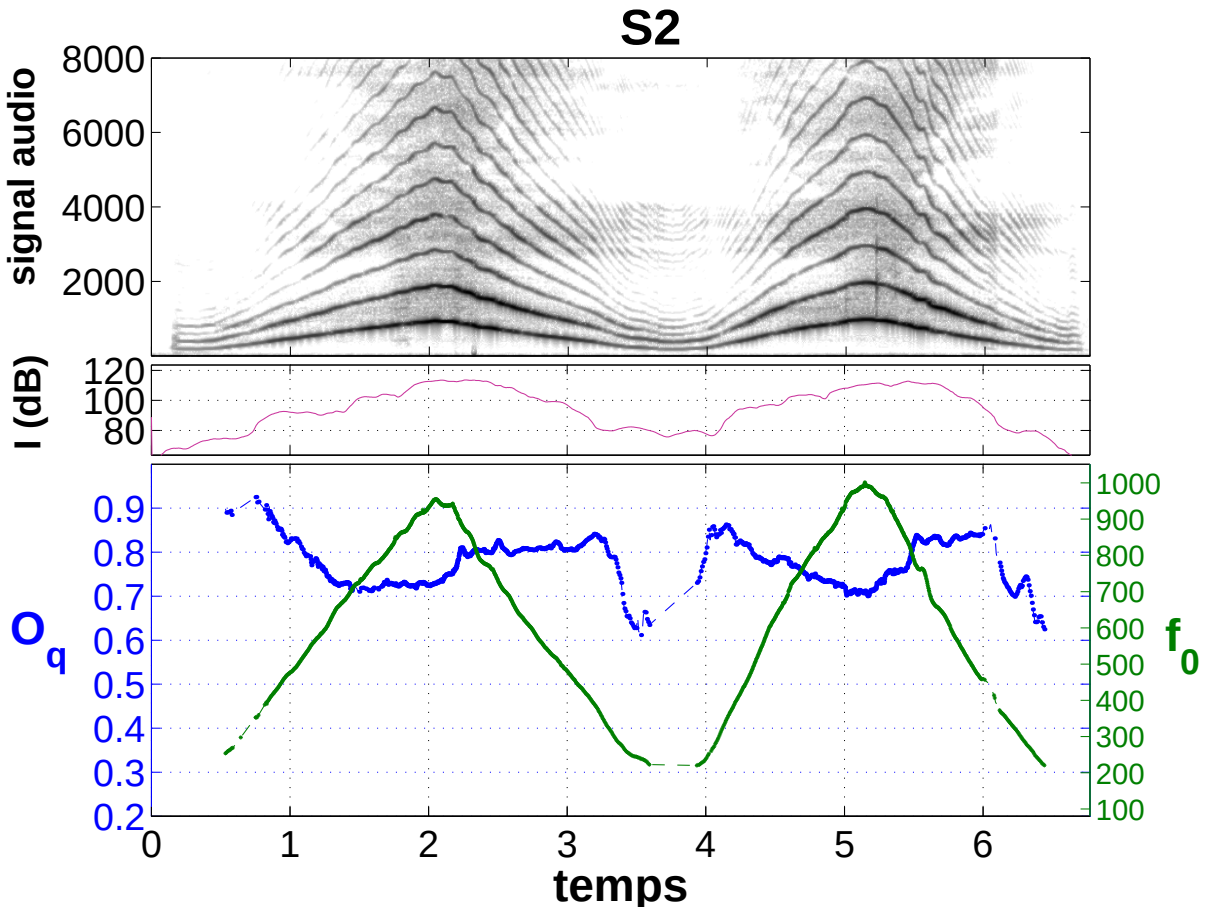


FIG. *F.27 – soprano S2, (CD, piste [55])*

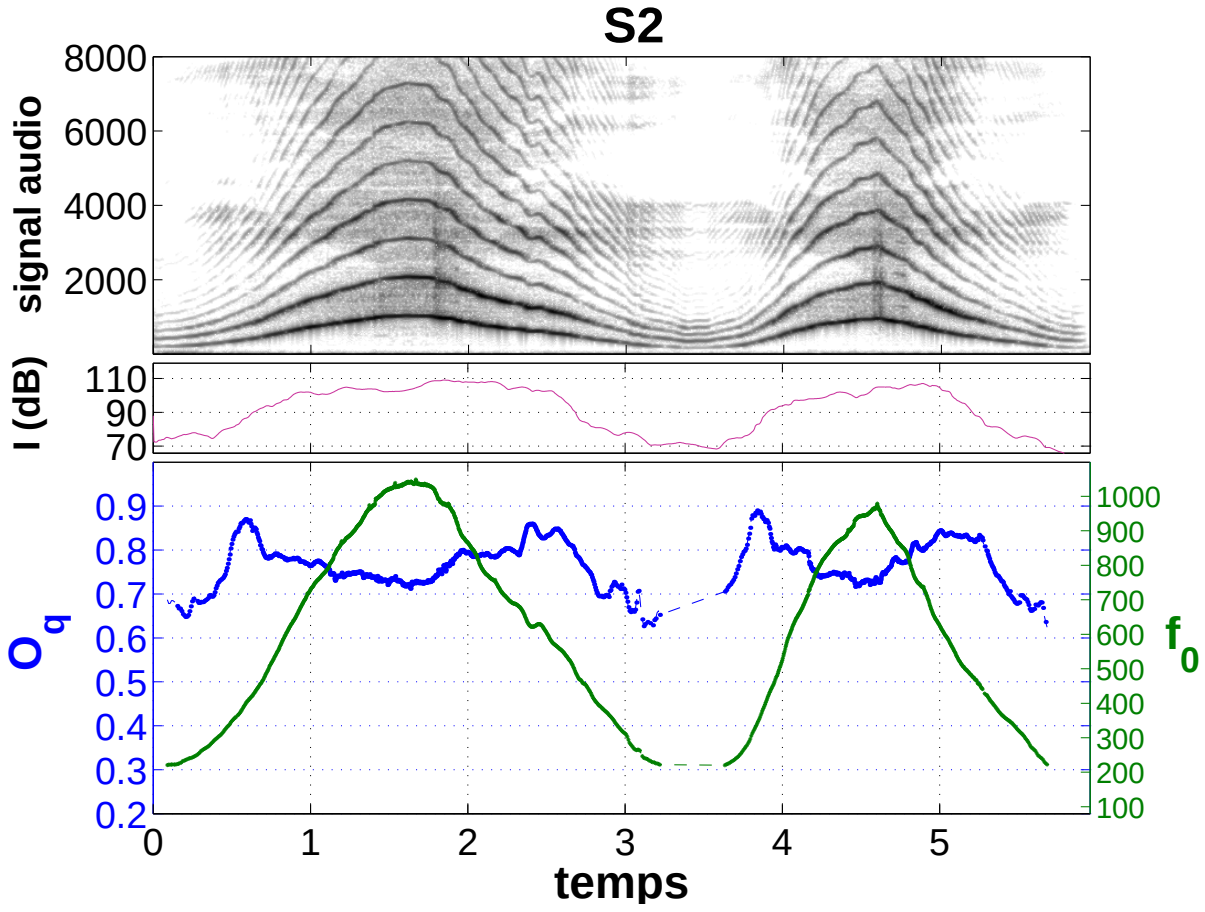


FIG. *F.28 – soprano S2, (CD, piste [56])*

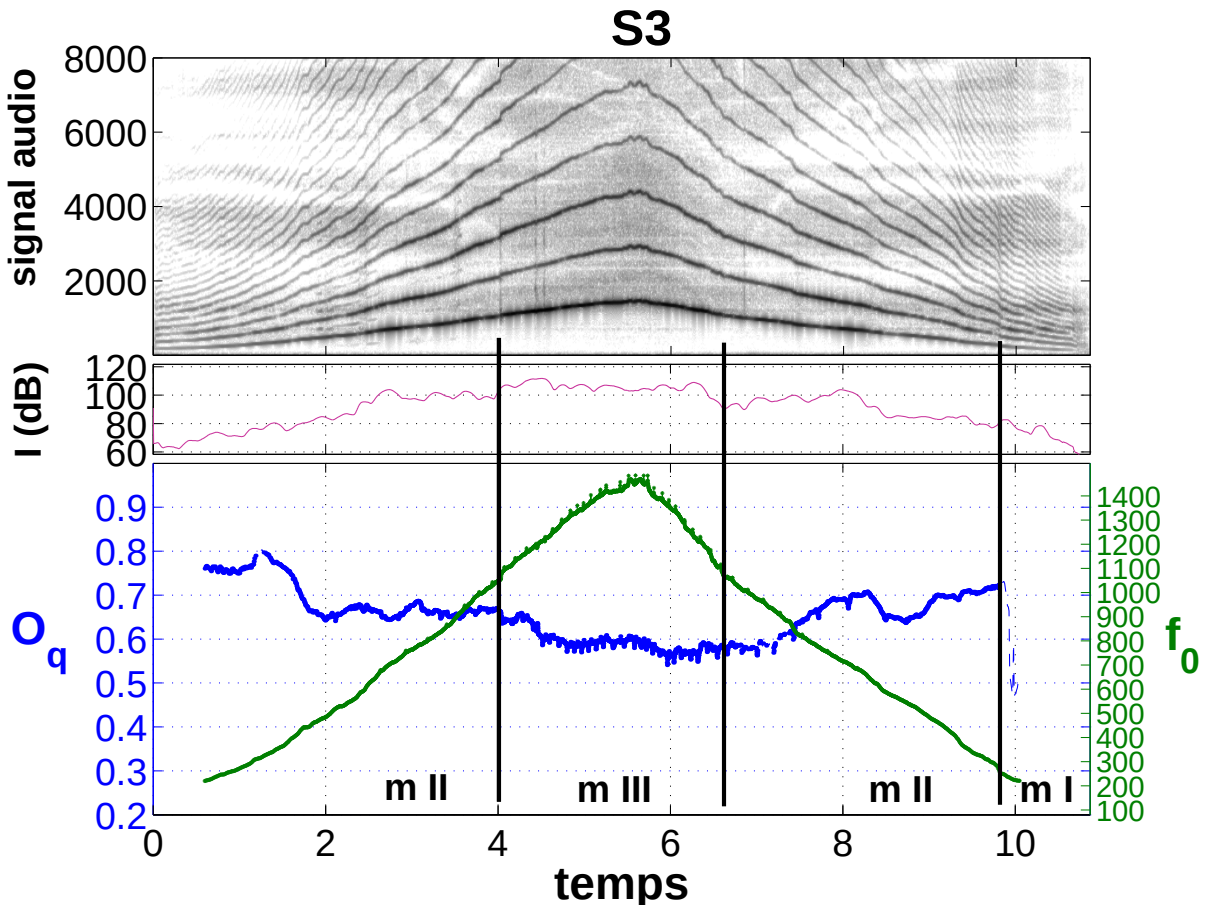


FIG. *F.29 – soprano S3, (CD, piste [57])*

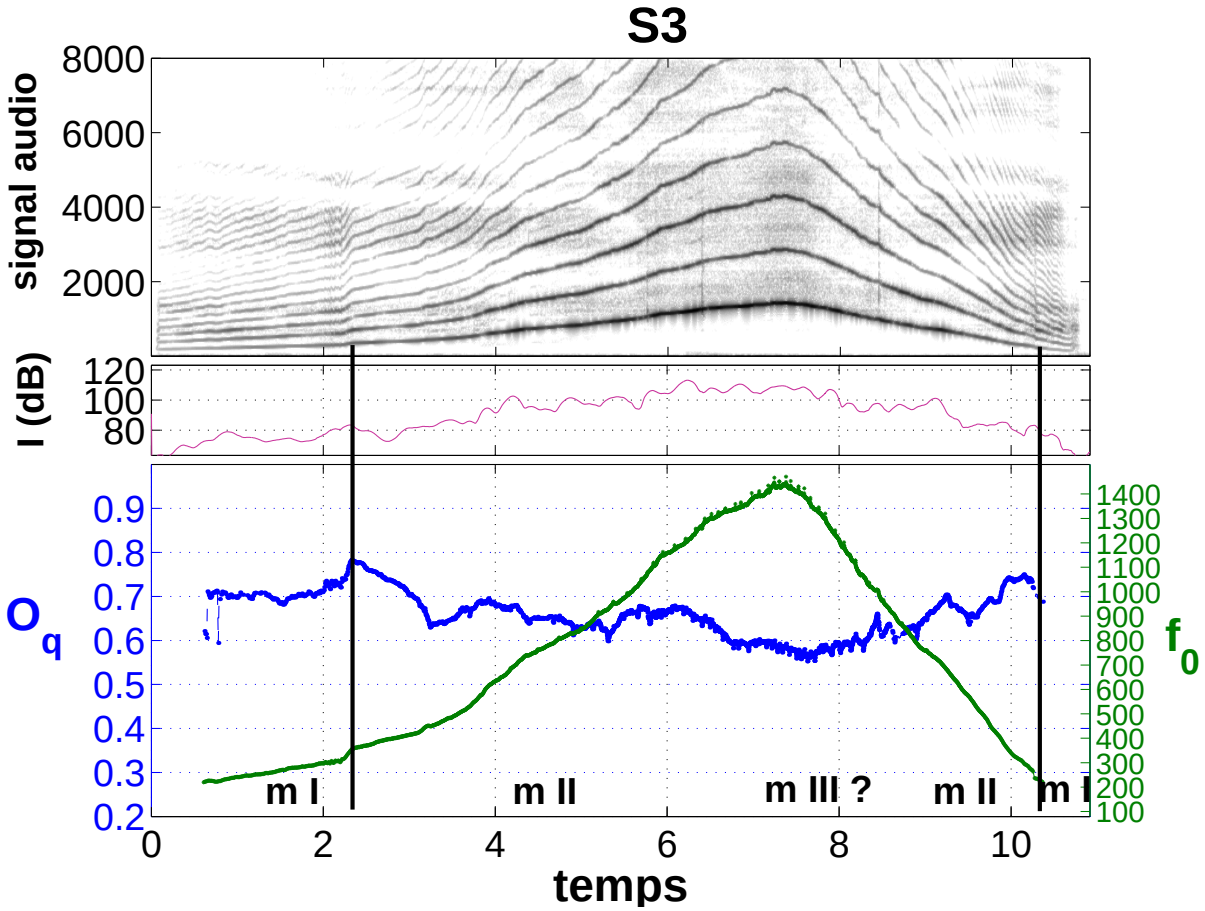


FIG. *F.30 – soprano S3, (CD, piste [58])*

Annexe G

Etude des premières mesures de l' Ave Maria de Gounod

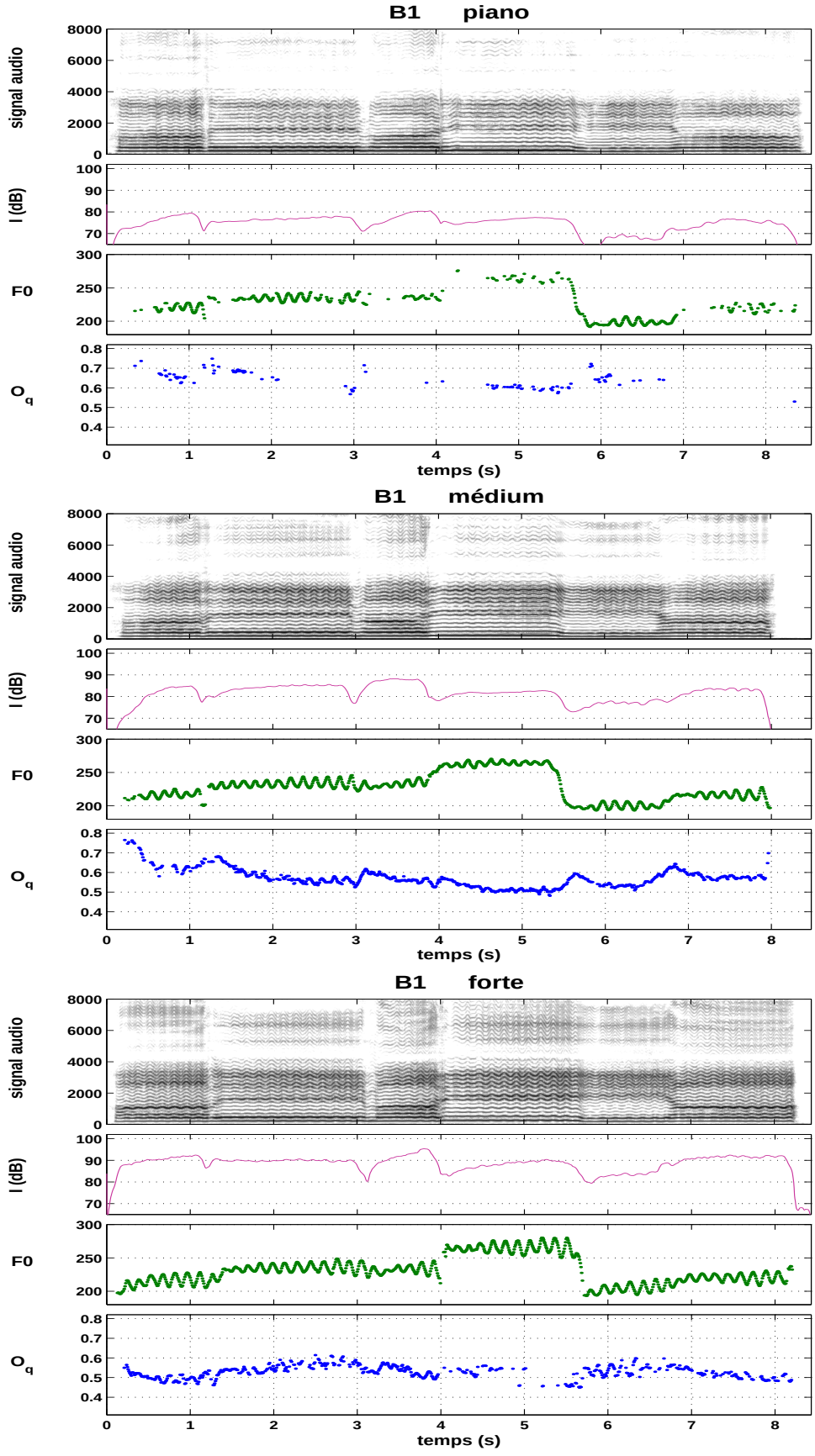


FIG. *G.1 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryton B1 (CD, piste [59])*

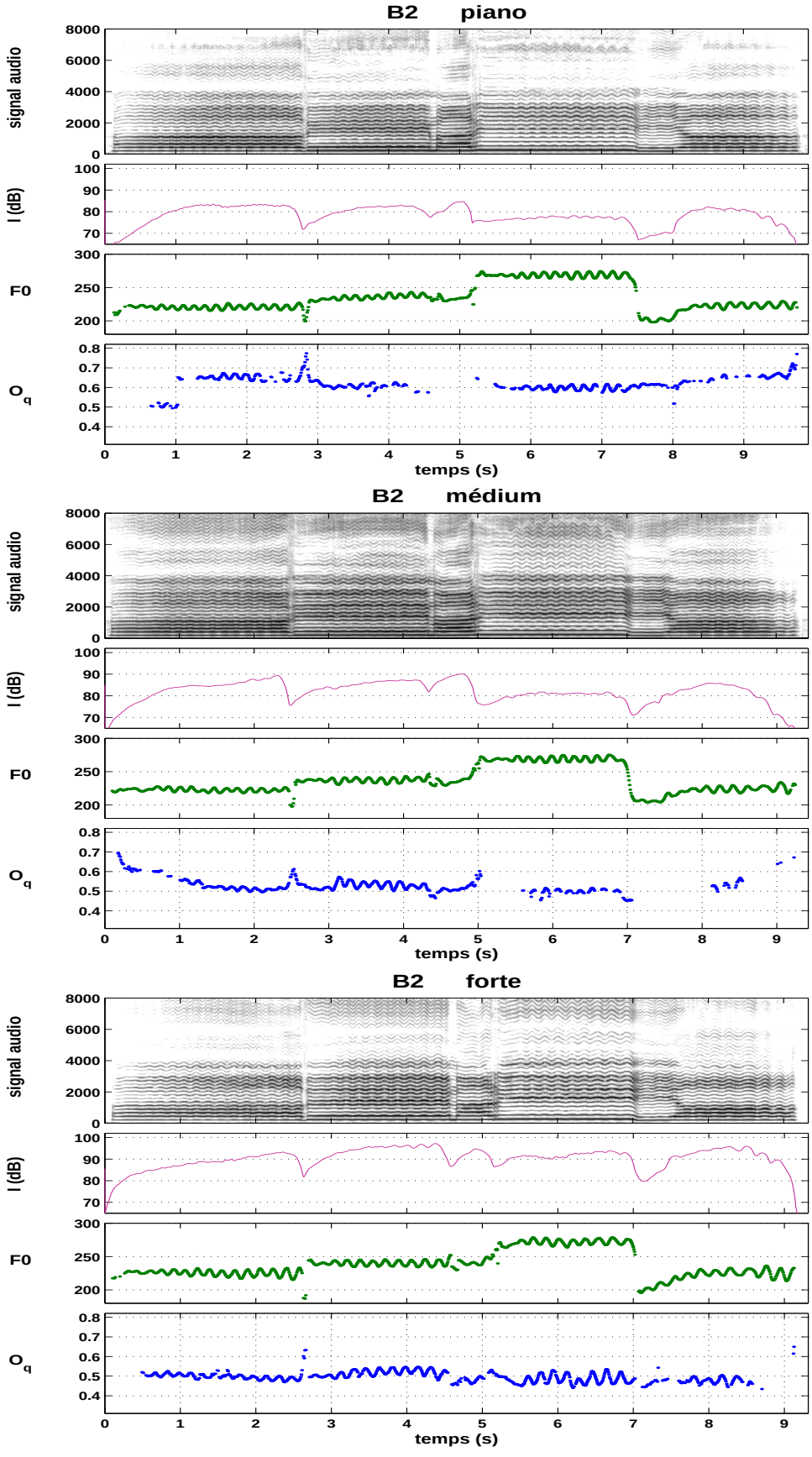


FIG. *G.2 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryton B2 (CD, piste [60])*

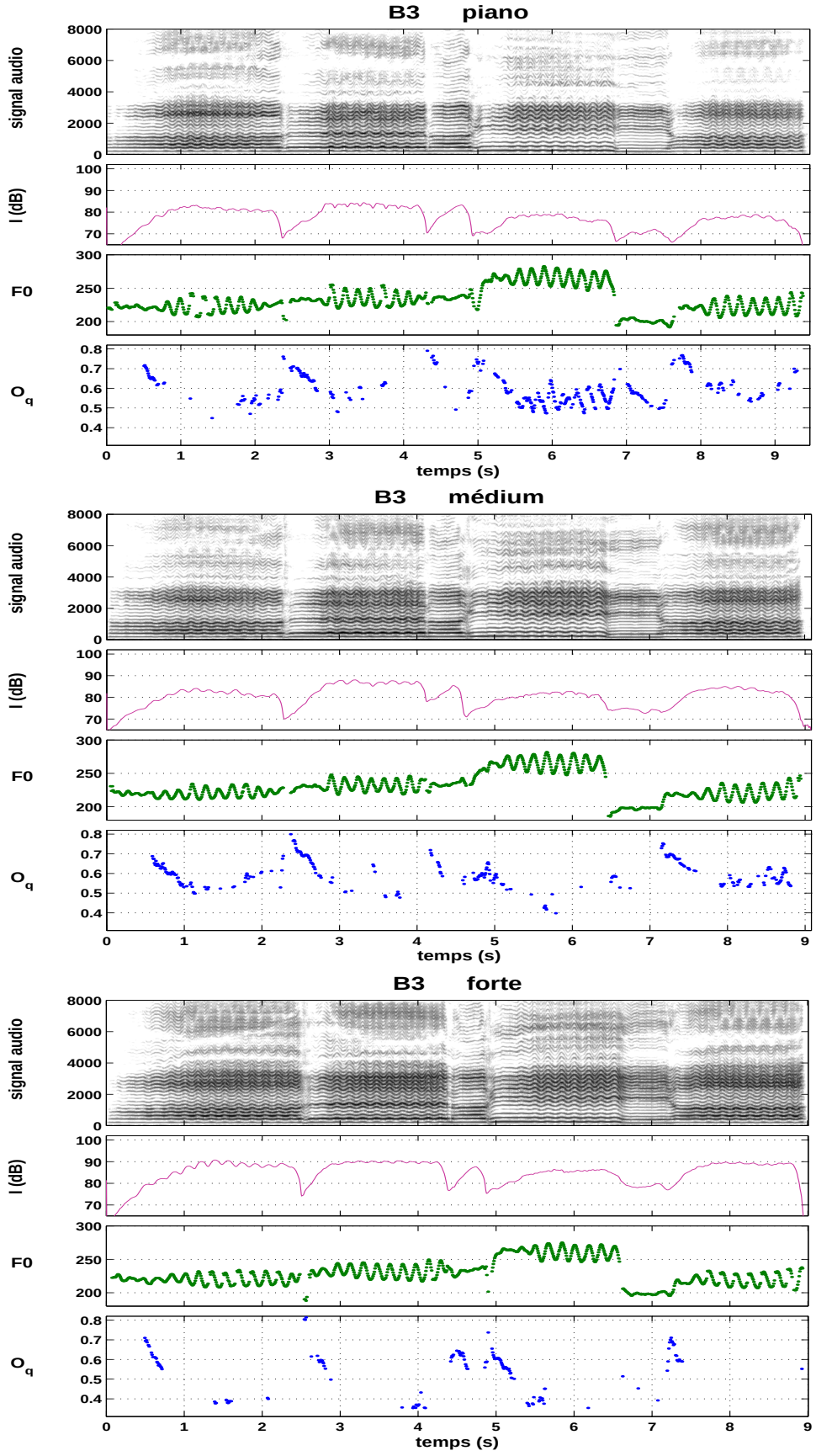


FIG. *G.3 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryont B3 (CD, piste [61])*

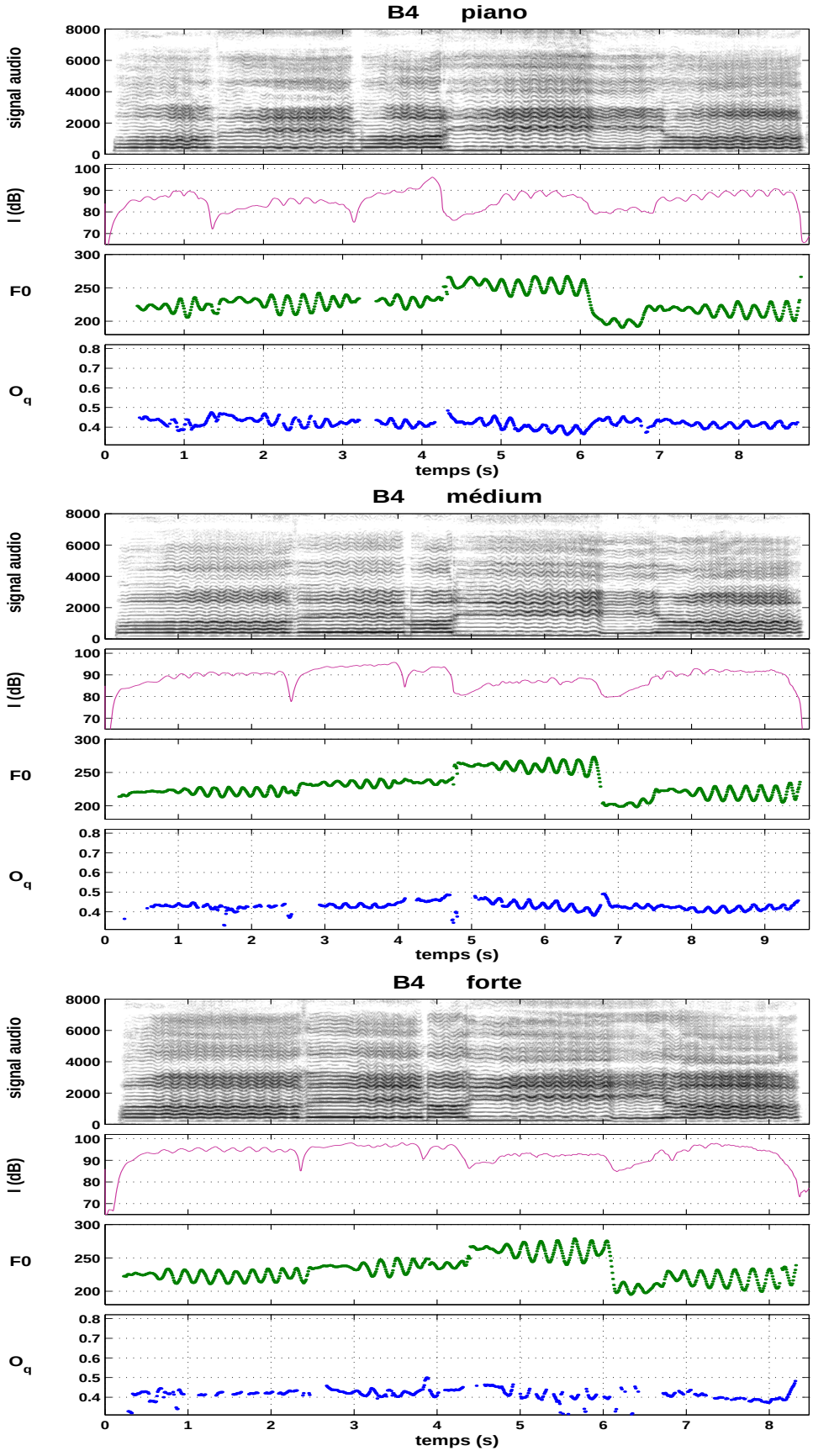


FIG. *G.4 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryton B4 (CD, piste [62])*

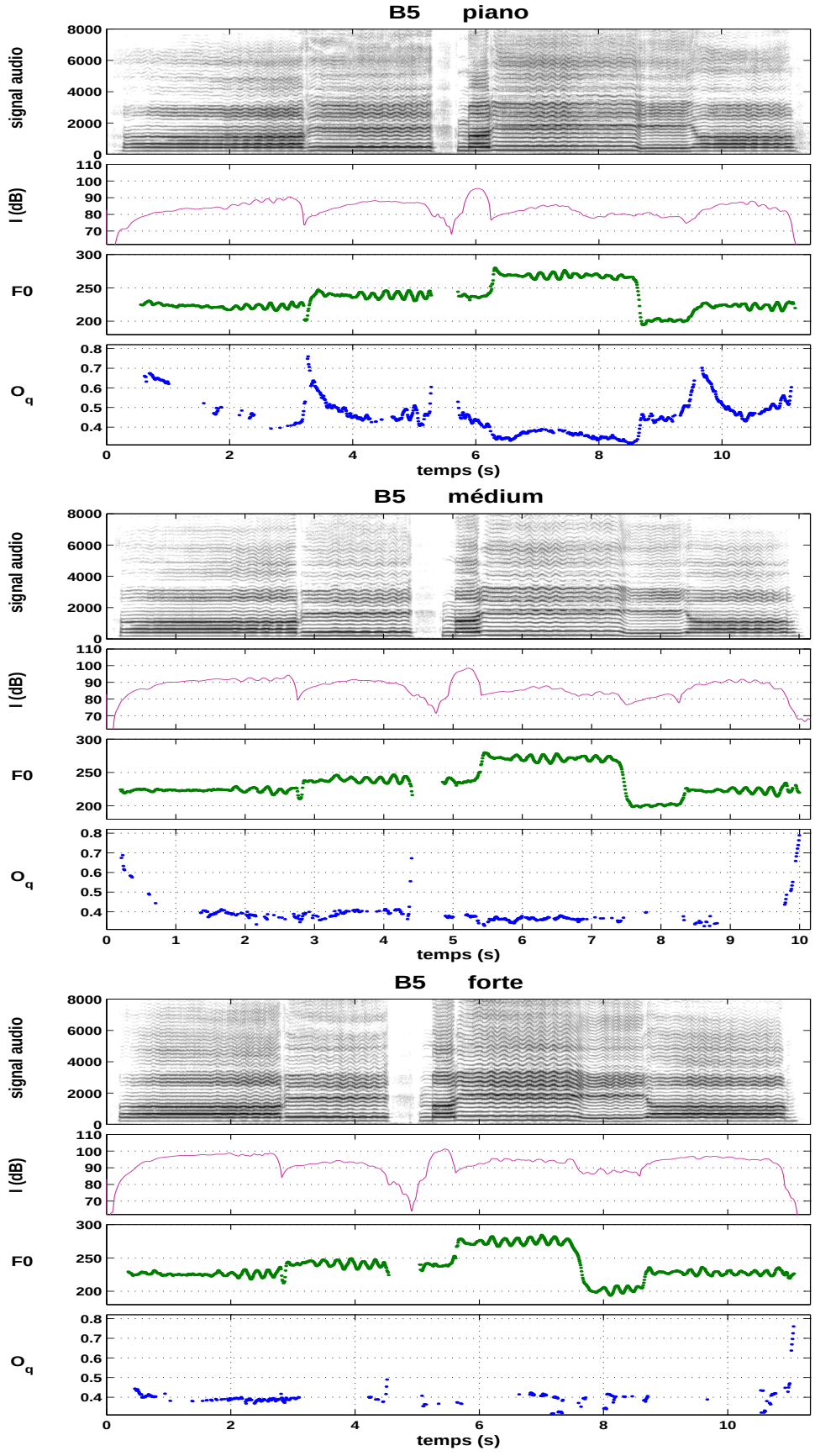


FIG. *G.5 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryton B5 (CD, piste [63])*

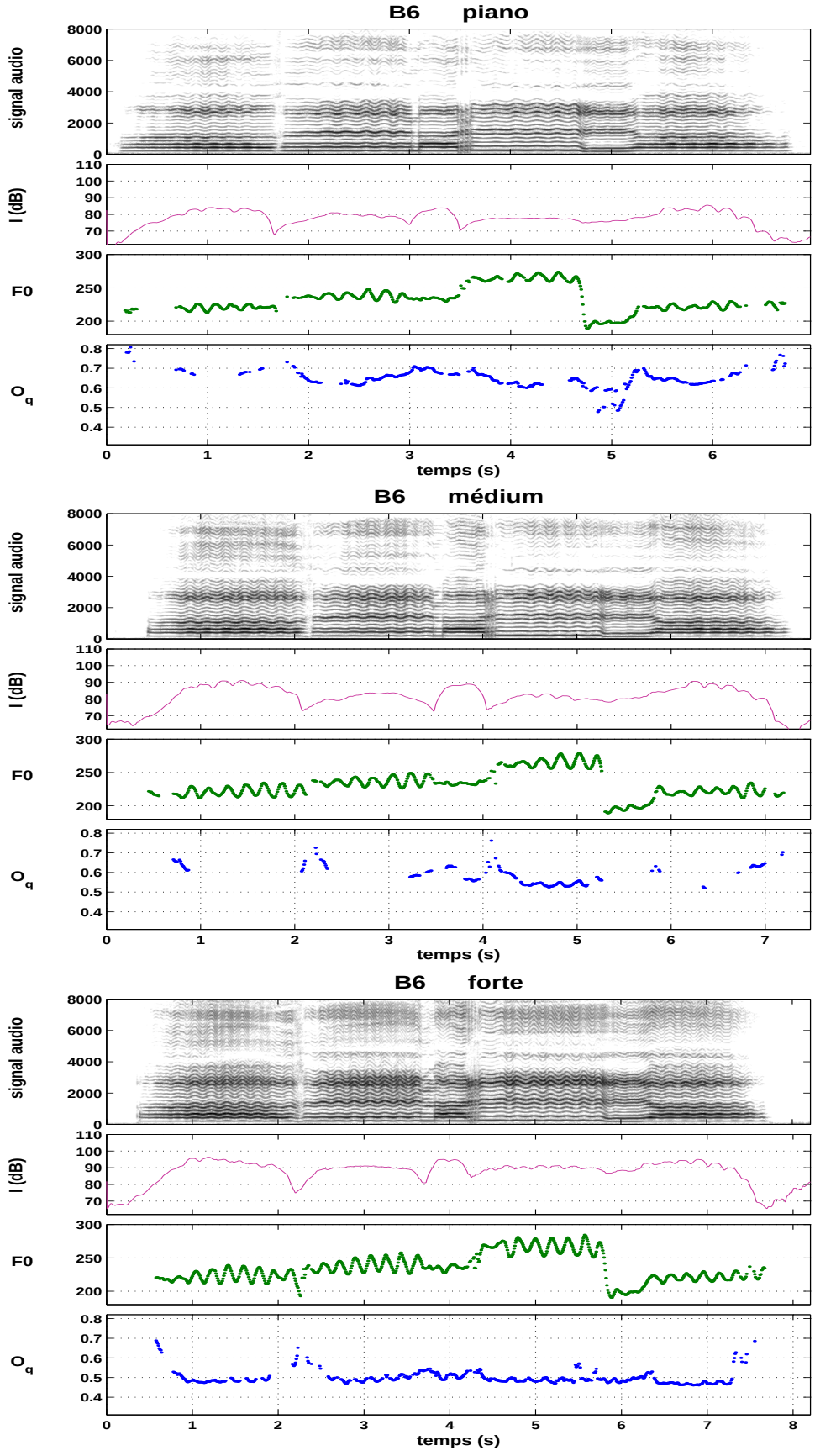


FIG. *G.6 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryton B6 (CD, piste [64])*

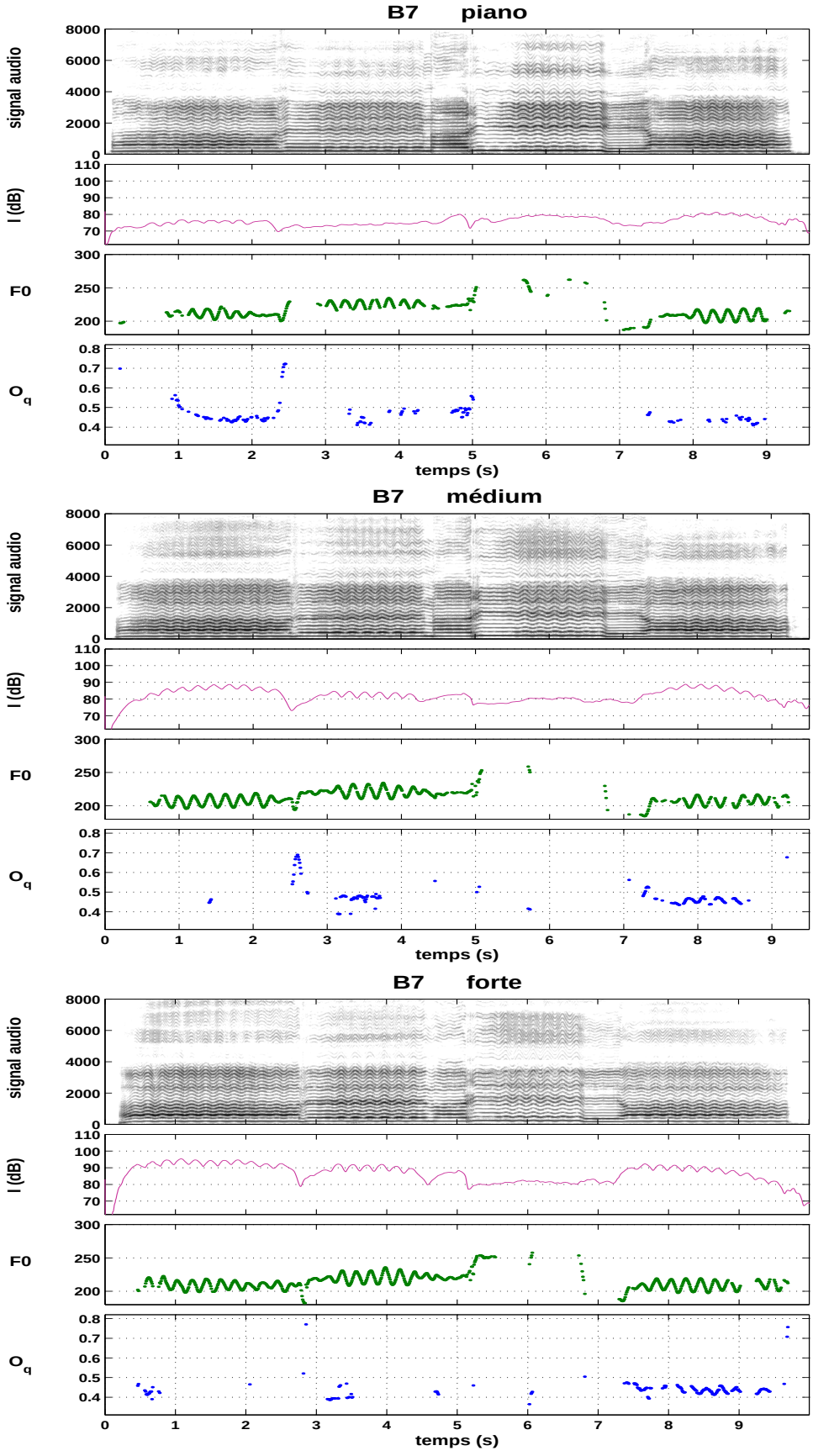


FIG. *G.7 – Premières mesures de l'Ave Maria de Gounod, interprétées par le baryton B7 (CD, piste [65])*

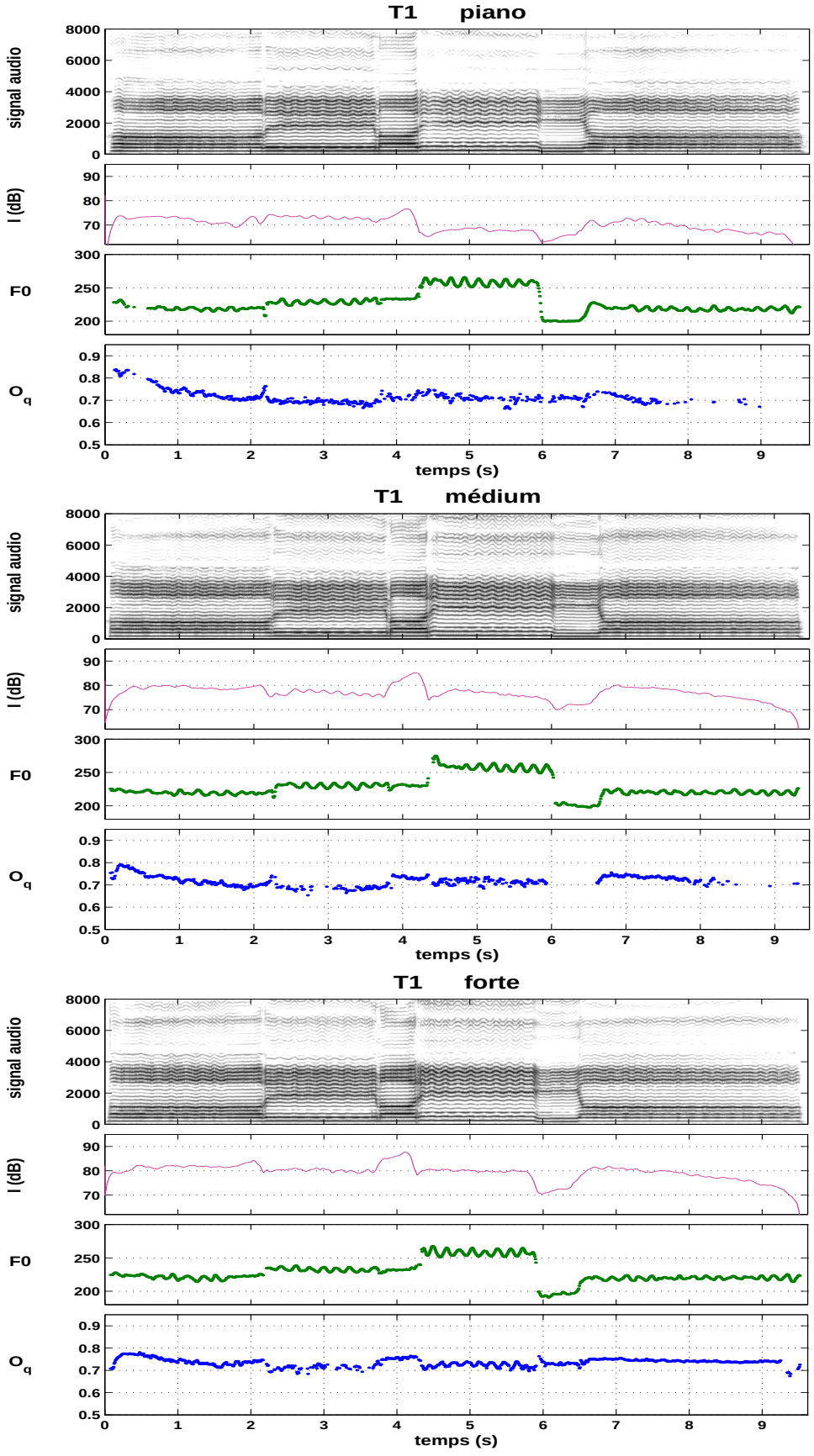


FIG. *G.8 – Premières mesures de l'Ave Maria de Gounod, interprétées par le ténor T1 (CD, piste [66])*

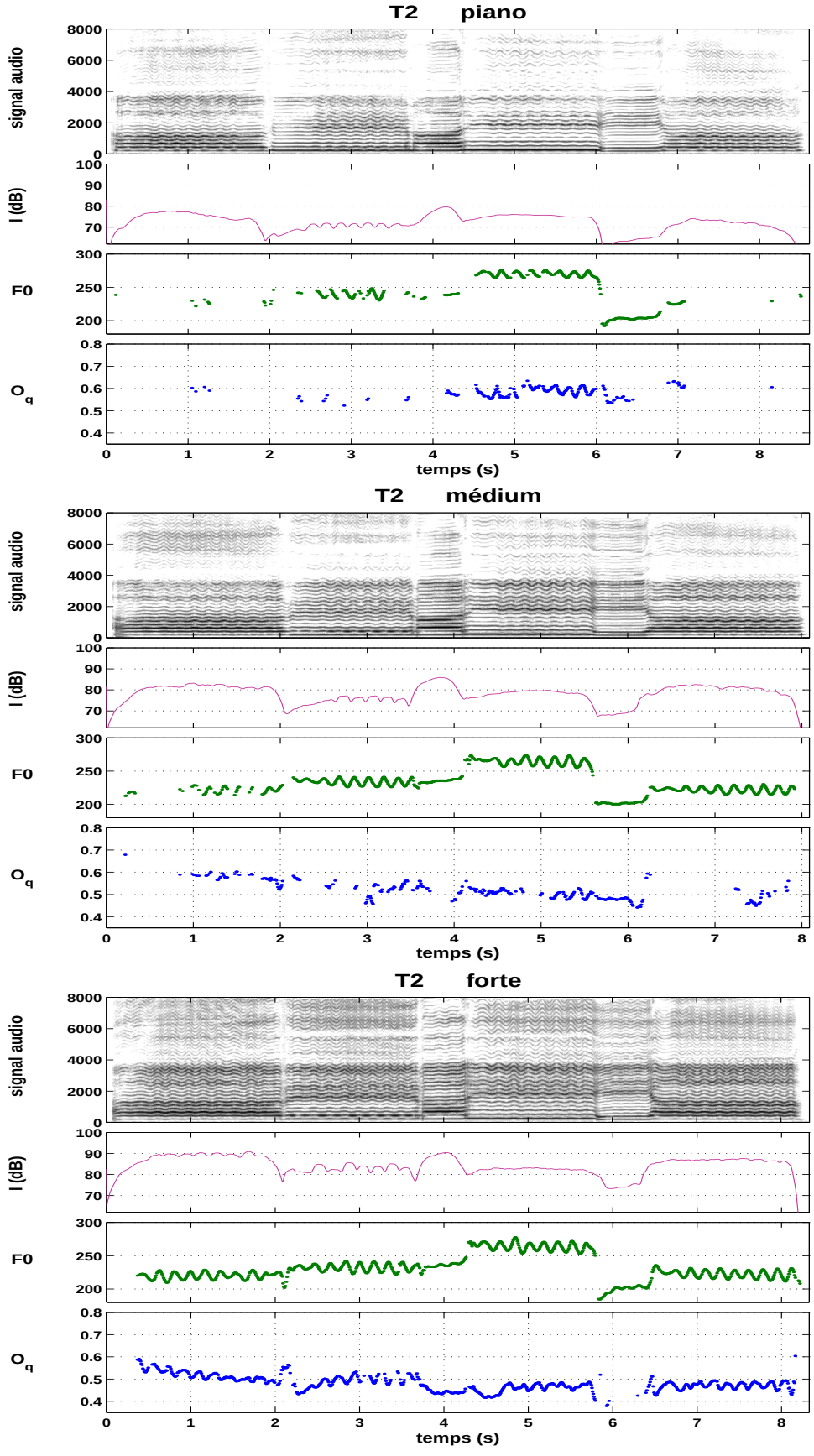


FIG. *G.9 – Premières mesures de l'Ave Maria de Gounod, interprétées par le ténor T2 (CD, piste [67])*

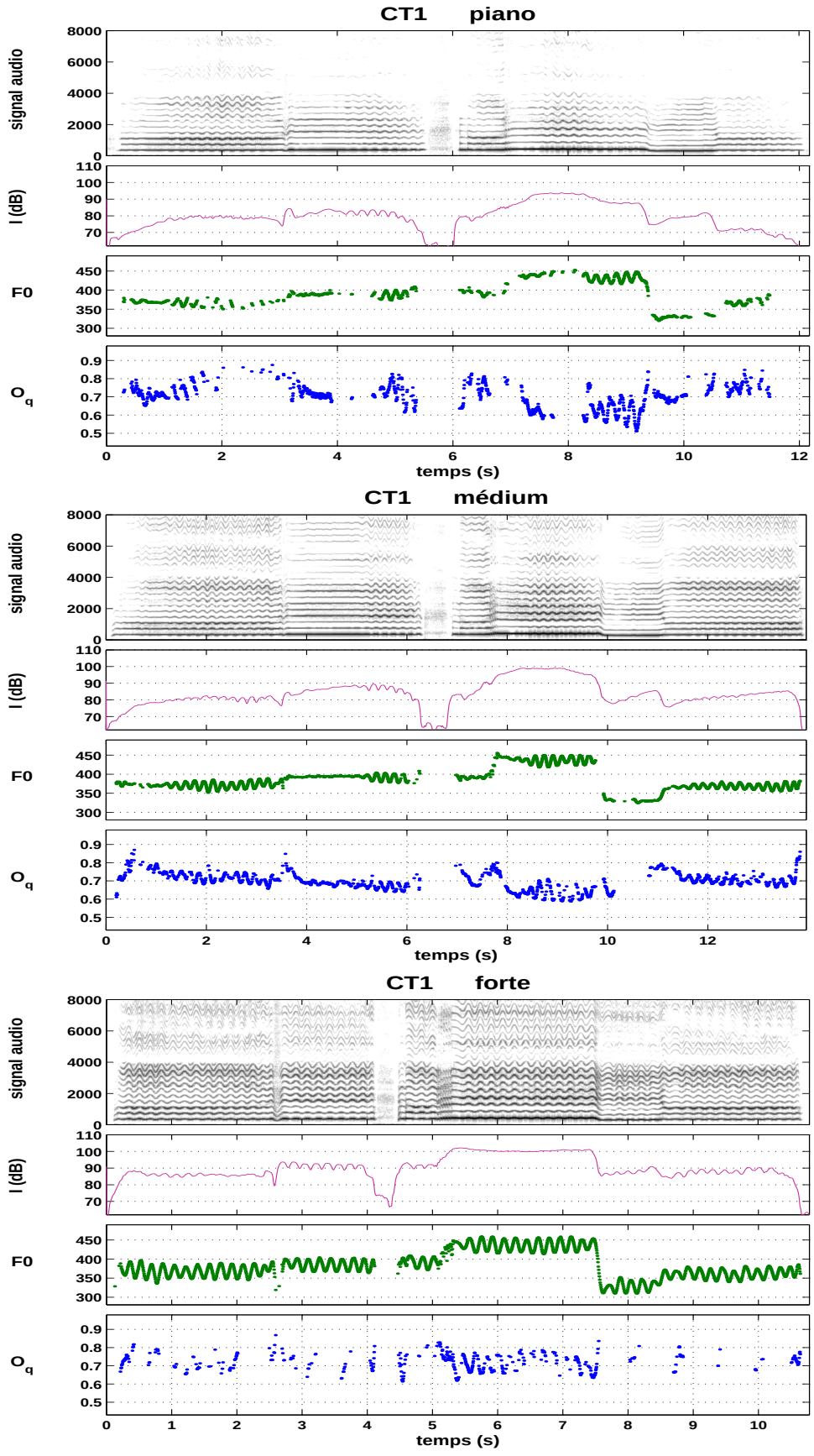


FIG. *G.10 – Premières mesures de l'Ave Maria de Gounod, interprétées par le contre-ténor CT1 (CD, piste [68])*

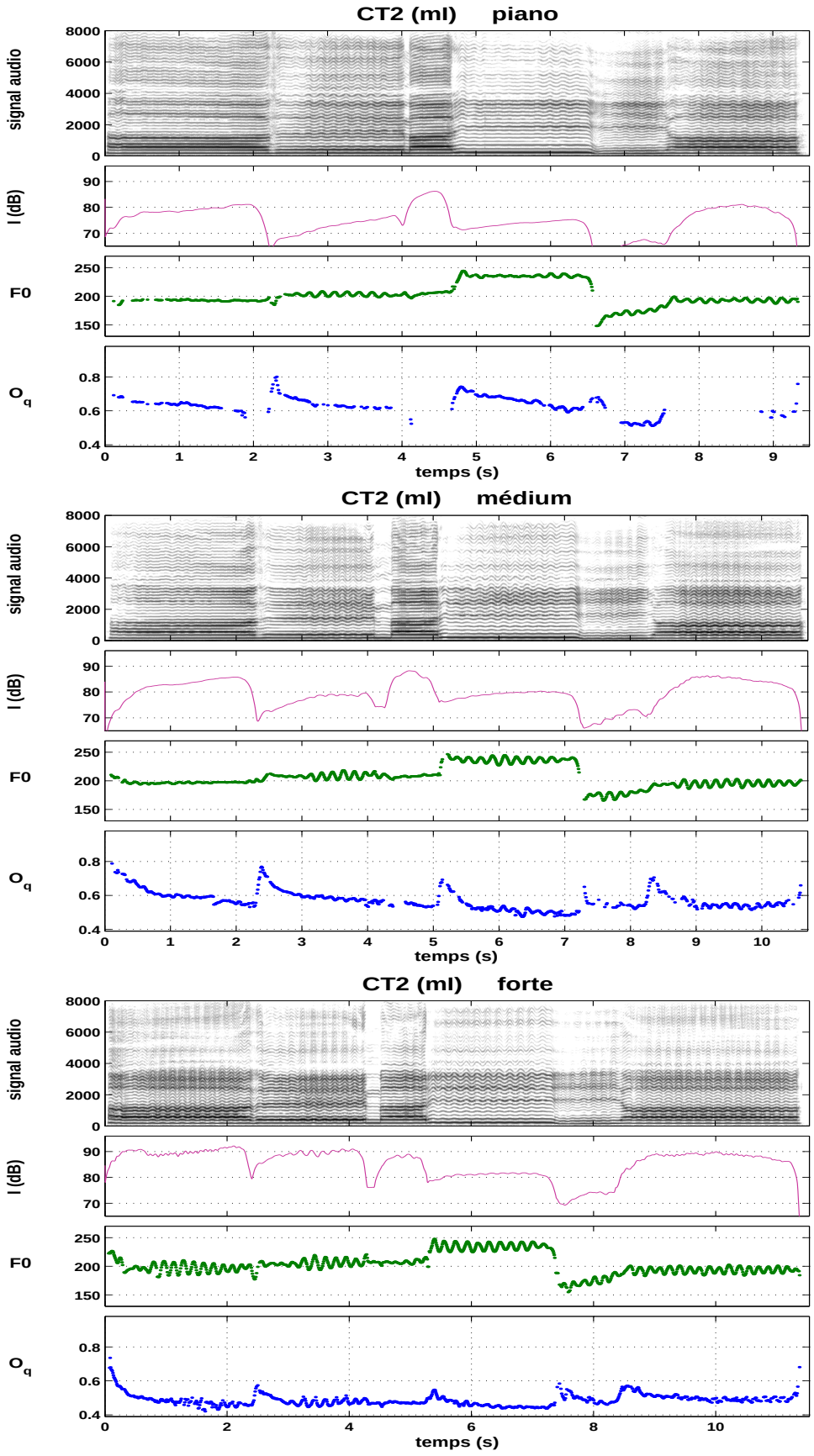


FIG. *G.11 – Premières mesures de l'Ave Maria de Gounod, interprétées par le contre-ténor CT2 en mécanisme I (CD, piste [69])*

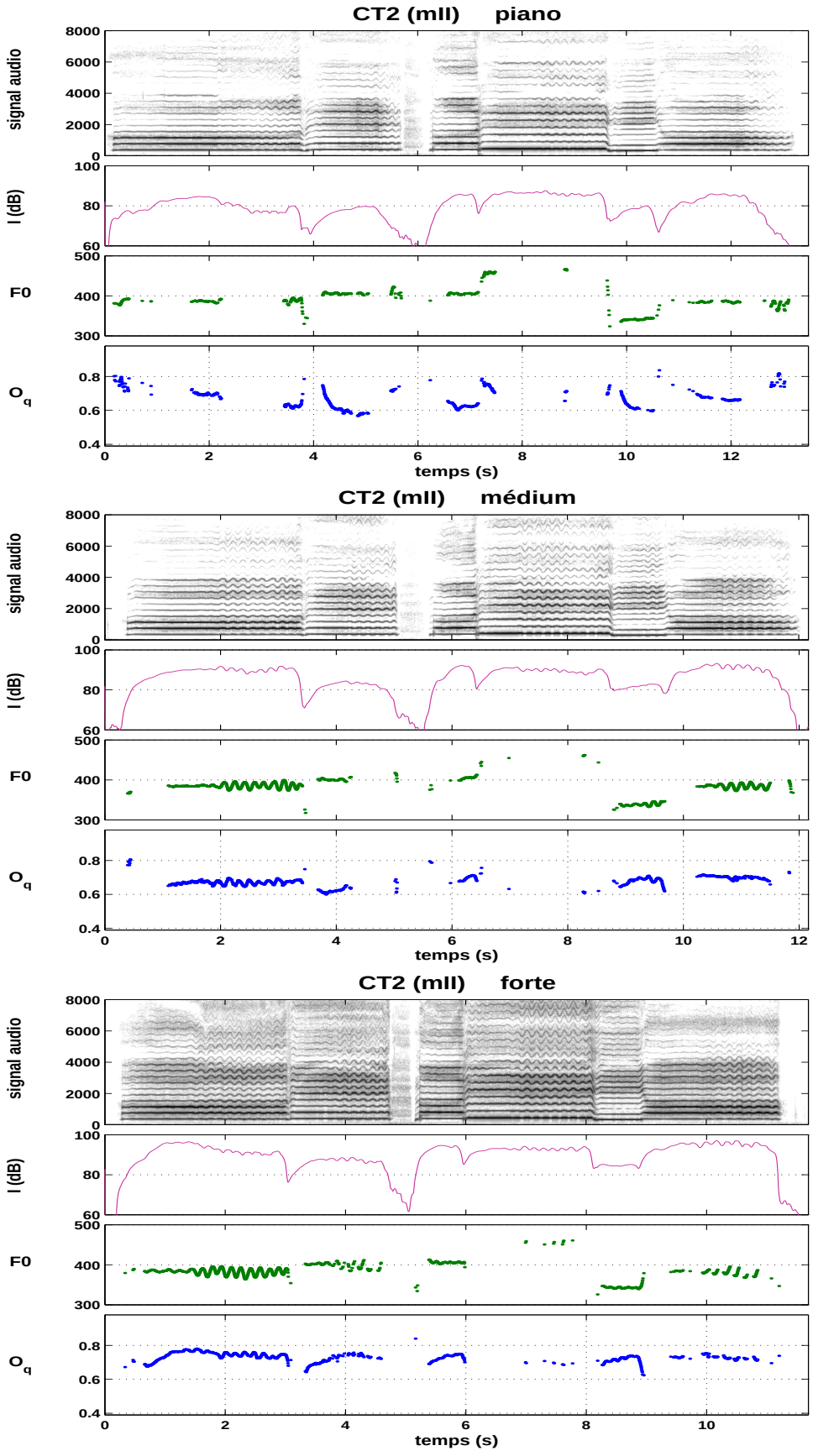


FIG. *G.12 – Premières mesures de l'Ave Maria de Gounod, interprétées par le contre-ténor CT2 en mécanisme II (CD, piste [70])*

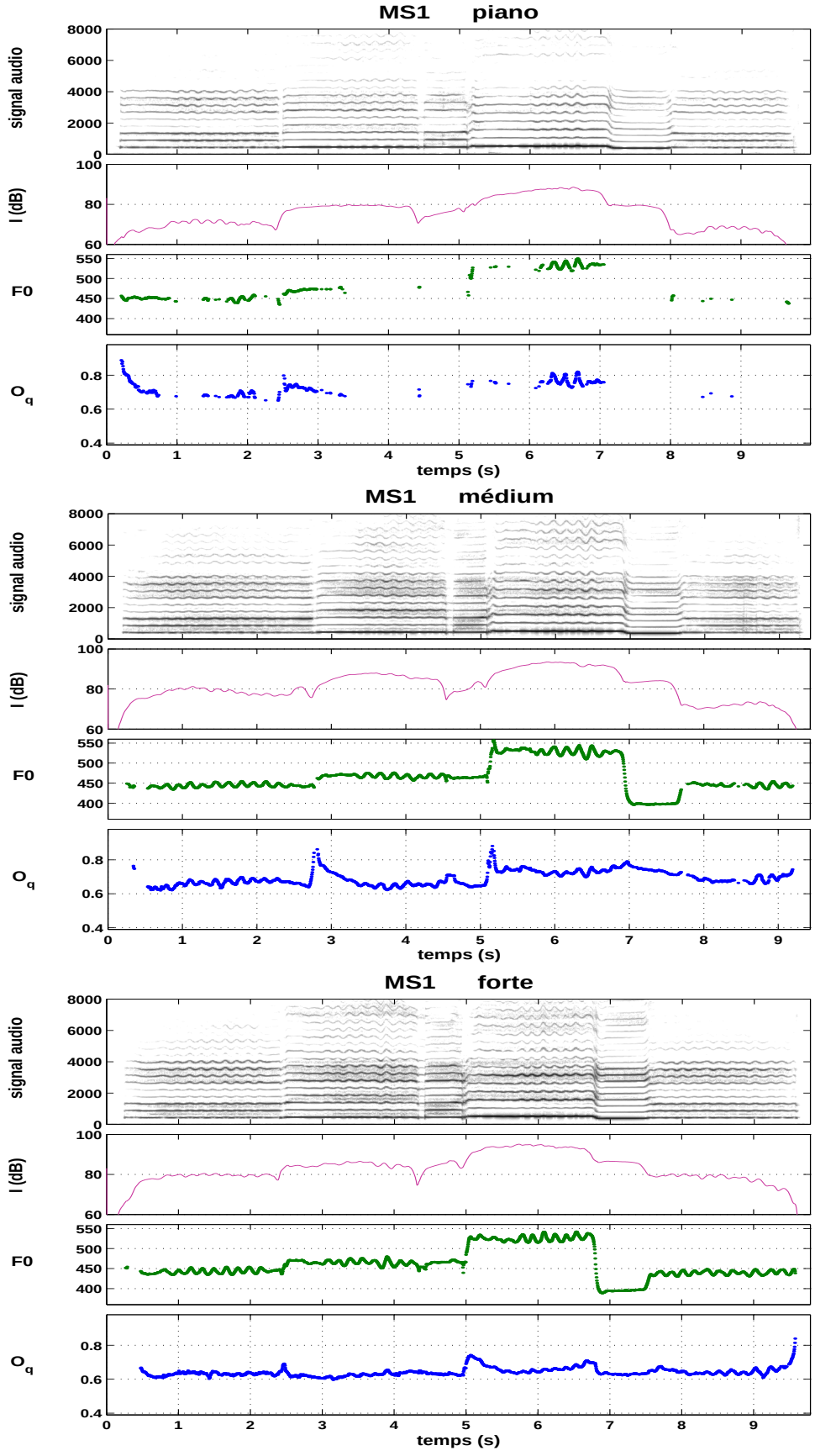


FIG. *G.13 – Premières mesures de l'Ave Maria de Gounod, interprétées par la mezzo-soprano MS1 (CD, piste [71])*

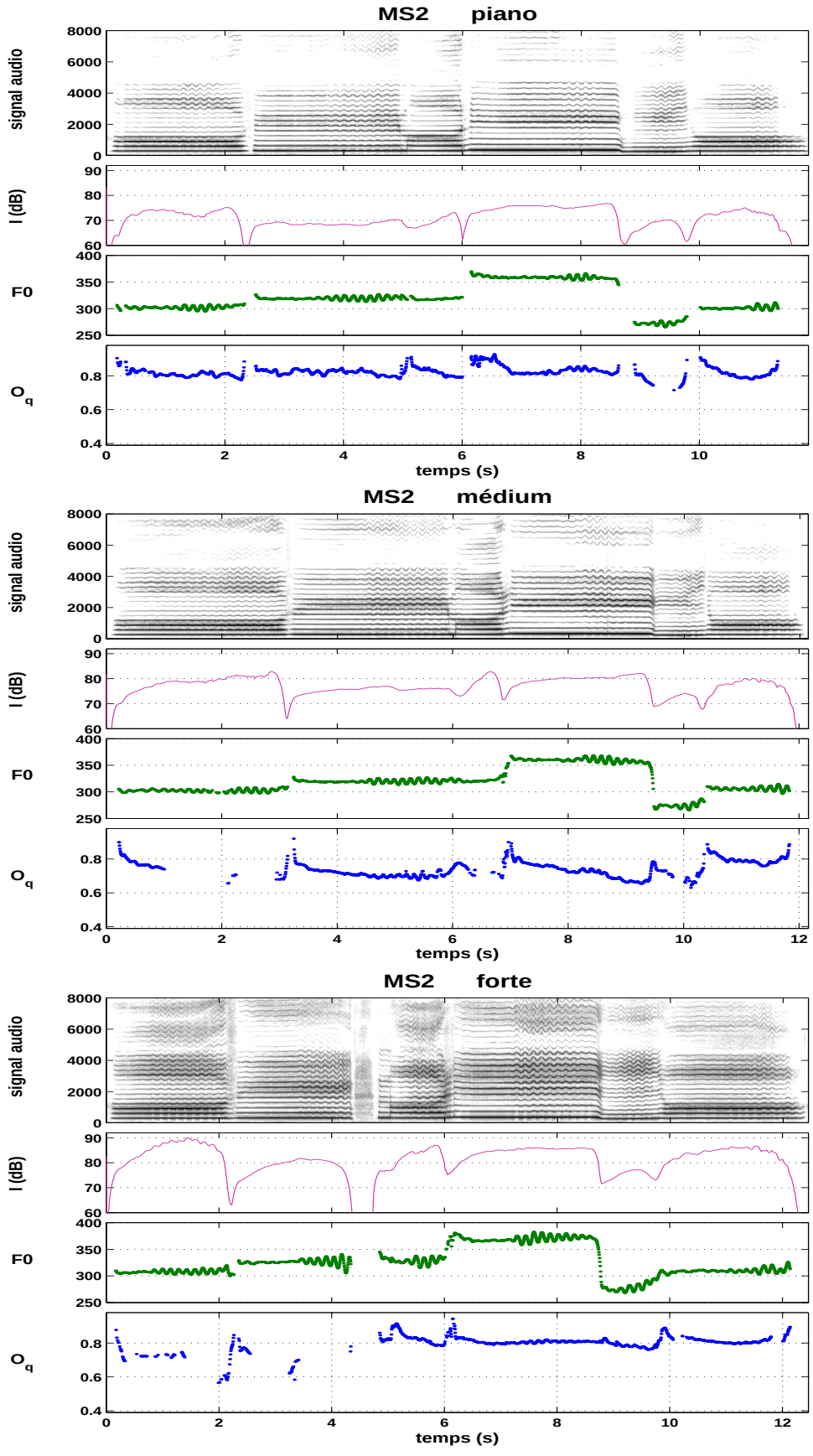


FIG. *G.14 – Premières mesures de l'Ave Maria de Gounod, interprétées par la mezzo-soprano MS2 (CD, piste [72])*

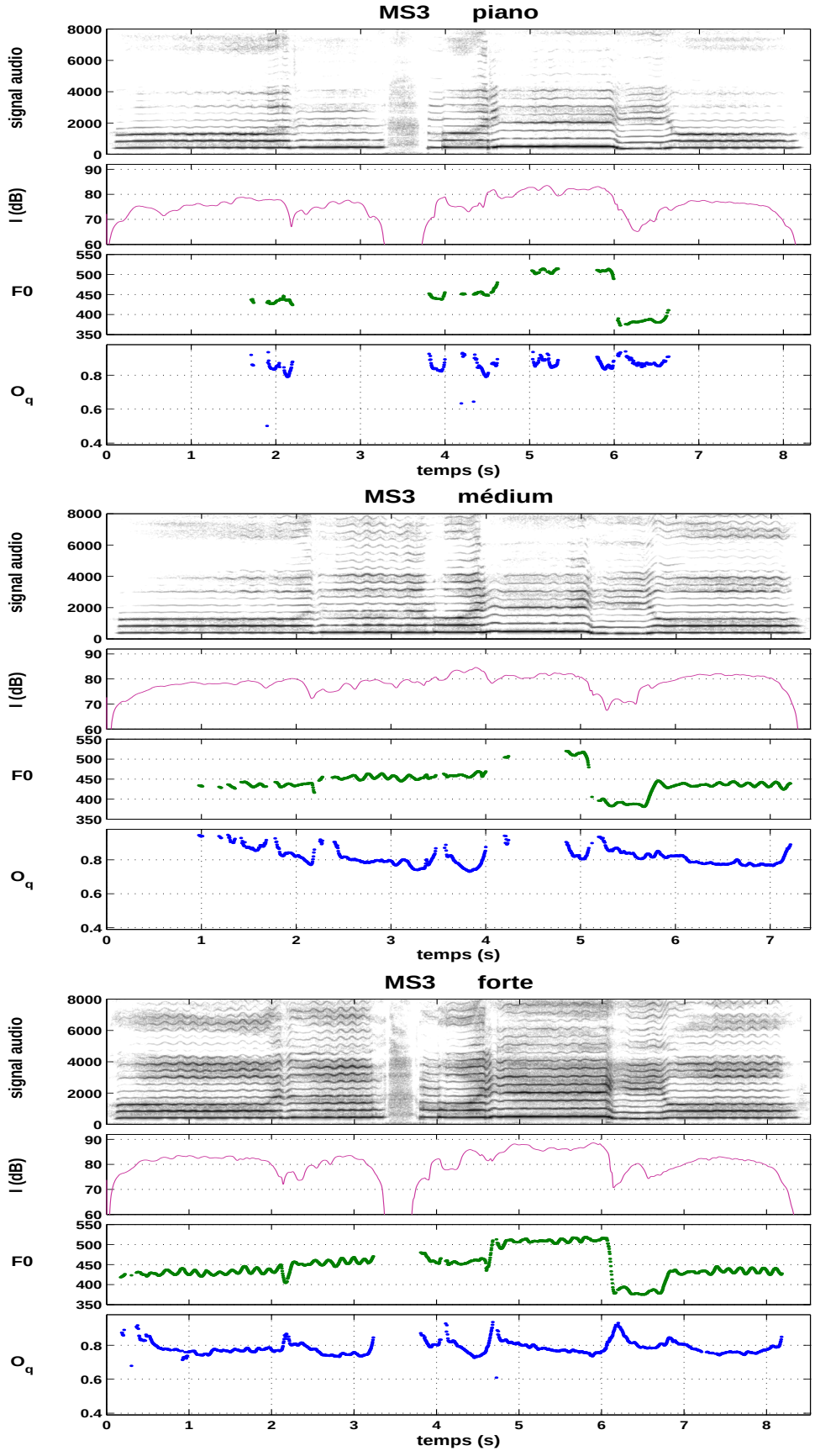


FIG. *G.15 – Premières mesures de l'Ave Maria de Gounod, interprétées par la mezzo-soprano MS3 (CD, piste [73])*

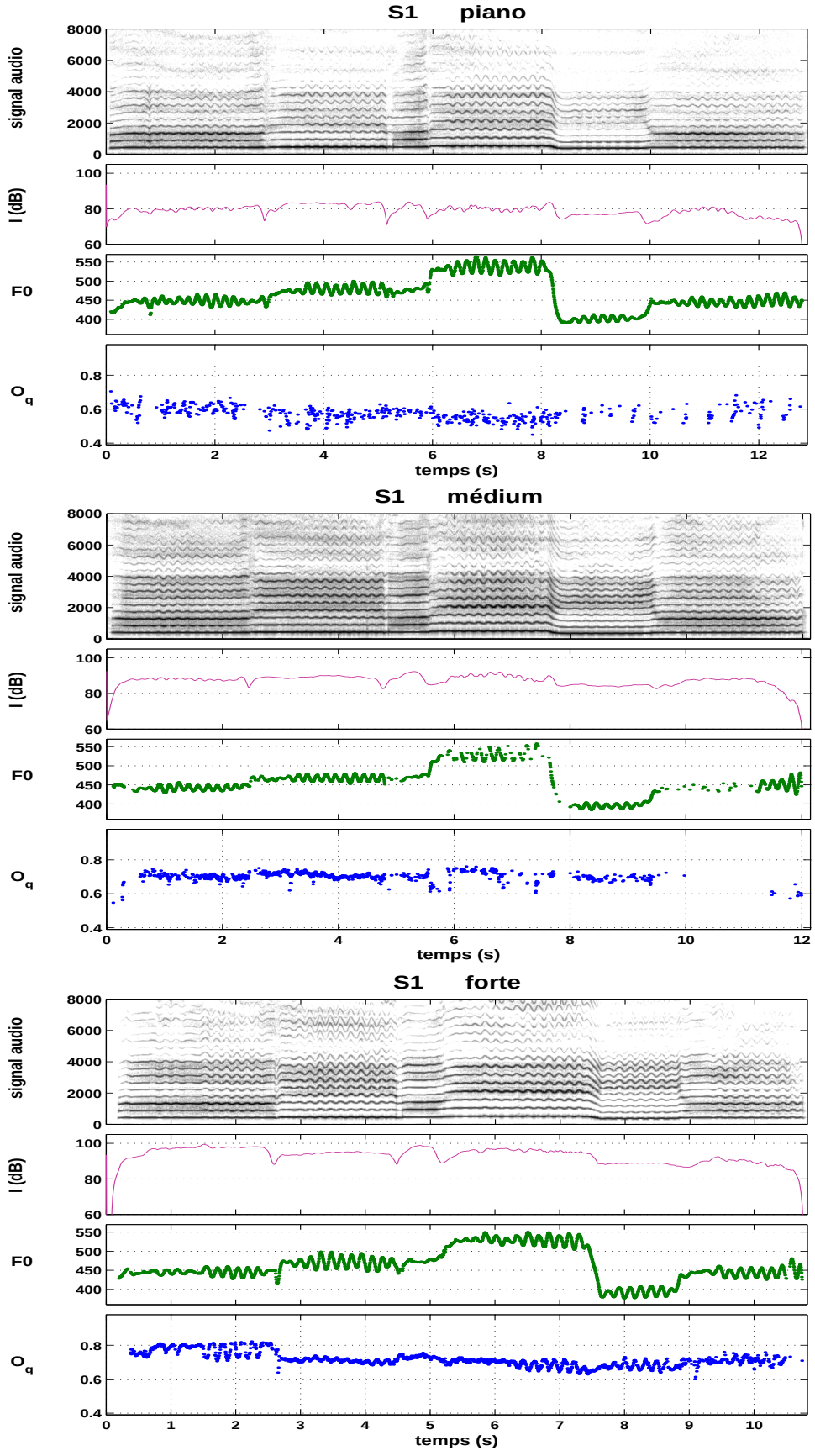


FIG. *G.16 – Premières mesures de l'Ave Maria de Gounod, interprétées par la soprano S1 (CD, piste [74])*

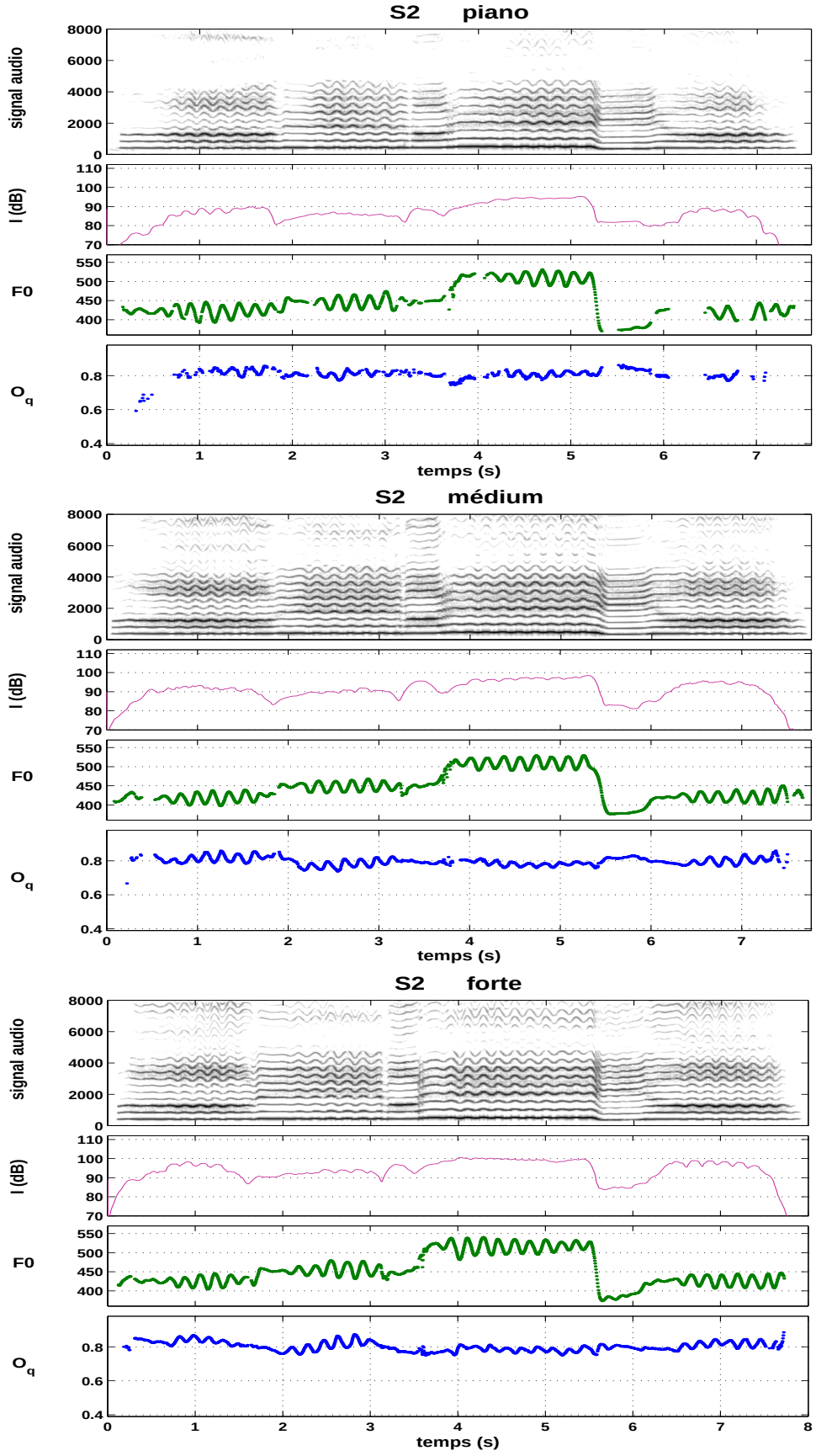


FIG. *G.17 – Premières mesures de l'Ave Maria de Gounod, interprétées par la soprano S2 (CD, piste [75])*

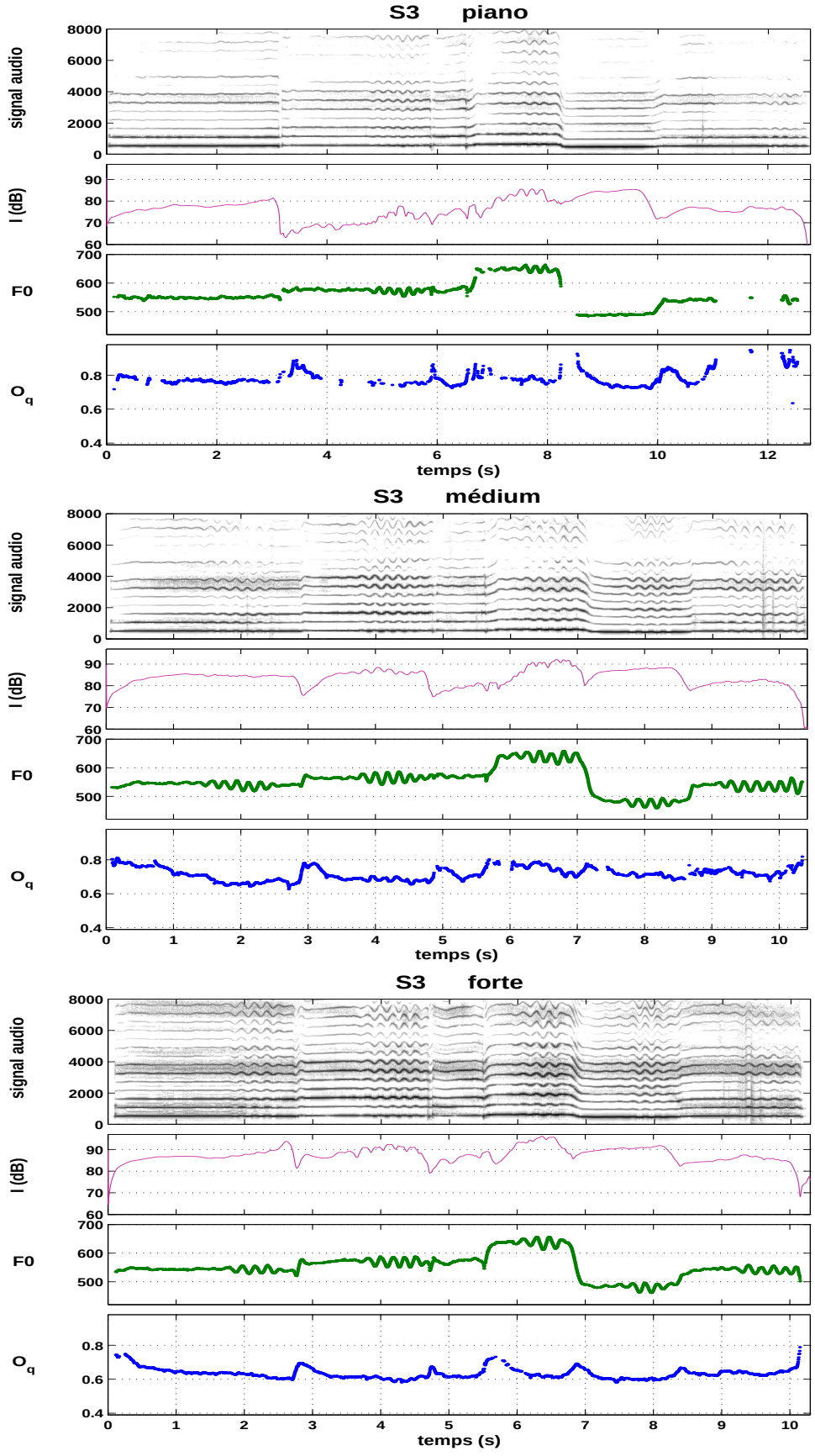


FIG. *G.18 – Premières mesures de l'Ave Maria de Gounod, interprétées par la soprano S3 (CD, piste [76])*

Bibliographie

[1] Alku P. (1992). Glottal wave analysis with pitch synchronous iterative adaptative inverse filtering. *Speech Communication 11*, 109–18.

[2] Alku P. (1992). Inverse filtering of the glottal waveform using the itakura-saito distorsion measure. In *Proc. Int. Conf. on Spoken Language Processing* (Banff, Canada, Oct. 1992), pp. 847–50.

[3] Alku P., Strik H., and Vilkman E. (1997). Parabolic spectral parameter - a new method for quantification of the glottal flow. *Speech Communication 22*, 67–79.

[4] Alku P., and Vilkman E. (1994). Estimation of the glottal pulseform based on discrete all-pole modeling. In *Proc. Int. Conf. on Spoken Language Processing* (Yokohama, Japan, Sept. 1994), pp. 1619–22.

[5] Alku P., and Vilkman E. (1996). Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering. *Speech Communication 18*, 131–8.

[6] Alku P., and Vilkman E. (1996). A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. *Folia Phoniatr. 48*, 240–54.

[7] Alku P., Vintturi J., and Vilkman E. (1999). On the linearity of the relationship between the sound pressure level and the negative peak amplitude of the differentiated glottal flow in vowel production. *Speech Communication 28*, 269–81.

[8] Allen D. R., and Strong W. J. (1985). A model for the synthesis of natural sounding vowels. *J. Acous. Soc. Am. 78*, 1, 58–69.

[9] Ananthapadmanabha T. V. (1984). Acoustic analysis of voice source dynamics. *STL-QPSR 2-3*, 1–24.

[10] Ananthapadmanabha T. V., and Fant G. (1982). Calculation of true glottal flow and its components. *Speech Communication 1*, 167–84.

[11] Anastaplo S., and Karnell M. P. (1988). Synchronized videostroboscopic and electroglottographic examination of glottal opening. *J. Acous. Soc. Am. 83*, 5, 1883–90.

[12] Askenfelt A., Gauffin J., Sundberg J., and Kitzing P. (1980). A comparison of microphone and electroglottograph for the measurement of vocal fundamental frequency. *J. Speech Hear. Res. 23*, 250–73.

[13] Atal B. S., and Hanauer S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *J. Acous. Soc. Am. 50*, 2, 637–55.

[14] Baer T., Löfqvist A., and McGarr N. S. (1983). Laryngeal vibrations : A comparison between high-speed filming and glottographic techniques. *J. Acous. Soc. Am. 73*, 1304–08.

[15] Baer T., Titze I. R., and Yoshioka H. (1983). Multiple simultaneous measures of vocal fold activity. In *Vocal fold physiology : Contemporary Research and Clinical Issues*, Bless and A. ed., Eds. College Hill Press, San Diego, pp. 229–37.

[16] Baken R. J. (1992). Electroglottography. *J. Voice 6*, 2, 98–110.

[17] Baken R. J., and Orlikoff R. F. (2000). *Clinical measurement of speech and voice*, 2^{nde} ed. Singular Publishing Group, San Diego, California.

[18] Behnke E. (1880). *The mechanism of the human voice.* J. Curwen & Sons Ltd., 12ième ed., London.

[19] Bell C. G., Fujisaki H., Heinz J. M., Stevens K. N., and House A. S. (1961). Reduction of speech spectra by analysis-by-synthesis techniques. *J. Acous. Soc. Am. 33*, 12.

[20] Bell Telephone Laboratories (1957). High speed motion pictures of the vocal cords. Tech. rep., Bell Telephone Laboratories Bureau of Publications, New York.

[21] Berke G. S., Moore D. M., Hantke D. R., Hanson D. G., Gerratt B. R., and Burstein F. (1987). Laryngeal modeling : theoretical, in vitro, in vivo. *Laryngoscope 97*, 871–81.

[22] Bonnet C. (1986). *Manuel pratique de psychophysique.* Armand Colin, Paris.

[23] Botte M. C. (1989). Perception de l'intensité sonore. In *Psychoacoustique et perception auditive*, S. A. INSERM, Ed. M. C. Botte, G. Canévet, L. Demany and C. Sorin, Paris, pp. 1–41.

[24] Bucella F., Hassid S., Beeckmans R., Soquet A., and Demolin D. (2000). Pression sous-glottique et débit d'air buccal des voyelles en français. In *XXIIIèmes Journées d'Etude sur la Parole* (Aussois, 2000), pp. 449–52.

[25] Castellengo M. (1986). Les deux principaux mécanismes de production de la voix humaine : leur étendue et leur utilisation musicale. In *Actes du colloque « La voix dans tous ses éclats »* (Paris, Avril 1986), pp. 1–4.

[26] Castellengo M. (1991). Continuité, rupture, ornementation, ou les bons usages de la transition entre deux modes d'émission vocale. In *La voix et les techniques vocales, Cahiers de musiques traditionnelles n°4*, A. d'ethnomusicologie, Ed. Genève.

[27] Castellengo M., and Roubeau B. (2001). *La notion de registre vocal.* à paraître.

[28] Castellengo M., Roubeau B., and Valette C. (1983). Study of the acoustical phenomena characteristic of the transition between chest voice and falsetto. In *Proc. SMAC 83, vol. 1* (Stockholm, Sweden, July 1983), pp. 113–23.

[29] Childers D. G. (2000). *Speech processing and synthesis toolboxes.* John Wiley & Sons, New York.

[30] Childers D. G., Hicks D. M., Moore G. P., and Alsaka Y. A. (1986). A model for vocal fold vibratory motion, contact area, and the electroglottogram. *J. Acous. Soc. Am. 80*, 5, 1309–1320.

[31] Childers D. G., Hicks D. M., Moore G. P., Eskenazi L., and Lalwani A. L. (1990). Electroglottography and vocal fold physiology. *J. Speech Hear. Res. 33*, 245–254.

[32] Childers D. G., and Hu T. H. (1994). Speech synthesis by glottal excited linear prediction. *J. Acous. Soc. Am. 96*, 4, 2026–36.

[33] Childers D. G., and Krishnamurthy A. K. (1985). A critical review of electroglottography. *CRC Critical Reviews in Biomedical Engineering 12*, 131–161.

[34] Childers D. G., and Larar J. N. (1984). Electroglottography for laryngeal function assessment and speech analysis. *IEEE Trans. Biomedical Engineering BME-31*, 12, 807–17.

[35] Childers D. G., and Lee C. K. (1991). Vocal quality factors : Analysis, synthesis, and perception. *J. Acous. Soc. Am. 90*, 2394–2410.

[36] Childers D. G., Moore G. P., Naik J. M., Larar J. N., and Krishnamurthy A. K. (1983). Assessment of laryngeal function by simultaneous, synchronized measurement of speech, electroglottography and ultra-high speed film. In *Transcripts of the eleventh symposium : Care of the professional voice.*, L. V., Ed. New York : Voice Foundation, pp. 234–44.

[37] Childers D. G., Naik J. M., Larar J. N., Krishnamurthy A. K., and Moore G. P. (1983). Electroglottography, speech and ultra-high speed cinematography. In *Vocal fold physiology and biophysics of voice*, I. Titze and R. Scherer, Eds. Denver, CO : Denver Center for the Performing Arts, pp. 202–20.

[38] Childers D. G., Smith A. M., and Moore G. P. (1984). Relationships between electroglottograph, speech, and vocal cord contact. *Folia Phoniatr. 36*, 105–18.

[39] Chuberre B. (2000). *Les registres et passages dans la voix chantée.* Mémoire de phoniatrie, Université de Nantes.

[40] Colton R. H., and Conture E. G. (1990). Problems and pitfalls of electroglottography. *J. Voice 4*, 10–24.

[41] Cookman S., and Verdolini K. (1999). Interrelation of mandibular laryngeal functions. *J. Voice 13*, 1, 11–24.

[42] Cranen B. (1991). Simultaneous modelling of EGG, PGG, and glottal flow. In *Vocal Fold Physiology : Acoustic, perceptual, and physiological aspects of voice mechanisms*, H. B. Gauffin J., Ed. Singular Publishing Group, San Diego, pp. 57–64.

[43] Dagnelie P. (1975). *Théorie et méthodes statistiques*, vol. 2. Les Presses Agronomiques de Gembloux, Belgique.

[44] d'Alessandro C., Darsinos V., and Yegnanarayana B. (1995). Evaluation of periodic/aperiodic decomposition for analysis of aperiodicities in the voice source. In *Intern. Symposium on Musical Acoustics* (Dourdan, France, Juil. 1995), pp. 446–52.

[45] d'Alessandro C., Darsinos V., and Yegnanarayana B. (1998). Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources. *IEEE Trans. Speech and Audio Processing 6*, 1, 12–23.

[46] d'Alessandro C., and Doval B. (1998). Voice quality modification using periodic-aperiodic decomposition and spectral processing of the voice source signal. In *Proc. 3rd Intern. Workshop on Speech Synthesis* (Jenolan Caves, Australia, Nov. 1998), pp. 277–82.

[47] d'Alessandro C., Yegnanarayana B., and Darsinos V. (1995). Decomposition of speech signals into deterministic and stochastic components. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Detroit, USA, Apr. 1995), pp. 760–63.

[48] Darsinos V., d'Alessandro C., and Yegnanarayana B. (1995). Evaluation of a periodic/aperiodic speech decomposition algorithm. In *Proc. European Conf. on Speech Communication and Technology* (Madrid, Spain, Sept. 1995), pp. 393–96.

[49] Daudin J. J., Robin S., and Vuillet C. (1999). *Statistique inférentielle. Idées, démarches, exemples.* Pratique de la statistique - Société Française de Statistique et Presses Universitaires de Rennes.

[50] de Cheveigné A., and Kawahara H. (2001). Comparative evaluation of f_0 estimation algorithms. In *Proc. Eurospeech 2001* (Aalborg, Denmark, Sept. 2001).

[51] de Cheveigné A., and Kawahara H. (2001). YIN, a fundamental frequency estimator for speech and music. *submitted to J. Acous. Soc. Am.*.

[52] Dejonckere P. (1981). Comparison of two methods of photoglottography in relation to electroglottoraphy. *Folia Phoniatr. 33*, 338.

[53] Demolin D., Hassid S., Ladefoged P., and Soquet A. (2000). Vowel and consonant effects on subglottal pressure. *à paraître*.

[54] Doval B., and d'Alessandro C. (1997). Spectral correlates of glottal waveform models : an analytic study. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Munich, Germany, Apr. 1997), pp. 446–52.

[55] Doval B., and d'Alessandro C. (1999). The spectrum of glottal flow models. *notes et documents LIMSI 99 - 07*.

[56] Doval B., d'Alessandro C., and Diard B. (1997). Spectral methods for voice source parameter estimation. In *Proc. Europseech'97* (Rhodes, Greece, Sept. 1997), pp. 533–36.

[57] Dromey C., Stathopoulos E. T., and Sapienza C. M. (1992). Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *J. Voice 6*, 1, 44–54.

[58] Dutoit T. (1997). *An introduction to text-to-speech synthesis.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

[59] El-Jaroudi A., and Makhoul J. (1991). Discrete all-pole modeling. *IEEE Trans. on Signal Processing 39*, 2, 411–423.

[60] Fabre P. (1957). Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation : glottographie de haute fréquence. *Bull. Nat. Med.*, 66–69.

[61] Fabre P. (1958). Etude comparée des glottogrammes et des phonogrammes de la voix humaine. *Ann. Oto. Lar. (Paris) 75*, 767, 767–75.

[62] Fabre P. (1959). La glottographie électrique en haute fréquence : particularités de l'appareillage. *C. R. Soc. Biol. (Paris) 153*, 1361–64.

[63] Fabre P. (1961). Glottographie respiratoire. *Ann. Oto. Lar. (Paris) 78*, 12, 814–824.

[64] Fant G. (1960). *Acoustic theory of speech production*. Mouton, La Hague.

[65] Fant G. (1979). Glottal source and excitation analysis. *STL-QPSR 1*, 85–107.

[66] Fant G. (1982). Preliminaries to analysis of the human voice source. *STL-QPSR 4*, 1–27.

[67] Fant G. (1986). Glottal flow : models and interaction. *Journal of Phonetics 14*, 393–9.

[68] Fant G. (1993). Some problems in voice source analysis. *Speech Communication 13*, 7–22.

[69] Fant G. (1995). The lf-model revisited. transformations and frequency domain analysis. *STL-QPSR 2-3*, 119–56.

[70] Fant G. (1997). The voice source in connected speech. *Speech Communication 22*, 125–39.

[71] Fant G. (2000). The source-filter frame of prominence. *Phonetica 57*, 113–27.

[72] Fant G., Liljencrants J., and Lin Q. (1985). A four-parameter model of glottal flow. *STL-QPSR 4*, 1–13.

[73] Flanagan J. L. (1965). *Speech analysis, synthesis and perception*. Springer Verlag, Berlin.

[74] Flanagan J. L., Ishizaka K., and Shipley K. L. (1975). Synthesis of speech from a dynamic model of the vocal cords and vocal tract. *Bell Syst. Tech. Journal 54*, 3, 485–506.

[75] Fourcin A. J. (1974). Laryngographic examination of vocal fold vibration. In *Ventilatory and phonatory function*, B. Wyke, Ed. Oxford University Press, London, pp. 315–26.

[76] Fourcin A. J. (1981). Laryngographic assessment of phonatory function. *ASHA Rep. 11*, 116–124.

[77] Fourcin A. J., and Abberton E. (1971). First applications of a new laryngograph. *Medical and Biological Illustration 21*, 172–182.

[78] Fröhlich M., Michaelis D., and Strube H. W. (2001). SIM-simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. *J. Acous. Soc. Am. 110*, 1, 479–88.

[79] Frokjaer-Jensen B. (1969). Construction and comparative tests of two different types of glottographs : (1) the photo-electric glottograph and (2) the fabre glottograph. *17th Natl. Congr. Otolaryngol., Denmark 1*.

[80] Frokjaer-Jensen B. *The electroglottograph, Type EG 830*, 1973. Manual.

[81] Frokjaer-Jensen B., and Thorvaldsen P. (1968). Construction of a fabre glottograph. *ARIPUC 3*, 1.

[82] Fujimura O., and Lindqvist J. (1971). Sweep-tone measurements of vocal-tract characteristics. *J. Acous. Soc. Am. 49*, 541–58.

[83] Fujisaki H., and Ljungqvist M. (1987). Estimation of voice source and vocal tract parameters based on arma analysis and a model for the glottal source waveform. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (1987), pp. 637–40.

[84] Garcia M. (1884). *Traité complet de l'Art du Chant*. Heugel & Cie, 8ième ed., Paris.

[85] Gauffin J., and Sundberg J. (1989). Spectral correlates of glottal voice source waveform characteristics. *J. Speech Hear. Res. 32*, 556–65.

[86] Gerrat B. R., Hanson D. G., and Berke G. S. (1988). Laryngeal configuration associated with glottography. *Am. J. Otolaryngol 9*, 173–9.

[87] Gilbert H. R., Potter C. R., and Hoodin R. (1984). Laryngograph as a measure of vocal fold contact area. *J. Speech Hear. Res. 27*, 178–82.

[88] Gobl C. (1988). Voice source dynamics in connected speech. *STL-QPSR 1*, 123–59.

[89] Gramming P., Sundberg J., Ternstrom S., Leanderson R., and Perkins W. (1988). Relationship between changes in voice pitch and loudness. *J. Voice 2*, 118–26.

[90] Green D. M., and Swets J. A. (1966). *Signal detection theory and psychophysics*. Huntington, NY : Krieger.

[91] Guerin B., Mryati M., and Carré R. (1976). A voice source taking account of coupling with the supraglottal cavities. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (1976), pp. 47–50.

[92] Guily J. L. S., and Roubeau B. (1994). Voies nerveuses et physiologie de la phonation. In *Encyclopédie Médico-Chirurgicale, Oto-rhino-laryngologie*, E. Techniques, Ed. Paris, France.

[93] Hacki T. (1996). Electroglottographic quasi-open quotient and amplitude in crescendo phonation. *J. Voice 10*, 342–347.

[94] Haji T., Horiguchi S., Baer T., Gould W. J., and Wilbur J. (1986). Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. *J. Acous. Soc. Am. 80*, 58–62.

[95] Hammarberg B., and Gauffin J. (1995). Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In *Vocal Fold Physiology : Voice quality control*, O. Fujimura and M. Hirano, Eds. Singular Publishing Group, San Diego, California, pp. 283–303.

[96] Hanson D. G., Gerratt B. R., and Berke G. S. (1990). Frequency, intensity and target matching effects on photoglottographic measures of open quotient and speed quotient. *J. Speech Hear. Res. 33*, 45–50.

[97] Hanson H. M. (1995). *Glottal characteristics of female speakers*. Ph.d. thesis, Harvard University.

[98] Hanson H. M. (1997). Glottal characteristics of female speakers : Acoustic correlates. *J. Acous. Soc. Am. 101*, 466–481.

[99] Hanson H. M., and Chuang E. S. (1999). Glottal characteristics of male speakers : Acoustic correlates and comparison with female data. *J. Acous. Soc. Am. 106*, 2, 1064–1077.

[100] Hedelin P. (1984). A glottal lpc-vocoder. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (1984), pp. 161–64.

[101] Henrich N., d'Alessandro C., Castellengo M., and Doval B. (2000). Mesures électroglottographiques de quotient d'ouverture glottique en voix parlée et chantée. In *XXIIIèmes Journées d'Etude sur la Parole* (Aussois, 2000), pp. 133–36.

[102] Henrich N., d'Alessandro C., and Doval B. (2001). Spectral correlates of voice open quotient and glottal flow asymmetry : theory, limits and experimental data. In *Eurospeech 2001* (Aalborg, Denmark, 2001).

[103] Henrich N., Doval B., and d'Alessandro C. (1999). Glottal open quotient estimation using linear prediction. In *Proc. Intern. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* (Firenze, Italy, 1999), pp. 12–17.

[104] Henrich N., Doval B., d'Alessandro C., and Castellengo M. (2000). Open quotient measurements on egg, speech and singing signals. In *Proc. 4th Intern. Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research* (Jena, Germany, 2000), pp. 93–101.

[105] Herzel H. (1993). Bifurcations and chaos in voice signals. *Appl. Mech. Rev. 46*, 7, 399–413.

[106] Hess M. H., and Ludwigs M. (2000). Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns. *J. Voice 14*, 2, 255–71.

[107] Hirano M., Vennard W., and Ohala J. (1970). Regulation of register, pitch and intensity of voice. *Folia Phoniatr. 22*, 1–20.

[108] Hollien H., and Michel J. F. (1968). Vocal fry as a phonational register. *J. Speech Hear. Res. 11*, 600–4.

[109] Holmberg E. B., Hillman R. E., and Perkell J. S. (1988). Glottal air flow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. *J. Acous. Soc. Am. 84*, 2, 511–29.

[110] Holmberg E. B., Hillman R. E., and Perkell J. S. (1989). Glottal air flow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. *J. Voice 3*, 4, 294–305.

[111] Holmberg E. B., Hillman R. E., Perkell J. S., Guiod P. C., and Goldman S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *J. Speech Hear. Res. 38*, 1212–23.

[112] Howard D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *J. Voice 9*, 2, 163–72.

[113] Howard D. M., Lindsey G. A., and Allen B. (1990). Toward the quantification of vocal efficiency. *J. Voice 4*, 3, 205–12.

[114] Husson R. (1962). *Physiologie de la phonation*. Masson, Paris.

[115] Isaksson A., and Millnert M. (1986). Inverse glottal filtering using a parameterized input model. *Signal Processing 18*, 435–45.

[116] Ishizaka K., and Flanagan J. L. (1972). Synthesis of voiced sounds from a two mass model of the vocal cords. *Bell Syst. Tech. Journal 50*, 6, 1233–68.

[117] Isshiki N. (1964). Regulatory mechanism of voice intensity variation. *J. Speech Hear. Res. 7*, 17–29.

[118] Itakura F., and Saito S. (1970). A statistical method for estimation of speech spectral density and formant frequencies. *Electron. Commun. Japan 53-A*, 36–43.

[119] Jacobs R. (1985). *La controverse sur le timbre du contre-ténor*. Actes Sud.

[120] Johnson D. M., Watson C. S., and Kelly W. J. (1984). Performance differences among the intervals in forced-choice tasks. *Perception & Psychophysics 35*, 6, 553–7.

[121] Jolicoeur P. (1991). *Introduction à la biométrie*. DECARIE / MASSON, Montréal / Paris.

[122] Kalman R. E. (1958). Design of a self-optimizing control system. *Trans. ASME 80*, 468–78.

[123] Karlsson I. (1986). Glottal waveforms for normal female speakers. *STL-QPSR 1*, 31–36.

[124] Karlsson I. (1992). *Analysis and synthesis of different voices with emphasis on female speech*. Ph.d. thesis, Royal Institute of Technology, Stockholm.

[125] Karnell M. P. (1989). Synchronized videostroboscopy and electroglottography. *J. Voice 3*, 1, 68–75.

[126] Kitzing P. (1977). Methode zur kombierten photo- und elektroglottographischen registrierung von stimmlippenschwingungen. *Folia Phoniatr. 29*, 249–60.

[127] Kitzing P. (1982). Photo- and electroglottographical recording of the laryngeal vibratory pattern during different registers. *Folia Phoniatr. 34*, 234–41.

[128] Kitzing P. (1983). Simultaneous photo- and electroglottographic measurements of voice strain. *In : Titze IR, Scherer RC, eds. Vocal Fold physiology. The Denver Center for the Performing Arts*, 221–9.

[129] Kitzing P. (1990). Clinical applications of electroglottography. *J. Voice 4*, 3, 238–49.

[130] Kitzing P., Carlborg B., and Löfqvist A. (1982). Aerodynamic and glottographic studies of the laryngeal vibratory cycle. *Folia Phoniatr. 34*, 216–24.

[131] Kitzing P., and Sonesson B. (1974). A photoglottographical study of the female vocal folds during phonation. *Folia Phoniatr. (Basel) 26*, 138–49.

[132] Klatt D., and Klatt L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acous. Soc. Am. 87*, 2, 820–857.

[133] Klatt D. H. (1980). Software for a cascade/parallel formant synthesizer. *J. Acous. Soc. Am. 67*, 979–95.

[134] Krishnamurthy A. K., and Childers D. G. (1986). Two-channel speech analysis. *IEEE Trans. Acoustics and Speech Signal Processing ASSP-34*, 4, 730–43.

[135] Ladefoged P., and McKinney N. P. (1963). Loudness, sound pressure and subglottal pressure in speech. *J. Acous. Soc. Am. 35*, 4, 454–60.

[136] Le Huche F., and Allali A. (1991). *La voix. Anatomie et physiologie des organes de la voix et de la parole. 2ième ed.* Masson, Paris.

[137] Lecluse F. (1977). *Elektroglottografie.* Thesis rotterdam, Drukkerig Elinkwijk, Utrecht.

[138] Lecluse F., and Brocaar M. (1977). Quantitative measurements in the electroglottogram. *17th International Congress of Logopedics and Phoniatrics.*

[139] Lecluse F. L. E., Brocaar M. P., and Verschuure J. (1975). The electroglottography and its relation to glottal activity. *Folia Phoniatr. 27*, 215–24.

[140] Lecuit V., and Demolin D. (1998). Relation entre pression sous-glottique et intensité : étude des voyelles du français. In *XXIIèmes Journées d'Etude sur la Parole* (Martigny, Suisse, 1998), pp. 299–302.

[141] Lecuit V., and Demolin D. (1998). The relationship between intensity and subglottal pressure with controlled pitch. In *ICSLP* (Sydney, 1998), pp. 3083–86.

[142] Levinson N. (1947). The wiener rms (root mean square) error criterion in filter design and prediction. *Journ. Math. Phys. 25*, 261–78.

[143] Levitt H. (1971). Transformed up-down methods in psychoacoustics. *J. Acous. Soc. Am. 49*, 2, 467–77.

[144] Liénard J. S., and Di Benedetto M. G. (1999). Effect of vocal effort on spectral properties of vowels. *J. Acous. Soc. Am. 106*, 1, 411–22.

[145] Liu J. (1993). *Amélioration de la décomposition source-filtre du signal vocal. Etude de la variabilité des paramètres de l'onde glottique. Application à la transformation de la voix.* thèse, Université Paris XI, Orsay.

[146] Ljungqvist M. G. (1986). *Speech analysis-synthesis based on modeling of voice source and vocal-tract characteristics.* Ph.d. thesis, University of Tokyo.

[147] Lous N. J. C., Hofmans G. C. J., Veldhuis R. N. J., and Hirschberg A. (1998). A symetrical two mass model coupled to vocal tract and trachea, with application to prothesis design. *Acta Acustica 84*, 1135–50.

[148] Lu H., and Smith J. O. (1999). Joint estimation of vocal tract filter and glottal source waveform via convex optimization. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (1999), pp. 79–82.

[149] Makhoul J. (1975). Linear prediction : A tutorial review. *Proc. IEEE 63*, 4, 561–580.

[150] Makhoul J. (1975). Spectral linear prediction : Properties and applications. *IEEE Trans. Acoustics and Speech Signal Processing ASSP-23*, 3, 283–296.

[151] Markel J. D., and Gray A. H. (1976). *Linear prediction of speech.* Springer-Verlag, Berlin.

[152] McAulay R. J. (1984). Maximum likelihood spectral estimation and its application to narrow-band speech coding. *IEEE Trans. Acoustics and Speech Signal Processing ASSP-32*, 2, 243–251.

[153] Milenkovic P. H. (1993). Voice source model for continuous control of pitch period. *J. Acous. Soc. Am. 93*, 2, 1087–96.

[154] Miller D. G. (2000). *Registers in singing.* Ph.d. thesis, Rijksuniversiteit Groningen, the Netherlands.

[155] Miller D. G., Svec J. G., and Schutte H. K. (2000). Measurement of characteristic leap interval between chest and falsetto registers. *Submitted to Journal of Voice.*

[156] Miller R. (1990). *La structure du chant.* éditions ipmc, La Vilette, Paris.

[157] Miller R. L. (1959). Nature of the vocal cord wave. *J. Acous. Soc. Am. 31*, 6, 667–677.

[158] Noll A. M. (1964). Short-time spectrum and "cepstrum" techniques for vocal pitch detection. *J. Acous. Soc. Am. 36*, 296–302.

[159] Oliveira L. C. (1993). Estimation of source parameters by frequency analysis. In *Proc. Eurospeech'93* (1993), pp. 99–102.

[160] Oliveira L. C. (1996). Text-to-speech synthesis with dynamic control of source parameters. In *Progress in Speech Synthesis*, J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, Eds. Springer-Verlag, pp. 27–39.

[161] Orlikoff R. F. (1991). Assessment of the dynamics of vocal fold contact from the electroglottogram : data from normal male subjects. *J. Speech Hear. Res. 34*, 1066–72.

[162] Orlikoff R. F. (1998). The uses and abuses of electroglottography. *Phonoscope 1*, 1, 37–53.

[163] Pabon J. P. H., and Plomb R. (1988). Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *J. Speech Hear. Res. 31*, 710–22.

[164] Papoulis A. (1984). *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York.

[165] Parsons T. W. (1987). *Voice and speech processing.* McGraw-Hill, New York.

[166] Pedersen M. F. (1977). Electroglottography compared with synchronized stroboscopy in normal persons. *Folia Phoniatr. 29*, 191–99.

[167] Rioux V. (2001). *Sound quality of flue organ pipes - An interdisciplinary study of the art of voicing.* Ph.d. thesis, Chalmers University of Technology, Gothenburg, Sweden.

[168] Rodet X. (1984). Time-domain formant-wave-function synthesis. *Computer Music Journal 8*, 3, 9–14.

[169] Rosenberg A. E. (1971). Effect of glottal pulse shape on the quality of natural vowels. *J. Acous. Soc. Am. 49*, 583–590.

[170] Rothenberg M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J. Acous. Soc. Am. 53*, 6, 1632–45.

[171] Rothenberg M. (1981). Some relations between glottal air flow and vocal fold contact area. *ASHA Rep. 11*, 88–96.

[172] Rothenberg M. (1992). A multichannel electroglottograph. *J. Voice 6*, 1, 36–43.

[173] Rothenberg M., and Mahshie J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *J. Speech Hear. Res. 31*, 338–51.

[174] Roubeau B. (1993). *Mécanismes vibratoires laryngés et contrôle neuro-musculaire de la fréquence fondamentale.* thèse, Université Paris XI, Orsay.

[175] Roubeau B., and Castellengo M. (1993). Revision of the notion of voice register. *XIXth International CoMeT Congress, Utrecht.*

[176] Roubeau B., Chevrie-Muller C., and Arabia C. (1991). Control of laryngeal vibration in register change. In *Vocal Fold Physiology : Acoustic, perceptual, and physiological aspects of voice mechanisms*, J. Gauffin and B. Hammarberg, Eds. Singular Publishing Group, San Diego, pp. 279–86.

[177] Roubeau B., Chevrie-Muller C., and Arabia-Guidet C. (1987). Electroglottographic study of the changes of voice registers. *Folia Phoniatr. 39*, 280–89.

[178] Roux J. L., and Gueguen C. (1977). A fixed point computation of partial correlation coefficients. *IEEE Trans. Acoustics and Speech Signal Processing 25*, 3, 257–59.

[179] Saporta G. (1990). *Probabilités, analyse des données et statistique.* Editions Technip, Paris.

[180] Scherer R. C., Arehart K. H., Guo C. G., Milstein C. F., and Horii Y. (1998). Just noticeable differences for glottal flow waveform characteristics. *J. Voice 12*, 1, 21–30.

[181] Schutte H. K. (1980). *The efficiency of voice production.* Ph.d. thesis, Rijksuniversiteit Groningen, the Netherlands.

[182] Schutte H. K., and Miller D. G. (1986). The effect of f_0/f_1 coincidence in soprano high notes on pressure at the glottis. *Journal of Phonetics 14*, 385–92.

[183] Stevens K. N. (1960). Toward a model for speech recognition. *J. Acous. Soc. Am. 32*, 1, 47–55.

[184] Stevens K. N., and Hanson H. M. (1995). Classification of glottal vibration from acoustic measurements. In *Vocal Fold Physiology : Voice quality control*, O. Fujimura and M. Hirano, Eds. Singular Publishing Group, San Diego, California, pp. 147–70.

[185] Sundberg J. (1973). Observations on a professional soprano singer. *STL-QPSR 1*, 14–24.

[186] Sundberg J. (1977). The acoustics of the singing voice. *Scientific American 236*, 3.

[187] Sundberg J. (1987). *The Science of the Singing Voice.* Northern Illinois University Press, Dekalb, Illinois.

[188] Sundberg J. (1995). Vocal fold vibration patterns and modes of phonation. *Folia Phoniatr. 47*, 218–28.

[189] Sundberg J. (1996). The human voice. In *Comprehensive Human Physiology, Vol. 1*, R. Greger and U. Windhorst, Eds. Springer-Verlag, Berlin Heidelberg, pp. 1095–1104.

[190] Sundberg J., Andersson M., and Hultqvist C. (1999). Effects of subglottal pressure variation on professional baritone singers' voice sources. *J. Acous. Soc. Am. 105*, 1965–71.

[191] Sundberg J., Cleveland T. F., Stone R. E., and Iwarsson J. (1999). Voice source characteristics in six premier country singers. *J. Voice 13*, 2, 168–83.

[192] Sundberg J., and Högset C. (1999). Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *STL-QPSR 3-4*, 65–74.

[193] Sundberg J., Titze I. R., and Scherer R. C. (1993). Phonatory control in male singing : a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *J. Voice 7*, 1, 15–29.

[194] Svec J. G., Schutte H. K., and Miller D. G. (1999). On pitch jumps between chest and falsetto registers in voice : Data from living and excised human larynges. *J. Acous. Soc. Am. 106*, 3, 1523–31.

[195] Tanaka S., and Gould W. J. (1983). Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects. *J. Acous. Soc. Am. 73*, 4, 1316–21.

[196] Teaney D., and Fourcin A. J. (1980). The electrolaryngography as a clinical tool for the observation and analysis of vocal fold vibration. *The Voice Foundation.*

[197] Timcke R., von Leden H., and Moore P. (1958). Laryngeal vibrations : measurements of the glottic wave. *AMA Arch. Otolaryngology 68*, 1–19.

[198] Titze I. (1990). Interpretation of the electroglottographic signal. *J. Voice 4*, 1, 1–9.

[199] Titze I. R. (1973). The human vocal cords : a mathematical model. *Phonetica 28*, 129–70.

[200] Titze I. R. (1984). Parameterization of the glottal area, glottal flow and vocal fold contact area. *J. Acous. Soc. Am. 75*, 570–80.

[201] Titze I. R. (1989). A four-parameter model of the glottis and vocal fold contact area. *Speech Communication 8*, 191–201.

[202] Titze I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acous. Soc. Am. 85*, 2, 901–6.

[203] Titze I. R. (1992). Phonation threshold pressure : a missing link in glottal aerodynamics. *J. Acous. Soc. Am. 91*, 5, 2926–35.

[204] Titze I. R. (1994). *Principles of Voice Production.* Prentice-Hall Inc., Englewood Cliffs, New York.

[205] Titze I. R. (1995). Definitions and nomenclature related to voice quality. In *Vocal Fold Physiology : Voice quality control*, O. Fujimura and M. Hirano, Eds. Singular Publishing Group, San Diego, California, pp. 335–42.

[206] Titze I. R., Baer T., Cooper D., and Scherer R. (1984). Automatic extraction of glottographic waveform parameters and regression to acoustic and physiologic variables. In *Vocal fold physiology : contemporary research clinical issues*, A. J. Bless DM, Ed. College Hall, San Diego, pp. 146–54.

[207] Titze I. R., Baken R. J., and Herzel H. (1993). Evidence of chaos in vocal fold vibration. In *Vocal Fold Physiology : Frontiers in Basic Science*, I. R. Titze, Ed. Singular Publishing Group, San Diego, California, pp. 143–88.

[208] Titze I. R., Mapes S., and Story B. (1994). Acoustics of the tenor high voice. *J. Acous. Soc. Am. 95*, 2, 1133–42.

[209] Titze I. R., and Sundberg J. (1992). Vocal intensity in speakers and singers. *J. Acous. Soc. Am. 91*, 5, 2936–46.

[210] Van den Berg J. (1958). Myoelastic-aerodynamic theory of voice production. *J. Speech Hear. Res. 1*, 227–44.

[211] Van den Berg J., Zantema J. T., and Doornenbal P. (1957). On the air resistance and the bernouilli effect of the human larynx. *J. Acous. Soc. Am. 29*, 5, 626–31.

[212] Veeneman D. E., and BeMent S. L. (1985). Automatic glottal inverse filtering from speech and electroglottographic signals. *IEEE Trans. Acoustics and Speech Signal Processing ASSP-33*, 2, 369–77.

[213] Veldhuis R. (1998). A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation. *J. Acous. Soc. Am. 103*, 566–571.

[214] Vennard W. (1967). *Singing : the mechanism and the technic*. Carl Fischer Inc., New York.

[215] Vu Ngoc T., and d'Alessandro C. (1999). Robust glottal closure detection using the wavelet transform. In *Proc. Eurospeech 1999* (Budapest, Hungary, Sept. 1999), pp. 2805–08.

[216] Vu Ngoc T., and d'Alessandro C. (2000). Glottal closure detection using EGG and the wavelet transform. In *Proc. 4th Intern. Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research* (Jena, Germany, Apr. 2000), pp. 147–54.

[217] Wong D. Y., Markel J. D., and Gray A. H. (1979). Least squares glottal inverse filtering from the acoustic speech waveform. *IEEE Trans. Acoustics and Speech Signal Processing ASSP-27*, 350–55.

[218] Yegnanarayana B., d'Alessandro C., and Darsinos V. (1998). An iterative algorithm for decomposition of speech signals into periodic and aperiodic components. *IEEE Trans. Speech and Audio Processing 6*, 1, 1–11.

Table des matières

INTRODUCTION **1**

Partie préliminaire : Caractérisation de la source glottique **7**

1 Description physiologique **11**

1.1 Description statique et dynamique du larynx 11

1.1.1 Anatomie du larynx et vibration laryngée 11

1.1.2 Actions musculaires 13

1.1.3 Quelques ordres de grandeurs 14

1.2 Les mécanismes laryngés 14

1.2.1 Illustration 14

1.2.2 Définition 15

1.2.3 Terminologie 16

1.2.4 Les transitions entre mécanismes 18

1.2.5 Contrôle de la fréquence fondamentale 19

1.3 Méthodes exploratoires 20

1.4 Conclusion 21

2 Modélisation et paramètres de source glottique **23**

2.1 Le modèle de signal source-filtre 23

2.1.1 Présentation 23

2.1.2 Les limites 25

2.1.3 Modélisation de la production vocale 25

2.2 Les paramètres de source glottique 26

2.2.1 Définition 26

2.2.2 Influence sur la qualité vocale 28

2.3 Conclusion 32

I Première partie : Analyse temporelle et spectrale de la source glottique 33

3 Modélisation de la source glottique 37

3.1 Unification des modèles temporels de source glottique et formalisme . . . 37
3.1.1 Les paramètres . . . 38
3.1.2 Expression générique dans le cas d'une fermeture abrupte . . . 38
3.2 Effet spectral des paramètres de source glottique . . . 42
3.2.1 Effet spectral du paramètre d'amplitude . . . 42
3.2.2 Effet spectral du quotient ouvert et du coefficient d'asymétrie . . . 42
3.2.3 Position spectrale du formant glottique . . . 48
3.2.4 Effet spectral du quotient de phase de retour . . . 48
3.3 Conclusion . . . 50

4 Estimation spectrale des paramètres de source glottique 55

4.1 Etude théorique de la différence spectrale $(H_1 - H_2)$ . . . 55
4.1.1 Introduction . . . 55
4.1.2 Etude théorique . . . 56
4.1.3 Discussion . . . 60
4.2 Estimation de la source glottique . . . 63
4.2.1 Les méthodes d'estimation par prédiction linéaire . . . 63
4.2.2 Application à l'estimation de la source glottique par filtrage inverse . . . 64
4.2.3 Evaluation des méthodes . . . 65
4.3 Mesures expérimentales de $(H_1 - H_2)$ . . . 72
4.3.1 Méthodes de mesure . . . 72
4.3.2 Application à la parole . . . 74
4.3.3 Etude de $(H_1 - H_2)$ dans le cas du chant . . . 76
4.3.4 Discussion . . . 78
4.4 Conclusion . . . 80

II Deuxième partie : Mesures acoustiques et électroglottographiques 83

5 Principe et application de l'électroglottographie 87

5.1 Description du dispositif . . . 87
5.2 Caractérisation des signaux électroglottographiques . . . 88
5.2.1 Une représentation du contact entre les cordes vocales . . . 88
5.2.2 Discussion sur la représentation d'un signal EGG . . . 89
5.2.3 Description d'une période de signal EGG . . . 90

5.3 Apport des signaux électroglottographiques 94
5.4 Conclusion 95

6 Constitution d'une base de données **97**
6.1 Le dispositif expérimental 97
6.2 Les chanteurs 99
6.3 Le protocole d'enregistrement 101
6.4 Conclusion 103

7 Etude des signaux DEGG **105**
7.1 Des signaux électroglottographiques à leurs dérivées 105
7.1.1 Discussion sur la validité des mesures 107
7.2 Analyse des signaux DEGG de la base de données 107
7.2.1 Pics bien marqués 107
7.2.2 Pics dédoublés ou peu marqués 109
7.2.3 Visualisation des passages entre mécanismes laryngés sur les signaux DEGG 116
7.3 Mesure de la fréquence fondamentale et du quotient ouvert 119
7.3.1 Détection directe par seuillage 119
7.3.2 Mesure par corrélation 119
7.3.3 Comparaison entre les méthodes de mesure de quotient ouvert 125
7.4 Conclusion 127

8 Mesures du quotient ouvert en voix parlée et chantée **129**
8.1 Revue bibliographique 129
8.1.1 Mesures par glottographie (électro-, photo- et visualisation cinématographique) . . 130
8.1.2 Mesures par filtrage inverse du signal de pression 131
8.1.3 Résumé 131
8.2 Effet des mécanismes laryngés 132
8.2.1 Illustration 132
8.2.2 Analyse en contexte (parlé, chanté) 136
8.2.3 Conclusion 138
8.3 Fréquence fondamentale et intensité vocale 138
8.4 Quotient ouvert et intensité vocale 139
8.4.1 Illustration sur un son filé 139
8.4.2 Production vocale en mécanisme I 140
8.4.3 Production vocale en mécanisme II 144
8.4.4 Analyse en contexte chanté 145
8.4.5 Conclusion 148

8.5 Quotient ouvert et fréquence fondamentale . . . 149
8.5.1 Production vocale en mécanisme I . . . 149
8.5.2 Production vocale en mécanisme II . . . 150
8.5.3 Conclusion . . . 150
8.6 Effet de la voyelle . . . 151
8.7 Comparaison parlé - chanté . . . 155
8.7.1 Exemple du baryton B2 . . . 155
8.7.2 Exemple de la soprano S2 . . . 158
8.8 Conclusion . . . 160

III Troisième partie : Etude perceptive des paramètres de source glottique 165

9 Seuils de perception des paramètres de source glottique 169
9.1 Problématique . . . 169
9.1.1 Le seuil différentiel . . . 169
9.1.2 Expérience de Scherer & al [180] . . . 170
9.2 Méthode adaptative à choix forcé . . . 171
9.2.1 Quelle méthode pour estimer un seuil différentiel ? . . . 171
9.2.2 Description de la méthode adaptative à choix forcé 1-up 2-down . . . 172
9.2.3 Implémentation de la méthode . . . 173
9.3 Mise en place du test . . . 176
9.3.1 Synthèse des stimuli . . . 177
9.3.2 Description des sessions . . . 178
9.3.3 Déroulement du test . . . 180
9.3.4 Les sujets . . . 181
9.4 Résultats . . . 183
9.4.1 Résultats généraux . . . 183
9.4.2 Mesure du seuil pour 3 valeurs de quotient ouvert . . . 185
9.4.3 Mesure du seuil pour 2 valeurs de coefficient d'asymétrie . . . 188
9.4.4 Variations du seuil différentiel de quotient ouvert en fonction du contexte . . . 190
9.5 Conclusion . . . 193

Exemples sonores 196

CONCLUSION GENERALE DE L'ETUDE 201

ANNEXES 207

A Expressions temporelles et spectrales des modèles d'ODG **209**

B Analyse par prédiction linéaire **213**

C Protocole d'enregistrement **221**

D Un peu de statistique ... **225**

E Variations de l'intensité vocale en fonction du quotient ouvert **229**

F Visualisation des glissandos **237**

G Etude des premières mesures de l' Ave Maria de Gounod **253**

BIBLIOGRAPHIE **273**

TABLE DES MATIERES **283**