

UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS VI

ÉCOLE DOCTORALE SMAER
SCIENCES MÉCANIQUES, ACOUSTIQUE, ÉLECTRONIQUE ET ROBOTIQUE
DE PARIS

THÈSE

Pour obtenir le titre de
DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE
SPÉCIALITÉ : SIGNAL, IMAGE, PAROLE, TÉLÉCOMS

Présentée par

Monsieur THOMAS HÉZARD

Production de la voix : exploration, modèles et analyse/synthèse

Thèse dirigée et encadrée par Thomas HÉLIE, René CAUSSÉ et Boris DOVAL,
préparée à l'Institut de Recherche et Coordination Acoustique/Musique,
CNRS UMR 9912 Sciences et Technologies de la Musique et du Son.

Date prévisionnelle de soutenance : 9 décembre 2013.

Composition du jury :

<i>Rapporteurs :</i>	Thierry DUTOIT	- Professeur, Faculté Polytechnique de Mons
	Xavier PELORSON	- Directeur de recherche, HDR, CNRS, GIPSA-lab
<i>Examineurs :</i>	Philippe DEPALLE	- Professeur, Schulich School of Music, McGill University
	Malte KOB	- Professeur, Hochschule für Musik Detmold
	Pierre-Yves LAGRÉE	- Directeur de recherche, HDR, CNRS, IJLRA, UPMC
	Christophe VERGEZ	- Chargé de recherche, HDR, CNRS, LMA
<i>(collaborateur)</i>	Boris DOVAL	- Maître de conférence, LAM, IJLRA, UPMC
<i>Co-directeur :</i>	Thomas HÉLIE	- Chargé de recherche, HDR, CNRS, IRCAM
<i>Co-directeur (invité) :</i>	René CAUSSÉ	- Directeur de recherche, HDR, IRCAM

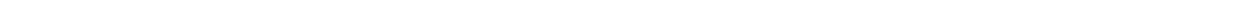
version 1.1 - 7 janvier 2014

contact : thomas.hezard@thz.fr

à mes parents.

Ce sont des voix qui restent en dernier, tout comme c'est la voix souvent qui, comme un parfum, précède et annonce l'entrée physique de quelqu'un dans votre vie.

—JEAN-JACQUES SCHUHL, INGRID CAVEN



Résumé

Ce travail porte sur l'étude de la production de la voix. L'objectif principal est de proposer des solutions pour son analyse-synthèse qui permettent une large représentation de la diversité des signaux biométriques (observés en pratique) dans le cadre du paradigme source-filtre ou de modèles physiques simplifiés. L'étude est menée selon trois axes, l'exploration de l'activité laryngée, la modélisation de la phonation et l'analyse-synthèse sonore.

Après une description de l'appareil phonatoire et de modèles d'analyse-synthèse actuels, la première partie s'intéresse à l'exploration de signaux biométriques. À partir du recensement des dispositifs actuels, nous proposons un outil de synchronisation qui révèle la complexité du mouvement 3D des plis vocaux. Une étude comparative avec un système acoustique à valve humaine similaire (lèvres d'un tromboniste) est menée et le développement d'un nouvel outil combinant deux techniques récentes d'imagerie électrique est proposé.

Dans la deuxième partie, deux algorithmes d'analyse sont proposés pour un modèle source-filtre fondé sur une représentation tous-pôles causale-anticausale du signal vocal. Un calcul du cepstre complexe est proposé, garantissant l'holomorphie sur une couronne de convergence dans le domaine en Z . Cet outil est alors exploité pour la séparation de la source et du filtre.

Enfin, des représentations paramétriques des signaux d'aire glottique sont exploitées pour piloter des modèles à géométrie forcée. Un modèle de synthèse, inspiré du modèle source-filtre mais intégrant l'effet du couplage entre la glotte et le conduit vocal, et des avancées sur la modélisation de conduit vocal à profil régulier sont présentées.

Mots clés : production de la voix, analyse-synthèse, séparation source-filtre, modèles physiques, mesures in vivo, impédancemétrie.



Abstract

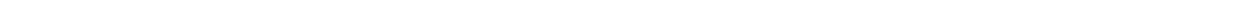
This work focuses on the study of voice production. The main objective is to propose solutions for its analysis-synthesis which enable a wide representation of the diversity of biometric signals (observed in practice) within the framework of source-filtre paradigm or simplified physical models. The study is conducted according to three axes : exploration of laryngeal activity, modelisation of the phonation and sound analysis-synthesis.

After a description of the vocal apparatus and current analysis-synthesis models, the first part focuses on the exploration of biometric signals. From an inventory of current devices, we propose a synchronisation tool which reveals the complexity of the 3D motion of vocal folds. A comparative study with a similar human-valve-based sound production system (lips of a trombonist) is conducted and the development of a new tool combining two recent electrical imaging techniques is proposed.

In the second part, two new analysis algorithms are proposed for a source-filter model based on a causal-anticausal all-pole representation of speech signals. A computation method of the complex cepstrum is proposed, which ensures the holomorphy on a convergence domain in the Z plane. This tool is then used for the separation of the source and the filter.

Finally, parametric representations of glottal area waveforms are used to drive models controlled by the geometry. A synthesis model, based on the source-filter model but including the effect of the coupling between the glottis and the vocal tract, and advances in the modeling of the vocal tract with regular profile are presented.

Keywords : voice production, analysis-synthesis, source-filter separation, physical models, in vivo measurements, impedancemetry.



Remerciements

Une thèse ne se fait pas seul et de nombreuses personnes m'ont accompagné pendant mes travaux. Je tiens à les remercier et à leur rendre hommage ici, en adressant par avance toutes mes excuses à ceux que je pourrais oublier.

Je tiens tout d'abord à exprimer toute ma gratitude à mes directeurs et encadrants Thomas Hélie, René Caussé et Boris Doval, pour leur accompagnement et leurs conseils scientifiques tout au long de mon travail. J'adresse une pensée particulière à Thomas Hélie, qui m'a encadré au quotidien pendant ces trois années, pour ses précieux conseils tant scientifiques que personnels, sa disponibilité sans faille et sa passion communicative pour les sciences. C'est grâce à sa présence que la notion de formation qui accompagne la thèse de doctorat a pris tout son sens.

Je remercie ensuite tous les membres du jury d'avoir accepté de s'investir dans la lecture de ce document. C'est grâce à la bonne volonté de chacun d'entre eux que j'ai réussi à réunir dans mon jury de telles compétences dans chacune des disciplines abordées.

Je tiens à remercier l'école doctorale SMAER qui m'a accordé sa confiance en choisissant de soutenir mon projet de recherche. Je ne saurais oublier les précieux conseils de Fleur Gire et Carlos Agon (M2R ATIAM), Boris Doval et Jean-Dominique Polack (IJLRA) et Hugues Vinet (IRCAM) lors de ma recherche de financement. Toujours à l'ED SMAER, j'adresse des remerciements chaleureux à Anne-Marie Aubin, qui œuvre au quotidien pour assister les doctorants dans leurs démarches administratives.

Cette thèse a reçu le soutien de la Société Française Acoustique et du Groupe Spécialisé en Acoustique Musicale, que je remercie pour les bourses qu'ils m'ont attribuées pour des participations à des congrès et séminaires.

Ma thèse fut l'occasion de nombreuses rencontres et collaborations scientifiques. Je tiens à remercier, pour leur apport dans mon travail et leur sympathie (dans l'ordre alphabétique), Erkki Bianco, Gilles Degottex, Vincent Fréour, Hervé Gagnon, Jean-Luc Gennisson, Nathalie Henrich, Malte Kob, Aude Lagier, Thierry Legou et Mickaël Tanter. J'ai eu de plus l'immense plaisir de participer à une journée de travail du Groupe Larynx organisée par Nathalie Henrich que je remercie pour sa participation active à la mise en place de rencontres interdisciplinaires autour des sciences de la voix.

J'ai eu l'opportunité de co-encadrer les stages de Mathias Le Borgne et Antonin Mauffrey à l'IRCAM. Je les remercie tous les deux pour leurs travaux, leur sympathie et l'intérêt qu'ils ont porté à leur sujet.

Je tiens à remercier l'équipe encadrante des projets mécatroniques de l'école Mines ParisTech, Yvon Gaignebet, Denis Brousse, Frédéric Bruyère et Jérémie Rupil ainsi que tous les étudiants qui ont participé à l'avancement du projet VoiceTronics pour leur investissement et leur sympathie.

Cette thèse s'est déroulée à l'IRCAM et je remercie ses directeurs scientifiques Hugues Vinet et Gérard Assayag pour leur accueil ainsi que les équipes Analyse-Synthèse des Sons et Acoustique Instrumentale et leurs directeurs respectifs Xavier Rodet puis Axel Röbel et René Caussé qui m'ont offert un cadre de travail privilégié. Je remercie toutes les personnes de l'IRCAM qui ont participé de près ou de loin à mon travail, notamment l'équipe administrative avec Martine Grospron, Carole Tan et Sylvie Benoit, Emmanuel Fléty pour son expertise en électronique analogique et numérique, Alain Terrier pour ses conseils et réalisations mécaniques ainsi que l'inimitable équipe informatique de l'IRCAM, Laurent Ghys, Arnaud Gomes et Ghislaine Montagne, en adressant une pensée vive et chaleureuse à Youcef Bensaid et à ses proches, sans oublier la régie bâtiment, Georges-Elie Giscard, Alain Nicolas et Chantal Vogel qui ont toujours accueilli mes plaintes récurrentes sur la climatisation ou les luminaires défectueux avec gentillesse. Je remercie plus généralement tout le personnel de l'IRCAM (et mes co-promus du master ATIAM) avec qui ce fut toujours un plaisir d'échanger. J'adresse une pensée particulière à mes compagnons de rédaction Louis Bigo et Benjamin Lévy, à Charles Picasso pour nos inoubliables discussions musicales et philosophiques, à Baptiste Vericel en souvenir des années d'étude passées ensemble, ainsi qu'à Béatrice Laroche et Valérie Dos Santos Martins pour leurs conseils réconfortants pendant la rédaction.

Sur un plan plus personnel, je tiens à profiter de ces quelques lignes pour remercier toutes les personnes qui m'ont encouragé dans la voie de la recherche et qui m'ont donné le goût des sciences pendant mes études, notamment Sylvain Reynal, Laurence Hafemeister, Michel Corneloup et Antoine Tauvel de l'ENSEA ainsi que toute l'équipe enseignante du M2R ATIAM et Rémi Mignot qui a co-encadré mes deux stages à l'IRCAM.

Je tiens à remercier tout particulièrement et très chaleureusement mes parents. Le travail présenté ici est en quelques sortes le point d'orgue de longues années d'études pendant lesquelles ils m'ont toujours soutenu, encouragé et conseillé. Votre appétit pour la vie ainsi que le courage et l'inébranlable humanité avec lesquels vous faites face aux épreuves de l'existence sont une constante source d'admiration et d'inspiration. Je vous dédie cet ouvrage avec toute mon affection. Je remercie plus généralement tous mes proches qui ont su m'entourer et me soutenir pendant les périodes les plus difficiles de cette aventure, avec une pensée particulière pour Armel, Marie, Didier et Françoise.

J'ai évidemment une pensée pour tous mes amis, leurs encouragements et les bons moments passés avec eux avant pendant ces trois années. Je remercie, dans le désordre et en en oubliant certainement quelques-uns à cette heure tardive, Maxime pour être toujours présent quand j'en ai besoin, Mélusine pour être à la fois la personne la plus horripilante et la plus attachante de la planète, Farah pour tous ses conseils musicaux qui ont accompagné mes longues heures de rédaction, Stéphane pour ses truculentes histoires du bout du monde, Vladimir pour sa bonne humeur inébranlable, Gisèle pour ses récits où la misanthropie croise l'humour avec toujours autant de brio, Daniel pour toujours trouver une bonne raison de râler, Bastien pour son naturel et ses voyages qui font rêver, Alexis et Sarah pour leur gentillesse simple qui fait du bien. Sans oublier Julien et Clara qui nous ont donné l'inspiration, Laetitia et Gilles pour les verres de champagne dans le jacuzzi sous les étoiles, leurs enfants si adorablement fatiguants, Émilie, Michaël, Julien, Thomas et toute la troupe niçoise. Et enfin tous les parisiens et ex-parisiens avec qui j'ai partagé de nombreux bons moments pendant ces trois années et qui n'ont jamais manqué une occasion pour prendre de mes nouvelles et m'encourager.

Je termine avec une pensée affectueuse adressée à François pour tout ce qu'il m'apporte au quotidien et pour son soutien pendant ces trois années, notamment pendant ces dernières semaines de rédaction malgré mes horaires impossibles et mon humeur aléatoire.

Notes liminaires

Le terme “cordes vocales” tend à disparaître des documents scientifiques au profit du terme “plis vocaux”, plus proche de la physiologie et correspondant à la terminologie anglophone de “vocal folds”. Pour ces raisons, le terme “plis vocaux” sera celui employé dans ce document.

Des notations, abréviations et expressions usuelles associées à la production de la voix sont utilisées tout au long de ce document. Elles seront généralement définies lors de leur première utilisation. Elles sont rappelées en fin de document, à partir de la page 259.

Mon travail de thèse, pluridisciplinaire par la nature de l’objet étudié, a été l’occasion de plusieurs collaborations (en supplément de l’encadrement de ma thèse) auxquelles il sera fait référence dans ce document. Je remercie et liste ici les personnes avec qui j’ai eu l’occasion de travailler.

- Nathalie Henrich, chargée de recherche CNRS au Gipsa-lab à Grenoble, France.
- Malte Kob, professeur à la Hochschule für Musik à Detmold, Allemagne.
- Thierry Legou, ingénieur de recherche au Laboratoire Parole et Language à Aix-en-Provence, France.
- Aude Lagier, chirurgienne ORL, chef de clinique assistante à l’Hôpital de La Timone à Marseille, France.
- Mickael Tanter, directeur de recherche INSERM à l’ESPCI à Paris, France.
- Jean-Luc Gennisson, chargé de recherche CNRS à l’ESPCI à Paris, France.
- Vincent Fréour, doctorant à l’université McGill à Montréal, Canada.

Nathalie Henrich, Malte Kob, Thierry Legou et Aude Lagier ont participé au projet VoiceTronics, initié par Thomas Hélie, Nathalie Henrich et moi-même au début de ma thèse, en septembre 2010. Ce projet fut l’occasion de rencontres et collaborations supplémentaires avec des étudiants et enseignants de l’école Mines ParisTech. Il est détaillé dans l’annexe A.

J’ai eu l’opportunité et le plaisir d’encadrer trois stages pendant ma thèse. Ces stages ont permis des avancées significatives dans plusieurs pans de mon travail et il y sera fait référence dans ce document. Je remercie ici mes stagiaires pour leurs travaux dans le cadre

- du stage de M2R ATIAM de Mathias Le Borgne du 01/03/2012 au 17/08/2012,
- du stage de fin de 2^{ème} année d’école d’ingénieur d’Antonin Mauffrey, étudiant à l’École Nationale Supérieure de l’Électronique et de ses Applications (ENSEA) du 18/06/2012 au 31/08/2012,
- du Projet de Fin d’Études de l’ENSEA d’Antonin Mauffrey du 01/04/2013 au 30/09/2013.

La bibliographie présentée en fin d’ouvrage est particulièrement riche en raison de (i) l’aspect pluridisciplinaire du travail présenté dans cette thèse, et (ii) une volonté d’exhaustivité de la bibliographie lors de l’utilisation d’outils et de méthodes préexistants.



Sommaire

Préface	i
Résumé	iii
Abstract	v
Remerciements	vii
Notes liminaires	ix
Sommaire	xi
Introduction générale	1
Partie préliminaire : Présentation de l'objet d'étude et position du problème	9
1 Production de la voix humaine	13
1.1 Anatomie de l'appareil phonatoire	14
1.2 Voix parlée, voix chantée et mécanismes laryngés	18
1.3 Méthodes exploratoires	21
1.4 Conclusion	23
2 Différents modèles pour l'analyse et la synthèse sonores de la voix	25
2.1 Signal vocal et quasi-stationnarité	26
2.2 Modèles physiques et modèles de signaux	26
2.3 Cuivres et voix, des liens privilégiés	31
2.4 Position du problème et organisation du document	31
I Première partie : Outils pour l'exploration et la mesure de signaux biométriques	35
3 Méthodes pour l'exploration laryngée	39
3.1 Ordres de grandeur temporel et spatial du mouvement glottique	40
3.2 Échographie	40
3.3 Vidéo-endoscopie ultra-rapide	41

3.4	Électroglottographie	42
4	Mesures multimodales sur les lèvres et la glotte, comparaison entre deux systèmes à valve humaine	45
4.1	Introduction	47
4.2	Experimental setup	49
4.3	Visualisation tool	52
4.4	Features extraction, analysis and comparisons	54
4.5	Conclusion	66
5	Imagerie électrique multicanal du larynx	69
5.1	Tomographie d'impédance électrique	71
5.2	Test préliminaire de validation du principe de l'EIT sur une maquette de larynx simplifiée	73
5.3	Conception d'un dispositif mixte EGG/EIT multicanal à géométrie variable	75
5.4	Développements et réalisations techniques	78
5.5	Conclusion et perspectives	86
II	Deuxième partie : Modèle source-filtre et analyse-synthèse	91
6	Le modèle source-filtre	95
6.1	Source glottique	96
6.2	Filtre vocal et rayonnement	100
6.3	Modèle source-filtre paramétrique complet étudié	102
7	Analyse du modèle source-filtre	103
7.1	État de l'art des méthodes d'estimation et de séparation	104
7.2	Premier algorithme : Séparation source-filtre pour la classe des signaux périodiques .	105
7.3	Deuxième algorithme : Analyse cepstrale et séparation source-filtre	115
7.4	Conclusion et perspectives	137
III	Troisième partie : Modèles à géométrie forcée et analyse-synthèse	141
8	Analyse-Synthèse de fonctions d'aire glottique	145
8.1	Construction d'un dictionnaire de test par estimation de fonctions d'aire glottique à partir de vidéo-endoscopies ultra-rapides	146
8.2	Modèle LF	147
8.3	Modèles de diversité en systèmes dynamiques non linéaires	157
8.4	Conclusion	166

9 Modèles à géométrie forcée	169
9.1 Modèle couplé inspiré du modèle source-filtre	170
9.2 Apport du couplage entre la source glottique et le conduit vocal	175
9.3 Modélisation de conduit vocal à profil régulier en guides d'ondes acoustiques 1D	179
9.4 Conclusion et perspectives	188
Conclusion générale	193
Bibliographie	199
Annexes	219
A Projet VoiceTronics	219
A.1 Projets mécatroniques à l'école Mines ParisTech	220
A.2 SP2 : Maquette de larynx robotisée à échelle réelle et à géométrie simplifiée	220
A.3 SP3 : Maquette de larynx robotisée à l'échelle 5 :1	221
B Modèles de pulse glottique par fonctions paramétriques définies par morceaux	225
B.1 Forme générique	225
B.2 Expressions analytiques des modèles	226
C Étude analytique du modèle LF	229
C.1 Écriture temporelle du modèle de dérivée du débit	229
C.2 Écriture temporelle adimensionnée du modèle de dérivée du débit	230
C.3 Équations implicites et bilan des degrés de liberté	230
C.4 Écriture temporelle adimensionnée du modèle de débit	231
C.5 Transformées de Fourier des modèles temporel et dérivé	232
D Dictionnaire de cepstres en temps discret	233
D.1 Impulsion de Dirac	233
D.2 Pôle	233
D.3 Zéro	234
D.4 Propriétés remarquables	235
E Modèles de fonction d'aire glottique	237
F Modele de Lur'e	239
F.1 Modèle de Lur'e en temps continu	239
F.2 Bases de fonctions pour la non-linéarité ψ	241

F.3 Première tentative d'analyse de fonctions d'aire glottique	242
G Propagation dans les tubes acoustiques	251
G.1 Modèle général et matrices de transfert acoustique	251
G.2 Cas des tubes droits	252
G.3 Observations sur les fonctions H_t et Z_{in}	255
Notations, acronymes et expressions usuelles	259
Liste des tableaux	263
Table des figures	265
Table des matières	271





Introduction générale

La communication est au cœur de l'activité humaine et de nos sociétés. Si les modes de communication tendent à se diversifier, la communication orale tient toujours une place privilégiée dans nos échanges. Depuis cinquante ans, les technologies permettant de transmettre la voix, de la faire voyager mais aussi de la synthétiser, de l'analyser et de la modifier sont en perpétuelle évolution. Ainsi aujourd'hui la synthèse des voix parlées et chantées est devenue commune et elle a envahi notre quotidien, depuis les halls de gare jusqu'à nos appareils électroniques (téléphones, logiciels de synthèse musicale, systèmes d'assistance à la conduite etc.), en passant par la publicité et le cinéma. Les systèmes de reconnaissance vocale ont fait leur apparition il y a quelques années, d'abord fondés sur l'apprentissage d'un vocabulaire très restreint, puis prenant en compte des vocabulaires plus complets et des formules syntaxiques plus élaborées.

Les mécanismes à l'origine de la production de la voix humaine sont d'une grande complexité. Le larynx, organe clé de la phonation qui abrite les plis vocaux, est constitué de plusieurs cartilages et de nombreux muscles permettant de contrôler leurs articulations. La vibration des plis vocaux à l'origine de la phonation est un phénomène d'auto-oscillation. Ceci signifie que leur oscillation n'est pas due à une activité musculaire périodique mais à leur mise en vibration par les effets conjugués des forces mécaniques et aérodynamiques qui s'exercent dans le larynx lorsque les poumons expulsent de l'air. Le fonctionnement du larynx, en temps que mécanisme dynamique oscillant, est assez similaire à celui des lèvres d'un musicien jouant d'un instrument à embouchure (famille des cuivres principalement). Le contrôle de la vibration laryngée est effectué par les muscles du larynx, selon des processus qui ne sont pas encore tous compris aujourd'hui. Pour produire les différents sons de la voix, les vibrations provoquées dans le larynx sont modulées par plusieurs résonateurs : la cavité buccale (dont la géométrie est pilotée par la position de la langue, de la mâchoire et des lèvres), la cavité nasale (dont l'ouverture est pilotée par le *velum*) et les sinus paranasaux.

La complexité de ces phénomènes est telle que plusieurs approches ont dû être développées pour modéliser, analyser, synthétiser ou transformer le son vocal, selon les buts visés. Ainsi la communauté scientifique s'intéresse à des approches fondées sur :

1. la concaténation d'unités de sons (diphones, syllables, mots, segments sémantiques),
2. les modèles de signaux (sinusoïdes plus bruit, formes d'ondes formantiques etc.),
3. la modélisation source-filtre,
4. des modèles physiques simplifiés (e.g. modèle de glotte à deux masses, représentation du conduit vocal par guides d'ondes ou impédances acoustiques),
5. des modélisations vibra-aéro-acoustiques complexes (mécaniques des fluides et des solides déformables couplées).

Aujourd'hui, les approches qui permettent à la fois des synthèses et analyses de qualité sont plutôt de type signal (1-3), alors que les approches plus physiques (4-5) permettent de comprendre le

fonctionnement et de prédire le comportement de la voix. Plus précisément, pour les modèles de type “source-filtre”, des méthodes d’estimation de paramètres sont disponibles [Deg10, Dru11], qui conduisent à une qualité sonore satisfaisante, mais le réalisme sonore est encore limité car des phénomènes significatifs (couplage glotte/conduit vocal, bruits de turbulence, double instant de fermeture de la glotte, etc.) ne sont pas représentés. A l’inverse, il est difficile d’obtenir des méthodes d’estimation et d’inversion robustes pour les modèles physiques (4-5), même si des résultats existent pour des cas plus simples que la voix [HVLR99, dNCH10].

Sujet de la thèse

Cette thèse a pour but de proposer des solutions pour l’analyse-synthèse de la voix qui permettent une large représentation de la diversité des signaux biométriques (observés en pratique) dans le cadre du paradigme source-filtre ou de modèles physiques simplifiés. La production vocale sera étudiée selon trois axes complémentaires :

- A1** l’exploration de l’activité laryngée,
- A2** la modélisation de la phonation,
- A3** l’analyse et la synthèse sonore.

A1 Les dispositifs d’exploration de la vibration laryngée seront exploités pour étudier le mouvement des plis vocaux et le comparer avec celui des lèvres dans le cas des cuivres. Un dispositif électronique d’exploration non invasive de l’activité laryngée, qui a pour but de repousser les limites des dispositifs actuels (information partielle ou dispositifs invasifs), sera étudié.

A2 Des systèmes de production de la voix dans le cadre du paradigme source-filtre puis des modèles physiques simplifiés seront développés. Nous nous appuyerons sur des observations de plis vocaux en vibration et sur les modèles existants pour construire un modèle piloté par la géométrie glottique. Ce travail se situe donc à la frontière entre les points 3 et 4 de la catégorisation ci-dessus.

A3 Des outils seront proposés pour l’analyse du modèle source-filtre et des modèles couplés.

L’objectif, à terme, est de proposer des modèles héritant à la fois du faible coût de l’approche source-filtre et du réalisme physiologique des modèles physiques simples ainsi que des outils pour l’inversion de ces modèles (estimation des paramètres à partir d’un son cible).

Organisation du document et présentation de la démarche

Ce document est organisé en neuf chapitres regroupés en quatre parties.

Partie préliminaire : présentation de l’objet d’étude et position du problème

La partie préliminaire présente notre objet d’étude : la voix, son exploration et sa modélisation.

- Le chapitre 1 présente l’appareil de production de la voix humaine selon trois points de vue. Tout d’abord, la compréhension de **l’anatomie de l’appareil phonatoire** est nécessaire au développement d’outils pour l’exploration de l’activité laryngée et de modèles inspirés du fonctionnement physique de la phonation humaine. Ensuite, une revue **des différents types de signaux vocaux** nous permet d’aborder la variété des sons que la voix humaine est capable de produire. Notamment, la notion de mécanisme laryngé (régime de vibration des plis vocaux) est rappelée. Enfin, un aperçu **des méthodes d’exploration** des plis vocaux en vibration et des mouvements des articulateurs nous permet de nous familiariser

avec certains outils actuels qui sont exploités dans la suite du document.

- Le chapitre 2 présente les différentes approches développées dans la littérature pour modéliser, synthétiser et analyser la voix humaine. Les deux paradigmes “modèles de signaux” (1-3 dans la catégorisation évoquée plus haut) et “modèles physiques” (4-5 dans la même catégorisation) sont présentés en détail. Les modèles qui nous intéressent particulièrement dans notre étude sont le modèle source-filtre et les modèles physiques simples. La littérature fait apparaître deux types de modèles physiques simples : les modèles à géométrie forcée et les modèles auto-oscillants. Je propose la classification suivante, correspondant à **une évolution dans les degrés de liberté du modèle de glotte**.

C1. Modèles **source-filtre** (source acoustique forcée).

C2. Modèles à **géométrie forcée** (système aéro-acoustique à géométrie pilotée).

C3. Modèles **auto-oscillants** (système vibro-acoustique non linéaire simplifié).

Le travail présenté dans cet ouvrage s’articule alors de la façon suivante :

Partie I. Nous nous intéressons d’abord à l’exploration de la vibration glottique en exploitant des outils existants puis en proposant le développement de nouveaux outils.

Partie II. Ensuite, l’analyse-synthèse de la voix est étudiée pour le paradigme source-filtre (acoustique forcée, **C1**) en focalisant nos efforts sur la construction de nouvelles méthodes de séparation source-filtre.

Partie III. Enfin, nous réintégrons des considérations physiques dans nos modèles, orientés par les mesures obtenues par les outils de la partie I, grâce au paradigme à géométrie forcée (**C2**).

Partie I : Outils pour l’exploration et la mesure de signaux biométriques

La première partie est consacrée à l’exploration de l’activité laryngée afin de nous familiariser avec les caractéristiques de la vibration glottique.

- Le chapitre 3 débute par un état de l’art des méthodes permettant de visualiser la vibration des plis vocaux et se poursuit avec des premiers tests. Deux techniques permettent aujourd’hui de répondre à ce problème délicat : la **vidéo-endoscopie ultra-rapide** et l’**électroglottographie** [Fab57]. Cependant, aucune de ces techniques ne permet une visualisation complète du mouvement tridimensionnel des plis vocaux. L’**échographie ultra-rapide** [SCTH99, TBSF02, DGTF06] a été proposée et testée pendant ma thèse mais les images obtenues se sont révélées inexploitable.
- Le chapitre 4 est consacré à une **étude comparée des plis vocaux en vibration** (dans le cadre de la phonation) et **des lèvres en vibration** (dans le cas des instruments de la famille des cuivres) basée sur des signaux vidéos ultra-rapides et électroglottographiques synchronisés. Je propose un **outil de visualisation des signaux synchronisés** qui permet notamment de mettre en évidence la présence du mouvement des plis vocaux (et des lèvres) dans la dimension perpendiculaire au plan glottique.
- La construction d’un **dispositif électronique de mesure non invasive** visant à accéder à la géométrie des plis vocaux dans cette dimension (effet de buse ayant une influence sur l’aéro-acoustique glottique) est proposée dans le chapitre 5. Son but est d’établir des modèles de pilotage géométrique pour les modèles à géométrie forcée. Ce dispositif combine (i) l’exploitation combinée de la technique de l’électroglottographie et de la tomographie d’impédance électrique [Hol10], (ii) un multiplexage fréquentiel afin de multiplier les canaux d’excitation, et (iii) un système d’électrodes à géométrie variable. Les principes du dispositif sont détaillés et un état d’avancement de la construction des circuits électroniques et de l’interface homme-machine est présenté.

Partie II : Modèle source-filtre et analyse-synthèse

L'objectif de la partie II est de proposer des outils pour l'estimation des paramètres du modèle et la séparation de la contribution de la source et du filtre dans le paradigme source-filtre (C1).

- Le modèle d'étude est présenté dans le chapitre 6. Dans le modèle proposé, le signal de source glottique est généré par un filtrage causal/anticausal d'un train d'impulsions de Dirac, dans l'idée du modèle CALM [DDH03]. Contrairement à la plupart des modèles rencontrés dans la littérature [CC95, DDH06, Deg10], je propose un modèle fondé sur une **représentation tous-pôles du signal vocal**. Cette démarche permet de séparer efficacement dans le modèle la contribution anticausale de la source de la contribution causale du filtre.
- Pour répondre au problème de la séparation source-filtre pour le modèle du chapitre 6, deux nouveaux algorithmes sont proposés dans le chapitre 7. Le premier algorithme repose sur une estimation des pôles du signal analysé permettant d'**isoler la contribution des pôles causaux et des pôles anticausaux** et une **estimation des paramètres sur le cepstre différentiel par un méthode à maximisation de vraisemblance**. Des limitations de cet algorithme m'ont amené à proposer une seconde approche. Le deuxième algorithme est fondé sur la **représentation des contributions causale et anticausale par le cepstre complexe**. Une attention particulière a été portée sur la définition et le calcul du cepstre complexe. Je propose une méthode de calcul de ce dernier qui garantit l'holomorphicité sur une couronne de convergence dans le domaine en Z .

Partie III : Modèles à géométrie forcée et analyse-synthèse

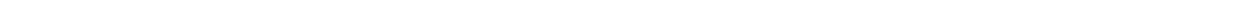
Dans la partie III, nous orientons le travail vers des modèles plus proches du fonctionnement physique de l'appareil phonatoire en se positionnant dans le cadre du paradigme à géométrie forcée (C2). Plus précisément, nous étudions des modèles pour la géométrie forcée de la glotte et du conduit vocal.

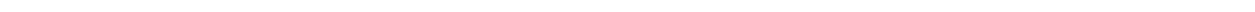
- Le chapitre 8 s'applique à proposer des modèles de fonctions d'aire glottique $A_g(t)$ pour le pilotage du modèle de synthèse (décrit dans le chapitre 9). Deux approches sont envisagées. Tout d'abord, des traitements et observations d'une banque de données de signaux $A_g(t)$ mesurés m'ont amené à proposer de **représenter les fonctions d'aire glottique par le modèle d'impulsion paramétrique de Liljencrants-Fant [FLL85]**. Un algorithme d'estimation des paramètres est proposé et validé. Ce modèle fournit une représentation paramétrique efficace des fonctions d'aire glottique. Toutefois, il ne permet pas d'accéder à la diversité des signaux de mesures observés dans la banque de données. C'est pourquoi je me suis attaché à trouver des modèles capables de générer une plus grande variété de régimes. Un travail qui a retenu notre attention pour traiter cette question est celui de De Feo et Hasler [De 01, DH03]. Suite à une discussion avec M. Hasler, nous nous sommes proposé d'adapter les **modèles de diversité** (systèmes dynamiques non linéaires de Lur'e) et la méthode de classification des signaux biométriques proposés dans [De 01] à notre problème de représentation des signaux d'aire glottique. Un premier essai d'exploitation directe de la méthode proposée a conduit à des résultats trop pauvres. Nous proposons alors d'injecter de la connaissance sur le fonctionnement physique de la vibration glottique dans le modèle dynamique pour améliorer les résultats d'analyse.
- Pour terminer, le chapitre 9 est consacré aux modèles à géométrie forcée (C2). Afin d'obtenir un modèle facilement implémentable, je propose d'abord de construire un **modèle de synthèse numérique inspiré du modèle source-filtre mais intégrant le couplage entre la source glottique et le conduit vocal**. L'influence de l'apport de ce couplage

aéro-acoustique simple est étudiée. Une étude est ensuite menée afin d'**améliorer la modélisation de la propagation acoustique dans les conduits à section variable**. Un modèle de type "équation de pavillons" raffiné permet de prendre en compte l'effet de la géométrie du conduit et de pertes [Hé02]. Des résultats récents permettent de construire une structure en guides d'ondes numériques simulable en temps discret à partir de ce modèle [Mig09]. Ces simulations étant instables dans le cas des tubes concaves¹, je propose une solution pour résoudre le problème de stabilité.

La conclusion dresse un bilan de l'étude et des contributions scientifiques apportées par cette thèse. Les avancées proposées par ce travail selon les trois axes **A1**, **A2** et **A3** sont résumées. La pertinence des méthodes d'exploration, de modélisation et d'analyse-synthèse développées et des résultats obtenus est étudiée et des pistes pour les travaux futurs sont proposées.

1. cf. page 261.





Partie préliminaire

Présentation de l'objet d'étude et position du problème



SOMMAIRE DE LA PARTIE

La partie préliminaire est consacrée à la découverte de notre objet d'étude : la voix, son exploration et sa modélisation. Les notions de base qui ont trait à la phonation seront posées dans cette partie et nous y ferons référence tout au long du document.

Le chapitre 1 s'applique à présenter l'appareil phonatoire humain selon trois axes.

- 1. Physiologie** : Anatomie de l'appareil phonatoire et situation du larynx, de ses cartilages, des plis vocaux et des composantes du conduit vocal. Principe du mouvement vibratoire des plis vocaux et de leur contrôle.
- 2. Signal vocal** : Différents types de voix et leurs caractéristiques. Mécanismes laryngés.
- 3. Exploration** : Méthodes d'exploration du larynx et du conduit vocal.

Le chapitre 2 est consacré à une revue condensée des démarches existantes pour modéliser, analyser et synthétiser la voix humaine. Nous verrons à cette occasion que l'appareil phonatoire et les instruments de la famille des cuivres présentent des liens privilégiés. La dernière section de ce chapitre sera consacrée à la position du problème qui sera étudié dans les parties suivantes.



Chapitre 1

Production de la voix humaine

Sommaire

1.1 Anatomie de l'appareil phonatoire	14
1.1.1 Présentation générale, situation du larynx et des plis vocaux	14
1.1.2 Auto-oscillation des plis vocaux	16
1.1.3 Articulations des cartilages et actions musculaires	17
1.1.4 Contrôle des résonateurs	18
1.2 Voix parlée, voix chantée et mécanismes laryngés	18
1.2.1 Voix parlée et voix chantée	18
1.2.2 Notion de qualité vocale	19
1.2.3 Mécanismes laryngés	19
1.3 Méthodes exploratoires	21
1.3.1 Visualisation des plis vocaux en vibration	21
1.3.2 Méthodes indirectes d'observation de l'activité glottique	22
1.3.3 Corpus de données multi-modales	22
1.3.4 Estimation de la géométrie et des propriétés acoustiques du conduit vocal	22
1.4 Conclusion	23

1.1 Anatomie de l'appareil phonatoire

Afin d'illustrer mes propos dans cette section, j'utiliserai trois outils.

Des planches anatomiques tirées d'ouvrages de médecine. Ces illustrations sont précises mais parfois difficiles à lire pour un œil extérieur au monde médical.

Des photos de moulages en plâtre des cartilages laryngés réalisés par Aude Lagier sur un larynx humain excisé. Chacune de ces photos présente une échelle en centimètres.

Des photos d'une maquette robotisée du larynx à l'échelle 5 :1 présentant les cartilages et leurs articulations simplifiées, réalisée par les étudiants du projet VoiceTronics (cf. annexe A). La réalisation de cette maquette, impulsée à la demande de l'équipe enseignante encadrant le projet, a été l'occasion de fabriquer un objet de démonstration pour aborder la physiologie du larynx et de support pour les discussions avec des spécialistes d'anatomie.

1.1.1 Présentation générale, situation du larynx et des plis vocaux

L'appareil phonatoire humain est présenté en figure 1.1 en coupe sagittale¹.

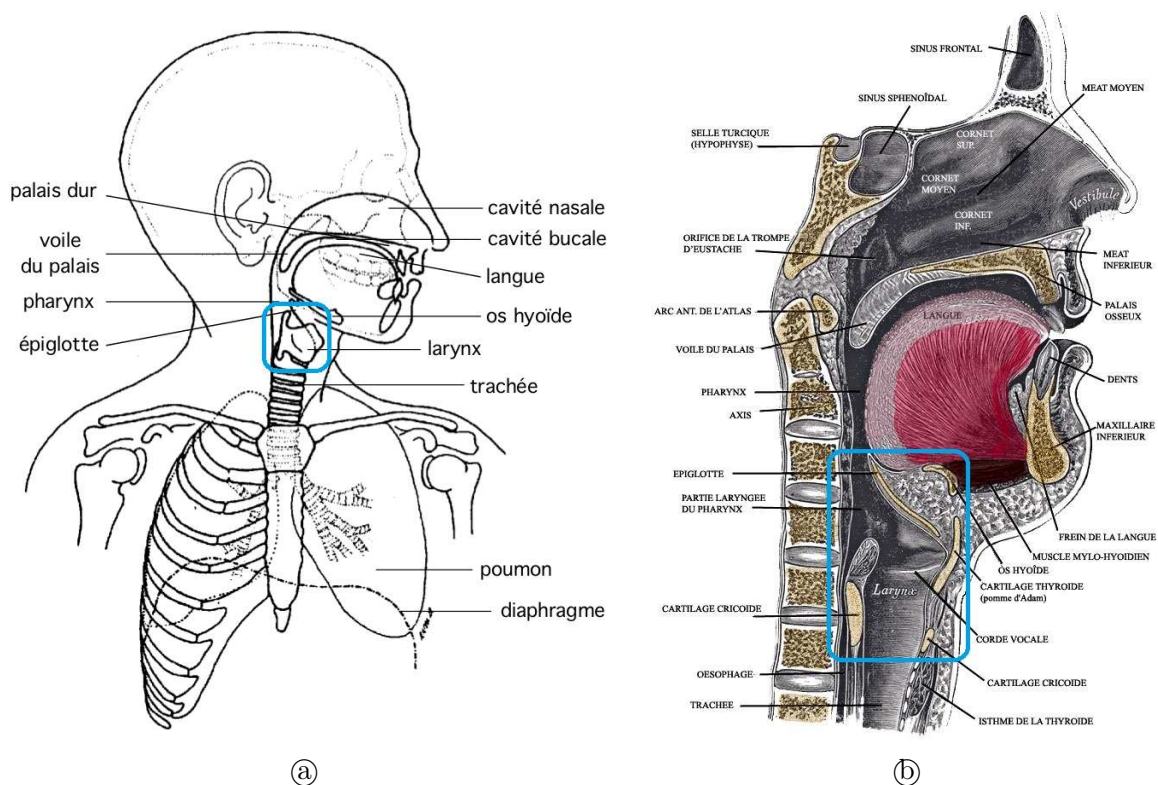


FIGURE 1.1 – Anatomie de l'appareil phonatoire humain, coupe sagittale. (a) d'après [L05]. (b) d'après [GL18]. Les rectangles bleus indiquent la position du larynx qui abrite les plis vocaux.

L'organe clé dans la production de la voix est le larynx, qui abrite les plis vocaux. L'anatomie du larynx est présentée en figure 1.2 et celle des plis vocaux en figure 1.3. Je préfère le terme

1. cf. p.261.

pli vocal au terme *cordes vocales*, plus commun, car il est plus proche de l'anatomie. En effet, comme on peut le voir sur la figure 1.2-①, les plis vocaux (dans le rectangle rouge) sont des bourrelets formant une constriction dans le conduit laryngé et non des cordes vibrantes. Nous désignons par *glotte* ou *ouverture glottique* l'espace situé entre les plis vocaux. Le larynx est composé de plusieurs cartilages, présentés en figure 1.2-②, recouverts de tissus mous. Le cartilage cricoïde est un anneau qui prolonge la trachée. Au cartilage cricoïde sont reliés les cartilages aryténoïdes et le cartilage thyroïde (dont la partie antérieure, proéminente chez l'homme, forme la pomme d'Adam). Le cartilage épiglottique (ou plus simplement l'épiglotte) est relié au cartilage thyroïde et permet, en se rabaissant, de fermer le larynx. Les plis vocaux, dont l'anatomie complexe est présentée en figure 1.3-①, sont attachés d'une part au cartilage thyroïde, à la base de l'épiglotte, et d'autre part aux pointes intérieures des cartilages aryténoïdes (processus vocaux). Les sons de la voix sont produits par vibration des plis vocaux sous l'effet de l'air expulsé par les poumons. Le son résultant de cette vibration résonne dans les cavités buccale et nasale (figure 1.1) et les voyelles sont articulées en modifiant la géométrie, et donc les caractéristiques acoustiques, de ces résonateurs, que nous regroupons sous le vocable *conduit vocal*.

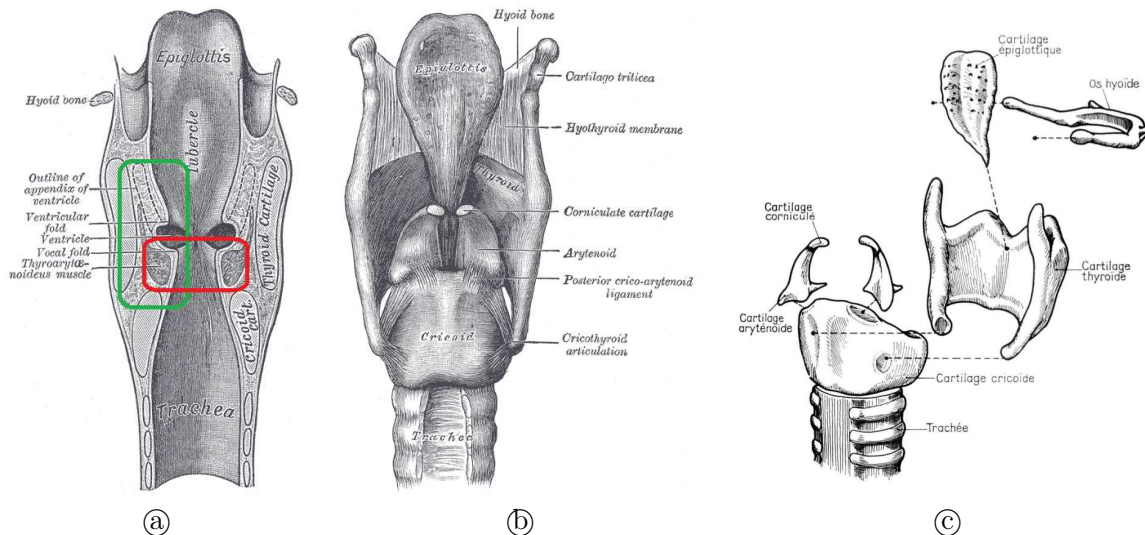


FIGURE 1.2 – Anatomie du larynx (rectangle bleu dans la figure 1.1). ① coupe frontale, d'après [GL18]. La position des plis vocaux est indiquée par le rectangle rouge. L'anatomie d'un pli vocal et d'une bande ventriculaire (rectangle vert) est détaillée en figure 1.3-①. ② cartilages et membranes, d'après [GL18]. ③ vue éclatée des cartilages, d'après [LPV88]. ④ moulage de cartilage thyroïde (sans les cornes supérieures). ⑤ moulage de cartilage cricoïde. ⑥ moulage des cartilages aryténoïdes. Les surfaces d'articulation entre les cartilages aryténoïdes et le cartilage cricoïde sont repérées par des couleurs.

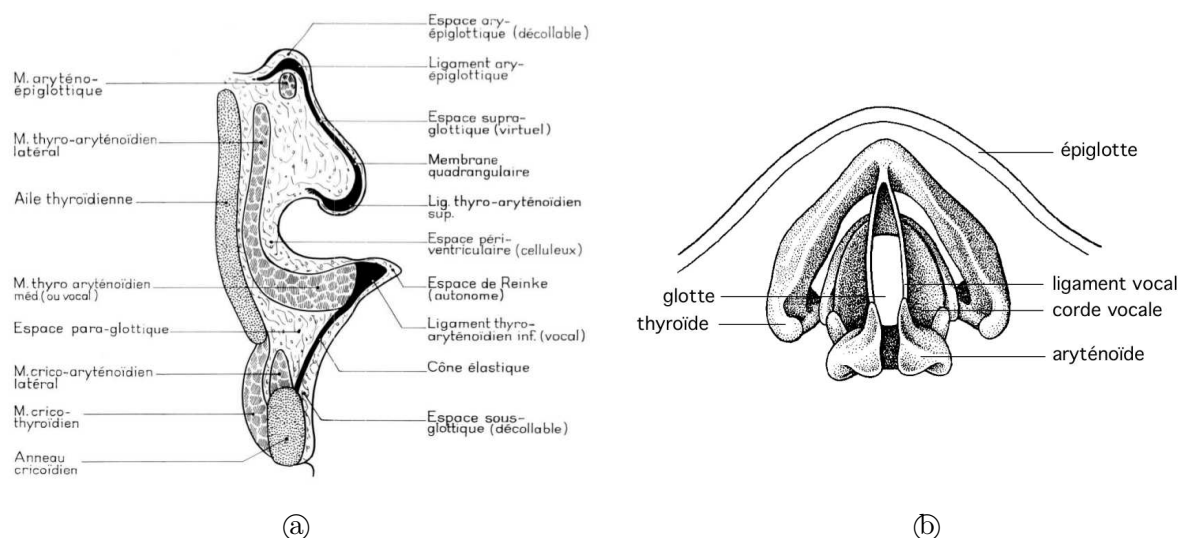


FIGURE 1.3 – Anatomie des plis vocaux. (a) coupe frontale (rectangle vert sur la figure 1.2-(a)), d’après [LPV88]. (b) vue supérieure, d’après [L05].

En dehors de la respiration silencieuse pendant laquelle le larynx est relâché et permet le passage de l’air vers ou depuis les poumons (glotte ouverte et épiglottite relevée), le larynx a deux fonctions. La première est d’assurer un rôle d’obturateur afin d’empêcher les aliments solides et liquides de pénétrer dans la trachée lors de la déglutition. Pour ceci, deux clapets sont mis en jeu. L’épiglottite (figure 1.1, à l’arrière de la langue et figure 1.2 dans la partie supérieure du larynx) s’abaisse pour boucher le larynx et les plis vocaux se mettent en contact afin de fermer l’ouverture glottique.

1.1.2 Auto-oscillation des plis vocaux

La deuxième fonction du larynx, qui nous intéresse ici, est la production de sons. Avant le début de la phonation, les plis vocaux sont mis en contact par translation et rotation des cartilages aryénoïdes (mouvement d’adduction, figure 1.4-(a)-(b)). La vibration des plis vocaux est expliquée par la théorie aérodynamique-myoélastique de Van den Berg [vdB58]. La glotte fermée oppose une résistance au passage de l’air venant des poumons, provoquant une augmentation de la pression sous-glottique jusqu’à forçage de l’ouverture glottique. Une fois la glotte ouverte, la pression sous-glottique retombe et les plis vocaux tendent à se rapprocher sous l’effet de leur élasticité et de l’aspiration provoquée par le passage du flux d’air à travers la glotte (effet de Bernoulli). La glotte se referme et le cycle se répète. On parle alors d’un phénomène d’auto-oscillation, la vibration n’est pas à proprement parler provoquée par une activité musculaire périodique mais par des forces mécaniques et aérodynamiques qui s’appliquent aux plis vocaux. A la fin de la phonation, les plis vocaux repassent en position de respiration par le mouvement d’abduction, illustré sur la figure 1.4-(c).

Les bandes ventriculaires sont les bourrelets situés au-dessus des plis vocaux, ils ont une forme similaire à ces derniers mais ne possèdent pas de muscle interne. Il a été montré que les bandes ventriculaires peuvent avoir un rôle important dans la phonation [Bai09]. Notamment, une mise en vibration des bandes ventriculaires peut intervenir dans certains types de chants comme les chants sardes, les chants sacrés tibétains, ou la gutture du chant métal.

La voix chuchotée est produite en fermant partiellement les plis vocaux afin de créer des turbu-

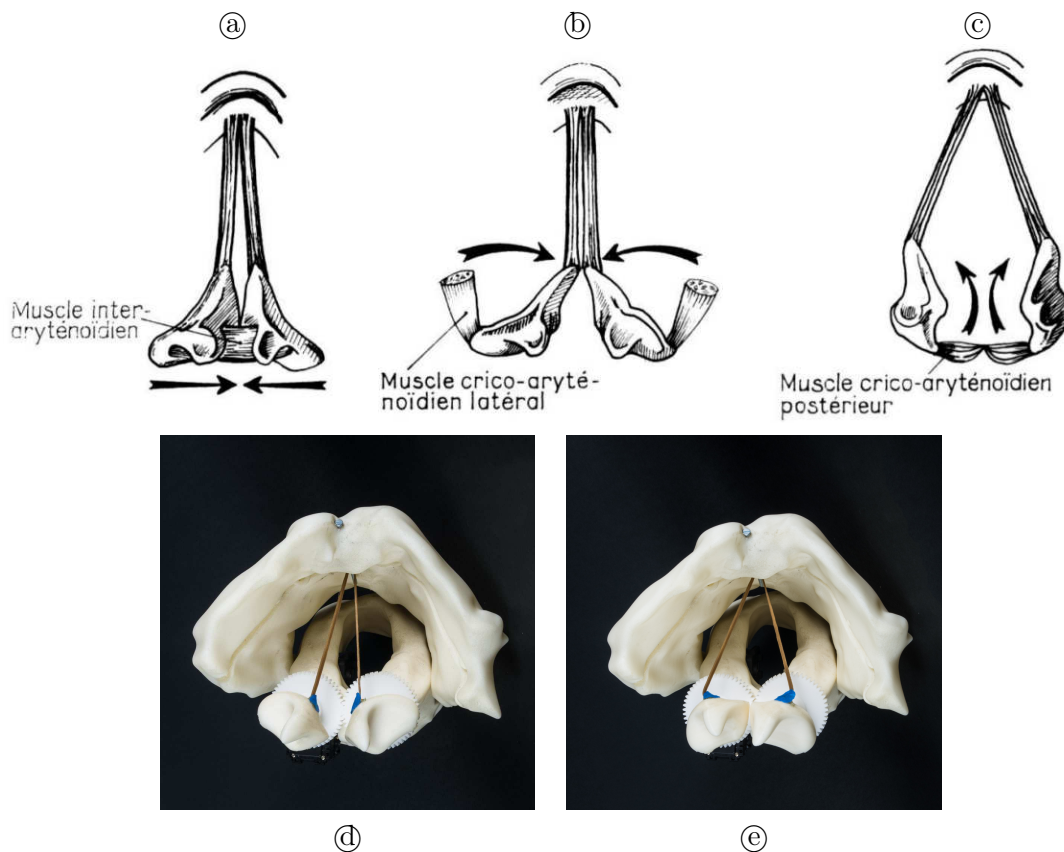


FIGURE 1.4 – Adduction (a) - (b) et abduction (c) des plis vocaux. D'après [LPV88]. (d) - (e) En première approximation, nous pouvons réduire le mouvement d'adduction / abduction à une rotation des cartilages aryténoïdes (ici sur une maquette robotisée à l'échelle 5 :1). Le relief des processus vocaux (points d'accroche du ligament vocal sur les cartilages aryténoïdes) étant peu marqué sur cette maquette, ils sont repérés avec des pastilles bleues.

lences dans le larynx sans mettre les plis vocaux en vibration. Ces turbulences agissent comme des sources acoustiques à spectre large bande et les ondes résultantes sont modulées par les résonateurs de la même façon que les ondes produites par vibration glottique.

1.1.3 Articulations des cartilages et actions musculaires

Si la vibration des plis vocaux n'est pas directement produite par une activité musculaire périodique, elle est en revanche contrôlée par un ensemble de muscles présents dans la région laryngée. Je présente ici les muscles principaux qui sont intéressants pour mon propos, le lecteur intéressé pourra se référer à [GL18, LPV88, L05, Fin08] pour plus de précisions à ce sujet.

En premier lieu, des muscles sont nécessaires pour l'adduction et l'abduction des plis vocaux. Ces muscles relient les aryténoïdes entre eux et au cricoïde et sont présentés en figures 1.4.

Lors de la phonation, deux types d'action permettent de modifier les propriétés mécaniques des plis vocaux et ainsi d'influer sur la fréquence fondamentale de la vibration glottique et ses caractéristiques spectrales. Le muscle thyro-aryténoïdien interne, aussi appelé muscle vocal ou muscle *vocalis* constitue le cœur des plis vocaux (figure 1.3 - (a)). En se contractant, ce muscle provoque

un rétrécissement et une rigidification des plis vocaux. A l'inverse, les muscles thyro-aryténoïdiens externes permettent une élongation et un amincissement des plis vocaux par rotation du cartilage thyroïde autour de l'articulation crico-thyroïdienne (figure 1.5).



FIGURE 1.5 – Articulazione crico-tiroideia. (a) d'après [LPV88]. (b)-(c) dimostrazione su una maquette robotizzata à l'échelle 5 :1. L'amplitude de rotation a été ici volontairement exagérée pour illustrer l'articulation.

Enfin, des muscles pilotent la translation verticale du larynx, ce qui permet essentiellement d'agrandir ou de raccourcir le conduit vocal et donc de modifier son comportement acoustique.

1.1.4 Contrôle des résonateurs

Afin de moduler les sons produits par l'activité glottique, l'humain possède la capacité de modifier la géométrie des résonateurs acoustiques que sont le pharynx, la cavité buccale, la cavité nasale et les sinus paranasaux (figure 1.1). La modification de la géométrie du conduit vocal entraîne, entre autres, une modification de ses fréquences de résonances. La position fréquentielle de ces résonances (on parle de "formant" pour la voix) permet de différencier les voyelles prononcées.

Tout d'abord, le voile du palais (aussi appelé *velum*) permet d'ouvrir et de fermer les cavités nasales. Outre encore une fois l'évitement du passage d'aliments lors de la déglutition, cette ouverture/fermeture permet de produire les voyelles dites nasales. Il a été montré que l'ouverture des fosses nasales et des sinus paranasaux provoquent l'apparition d'anti-résonances dans la fonction de transmittance acoustique du conduit vocal [Mae82b, NKvS05].

Dans la bouche, la langue possède une grande mobilité et permet de moduler avec finesse la forme de la cavité buccale. Enfin, l'ouverture de la mâchoire et la position des lèvres permettent elles aussi de modifier le comportement acoustique du conduit vocal.

1.2 Voix parlée, voix chantée et mécanismes laryngés

1.2.1 Voix parlée et voix chantée

Nous pouvons différencier deux utilisations principales de la voix : la parole et le chant. Ces deux utilisations présentent des différences dans le contrôle de la vibration laryngée. Nous pouvons notamment noter les points suivants.

- La fréquence fondamentale est contrôlée très finement et reste stable dans le cas de la voix chantée (mélodie) alors qu'elle est plus libre et varie en général plus rapidement dans le cas de la voix parlée (prosodie).
- L'ambitus entre la fréquence fondamentale (la note) la plus élevée et la plus basse est en général beaucoup plus important pour le chant que pour la parole.
- Les variations d'intensité sont plus importantes dans la voix chantée (de *pp* à *ff*).
- La rapport entre la durée totale des parties voisées (plis vocaux en vibration) et non voisées (sons turbulents) est beaucoup plus important pour le chant que pour la parole. De ce fait, le rythme respiratoire est modifié. Les inspirations sont raccourcies tandis les expirations sont considérablement allongées et contrôlées pour être les plus régulières possible.

1.2.2 Notion de qualité vocale

Outre l'aspect voisé/non voisé (plis vocaux en vibration ou non), la fréquence fondamentale et l'intensité sonore, de nombreux autres paramètres peuvent être contrôlés lors de la phonation. L'étude de la qualité vocale [HBE⁺08] vise à définir de façon perceptive les sons vocaux (en parole ou en chant) en liaison avec les caractéristiques physiologiques de la phonation, notamment de l'activité laryngée. On parle par exemple de quantité de souffle dans la voix ("breathiness" en anglais, liés à des fermetures glottiques non complètes, on parle parfois de "voix blanches" en français), rugosité, effort vocal (voix criée), efficacité vocale, voix tendue/relâchée etc. Ces paramètres perceptifs sont à mettre en lien avec les contrôles mécaniques et acoustiques de l'appareil phonatoire.

1.2.3 Mécanismes laryngés

Une notion importante, qui fait le lien entre évaluation perceptive et fonctionnement physiologique est la notion de mécanisme laryngé, bien que les chanteurs parlent plus volontiers de registres et que la littérature abondante n'est pas toujours unanime sur les définitions et propriétés de ces mécanismes (voir [Lam10, RHC09, Hen06, Hen01, Rou93] pour une revue complète sur le sujet). La littérature nous enseigne l'existence de quatre mécanismes de vibration laryngée.

Mécanisme 0 ou voix de *fry*. Les plis vocaux sont courts, épais et peu tendus. La durée d'ouverture des plis vocaux est faible comparée à la période. Ce mécanisme permet de réaliser des sons aux fréquences les plus graves.

Mécanisme 1. Les plis vocaux sont épais et vibrent sur toute leur longueur. La masse et l'amplitude de vibration sont importantes et le muscle vocal est entraîné dans le mouvement vibratoire. Le ratio entre la durée d'ouverture et la période est en général plus élevé qu'en mécanisme 0 mais reste inférieur à 0,5. C'est le mécanisme privilégié de la voix parlée d'homme, des voix chantées des basses, barytons, ténors et altos féminins et des voix chantées masculines et féminines dans la variété.

Mécanisme 2. Le mécanisme 2 est caractérisé par le découplage du muscle vocal de la vibration des plis vocaux [LSGR94]. Les plis vocaux sont fins et ne vibrent plus que sur les 2/3 de leur longueur, en raison d'une compression accentuée des cartilages aryténoïdes (extrémités antérieure des plis vocaux). Le ratio entre la durée d'ouverture et la période est en général supérieur à 0,5. Le mécanisme 2 est le mécanisme privilégié de la voix parlée de femme, tandis que l'homme l'utilise occasionnellement pour émettre des sons aigus. Les chanteuses mezzos-sopranos et sopranos ainsi que les chanteurs altos et haute-contre utilisent presque

exclusivement ce mécanisme. Le mécanisme 2 est plus utilisé en voix parlée par les hommes dans d'autres cultures (Asie, Afrique).

Mécanisme 3 ou voix de *sifflet*. Les plis vocaux sont très fins, allongés et tendus et l'amplitude de vibration est très faible. La durée de fermeture complète est quasi-nulle.

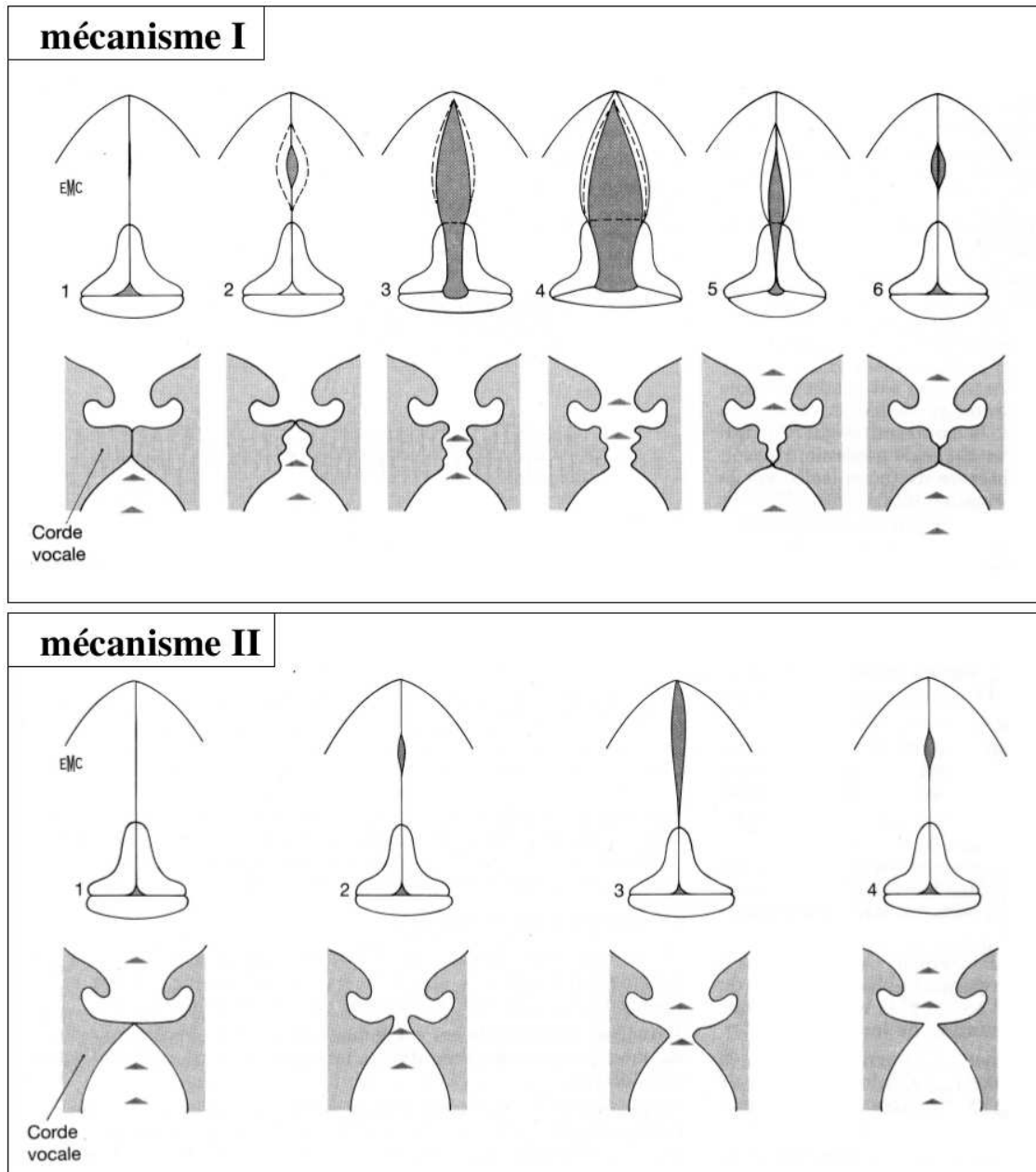


FIGURE 1.6 – Mécanismes laryngés 1 et 2. D'après [Hen01], adapté de [Ven67].

Les mécanismes 1 et 2 sont très majoritairement utilisés, en voix parlée comme en voix chantée. Un cycle de vibration glottique de chacun de ces mécanismes est présenté en figure 1.6. Sont schématisées une vue supérieure, comme sur les figures 1.3-Ⓣ et 1.4, ainsi qu'une vue en coupe frontale, comme sur les figures 1.2-ⓐ et 1.3-ⓐ. Nous pouvons observer sur cette figure le mouvement dans l'épaisseur des plis vocaux dans le cas du mécanisme 1. La surpression sous-glottique provoque tout d'abord un décollement du bas vers le haut, puis les forces aérodynamiques entraînent un acco-

lement des plis vocaux du bas vers le haut. Dans la phase d'ouverture comme dans la phase de fermeture, on observe un déroulement des plis vocaux dans leur épaisseur. Ce déroulement n'est pas présent dans le cas du mécanisme 2, en raison de la finesse et de la rigidité des plis vocaux. De plus, cette figure illustre l'action des aryténoïdes, qui participent à la vibration en mécanisme 1 et qui n'y participent pas en mécanisme 2, ainsi que la différence de longueur vibrante qui en résulte.

1.3 Méthodes exploratoires

1.3.1 Visualisation des plis vocaux en vibration

De par sa position, sa sensibilité et sa fragilité, le larynx est un organe qu'il est difficile d'explorer *in vivo*. Nous devons les premières visualisations d'un larynx en vibration au chanteur et professeur de chant Manuel Garcia qui réussit à observer les plis vocaux en vibration à l'aide d'un petit miroir fixé au bout d'un tige et inséré dans la bouche d'un chanteur (une source lumineuse est renvoyée vers le larynx à l'aide d'un jeu de miroirs) [Gar54, Gar94], voir figure 1.7. Il appela son instrument le *laryngoscope*, qui fut l'objet de nombreuses améliorations visant à observer le plus précisément possible les mouvements des plis vocaux [BO00, Rad05].

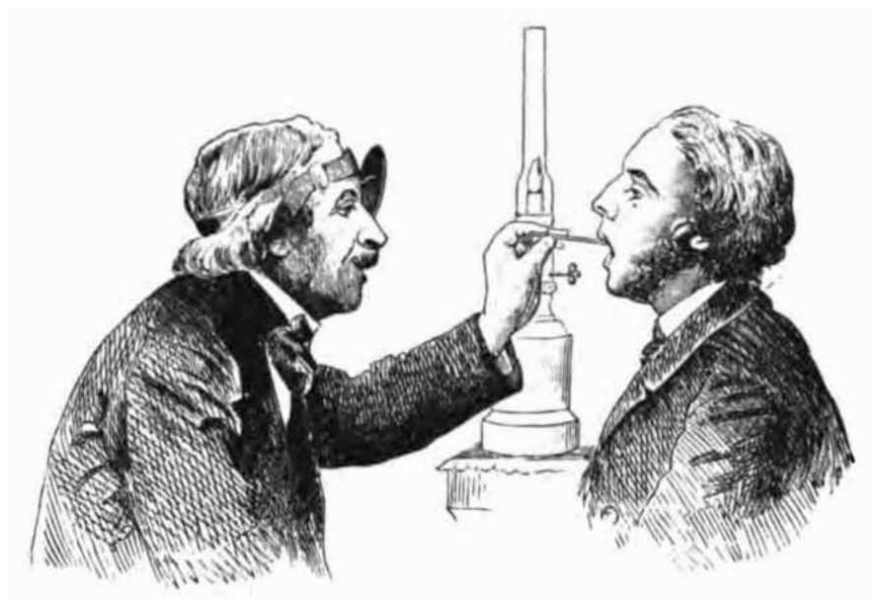


FIGURE 1.7 – Laryngoscope de Manuel Garcia, d'après [Gar94].

Trois techniques principales sont issues de la laryngoscopie de Garcia.

Laryngoscopie stroboscopique. Le mouvement des plis vocaux est ralenti en l'observant (ou en l'éclairant) de façon discontinue et périodique, à une période proche de la période fondamentale de vibration. Cela permet de reconstruire un cycle de vibration glottique avec une période visible par l'œil humain, à partir d'échantillons provenant de plusieurs périodes consécutives, dans le cas où la vibration est relativement stable. Les transitoires et mouvements rapides ne peuvent pas être observés par cette technique.

Enregistrement vidéo haute vitesse. L'enregistrement vidéo haute-vitesse des plis vocaux (initié par des travaux aux Bell Laboratories [HF38, Far40]) permet, en enregistrant plusieurs

dizaines d'images par cycle et en les repassant à vitesse lente, d'observer finement le détail de chaque cycle de vibration.

Vidéokymographie. La vidéokymographie, appliquée pour la première fois à la voix par Jan Švec [Š00], propose d'observer une seule ligne du plan glottique. Cette technique permet d'accéder à des hautes fréquences d'échantillonnage avec des caméras classiques. Elle permet de plus une représentation 2D d'un enregistrement vidéo.

Ces techniques présentent le désavantage d'une observation partielle de la vibration laryngée. Toutes sont fondées sur une observation des plis vocaux depuis le pharynx. Par conséquent, aucune information sur le mouvement dans l'épaisseur des plis vocaux n'est accessible. Nous reparlerons de ce problème dans ce document.

1.3.2 Méthodes indirectes d'observation de l'activité glottique

Une des méthodes les plus utilisées pour l'étude la vibration glottique est l'**électroglottographie**, ou EGG. Cette méthode, initialement proposée par Fabre en 1957 [Fab57] consiste à estimer la conductance du larynx en faisant passer un courant haute fréquence à travers le larynx et en mesurant la tension résultante. Nous reviendrons en détail sur cette technique dans la partie I.

Enfin, des techniques permettent de mesurer directement la pression sous-glottique [KL75, SH81, PH98]. Cependant, ces techniques sont invasives et risquées, et nécessitent d'être effectuées en milieu hospitalier sous le contrôle de médecins.

1.3.3 Corpus de données multi-modales

Nous disposons à l'IRCAM du corpus de données *USC_2008_02* regroupant des mesures effectuées pendant la phonation de voyelles tenues sur un ensemble de participants par Gilles Degottex et Erkki Bianco [DBR08b]. Ces mesures contiennent des enregistrements simultanés

- du son produit pris avec un microphone placé devant les lèvres (à une distance non précisée),
- du signal EGG,
- d'une vidéo-endoscopie ultra-rapide des plis vocaux à 4000 images par seconde.

Ce corpus est constitué de 126 mesures effectuées sur des hommes et des femmes et présente une variété intéressante en termes de mécanismes laryngés et de qualité vocale. Cependant, la fréquence d'échantillonnage vidéo ne permet d'explorer les mouvements laryngés qu'avec une bande passante de 2 kHz. Seules les mesures faites sur des sons dont la fréquence fondamentale est en-dessous de 500 Hz présentent donc des vidéos exploitables. Ce corpus représente la principale source de données dont je dispose et il y sera régulièrement fait référence tout au long de ce document.

1.3.4 Estimation de la géométrie et des propriétés acoustiques du conduit vocal

L'estimation des propriétés géométriques et acoustiques du conduit vocal à partir du son est un domaine d'étude à part entière et se rapproche de l'analyse-synthèse du signal vocal. Des méthodes mises en œuvre pour ce problème dans le cadre du traitement du signal de parole seront présentées dans le chapitre 7. Je présente ici les méthodes permettant d'accéder directement aux mouvements articulatoires.

Les interfaces de parole silencieuse (*silent speech interfaces*) ont pour but de générer des sons de parole intelligibles à partir des mouvements des articulateurs effectués sans émission de son.

[DSH⁺10] présente une revue complète et historique des méthodes et techniques mises en jeu pour résoudre ce problème. Les interfaces de parole silencieuse ont nécessité le développement de méthodes d'estimation des mouvements des articulateurs qui nous intéressent ici. Plusieurs stratégies ont été mises en place :

- captation des mouvements de points fixes sur les articulateurs à l'aide d'articulographes électromagnétiques [FEG⁺08],
- captation des mouvements de la langue et des lèvres par ultrasons et caméra [HBD12],
- utilisation des microphones NAM (Non-Audible Murmur) [TBLT10],
- électromyographie de surface (sEMG) sur les muscles articulateurs [SW10].

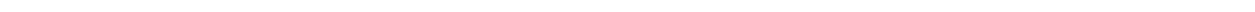
Nous pouvons aussi citer les méthodes de visualisation des mouvements des articulateurs par imagerie fonctionnelle, notamment à l'aide de rayons X [SHL⁺11] et d'imagerie par résonance magnétique (les travaux fondateurs de Rokkaku et al. [RHI⁺86] et ses suites).

A ces méthodes d'estimation de la position des articulateurs s'ajoutent les méthodes d'estimation des propriétés acoustiques du conduit vocal. Nous verrons au chapitre 7 que la LPC (Linear Predictive Coding) est une méthode historique permettant une estimation efficace de la géométrie du conduit vocal à partir d'un extrait de son de parole. Kob et Neuschaefer-Rube proposent dans [KNR02] une méthode originale permettant une mesure des résonances du conduit vocal. L'impédance du conduit vocal est mesurée au niveau de la bouche pour des voyelles tenues. Pour ceci, un haut-parleur permet de générer un sweep au niveau des lèvres afin de balayer la plage de fréquences étudiée et le couple pression/vitesse résultant aux lèvres est mesuré avec un équipement adéquat. Cette méthode permet une évaluation acoustique précise et fiable de la localisation des résonances du conduit vocal. Cependant, la mesure n'est pas instantanée et nécessite donc que le sujet maintienne une configuration la plus stable possible.

1.4 Conclusion

Nous avons vu dans ce chapitre que la physiologie de l'appareil phonatoire humain est complexe. Le contrôle de la vibration glottique et des articulateurs met en jeu un grand nombre de muscles. Le rôle exact de chacun des muscles de la région laryngée et les conséquences de leur activité sur la vibration glottique et les sons émis ne sont pas encore totalement compris. De plus, ces mouvements sont particulièrement difficiles à mesurer *in vivo*, en raison de la position du larynx. Des méthodes existent pour observer ou estimer les mouvements du larynx et des articulateurs mais chacune présente des inconvénients. La seule méthode permettant aujourd'hui la visualisation directe des plis vocaux en vibration est la vidéo-endoscopie ultra-rapide mais elle est relativement invasive, même si les endoscopes souples actuels permettent à des chanteurs entraînés de chanter de façon naturelle malgré leur présence. De plus, les fréquences d'échantillonnage des vidéo-endoscopes restent encore basses comparées aux fréquences mises en jeu par la vibration glottique.

La première partie de mon étude sera donc consacrée à l'étude du système phonatoire et aux méthodes d'exploration de l'activité glottique. Dans la chapitre 3, nous passerons en revue ces méthodes dans le détail et nous verrons des exemples de données extraites du corpus USC_2008_02 ou de mesures que j'ai moi-même réalisées. Le chapitre 4 sera consacré à une étude comparée du comportement vibratoire de la glotte en phonation et des lèvres de joueurs de cuivre pendant le jeu à partir d'une méthode de mesure et d'analyse communes. Nous verrons en effet dans le chapitre suivant que cuivres et voix tissent entre eux des liens privilégiés. Enfin, le chapitre 5 sera consacré à la construction d'un nouveau dispositif pour la mesure de l'activité glottique, visant à pallier les limites des appareils actuels en terme de résolution temporelle et spatiale.



Chapitre 2

Différents modèles pour l'analyse et la synthèse sonores de la voix

Sommaire

2.1	Signal vocal et quasi-stationnarité	26
2.2	Modèles physiques et modèles de signaux	26
2.2.1	Modèle source-filtre	28
2.2.2	Modèles physiques simples	29
2.2.3	Différents choix pour la modélisation de la phonation	30
2.3	Cuivres et voix, des liens privilégiés	31
2.4	Position du problème et organisation du document	31

2.1 Signal vocal et quasi-stationnarité

Le son vocal émis par l'appareil phonatoire que nous venons de décrire, est, dans le cas voisé, un signal quasi-périodique dont la fréquence fondamentale varie en fonction du temps. Un extrait d'enregistrement de voix parlée est présenté en figure 2.1. Le spectrogramme correspond à une phrase entière, faisant apparaître des zones voisées (fréquence fondamentale identifiable) et des zones non voisées (spectre continu). Le signal temporel présente un extrait de cette phrase correspondant à un segment voisé.

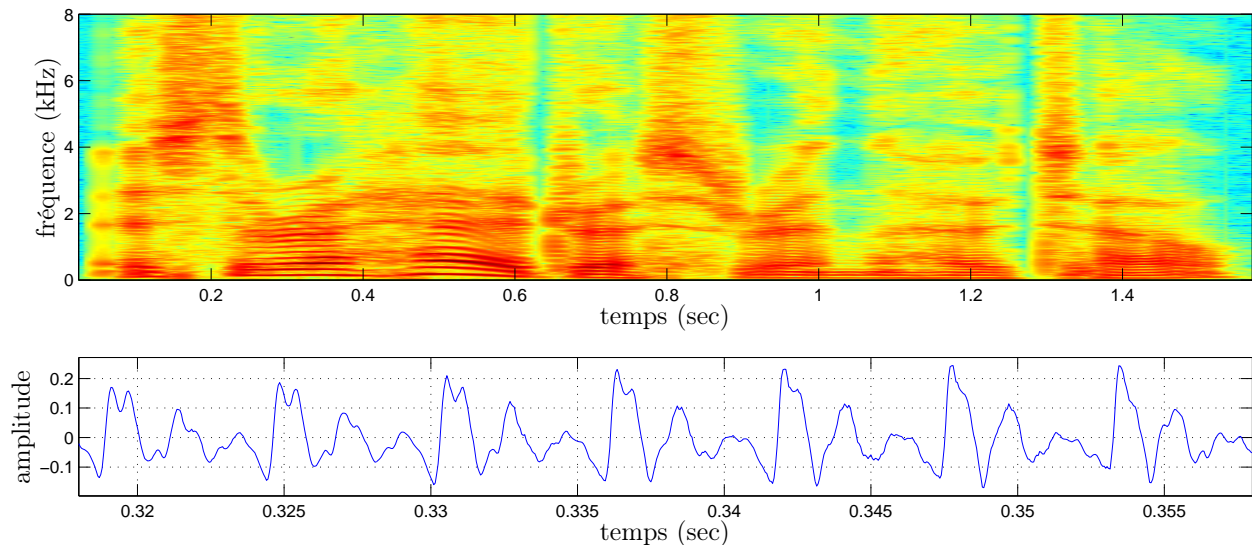


FIGURE 2.1 – Exemple de signal vocal (voix parlée). En haut : spectrogramme de la phrase anglaise “Is there a hairdresser in the hotel?” prononcée par un locuteur masculin (voix parlée, mécanisme 1). En bas : extrait du signal temporel.

En traitement de la parole, il est généralement admis que le signal est stationnaire sur une fenêtre d'environ 15 à 20 ms. Cela revient à dire que le mouvement des muscles laryngés et des articulateurs est négligeable sur cette durée. Cette hypothèse de quasi-stationnarité a été abondamment discutée dans la littérature [MG76, CT89, RS11] et sera reprise dans les travaux présentés ici.

2.2 Modèles physiques et modèles de signaux

Les deux grandes approches développées historiquement pour modéliser la production de la voix humaine sont les **modèles de signaux** et les **modèles physiques**. Les modèles de signaux permettent en général une implémentation faible coût et des méthodes d'inversion (recherche des paramètres du modèle à partir d'un son cible) efficaces. En revanche, leur réalisme est limité. De leur côté, les modèles physiques offrent en général plus de réalisme mais l'implémentation de ces modèles est plus coûteuse et leur inversion plus délicate. Une classification des principaux modèles pour l'analyse-synthèse de la voix que l'on trouve dans la littérature est proposée en figure 2.2 selon un axe allant des modèles de signaux aux modèles physiques. Nous regroupons les modèles selon cinq grandes familles : (i) la synthèse concaténative, (ii) la synthèse additive, (iii) les modèles source-filtre, (iv) les modèles physiques simples, (v) les modèles physiques complexes.

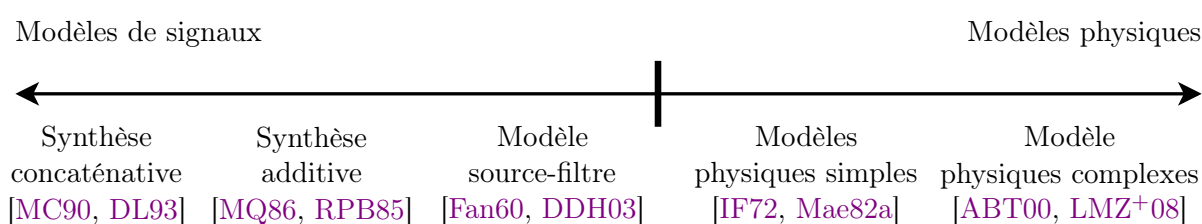


FIGURE 2.2 – Classification des modèles de production de la voix.

Je présente ici une revue succincte de ces modèles. Dans mes travaux, je me propose de travailler sur des modèles à l’interface entre les modèles de signaux et les modèles physiques, espérant tirer le meilleur des deux approches, à savoir introduire une partie du réalisme des modèles physiques et préserver le faible coût et l’irréversibilité des modèles de signaux. Ainsi, les travaux présentés dans cet ouvrage se situent à l’interface entre les modèles source-filtre (présentés en section 2.2.1) et les modèles physiques simples (présentés en section 2.2.2).

Dans la classification proposée, les synthèses concaténative et additive se positionnent à part du reste des modèles. La synthèse concaténative est fondée sur la concaténation de diphtongues pré-enregistrés [CS86, CM88], et non sur la synthèse de signaux à l’aide de modèles génératifs. Cette technique a inspiré de nombreuses applications, pour la synthèse ou la modification de la parole, notamment la méthode PSOLA et ses déclinaisons TD-PSOLA, FD-PSOLA, LP-PSOLA et MBR-PSOLA [MC90, VET92, DL93]. La synthèse additive vise quant à elle à reproduire le spectre d’un signal de parole grâce à des représentations sinusoïdales ou formantiques [MQ86, RPB85].

Le modèle source-filtre et les modèles physiques sont souvent présentés comme antagonistes. Les modèles source-filtre, que nous étudierons en détail dans le chapitre 6, s’appliquent à modéliser les caractéristiques temporelles et fréquentielles du signal vocal. Les modèles physiques, en revanche, s’intéressent à l’activité laryngée et à la configuration du conduit vocal à l’origine de la voix. Les modèles de signaux s’intéressent donc aux conséquences de la phonation (le son) et les modèles physiques aux causes (vibrations laryngées et mouvements articulatoires).

Cependant, le paradigme source-filtre et le paradigme modèle physique peuvent être considérés comme différents degrés de raffinement dans la modélisation de la phonation. Comme nous l’avons vu dans le chapitre précédent, la vibration glottique est le résultat d’un couplage entre les forces aérodynamiques et mécaniques qui s’exercent sur les plis vocaux (solides mous). Le conduit vocal agit ensuite sur ces derniers comme une charge acoustique, avec ses résonances et anti-résonances, et interagit donc avec l’acoustique glottique. Une modélisation complète de la production de la voix implique une résolution des équations aérodynamiques de Navier-Stokes depuis l’intérieur des poumons jusqu’aux rayonnements aux lèvres et aux narines en prenant en compte les interactions mécaniques avec

- les plis vocaux,
- les poumons, la trachée, le conduit laryngé et le conduit vocal (qui ne constituent pas un conduit rigide, ils possèdent eux aussi des propriétés mécaniques non triviales et leur dynamique est couplée à l’acoustique interne), et
- les interactions acoustiques avec le milieu extérieur (rayonnement).

Depuis le début des années 2000, plusieurs travaux ont été entrepris afin d’aborder la résolution de ce problème avec divers degrés de simplification [ABT00, Gun03, dORPGA03, LMZ+08]. Ils correspondent aux modèles physiques complexes dans la classification ci-dessus. Ces modèles, principalement en raison du coût de calcul élevé que représente leur simulation, ne sont pas encore utilisables en tant que modèle de synthèse.

2.2.1 Modèle source-filtre

Le modèle source-filtre propose de considérer le signal vocal comme résultant du filtrage d'un signal de source [Fan60]. Autrement dit, dans ce paradigme, la glotte est considérée comme une source idéale et le conduit vocal est réduit à un filtre linéaire. Cette modélisation correspond à une simplification drastique du fonctionnement de l'appareil phonatoire humain pour laquelle (nous détaillerons ces analogies dans les chapitres 6 et 9)

- la source glottique ne subit aucune influence du conduit vocal, elle est considérée comme une source de débit idéale et est entièrement représentée par un signal temporel $g(t)$,
- le conduit vocal est considéré comme une concaténation de tubes droits dans lesquelles les ondes acoustiques se propagent en ondes planes sans pertes, dont la modélisation aboutit à un filtre linéaire que nous appellerons *filtre vocal* dans la suite de cet ouvrage,
- le rayonnement aux lèvres est modélisé par une simple dérivation temporelle.

Le modèle source-filtre s'exprime finalement comme un signal source filtré, selon le schéma présenté en figure 2.3 : le signal de source glottique est filtré par le filtre vocal pour obtenir le signal aux lèvres, qui est à son tour filtré par le filtre de rayonnement pour obtenir le signal émis.

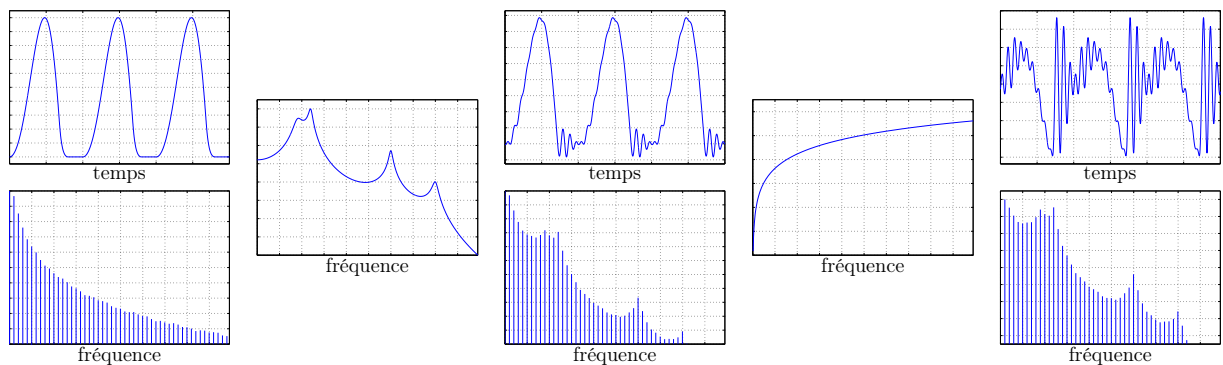


FIGURE 2.3 – Modèle source-filtre général. De gauche à droite : source glottique (temporel et fréquentiel), filtre vocal (fréquentiel), signal aux lèvres (temporel et fréquentiel), filtre représentant le rayonnement (fréquentiel), son rayonné (temporel et fréquentiel).

Étant données la linéarité et la stationnarité (locale) du système ainsi construit, les filtres peuvent être permutés sans conséquence sur le signal de sortie. Il est alors habituel de combiner le signal de source et le filtre de rayonnement (filtre dérivateur) pour obtenir un signal de source dérivé et ainsi simplifier le système, comme présenté dans la figure 2.4 : le signal de source glottique dérivé est directement filtré par le filtre vocal pour obtenir le signal émis.

Une variante du paradigme source-filtre a été introduite par Doval et al. dans [DDH03] avec le modèle CALM (Causal-Anticausal Linear Model) de la source glottique. La source glottique dérivée n'est plus ici générée par une fonction paramétrique périodique comme c'est habituellement le cas, mais par filtrage d'un train d'impulsions de Dirac. Ce filtre est appelé *filtre glottique* et le système de la figure 2.4 est modifié pour obtenir celui de la figure 2.5 : le train d'impulsions périodiques est filtré par le filtre glottique pour obtenir le signal de source dérivé, qui est à son tour filtré par le filtre vocal pour obtenir le signal émis. Cette version du modèle source-filtre permet donc de modéliser le signal vocal à l'aide d'un unique filtre linéaire excité par un train d'impulsions périodiques.

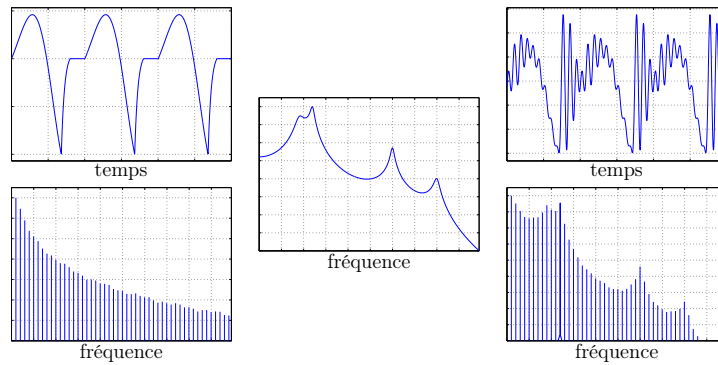


FIGURE 2.4 – Modèle source-filtre simplifié. De gauche à droite : source glottique dérivée (temporel et fréquentiel), filtre vocal (fréquentiel), son rayonné (temporel et fréquentiel).

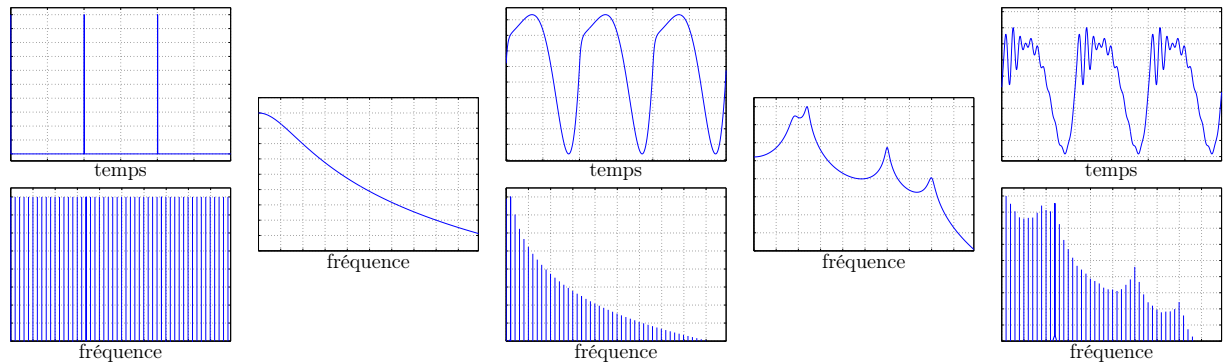


FIGURE 2.5 – Modèle source-filtre, CALM. De gauche à droite : train d’impulsion périodique (temporel et fréquentiel), filtre glottique (fréquentiel), source glottique dérivée (temporel et fréquentiel), filtre vocal (fréquentiel), son rayonné (temporel et fréquentiel).

2.2.2 Modèles physiques simples

Les modèles que je regroupe sous l’appellation “modèles physiques simples” rétablissent l’interaction entre la source glottique et le filtre vocal [Fla68, Rot81a, FL87] sans aller jusqu’aux raffinements des modèles physiques complexes déjà évoqués. Deux approches sont envisagées pour prendre en compte l’influence du conduit vocal sur l’activité glottique.

La première consiste à utiliser des modèles mécaniques auto-oscillants de plis vocaux. Plusieurs modèles ont été proposés (voir par exemple [Vil02, Rut07, Cis08, Bai09]) parmi lesquels nous pouvons notamment citer le modèle à deux masses d’Ishizaka et Flanagan [IF72] qui a été utilisé pour de nombreuses applications en synthèse et analyse de la voix. Dans ce modèle, les plis vocaux sont représentés par deux masses parallélépipédiques reliées entre elles par un ressort et reliées chacune au reste du corps par un ressort et un amortisseur, comme présenté dans la figure 2.6-Ⓐ. Les équations mécaniques, aérodynamiques et acoustiques régissant les différents composants du système complet permettent de construire l’équivalent électrique du système présenté en figure 2.6-Ⓑ, plus simple à implémenter sous forme électronique ou en simulation numérique.

La deuxième approche constitue une étape intermédiaire entre le modèle source-filtre et les modèles auto-oscillants : on impose la géométrie du conduit laryngé (qui varie au cours du temps), la pression sous-glottique et la géométrie du conduit vocal. Nous appellerons ces modèles les modèles

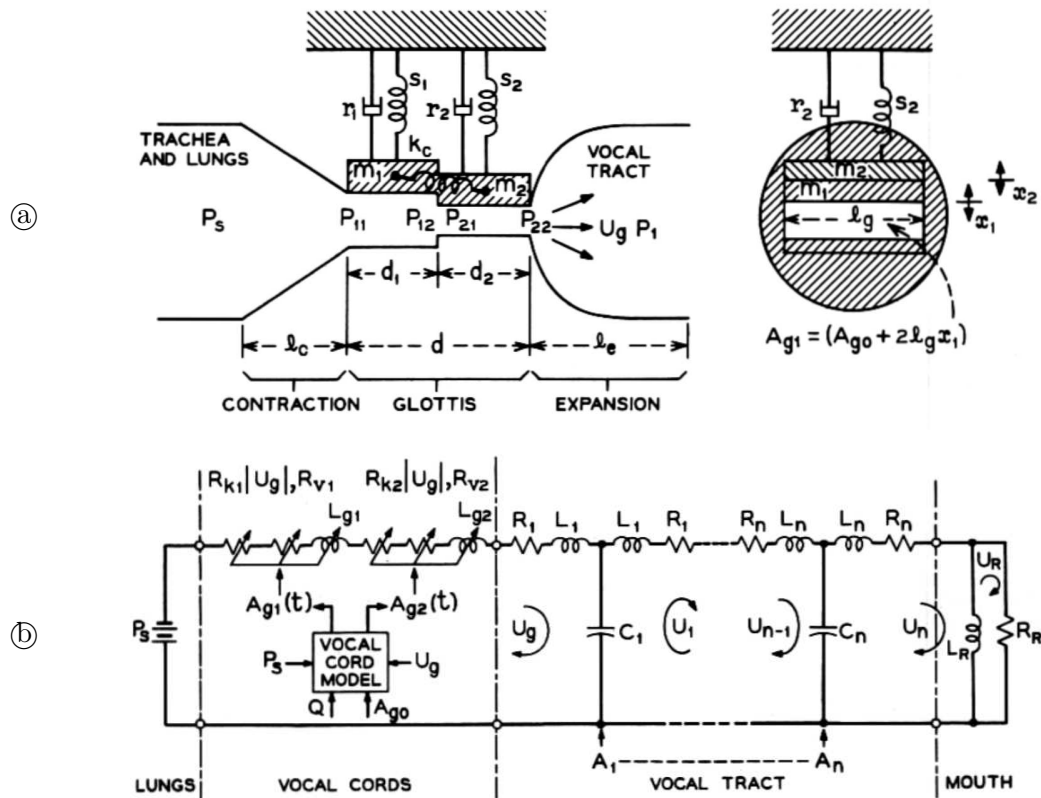


FIGURE 2.6 – Modèle à deux masses d'Ishizaka et Flanagan, d'après [IF72]. (a) modèle mécanique, (b) équivalent électrique.

à géométrie forcée. Les grandeurs acoustiques (pression et débit glottique, pression et débit dans le conduit vocal, pression et débit rayonné par les lèvres et les narines) sont calculées à partir de ces données de commande du modèle et des équations régissant l'acoustique à l'intérieur des conduits. Plusieurs modèles ont été développés suivant l'approche à géométrie forcée [AF82, AS85, GC86, TS97], parmi lesquels celui de Maeda [Mae82a] reste un de ceux les plus utilisés. Ce modèle est présenté en figure 2.7.

2.2.3 Différents choix pour la modélisation de la phonation

Les trois approches que je viens de présenter (modèle source-filtre, modèle à géométrie forcée et modèle auto-oscillant) correspondent à une évolution dans les degrés de liberté de modèle de glotte :

- C1.** Modèles **source-filtre** : la source acoustique est forcée.
- C2.** Modèles à **géométrie forcée** : la géométrie pilote un système aéro-acoustique.
- C3.** Modèles **auto-oscillants** : l'oscillation du système vibro-acoustique non linéaire est pilotée par les paramètres mécaniques mais n'est pas forcée.

Les travaux présentés dans ce document suivent cette évolution. La partie II est consacrée au choix **C1** afin de construire des outils d'analyse et de séparation source-filtre dans un cadre linéaire. Nous passerons ensuite au choix **C2** dans la partie III afin de rétablir l'interaction et étudier son intérêt pour la synthèse.

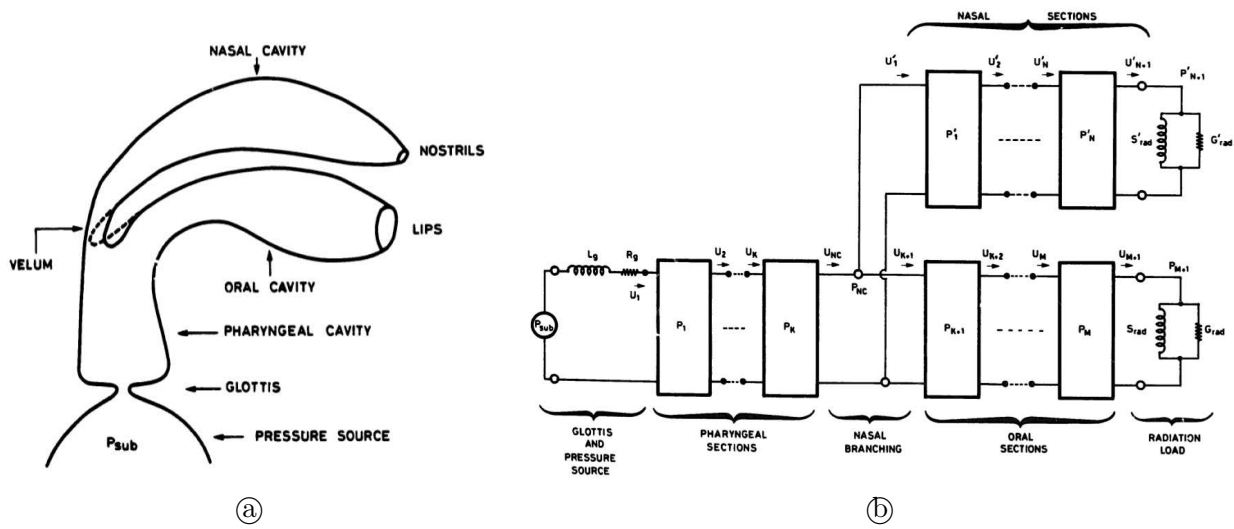


FIGURE 2.7 – Modèle de Maeda, d’après [Mae82a]. (a) conduit à simuler, (b) équivalent électrique.

2.3 Cuivres et voix, des liens privilégiés

Nous portons ici notre attention sur une remarque singulière. L’appareil phonatoire humain et les instruments de musique de la famille des cuivres sont des systèmes qui partagent des points communs forts dans leur fonctionnement physique. Ils sont tous deux constitués d’une “valve humaine” (constriction dans un tube) auto-oscillante couplée à un résonateur acoustique. Aux plis vocaux qui oscillent sous l’effet de la pression sous-glottique correspondent les lèvres du musicien qui oscillent, selon le même principe que les plis vocaux, sous l’effet de la pression dans la bouche. Au conduit vocal correspond le corps de l’instrument. Une étude de la littérature fait d’ailleurs apparaître que des modèles similaires sont utilisés pour simuler le comportement de ces deux systèmes.

Une comparaison du comportement des oscillateurs de ces deux systèmes (lèvres et plis vocaux) sera menée dans le chapitre 4.

2.4 Position du problème et organisation du document

Après avoir étudié le fonctionnement de l’appareil phonatoire humain et notamment le phénomène de vibration laryngée dans la partie I grâce à

- des méthodes exploratoires existantes dans le chapitre 3,
 - une étude comparative du comportement vibratoire de la glotte en phonation et des lèvres du joueur de cuivre pendant le jeu dans le chapitre 4,
 - l’élaboration d’un nouvel outil pour la mesure de l’activité glottique dans le chapitre 5,
- mon étude sera menée selon la progression proposée en section 2.2.3.

La partie II est consacrée à une étude du modèle source-filtre dans le cadre du modèle tous-pôles décrit dans la section 2.2.1 (choix C1 dans la section 2.2.3). Le choix s’est porté dans mon travail sur un modèle causal/anticausal, présenté au chapitre 6, particulièrement adapté au problème de la séparation source-filtre, abordé dans le chapitre 7.

La partie III est dédiée à l'étape suivante dans les choix de la section 2.2.3 : des modèles d'analyse/synthèse de la voix à géométrie forcée (choix **C2**). Après avoir étudié l'analyse/synthèse des fonctions d'aire glottique (pilotage géométrique du modèle, chapitre 8), un modèle simplifié, proche du modèle source-filtre, est établi dans le chapitre 9 afin d'évaluer l'apport de l'information d'interaction entre source glottique et conduit vocal sur la synthèse sonore. Un modèle de conduit vocal plus fin est ensuite étudié et des perspectives sont proposées pour faire évoluer le modèle de glotte.



Première partie

Outils pour l'exploration et la mesure de
signaux biométriques



SOMMAIRE DE LA PARTIE

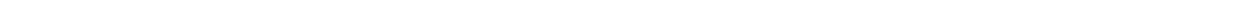
La partie **I** est consacrée à l’exploration de l’activité glottique et à ses outils.

Le chapitre **3** présente les différentes méthodes utilisées pour visualiser et caractériser les mouvements des plis vocaux et du larynx, à savoir

- l’imagerie par rayons X, scanner et résonance magnétique (mouvements du larynx),
- l’échographie (mouvements du larynx et anatomie),
- la vidéo-endoscopie ultra-rapide (vibration glottique),
- l’électroglottographie (mouvements du larynx et vibration glottique).

A l’aide des outils présentés dans le chapitre **3**, une étude comparative du comportement des plis vocaux pendant la phonation et des lèvres du joueur de cuivre est menée dans le chapitre **4**. Les résultats présentés dans ce chapitre sont issus de travaux réalisés en collaboration avec Vincent Fréour, doctorant à l’université McGill de Montréal, et le texte du chapitre, en anglais, provient d’un article soumis au journal *Acta Acustica united with Acustica* en Août 2013. Un résumé en français est proposé en préambule.

Enfin, le chapitre **5** est consacré à la construction d’un nouvel outil non invasif de mesure de l’activité glottique, qui vise à pallier les limites des méthodes actuelles. Le principe d’un appareil inspiré de l’électroglottographie et exploitant la technique de la tomographie d’impédance électrique est présenté et validé. La construction d’un premier prototype a été initiée pendant le stage de fin d’études d’ingénieur en électronique d’Antonin Mauffrey, co-encadré par moi-même, Thomas Hélie et Emmanuel Fléty, et les éléments matériels et logiciels réalisés sont présentés.



Chapitre 3

Méthodes pour l'exploration laryngée

Sommaire

3.1	Ordres de grandeur temporel et spatial du mouvement glottique	40
3.2	Échographie	40
3.3	Vidéo-endoscopie ultra-rapide	41
3.4	Électroglottographie	42

3.1 Ordres de grandeur temporel et spatial du mouvement glottique

L'observation des plis vocaux en vibration est difficile pour trois raisons.

Le larynx est un organe interne délicat. Il est impossible d'y accéder directement sans modifier les conditions de vibration.

Les plis vocaux bougent rapidement. Les fréquences fondamentales de vibration sont d'environ 80 – 160 Hz pour les hommes et 150 – 300 Hz pour les femmes en voix parlée [CT89]. En voix chantée, la plage de fréquence est plus importante, descendant jusqu'à 60 Hz environ pour les basses et montant jusqu'à 1400 Hz environ pour les sopranos [Mic88]. A ces fréquences de vibration s'ajoutent les mouvements du larynx (déglutition et mouvements pendant la phonation) qui peuvent descendre jusqu'à 0.1 Hz. Il apparaît donc nécessaire de pouvoir observer le mouvement avec une bande passante d'au moins 0,1 – 10 kHz, soit une fréquence d'échantillonnage d'au moins 20 kHz, pour être capable de rendre compte de toute la mobilité du larynx.

Les dimensions mises en jeu sont petites. Les plis vocaux mesurent entre 13 et 17 mm chez les femmes et entre 17 et 24 mm chez les hommes [Tit94]. L'amplitude de la vibration glottique pendant la phonation est d'environ 3 mm [LSGR94].

Pour ces raisons, les méthodes d'imagerie classiques telles que les rayons X, l'imagerie par résonance magnétique et le scanner sont inefficaces pour mesurer l'activité glottique. En revanche, elles peuvent être de bons outils pour explorer l'anatomie des cartilages du larynx ou bien les mouvements lents des articulateurs et du larynx dans son ensemble.

3.2 Échographie

L'échographie est utilisée fréquemment par les oto-rhino-laryngologues, notamment pour l'examen de la thyroïde, mais n'a pas encore été utilisée à notre connaissance pour l'examen des plis vocaux, principalement du fait de la faible fréquence d'échantillonnage temporelle des appareils classiques, de la présence du cartilage thyroïde qui gêne le passage des ultra-sons et de la difficulté de contrôler la position et l'angle du plan de coupe en synchronisation avec les mouvements du larynx. J'ai réalisé, avec l'aide d'Aude Lagier, des acquisitions en échographie standard sur mon larynx. Grâce à une faible calcification du cartilage thyroïde (qui a tendance à augmenter avec l'âge), les images se sont révélées lisibles pour un œil expert. Elles sont cependant inutilisables pour faire des relevés d'aire glottique ou de conduit laryngé. Des exemples d'images obtenues dans les plans transverse et frontal sont présentés en figure 3.1.

La technique récente de l'échographie ultra-rapide [SCTH99, TBSF02, DGTF06] permet d'obtenir jusqu'à 18000 images échographiques par seconde. Mickaël Tanter et Jean-Luc Gennisson nous ont permis de faire des essais d'imagerie ultra-rapide sur le larynx avec leur matériel au Laboratoire Ondes et Images de l'Institut Langevin, toujours sous la supervision d'Aude Lagier. Malgré les potentialités prometteuses de cette technique (appliquée notamment à l'estimation de l'élasticité des tissus utilisée dans le cadre de la détection de tumeurs cancéreuses [BCT+03]), les images obtenues se sont révélées illisibles en l'état. Un exemple est présenté sur la figure 3.2. Ici encore, le cartilage thyroïde est vraisemblablement en partie responsable de la dégradation de la résolution de l'image et de la montée du bruit.

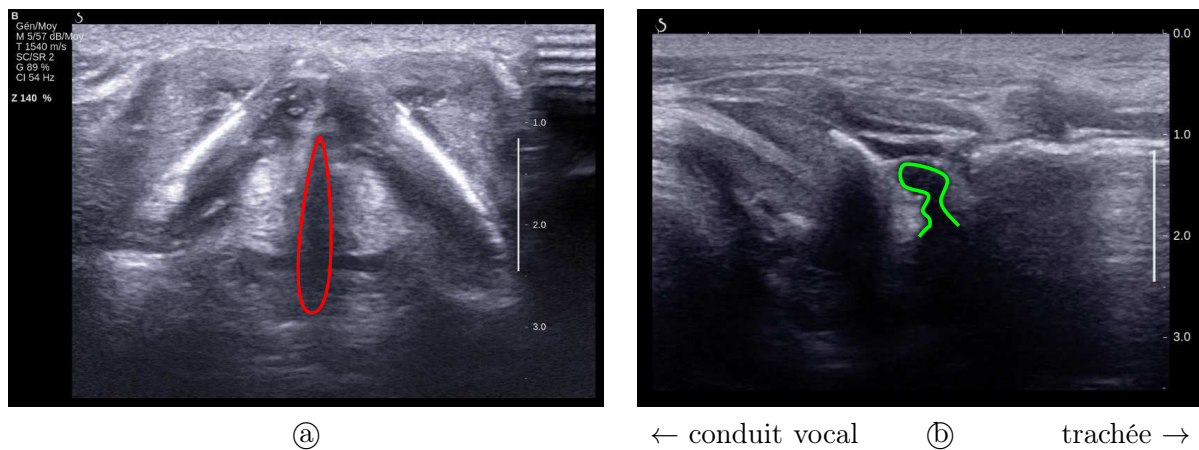


FIGURE 3.1 – Images échographiques des plis vocaux. (a) plan transverse, la position présumée de la glotte est indiquée en rouge. (b) plan frontal, la position présumée du ventricule du larynx est indiquée en vert. Images acquises et annotées avec l’aide d’Aude Lagier.

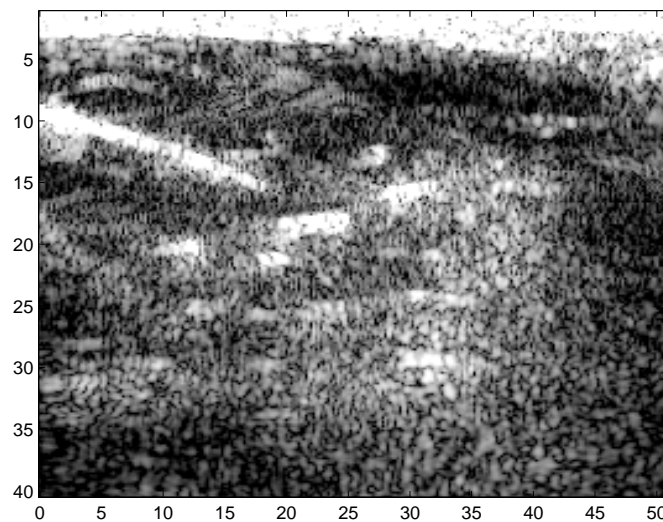


FIGURE 3.2 – Image obtenue par échographie ultra-rapide (plan transverse). L’image est impossible à interpréter. Acquisition réalisée avec l’appareil Aixplorer® de Supersonic Imagine™ [aix], mis à disposition par le Laboratoire Ondes et Images, Institut Langevin, ESPCI.

3.3 Vidéo-endoscopie ultra-rapide

La vidéo-endoscopie, héritière de la laryngoscopie de Manuel Garcia, représente aujourd’hui la méthode la plus efficace pour observer les plis vocaux, en phonation ou non. La figure 3.3 présente des images extraites à intervalle régulier d’un film endoscopique ultra-rapide. L’intervalle temporel couvre un peu moins d’un cycle glottique (la phase fermée, longue, a été tronquée). Ce type de données est d’un grand intérêt pour le domaine médical (phoniatrie, chirurgie ORL, etc.) mais aussi pour la pédagogie du chant [DP03, DPB⁺08]. Pour le travail décrit dans la suite de ce document,

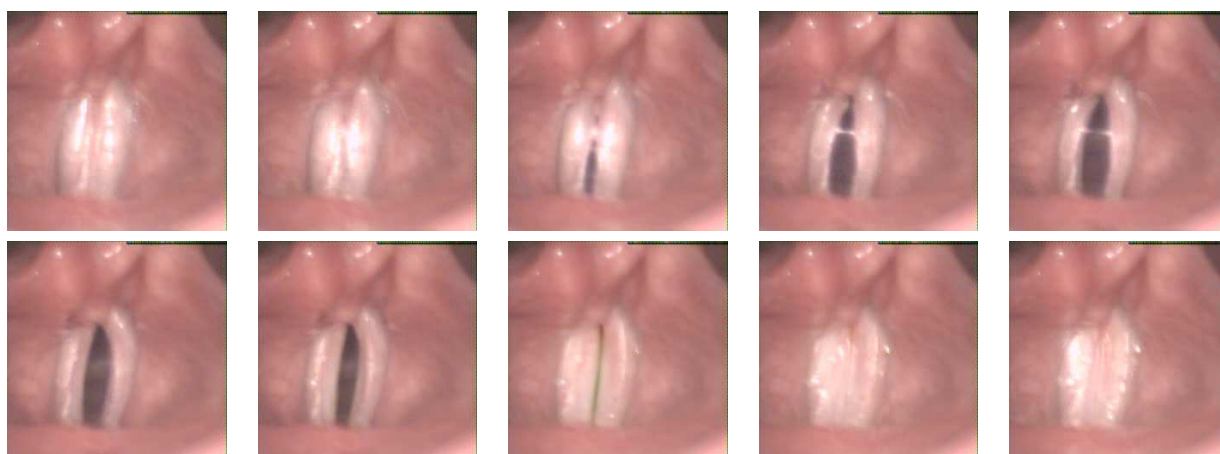


FIGURE 3.3 – Vidéo-endoscopie ultra-rapide des plis vocaux. Données extraites du corpus USC_2008_02.

l'intérêt principal de ces données est de pouvoir extraire des données d'aire d'ouverture glottique dynamiques. Une revue des algorithmes proposés pour extraire l'évolution de l'aire glottique à partir d'une vidéo-endoscopie ultra-rapide est proposée dans [Kar10]. J'ai pu tester, sur les signaux disponibles dans le corpus USC_2008_02, l'algorithme proposé par Deliyski dans [Del06] ainsi que celui proposé par Degottex et al. dans [DBR08a]. Le deuxième a l'avantage de donner une mesure en 2D de l'aire glottique pour chaque image mais demande un travail de réglage fin sur chaque vidéo et je n'ai pas réussi à obtenir des résultats satisfaisants sur l'ensemble des mesures. Quelques vidéos ont pu être analysées avec cet algorithme et les résultats obtenus seront l'objet de l'étude comparative présentée dans le chapitre 4. Le premier algorithme ne permet d'estimer qu'une valeur mono-dimensionnelle de la fonction d'aire glottique et les résultats obtenus sont pollués dans les basses fréquences par la luminance moyenne de la vidéo. Néanmoins, ces résultats ne présentent quasiment pas d'estimées aberrantes et nous verrons dans le chapitre 8 que nous pouvons aisément nous débarrasser du problème en basses fréquences.

L'inconvénient principal de la vidéo-endoscopie est que la vision des plis vocaux est limitée à leur face supérieure. Outre le fait que les mouvements de l'épiglotte et des bandes ventriculaires peuvent parfois partiellement cacher les plis vocaux, les mouvements dans l'épaisseur, comme ceux présentés dans la figure 1.6 pour le mécanisme 1, ne peuvent pas être observés par cette méthode. A l'heure actuelle, d'après nos connaissances, la seule technique permettant d'approcher ce mouvement est l'électroglottographie.

3.4 Électroglottographie

L'ElectroGlottoGraphe (EGG) est un appareil électrique communément utilisé par le monde médical qui traite de la voix. Cet appareil, proposé en 1957 par Philippe Fabre [Fab57], permet d'estimer la conductance du larynx en faisant passer un courant alternatif d'amplitude constante entre deux électrodes situées sur la gorge du patient de part et d'autre du larynx, et en mesurant la tension résultante, comme présenté sur la figure 3.4-@. Les mesures sont faites à haute fréquence (autour de 1 ou 2 MHz en général), d'une part pour éviter l'électrolyse du sang qui s'opère lorsque le corps humain est soumis à un courant continu, d'autre part car la résistance moyenne mesurée

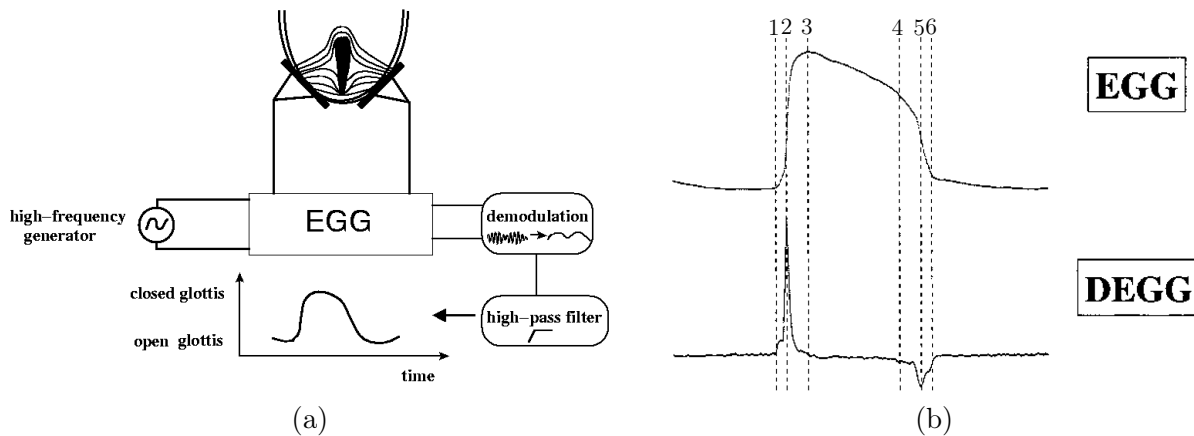


FIGURE 3.4 – (a) Description d'un appareil EGG. (b) Exemple de signaux EGG et DEGG sur une période de vibration glottique. 1-3 : phase de fermeture, 2 : instant de fermeture, 3-4 : phase fermée, 4-6 : phase d'ouverture, 5 : instant d'ouverture, 6 au 1 suivant : phase ouverte. D'après [HDDC04].

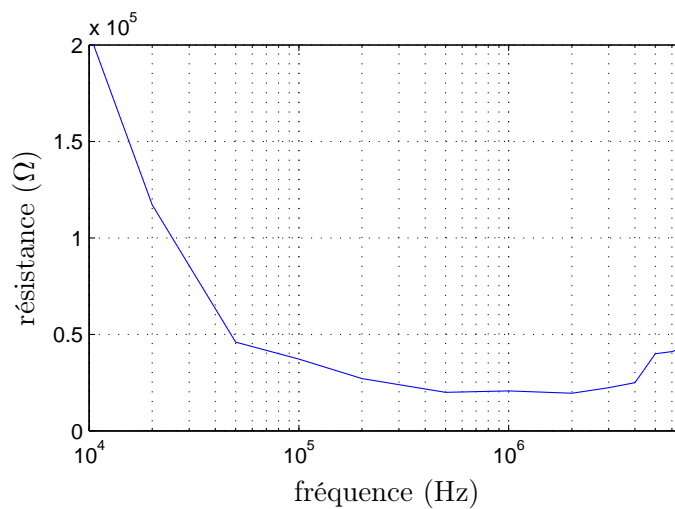


FIGURE 3.5 – Résistance du larynx mesurée en fonction de la fréquence porteuse. Mesures effectuées par Antonin Mauffrey.

entre deux électrodes disposées de part et d'autre du larynx présente un minimum dans cette zone, comme le montre la figure 3.5 (mesures effectuées par Antonin Mauffrey, en accord avec les résultats de la littérature).

Un exemple typique de signal EGG et de sa dérivée sur une période de vibration glottique est présenté en figure 3.4-ⓑ. Par principe, le signal EGG renseigne sur le degré de contact des plis vocaux [Orl91]. La conductance est élevée lorsque les plis vocaux sont en contact (glotte fermée), et faible lorsque les plis vocaux ne le sont pas (glotte ouverte). Les signaux EGG sont utilisés, entre autres, pour

- détecter la phase fermée dans le but d'améliorer les performances de l'estimation du filtre vocal par filtrage inverse [Kri83],
- détecter les segments voisés des segments non voisés et estimer la fréquence fondamentale dans le cas voisé [Kri83]

- estimer le quotient ouvert, défini comme le ratio entre la durée de la phase fermée et la période [HDDC04],
- étudier les mécanismes laryngés [RHC09].

La littérature est riche concernant les analyses comparées entre signaux EGG et vidéo-endoscopies ultra-rapides [Kri83, BLM83, CSM84, CK85, CHM⁺90, EDS⁺10, Kar10]. Je présenterai au chapitre 4 une étude comparée entre vibration de la glotte en phonation et vibration des lèvres d'un joueur de cuivres fondée sur l'analyse des signaux vidéo-endoscopiques ultra-rapide et EGG synchronisés.

Rothenberg propose en 1992 un EGG à deux canaux [Rot92]. Cet appareil donne une estimation de la hauteur du larynx et permet donc un accès à un signal EGG dont l'amplitude est corrigée et un placement plus efficace des électrodes. L'électroglottographe commercial *EG-2* de *Glottal Enterprise*, aujourd'hui très répandu, repose sur le principe proposé par Rothenberg.

Malgré l'amélioration apportée par les résultats de Rothenberg, l'EGG reste un instrument de mesure qui fournit une description uni-dimensionnelle de la vibration glottique, dont il est impossible d'extraire toute l'information sur les mouvements tridimensionnels des cartilages du larynx et des plis vocaux présentés dans le chapitre 1. Afin d'augmenter la quantité d'information que peut apporter un EGG, Kob et Frauenrath proposent en 2009 un EGG à 36 canaux (6 électrodes sont placées de chaque côté du larynx, fournissant $6 \times 6 = 36$ chemins) [KF09]. Ce système permet la mesure simultanée de la position en 2D du larynx et du signal EGG dont les variations d'amplitude dues aux mouvements du larynx sont compensées.

Ce dernier outil ne permet malheureusement pas un accès plus complet aux caractéristiques de la vibration laryngée en l'état. De plus, l'accès aux 36 canaux se fait par multiplexage temporel, ce qui réduit considérablement la bande passante des signaux mesurés. Je présenterai au chapitre 5 un prototype d'appareil de mesure biométrique développé pendant mes travaux de thèse qui tente de répondre à ces problèmes.

Chapitre 4

Mesures multimodales sur les lèvres et la glotte, comparaison entre deux systèmes à valve humaine

Sommaire

4.1 Introduction	47
4.1.1 Context	47
4.1.2 Problem statement and document organisation	48
4.2 Experimental setup	49
4.2.1 Human voice production system	50
4.2.2 Brass instruments	51
4.3 Visualisation tool	52
4.3.1 Opening area extraction	52
4.3.2 Synchronous display of multimodal measurements	53
4.3.3 First observations from videos visualisation	54
4.4 Features extraction, analysis and comparisons	54
4.4.1 Qualitative comparison between vocal-fold & lip EI signals (Obs. 1 & 2)	54
4.4.2 Qualitative comparison between vocal-fold & lip DEI signals (Obs.1 &3)	55
4.4.3 Lip and vocal-fold OA and DOA signals and parametric representation (Obs. 4 & Prop. 1)	58
4.4.4 Geometrical parameters of lip and glottis opening areas (Obs. 2)	61
4.4.5 Opening geometries and electrical impedance signals (Obs. 2)	63
4.5 Conclusion	66

Il a été mentionné dans la section 2.3 que le système de production de la voix humaine partage de nombreuses similitudes avec les cuivres. Durant mon travail de thèse, j’ai mis en place une collaboration avec Vincent Fréour, doctorant au laboratoire *Computational Acoustic Modeling Laboratory, Centre for Interdisciplinary Research in Music Media and Technology* de l’université McGill à Montréal au Canada, afin d’étudier les similarités et différences entre les oscillateurs de ces deux systèmes à l’aide de méthodes de mesures communes adaptées aux lèvres d’un musicien pendant le jeu et à la glotte en phonation. Le texte de ce chapitre, en anglais, présente les résultats de cette étude et est extrait d’un article que j’ai rédigé avec Vincent Fréour, René Caussé, Thomas Hélie et Gary P. Scavone, soumis au journal *Acta Acustica united with Acustica* le 10/08/2013. Un résumé en français est proposé ci-dessous.

Résumé :

L’ensemble musicien-cuivre et le système de production de la voix humaine sont deux systèmes composés d’une “valve humaine”, une constriction dans un tube, oscillante (lèvres du musicien ou plis vocaux du chanteur) couplée à un résonateur acoustique (colonne d’air dans l’instrument ou dans le conduit vocal). Les modèles de synthèse sonore développés pour ces deux systèmes sont très similaires. Tous peuvent se résumer au même ensemble d’équations : (a) équation mécanique représentant la dynamique de l’oscillateur, (b) équation fluide reliant l’écoulement et la chute de pression à travers la valve et l’acoustique dans les résonateurs et (c) équation acoustique représentant l’effet des résonateurs. Cependant, les grandeurs caractéristiques de ces systèmes (masse et élasticité des tissus notamment) ont des ordres de grandeur différents. Ceci implique des différences dans le comportement vibratoire de la glotte et des lèvres. En particulier, le couplage entre le résonateur et l’oscillateur est en général significativement plus important dans le cas des cuivres.

Afin d’étudier et de comparer ces deux systèmes, une structure commune de mesures et d’analyse de données permettant d’observer *in vivo* et de façon relativement non intrusive les comportements vibratoires des lèvres et de la glotte est proposée. Notre méthode s’appuie sur trois outils d’exploration : (i) la mesure impédancemétrique avec l’électroglottographie (EGG) et son équivalent l’électrolabiographie (ELG [FS11, FS12]), (ii) l’enregistrement vidéo haute vitesse de l’oscillateur en vibration et (iii) l’enregistrement audio du son rayonné. Les systèmes de mesure associés aux lèvres et aux plis vocaux sont présentés en détail et nous introduisons un outil de visualisation permettant d’observer les différents signaux de mesure synchronisés. Nous nous intéressons ensuite à l’extension des outils standards d’analyse des signaux électroglottographiques aux signaux électrolabiographiques. Enfin, nous explorons les similarités et différences entre (a) les signaux EGG et ELG, (b) les signaux d’ouverture glottique et d’ouverture des lèvres, et (c) les relations EGG/aire d’ouverture glottique et ELG/aire d’ouverture des lèvres.

Cette étude comparative nous mène aux résultats suivants.

R1 Les signaux de type EGG et ELG montrent des signaux de nature similaire, ce qui nous mène à étendre la notion de quotient ouvert Oq (ratio entre la durée de la phase ouverte et la période, estimé à partir des signaux EGG) au cas des signaux ELG. L’algorithme DECOM [HDDC04] est utilisé pour estimer la valeur de Oq sur les deux types de signaux. Nous montrons que

les quotients ouverts sont bien plus importants pour le cas des lèvres que pour la glotte, ce qui confirme les résultats obtenus dans [BCG10] à partir d'analyse de vidéos ultra-rapides.

- R2** Une catégorisation des signaux ELG selon la fréquence de jeu apparaît dans nos mesures. Si les signaux ELG en basse fréquence sont assez différents des signaux EGG produits en mécanisme M1, les signaux ELG en haute fréquence présentent les mêmes caractéristiques que les signaux EGG produits en mécanisme M2, notamment en termes d'asymétrie du pulse.
- R3** Les pics multiples sont beaucoup plus fréquents sur les dérivées des signaux ELG que sur les dérivées des signaux EGG, indiquant des fermetures séquentielles le long des lèvres (fermetures en zip), plus douces que pour les plis vocaux.
- R4** Le résultat **R3** est confirmé par l'analyse des vidéos ultra-rapides. Les phases de fermeture apparaissent abruptes et courtes dans le cas des plis vocaux tandis qu'elles sont beaucoup plus progressives et longues dans le cas des lèvres. Une conséquence de cette observation est que le modèle de Liljencrants-Fant, qui sera proposé pour représenter de façon paramétrique les signaux d'ouverture glottique dans le chapitre 8, est beaucoup moins performant pour représenter les signaux d'ouverture labiale que les signaux d'ouverture glottique.
- R5** La fermeture plus douce des lèvres est en partie expliquée par un mouvement plus progressifs de coins de l'ouverture. Ce mouvement, moins prononcé et plus abrupt dans le cas de l'ouverture glottique, explique aussi les différences qui apparaissent dans les propriétés des signaux EGG et ELG et qui rend les signaux ELG plus difficilement interprétables que les signaux EGG.
- R6** Enfin, nous concluons sur la mise en évidence, notamment grâce à la représentation des signaux d'aire d'ouverture en fonction des signaux impédancemétriques, d'un mouvement dans la dimension perpendiculaire au plan d'ouverture. Ce mouvement n'est pas accessible par enregistrement vidéo et n'est pas entièrement caractérisable par la mesure impédancemétrique mono-canal. Un instrument de mesure électrique multi-canal permettant d'accéder aux mouvements dans cette dimension est actuellement à l'étude et sera présenté dans le chapitre 5.
-

4.1 Introduction

4.1.1 Context

Brass instrument-player and the human voice production systems are both composed of a vibrating “human valve” –constriction in a pipe– coupled to an acoustic resonator : lips coupled to the instrument air column or vocal folds coupled to the vocal tract. In both cases, the aeroacoustic coupling is responsible for the self-oscillations. The various sounds and regimes are obtained by modifying the resonator geometry and the mechanical properties of the constriction. Hence, similar dynamical systems involving lumped element physical models (mass, spring and dampers) have been developed for the simulation of brass instruments and voice.

The dynamics of these two structures are commonly represented by a set of three coupled equations : a mechanical equation representing the dynamics of the nonlinear oscillator, a flow equation

relating the acoustic pressure drop across the lips or glottis to the volume flow generated in the upstream and downstream resonators –vocal tract and instrument for brass, trachea and vocal tract for voice–, and the acoustic equations simulating the effect of the resonators. Physical representations of lip-reed valves commonly consider the lips as identical and placed symmetrically against the mouthpiece, then reducing the problem to the modelling of one lip. Common lip models include parametric one-mass one-dimensional models allowing the lips to strike according to either an outward or inward mechanism [Hel54, AS95, Cam04]. Refined versions of lip models have been proposed including sideways striking valve models [Fle93, CGC00], two-dimensional models combining the effect of the pressure on the internal, external and channel faces of the lips [AS96], or two-mass models involving two coupled oscillators [CGC00, Ric03]. Similar approaches have been studied for the vocal folds, the most common physical representation being the n -mass model, also parametrized as one or two-dimensional [FL68, IF72, PHvH⁺94, Hvv05, Tit73, Tit74, Cve12]. The pressure drop and flow equations between the lips or vocal folds can be of various complexity, from simplified Bernoulli to complete Navier-Stokes equations [MG76, VPF⁺04, LBD⁺05]. Finally, propagation inside the upstream and downstream resonators is simulated using a linear model connected to radiation impedances at both ends [KL62, Mae82a, Sto05, EDCG12]. Contrary to the voice production system, the downstream air-column of a brass instrument presents a number of strong impedance resonances. In order to select a proper sound frequency, the resonance frequency and damping factor of the lip-valve system must be accurately adjusted, along with the driving quasi-static pressure in the player's mouth.

Brass instrument-player and the human voice production systems also share difficulties for the in-vivo exploration of the oscillator activity. In both cases, it is difficult to monitor directly the oscillator vibration without invasive methods likely to perturb sound production mechanisms.

Although they share strong similarities and can be modelled with the same class of systems, important differences exist between lips and glottis. The physiological characteristics –size, stiffness of the tissue of the oscillator and the resonator– are very different between brass instruments and human voice. These differences lead to important dissimilarities in terms of acoustic coupling strength. In particular, the coupling can be very strong in brass instruments primarily due to the geometry of downstream resonator [CGC00], whereas it is much more moderate in voice production.

In order to study and compare these two excitation mechanisms, a measurement and analysis framework allowing in-vivo investigations of the behaviour of the lips and glottis is needed. The non-invasive assessment of vibrating lips and vocal-fold motion has been the object of significant research effort, yet, to our knowledge, only three techniques give in-vivo relatively non-intrusive access to the motion of lips and vocal folds with a sufficient frequency bandwidth : high-speed video recording, electrical impedance measurement and audio recording. Thus, our measurement framework is based on these three technologies.

4.1.2 Problem statement and document organisation

Several studies have been conducted on electroglottography and high-speed video endoscopy of the glottis [BLM83, CNL⁺83, CL84, CSM84, CK85, CHM⁺90, DBR08b, GDOM09, Kar10]. This paper aims at 1) building a tool for a common relatively non-intrusive in-vivo measurement method of lips and glottis vibration, 2) extending classic electroglottographic analysis tools to the electrolabiograph, 3) exploring similarities and differences between lips and glottis openings, 4) exploring the relationships between electrical impedance signals and lips and glottis openings.

The method to measure simultaneously and synchronously high-speed video, electrical impedance and radiated sound for the lips and glottis is described in section 4.2. A visualisation tool allowing all measurements to be synchronously displayed is presented in section 4.3, leading to a first visual exploration of the synchronised data. Section 4.4 is dedicated to the presentation of the analysis framework and the study of the signals database. At first, electroglottographic and electrolabiographic signals are studied and compared in sections 4.4.1 and 4.4.2, highlighting the interpretability of the electrolabiographic signals with regard to standard electroglottographic features. Secondly, open area waveforms of the lips and glottis are compared in section 4.4.3 and 4.4.4. It has been shown that an efficient parametric representation of glottis open area waveforms is possible [HHCD12]. This approach is extended to the lips and the parametric modelling of lip open area waveform is evaluated. Finally the relationships between electrical impedance signals and open area are explored in section 4.4.5.

4.2 Experimental setup

In 1894, Manuel Garcia managed for the first time to visualize the vibrating glottis in a living human using his laryngoscope [Gar54, Gar94]. The first laryngoscope inspired many other tools, including modern fiberscope and video endoscopy. In our work, we chose to record the motion of vocal folds using a high-speed video endoscope. This tool gives a direct access to the complete motion –unlike stroboscopic techniques– and, after some practice, allows singers to sing as in natural conditions using a modern video endoscope recording the vocal-fold motion through the mouth or the nose. Section 4.2.2 will present how one can record the lips motion during brass instruments playing with a standard high-speed camera and a transparent mouthpiece.

The second selected device for the non-invasive in-vivo exploration of lips and glottis is the ElectroGlottoGraph –EGG–, an electric device widely used by voice pathologists. The first electroglottograph was developed by Philippe Fabre in 1957 [Fab57]. This technology is based on impedance analysis of the larynx; two electrodes are placed on each side of the throat, a low-intensity high-frequency modulated current (commonly around 2 MHz) is generated and applied to the laryngeal level, and the impedance between the two electrodes is measured. This impedance is approximately proportional to the vocal-fold contact area. A basic EGG device is schematically presented in Fig. 4.1-a. Fig. 4.1-b presents a typical EGG signal and its derivative with common temporal features. For a more detailed review about EGG technology, we refer the reader to [CK85, CHM⁺90, HDDC04].

Unlike many imaging technologies (MRI, X-ray, echography), the frequency range of the EGG is sufficiently large to allow monitoring of vocal-fold vibration : from 0.2 Hz (for swallowing) to F_{max} where F_{max} is between 4 kHz (telephone bandwidth) and 20 kHz (audible bandwidth). The output of this measurement system is a time-domain one-dimensional signal, which, as we'll see in section 4.4.5, does not reflect the whole spatial characteristics of vocal-fold vibratory movement. EGG is used in both medical and research fields. Typical applications of EGG include extraction of vocal-fold vibration features and identification of the laryngeal mechanism [HDDC04, RHC09]. Fréour and Scavone proposed a modified version of the electroglottograph adapted for the lips of a trombone player [FS11, FS12].

Finally, the radiated sound is recorded with a standard microphone. As we'll see in sections 4.2.1 and 4.2.2, a special attention has to be given to the synchronisation of the measurements, as high-speed video, electrical impedance and audio are not all recorded with the same sampling

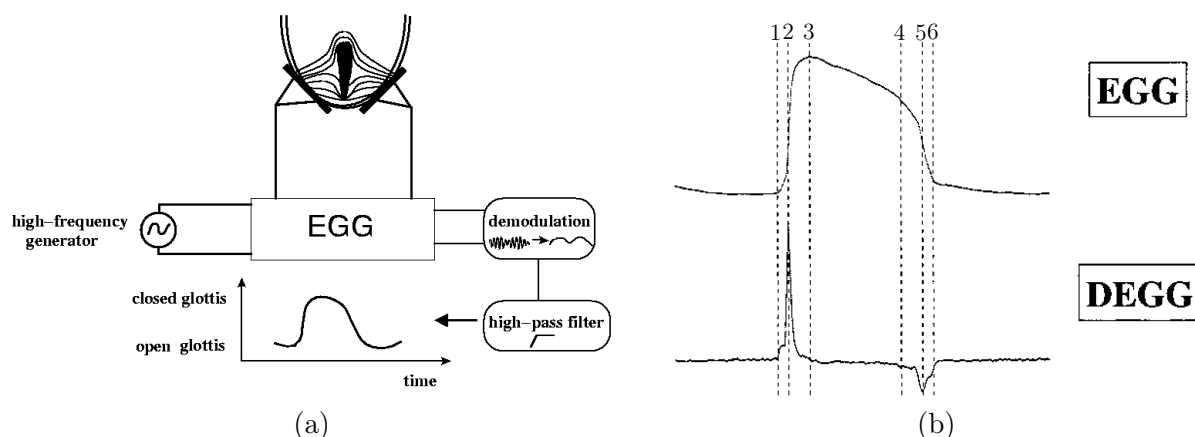


FIGURE 4.1 – (a) Schematic description of an EGG device. (b) Typical glottal cycle on EGG and EGG derivative signals. 1-3 : closing phase, 2 : closing instant, 3-4 : closed phase, 4-6 : opening phase, 5 : opening instant, 6 to the next 1 : open phase. From [HDDC04].

frequency.

4.2.1 Human voice production system

The vocal data exhibited in this article are extracted from the corpus “USC_2008_02” of multimodal measurements built at IRCAM by Gilles Degottex and Erkki Bianco [DBR08b]. An electroglottographic signal comes from a portable electroglottograph (EG-90 from F-J Electronics) and is recorded with a sampling frequency of 44150 Hz. Vocal-fold motion is filmed in color using a rigid endoscope, which passes through the mouth, connected to a high speed camera (HreS ENDOCAM 5562) recording 4000 frames per second with a 256×256 pixel resolution. Finally, the radiated sound outside the mouth is recorded with a standard microphone at a sampling frequency of 44150 Hz. All these measurements are synchronized using Richard Wolf’s HreS Endocam system [End]. The system specifies that the synchronization between high frame rate signals (electroglottograph and audio) and high-speed video shows a temporal precision of one video frame, that is $250 \mu s$. This delay is relatively small regarding the usual speaking frequencies of male speakers but must be considered in our study, especially when the fundamental frequency goes up. Furthermore, the Shannon frequency corresponding to the high-speed video frame rate, 2000 Hz, is also relatively restrictive for the study of signals extracted from the video and must be considered. Moreover, no precise information about the microphone position and the vocal tract length is available. Hence, we can’t properly synchronise the audio data to video and EGG recordings.

All the measurements exhibited here were performed during sustained vowels. In this article, we introduce five recordings –labelled G1 to G5 in the different figures and in Table 4.2– corresponding to three different speakers. These five measurements present an interesting variety of vocal qualities, although they correspond to the same French vowel /œ/ in *mezzo-forte* dynamic. The three speakers are men singing at relatively low frequencies so that the spectral content of the glottal activity is captured with a relatively good accuracy regarding the camera frame rate. Unfortunately this implies that the recorded regimes were very similar. In sections 4.4.1 and 4.4.2, we introduce four other EGG measurements G6 to G9 with higher fundamental frequencies, but without video recordings.

4.2.2 Brass instruments

Video acquisition is performed using a transparent mouthpiece developed by Castellango et al. [CCS83]. This mouthpiece is made of a cylindrical cup closed by a flat surface perpendicular to the cylinder axis. The shank of the mouthpiece is oriented laterally allowing an open field for lip visualisation. The mouthpiece is fixed on a stand and connected to a flexible plastic tube of the same inside diameter and length as the trombone slide in closed position. The end of the tube is connected to the bell section of the instrument. This set-up therefore enables one to easily keep the mouthpiece in fixed position while maintaining the acoustical properties of the downstream resonator close to a normal instrument.

Two electrodes made of tin-plated copper foil shielding tape are glued on the rim of the mouthpiece, as shown in Fig. 4.2, and connected to a commercial electroglottograph signal conditioner (Voce Vista) as proposed by Fréour and Scavone [FS11, FS12]. This enables lip electrical impedance to be monitored during measurement without impairing the lip filming. The output of the signal conditioner is connected to an acquisition board (Qualysis USB-2533) running at 48 kHz. Prior to experiments, the latency of the EGG signal conditioner was measured using a controllable variable resistance (FET optocoupler) mounted at the electrodes. A group delay of $180\mu s$ was quantified between actual variation of the electrical resistance at the electrodes and the signal conditioner output. This delay is then systematically compensated in the recorded data. In the following, the signal obtained with this method will be called ELG, for ElectroLabioGraph, as proposed in [FS11, FS12].



FIGURE 4.2 – Transparent mouthpiece developed by Castellango et al. [CCS83] equipped with copper foil shielding tape electrodes.

Lip motion is captured using a high-speed camera (Qualisys Oqus 310) running at 6000 frames per second. A strong cool light source is used to compensate for the short exposure time required at this frame rate. The clock signal of the camera is shared with the USB acquisition board to guarantee synchronisation between both acquisition systems. In addition, an LED, attached to the mouthpiece and powered by a 1 Hz square wave from an external signal generator, is filmed during experiments. The voltage at the LED is simultaneously monitored by the USB-2533 acquisition board. This procedure allows for quantification of a possible small delay between the camera and acquisition board at the start of acquisition since communication with both systems is performed with the same computer but through different ports : Ethernet for the camera and USB for the acquisition board. The radiated sound is measured using a standard studio microphone located 30 cm from the bell and connected to the USB interface.

The complete installation is presented in Fig. 4.3.

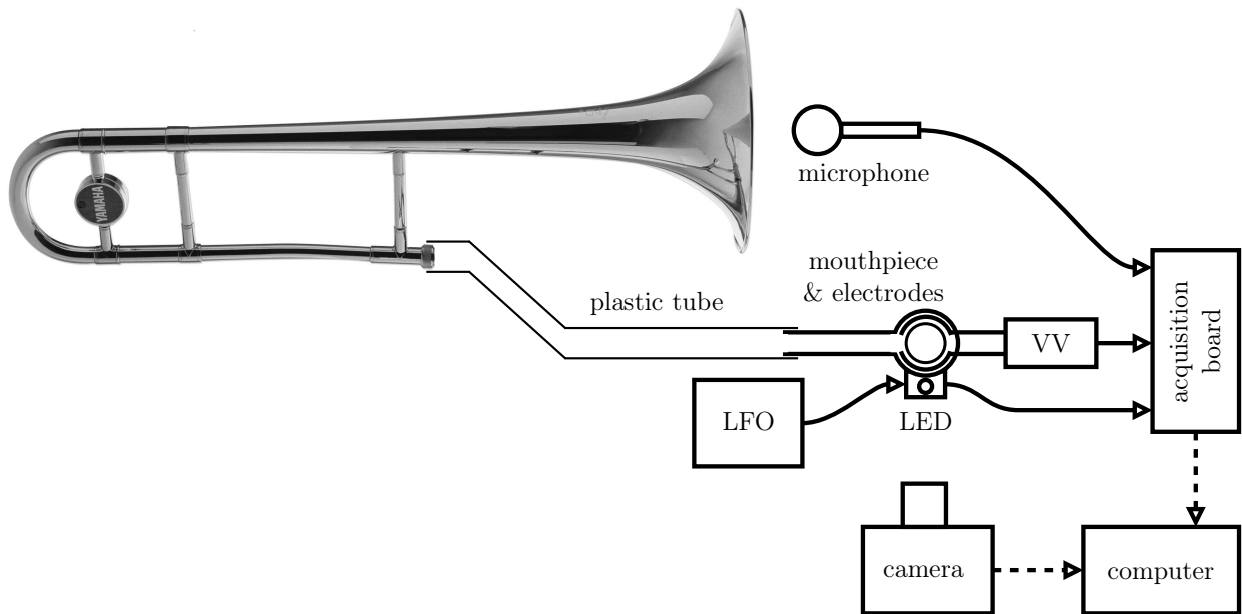


FIGURE 4.3 – Data acquisition set-up for brass instruments. VV stands for Voce Vista signal conditioner.

In this article, we report results from one player playing four tones (Bb2, Bb3, F3, and D4, labelled L1 to L4 in the following) corresponding to the second, third, fourth and fifth peaks of the trombone input impedance. The player was instructed to play at a stable and comfortable dynamic. Results for higher tones are not presented since the lip opening area is too small to be properly detected and extracted from video acquisitions.

4.3 Visualisation tool

In order to run some analysis algorithms indifferently on lips or glottis data, a general data organisation is needed. The following abbreviations will be used :

- **HSV** stands for High-Speed Video and refers to the video recording of lips or glottis,
- **EI** stands for Electrical Impedance and refers to the **EGG** or **ELG** signal, **DEI** stands for its temporal derivative,
- **RS** stands for Radiated Sound and refers to the audio recording of the radiated sound in both cases,
- **OA** stands for Opening Area and refers to the signal representing open surfaces (in pixels or in mm^2) between the lips or the vocal folds, **DOA** stands for its temporal derivative. **2D-OA** stands for the 2D signal representing the shape of the opening area over time. In the voice case, 2D-OA is a direct representation of the *rima glottidis* shape seen from above.

4.3.1 Opening area extraction

The 2D-OA signals are extracted from the HSV using adapted algorithms. For the glottis, an algorithm from G. Degottex [DBR08a] is used. In the case of lips, we developed a simple, but

yet perfectly effective, algorithm based on luminance thresholding. These two algorithms give us accurate estimation of the 2D opening areas over time. An example of detected 2D-OA on one frame is given in Fig. 4.10. OA signals are simply deduced from the 2D-OAs by computing the number of pixels inside the estimated area on each frame. Note that, for the voice case, the distance between the camera and the glottis can be variable due to the vertical motion of the larynx and is not estimated. Hence, neither scaling information nor larynx position are available and vertical motion of the larynx can modulate the size of the estimated 2D-OA signals. Some methods for the dynamic estimation of glottis actual size on HSVs exist (see for example [KBMK13, KF09]) but none has been implemented for the USC_2008_02 corpus. For the trombone, the distance between the camera and the lips is fixed during one measurement and the scaling factor is retrieved by detecting the pixel-size of the radius of the mouthpiece.

4.3.2 Synchronous display of multimodal measurements

A video tool was designed to synchronously display HSV, EI, DEI, OA, DOA and RS signals in both time and frequency domains. This tool also displays an instantaneous fundamental frequency estimation (computed with the YIN algorithm [dCK02] on RS), the 2D-OA estimation and the temporal position. Two screen shots of these videos are shown in Fig. 4.4. Video samples (in color) are available on the corresponding author's website [Hez].

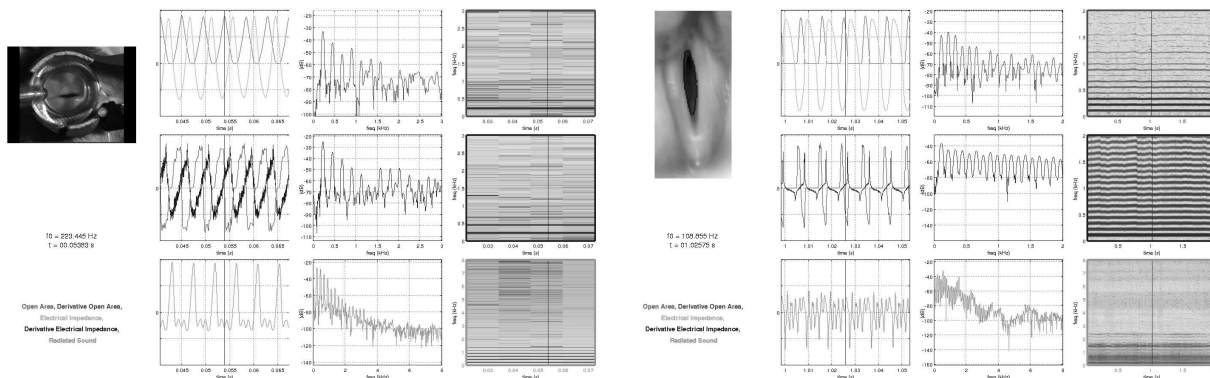


FIGURE 4.4 – Synchronous display of multimodal measurements of lips and glottis. Video samples are available on the corresponding author's website [Hez].

These videos allow to synchronously visualise the temporal evolution of the different signals. One can see OA and its derivative DOA along with its local magnitude spectrum, EI and its derivative DEI along with its local magnitude spectrum and RS along with its local magnitude spectrum. Signals are identified with a captioned color code. It is a useful tool to visually explore the behaviour of lips or glottis in different regimes. One can graphically estimate standard EGG features on the EI and DEI signals, such as the open quotient or the asymmetrical coefficient.

To our knowledge, there is no tool allowing simultaneous and synchronous visualisation of all these measurements. This tool could therefore be the object of further development for medical and pedagogical applications.

4.3.3 First observations from videos visualisation

The visualisation of the generated videos led to the following observations and propositions.

Obs. 1 Similar low / high frequencies categorisation appear for lips and glottis EIs and OAs.

Obs. 2 Strong variations occur during the closed phase for both systems. Also, HSVs and OAs show that the completely closed phase is much shorter for the lips than for the glottis and is characterized by more activity in lip EIs than in glottis EIs. This raises the question of the motion of lips and glottis along the dimension perpendicular to the vocal-fold or lip plane –it will be called “thickness dimension” in the following–, which will be assessed in section 4.4.5.

Obs. 3 Lip DEI signals show generally much less defined peaks and many more instances of double or triple-peaks (as described in [HDDC04]) than the glottis DEI. This feature is studied in section 4.4.2.

Obs. 4 We observe that OA signals tend to all have a similar pulse shape.

Prop. 1 In section 4.4.3, we will investigate a common parametric representation of DOA signals in both cases.

4.4 Features extraction, analysis and comparisons

4.4.1 Qualitative comparison between vocal-fold & lip EI signals (Obs. 1 & 2)

Four representative excerpts of ELG and EGG signals are presented in Fig. 4.5. The four ELG examples are extracted from sustained tones with instantaneous fundamental frequency of 109.5 Hz, 165.2 Hz, 223.5 Hz and 276.6 Hz. The four EGG examples are extracted from sustained vowels with instantaneous fundamental frequency of 143.5 Hz, 191.4 Hz, 271.0 Hz and 422.3 Hz. According to the convention presented in [HDDC04], the two top right excerpts correspond to an M1 laryngeal mechanism while the two bottom right correspond to an M2 mechanism.

Overall, ELG and EGG signals show different waveforms. The open phases in ELGs are marked by significant variations of the signal, while EGGs remain almost constant in both M1 and M2. Moreover, the four ELG excerpts fall in two categories according to the fundamental frequency f_0 ; the two top left signals show a relatively triangular waveform, particularly asymmetrical and closer to a sawtooth for the lower pitch. The increasing phase of the signal, corresponding to an increase in contact between the lips, is characterized by an irregular evolution, while the decreasing phase is short and linear. On the contrary, the two top right signals show a rather sinusoidal and regular waveform.

Although the two lowest frequency ELG signals do not share particular features with M1 EGG signals, the two higher frequency ELGs show common characteristics with M2 EGGs. The pulse-like shape characteristic of EGG in M2 is apparent in ELG signals, with a similar asymmetrical pattern (L4 and G9). The main difference appears in the distinct activity during opening (at ELG and EGG minimum); although EGG signal is relatively constant during the open phase, ELGs show more variations, possibly due to the complex characteristics of the contact between the lips during opening.

We defined the “OA corners” of lip and glottal openings as the position of the corner extremities of the opening area. Looking at the high speed video capture of the lips, we clearly observe that contrary to the vocal folds, the position of the OA corners at the extremity of the lip opening area

varies along the length of the lip contact. This attribute is not verified in the vocal folds whose OA corners are characterized by an almost constant position located at the extremity of the folds. In other words, lips tend to close progressively along the lip length (zipper-like closure), while the whole length of the vocal folds tends to close within a much shorter time frame. This contrast in the opening mechanism between the two valve systems may contribute to explain the differences observed between ELG and EGG waveforms, particularly with respect to the characteristics of ELG signals during the open phase.

4.4.2 Qualitative comparison between vocal-fold & lip DEI signals (Obs.1 &3)

The derivative of the ELG and EGG signals presented in Fig. 4.5 are displayed in Fig. 4.6. Similarly to DEGG, DELG signals display peaks at closing and opening instants. Although these peaks are clearly visible for the two lowest tones, the change in ELG waveform for the two highest tones produce a sawtooth-like DELG waveform making the identification of closing and opening peaks in the DELG less accurate. Again, DELG signals can be categorized according to f_0 ; the two lowest tones show a similar rectangular waveform including a plateau during the open phase as observed in DEGG signals, but with non-negligible positive value while DEGGs are approximately null during open phases. On the contrary, the two highest tones are closer to a triangle or sawtooth signal and the open phase is marked by important variations in DELG. In all DELG signals, the distance between closing and opening peaks is relatively small compared to DEGG, suggesting shorter closed phases for the trombone player's lips than for the singer's glottis.

The double-peak feature observed in voice (clearly visible in G9, see [HDDC04]) is also present in some DELG signals, especially visible in L2. Instances of a triple-peak are even visible in this signal. Analogously to voice, this attribute observed on the closing peaks suggests an irregular or sequential closing phase along the length of the lip opening area. This feature is indeed visible in high speed video captures of the lips at the frequency of signal L2.

The open quotient (Oq) is defined for the glottis as the ratio of the duration of the open phase to the wave period. Efficient methods have been developed to estimate the open quotient from EGG signals or its derivative [CHM⁺90, HGB90, Lec77, LB77, MvS02, Or191, HDDC04]. In this study, the DECOM algorithm [HDDC04] is used to extract Oq values from lip and glottis DEI signals. In addition, DECOM allows the number of opening and closing peaks to be estimated for each period of the DEI signals. The average Oq , the average number of closing peaks n_{cl} , as well as the average number of opening peaks n_{op} are presented in Table 4.2 for all measurements.

Open quotient values are significantly higher for the lips than for the vocal folds. In agreement with previous OA estimations in trombone players' lips [BCG10], this difference is clearly visible in the HSV and OA signals; the completely closed phase is very short in the trombone but can be rather long in voice. This may be partly explained by the fact that, unlike the vocal folds, the lips do not close and open along their thickness. Moreover, instances of multiple closing and opening peaks are more present in the trombone. This observation is in agreement with variation of the position of OA corners during opening and closing of the lips, which possibly results in an more irregular evolution of the contact than for the vocal folds during the opening and closing phases.

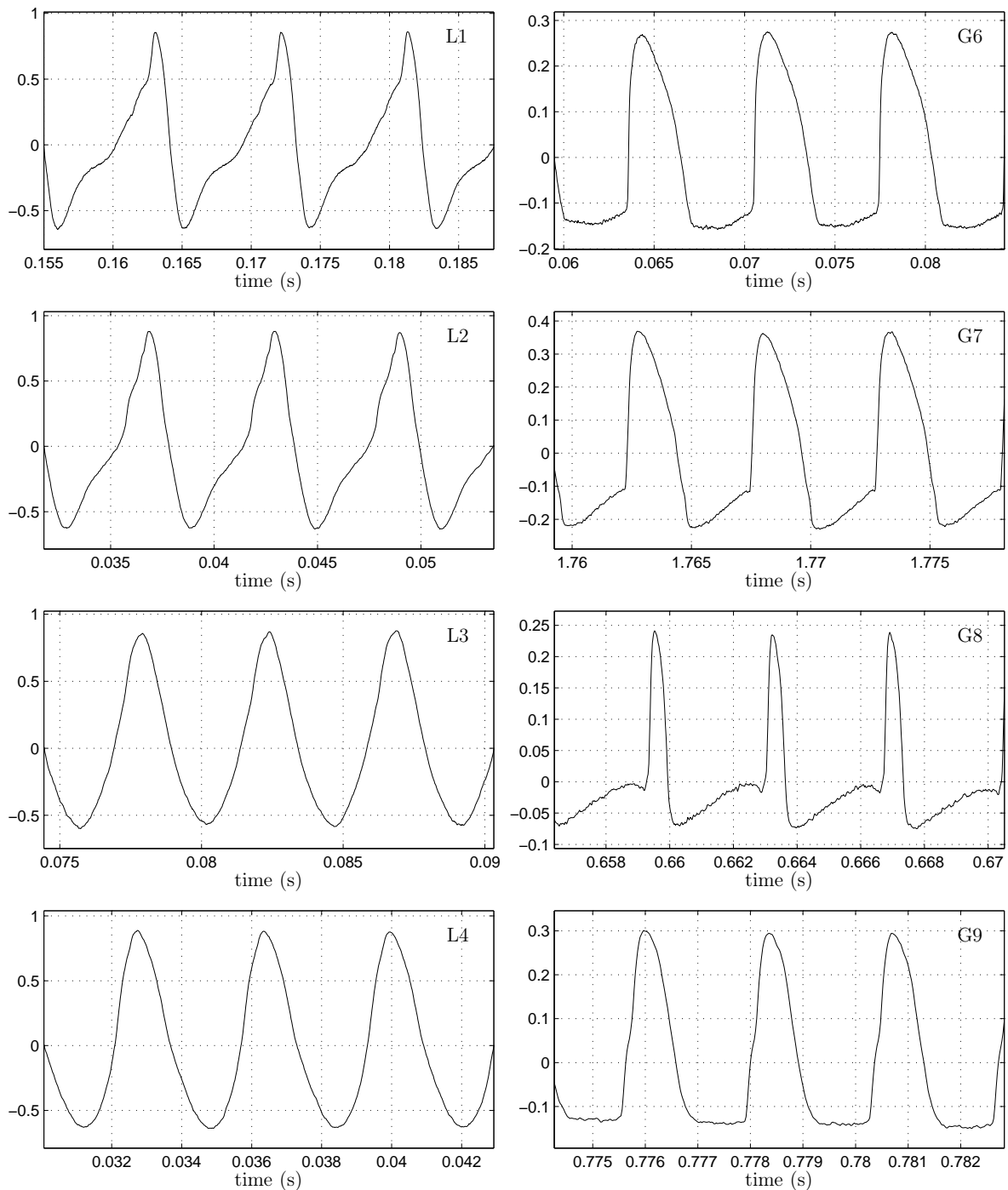


FIGURE 4.5 – Extracts of EI signals of the lips (left column, instantaneous fundamental frequency of 109.5 Hz, 165.2 Hz, 223.5 Hz and 276.6 Hz from top to bottom) and the vocal folds (right column, instantaneous fundamental frequency of 143.5 Hz, 191.4 Hz, 271.0 Hz and 422.3 Hz from top to bottom). Vocal-fold EI signals correspond to, from top to bottom, a male voice in mechanism M1, a female voice in M1, a female voice in M2 and a male voice in M2. According to the definition of EI, higher value correspond to higher degree of contact between the lips or vocal folds.

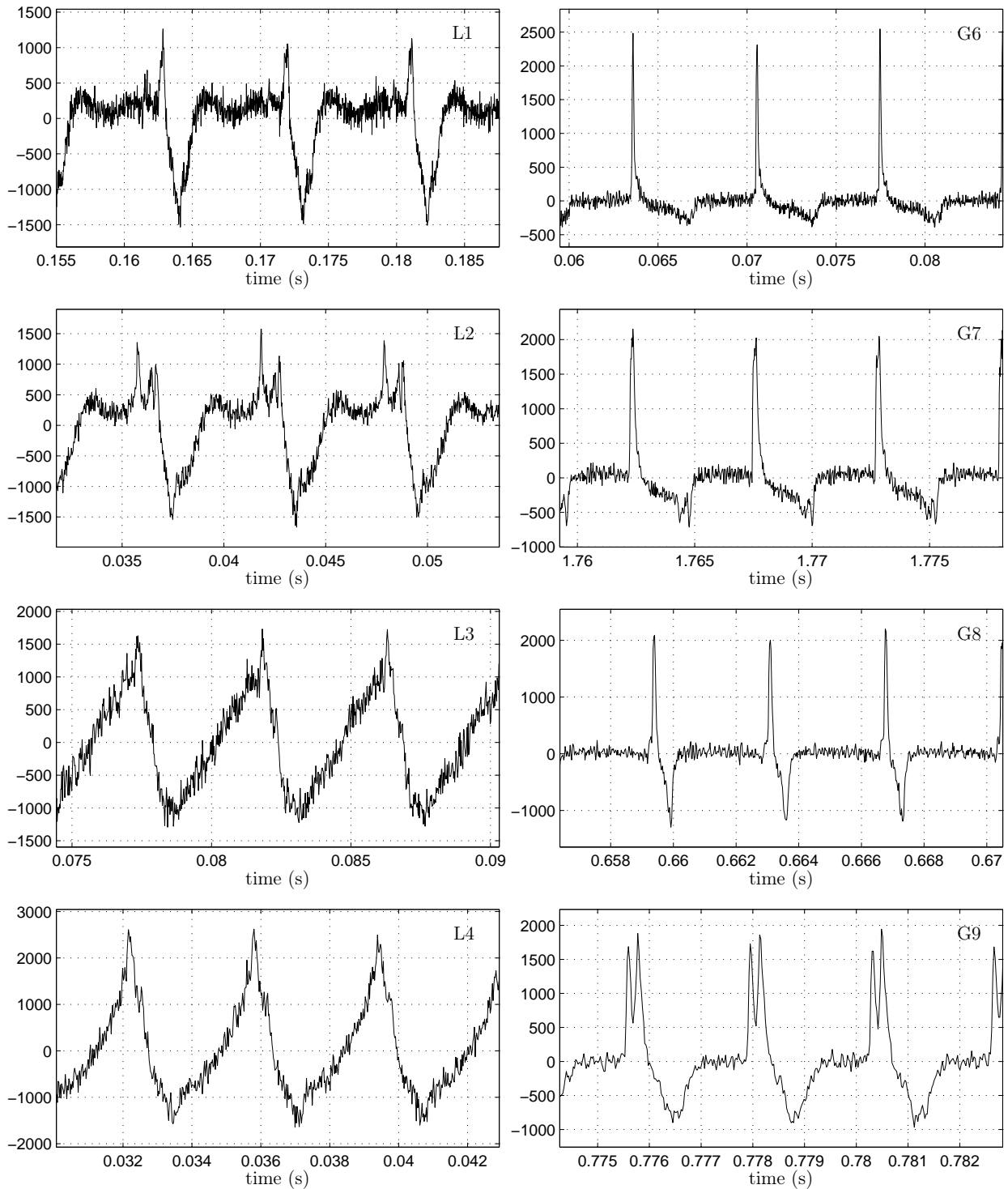


FIGURE 4.6 – Electrical impedance derivative DEI signals of the lips (left column) and the vocal folds (right column), corresponding to EI waveforms presented in Fig. 4.5.

4.4.3 Lip and vocal-fold OA and DOA signals and parametric representation (Obs. 4 & Prop. 1)

Fig. 4.7 presents eight representative excerpts of OA signals from L1 to L4, G1 and G3 to G5 measurements.

As expected, these plots clearly confirm that the Oq value is significantly higher for the lips than for the glottal folds, and therefore validates the use of the DECOM Oq estimator on ELG signals. The categorisation of the lip vibration according to the fundamental frequency is once again obvious. Low frequency and high frequency lip OA signals tend to have on opposite asymmetry, as already seen in the EIs. Moreover, the difference in the roughness of the closure appears very clearly on the OA signals. The closing phase (transition between decreasing and null OA values) is abrupt and angular in the voice case while it is much smoother in the lips case, except maybe for L3. Finally, the irregular closure of lips is visible in the decreasing phase of OA signals, especially in L1 and L2.

Despite the huge variability of the extracted OA waveforms, we observed that most of the signals show a rather similar periodic pulse shape. A very common pulse model used in voice production models is the Liljencrants-Fant model (LF model) [FLL85]-[Fan95]. This model is described in its derivative form by

$$s'(t) = \begin{cases} s'_1(t) & 0 \leq t \leq t_e \\ s'_2(t) & t_e < t \leq t_c \\ s'_3(t) & t_c \leq t < T \end{cases}, \quad (4.1)$$

where

$$s'_1(t) = -E e^{\alpha(t-t_e)} \frac{\sin(\pi \frac{t}{t_p})}{\sin(\pi \frac{t_e}{t_p})}, \quad (4.2)$$

$$s'_2(t) = -\frac{E}{\varepsilon t_a} \left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)} \right), \quad (4.3)$$

$$s'_3(t) = 0. \quad (4.4)$$

It is driven by one temporal parameter T (the period), one amplitude parameter E and four shape parameters $\{t_p, t_e, t_a, t_c\}$. Two parameters α and ε are determined by the implicit equations

$$\lim_{t \rightarrow t_e} s'_2(t) = s'_1(t_e) \text{ and} \quad (4.5)$$

$$\int_0^T s'(t) dt = 0. \quad (4.6)$$

Its general shape is presented on Figure 4.8.

Initially conceived as a glottal flow model, this model appears to be relevant as an OA model for the glottis case, as shown in [HHCD12]. We have developed an estimation method of the LF parameters for DOA signals. The goal of this algorithm is to estimate the LF parameters which provide the best temporal estimation of a target DOA signal.

Supposing that the target signal contains N periods, the estimation algorithm consists of N steps :

- rough estimation of the parameters,
- $1 \leq n \leq N - 1$: optimisation of the LF parameters of periods n and $n + 1$.

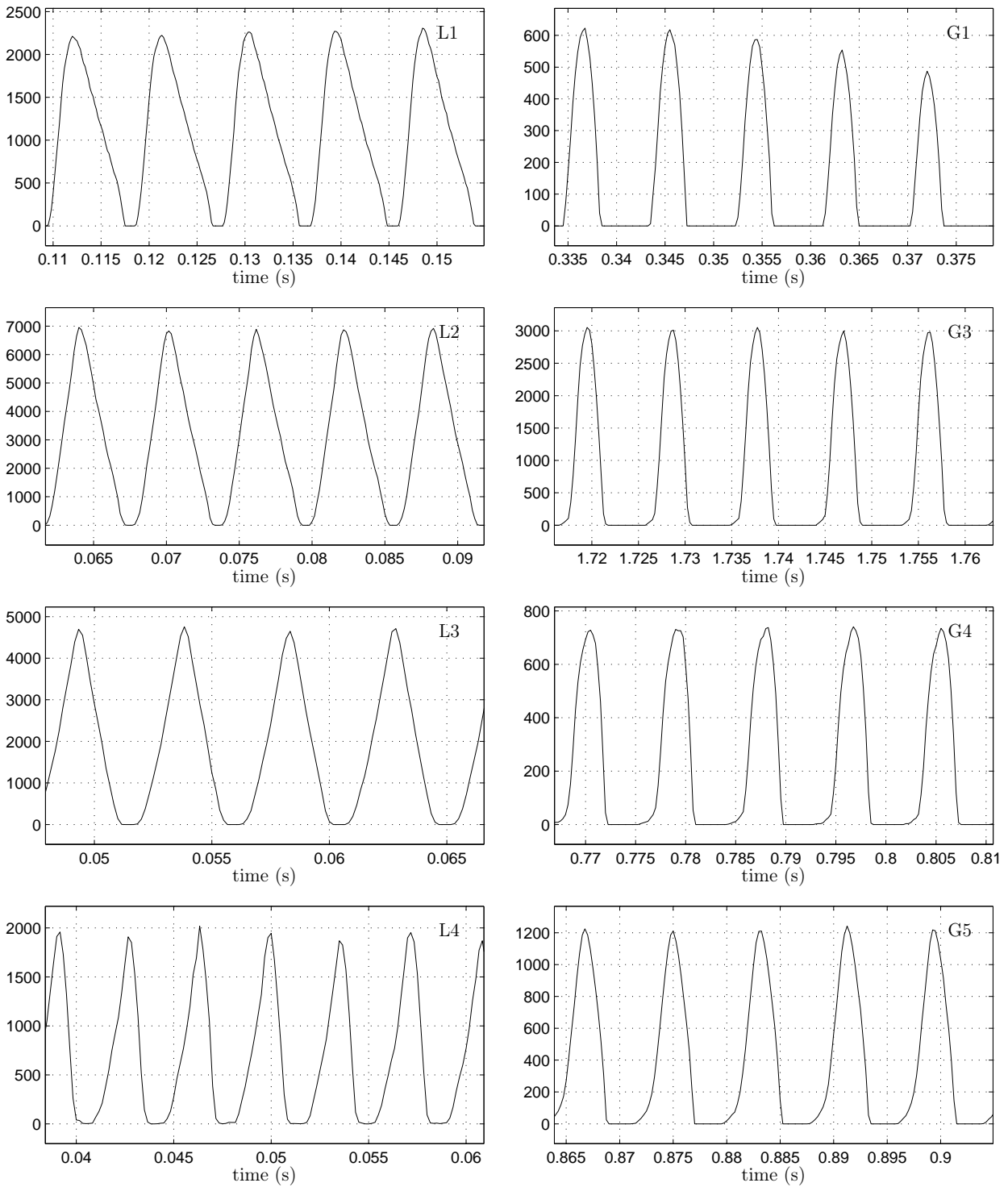


FIGURE 4.7 – Open area (OA) signals of the lips (left column) and the vocal-folds (right column).

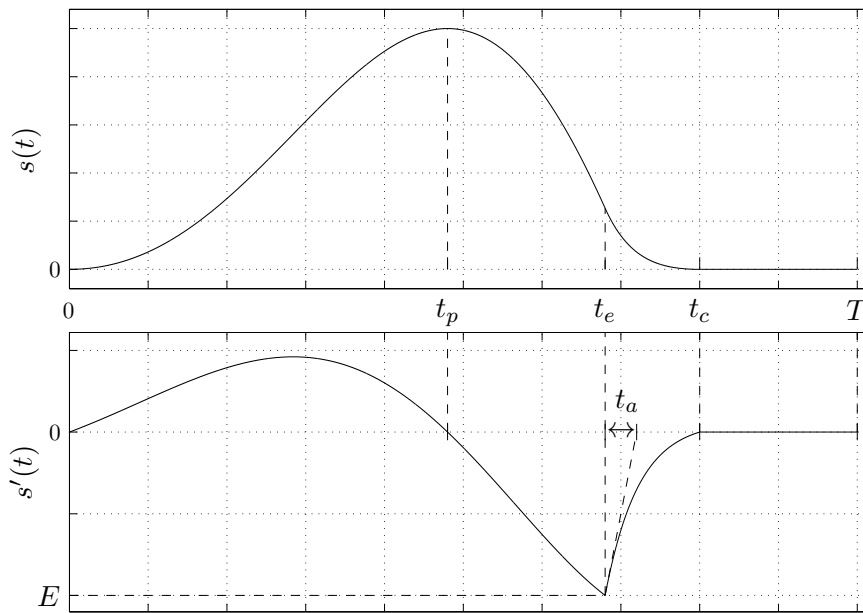


FIGURE 4.8 – One period of LF model. This model has one temporal parameter T (the period), one amplitude parameter E and four shape parameters $\{t_p, t_e, t_a, t_c\}$.

	LF with t_c	LF without t_c
ε	0.2070	0.2097
T_{comp}	9.40 sec	7.37 sec

TABLE 4.1 – Performances of the two versions of the LF parameters estimation algorithm (complete LF model and LF model without parameter t_c). T_{comp} is the normalized computation time for one period.

The first step is achieved thanks to the event detection method described in [KAZ00]. After event detection (which determines N) has been performed, the global maximum and the global minimum are localised between each pair of adjacent events. Then, the zero crossing points are detected between adjacent minima and maxima. The set of detected minima, maxima and zero crossing points give access to a rough estimation of every LF parameter for each period. The N following optimisations are performed using the simplex method [LRWW98], the criterion being the quadratic distance between the temporal model $s'(t)$ and the DOA target.

Two versions of the algorithm have been considered. The first one models the target with the complete LF model, the second one removes the parameter t_c , i.e. fixes $t_c = T$ or removes the exactly null phase at the end of each period. Table 4.1 presents the mean quadratic error ε and computation time T_{comp} for LF parameter estimation on a set of DOA signals extracted from the corpus USC_2008_02, containing approximately 1500 periods with various voice qualities. As it can be seen in Table 4.1, removing the t_c parameter offers a significant gain in computation time for a negligible performance loss. Moreover, we observed that the parameter t_c is often mis-estimated. Hence, the LF model without the parameter t_c will be considered in our study.

Figure 4.9 presents some DOA signals with the results of this estimation method and Table 4.2 indicates the quadratic error ε between the LF model obtained by the algorithm and the target DOA signals for each measurement of the two corpora. One can clearly see that, if the LF model

	L1	L2	L3	L4	G1	G2	G3	G4	G5
f_0	109.1	165.3	223.5	276.6	108.9	109.5	110.3	114.1	123.4
ϵ	0.454	0.499	0.479	0.469	0.400	0.327	0.337	0.294	0.356
Oq	0.8521	0.7278	0.6050	0.6209	0.3967	0.4301	0.4473	0.5029	0.4319
n_{cl}	1.000	3.100	1.918	1.289	1.549	1.000	1.000	1.000	1.003
n_{op}	1.167	1.102	1.721	1.087	1.235	1.000	1.028	1.752	1.088

TABLE 4.2 – Numerical results : fundamental frequency f_0 (Hz), quadratic error for LF-estimation ϵ , open quotient Oq , mean number of closing peaks during one closing phase n_{cl} , mean number of opening peaks during one opening phase n_{op} .

is relevant for glottis DOA signals, it does not seem appropriate when applied to lip DOA signals. Hence, new models must be developed for the lip case. However, the results on voice show promising benefits for parametric sound analysis and synthesis, as presented in [HHCD12].

Observation of DOA signals from the database, as well as of the short representative segments in Fig. 4.9, highlights significant differences between lip and glottis OA waveforms. The minima of the DOA signals, corresponding to maxima of the closing velocity, is very precise and well temporally defined in the glottis case, whereas in all lips measurements but L4, a noisy relatively flat phase appears around the minimum, corresponding once again to the irregular closure of the lips. This property of the lip motion may predominantly explain the poor performance of the LF model in representing lip DOA signals. This may also be emphasized by ripples in the positive phase of lip DOA signals, which cannot be represented by the LF model.

4.4.4 Geometrical parameters of lip and glottis opening areas (Obs. 2)

Two geometrical parameters derived from the 2D-OA signals are introduced in order to describe efficiently the main characteristics of the 2D-OA with only two signals. For each frame, the smallest rectangular box containing the detected opening area is computed and the two dimensions of the box are extracted. The largest dimension of the box is noted D and the smallest dimension is named d as illustrated in Fig. 4.10.

D and d are represented as a function of OA for two lip signals (L1 and L3) and two glottis signals (G2 and G5) in Fig. 4.11. This representation allows the influence of the two main dimensions of the opening area to be evaluated with respect to OA. In other words, these figures may be useful tools in order to detect zipper-like closures, as well as abrupt closures occurring on the whole length of the lips or vocal folds.

Overall, relatively similar relationships are observed for both valve systems. The relationship $OA \mapsto d$ is relatively linear. On the contrary, the $OA \mapsto D$ shows a two-linear-segment shape, a first short segment of high slope and a second segment of much smaller slope. This feature is particularly visible in the glottis signals (G2 and G5) and can be related to an abrupt opening and closure of the vocal folds. Indeed, during the opening and closing phases, the two vocal folds move rapidly toward or away from each other, as depicted by the first short segment of high slope in $OA \mapsto D$. During the open phase, vocal folds are not in contact and the variation of OA is mainly caused by variations in d , as shown by the second almost constant segment in $OA \mapsto D$.

This behaviour is not as striking for the lips (L1 and L4). In lowest tones (L1), the relationship $OA \mapsto D$ tends to show a similar two-segment structure as observed in the voice. However, for

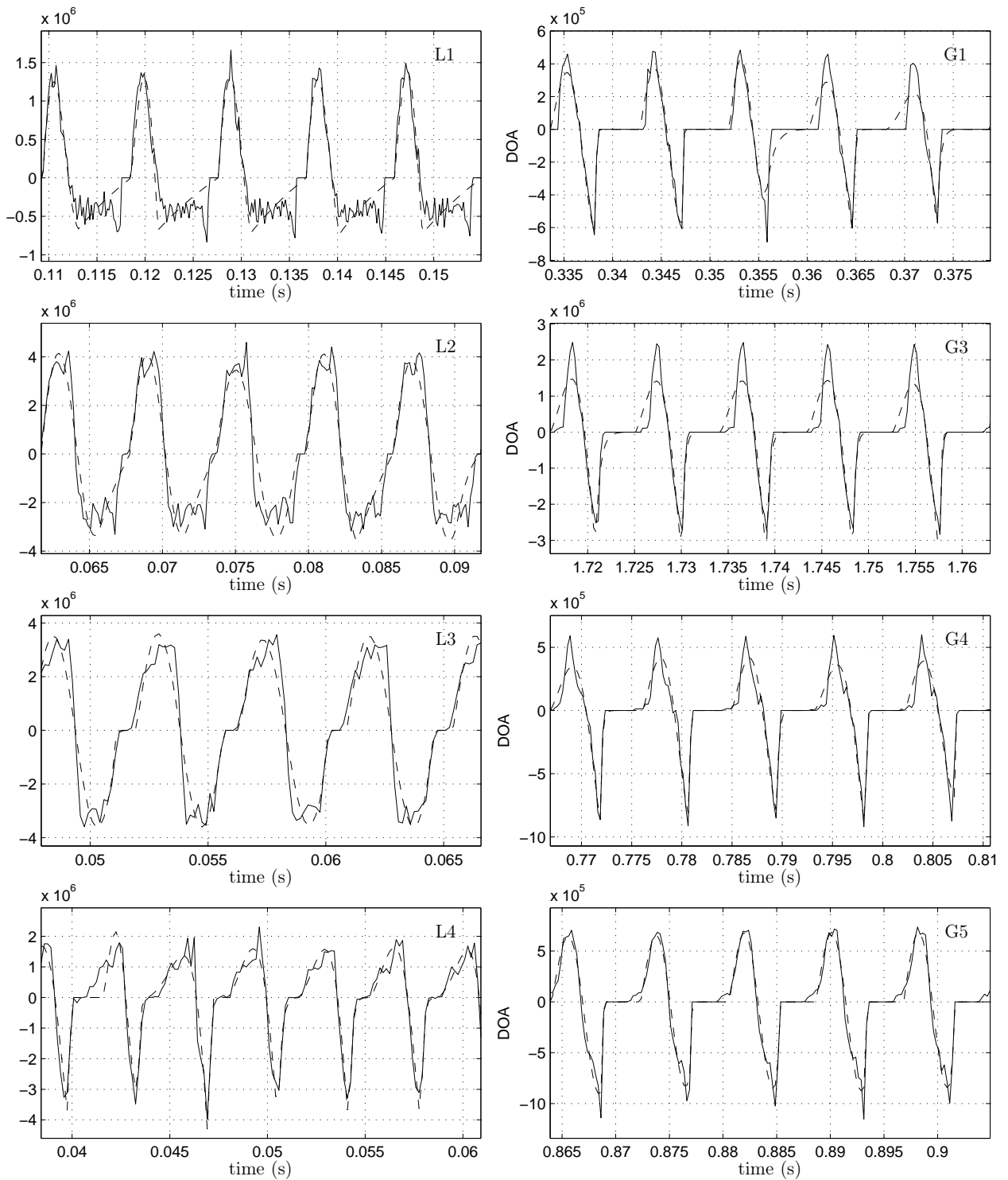


FIGURE 4.9 – LF estimation examples. (—) measured DOA signal, (---) LF model (without t_c).

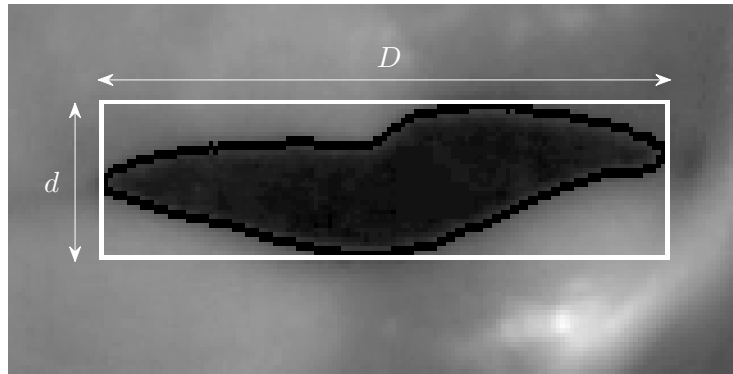


FIGURE 4.10 – Geometrical parameters D and d on a lip frame. Detected 2D-AO is represented in black.

the highest tones (L4), the $OA \mapsto D$ and $OA \mapsto d$ relationships are closer, indicating a balanced contribution of D and d to the opening area OA , possibly correlated to the zipper-like opening and closure of the lips. This logarithmic shape of the $OA \mapsto d$ relation is in agreement with previous quantitative analysis of the relationship between both variables in artificial and real trombone players' lips [BCG10].

4.4.5 Opening geometries and electrical impedance signals (Obs. 2)

The relationship between EI and OA in eight excerpts of lips and glottis vibration measurements is represented in Fig. 4.12.

As predicted by $OA \mapsto D$ and $OA \mapsto d$ relationships, $EI \mapsto OA$ relations are marked by strong hysteresis behaviours, apparently uncorrelated with the playing frequency and particularly emphasized in the case of the lips. The two-step sequential relationship $OA \mapsto D$ observed for the vocal folds in Fig. 4.11 is clearly visible in the (EI,OA). It indicates a predominant influence of D on EI for the glottis. Unlike vocal-fold results, this predominant influence of one geometrical dimension on the EI signal is not apparent for the lips. Nevertheless, this observation is in agreement with previous evidence of zipper-like lip closures and openings in trombone performance.

In these figures, it clearly appears that OA maximum does not coincide with EI minimum, especially for the lips. Added to the hysteresis already mentioned, it highlights the lack of information provided by the EI signal during the open phase. Nevertheless, in the voice case, the steadiness of the OA corners tends to reduce the hysteresis and to establish a more systematic synchronicity between maximum of the OA and minimum of the EI.

We verify here, as it can be seen on the videos, that the EI signal brings information during the completely closed phase which cannot be interpreted with the present tools. We strongly believe that this information is precisely related to the zipper-like opening of the vocal folds along their thickness in the voice case and to the upward-backward motion of the lips in the brass instrument case, but, for now, we do not have access to technologies which will help us verify this hypothesis.

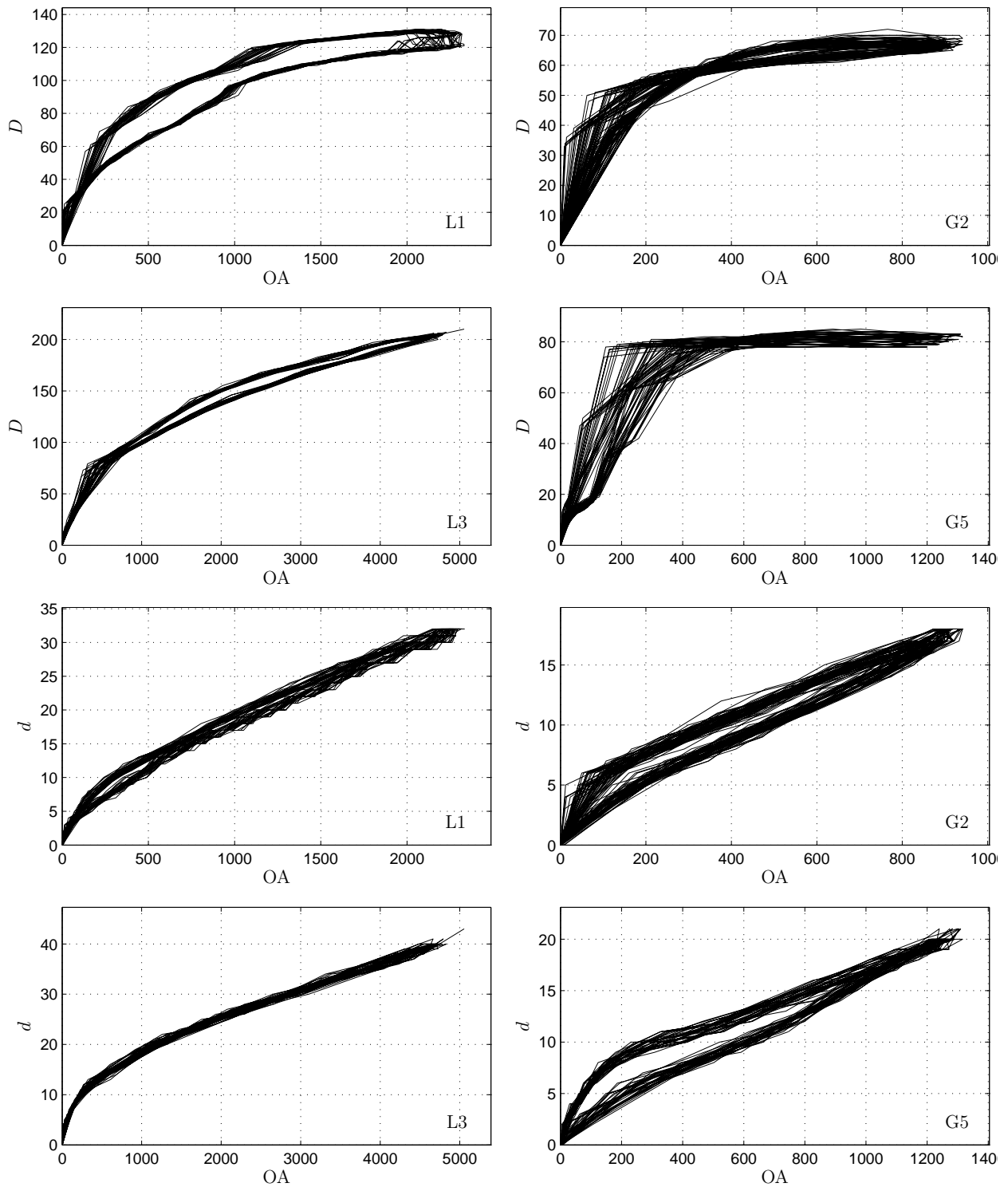


FIGURE 4.11 – Opening Area OA versus geometrical parameters D (two top lines) and d (two bottom lines) for lips (left column) and glottis (right column). OA, D and d are in pixels.

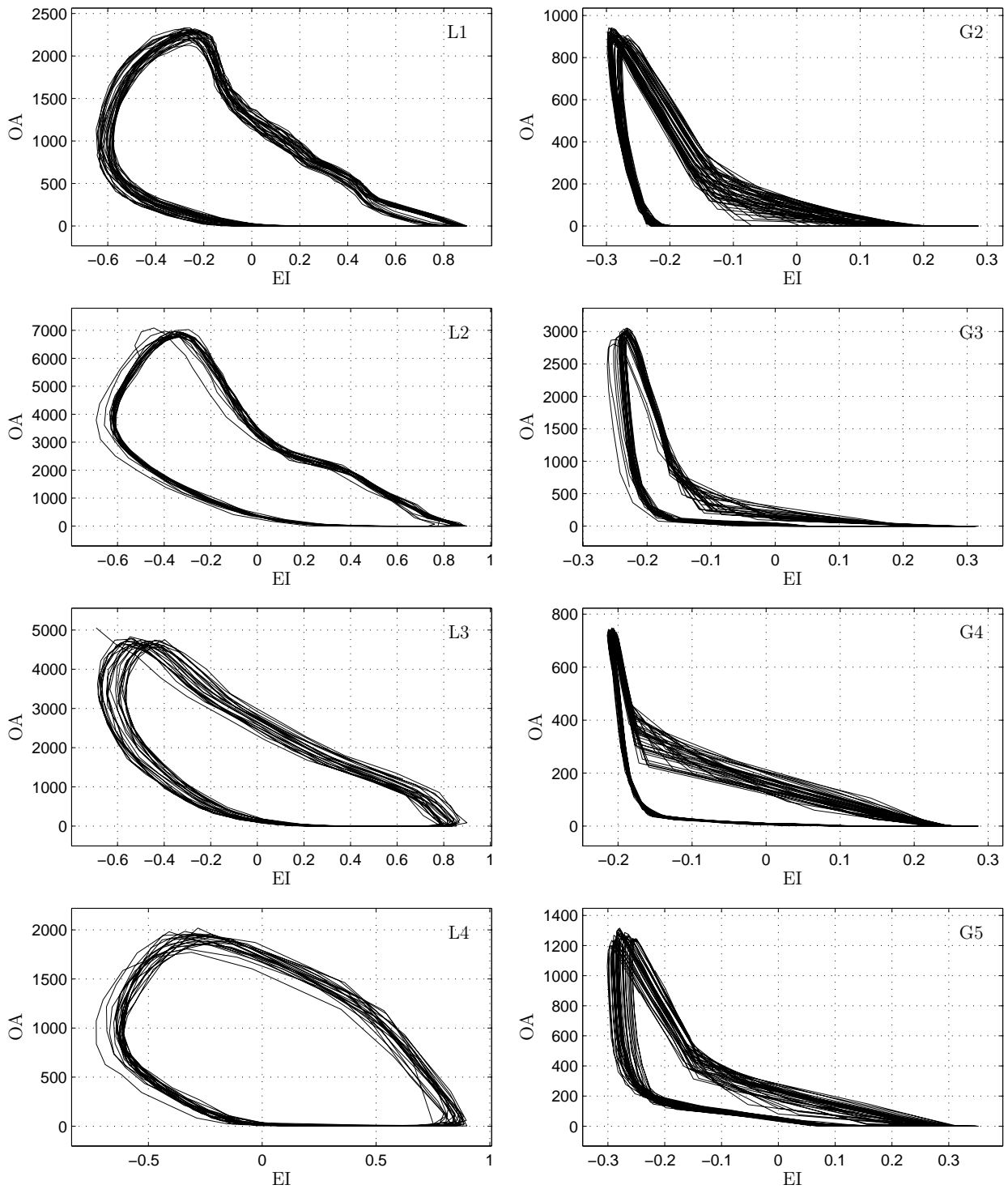


FIGURE 4.12 – Electrical impedance EI versus opening area OA. Left column : lips, right column : glottis. EI is normalized, OA is in pixels.

4.5 Conclusion

An *in vivo* and synchronous measurement and analysis framework was presented for the comparison of the vibratory mechanisms in oscillating vocal folds and lips. A visualisation tool was developed to allow the synchronous examination of the multimodal data. Moreover, this visualisation tool presents some obvious interests for medical uses, as well as vocal and brass instrument pedagogy. Analyses of the experimental data was conducted with respect to : 1) comparisons of EI and DEI signals, 2) parametric LF modelling of the DOA signals in both lips and glottis, 3) geometrical characteristics of lip and glottis opening areas, 4) correlations between OA geometrical parameters and EI signals ; and lead to the following results.

(R1) The transposition of the EGG device to the ELG device for monitoring the lips of a brass instrument player by Fréour and Scavone [FS11, FS12] proved to be efficient and able to deliver interpretable signals.

(R2) Analogously to the vocal folds, lips EI signals can be categorised according to their fundamental frequency. They show more complex waveforms for the lowest tones (analogously to a M1 laryngeal mechanism), while displaying more sinusoidal shapes in the highest ones (such as a M2 glottal mechanism), corroborating previous evidence of a more sinusoidal motion of the lips with increasing pitch [Mar42, CS96, YM03, BCG10]. Moreover, lip EI signals show much activity during the open phase, whereas EGG signals are rather constant. This could be related to emphasized variations in OA corners positions during opening and closing phases for the lips. The differentiation observed in lips EI signals according to the playing frequency is well visible in DEI waveforms ; a transition from a rectangular to triangular waves is observed with increase in pitch.

(R3) Analogously to the glottis, lips DEI signals show closing and opening peaks, especially visible in low frequencies and including occurrences of multiple closing and opening peaks. This similarity makes possible the use of standard EGG features extractors on DELG signals, especially Oq estimators. Due to the frequent occurrence of multiple opening and closing peaks in lip DEI signals, classical Oq estimators [RM88, HLA90, How95] are not able to provide a proper measurement of Oq for lips EI signals. However, the DECOM algorithm [HDDC04] is able to detect and enumerate the multiple opening and closing peaks in DEI signals and take them into account for the Oq estimation. Thus, this latter algorithm allows to properly estimate the open quotient for either EGG or ELG signals, which is logically revealed to be much higher for the lip case than for the glottis.

(R4) The HSV recordings and the 2D-OA extraction algorithms offer the ability to compare glottis and lips 2D-OA and OA signals. Glottis OA signals show good agreement with the parametric LF pulse model, whereas less satisfactory results are obtained with lip OA signals. As shown in [HHCD12], this result gives interesting perspectives for the analysis and synthesis of voice sounds.

(R5) Geometrical parameters extracted from OA signals reveal an overall linear correlation between opening large dimension and OA in both lips and glottis. On the contrary, the influence of the opening small dimension on OA is more emphasized in the lips than in the glottis. This is in agreement with variations in OA corners position occurring for the lips and in a much lesser extent for the glottis, which can explain the poor results of the LF model representation on lips DOA signals in **(R4)**.

This work opened several important perspectives. At first, extended corpora with more recordings would allow a statistical study to be conducted on the different measurements and their relationships for lips and glottis. Moreover, a low/high frequency categorisation of the lip motion

seems to appear from the EI and OA signals but could not be studied in a statistically meaningful way without an extended corpora of lip measurements. Moreover, the influence of dynamic was not investigated in this study. Although brassy regimes do not seem to induce significant changes in the behaviour of the lip opening area [SCB⁺09], further investigations on the relationship between loudness and lip vibration could be the object of future work.

In conclusion, as it has been mentioned several times in this article, the main limitation of the exploration of the lips and glottis are the impossibility to monitor the motion within the thickness dimension, perpendicular to the lip or glottis plane. We saw in this article that the roughness and (ir)regularity of the closure has a drastic impact on the data and can be easily detected and characterized with EI signals. However, on one hand, it has been shown that the lips can have an upward/backward motion and that the glottis (at least in M1 mechanism) has a zipper-like opening along its thickness [CK85]. On the other hand, EI signals show much activity during the completely closed phase. The authors strongly believe that this activity is mainly due to the motion of lips and glottis along the thickness dimension, to which high-speed video cannot give access. A new multichannel EGG/EIT device [HHD⁺12] is currently in development and we have good hope that this equipment will give access to these motions.



Chapitre 5

Imagerie électrique multicanal du larynx

Sommaire

5.1	Tomographie d'impédance électrique	71
5.2	Test préliminaire de validation du principe de l'EIT sur une maquette de larynx simplifiée	73
5.3	Conception d'un dispositif mixte EGG/EIT multicanal à géométrie variable	75
5.3.1	Mesures mixtes EGG/EIT	75
5.3.2	Multiplexage fréquentiel et multiplexage temporel	75
5.3.3	Électrodes à géométries d'activation configurables	75
5.3.4	Vue d'ensemble et mise en équation	75
5.3.5	Choix et dimensionnement	78
5.4	Développements et réalisations techniques	78
5.4.1	Vue générale du dispositif	78
5.4.1.1	Identification des contraintes et des problèmes techniques	78
5.4.1.2	Architecture du dispositif et état d'avancement	79
5.4.2	Éléments matériels	80
5.4.2.1	Génération de signaux isolés galvaniquement	80
5.4.2.2	Routeur analogique	81
5.4.2.3	Accès protégé aux électrodes	83
5.4.2.4	Traitement synchronisé des ordres de configuration et des mesures	84
5.4.3	Premier élément logiciel et premiers tests	85
5.5	Conclusion et perspectives	86

Ce chapitre est consacré à l'élaboration d'un dispositif d'imagerie électrique multicanal du larynx. Nous avons vu dans le chapitre 3 que l'EGG est une technique d'imagerie simple à mettre en œuvre et qu'elle donne accès à des informations sur l'activité laryngée sans être invasive. Cependant, comme il l'a été souligné dans le chapitre 4, cette information est partielle et même les dispositifs expérimentaux récents (comme celui de Kob et Frauenrath [Fra06, KF09]) ne parviennent pas, à l'heure actuelle, à donner plus qu'une information monodimensionnelle sur le mouvement tridimensionnel des plis vocaux.

La *Tomographie d'Impédance Électrique* (TIÉ ou EIT pour *Electrical Impedance Tomography*, anciennement appelée *Applied Potential Tomography*) est une technique d'imagerie électrique multicanal [HW78, BB84, Hol10]. Contrairement à l'EGG qui consiste à estimer la conductance globale d'un corps (le larynx) entre deux points, le principe de l'EIT est de reconstruire la carte de conductivité interne d'un corps soumis à une excitation électrique entre deux points de sa surface, à partir de la carte de potentiel mesurée à sa surface. La technique de l'EIT est présentée succinctement dans la section 5.1. Cette technique est utilisée en imagerie médicale notamment pour (i) le monitoring pulmonaire [Fre00], (ii) la mesure de l'activité cérébrale [BBP+78, BLB+03], (iii) la détection de tumeur cancéreuse [WTMD01, ZG03], et (iv) l'exploration gastrique [MBA+87]. Elle n'a, à ma connaissance, jamais été utilisée dans le cadre de la mesure de l'activité glottique mais des résultats préliminaires ont été obtenus dans le cadre

- de l'imagerie de la déglutition [HLG+94, HLG+96], et
- d'une étude *in silico* de l'EIT pour la détection de l'ouverture glottique [SNK+11].

Cette dernière étude conclue que l'EIT effectuée à l'aide d'un collier d'électrodes disposé à la surface du cou devrait être sensible aux variations de l'ouverture glottique.

Après un test de validation du principe de l'EIT sur une maquette de larynx simplifiée dans la section 5.2, je propose dans la section 5.3 un dispositif mixte EGG/EIT multicanal à électrodes à géométrie variable pour la mesure de l'activité glottique. L'élaboration de ce dispositif a été initiée dans le cadre du projet VoiceTronics (cf. annexe A), sous l'impulsion de Thomas Hélie, Nathalie Henrich et moi-même. Les travaux préliminaires des étudiants du projet ont permis d'établir un cahier des charges précis. La construction d'un premier prototype a été entamée lors du stage d'Antonin Mauffrey, co-encadré par Thomas Hélie, Emmanuel Fléty et moi-même à l'IRCAM. Une partie du texte et des figures de ce chapitre proviennent de son rapport de stage de fin d'études d'ingénieur [Mau13]. Ce prototype, à vocation expérimentale, est fondé sur trois ingrédients.

Dispositif mixte EGG/EIT Des modules de mesure de type EGG (mesure de courant) et EIT (mesure de potentiel) sont construits pour permettre d'expérimenter une combinaison des deux techniques.

Multiplexage temporel et fréquentiel Le multiplexage fréquentiel permet un accès simultané à plusieurs canaux tout en préservant la bande passante. Le multiplexage temporel permet d'envisager un balayage des électrodes disposées sur le cou.

Électrodes à géométrie variable Les électrodes peuvent (i) être ré-assignées aux modules d'excitation et de mesure pendant l'expérience et (ii) être regroupées afin de dessiner des motifs d'activation.

Des éléments concernant la réalisation technique du prototype, les contraintes et difficultés rencontrées et les solutions proposées, ainsi qu'un bilan de l'état d'avancement du prototype sont présentés succinctement dans la section 5.4. Le lecteur intéressé pourra se reporter au rapport de stage d'Antonin Mauffrey [Mau13].

5.1 Tomographie d'impédance électrique

Les premiers travaux sur la tomographie d'impédance électrique sont en général attribués à Barber & Brown et leur article fondateur de 1984 [BB84], bien que des premières traces de cette technique apparaissent dès 1978 dans les travaux de Henderson & Webster [HW78]. Depuis cette époque, la tomographie d'impédance électrique est un domaine de recherche en perpétuelle évolution. Parmi la bibliographie extrêmement dense sur le sujet, un ouvrage de référence récent est [Hol10]. En complément de cet ouvrage, les articles de revue [CIN99, Bor02] permettent d'aborder plus rapidement la théorie et la technique de l'EIT.

Le principe de l'EIT est la reconstruction d'une carte de conductivité interne d'un corps soumis à un courant entre deux points de sa surface à partir de la mesure de la carte de potentiel à sa surface. Ce problème se formalise sous la forme du problème inverse de Calderón [Cal80, Uhl09]. Soit $\Omega \in \mathbb{R}^n$ un domaine borné à frontière $\partial\Omega$ régulière, $\sigma(x)$ la conductivité au point $x \in \Omega$ et \vec{n} la normale à la surface du domaine (cf figure 5.1).

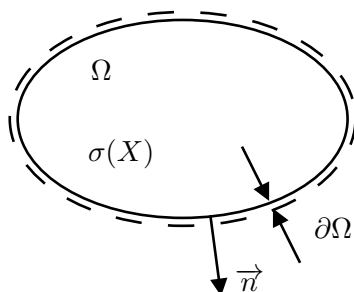


FIGURE 5.1 – Problème de Calderón.

En l'absence de source de courant à l'intérieur de Ω , le potentiel électrique $u(x)$ est gouverné par la loi de Kirchoff continue

$$\nabla \cdot [\sigma(x)\nabla u(x)] = 0 \quad \forall x \in \Omega. \quad (5.1)$$

La densité de courant $j(x)$ sur la frontière $\partial\Omega$ s'écrit grâce à la loi d'Ohm continue

$$j(x)|_{\partial\Omega} = -\vec{J} \cdot \vec{n} = \sigma(x)\nabla u(x) \quad \forall x \in \partial\Omega, \quad (5.2)$$

où \vec{J} est le vecteur densité de courant sur Ω . De plus, la conservation de la charge implique que

$$\int_{\partial\Omega} j(x) = 0, \quad (5.3)$$

et nous définissons une référence de potentiel, par exemple en écrivant

$$\int_{\partial\Omega} u(x) = 0. \quad (5.4)$$

Le système (5.1)-(5.2)-(5.3)-(5.4) définit le *modèle continuum* de l'EIT. Le résultat de Calderón [Cal80, Uhl09] stipule que l'on peut retrouver la valeur de $\sigma(x)$ pour ce modèle à partir de la carte de potentiel à la surface $u(x)|_{\partial\Omega}$ (pour des conditions aux limites de Dirichlet) ou bien de la densité de courant à la surface $j(x)|_{\partial\Omega}$ (pour des conditions aux limites de Neumann).

La pratique de l'EIT s'éloigne des conditions du problème inverse de Calderón. Par l'emploi d'électrodes pour injecter du courant entre deux points de la surface et pour mesurer le potentiel sur la surface, la connaissance de la condition aux limites est forcément incomplète (le potentiel n'est pas connu en la totalité des points de la surface, cf figure 5.2). De plus, les électrodes ne sont pas ponctuelles et il est nécessaire de modéliser leur comportement électrique (distribution de potentiel et de courant sur la surface de l'électrode, impédance de contact etc.). Ces connaissances partielles et la difficulté de modéliser précisément le contact électrode-peau rend le problème de l'EIT sévèrement mal posé, en raison de sa grande sensibilité au bruit et aux erreurs de mesure. Néanmoins, des stratégies permettent de régulariser le problème grâce à des techniques de régularisation (Tikhonov, moindres carrés, optimisation itérative etc. voir [Hol10, chap. 1]). Une piste complémentaire pour régulariser ce problème inverse, peu exploitée dans les applications d'EIT, est l'utilisation de modèles paramétriques de conductivité $\sigma(x)$.

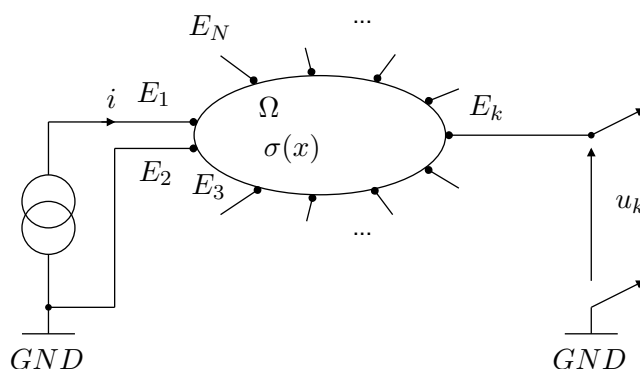


FIGURE 5.2 – Principe de la tomographie d'impédance électrique.

Des applications de l'EIT ont été proposées pour

- le monitoring pulmonaire [Fre00],
- la mesure de l'activité cérébrale [BBP⁺78, BLB⁺03],
- la détection de tumeur cancéreuse [WTMD01, ZG03, HGG12], et
- l'exploration gastrique [MBA⁺87].

L'EIT n'a jamais, à ma connaissance, été appliquée à l'imagerie de l'activité glottique, mais deux travaux en lien avec cette application sont à noter. Premièrement, Hughes et al. sont parvenus à détecter le passage d'un bol alimentaire à travers le pharynx à l'aide d'un collier d'électrode placé sur le cou [HLG⁺94, HLG⁺96]. Ces travaux montrent que les mouvements du larynx et du pharynx peuvent être détectés grâce à la technique de l'EIT.

Deuxièmement, Seppänen & al. ont réalisé une simulation *in silico* de la technique de l'EIT appliquée à la mesure de l'activité laryngée [SNK⁺11]. La figure 5.3 présente des cartes de conductivité en coupe transverse au niveau du plan glottique, simulées et reconstruites à l'aide d'un modèle complet (incluant un modèle d'électrodes), extraites de [SNK⁺11]. Comme nous pouvons le voir sur ces images, ces travaux concluent que l'EIT devrait être sensible à l'activité glottique.

Notons de plus que, si la plupart des applications de l'EIT s'intéressent à des reconstructions bidimensionnelles, de nombreux travaux se sont penchés sur la question de la reconstruction 3D pour l'EIT, en théorie et en pratique [VVS⁺99, BIN00, DSH⁺05, BHW⁺09]. Enfin, des travaux ont montré la faisabilité et l'intérêt de l'EIT par multiplexage fréquentiel [RGV92, BSM01, GIR07].

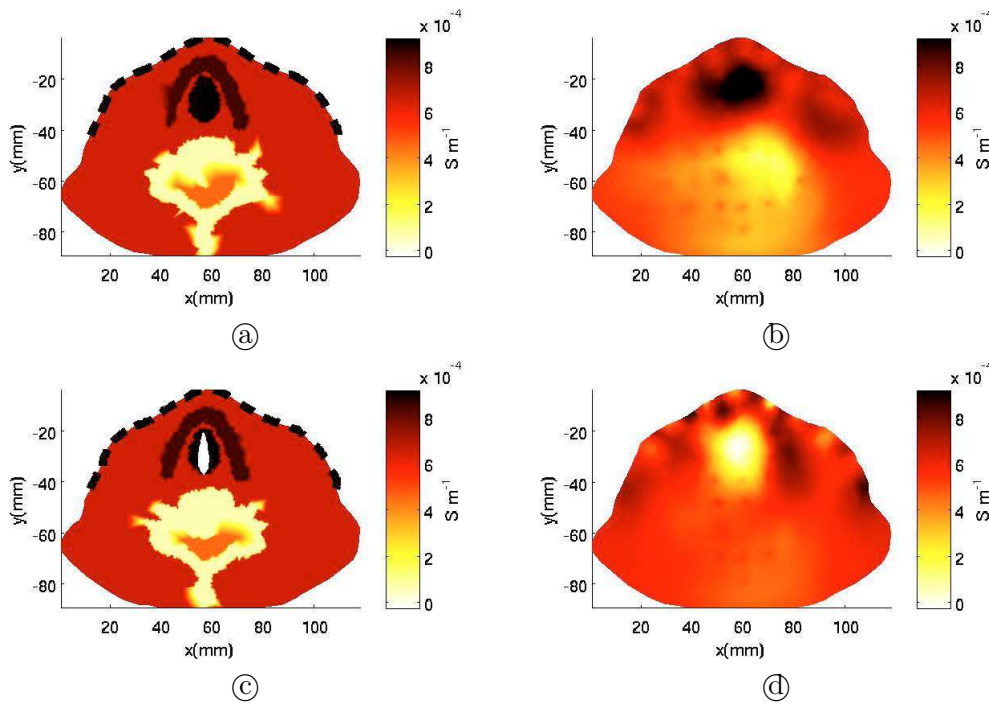


FIGURE 5.3 – Simulation de l'EIT appliquée à la mesure de l'activité glottique : carte de conductivité en coupe transverse au niveau du plan glottique, d'après [SNK⁺11]. (a)-(b) simulation et reconstruction en glotte fermée. (c)-(d) simulation et reconstruction en glotte ouverte.

5.2 Test préliminaire de validation du principe de l'EIT sur une maquette de larynx simplifiée

D'après l'étude bibliographique de la section précédente, les développements techniques et théoriques de l'EIT montrent qu'il est envisageable d'utiliser cette technique dans le cadre de l'imagerie des plis vocaux en vibration. Avant d'étudier la conception d'un dispositif, je propose un test préliminaire permettant de valider l'utilisation de l'EIT sur une maquette de larynx à échelle réelle et à géométrie simplifiée. Cette maquette a été réalisée dans le cadre du projet VoiceTronics pour ce test et pour la validation du dispositif présenté dans la suite de ce chapitre. Elle est constituée d'un hydrogel conducteur aux propriétés électriques proches de celles des tissus humains. Sa géométrie est simplifiée mais elle permet l'ouverture et la fermeture des plis vocaux sur une amplitude de quelques millimètres, comme dans le cas réel (cf annexe A pour plus de détails sur cette maquette).

Pour tester l'utilisation de l'EIT sur cette maquette, j'ai effectué deux mesures de type EIT en position statique : en glotte ouverte et en glotte fermée. Une photo du dispositif expérimental est présentée en figure 5.4. Un courant généré par un composant de type DDS (voir 5.4.2.1) est injecté à travers deux électrodes distantes d'environ 5 mm, au niveau du plan glottique, proche de la barre d'entraînement du pli vocal (figure 5.4-(b)). Deux électrodes A et B distantes d'environ 1 cm disposées à l'opposé (figure 5.4-(c)) permettent de mesurer les modifications de la différence de potentiel V_{AB} résultantes à la surface de la maquette. Cette mesure de tension est effectuée à l'aide d'un oscilloscope sur une entrée haute impédance (courant nul), comme dans le cas d'une mesure de type EIT (cf figure 5.2). Les résultats obtenus sont présentés sur le tableau 5.1. La chute de tension entre les électrodes de mesure lors de la fermeture des plis vocaux est faible mais détectable.

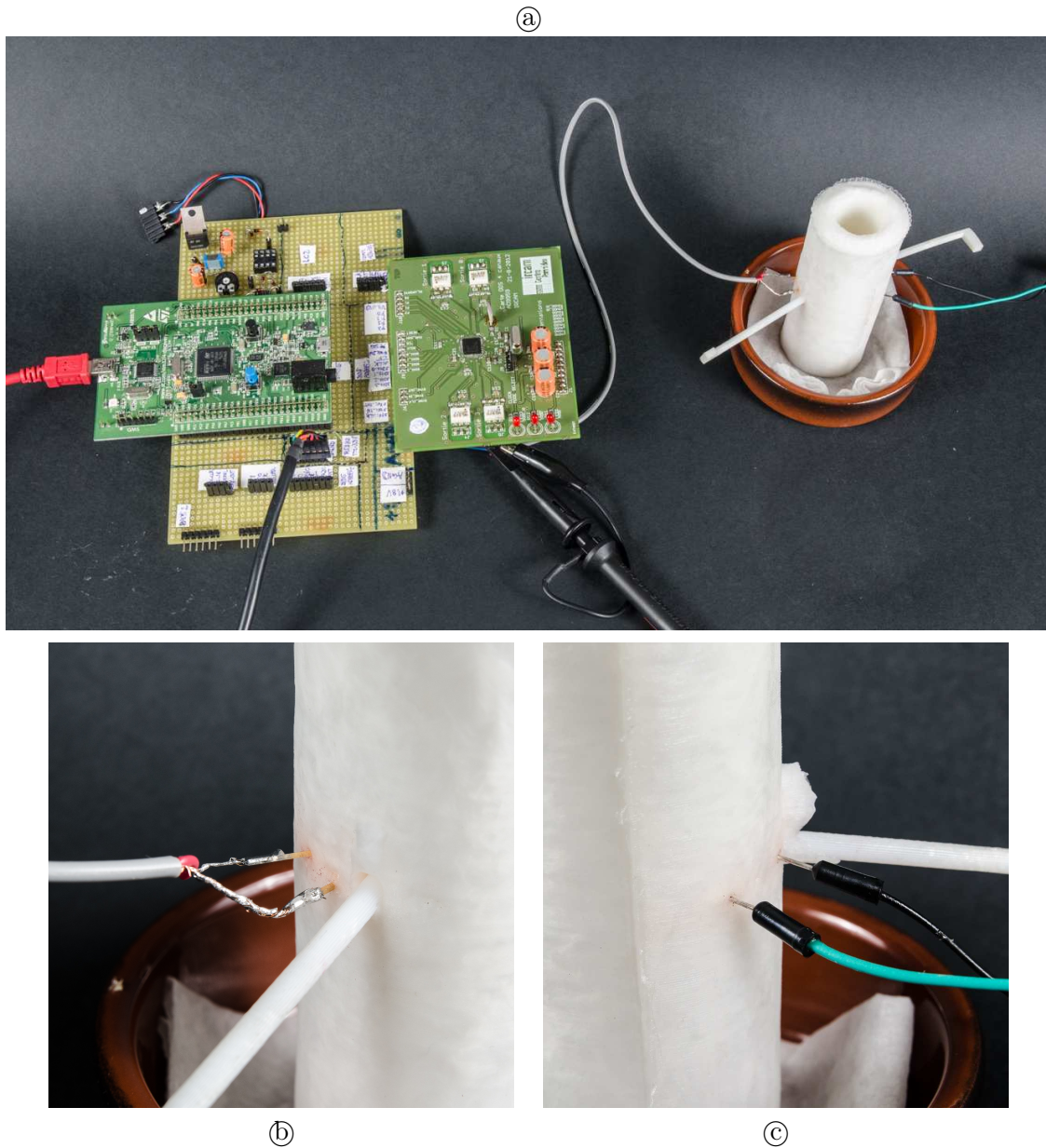


FIGURE 5.4 – Test qualitatif de validation du principe de l’EIT sur une maquette de larynx simplifiée. (a) dispositif expérimental. (b) électrodes d’injection de courant. (c) électrodes de mesure de tension.

	U_{inj}	U_{mes}
glotte ouverte	8.12 V	1.18 V
glotte fermée	8.12 V	1.15 V

TABLE 5.1 – Test qualitatif de validation du principe de l’EIT sur une maquette de larynx simplifiée : résultats numériques. U_{inj} est la tension crête-à-crête mesurée entre les électrodes d’injection de courant. U_{mes} est la tension crête-à-crête mesurée entre les électrodes de mesure de tension.

5.3 Conception d'un dispositif mixte EGG/EIT multicanal à géométrie variable

Dans le but de construire un appareil de mesure électrique capable de donner une information plus complète que les électroglottographes existants, je propose de concevoir un circuit capable d'exploiter à la fois la technique de l'EGG et de l'EIT. Le dispositif est aussi capable de faire du multiplexage temporel et fréquentiel et de modifier la géométrie de ses électrodes en cours de manipulation. De plus, ce système est conçu de façon modulaire afin de pouvoir ajouter ou retirer des éléments simplement et adapter le système selon les résultats des premières mesures.

5.3.1 Mesures mixtes EGG/EIT

Afin d'effectuer des mesures de type EGG et de type EIT, le système doit inclure deux types de modules de mesure. Des modules de mesure de courant, présentant une entrée basse impédance (résistance de mesure calibrée), sont présents pour effectuer des mesures de type électroglottographie. Des modules de mesures de tension, présentant une entrée haute impédance, sont utilisés pour les mesures de type tomographie d'impédance électrique.

5.3.2 Multiplexage fréquentiel et multiplexage temporel

L'appareil doit pouvoir réaliser un multiplexage fréquentiel afin de pouvoir exploiter plusieurs canaux d'excitation simultanément tout en préservant la bande passante des signaux mesurés. Pour ceci, plusieurs modules d'excitation indépendants et déconnectés doivent être présents. Le dispositif permet aussi de réaliser un multiplexage temporel à l'aide d'un routeur analogique qui relie les modules d'excitation et de mesure aux électrodes. Ce multiplexage temporel permet d'effectuer un balayage des électrodes ou de concentrer l'énergie sur une zone d'intérêt.

5.3.3 Électrodes à géométries d'activation configurables

Le routeur analogique permet de connecter les électrodes aux modules d'excitation et de mesure selon les deux règles suivantes :

- une électrode est soit déconnectée du système, soit connectée à un module au plus,
- un module peut être connecté à une ou plusieurs électrodes.

La première condition permet d'éviter les court-circuits, notamment entre deux modules d'excitation. La deuxième condition permet de dessiner des motifs d'activation sur les électrodes. La figure 5.5 illustre ce mode de fonctionnement.

5.3.4 Vue d'ensemble et mise en équation

Les points 5.3.1, 5.3.2, 5.3.3 nous conduisent au système décrit par la figure 5.6. Il communique avec un ordinateur qui commande le système et analyse les données de mesurées.

Ce système est composé de K modules d'excitation \mathcal{E}_k qui génèrent chacun un courant sinusoïdal à une fréquence différente, N modules de mesure de courant (type EGG) \mathcal{R}_n et L modules de mesures de tension (type EIT) \mathcal{V}_l . L'ensemble est relié, par l'intermédiaire d'un routeur analogique,

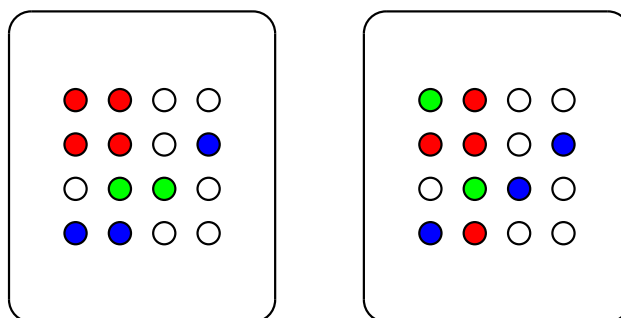


FIGURE 5.5 – Électrodes à géométrie variable. Les électrodes de même couleur sont reliées au même module afin de dessiner des motifs d’activation.

à M électrodes disposées à la surface du cou. L’ordinateur pilote les différents modules, le routeur analogique et le système de démodulation et d’échantillonnage des données de mesure. Il reçoit les données démodulées et échantillonnées pour les analyser.

Les modules d’excitation et de mesure sont présentés sur la figure 5.7. Les tensions $u_k(t)$, $v_n(t)$ et $V_l(t)$ sont mesurées (dispositif haute impédance), puis démodulées, échantillonnées et numérisées. Le démodulation est opérée en module et en phase par rapport à chaque porteuse (chaque signal d’excitation).

Chaque module d’excitation \mathcal{E}_k produit un courant électrique sinusoïdal, de la forme

$$i_k(t) = I_0 \cdot \sin(2\pi f_k t), \quad \text{où} \quad (5.5)$$

- I_0 est l’amplitude maximale, qui est identique pour tous les modules d’excitation,
- f_k est la fréquence porteuse propre au module \mathcal{E}_k .

Les modules de mesure de courant \mathcal{R}_n sont constitués d’une résistance de précision R_n . Cette résistance est dans un premier temps fixe et calibrée. Elle pourra à l’avenir être ajustable numériquement. Le courant $i_n^R(t)$ qui traverse le module de mesure \mathcal{R}_n est retrouvé grâce à la loi d’Ohm

$$i_n^R(t) = \frac{v_n(t)}{R_n}. \quad (5.6)$$

Les modules de mesure de tension \mathcal{V}_l présentent une entrée haute impédance, le courant qui les traverse est nul. Ils mesurent la valeur des tensions $V_l(t)$.

Afin de prendre en compte les imperfections des générateurs de courant dans les modules d’excitation, les courants traversant ces derniers sont mesurés. Les signaux à démoduler sont donc

- 2 signaux pour chaque module d’excitation \mathcal{E}_k afin de les caractériser en courant et en tension,
- 1 signal pour chaque module de mesure de courant \mathcal{R}_n ,
- 1 signal pour chaque module de mesure de tension \mathcal{V}_l .

Chaque signal est démodulé en module et en phase sur chacune des porteuses (les K signaux $i_k(t)$ (5.5) des modules d’excitation). Le nombre de canaux de données à numériser et à envoyer à l’ordinateur est donc

$$N_c = 2K(2K + N + L). \quad (5.7)$$

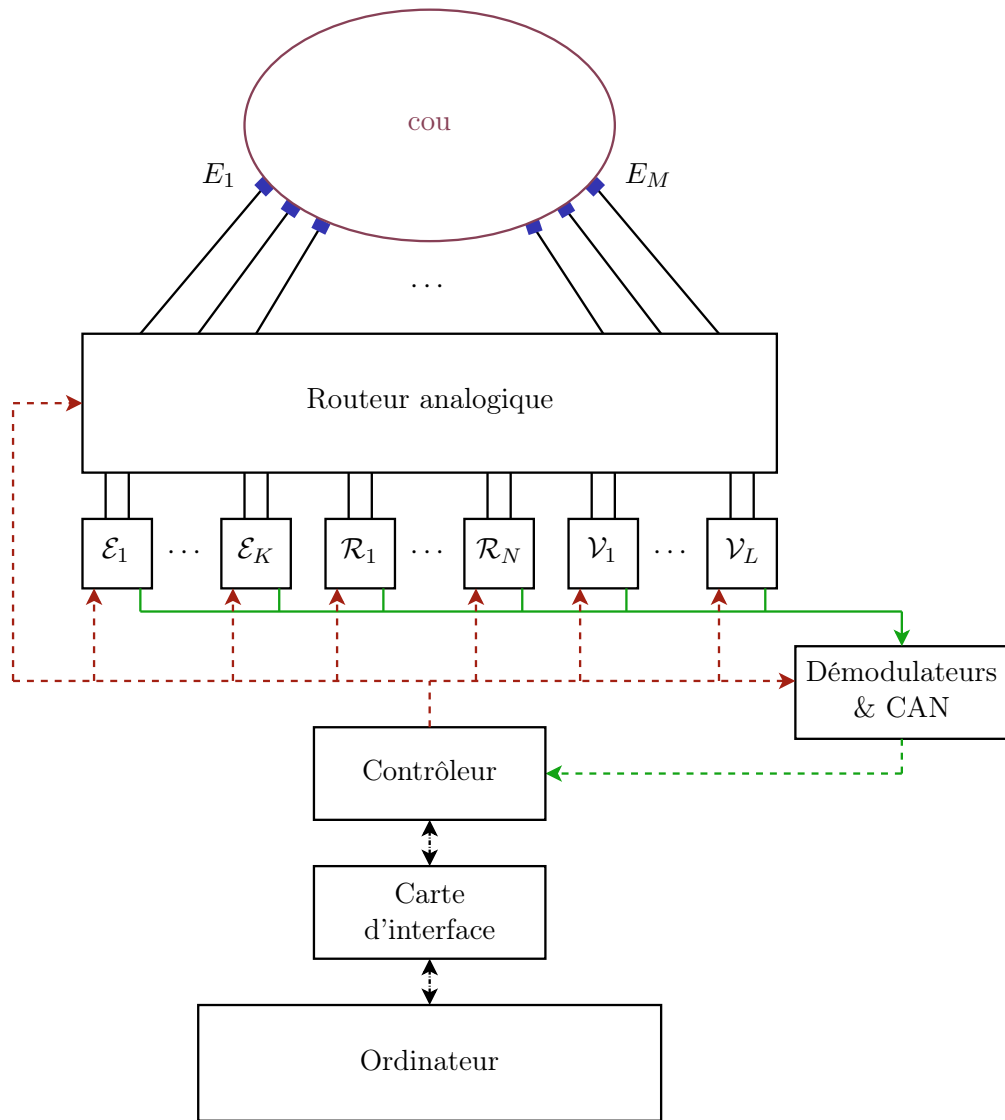


FIGURE 5.6 – Système complet (signaux de commande et de configuration en rouge, signaux de mesure en vert, signaux numériques en pointillés, signaux analogiques en traits pleins).

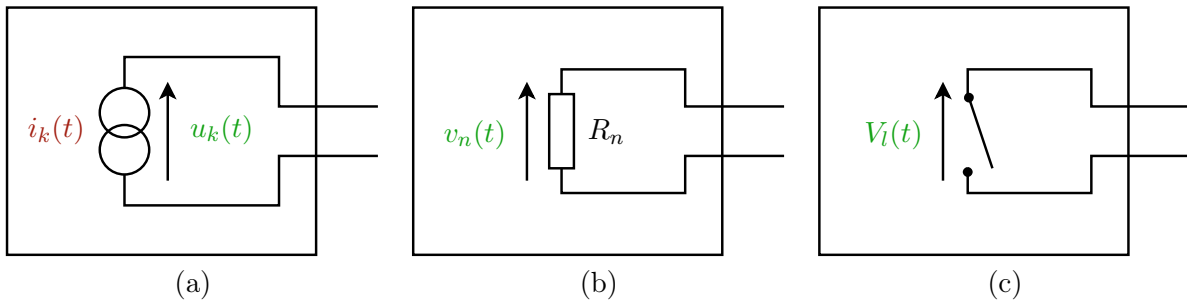


FIGURE 5.7 – (a) Module d'excitation \mathcal{E}_k (b) Module de mesure de courant \mathcal{R}_n (c) Module de mesure de tension \mathcal{V}_l .

5.3.5 Choix et dimensionnement

Les contraintes techniques, de coût, de temps de développement, et de capacité de fabrication nous ont menés aux choix suivants.

1. Le premier prototype est composé de
 - $K = 2$ modules d'excitation,
 - $N = 2$ modules de mesure de courant,
 - $L = 4$ modules de mesure de tension,
 - $M = 32$ électrodes.
2. La carte d'interface (cf. figure 5.6) est une carte audio-numérique prenant en charge les canaux MADI¹ sur fibre optique.
3. Le contrôleur est basé sur un composant de type FPGA. Une interface physique permettant la sérialisation/dé-sérialisation des données MADI et l'adaptation des débits pour le FPGA est insérée entre la carte audio et le contrôleur.
4. Les fréquences porteuses sont réglables dans l'intervalle $[0.1, 20]$ MHz.
5. L'échantillonnage des mesures est effectué à 48 kHz sur 24 bits.

5.4 Développements et réalisations techniques

Je présente ici succinctement l'état d'avancement du prototype et les développements techniques réalisés. Pour plus de détails, le lecteur pourra se référer au rapport de stage d'Antonin Mauffrey [Mau13].

5.4.1 Vue générale du dispositif

5.4.1.1 Identification des contraintes et des problèmes techniques

Lors de la conception du dispositif, une attention particulière a été portée sur

- la sécurité du dispositif, destiné à être utilisé sur des patients humains,
- le rejet des perturbations électromagnétiques, des modules et électrodes entre eux et de/vers l'environnement, en raison de la finesse des phénomènes que nous cherchons à observer.

Le système effectue des mesures électriques sur le corps humain. Il est donc soumis à la norme de sécurité ANSI / AAMI – 1993. Cette norme stipule que l'appareil doit respecter, entre autres, les points suivants.

1. Aucune tension continue ou courant continu ne doit traverser le corps humain.
2. Les courants alternatifs qui traversent le corps humain ne doivent pas excéder une intensité efficace de 10 mA.
3. Les électrodes qui sont au contact du corps humain doivent être séparées galvaniquement du secteur et de la Terre.

1. Protocole audio-numérique permettant des très hauts débits (64 canaux échantillonnés à 48 kHz sur 24 bits sur une seule fibre optique). Le choix d'une connexion audio-numérique est motivé par l'expertise et les logiciels disponibles à l'IRCAM.

5.4.1.2 Architecture du dispositif et état d'avancement

L'architecture du dispositif et les différentes cartes qui le composent sont présentées sur la figure 5.8.

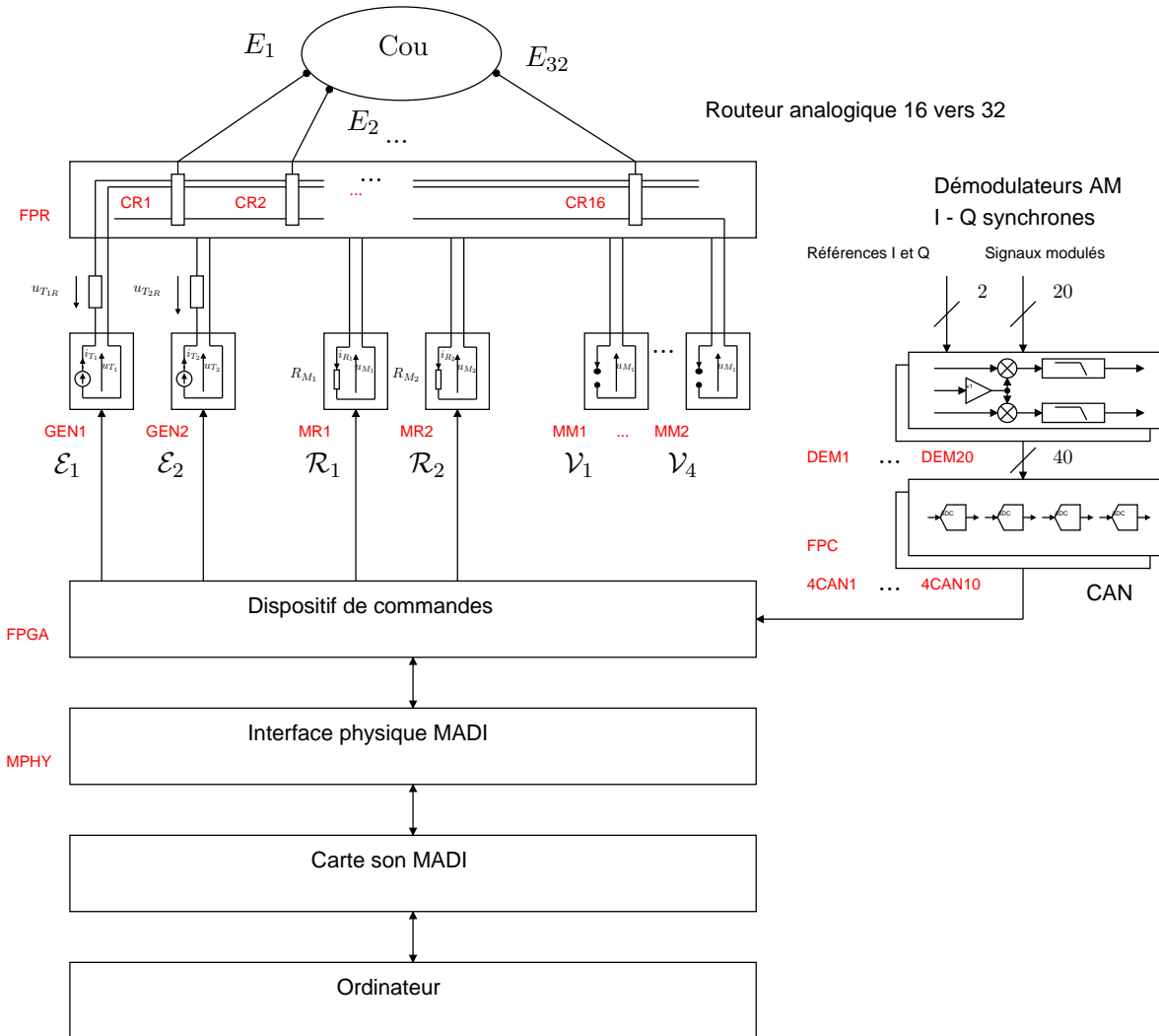


FIGURE 5.8 – Architecture détaillée du dispositif. Chaque label en rouge correspond à une carte électronique. Sont déjà réalisées les cartes FPGA, MPHY, GEN1-GEN2, FPR, CR1-CR8 ainsi qu'un prototype de la carte 4CAN1. Restent à réaliser MRx, MMx, DEMx, FPC, CR9-CR16 et les 4CANx définitives.

Pour des raisons pratiques, les modules d'excitation \mathcal{E}_k réalisés sont des sources de tension alternative et non des sources de courant (ce qui est sans conséquence puisqu'ils sont caractérisés en courant et en tension). Ils sont présentés dans la section 5.4.2.1. Les tensions aux bornes de chaque module sont mesurées par des amplificateurs d'instrumentation, sous haute impédance, puis démodulés et numérisés. L'accès aux électrodes est effectué par un routeur analogique 16 vers 32, qui autorise la connexion d'une électrode à un unique ou aucun module. Chaque module peut être connecté à une ou plusieurs électrodes. Une attention particulière a été portée sur la conception de ce routeur, comme détaillé en section 5.4.2.2. La section 5.4.2.3 présente la stratégie développée pour éviter la perturbation des canaux analogiques.

Les données de commande du dispositif sont

- la fréquence porteuse des deux modules d'excitation et
- l'état du routeur analogique.

L'envoi des commandes se fait à travers la connexion MADI. Un protocole *ad-hoc* est mis en place pour empaqueter les données de commande dans les canaux MADI et assurer la synchronisation des ordres de configuration et des mesures. Ce protocole est présenté dans la section 5.4.2.4.

Sont aujourd'hui réalisés (voir figure 5.8)

- l'interface physique permettant la sérialisation/dé-sérialisation des données MADI,
- les deux modules d'excitation \mathcal{E}_1 et \mathcal{E}_2
- un routeur analogique 8 vers 32 constitué d'une carte fond de panier et de 8 cartes 1 vers 32 (le fond de panier est prévu pour 16 cartes 1 vers 32),
- un prototype d'une carte de conversion analogique-numérique 4 canaux,
- le protocole de communication permettant de piloter le routeur.

Reste aujourd'hui à réaliser

- les modules de mesure de courant \mathcal{R}_n et les modules de mesure de tension \mathcal{V}_l ,
- les cartes de démodulation,
- les cartes de conversion analogique-numérique définitives,
- le protocole de pilotage des modules d'excitation.

5.4.2 Éléments matériels

5.4.2.1 Génération de signaux isolés galvaniquement

Les modules d'excitation sont constitués d'un composant de type DDS (Direct Digital Synthesizer). Ces composants permettent de contrôler numériquement la fréquence des signaux générés au Hertz près. Le composant choisi, AD9959 fabriqué par Analog Device, possède 4 sorties indépendantes et permet donc de générer, en plus du signal d'excitation, deux sinusoïdes de référence (I et Q, pour la démodulation en module et en phase) qui sont envoyées vers les cartes de démodulation sans être perturbées par la charge des électrodes.

Afin d'isoler galvaniquement le DDS (alimenté) des électrodes disposées sur le cou du patient, un transformateur 1 : 1 est disposé entre le DDS et le routeur, selon le schéma indiqué en figure 5.9. Les mesures de courant et de tension sont effectuées au secondaire, au plus près des électrodes, pour avoir des mesures les plus fiables possible. Le module ainsi construit est équivalent à un générateur de tension.

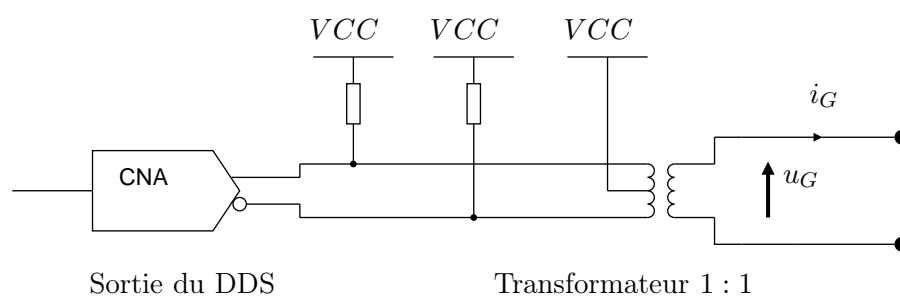
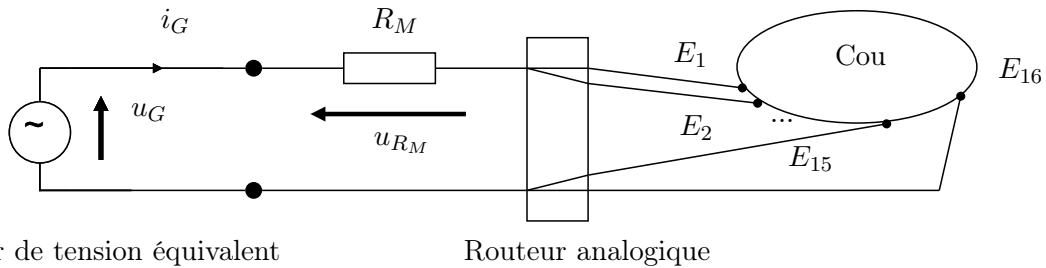


FIGURE 5.9 – DDS et transformateur d'isolement.

Notons qu'un tel dispositif est suffisant pour réaliser une EGG standard, comme présenté sur la figure 5.10.



Générateur de tension équivalent

Routeur analogique

FIGURE 5.10 – EGG à un canal réalisé à l'aide d'un module d'excitation.

Un des deux modules d'excitation réalisés est présenté sur la figure 5.11.

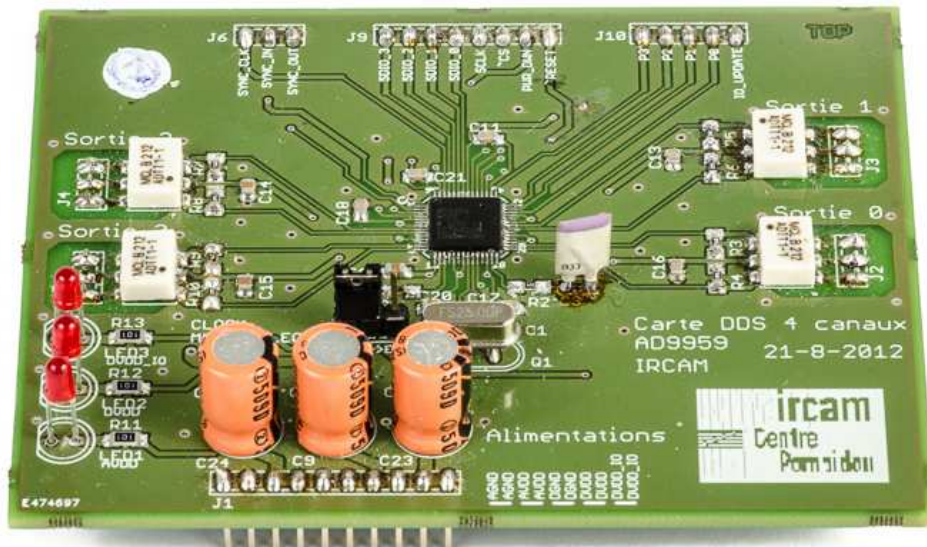


FIGURE 5.11 – Module d'excitation réalisé.

5.4.2.2 Routeur analogique

De fortes contraintes pèsent sur les commutateurs analogiques destinés à assurer le routage analogique entre les modules d'excitation et de mesure d'une part, et les électrodes d'autre part.

1. Les commutateurs doivent permettre le passage de signaux alternatifs.
2. Les commutateurs doivent permettre un passage direct du signal analogique (pas de suiveur).
3. Ils doivent présenter une résistance inférieure à 5Ω .
4. La résistance d'un canal activé doit être connue très précisément, car l'ordre de grandeur des variations de résistance que nous souhaitons mesurer est de 1Ω .

5. Les capacités vues en entrée et en sortie des commutateurs ne doivent pas charger les lignes vues par les électrodes, sous peine de rendre impossible les mesures “haute impédance” aux fréquences de travail.
6. Les temps de transition lors des commutations doivent être inférieurs à 10 % du temps de l'échantillon (soit environ $2 \mu\text{s}$).
7. Les commutations doivent être progressives, afin de garantir un passage pour le courant tout au long de la transition.
8. L'isolation entre deux canaux spatialement proches doit être suffisante pour éviter les perturbations.
9. Les commutateurs doivent être synchronisables entre eux.

Étant donné l'accumulation des contraintes, le choix s'est porté sur une matrice d'interconnexion entièrement réalisée par nos soins. Le routeur est composé de 16 cartes 1 vers 32 branchées sur une carte “fond de panier”. Chaque carte 1 vers 32 est composée de 4 composants ADG714, fabriqués par Analog Device, chacun constitué de 8 interrupteurs. La structure d'une carte 1 vers 32 est présentée en figure 5.12.

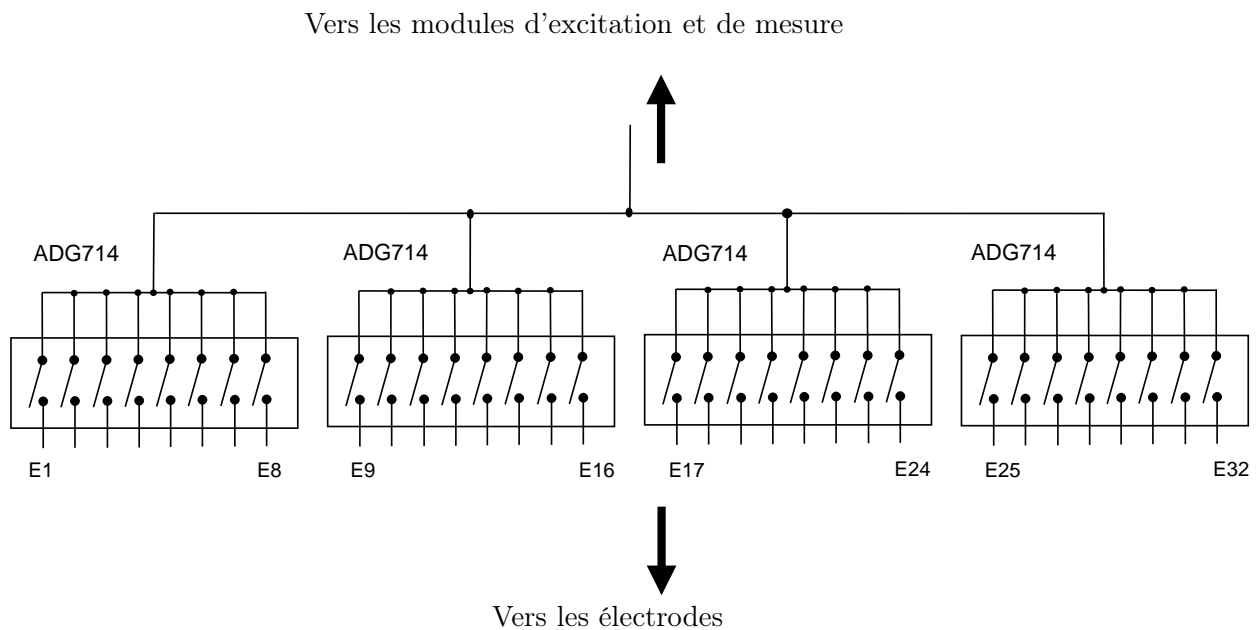


FIGURE 5.12 – Carte de commutation 1 vers 32.

Les commutations progressives garantissant le passage du courant d'excitation en tout instant sont réalisées par un fonctionnement “Make-Before-Break”. Le passage d'une configuration à une autre se fait en deux passes.

Première passe Les interrupteurs ouverts qui doivent être fermés dans la nouvelle configuration sont actionnés.

Deuxième passe Les interrupteurs fermés qui doivent être ouverts dans la nouvelle configuration sont actionnés.

5.4.2.3 Accès protégé aux électrodes

Un soin particulier doit être apporté à la conception des cartes du routeur et des câbles des électrodes afin de prévenir les perturbations des canaux analogiques entre eux et les perturbations apportées par l'environnement. De plus, les câbles des électrodes sont susceptibles d'apporter des effets capacitifs.

Lors de l'utilisation d'un câble coaxial permettant de protéger le signal, voyageant dans l'âme du câble, des perturbations extérieures, une capacité apparaît entre l'âme et le blindage, comme présenté sur la figure 5.13 en haut. Afin de réduire l'influence de la capacité parasite, nous plaçons des suiveurs pour forcer le blindage au même potentiel que l'âme du câble. Afin de réduire la génération de perturbations, susceptibles de parasiter les câbles voisins, un blindage supplémentaire, au potentiel libre ou relié à la masse du système, est ajouté (figure 5.13 en bas).

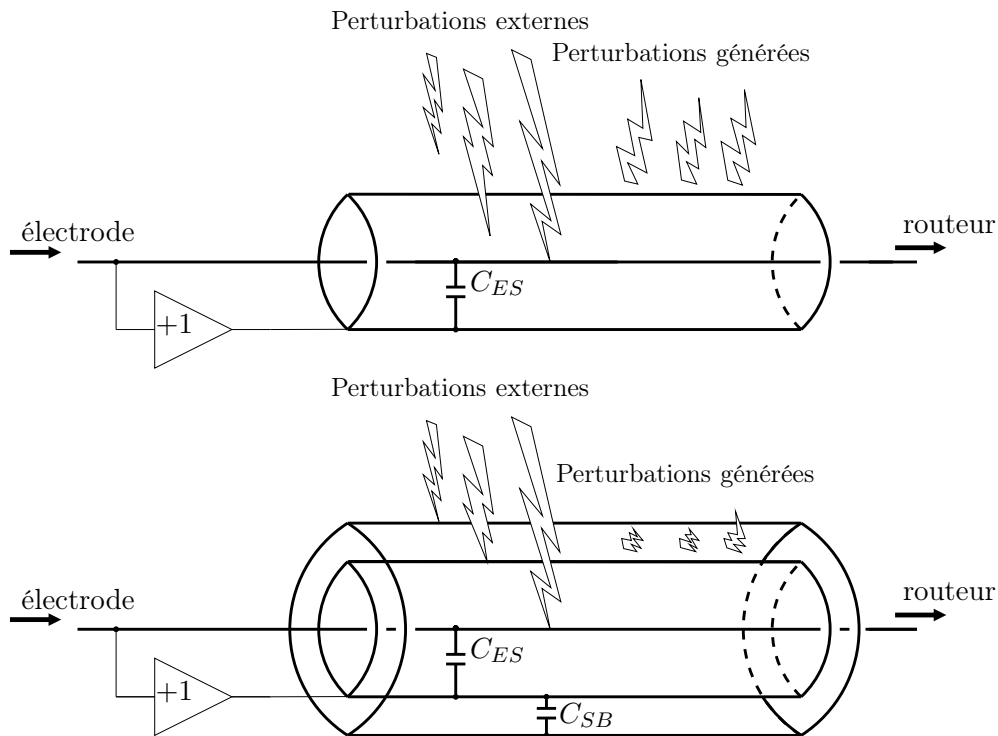


FIGURE 5.13 – Câble coaxial et triaxial. En haut : le blindage permet une protection des perturbations externes. Le suiveur permet de réduire l'influence de la capacité de couplage C_{ES} . En bas : Le blindage externe non suivi permet en plus de bloquer les perturbations générées par le câble.

La même stratégie de blindage suivi est mise en œuvre dans les cartes électroniques à l'aide de la méthode dite du "Stitching". Cette méthode consiste, dans une carte constituée de plusieurs couches, à entourer la piste véhiculant le signal analogique de pistes (une piste sur les couches directement inférieure et supérieure et des ponts traversants reliant ces deux pistes) mises au même potentiel par l'intermédiaire d'un suiveur. Elle permet de constituer une sorte de cage de Faraday autour de la piste à protéger au sein même des cartes électroniques. Cette technique, illustrée sur la figure 5.14 a été mise en œuvre lors de la conception des cartes 1 vers 32 et de la carte fond de panier du routeur, réalisées sur des cartes 4 couches.

La carte fond de panier sur laquelle est montée huit cartes 1 vers 32 est présentée en figure 5.15.

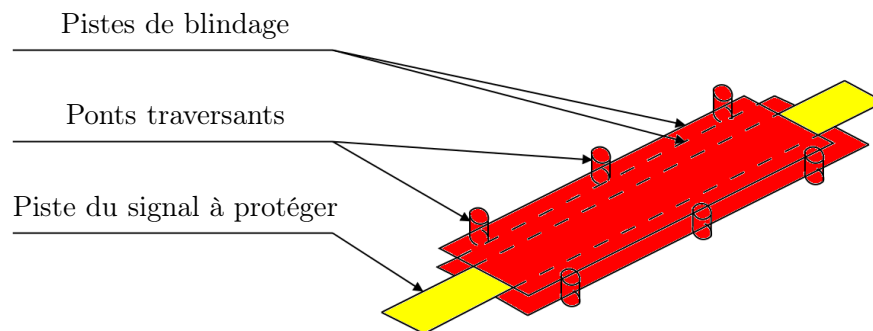


FIGURE 5.14 – Blindage des pistes du routeur analogique par la méthode du *Stitching*.

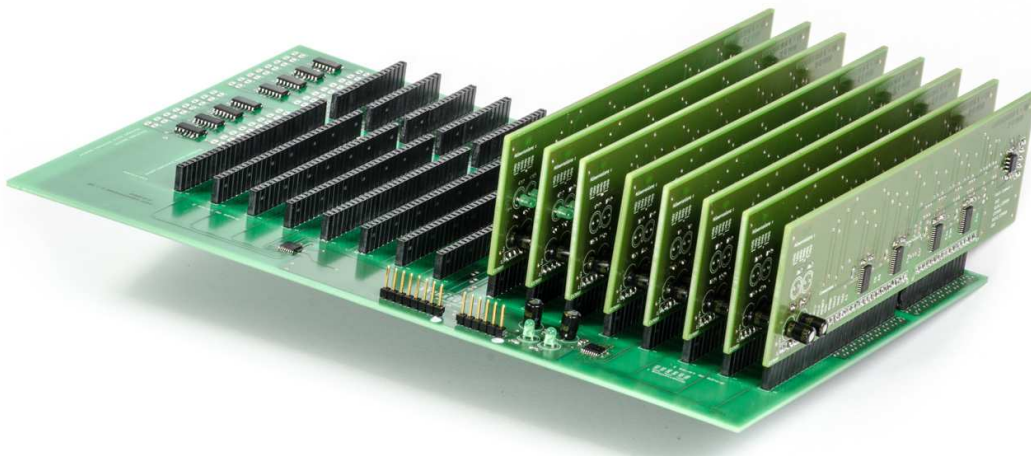


FIGURE 5.15 – Carte fond de panier et huit cartes 1 vers 32 (4 couches). On y voit les 4 composants ADG714 sur la première carte 1 vers 32 et les suiveurs à l’extrémité de la carte fond de panier.

5.4.2.4 Traitement synchronisé des ordres de configuration et des mesures

D’après l’équation (5.7) et les choix de dimensionnement présentés dans la section 5.3.5, notre système doit être capable d’échantillonner et d’envoyer 40 signaux de données à une fréquence de 48 kHz. De plus, le dispositif de commande doit pouvoir piloter le routeur analogique en ayant la possibilité de réaliser une modification de configuration à chaque échantillon en deux passes. Les CAN sont eux aussi pilotés par le dispositif de commande.

Le choix pour le dispositif de commande s’est porté sur un composant de type FPGA, capable de traiter parallèlement l’intégralité des ordres de commande et des canaux de données à haute vitesse et présentant un grand nombre d’entrées/sorties. Une interface physique est intercalée entre la carte audio-numérique et le FPGA afin de (i) réduire le débit de données entrant par la fibre optique au format MADI, trop important pour le FPGA, en dé-sérialisant les données brutes (ii) décharger le FPGA en effectuant la mise en forme et la sérialisation des données au standard MADI. Cette carte réalisée par nos soins, est présentée en figure 5.16.

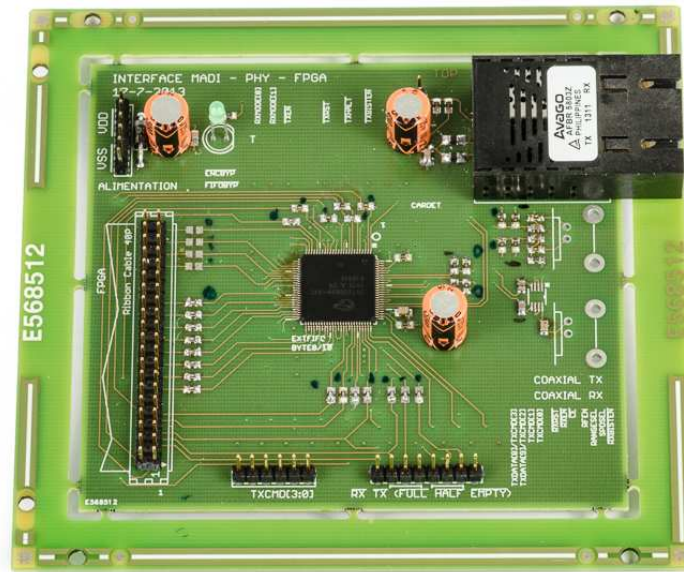


FIGURE 5.16 – Accès au lien MADI physique : carte réalisée avec entrée/sortie optique.

Le contrôleur FPGA a pour rôle de piloter et de synchroniser l'intégralité du système. Une attention particulière doit être portée au problème de la synchronisation, pour deux raisons.

1. Les effets capacitifs ne peuvent être entièrement supprimés et font apparaître des transitoires après une modification de configuration. Il est donc nécessaire d'effectuer la mesure le "plus tard possible", juste avant une nouvelle commutation des routeurs.
2. Les communications MADI montante et descendante ne sont pas synchronisées. Un délai non contrôlable peut apparaître entre les trames qui arrivent dans le dispositif (contenant les signaux de commandes) et celles partant du dispositif (contenant les données mesurées démodulées et échantillonnées).

La stratégie pour répondre à ces problèmes est présentée en figure 5.17. Pour répondre au premier problème, la mesure (effectuée en bloquant l'entrée des CAN) est réalisée juste avant l'établissement d'une nouvelle configuration (qui, pour rappel, est effectuée en deux passes, selon la procédure Make-Before-Break). Pour répondre au deuxième problème, une donnée d'identification de configuration est ajoutée à chaque trame de signaux de commande envoyée par l'ordinateur. Cette donnée est mis en mémoire tout au long du processus de configuration du dispositif et est renvoyée à l'ordinateur en même temps que les échantillons des données de mesure qui lui correspondent. Le traitement séquentiel et parallèle des données de configuration est présenté sur la figure 5.17.

5.4.3 Premier élément logiciel et premiers tests

Afin de tester les cartes et protocoles développés, un premier élément logiciel a été programmé pour envoyer des données de commande au routeur et recevoir des données échantillonnées. Le choix a été fait d'utiliser le logiciel de programmation graphique Max/MSP, bien adapté au traitement des signaux audio-numériques et disponible à l'IRCAM.

Les données de pilotage du routeur sont codées sur les 32 premiers canaux du lien MADI allant de l'ordinateur au dispositif. Chaque carte de commutation 1 vers 32 est constituée de 32 interrupteurs.

L'état de ces 32 interrupteurs est codé par 32 bits (un 0 correspond à un interrupteur ouvert, un 1 correspond à un interrupteur fermé), répartis sur les 16 bits de poids fort de deux canaux². Ainsi les canaux 1 et 2 du lien MADI pilotent la carte de commutation numéro 1, les canaux 2 et 3 la carte numéro 2 etc. Ce codage n'est pas optimal pour l'occupation de la bande passante dans le lien MADI mais il permet un décodage rapide par le FPGA et seule la moitié des 64 canaux disponibles est utilisée, laissant de l'espace pour les développements futurs.

J'ai réalisé un patch Max/MSP permettant de contrôler directement la matrice de commutation et effectuant le codage de celle-ci selon le protocole qui vient d'être décrit. L'interface utilisateur est présentée en figure 5.18. L'utilisateur peut, au choix, définir un numéro de modules pour chaque électrode (entre 0 et 16, 0 signifie que l'électrode est déconnectée). Ce patch a permis de valider le fonctionnement du routeur et du protocole de communication en pilotant des LEDs s'allumant ou s'éteignant selon la configuration de la matrice de pilotage du routeur.

Nous avons pu valider le fonctionnement de la carte CAN 4 canaux réalisée et de la mise en forme des données au standard MADI en récupérant dans Max/MSP des signaux analogiques envoyés sur les entrées de la carte CAN. La qualité de l'échantillonnage est particulièrement satisfaisante et la dynamique des 24 bits de codage est bien exploitée.

Enfin, le fonctionnement des cartes DDS a déjà été validé lors de leur réalisation dans le cadre d'un précédent stage [Mau12].

La fin du stage (30 septembre 2013) et la fin de ma thèse n'ont pas permis d'aller jusqu'à la mise en place des démodulateurs. Ce dernier ingrédient permettra (i) de valider le fonctionnement EGG standard et (ii) faire des premières mesures lorsque les modules \mathcal{R}_k et \mathcal{V}_k (plus simple à concevoir) seront réalisés.

5.5 Conclusion et perspectives

J'ai présenté dans ce chapitre un dispositif mixte EGG/EIT multicanal à électrodes à géométrie variable pour l'exploration de l'activité glottique. Nous avons vu en préambule, grâce à une étude bibliographique et à un test préliminaire, que l'utilisation de l'EIT pour l'exploration glottique est envisageable et devrait permettre d'améliorer les performances des outils d'exploration actuels, notamment pour la reconstruction du mouvement dans l'épaisseur des plis vocaux. Sur la base de ces considérations un nouveau dispositif est proposé. Il exploite

1. des mesures mixtes EGG/EIT afin de disposer des deux techniques et de pouvoir les combiner,
2. un multiplexage fréquentiel afin d'exciter plusieurs chemins simultanément,
3. un routeur analogique permettant de balayer les électrodes, de concentrer l'énergie en une zone d'intérêt, et de dessiner des motifs d'activation en regroupant plusieurs électrodes.

De nombreux composants du dispositif (modules d'excitation, communication MADI, routeur analogique) ont été fabriqués/programmés et validés. Le routeur nécessite des tests plus complets pour vérifier l'absence de diaphonie (perturbation des canaux entre eux) et l'efficacité de l'isolation des canaux par la méthode du stitching. Des premiers tests du routeur ont fait apparaître des dysfonctionnements, vraisemblablement dus à des défauts de fabrication. Ce problème devra être étudié en détail et résolu. Des composants restent à réaliser (modules de démodulation et version définitive des cartes d'échantillonnage, modules de mesure et pilotage des modules d'excitation par

2. Les canaux sont codés sur 24 bits, le reste des bits n'est pas utilisé.

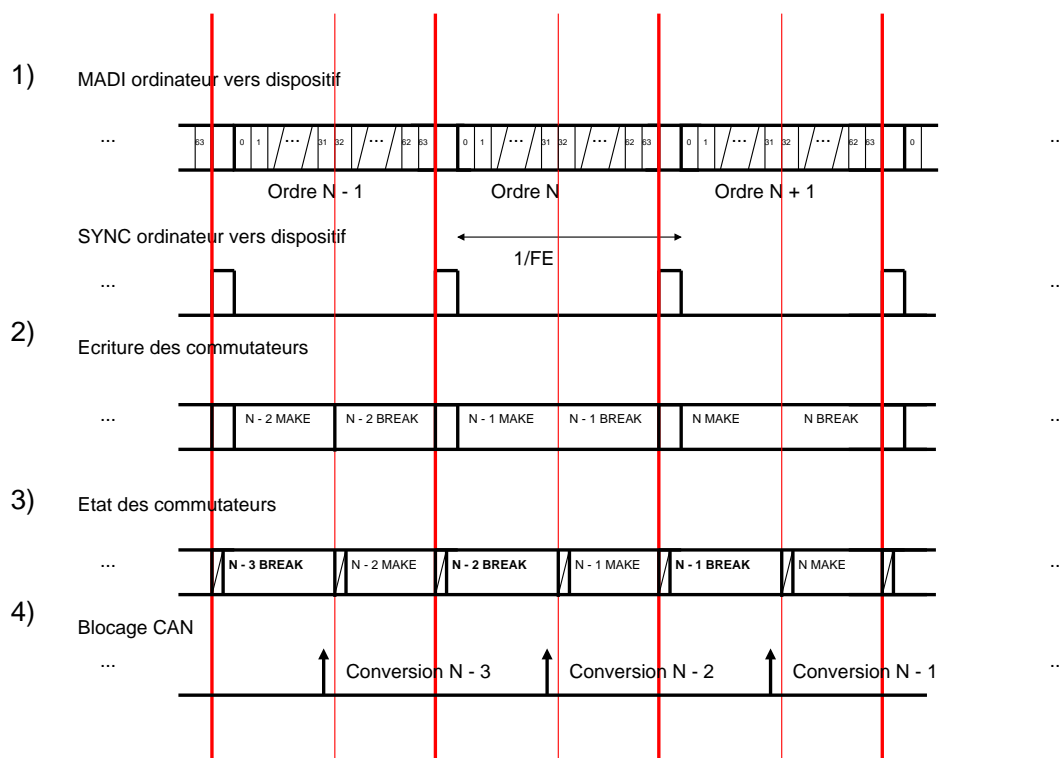


FIGURE 5.17 – Traitement séquentiel des ordres de configuration. Les ordres de configuration sont d’abord lu et mis en mémoire puis envoyés au commutateurs du routeur. Ces derniers sont ensuite activés tous en même temps. Cette procédure est effectuée deux fois par période (Make-Before-Break) et les CAN ne sont activés qu’au dernier moment, juste avant le Make de l’échantillon suivant. La donnée SYNC (identification des ordres de configuration) de la configuration N est mise en mémoire pendant toute la procédure et est renvoyée en même temps que les données de mesure correspondant à cette configuration.

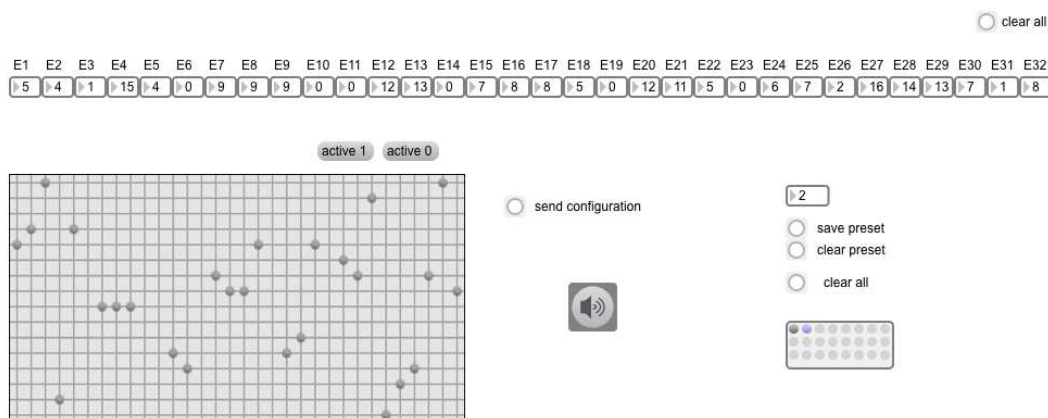


FIGURE 5.18 – Programme MAX/MSP pour le pilotage du routeur.

le FPGA). Des démodulateurs standards utilisés dans le domaine de télécommunications ont été testés pendant le stage d'Antonin Mauffrey. Ces composants ne donnent pas satisfaction en raison du niveau de bruit résiduel et des basculements entre les voies I et Q qui peuvent apparaître. Ces limitations ne sont pas un problème en télécommunications (numériques) mais rendent ces solutions inutilisables dans notre cas.

Lorsque ces composants seront réalisés, des tests plus complets seront menés sur la maquette à l'échelle 1 :1 afin de valider le fonctionnement du dispositif dans son ensemble. Cette maquette est prévue pour être pilotée en vibration, des tests pourront donc être menés en position statique comme en dynamique.

Deux pistes sont envisagées pour régulariser le problème de l'EIT dans notre cas.

- La résolution d'un problème paramétrique, en prenant en compte un modèle paramétrique de la carte de conductivité à l'intérieur du larynx. Ceci revient à injecter de la connaissance a priori sur la géométrie du corps étudié. Notre problème est a priori bien adapté à ce type de méthodes car nous cherchons à détecter la géométrie d'un conduit rempli d'air (très peu conducteur) au milieu de tissus humains (conducteurs).
- L'utilisation des transformations conformes, qui ont déjà mené à des résultats probants dans le cadre de l'EIT [HK05, HK06, HK10, BHBH10].

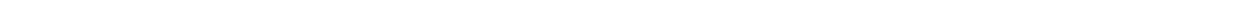
Ce projet a été présenté, dans son état actuel, lors des *Journées Larynx* organisées par Nathalie Henrich les 12 et 13 septembre 2013 au *Laboratoire d'Anatomie des Alpes Françaises*, à des spécialistes du larynx (scientifiques, phoniâtres et chirurgiens). Les nouvelles possibilités que permettrait ce dispositif ont suscité des réactions positives et enthousiastes de la part de la communauté médicale, qui nous encourage à poursuivre nos efforts pour aller vers un dispositif qui permettra, nous l'espérons, d'aboutir à une imagerie 3D des plis vocaux en vibration.





Deuxième partie

Modèle source-filtre et analyse-synthèse



SOMMAIRE DE LA PARTIE

La partie **II** est consacrée à l'analyse-synthèse du modèle source-filtre (choix **C1** dans la section 2.2.3, acoustique glottique forcée).

Le modèle tous-pôles évoqué dans la section 2.2.1 permet une représentation du signal vocal par filtrage linéaire d'un train d'impulsions de Dirac. Les modèles causaux-anticausaux sont particulièrement adaptés au problème de la séparation source-filtre car ils permettent de différencier dans le signal étudié la contribution anticausale de la source glottique de la contribution causale du filtre vocal. Après un rappel historique des modèles paramétriques de source glottique et de filtre vocal existants, le modèle tous-pôles défini pour mon étude sera présenté dans le chapitre 6.

La question de l'estimation des paramètres de ce modèle et de la séparation source-filtre à partir de signaux de voix est abordée dans le chapitre 7 où deux algorithmes sont proposés.

Premier algorithme : Les pôles causaux du filtre et anticausaux de la source sont retrouvés grâce à un opérateur permettant de transformer en zéros les pôles d'un signal parfaitement périodique construit à partir d'un extrait du signal étudié. Cet algorithme permet de retrouver tous les paramètres dans le cas des signaux de synthèse non bruités mais aboutit à des résultats insuffisamment robustes pour traiter des signaux réels. Des améliorations possibles sont proposées mais laissées en perspective car le choix a été fait de porter mes efforts de recherche sur la construction d'un second algorithme.

Deuxième algorithme : Le second algorithme repose sur (i) la prise en compte de la robustesse par une méthode à maximisation de vraisemblance et (ii) la représentation des paramètres du modèle (pôles causaux/anticausaux, gain et position du GCI) par le cepstre complexe. Le problème du calcul du cepstre complexe est préalablement abordé en portant une attention particulière à la définition et aux propriétés du logarithme complexe.



Chapitre 6

Le modèle source-filtre

Sommaire

6.1	Source glottique	96
6.1.1	Modèles paramétriques de source glottique	96
6.1.1.1	Modèles paramétriques définis par morceaux	97
6.1.1.2	CALM (Causal-Anticausal Linear Model)	97
6.1.2	Choix du modèle d'étude	98
6.1.2.1	Pôles causaux et pôles anticausaux	99
6.1.2.2	Définition du modèle de source glottique	100
6.2	Filtre vocal et rayonnement	100
6.2.1	Conduit vocal et filtre tous-pôles	101
6.2.2	Filtre pôles-zéros	101
6.2.3	Filtre de rayonnement	101
6.2.4	Définition du modèle d'étude	101
6.3	Modèle source-filtre paramétrique complet étudié	102

6.1 Source glottique

Le modèle source-filtre repose sur l'hypothèse de non-interaction entre la source glottique et le conduit vocal (acoustique glottique forcée). De ce fait, le modèle se résume, comme son nom l'indique, à un signal (la *source*) filtré (par le *filtre*). D'un point de vue physique, on peut considérer le signal de source comme une source de débit idéale. Après une revue des modèles existants, le modèle d'étude est défini dans la section 6.1.2.

6.1.1 Modèles paramétriques de source glottique

De nombreux modèles temporels de source glottique ont été proposés dans la littérature. La plupart d'entre eux modélisent une période de source glottique $g(t)$ ou de sa dérivée $g'(t)$ à l'aide de fonctions paramétriques définies par morceaux. Une étude récente propose une unification partielle de l'écriture de ces différents modèles [DDH06]. Je propose dans l'annexe B une écriture de ces modèles en fonction des mêmes paramètres temporels $\{t_i, t_p, t_e, t_c, d_a\}$ (paramètres de forme), de la durée totale du pulse T (période) et d'un gain \mathcal{G} . Ces modèles proposent de représenter l'impulsion glottique $g(t)$ et/ou sa dérivée $g'(t)$ à l'aide de quatre fonctions paramétriques, comme présenté sur la figure 6.1 : une phase montante sur $[0, t_p]$ (avec un maximum de dérivée en t_i), une phase descendante $[t_p, t_e]$, une phase de retour $[t_e, t_c]$ paramétrée par d_a , et une phase nulle $[t_c, T]$, où T est la période glottique. Les gains $\mathcal{G} = \max(-g'(t))$ et $\tilde{\mathcal{G}} = \max(g(t))$ de la figure 6.1 sont liés pour chaque modèle. Lorsqu'un des deux gains est choisi, le deuxième est aisément calculé en écrivant les égalités $\tilde{\mathcal{G}} = g(t_p)$ et $\mathcal{G} = g'(t_e)$. Le paramètre d_a permet, comme on le voit sur les courbes de $g'(t)$ dans la figure 6.1, d'assurer la continuité de la dérivée de la source glottique. Cela revient à modéliser une fermeture douce des plis vocaux, la source glottique $g(t)$ revient doucement à zéro entre t_e et t_c et ne présente pas de discontinuité de pente. Nous parlons alors de fermeture douce, et de fermeture abrupte en cas d'absence de phase de retour ($d_a = 0$). Notez que le cas de la fermeture abrupte rend le paramètre t_c caduque (cas dégénéré). Certains modèles n'incluent pas de phase de retour $[t_e, t_c]$ et d'autres n'incluent pas de phase nulle $[t_c, T]$.

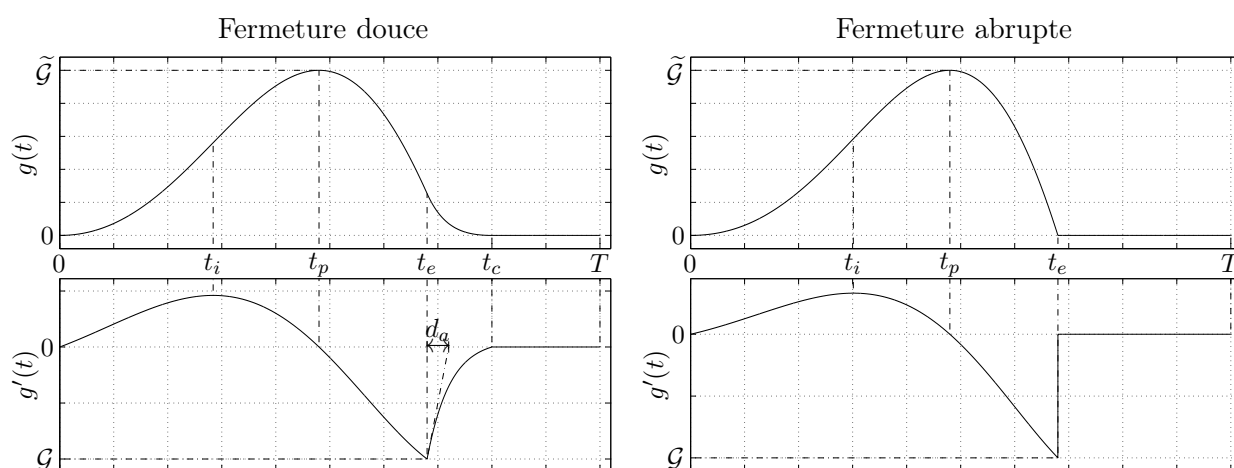


FIGURE 6.1 – Forme générique d'un pulse glottique paramétrique de durée T (les courbes présentées ici sont issues du modèle LF). La forme de l'impulsion est paramétrée par les instants $\{t_i, t_p, t_e, t_c\}$ et la durée d_a . \mathcal{G} (ou $\tilde{\mathcal{G}}$, les deux sont liés) est le paramètre de gain.

L'influence des paramètres de forme est facilement interprétable. Le rapport $\frac{t_p}{t_e}$ permet de régler l'asymétrie de l'impulsion glottique. La durée d_a est directement liée à la douceur de la fermeture glottique comme nous l'avons évoqué. Enfin, le paramètre t_c règle la longueur de la phase nulle à la fin de la période glottique. Pour une revue plus détaillée de l'influence de ces paramètres sur le pulse glottique et son spectre, le lecteur pourra se référer à [DDH06] et [Deg10].

Le modèle CALM [DDH03] propose une approche différente. Dans ce modèle, la glotte est modélisée comme un filtre, autrement dit le pulse glottique est considéré comme étant la réponse impulsionnelle d'un "filtre glottique".

Dans le cadre de mes travaux, nous nous intéresserons à une approche similaire au modèle CALM. La glotte est représentée par une paire de pôles complexes conjugués anticausaux et (éventuellement) un pôle réel causal correspondant à la phase de retour.

Les dénominations des modèles de source glottique utilisées ici sont celles couramment utilisées dans la littérature. Je présente ici les modèles principaux et leurs propriétés. Les expressions analytiques paramétriques des segments présentés sur la figure 6.1 sont présentées en annexe B. Trois familles de modèles sont distinguées :

- le modèle Rosenberg-B [Ros71] et ses dérivés KLGLOTT88 [KK90] et R++ [Vel98],
- le modèle de Fant [Fan79] et ses dérivés LF [FLL85, FKL94] et LF-Rd [Fan95],
- et enfin le modèle CALM [DDH03] et le modèle utilisé dans mes travaux qui en est inspiré.

6.1.1.1 Modèles paramétriques définis par morceaux

Le nombre de paramètres de forme (auxquels s'ajoutent ceux de gain et de période), la présence d'une phase de retour et la présence d'une phase nulle de chaque modèle sont présentés dans le tableau 6.1. Le modèle LF (et sa version réduite à une paramètre LF-Rd) reste aujourd'hui le modèle le plus utilisé. Une étude analytique de ce modèle est détaillé en annexe C.

Modèle	<i>Rosenberg-B</i>	<i>KLGLOTT88</i>	<i>R++</i>	<i>Fant</i>	<i>LF</i>	<i>LF-Rd</i>
Nb de paramètres de forme	2	2	3	2	4	1
Phase de retour	non	oui	oui	non	oui	oui
Phase nulle	oui	non	non	non	oui	oui

TABLE 6.1 – Modèles paramétriques d'impulsion glottique définis par morceaux. De haut en bas : nom du modèle, nombre de paramètres de formes (auxquels s'ajoutent ceux de gain et de période), présence d'une phase de retour, présence d'une phase nulle.

6.1.1.2 CALM (Causal-Anticausal Linear Model)

Le modèle CALM [DDH03] qui a inspiré notre démarche est fondamentalement différent des modèles paramétriques présentés ci-dessus. Ce modèle représente la source glottique comme la réponse impulsionnelle d'un filtre, le filtre glottique. Il est cependant inspiré du modèle LF. En effet, le modèle LF est constitué d'une portion de sinusoïde croissante exponentiellement entre 0 et t_e et d'une portion d'exponentielle décroissante entre t_e et t_c . Si on prolonge à gauche la sinusoïde croissante et à droite l'exponentielle décroissante, on obtient la réponse impulsionnelle d'un filtre composé d'une paire de pôles complexes conjugués anticausaux stables (module plus grand que 1) et d'un pôle réel causal stable (module compris entre 0 et 1) qui correspond à la phase de retour.

Le modèle CALM de dérivée d'impulsion glottique s'exprime donc dans le domaine en Z

$$G(z) = \frac{b_1 z}{1 + a_1 z + a_2 z^2} \frac{b_{TL}}{1 - a_{TL} z^{-1}}, \quad (6.1)$$

avec les coefficients

$$\begin{aligned} a_1 &= -2 e^{-a_p/f_s} \cos(b_p/f_s), & a_2 &= e^{-2a_p/f_s}, & b_1 &= \mathcal{G} \frac{\pi^2}{b_p^3} e^{-a_p/f_s} \sin(b_p/f_s) \\ a_p &= -\frac{\pi}{O_q T \tan(\pi\alpha_m)}, & b_p &= \frac{\pi}{O_q T}, \\ a_{TL} &= \nu - \sqrt{\nu^2 - 1}, & b_{TL} &= 1 - a_{TL}, & \nu &= 1 - \frac{1}{\eta}, & \eta &= \frac{1}{\frac{e^{-TL/10 \cdot \log(10)} - 1}{\cos\left(2\pi \frac{300}{f_s}\right) - 1}}, \end{aligned}$$

où f_s est la fréquence d'échantillonnage. Les paramètres de ce modèle sont

- \mathcal{G} le gain et T la période,
- O_q le quotient ouvert et α_m le coefficient d'asymétrie,
- TL l'atténuation du filtre passe-bas (pôle causale).

L'origine temporelle de ce modèle correspond à l'instant t_e des modèles précédents. Un retard est ensuite ajouté pour positionner temporellement l'impulsion glottique.

6.1.2 Choix du modèle d'étude

La liste de modèles présentée dans ce document est loin d'être exhaustive. Le lecteur pourra se référer à [CC95, DDH06, Deg10] pour un aperçu plus complet des modèles de source glottique existants. Néanmoins, il apparaît deux grandes familles de modèles. L'approche historique consiste à modéliser la source glottique à l'aide de fonctions temporelles définies par morceaux. Ils sont tous définis par des fonctions polynomiales, trigonométriques ou exponentielles tronquées. Ces troncatures génèrent des discontinuités dans la dérivée d'ordre n du signal de source glottique ($n = 2$ dans le cas d'une fermeture douce et $n = 1$ dans le cas d'une fermeture abrupte pour les modèles présentés ici). Or la régularité temporelle d'un signal et sa décroissance asymptotique en hautes fréquences dans le domaine de Fourier sont reliées comme suit.

Théorème 1. *D'après [Mal00, p.28-29]. Si $s \in \mathcal{C}^n$ et que $\forall k \in [1, n], s^{(k)} \in L^1(\mathbb{R})$ alors*

$$|S(f)| \leq \frac{\|s^{(n)}\|_1}{(2\pi|f|)^n}.$$

Ainsi la phase de retour dans les modèles KLGLOTT88, R++ et LF permet de régulariser le signal de source glottique et a pour effet de réduire son contenu haute fréquence.

Une approche plus récente, celle du modèle CALM, consiste à modéliser la source glottique comme la réponse impulsionnelle d'un filtre ayant une partie causale et une partie anti-causale. Ici la régularisation est naturellement assurée par l'absence de troncature dans la définition. Cette modélisation amène à la définition d'une notion bien connue en analyse du signal vocal : l'instant de fermeture glottique, que nous appellerons ici GCI, pour son équivalent anglais Glottal Closure Instant, courant dans la littérature sur le sujet [NKGB07, Deg10, Dru11]. Le GCI est généralement défini comme l'instant où la valeur absolue de la dérivée de la source glottique est maximale. Dans

tous les modèles de source glottique, cet instant correspond à l'instant de transition entre la phase descendante et la phase de retour (t_e en figure 6.1). Cet instant est considéré comme une origine temporelle pour le modèle CALM.

6.1.2.1 Pôles causaux et pôles anticausaux

Lorsque nous parlons de pôles causal et anticausal, il est nécessaire de préciser si la causalité et l'anticausalité sont strictes. Le tableau 6.2 présente les formes temporelle et spectrale (transformée en Z avec son domaine de convergence) des pôles causaux strict et non strict, et des pôles anticausaux strict et non strict, ainsi que les conditions de stabilité.

	causal		anticausal	
	strict	non strict	strict	non strict
$s(n)$	$-\mathbb{1}_{\mathbb{N}^*}(n) \cdot \gamma^n$	$\mathbb{1}_{\mathbb{N}}(n) \cdot \gamma^n$	$-\mathbb{1}_{\mathbb{Z}^*}(n) \cdot \alpha^n$	$\mathbb{1}_{\mathbb{Z}^-}(n) \cdot \alpha^n$
$S(z)$	$\frac{1}{1 - z/\gamma}$	$\frac{1}{1 - \gamma/z}$	$\frac{1}{1 - \alpha/z}$	$\frac{1}{1 - z/\alpha}$
domaine de convergence	$ z > \gamma $		$ z < \alpha $	
condition de stabilité	$ \gamma < 1$		$ \alpha > 1$	

TABLE 6.2 – Pôles causaux/anticausaux stricts/non stricts.

Dans la littérature, le choix habituel est d'utiliser des pôles causaux non stricts avec des pôles anticausaux stricts [OS75]. L'intérêt de ce choix est l'unicité de l'écriture de la transformée en Z (cf tableau 6.2) et le non-recouvrement temporel lors de l'addition d'un pôle causal et d'un pôle anti-causal (orthogonalité des signaux).

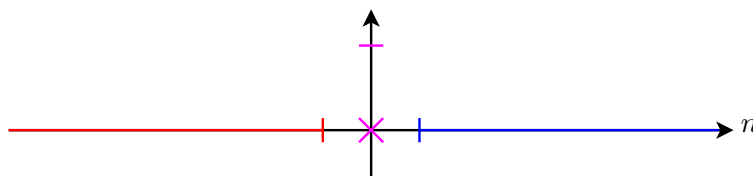
Dans l'étude présentée ici, nous choisissons de travailler avec des pôles causaux et anti-causaux non stricts afin de

1. différencier aisément les pôles causaux et anticausaux dès l'écriture,
2. symétriser le rôle des pôles causaux et anticausaux, notamment par retournement temporel.

Considérons un signal dont la transformée en Z est égale au produit d'un pôle causal (non strict) stable γ (avec $|\gamma| < 1$) et d'un pôle anticausal (non strict) stable α (avec $|\alpha| > 1$). Nous pouvons décomposer ce signal en

- un gain qui dépend des deux pôles,
- les échantillons en temps négatifs où seul le pôle anticausal influe,
- les échantillons en temps positifs où seul le pôle causal influe,
- l'échantillon à $n = 0$ qui dépend des deux pôles.

$$\begin{aligned}
 S(z) &= \frac{1}{\left(1 - \frac{\gamma}{z}\right) \left(1 - \frac{z}{\alpha}\right)} \\
 &= \frac{1}{1 - \frac{\gamma}{\alpha}} \left(\frac{z/\alpha}{1 - \frac{z}{\alpha}} + 1 + \frac{\gamma/z}{1 - \frac{\gamma}{z}} \right), \\
 s(n) &= \frac{1}{1 - \frac{\gamma}{\alpha}} \left(\mathbb{1}_{\mathbb{Z}^*}(n) \cdot \alpha^{n+1} + \delta(n) + \mathbb{1}_{\mathbb{N}^*}(n) \cdot \gamma^{n-1} \right).
 \end{aligned}$$



Remarque 1. Retournement temporel

La transformation $(z, \alpha, \gamma) \mapsto (z^{-1}, \alpha^{-1}, \gamma^{-1})$ et, en correspondance dans le domaine temporel, $(n, \alpha, \gamma) \mapsto (-n, \alpha^{-1}, \gamma^{-1})$ correspond à retournement temporel.

6.1.2.2 Définition du modèle de source glottique

Le choix du modèle d'étude s'est porté sur une adaptation du modèle CALM qui écrit les pôles causaux et anticausaux selon les choix de la section 6.1.2.1. Ce modèle définit la dérivée de l'impulsion glottique comme la réponse impulsionnelle d'un filtre dont la transformée en Z s'écrit

$$G(z) = \frac{z^{-d}}{\left(1 - \frac{a}{z}\right) \left(1 - \frac{z}{b}\right) \left(1 - \frac{z}{\bar{b}}\right)}, \quad (6.2)$$

où d représente le GCI, a est un pôle réel causal stable ($|a| < 1$) et b est un pôle anticausal stable ($|b| > 1$). Le paramètre a règle la phase de retour (étendue à l'infinie à droite). Il a exactement la même interprétation que le paramètre a_{TL} du modèle CALM. Le paramètre b règle la forme de l'impulsion glottique.

Les modèles périodiques de source glottique permettent de synthétiser uniquement des sons voisés. Pour les sons non voisés et les voix chuchotées, le signal de source glottique est en général remplacé par du bruit. Dans un premier temps, nous restreignons notre étude aux sons voisés.

6.2 Filtre vocal et rayonnement

Le deuxième élément constitutif du modèle source-filtre de la production vocale après la source glottique est le filtre vocal. Le filtre vocal modélise l'effet de filtrage du conduit vocal sur le signal de source glottique. Dans le paradigme source-filtre, cette influence non interactive (une modification du conduit vocal n'a pas d'influence sur le signal de source) est simplement représentée par un filtre linéaire. La littérature fait apparaître deux types de filtres vocaux, les filtres tous-pôles et les filtres pôles-zéros. Nous rappelons ci-après leurs définitions et propriétés principales, avant de motiver le choix de notre modèle.

6.2.1 Conduit vocal et filtre tous-pôles

Comme nous le verrons en détail dans la partie III, la propagation d'une onde acoustique dans le conduit vocal avec prise en compte d'un rayonnement idéalisé peut être modélisée par un filtre tous-pôles à coefficients réels dont les pôles sont tous causaux et stables [MG76]. La transformée d'un tel filtre s'écrit

$$V(z) = \frac{1}{\prod_{k=1}^K \left(1 - \frac{\alpha_k}{z}\right) \left(1 - \frac{\bar{\alpha}_k}{z}\right)}, \quad (6.3)$$

où $|\alpha_k| < 1 \forall k$. Chaque paire de pôles correspond à un formant, c'est-à-dire à un résonance du filtre vocal. La synthèse des différentes voyelles se fait alors par réglage de la fréquence et du facteur de qualité associés à chaque paire de pôles.

6.2.2 Filtre pôles-zéros

Il a été montré que l'ouverture des cavités nasales (cas des voyelles et consonnes nasales) peut être modélisée par des paires pôles/zéros dans le filtre vocal [NKvS05]. De plus, d'autres cavités connectées aux cavités nasales existent dans le crâne humain (sinus paranasaux) et sont susceptibles d'apporter des anti-résonances [Mae82b]. Néanmoins, les filtres tous-pôles sont encore largement utilisés dans l'analyse-synthèse du signal vocal et permettent l'utilisation de méthodes que la présence de zéros dans le filtre vocal rend impossible à mettre en œuvre (voir chapitre 7).

6.2.3 Filtre de rayonnement

Flanagan propose dans [Fla72] de modéliser le rayonnement aux lèvres par celui d'une sphère pulsante et obtient l'expression du son rayonné à une distance l de la bouche

$$S(l, \omega) = \frac{i\omega\rho}{4\pi l} \cdot e^{-\frac{i\omega l}{c}} \cdot U_L(\omega), \quad (6.4)$$

où $u_L(t)$ représente le débit aux lèvres obtenu par filtrage du débit glottique par le filtre vocal. Ne nous intéressant pas à la position précise du point de captation du son (en général non connu dans les enregistrements dont nous disposons), nous réduisons cette expression pour obtenir l'habituel filtre dérivateur modélisant le rayonnement aux lèvres [MG76, Deg10, RS11].

6.2.4 Définition du modèle d'étude

Notre choix porte sur

1. un modèle tous-pôles causal du filtre vocal, du type (6.3). Ce choix offre un cadre d'étude simple et permet l'utilisation d'algorithmes dédiés aux signaux tous-pôles, comme le premier algorithme que je proposerai dans le chapitre suivant.
2. Puis sur un modèle pôles-zéros pour réfléchir l'extension des algorithmes proposés à une classe plus réaliste.

6.3 Modèle source-filtre paramétrique complet étudié

L'étude que nous présentons ici est restreinte au cas des signaux discrets périodiques qui correspondent aux sons voisés quasi-stationnaires. En exploitant la linéarité du système, nous opérons le regroupement habituel de la source glottique et du filtre de rayonnement (cf. section 2.2.1) pour obtenir la chaîne simplifiée

$$\begin{aligned}
 S(z) &= \mathcal{G}G(z)V(z)E(z) \\
 &= \mathcal{G} \frac{z^{-d}}{\left(1 - \frac{a}{z}\right) \left(1 - \frac{z}{b}\right) \left(1 - \frac{z}{\bar{b}}\right)} \frac{1}{\prod_{k=1}^K \left(1 - \frac{\alpha_k}{z}\right) \left(1 - \frac{\bar{\alpha}_k}{z}\right)} E(z)
 \end{aligned} \tag{6.5}$$

où $E(z)$ est la transformée en Z du signal d'excitation $e(n)$, en général un peigne de Dirac de période T_0 $\text{III}_{T_0}(n)$.

Le vecteur de paramètres de ce modèle est donc

$$\theta = [\mathcal{G}, d, a, b, K, \{\alpha_k\}_{k \in [1, K]}, T_0]^T. \tag{6.6}$$

Dans cette partie, nous nous restreignons au cas des sons voisés non bruités. Le bruit sera systématiquement considéré comme une perturbation du signal. Ce point de vue est motivé par l'objectif des algorithmes qui est de donner une représentation paramétrique des signaux. Le bruit pourra être pris en compte par la suite pour des applications d'analyse-resynthèse.

Le chapitre suivant a pour objet l'estimation du vecteur de paramètres θ du modèle complet à partir d'extraits de signaux réels ou de synthèse.

Chapitre 7

Analyse du modèle source-filtre

Sommaire

7.1	État de l'art des méthodes d'estimation et de séparation	104
7.2	Premier algorithme : Séparation source-filtre pour la classe des signaux périodiques	105
7.2.1	General description of the algorithm	107
7.2.2	Space Π of infinitely periodic signals	108
7.2.3	Turning poles into zeros	109
7.2.4	Extracting the desired poles	109
7.2.5	Estimating the parameters d and \mathcal{G}	109
7.2.6	General remarks on the algorithm	110
7.2.7	Tests on synthetic signals	110
7.2.8	Discussion on robustness	111
7.2.9	Conclusion and perspectives	112
7.3	Deuxième algorithme : Analyse cepstrale et séparation source-filtre	115
7.3.1	Considérations générales sur le cepstre	115
7.3.1.1	Cepstre réel, cepstre complexe et déconvolution causale-anticausale	115
7.3.1.2	(re)Définition du cepstre complexe	117
	Rappels sur le logarithme complexe	117
	Cepstre complexe	120
	Méthodes de calcul	122
	Cepstre complexe et cepstre différentiel	124
7.3.2	Cepstre d'un signal tous-pôles causal/anti-causal	125
	Cepstre complexe	125
	Cepstres différentiels causaux et anti-causaux	126
7.3.3	Algorithme d'estimation des paramètres du modèle causal-anticausal	127
	Présentation de l'algorithme	127
	Maximisation de la vraisemblance des cepstres différentiels	128
	Remarques générales sur le deuxième algorithme.	129
	Ingédients à ajouter pour traiter des signaux vocaux mesurés.	129
7.3.4	Performances de l'algorithme d'estimation	129
7.3.4.1	Signaux de synthèse	129
7.3.4.2	Robustesse de l'algorithme	130
7.3.4.3	Signaux réels	136
7.4	Conclusion et perspectives	137

7.1 État de l'art des méthodes d'estimation et de séparation

Les problèmes de l'estimation des paramètres d'un modèle source-filtre et de la séparation source-filtre sont intrinsèquement liés. Considérons un modèle source-filtre décrit dans le domaine de Fourier par ¹

$$S(\omega) = E(\omega)G(\omega)V(\omega)L(\omega) + N(\omega). \quad (7.1)$$

Dans ce modèle,

- ω peut représenter la pulsation (temps continu) ou la fréquence réduite (temps discret),
- $S(\omega)$ est le son,
- $E(\omega)$ est l'excitation (en général un peigne de Dirac combiné à un retard pur),
- $G(\omega)$ est le filtre glottique,
- $V(\omega)$ est le filtre vocal,
- $L(\omega)$ est le filtre correspondant au rayonnement aux lèvres,
- $N(\omega)$ est la composante de bruit dans le signal.

L'identification des différentes composantes de ce modèle linéaire à partir du seul son émis ($S(\omega)$) permet évidemment d'opérer la séparation source-filtre a posteriori. Nous verrons cependant que les méthodes de séparation source-filtre ne nécessitent pas toutes d'estimer la totalité des composantes du modèle. Par ailleurs, comme évoqué au chapitre 6, le rayonnement aux lèvres, souvent modélisé par un filtre dérivateur, est combiné à la source glottique pour obtenir la source glottique dérivée dans de nombreux modèles.

Des méthodes ont été proposées pour répondre au problème de la séparation source-filtre, les principales sont présentées ci-après puis deux méthodes originales sont proposées.

La plus évidente et la plus utilisée est la technique dite de filtrage inverse [Mil59], qui permet, dans le cas non bruité, de retrouver une estimation de la source glottique à partir du filtre vocal et du filtre de rayonnement en écrivant

$$E(\omega)G(\omega) = \frac{S(\omega)}{V(\omega)L(\omega)}.$$

Plusieurs méthodes ont été proposées pour estimer le filtre vocal. Une des méthodes les plus répandues dans les applications d'analyse du signal vocal est la prédiction linéaire d'un modèle AR (autoregressif). Le modèle AR correspond à un modèle tous-pôles du filtre $G(\omega)V(\omega)L(\omega)$ sans excitation (l'excitation est entièrement codée dans les conditions initiales). La prédiction linéaire vise à estimer les coefficients du filtre AR, ou autrement dit, d'estimer une enveloppe spectrale paramétrique du signal analysé. Deux méthodes sont proposées pour résoudre ce problème : analyse de la fonction d'autocorrélation ou analyse de la matrice de covariance du signal d'analyse [MG76]. Cette méthode est abondamment utilisée, notamment dans le domaine des télécommunications (LPC ou Linear Predictive Coding) et plusieurs améliorations ont été proposées, par exemple pour réduire les dimensions du modèle AR [Per01] ou pour prendre en compte des conduits vocaux non stationnaires [SL08, SL09]. D'autres méthodes d'estimation des formants et de leurs caractéristiques sont passées en revue dans [RS11, p.649 et suivantes].

Une variante consiste à prendre en compte un modèle ARMA, c'est-à-dire un modèle avec pôles et zéros, qui permet, comme nous l'avons vu au chapitre précédent, de prendre en compte les zéros apportés par la nasalisation. Cependant, les zéros du modèle ARMA rendent le problème de prédic-

1. Les modèles de source glottique par fonctions paramétriques peuvent être mis sous la forme de filtres tous-zéros.

tion non linéaire, ce qui rend la résolution délicate [Ste77]. Néanmoins, plusieurs approches existent pour adapter la prédiction linéaire au cas ARMA [Mak75, KM79]. Il existe aussi des méthodes d'estimation du filtre par estimation de l'enveloppe spectrale qui passe par les pics de la transformée de Fourier [Sty95, CO98, VRR06].

L'algorithme STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum) proposé par Kawahara dans [Kaw97] propose une analyse du signal de parole dans laquelle la fenêtre d'analyse dépend de la période fondamentale. Cet algorithme permet une analyse adaptée à la transformation des signaux de voix avec un minimum d'artefacts.

Parmi les méthodes paramétriques, de nombreuses exploitent les modèles de source glottique paramétriques présentés au chapitre précédent. Nous pouvons notamment citer parmi ceux-là l'analyse des modèles ARX et ARMAX (AR et ARMA avec source exogène) pour lesquels la démarche de prédiction linéaire est adaptée au cas d'une source non triviale [FMT99]. D'autres méthodes paramétriques, comme celle proposée par Degottex [Deg10] permettent de retrouver les paramètres de la source glottique et/ou du filtre vocal à partir du son rayonné.

Enfin, un dernier type de méthodes de séparation source-filtre nous intéresse particulièrement : la séparation causale-anticausale. Si l'on omet le pôle réel représentant la phase de retour de la source glottique dans le modèle (6.5), le filtre glottique devient entièrement anticausal et le filtre vocal entièrement causal. Plusieurs méthodes proposent ainsi, pour répondre au problème de la séparation source-filtre, de séparer la composante causale de la composante anticausale du signal analysé [Qua02, RS11, DBD09, DBD11]. Deux approches principales sont développées pour effectuer cette séparation, la factorisation de la transformée en Z (ZZT [Boz05]) et la décomposition du cepstre complexe [Opp65].

A l'instar de ces dernières méthodes, je propose un algorithme exploitant la séparation intrinsèque de la source et du filtre à notre modèle d'étude causal-anticausal. Cet algorithme, présenté en section 7.2, part du constat qu'en traitement du signal numérique, les zéros sont plus faciles à identifier que les pôles. Un opérateur est alors proposé pour transformer en zéros les pôles d'un signal parfaitement périodique construit à partir d'un extrait du signal analysé. Cet algorithme donne satisfaction, il permet de retrouver tous les paramètres du modèle dans le cas des signaux de synthèse non bruités mais manque de robustesse, et aboutit à des résultats non fiables dans le cas des signaux réels. Des améliorations seront laissées en perspective, mes réflexions m'ayant conduit à choisir de poursuivre mes efforts de recherche sur la construction d'un second algorithme. Celui-ci s'appuie sur une nouvelle étude du cepstre complexe combiné à une méthode d'estimation incluant la prise en compte de la robustesse. Cet algorithme sera présenté en section 7.3.

7.2 Premier algorithme : Séparation source-filtre pour la classe des signaux périodiques

Rappelons tout d'abord le modèle d'étude. Nous considérons l'ensemble des signaux discrets générés par filtrage d'un train de Dirac idéal par le filtre glottique et le filtre vocal et amplifié par un gain \mathcal{G} . La transformée en Z d'un tel signal $s(n)$ s'écrit donc

$$\begin{aligned} S(z) &= \mathcal{G}V(z)G(z)\widetilde{\Pi}_{T_0}(z) \\ &= \mathcal{G} \cdot \frac{1}{\prod_{k=1}^K \left(1 - \frac{a_k}{z}\right)} \cdot \frac{z^{-d}}{\left(1 - \frac{a}{z}\right) \left(1 - \frac{z}{b}\right) \left(1 - \frac{z}{\bar{b}}\right)} \cdot \widetilde{\Pi}_{T_0}(z) \end{aligned} \quad (7.2)$$

où $\widetilde{\text{III}}_{T_0}(z)$ représente la transformée en \mathbb{Z} du peigne de Dirac $\text{III}_{T_0}(t)$ ². Les paramètres de ce signal sont les $K + 1$ pôles causaux stables (a et $\{a_k\}$), la paire de pôles anticausaux stables $\{b, \bar{b}\}$, le retard d qui définit le GCI, le gain \mathcal{G} et la période fondamentale T_0 . Le but de l'algorithme décrit ci-dessous est de retrouver les paramètres d'un tel signal à partir d'un extrait de longueur finie. Dans la suite, le vecteur contenant ces paramètres sera noté

$$\theta = \{a, b, \{a_k\}_{k \in [1, K]}, d, \mathcal{G}, T_0\}. \quad (7.3)$$

Je présente ici un algorithme que j'ai développé pour résoudre ce problème. Le texte de cette section, en anglais, est extrait d'un article présenté au congrès Interspeech 2013 et publié dans les actes du congrès [HHD13]. Un résumé en français est proposé en préambule.

Nota bene : Dans cette section uniquement, la TZ d'un signal $s(n)$ sera notée $\tilde{s}(z)$.

Résumé :

L'algorithme d'estimation présenté ici a pour but de retrouver les paramètres θ (7.3) du modèle (7.2) qui décrivent au mieux un extrait de signal quasi-stationnaire. Le schéma général de l'algorithme est présenté en figure 7.1. Il est constitué de 4 étapes et prend en entrée un signal de durée finie $y(n)$:

- E1** Détection de la période fondamentale T_0 .
- E2** Construction d'un signal infiniment périodique $s(n)$ à partir de $y(n)$.
- E3** Transformation des pôles de $s(n)$ en zéros.
- E4** Extraction des pôles du modèle par calcul de racines, du GCI d et du gain \mathcal{G} .

L'étape **E1** est réalisée par un algorithme de détection de périodicité standard. Nous avons utilisé ici l'algorithme YIN [dCK02].

Dans l'étape **E2**, nous construisons un signal $s(n)$ infiniment périodique à partir du signal d'entrée $y(n)$. Pour ceci, nous définissons l'opérateur \mathcal{P} qui consiste à sélectionner un extrait du signal $y(n)$ à l'aide d'une fenêtre $w_{T_0}(n)$ puis à le périodiser en le convoluant avec un peigne de Dirac idéal de période T_0 . Le choix de la fenêtre est libre mais doit respecter la condition standard d'addition-recouvrement $\sum_{k=-\infty}^{+\infty} w_{T_0}(n - kT_0) = 1 \forall n \in \mathbb{Z}$. Parmi les choix possibles, nous pouvons citer la sélection d'une des périodes présentes dans le signal $y(n)$, la moyenne de deux périodes adjacentes, ou la moyenne de toutes les périodes.

La transformation des pôles du signal parfaitement périodique $s(n)$ en zéros dans l'étape **E3** est assurée par l'opérateur \mathcal{H} défini par $\mathcal{H} : s \in \Pi \mapsto TFD^{-1}[1/TFD[s]] \in \Pi$, où Π désigne l'espace des signaux infiniment périodiques. Cette formulation de l'opérateur \mathcal{H} permet d'utiliser des algorithmes rapides de transformée de Fourier et de transformée de Fourier inverse sur une période

2. En toute rigueur, $\widetilde{\text{III}}_{T_0}(z)$ n'a de sens que pour $z \in \mathbb{S}$. Nous conservons néanmoins cette écriture car ceci n'aura pas de conséquences dans la suite.

du signal $s(n)$. L'opérateur \mathcal{H} peut être exprimé de façon exactement équivalente en remplaçant la TFD par la transformée en Z . Ainsi, les zéros de $\mathcal{H}[s]$ sont les pôles de s .

La dernière étape **E4** se déroule en trois sous-étapes. (a) La première sous-étape consiste à calculer les zéros de $\mathcal{H}[s](n)$ par factorisation de sa transformée en Z . La période du signal $s(n)$ contient T_0 échantillons, ainsi la factorisation de sa transformée en Z permet d'obtenir $T_0 - 1$ zéros. (b) La deuxième sous-étape consiste à effectuer une sélection parmi les zéros obtenus pour retrouver les $K + 3$ pôles correspondant au modèle (7.2) (sous la condition, toujours satisfaite en pratique, $K + 3 < T_0 - 1$). Trois stratégies sont proposées, consistant à minimiser soit l'erreur de reconstruction, soit la norme des coefficients non pris en compte lors de la factorisation, soit les résidus des pôles non sélectionnés. La première permet d'obtenir des meilleurs résultats mais nécessite plus de ressources. Les deux suivantes permettent un réglage du paramètre K manuel ou automatique. (c) La dernière sous-étape consiste à extraire le GCI (paramètres d) et le gain \mathcal{G} . Le GCI est retrouvé à l'aide de la position du premier coefficient sélectionné pour la factorisation et du nombre de pôles anti-causaux. Le gain est extrait par estimation de la puissance du signal d'entrée.

L'algorithme fournit des résultats satisfaisants pour des signaux de synthèse non bruités. En revanche, comme on s'y attend avec la description E1-E4, il est sensible à (P1) la qualité de l'estimation de la période fondamentale et la qualité de la période extraite du signal et (P2) au bruit et erreur de modélisation lors de lu calcul de $\mathcal{H}(s)$.

Cette limitation de robustesse se confirme pour les signaux réels. Des améliorations sont laissées en perspective (par exemple une régularisation de l'opérateur $\mathcal{H} \circ \mathcal{P}$ qui vise à réduire l'influence du bruit pour (P2)). Le choix a été fait dans mes travaux de porter mes efforts sur la construction d'un second algorithme dans lequel

- (a) la robustesse est prise en charge par une méthode à maximisation de vraisemblance,
- (b) les informations des pôles et du retard sont portées par le cepstre complexe, dont nous avons extrait la version analytique sur la couronne de convergence de $S(z)$.

Cette méthode est présentée dans le section 7.3.

7.2.1 General description of the algorithm

The main difficulty of poles estimation comes from the infinite-length of the support of pole-type signals. Finite-length signals have all-zeros Z -transforms. Hence, as we work in practice with finite-length signals, poles estimation seems doomed to fail. Another way of seeing the problem is that, as it has been highlighted in [Boz05] and [Dru11], windowing the signal has a drastic influence on the Z -transform that is extremely difficult to study analytically. The algorithm we propose offers a way to solve this problem by transforming poles into zeros in the class of periodic signals, making the support length finite.

The first step is to build an infinitely periodic signal s from the original extract. The second step is to turn the poles into zeros with an appropriate operator \mathcal{H} . The third step is to factorise the Z -transform of $\mathcal{H}[s]$ to compute its zeros, which are the poles of the Z -transform of s . Then, a selection of the meaningful zeros is performed. Finally, we can extract from the factorisation an estimation of the parameter d . A schematic representation of this algorithm is presented in Figure 7.1.

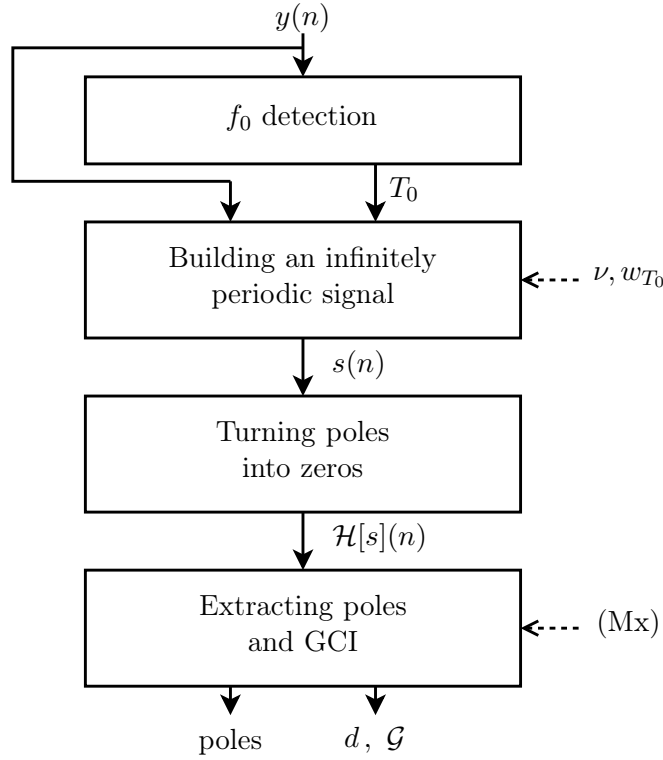


FIGURE 7.1 – General scheme of the all-pole causal/anticausal decomposition algorithm

7.2.2 Space Π of infinitely periodic signals

In order to build an infinitely periodic signal $s(n)$, we need to know the periodicity of the input signal $y(n)$. This can be done with the autocorrelation technique, the YIN algorithm [dCK02], or any other f_0 estimation algorithm. In the actual version of the algorithm, we consider only integer periods T_0 (expressed in samples).

Signal $s(n)$ of space Π is defined from y by

$$s(n) = \mathcal{P}[y](n) \triangleq (y(n) \times w_{T_0}(n)) * \text{III}_{T_0, \nu}(n) \quad (7.4)$$

where $*$ stands for the convolution operator and ν is a chosen instant representing the beginning of one period. $w_{T_0}(n)$ is a window that can be chosen to “mean” several periods of the signal $y(n)$ but must verify the property

$$\sum_{k=-\infty}^{+\infty} w_{T_0}(n - kT_0) = 1 \quad \forall n \in \mathbb{Z}. \quad (7.5)$$

This property ensures that, if $y(n)$ is a truncated version of an infinitely periodic signal, $s(n)$ retrieves the exact original infinitely periodic signal (standard overlap-add condition [OSB99]). For example, one can choose

$$(W1) \quad w_{T_0}(n) = \mathbb{1}_{[\nu, \nu+T_0-1]}(n) \quad \text{or}$$

$$(W2) \quad w_{T_0}(n) = \cos^2\left(\frac{(n-\nu)\pi}{2T_0}\right) \mathbb{1}_{[\nu-T_0, \nu+T_0-1]}(n).$$

Note that (W1) selects one period of the signal beginning at the instant ν , (W2) averages two periods of the signal around the instant ν .

7.2.3 Turning poles into zeros

Turning poles into zeros can be done using the operator

$$\mathcal{H} : s \in \Pi \mapsto DFT^{-1} \left(\frac{1}{DFT(s)} \right) \in \Pi \quad (7.6)$$

where s is the infinitely periodic signal (7.4), DFT stands for the Discrete Fourier Transform and DFT^{-1} for its inverse. It is obvious that $\widetilde{\mathcal{H}}[s](z)$ is the inverse of $\tilde{s}(z)$. Hence, poles of $\tilde{s}(z)$ are the zeros of $\widetilde{\mathcal{H}}[s](z)$ and vice versa.

7.2.4 Extracting the desired poles

Computing the zeros of the signal $\mathcal{H}[s](n)$ is possible with numerical methods. As we do not want that the periodicity of the signal $s(n)$ interfere with the poles research, we simply have to factorise the Z-transform of one period of $\mathcal{H}[s](n)$, noted $\widetilde{\mathcal{H}}[s](z)$. In practice, $\mathcal{H}[s]$ is computed over one period of $s(n)$ using the FFT algorithm. The resulting signal $\mathcal{H}[s](n)$ is then factorised with a numerical roots finder to compute its zeros. These zeros are the poles of $\tilde{s}(z)$.

The only question left is how to select the poles. The complete factorisation of $\widetilde{\mathcal{H}}[s](z)$ gives $T_0 - 1$ poles. However, the model we proposed has $K + 3$ poles. Three methods are proposed to reduce and/or impose the number of poles.

- (M1) The factorisation is performed on $K + 4$ consecutive coefficients in $\widetilde{\mathcal{H}}[s](z)$. The selection of these $K + 4$ coefficients is done by minimising the reconstruction error, which is defined as the 2-norm of the difference between the complex spectra of the original signal and the reconstructed signal.
- (M2) The factorisation is performed on $K + 4$ consecutive coefficients in $\widetilde{\mathcal{H}}[s](z)$ (consecutive in the sense of circular permutations), ensuring to retrieve $K + 3$ poles. The selection of these $K + 4$ coefficients is done by minimising the norm (n -norm, $n \in \bar{\mathbb{N}}^*$) of the unselected coefficients in $\widetilde{\mathcal{H}}[s](z)$. This method amounts to select the most influential coefficients.
- (M3) The factorisation is performed on the whole $\widetilde{\mathcal{H}}[s](z)$. The $K + 3$ poles with the maximum residues are selected, which amounts to select the most influential poles.

Note that (M2) and (M3) gives us the liberty to let the algorithm decide the number of poles. It simply needs to replace the minimisation of the remaining coefficients (or residues) by thresholding the remaining coefficients (or residues).

Finally, the separation between the causal and the anticausal component is automatically achieved by selecting the pole inside the unit circle for the causal component and outside the unit circle for the anticausal component. Hence, it is possible to separate the anticausal part of the glottal source from the rest of the signal.

7.2.5 Estimating the parameters d and \mathcal{G}

Estimating the parameter d amounts to detecting the position of the unique GCI inside the period of $s(n)$. It appears that this is automatically done by the algorithm in the previous step. The factorisation of $\widetilde{\mathcal{H}}[s](z)$ gives M_c causal poles (of absolute value smaller than one) and M_a anticausal poles (of absolute value greater than one). Using the definition of the causality, we can

recover d ,

$$d = \nu + d_{opt} + M_a, \quad (7.7)$$

where ν is the window position chosen in section 7.2.2 and d_{opt} is the position of the first coefficient selected by the (Mx) method in section 7.2.4.

The gain \mathcal{G} is retrieved by estimating the power of the input signal.

7.2.6 General remarks on the algorithm

As we'll see on section 7.2.7, this algorithm lets us exactly recover the parameters for signals corresponding to the model, in the “ideal case”. Moreover, one can easily show that, in this case, the algorithm results are independent of the choice of

- w_{T_0} as long as it verifies (7.5),
- ν as long as the support of y contains the support of w_{T_0} .

However, these choices (and the choice between (M1), (M2) and (M3)) can be very important in other cases. In particular, we observed that (M1) gives the best results but it's also the most resources-consuming method. It is interesting to highlight that in the ideal case, $\tilde{s}(z)$ is precisely the same as $GV(z)H(z)$ (for the right choice of ν). Another way of saying this is that one period of $s(n)$ is a periodic summation of the impulse response to the filter $GV(z)H(z)$ and then contains the whole information about the filter.

7.2.7 Tests on synthetic signals

We build an “almost ideal case” by filtering a very long Dirac comb of period T_0 and then extract a few periods in the middle of the filtered signal. This ensures that the part of the infinite-response of the filter which is not taken into account is negligible. The glottal source and filter parameters are chosen within classic human speech values. Complex conjugate pairs of poles are generated from frequency and Q factor values. Figure 7.2 illustrates the behaviour of the algorithm with choices

- ν set at the half of the length of signal y ,
- (W2),
- (M2) with 2-norm,

on a synthetic signal with the parameters

- $Fs = 10$ kHz, $f_0 = 200$ Hz, $G = 1$, $K = 2 \times 4$,
- $a = 0.8$, $f_b = 300$ Hz, $Q_b = 3$,
- $\{f_k\} = \{0.9, 1.2, 3, 4\}$ kHz, $\{Q_k\} = \{5, 15, 40, 15\}$.

Results of the algorithm are presented on Figure 7.3. One can see that the reconstruction is perfect for this almost ideal case. Poles are exactly estimated and the signal is very precisely reconstructed. Note that the GCIs are perfectly estimated.

The algorithm is working perfectly in the ideal case, we tested the algorithm with slight shifts from the ideal case. Still in synthetic speech, we tested the influence of

- a bad estimation of the fundamental period T_0 ,
- the presence of noise in the signal,
- the number of estimated poles.

Unfortunately, the two first bring drastically down the performances of the algorithm. A slight error in the T_0 estimation makes the poles estimations far from the truth. Introducing some noise in the signal makes the reconstructed signal tend to a flat-spectrum signal. As for the last point,

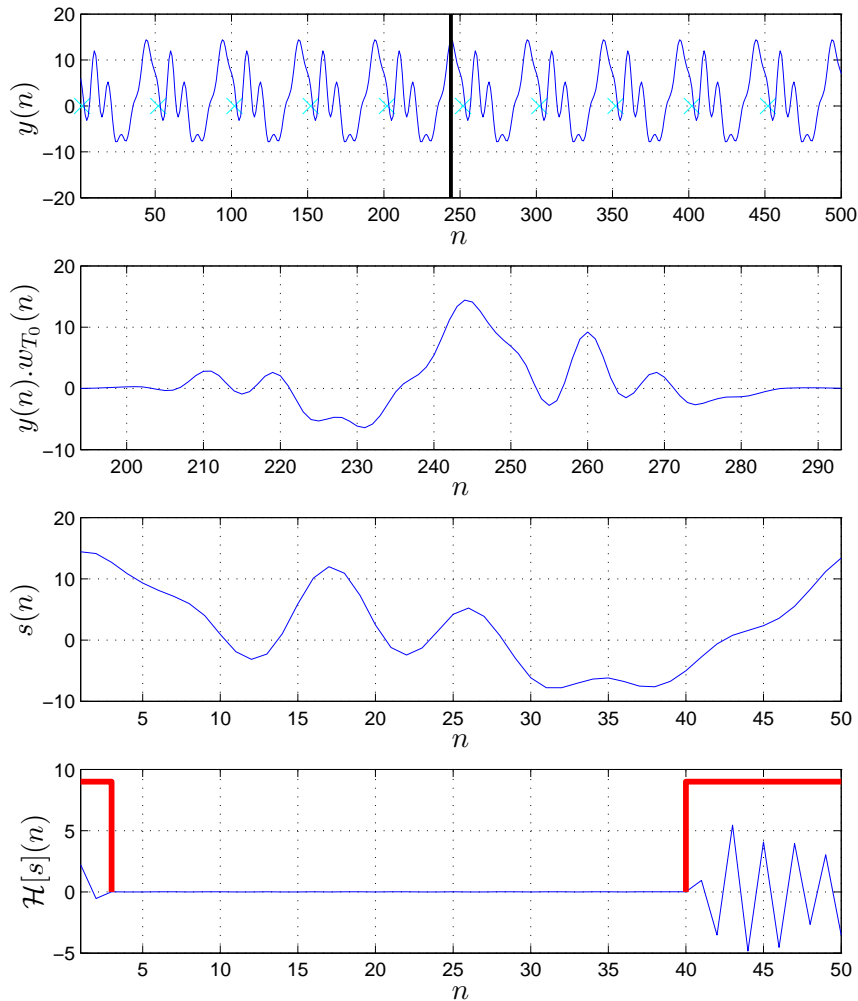


FIGURE 7.2 – Illustration of the algorithm. From top to bottom : 10 periods of an “almost ideal case” signal with GCIs (x) and the choice of ν (|), $y(n) \cdot w_{T_0}(n)$ with the choice (W2), one period of $s(n)$, $\mathcal{H}[s](n)$ with the optimal choice of $K + 3$ coefficients.

the algorithm can't reconstruct the signal if the number of estimated poles is smaller than the real number of poles but gives perfect results if the number of estimated poles is greater than the real number of poles. In this latter case, the algorithm finds some poles that are not in the original signal but with either very small absolute value or very small residue, so that their influence on the reconstruction is negligible.

7.2.8 Discussion on robustness

As one can guess given the results with non ideal cases, our algorithm is still very uncertain for real speech signals. We tested it on sustained vowels pronounced by a male speaker in a low register. Figures 7.4 and 7.5 present results of the analysis on a segment of a vowel /e/ with a fundamental frequency of 73 Hz. Figure 7.4 presents the results obtained using a model with 33 poles

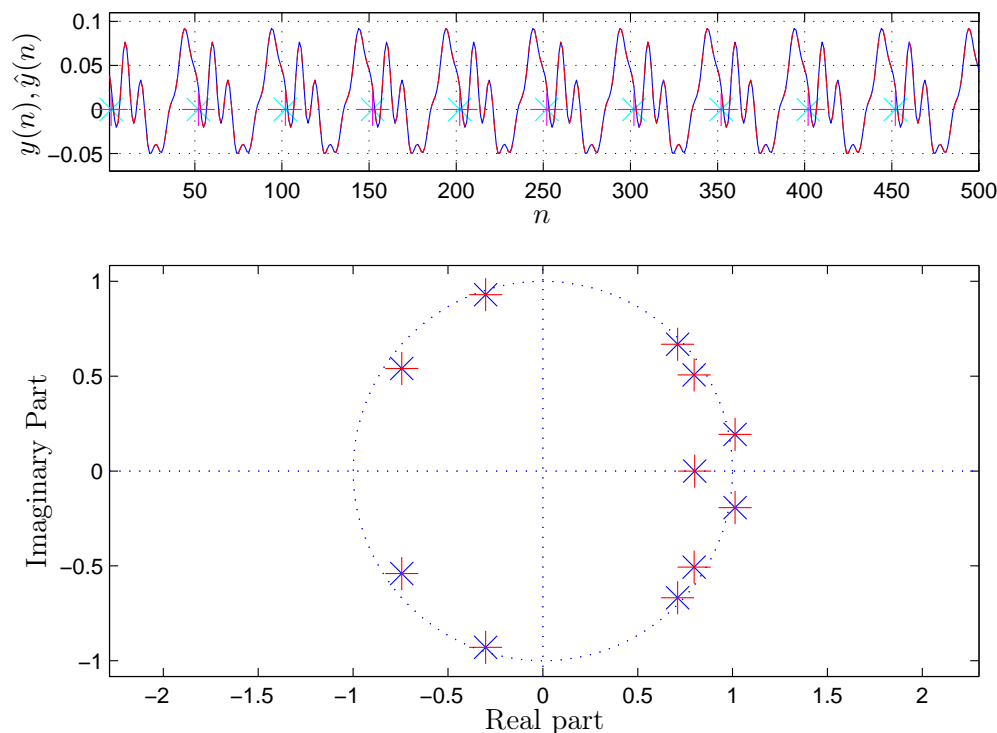


FIGURE 7.3 – Results of the algorithm. From top to bottom : signal $y(n)$ (-) and reconstructed signal $\hat{y}(n)$ (--) along with GCIs (x) and estimated GCIs (+), poles of the model (x) and estimated poles (+).

and Figure 7.5 presents the results obtained using a model with 13 poles. This latter corresponds to the classic choice described in chap 6. If the performances are globally poor, we can observe several informative behaviours of the algorithm. At first, with a low order model, the causal/anticausal decomposition tends to correspond to a low frequency / high frequency decomposition, which is globally coherent with the glottal/source decomposition. Then, we can observe that the glottal formant frequency seems to be well estimated in each case. Finally, we can observe that the reconstruction is much better for the low frequencies than for the high frequencies. This is probably due to the sensitivity of the algorithm to the noise.

7.2.9 Conclusion and perspectives

We presented a new algorithm for parametric causal/anticausal decomposition of speech signals. We demonstrated that the algorithm is perfectly effective for signals inside the space Π . This algorithm led us to define a new operator on Π : operator \mathcal{H} which turns poles into zeros and is easy to implement. Unfortunately, we saw that the algorithm is severely sensitive to noise and errors on the T_0 estimation. However, some regularisation techniques are under consideration and may improve the performances of this algorithm. Firstly, the operator \mathcal{H} could be computed using a Wiener deconvolution to decrease the sensitivity to noise. Another perspective to make the method more robust would consist of regularising $\mathcal{H} \circ \mathcal{P}$ by considering a model with a reduced number of poles from the beginning. At last, a variant of this algorithm taking into account the noise in the signal and using a likelihood maximisation on the complex cepstrum is currently being developed.

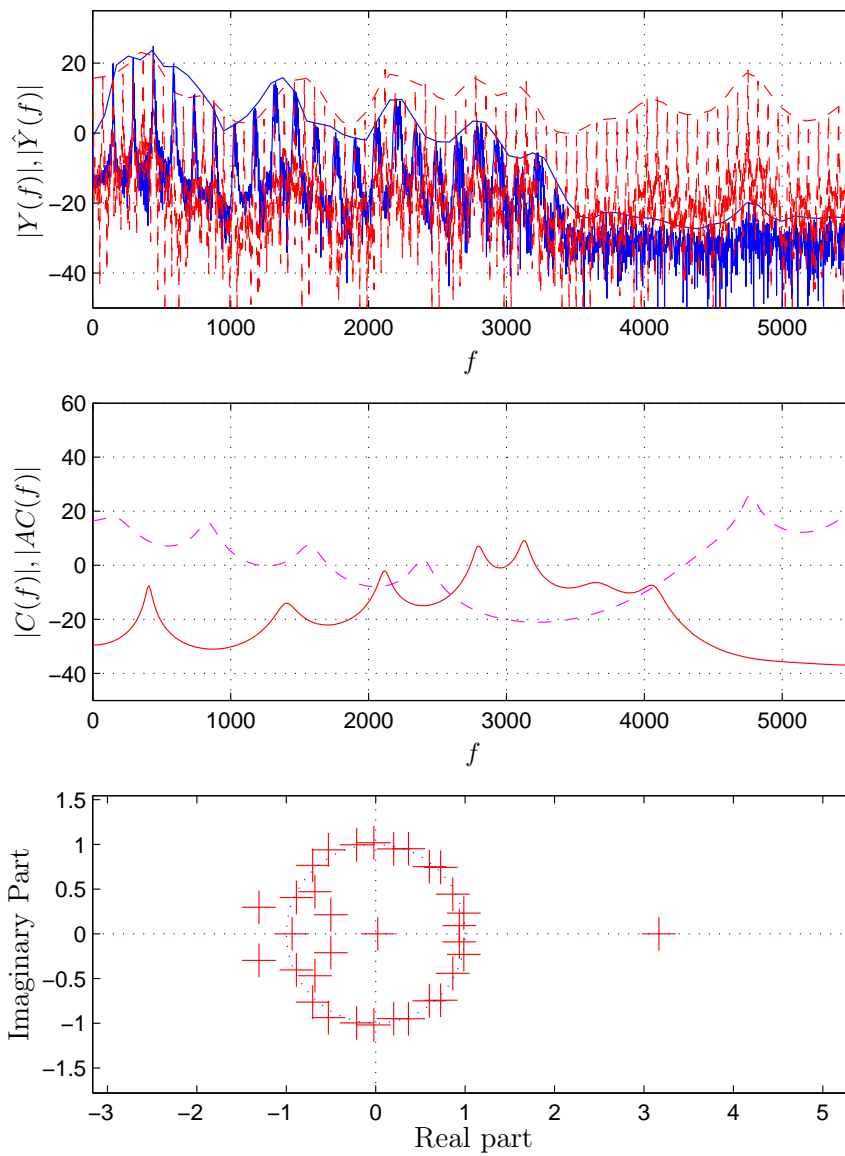


FIGURE 7.4 – Results of the algorithm for real speech with high order model. Top : Spectrum of original (-) and reconstructed signal (--). Middle : Spectrum of causal (--) and anticausal (-) components. Bottom : estimated poles.

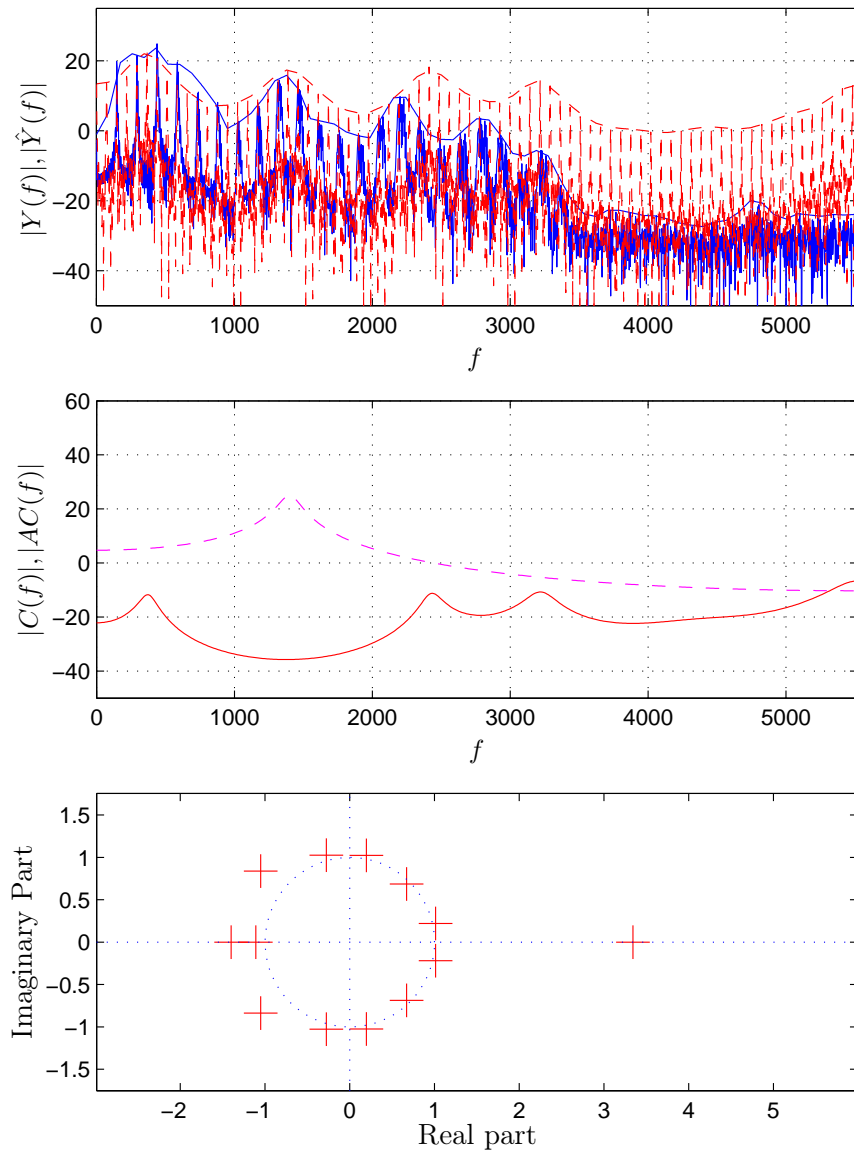


FIGURE 7.5 – Results of the algorithm for real speech with low order model. See Fig. 7.4.

7.3 Deuxième algorithme : Analyse cepstrale et séparation source-filtre

Je propose ici d'améliorer la robustesse en exploitant deux ingrédients.

I1 Une maximisation de vraisemblance paramétrique sur

I2 une mise en forme du signal qui isole bien les contributions des pôles anticausaux, des pôles causaux et de la position du GCI : le cepstre complexe.

Tout d'abord, une difficulté que je résous, dans la section 7.3.1, est la construction du cepstre complexe. L'approche consiste à chercher une version analytique du logarithme de $S(z)$ sur la couronne de convergence de $S(z)$ (qui inclut le cercle unité) plutôt que sur le cercle unité seul afin de résoudre le problème de définition multiple du logarithme complexe. Cette approche, à ma connaissance, n'a pas été utilisée dans la littérature. Elle conduit à un résultat plus simple que les calculs de logarithme complexe par factorisation de la transformée en Z et plus robuste que le déroulement de la partie imaginaire du logarithme de la transformée de Fourier (fréquemment utilisés, voir [Dru11, RS11]) : le calcul du cepstre complexe à l'aide du cepstre différentiel. Le cepstre complexe du modèle tous-pôles d'étude est calculé dans la section 7.3.2 dans le cas d'un signal non périodique (réponse impulsionnelle) et nous en déduisons un algorithme d'estimation fondé sur les ingrédients **I1** et **I2**, présenté dans la section 7.3.3. Ce travail conduit à une estimation performante des paramètres du modèles. Une première application à l'estimation de la source glottique donne des résultats encourageants et des perspectives pour la poursuite des travaux seront proposées.

7.3.1 Considérations générales sur le cepstre

7.3.1.1 Cepstre réel, cepstre complexe et déconvolution causale-anticausale

Le cepstre est outil de traitement du signal introduit en 1963 par Bogert, Healy and Tuckey [BHT63]. La définition originale du cepstre (réel) d'un signal par Bogert et al. peut se résumer en la transformée de Fourier du logarithme du module de la transformée de Fourier du signal. Suite à ces premiers travaux, une définition précise et rigoureuse du cepstre d'un signal discret était nécessaire. La définition du cepstre réel $c(n)$ d'un signal discret $s(n)$ la plus communément utilisée aujourd'hui provient des travaux de Oppenheim et Schafer [OSS68, Sch69, OSB99] et peut se résumer comme suit

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |S(\omega)| e^{i\omega n} d\omega, \quad (7.8)$$

$$S(\omega) = \sum_{n=-\infty}^{\infty} s(n) e^{-i\omega n}. \quad (7.9)$$

La définition est très proche de la définition initiale, à ceci près que le logarithme est transformé en Fourier inverse et non en Fourier direct. Nous utiliserons dans la suite de ce chapitre le vocabulaire français habituel associé au cepstre appelé dans le tableau 7.1.

Les premières utilisations du cepstre par Bogert et al. puis Schafer se sont concentrées sur la détection et la suppression d'écho. Les premières applications de l'analyse cepstrale dans le contexte du signal de parole ont été proposées par Noll [Nol64, Nol67]. Diverses applications dans le domaine du traitement du signal et plus particulièrement du signal vocal ont suivi. Le lecteur intéressé peut se référer à [CSK77, OSB99, RS11] pour une revue complète des différentes applications.

spectre	Transformée de Fourier (TF) d'un signal
cepstre	Transformée de Fourier Inverse du log de la TF d'un signal
analyse	Calcul du spectre d'un signal
alalyse	Calcul du cepstre d'un signal
filtrage	Opération linéaire effectuée sur un spectre
liftrage	Opération linéaire effectuée sur un cepstre
fréquence	Variable indépendante du spectre
quéfrence	Variable indépendante du cepstre
harmonique	Multiple de la fréquence fondamentale
rahmonique	Multiple de quéfrence fondamentale

TABLE 7.1 – Vocabulaire de l'analyse cepstrale.

Les recherches d'Oppenheim sur le cepstre l'ont mené à développer la théorie des systèmes homomorphiques [Opp65, OSS68] et à définir le cepstre complexe. Le cepstre complexe $\hat{s}(n)$ d'un signal discret $s(n)$ peut être défini de manière analogue à la définition du cepstre réel :

$$\hat{s}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[S](\omega) e^{i\omega n} d\omega, \quad (7.10)$$

$$S(\omega) = \sum_{n=-\infty}^{\infty} s(n) e^{-i\omega n}, \quad (7.11)$$

où $\log[S](\omega)$ représente un logarithme complexe de $S(\omega)$. L'intérêt du cepstre complexe, tient en une propriété fondamentale, héritée du logarithme, présentée par Oppenheim dès le début de ses travaux sur les systèmes homomorphiques [Opp65, OSS68],

$$\text{le cepstre transforme la convolution dans le domaine temporel en} \\ \text{somme dans le domaine quéfrentiel.} \quad (7.12)$$

Cette propriété permet de simplifier le problème de déconvolution de la source et du filtre en l'abordant dans le domaine cepstral et constitue la base de l'algorithme présenté ci-après. Cependant, elle s'oppose au problème de la définition et de l'unicité du logarithme complexe dans le calcul du cepstre complexe [OSS68, OS75, OSB99, RS11]. Pour illustrer ce problème, prenons l'exemple d'une transformée en Z constituée de deux pôles définie sur sa couronne de convergence \mathcal{C}_S

$$S(z) = \frac{1}{(z - z_1)(z - z_2)} \quad \forall z \in \mathcal{C}_S. \quad (7.13)$$

Les parties imaginaires de $\log\left(\frac{1}{z-z_1}\right)$ et $\log\left(\frac{1}{z-z_2}\right)$ sont présentées sur la figure 7.6- (a) et 7.6- (b). La partie imaginaire de $\log(S(z))$ est présentée en 7.6- (c) et celle de $\log\left(\frac{1}{z-z_1}\right) + \log\left(\frac{1}{z-z_2}\right)$ en 7.6- (d). Nous pouvons constater que $\log(S(z))$ n'est pas égal à $\log\left(\frac{1}{z-z_1}\right) + \log\left(\frac{1}{z-z_2}\right)$.

Je propose ici d'étudier ce problème en portant une attention particulière à la définition du logarithme complexe, ce qui nous mènera à étudier l'analyticité et l'holomorphicité du logarithme complexe en tant qu'opérateur. Un calcul du cepstre complexe permettant d'assurer la propriété $\log(S_1(z)S_2(z)) = \log(S_1(z)) + \log(S_2(z))$ (à l'origine de la propriété 7.12) est proposé.

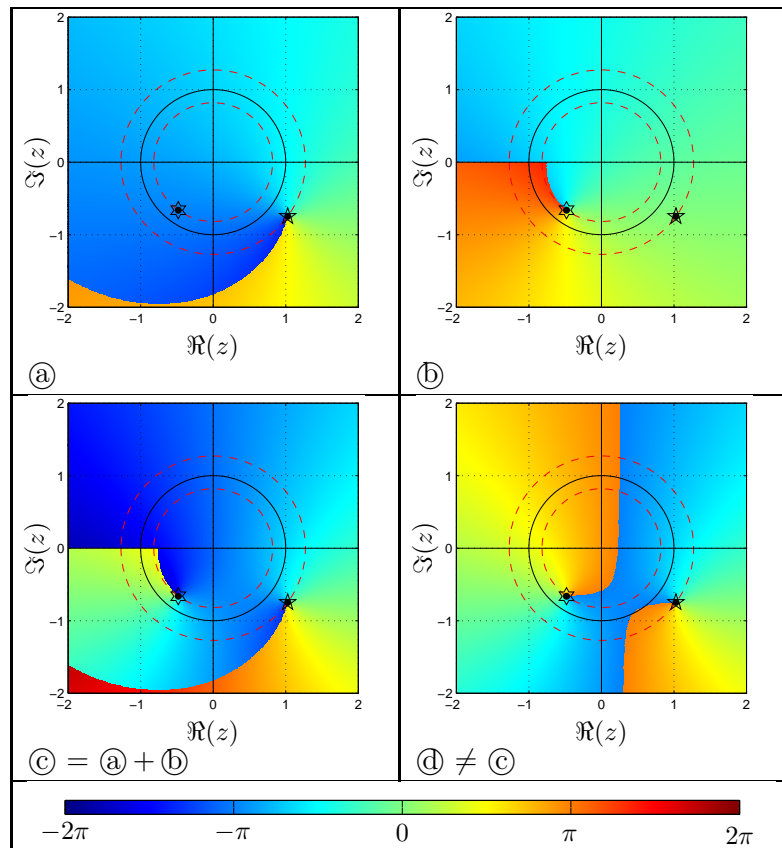


FIGURE 7.6 – Partie imaginaire du logarithme de la transformée du signal à deux pôles (7.13).
 (a) $\log\left(\frac{1}{z-z_1}\right)$. (b) $\log\left(\frac{1}{z-z_2}\right)$. (c) $\log\left(\frac{1}{z-z_1}\right) + \log\left(\frac{1}{z-z_2}\right)$. (d) $\log(S(z)) \neq \log\left(\frac{1}{z-z_1}\right) + \log\left(\frac{1}{z-z_2}\right)$.

7.3.1.2 (re)Définition du cepstre complexe

Le cepstre reposant sur le calcul d'un logarithme de la transformée de Fourier d'un signal, il est important de rappeler quelques définitions et propriétés du logarithme complexe.

Rappels sur le logarithme complexe

On appelle logarithme d'un nombre complexe a non nul (cf. par exemple [Sar07, p.40]), tout nombre complexe b tel que

$$e^b = a. \quad (7.14)$$

Le logarithme complexe général Log , tel qu'il est défini dans [AS64, p.67], est déduit de (7.14) et s'écrit

$$\text{Log}(r e^{i\theta}) \triangleq \ln(r) + i(\theta + 2k\pi) \quad \forall (r, \theta) \in \mathbb{R}^{+*} \times]-\pi, \pi], k \in \mathbb{Z}. \quad (7.15)$$

Pour chaque valeur de k , on parle d'une branche du logarithme.

On appelle détermination principale du logarithme complexe le logarithme dont la partie imaginaire se trouve dans l'intervalle $]-\pi, \pi]$. Nous le noterons par la suite \log . Il peut être défini

par

$$\log(z) \triangleq \ln |z| + i \arg(z) \quad \forall z \in \mathbb{C}^*, \quad (7.16)$$

où $\arg(z)$ représente la mesure principale de l'argument du complexe z . Ce logarithme correspond au choix $k = 0$ dans (7.15). La fonction \log définie en (7.16) est coupée sur \mathbb{R}^- , c'est-à-dire qu'elle présente une infinité de singularités (de type "saut") sur cette demi-droite.

La fonction \log (7.16) n'est évidemment pas le seul logarithme. Il est montré que tous les logarithmes complexes, c'est-à-dire les fonctions f telles que $e^{f(z)} = z$, sont définis sur \mathbb{C} privé d'une demi-droite issue de l'origine quelconque [Sar07]. Pour définir un logarithme, il suffit de choisir l'angle de cette droite par rapport à la demi-droite \mathbb{R}^- (ou une autre référence) et de définir la valeur du logarithme en un point [Sar07]. On peut alors modifier la définition (7.16) pour obtenir

$$\log_{\theta_0}(z) \triangleq \ln |z| + i \arg_{\theta_0}(z) \quad \forall z \in \mathbb{C} \setminus e^{i\theta_0}\mathbb{R}^-, \quad (7.17)$$

où $\theta_0 \in \mathbb{R}$ et $\arg_{\theta_0}(z)$ est la détermination de l'argument de z à valeur dans $]\theta_0 - \pi, \theta_0 + \pi]$. Cette définition est simple à visualiser sur la surface de Riemann associée au logarithme complexe [Val66, p.350 et suivantes]. La figure 7.7 présente cette surface de Riemann ainsi que les branches du logarithme correspondant à deux choix de θ_0 , le choix canonique $\theta_0 = 0$ et un choix arbitraire $\theta_0 = -1, 2\pi$.

Remarque 2. La fonction \log définie en (7.16) est une fonction holomorphe³ de $\mathbb{C} \setminus \mathbb{R}^-$ dans \mathbb{C} . La fonction \log_{θ_0} définie en (7.17) est une fonction holomorphe de $\mathbb{C} \setminus e^{i\theta_0}\mathbb{R}^-$ dans \mathbb{C} . Ces deux fonctions ont pour dérivée $z \mapsto 1/z$ sur leur ensemble de définition.

Considérons maintenant une fonction f holomorphe et non nulle sur un ouvert simplement connexe U de \mathbb{C} . Par définition [Val66, p.350 et suivantes], g un logarithme de f sur U est holomorphe sur U et tel que

$$e^g = f. \quad (7.18)$$

Autrement dit, il s'écrit

$$\log[f](z) = \log(f(z)) + 2k(z)i\pi \quad \forall z \in U, \quad (7.19)$$

avec $k(z) \in \mathbb{Z}$ de sorte que la continuité (et l'holomorphie) de g sur U soit assurée (un exemple est proposé plus loin, illustré par la figure 7.8). Une fois encore, une infinité de choix (de branches) est possible dans la définition de $\log[f]$. Sauf mention contraire, nous choisirons dans la suite systématiquement la branche correspondant à (7.19) telle que

$$\log[f](z) \in \mathbb{R} \text{ lorsque } f(z) \in \mathbb{R}^{+*}. \quad (7.20)$$

Remarque 3. On alors $\log[f]' = \frac{f'}{f}$.

Remarque 4. Les logarithmes (7.15), (7.16) et (7.17) ne sont autres que des déterminations ou des ensembles de branches du logarithme de l'identité.

3. Une fonction f à valeur complexe est holomorphe sur un ouvert U de \mathbb{C} si et seulement, en tout point z_0 de U , la limite suivante, appelée dérivée de f en z_0 , existe :

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}.$$

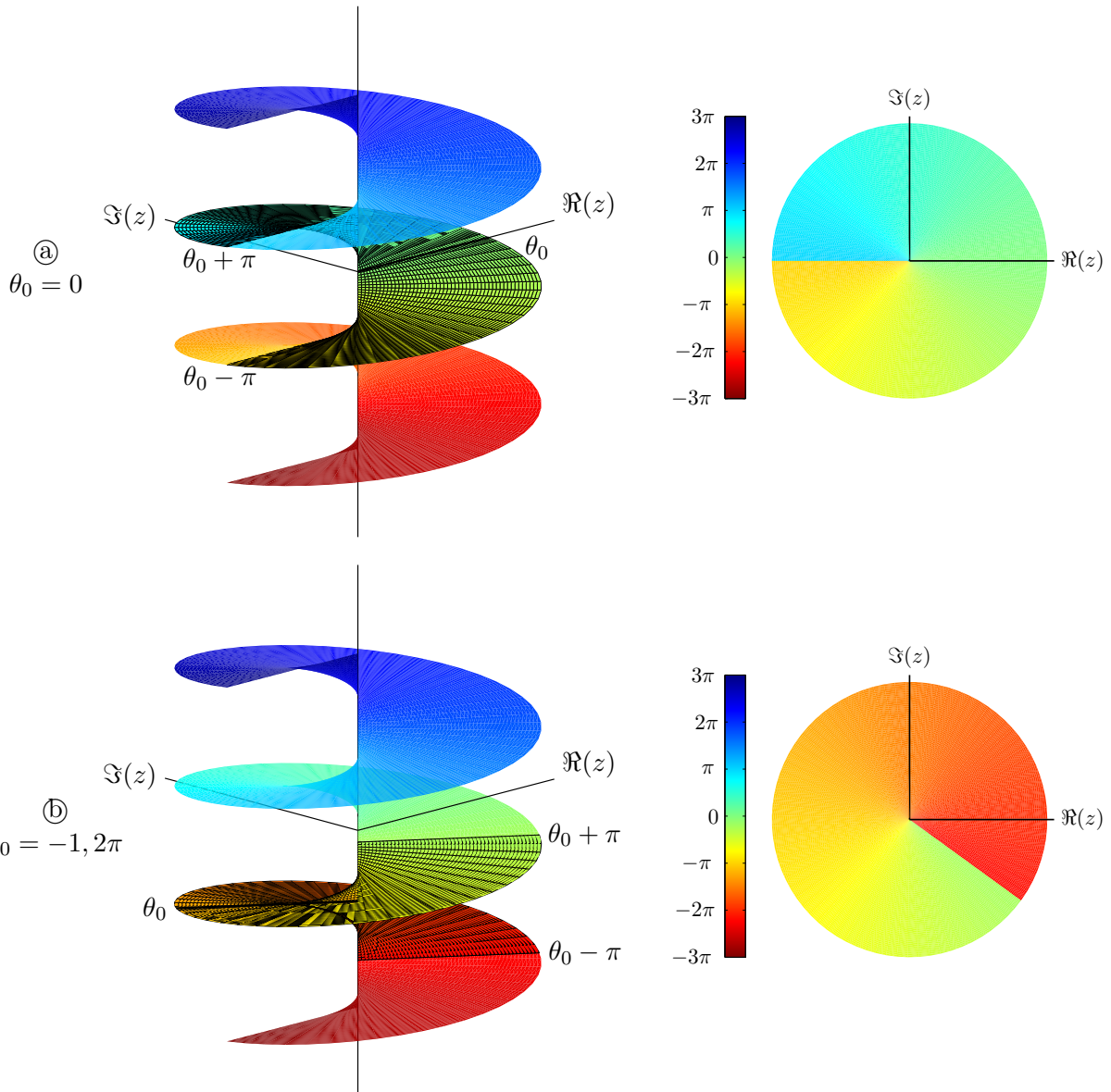


FIGURE 7.7 – Surface de Riemann associée au logarithme et choix d’une branche. A gauche, la partie hachurée de la surface de Riemann correspond à la branche du logarithme choisie. La partie imaginaire du logarithme correspondant est présentée sur le plan complexe à droite. On remarque que, pour le choix ②, la partie imaginaire du logarithme d’un nombre réel positif n’est pas nulle.

<i>signal</i>	<i>cepstre</i>
$\delta(n)$	0
$a^n H(n), a < 1$	$\frac{a^n}{n} H(n-1)$
$a^n H(-n), a > 1$	$-\frac{a^n}{n} H(-n-1)$
$\delta(n) - b\delta(n-1), b < 1$	$-\frac{b^n}{n} H(n-1)$
$\delta(n) - \frac{1}{b}\delta(n+1), b > 1$	$\frac{b^n}{n} H(-n-1)$
$s_1 * s_2(n)$	$\hat{s}_1(n) + \hat{s}_2(n)$
$\alpha s(n), \alpha \in \mathbb{C}^*$	$\ln(\alpha)\delta(n) + \hat{s}(n)$
$s(\alpha n), \alpha \in \mathbb{R}^*$	$-\ln(\alpha)\delta(n) + \alpha\hat{s}(\alpha n)$

TABLE 7.2 – Dictionnaire de cepstres utiles.

Cepstre complexe

Soit $s(n)$ un signal discret tel que sa transformée en Z $S(z)$ est holomorphe et non nulle sur un ouvert simplement connexe U de \mathbb{C} contenant le cercle unité \mathbb{S} (il s'agit en général d'une couronne de convergence). Alors nous définissons le cepstre complexe

$$\hat{s}(n) = \text{TZ}^{-1} \left[\log[\text{TZ}[s]](z) \right] (n), \quad (7.21)$$

qui est défini de façon unique grâce à (7.19-7.20). Avec cette définition, le calcul de $\log[S](\nu)$ pour l'inversion de la TZ ne fait pas apparaître de singularité dans sa partie imaginaire sur U car l'holomorphie de S et de $\log(S)$ sur U est respectée. De plus, comme \mathbb{S} est contenu dans U , on remarque que (7.21) est équivalent à (7.10-7.11) avec la définition du logarithme que nous venons de rappeler.

Le tableau 7.2 présente un tableau récapitulatif des cepstres complexes et des propriétés du cepstre complexe qui seront utiles dans la suite de notre étude. Le détail des calculs est présenté en annexe D.

Nous comprenons ici que la notion centrale dans la définition du cepstre complexe est la préservation de l'holomorphie dans le calcul de $\log[S](z)$. Afin d'illustrer l'enjeu de cette notion, étudions l'exemple d'un signal tous-pôles à temps continu $s(t)$ dont la transformée de Laplace $S(p)$ est composée de N_C pôles causaux (i.e à partie réelle négative) notés $\{c_i\}_{i \in [1, N_C]}$ et de N_A pôles anticausaux (i.e à partie réelle positive) notés $\{a_i\}_{i \in [1, N_A]}$. La bande de convergence de cette transformée de

Laplace, ensemble simplement connexe de \mathbb{C} sur lequel $S(p)$ est holomorphe et non nul, s'écrit

$$\begin{aligned} \mathcal{B} &= \{s \in \mathbb{C} / \zeta_C < \Re(s) < \zeta_A\} \quad \text{avec} & (7.22) \\ \zeta_C &= \max_{i \in [1, N_A]} (\Re(c_i)) , \\ \zeta_A &= \min_{i \in [1, N_C]} (\Re(a_i)) . \end{aligned}$$

La version discrétisée du signal a une transformée en Z construite à partir de la transformée de Laplace en prenant $z = e^{pT_e}$ dans la bande de convergence (avec p la variable indépendante complexe du domaine de Laplace et T_e la période d'échantillonnage). Ainsi les pôles causaux $\{c_i\}_{i \in [1, N_C]}$ deviennent $\{\gamma_i\}_{i \in [1, N_C]}$, pôles à l'intérieur du cercle unité, les pôles anticausaux $\{a_i\}_{i \in [1, N_A]}$ deviennent $\{\alpha_i\}_{i \in [1, N_A]}$, pôles à l'extérieur du cercle unité et la bande de convergence \mathcal{B} devient la couronne de convergence

$$\begin{aligned} \mathcal{C} &= \{z \in \mathbb{C} \setminus \mathbb{R}^- / \rho_C < |z| < \rho_A\} \quad \text{avec} & (7.23) \\ \rho_C &= \max_{i \in [1, N_A]} (|\gamma_i|) , \\ \rho_A &= \min_{i \in [1, N_C]} (|\alpha_i|) . \end{aligned}$$

La figure 7.8-Ⓐ présente la partie imaginaire de $\log[S](p)$, telle que définie par (7.19), pour un signal tous-pôles à deux pôles causaux et deux pôles anticausaux. Chaque pôle fait apparaître une coupure en forme de demi-droite issue du pôle et portée par la droite reliant l'origine au pôle. Pour comparaison, la figure 7.8-Ⓑ présente la partie imaginaire de $\log(S(p))$, détermination principale du logarithme (7.16) de $S(p)$ en chaque point. Les coupures sont ici courbées et traversent la bande de convergence \mathcal{B} ainsi que l'axe de Fourier $i\mathbb{R}$. L'holomorphie sur \mathcal{B} , et plus particulièrement sur $i\mathbb{R}$, n'est plus respectée, ce qui pose problème pour l'inversion de la transformée de Laplace sur \mathcal{B} , ou de la transformée de Fourier sur $i\mathbb{R}$, nécessaire à l'obtention du cepstre.

Les figures 7.8-Ⓒ et 7.8-Ⓓ présentent la même comparaison pour la version discrète (transformée en Z). Dans la figure 7.8-Ⓒ, en raison de la transformation $p \mapsto e^{pT_e}$, les coupures ne sont plus des demi-droites mais des portions de spirales reliant l'origine aux pôles. Les coupures de $\log(S(z))$ dans 7.8-Ⓓ traversent la couronne de convergence \mathcal{C} et le cercle unité \mathbb{S} , posant les mêmes problèmes d'inversion de transformées sur \mathcal{C} et \mathbb{S} , rendant impossible le calcul du cepstre.

Enfin, les figure 7.8-Ⓔ et 7.8-Ⓕ présentent les parties imaginaires de $\log[S](p) - \log(S(p))$ et de $\log[S](z) - \log(S(z))$. Cette figure fait apparaître les valeurs du $2k\pi$ de (7.19) où k dépend de p (ou z dans le cas discret). Ce terme peut être interprété comme le terme qu'il manque à $\log(S(z))$ (7.16) pour obtenir une détermination holomorphe du logarithme de la transformée dans la bande de convergence (on notera que les valeurs de k sont bien entières). Sur ces figures, à chaque saut de k dans le plan complexe correspond un changement de branche dans la surface de Riemann.

Remarque 5. Dans le cas d'une transformée ne contenant qu'un seul pôle, la partie imaginaire de $\log[S](p)$ est contenue dans l'intervalle $]-\pi, \pi]$. $\log[S](p)$ et $\log(S(p))$ sont alors confondus. Cette propriété permet de calculer $\log[S](p)$ par somme de la détermination principale du logarithme de chaque pôle lorsque les pôles sont connus.

Remarque 6. L'extension au cas des signaux pôles/zéros est immédiate, à condition de considérer la bande ou couronne de convergence ne contenant pas de zéros (points en lesquels le logarithme n'est pas défini).

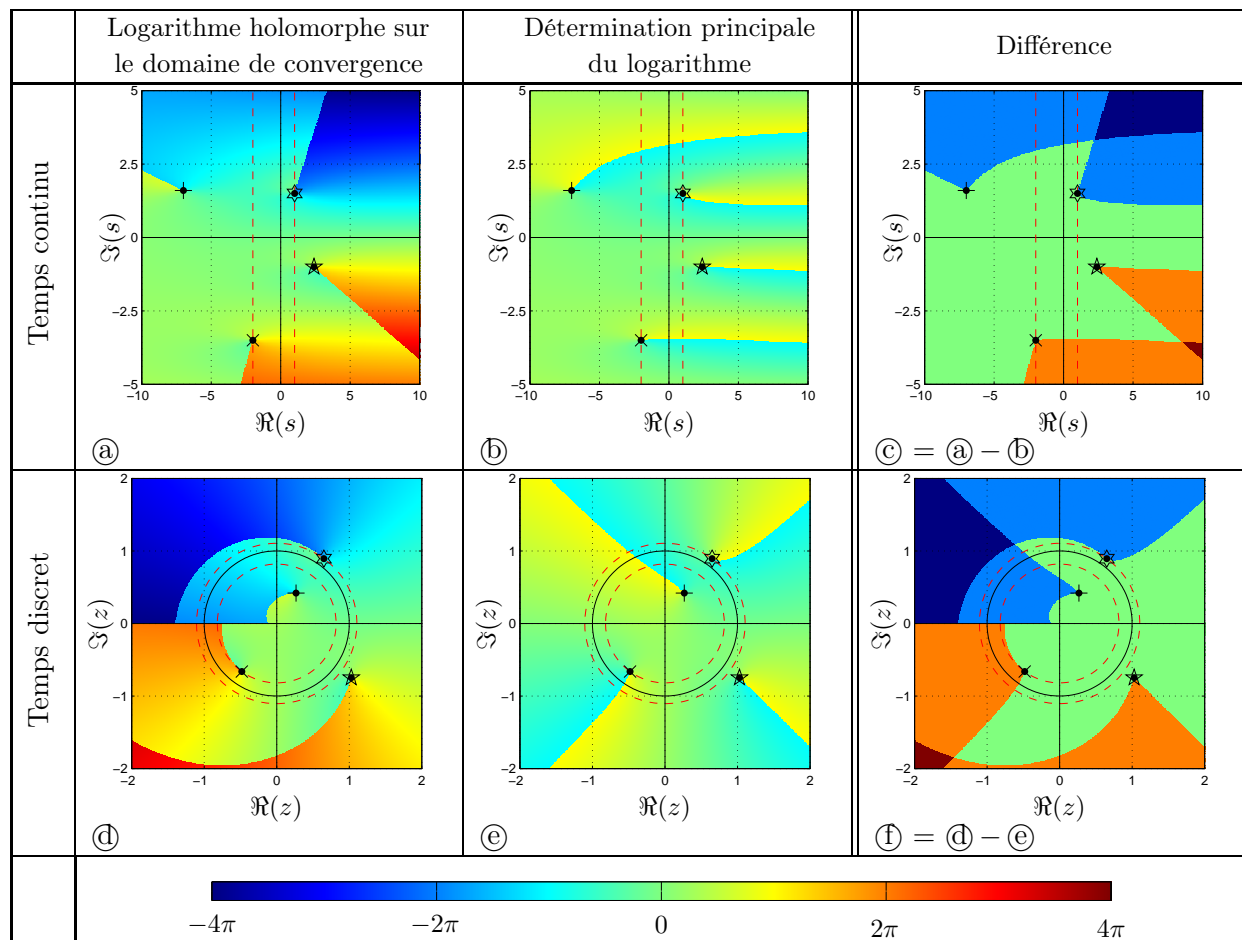


FIGURE 7.8 – Partie imaginaire du logarithme de la transformée d'un signal tous-pôles, comparaison de la détermination principale du logarithme $\log(S(p))$ ((b) et (e)) et de la détermination holomorphe du logarithme $\log[S](p)$ ((a) et (d)). Les pointillés rouges indiquent les frontières des domaines de convergence des transformées. Les symboles utilisés pour représenter les pôles se correspondent en temps continu et en temps discret.

Méthodes de calcul

La définition (7.21) peut être ré-écrite de façon équivalente à l'aide de la transformation de Fourier (à condition que le domaine d'holomorphic de $S(z)$ contienne \mathbb{S})

$$\hat{s}(n) = \text{TFTD}^{-1} \left[\log \left[\text{TFTD}[s] \right] (\nu) \right] (n), \quad (7.24)$$

où TFTD correspond à la transformée de Fourier à temps discret et TFTD^{-1} à sa réciproque.

Ainsi apparaissent deux méthodes de calcul du cepstre complexe. La première, fondée sur la définition (7.24) utilise le calcul de la transformée de Fourier discrète du signal. La deuxième, fondée sur la définition (7.21) et la remarque 5, utilise la factorisation de la transformée en Z (ou ZZT [Boz05]). Ces deux méthodes sont étudiées et comparées en détail dans [RS11, Dru11], aussi seules leurs propriétés principales sont présentées ici. Nous verrons ensuite une troisième méthode de calcul qui exploite les propriétés du cepstre différentiel [PF81].

Le calcul du cepstre complexe par transformée de Fourier permet un calcul rapide en remplaçant la TFTD par un algorithme de calcul de transformée de Fourier discrète (FFT [CT65]) dans (7.24). En revanche, cette méthode est source d'artefacts liés au repliement fréquentiel causé par le fenêtrage intrinsèque à l'algorithme de FFT [RS11, p.443 et suivantes]. De plus, cette méthode impose de calculer directement le logarithme de la transformée de Fourier échantillonnée. Or le calcul numérique du logarithme complexe en tant qu'opérateur préservant l'holomorphicité (7.19) n'est pas possible. Il est alors nécessaire de calculer la détermination principale du logarithme complexe (7.16) de chaque point fréquentiel puis d'effectuer une recherche numérique des sauts de $2k\pi$ dans la partie imaginaire du $\log(S(k))$ obtenu, c'est-à-dire des coupures artificielles traversant \mathbb{S} ayant été introduites par le non respect de l'holomorphicité du logarithme. Cette opération, plus connue sous son nom anglais de "unwrapping", est assurée par des algorithmes qui ne peuvent pas être parfaits en raison du caractère discret de la TFD.

Le calcul du cepstre complexe par factorisation de la transformée en Z [SD77] repose sur le caractère fini du support des signaux discrets qui sont étudiés en traitement numérique. En effet, le traitement informatique du signal impose de travailler avec des signaux à support fini, ce qui implique le plus souvent un fenêtrage du signal d'étude. Comme il l'a déjà été dit dans la section 7.2, les signaux à support fini ont des transformées en Z ne contenant que des zéros. Ainsi la transformées en Z d'un signal de L échantillons, que nous numérotions de 1 à L , peut se mettre sous la forme

$$S(z) = G \cdot z^{-K} \cdot \prod_{n=1}^N \left(1 - \frac{a_n}{z}\right) \cdot \prod_{m=1}^M \left(1 - \frac{z}{b_m}\right), \quad (7.25)$$

où $\{a_n\}_{n \in [1, N]}$ est l'ensemble des zéros de module inférieur à 1, $\{b_m\}_{m \in [1, M]}$ est l'ensemble des zéros de module supérieur à 1 et K représente la position de l'origine temporelle (au sens d'un point de jonction causal/anti-causal, soit le GCI dans le cas d'un signal vocal). Par propriété du logarithme complexe (7.19),

$$\log[S](z) = \log(G) + \log z^{-K} + \sum_{n=1}^N \log \left(1 - \frac{a_n}{z}\right) + \sum_{m=1}^M \log \left(1 - \frac{z}{b_m}\right).$$

Lors d'un calcul numérique par factorisation, la référence temporelle n'est en général pas située à l'origine temporelle qui n'est pas connue a priori. Elle est le plus souvent choisie au dernier échantillon de la séquence étudiée pour des raisons de simplicité algorithmique (cas de la fonction `roots` de Matlab®). Ainsi $-K$ devient K' dans (7.25) et il est possible de retrouver l'origine temporelle :

$$K = L - N - K'. \quad (7.26)$$

Nous pouvons finalement écrire le cepstre complexe du signal à partir du tableau 7.2, en notant $\hat{\delta}_d(K)$ le terme dû au retard pur (nous étudierons ce terme plus loin),

$$\hat{s}(n) = \log(G)\delta(n) + \hat{\delta}_d(K) + \sum_{i=1}^N \frac{a_i^n}{n} H(n-1) - \sum_{j=1}^M \frac{b_j^n}{n} H(-n-1). \quad (7.27)$$

Le principal désavantage de cette méthode est qu'elle présente un coût de calcul élevé en raison de la nécessité de calculer les racines d'un polynôme d'ordre très élevé. Cependant, elle présente l'immense avantage de préserver l'holomorphicité de la TZ du cepstre sur U . En effet, lors du calcul

numérique d'un cepstre par factorisation de la transformée en Z , il suffit de calculer le cepstre à l'aide de l'expression (7.27) après factorisation, sans qu'il ait été nécessaire d'effectuer un calcul de logarithme complexe. De plus cette méthode est assez robuste au fenêtrage.

Cette méthode est à relier aux remarques 5 et 6. La factorisation de la ZZT, dans le cas d'un signal à support fini (donc tous-zéros), nous donne un accès direct et complet aux zéros (équation (7.25)). $\log[S](z)$ est donc égale à la somme des déterminations principales du logarithme de la TZ de chacun des zéros (transposition de la remarque 5 dans le cas tous-zéros).

Cepstre complexe et cepstre différentiel

Une troisième méthode pour le calcul du cepstre complexe, peu utilisée dans le traitement numérique de la parole, a été proposée par Reddy et Rao en 1987 [RR87]. Cette méthode consiste à passer par le cepstre différentiel (défini en multipliant le cepstre par la variable indépendante qu'éfrentielle [PF81]) qui, nous le verrons, peut être calculé avec la transformée de Fourier tout en préservant l'holomorphie sur le cercle unité.

Nous avons vu dans la remarque 3 que, lorsque le logarithme de la TZ $S(z)$ d'un signal $s(n)$ est bien posé (i.e. holomorphe sur un domaine U contenant \mathbb{S} sur lequel $S(z)$ est holomorphe et non nul), la dérivée de $\log[S](z)$ est égale à $\frac{S'(z)}{S(z)}$ et est holomorphe sur U , et ce quelque soit le choix (de la branche) du logarithme. Or, nous savons que

$$\text{TZ}[\check{s}](z) = -zS'(z), \text{ avec } \check{s} = n \cdot s(n).$$

Ainsi apparaît une nouvelle formulation du logarithme complexe préservant l'holomorphie de $\log[S](z)$ tout en évitant le calcul direct du logarithme qui pose problème,

$$\hat{s}(n) = \frac{\text{TZ}^{-1} \left[\frac{\text{TZ}[\check{s}](z)}{\text{TZ}[s](z)} \right] (n)}{n} \quad \forall n \in \mathbb{Z}^*. \quad (7.28)$$

Cette formulation du cepstre complexe est aisément implémentable en exploitant les algorithmes de transformée de Fourier rapide sous la forme

$$\hat{s}(n) = \frac{\text{IFFT} \left[\frac{\text{FFT}[ns(n)]}{\text{FFT}[s(n)]} \right] (n)}{n} \quad \forall n \in \mathbb{Z}^*. \quad (7.29)$$

La formule (7.29) permet un calcul numérique rapide du cepstre complexe tout en préservant l'holomorphie du logarithme, propriété fondamentale dans la définition du logarithme complexe et donc dans l'unicité du cepstre complexe. Le point négatif de cette formulation réside dans la perte de l'information du gain. En effet, le gain apparaît dans (7.27) uniquement dans l'échantillon du cepstre de qu'éfrence nulle, autrement dit dans la composante continue de $\log[S]$. Comme cette dernière méthode de calcul passe par la dérivée de $\log[S]$, il est impossible de retrouver l'information de gain. Autrement dit, nous calculons $\log[S]$ à une constante près (ce qui revient à la remarque 3, toutes les branches de $\log[S]$ ont la même dérivée) donc nous ne pouvons pas reconstruire l'échantillon du cepstre de qu'éfrence nulle. L'estimation du gain est cependant rarement problématique en traitement du signal. Aussi, en raison de la combinaison du faible coût de calcul et de la préservation de l'holomorphie du logarithme, cette méthode sera utilisée dans la suite des travaux présentés ici.

7.3.2 Cepstre d'un signal tous-pôles causal/anti-causal

Cepstre complexe

Nous considérons dans cette section le modèle suivant. Le signal discret $s(n)$ est généré par filtrage tous-pôles causal (index +) / anticausal (index -) d'un ensemble d'impulsions de Dirac. La transformée en Z d'un tel signal s'écrit

$$S(z) = \mathcal{G} \cdot \frac{1}{\prod_{k=1}^{K^+} \left(1 - \frac{a_k^+}{z}\right)} \cdot \frac{1}{\prod_{k=1}^{K^-} \left(1 - \frac{z}{a_k^-}\right)} \cdot E(z), \quad (7.30)$$

avec $|a_k^+| < 1 \forall k \in [1, K^+]$ et $|a_k^-| > 1 \forall k \in [1, K^-]$, et où $E(z)$ est la transformée en Z d'un signal de type

$$e(n) = \sum_{k=1}^M e_k \delta(n - d_k), \quad (7.31)$$

avec $M \in \mathbb{N}^*$ et $d_k \in \mathbb{Z} \forall k$. Nous considérons dans un premier temps une impulsion isolée ($M = 1$, $e_1 = 1$ et on note $d_1 = d$). Autrement dit, nous nous intéressons à la réponse impulsionnelle du filtre composé par le filtre glottique (anticausal) et le filtre vocal (causal) en série à une impulsion de Dirac à l'instant d . La transformée en Z de cette réponse s'écrit donc

$$S(z) = \mathcal{G} \cdot z^{-d} \cdot \frac{1}{\prod_{k=1}^{K^+} \left(1 - \frac{a_k^+}{z}\right)} \cdot \frac{1}{\prod_{k=1}^{K^-} \left(1 - \frac{z}{a_k^-}\right)}. \quad (7.32)$$

Nous notons θ le vecteur de paramètres de ce modèle,

$$\theta = \{\mathcal{G}, d, \{a_k^+\}_{k \in [1, K^+]}, \{a_k^-\}_{k \in [1, K^-]}, K^+, K^-\}. \quad (7.33)$$

Le domaine de convergence de cette transformée en Z est la couronne comprise entre le plus petit module des pôles anti-causaux et le plus grand module des pôles causaux [OS75]. De plus, cette fonction étant une fonction rationnelle, nous savons qu'elle est holomorphe sur l'ensemble du plan complexe privé des singularités [Val66], qui contient cette couronne. Afin de pouvoir étudier le logarithme complexe de cette fonction, nous nous intéressons à \mathcal{D}_S , cette couronne privée de la demi-droite des réels négatifs, ouvert simplement connexe de \mathbb{C} sur lequel $S(z)$ est holomorphe.

$$\mathcal{D}_S = \{z \in \mathbb{C} \setminus \mathbb{R}^- / \zeta^+ < |z| < \zeta^-\}, \quad (7.34)$$

$$\text{avec } \zeta^+ \triangleq \max_{z \in \mathcal{Z}^+} |z| \quad (7.35)$$

$$\text{et } \zeta^- \triangleq \min_{z \in \mathcal{Z}^-} |z|, \quad (7.36)$$

où \mathcal{Z}^+ représente l'ensemble des pôles causaux, i.e. $\mathcal{Z}^+ = \{a_k^+\}_{k \in [1, K^+]}$, et \mathcal{Z}^- représente l'ensemble des pôles anti-causaux, i.e. $\mathcal{Z}^- = \{a_k^-\}_{k \in [1, K^-]}$. Remarquons que \mathcal{D}_S contient bien le cercle unité \mathbb{S} , de sorte que le calcul de sa transformée de Fourier soit bien posée.

Le cepstre complexe de $s(n)$ s'écrit dans le domaine $z \in \mathcal{D}_S$

$$\log[S](z) = \log(\mathcal{G}) + \log(z^{-d}) - \sum_{k=1}^{K^+} \log\left(1 - \frac{a_k^+}{z}\right) - \sum_{k=1}^{K^-} \log\left(1 - \frac{z}{a_k^-}\right).$$

Comme le cercle unité \mathbb{S} est inclus dans \mathcal{D}_S ⁴, nous pouvons calculer la transformée en \mathbb{Z} inverse de cette expression. D'après le tableau 7.2, nous obtenons le cepstre (nous notons encore ici $\hat{\delta}_d(n)$ la contribution du terme de retard pur, nous étudierons sa valeur au paragraphe suivant)

$$\hat{s}(n) = \log(\mathcal{G})\delta(n) + \hat{\delta}_d(n) + \sum_{k=1}^{K^+} \frac{(a_k^+)^n}{n} H(n-1) - \sum_{k=1}^{K^-} \frac{(a_k^-)^n}{n} H(-n-1). \quad (7.37)$$

Cette expression, rapportée à l'expression (7.32), illustre la propriété fondamentale (7.12). Ainsi, chaque facteur dans (7.32) (pôles causal ou anticausal, gain et retard pur) devient un terme de la somme dans (7.37). De plus, la séparation causale/anti-causale est intrinsèque au calcul du cepstre complexe, comme il l'a déjà été montré dans les méthodes de séparation source-filtre exploitant le cepstre complexe [BD03, Dru11]. En effet, l'expression (7.37) montre que l'influence des pôles causaux est restreinte aux quéfrenes positives, tandis que l'influence des pôles anti-causaux est restreinte aux quéfrenes négatives. Enfin, le gain est intégralement représenté par l'échantillon du cepstre de quéfrence nulle.

Cepstres différentiels causaux et anti-causaux

L'algorithme que je présente ici se fonde sur l'analyse du cepstre différentiel. En multipliant $\hat{s}(n)$ par n (la variable indépendante quéfrentielle) dans (7.37), on obtient le cepstre différentiel

$$\check{s}(n) \triangleq n\hat{s}(n) = \check{\delta}_d(n) + \sum_{k=1}^{K^+} (a_k^+)^n H(n-1) - \sum_{k=1}^{K^-} (a_k^-)^n H(-n-1). \quad (7.38)$$

Nous avons vu plus haut que le cepstre différentiel peut être calculé directement par

$$\check{s}(n) = \text{TZ}^{-1} \left[\frac{\text{TZ}[\check{s}](z)}{\text{TZ}[s](z)} \right] (n) \quad \forall n \in \mathbb{Z}. \quad (7.39)$$

Cette définition nous permet de calculer le cepstre différentiel du retard pur $\check{\delta}_d(n)$.

$$\begin{aligned} \text{TZ}[\delta(n-d)](z) &= z^{-d}, \\ \text{TZ}[n \cdot \delta(n-d)](z) &= d \cdot z^{-d}, \end{aligned}$$

d'où, par (7.39),

$$\check{\delta}_d(n) = d \cdot \delta(n). \quad (7.40)$$

4. En toute rigueur, seul $\mathbb{S} \setminus \{-1\}$ est inclus dans \mathcal{D}_S , mais ceci ne gêne en rien le calcul de la TZ inverse.

Nous obtenons finalement l'expression complète du cepstre différentiel du signal (7.32),

$$\check{s}(n) = d\delta(n) + \sum_{k=1}^{K^+} (a_k^+)^n H(n-1) - \sum_{k=1}^{K^-} (a_k^-)^n H(-n-1). \quad (7.41)$$

Nous définissons le cepstre différentiel causal $\check{s}^+(n)$ et le cepstre différentiel anti-causal $\check{s}^-(n)$

$$\check{s}^+(n) \triangleq \check{s}(n) \quad \forall n \in \mathbb{N}^* \text{ et} \quad (7.42)$$

$$\check{s}^-(n) \triangleq \check{s}(-n) \quad \forall n \in \mathbb{N}^*. \quad (7.43)$$

Dans le cas de notre signal, nous obtenons

$$\check{s}^+(n) = \sum_{k=1}^{K^+} (a_k^+)^n H(n-1), \text{ et} \quad (7.44)$$

$$\check{s}^-(n) = \sum_{k=1}^{K^-} \frac{1}{(a_k^-)^n} H(n-1). \quad (7.45)$$

On remarque que ces deux cepstres différentiels font apparaître (dans le domaine quéfrentiel) la somme de séries géométriques comme c'est le cas de la réponse impulsionnelle du filtre composé par chacun des pôles causaux et l'inverse de chacun des pôles anti-causaux du modèle (7.32). C'est cette représentation (bien connue dans le cas des représentations temporelles) que nous allons exploiter pour l'estimation dans l'algorithme présenté ci-après.

7.3.3 Algorithme d'estimation des paramètres du modèle causal-anticausal

Présentation de l'algorithme

L'estimation des paramètres se fait par recherche des pôles causaux et anticausaux permettant au mieux de décrire les cepstres différentiels causal et anticausal du signal étudié. L'algorithme mis en place est fondé sur une maximisation de vraisemblance afin de prendre en compte la présence de bruit dans le cepstre différentiel estimé du signal étudié.

Comme il l'a déjà été mentionné, les signaux disponibles pour un traitement numérique automatique sont des signaux à support fini. Ainsi nous considérerons un extrait temporel de la réponse impulsionnelle (7.32) dont nous cherchons à retrouver les paramètres. Nous verrons dans la section suivante quelles limites induisent le fenêtrage de la réponse.

Considérons donc un signal $y(n)$ de longueur fini L (échantillons 1 à L) extrait d'une réponse impulsionnelle ayant une transformée en Z de la forme (7.32). Nous postulons que l'instant d'excitation (le GCI) se trouve dans l'horizon temporel, ce qui est nécessaire pour pouvoir observer à la fois les parties causale et anti-causale de la réponse impulsionnelle. L'algorithme d'estimation des paramètres se déroule en cinq étapes :

1. calcul du cepstre différentiel $\check{y}(n)$,
2. estimation du GCI (paramètre d),
3. construction des signaux $\check{y}^+(n)$ et $\check{y}^-(n)$,

4. estimation des pôles causaux et anti-causaux du modèle,
5. estimation du gain.

L'étape 1 est réalisée par la méthode (7.39) en remplaçant la TZ et son inverse par la FFT et son inverse. L'étape 2 est réalisée par évaluation de l'échantillon de quéfrence nulle dans le cepstre différentiel obtenu, comme suggéré par l'équation (7.41). L'étape 3 est réalisée à l'aide des équations (7.42-7.43) à partir du cepstre différentiel estimé à l'étape 1. L'étape 4 est réalisée par maximisation de vraisemblance sur les signaux $\check{y}^+(n)$ et $\check{y}^-(n)$. Cette étape est détaillée ci-après. L'étape 5 est réalisée par estimation de la puissance du signal.

Maximisation de la vraisemblance des cepstres différentiels causaux et anticausaux

Afin de prendre en compte la présence de bruit dans les cepstres différentiels du signal d'étude, nous préférons maximiser la vraisemblance des cepstres différentiels observés plutôt que minimiser la distance entre les cepstres différentiels observés et ceux du modèle. Nous calculons pour ceci la fonction de vraisemblance des cepstres différentiels estimés connaissant les paramètres du modèle.

Soient $\check{y}^+(n)$ et $\check{y}^-(n)$ les cepstres différentiels pour $n \in [1, N]$ calculés à l'aide des équations (7.39-7.42-7.43) à partir des L échantillons du signal d'entrée $y(n)$. Nous considérons que les signaux $\check{y}^\pm(n)$ sont égaux au modèle plus un bruit blanc gaussien centré $\varepsilon^\pm(n)$ de variance $(\sigma^\pm)^2$,

$$\check{y}^\pm(n) = \pm \sum_{k=1}^{K^\pm} (a_k^\pm)^{\pm n} + \varepsilon^\pm(n).$$

La fonction de vraisemblance des N échantillons de \check{y}^\pm connaissant le vecteur de paramètres du modèle θ et la variance du bruit σ^\pm s'écrit alors

$$\begin{aligned} f_{[1,N]}^\pm(\theta, \sigma^\pm) &\triangleq p(\check{y}^\pm([1, N]) / \theta, \sigma^\pm) \\ &= \prod_{n=1}^N \frac{1}{\sqrt{2\pi(\sigma^\pm)^2}} \exp \left[-\frac{(\check{y}^\pm(n) \mp \sum_{k=1}^{K^\pm} (a_k^\pm)^{\pm n})^2}{2(\sigma^\pm)^2} \right] \\ &= (2\pi)^{-\frac{N}{2}} \cdot (\sigma^\pm)^{-N} \cdot \exp \left[-\sum_{n=1}^N \frac{(\check{y}^\pm(n) \mp \sum_{k=1}^{K^\pm} (a_k^\pm)^{\pm n})^2}{2(\sigma^\pm)^2} \right]. \end{aligned} \quad (7.46)$$

La log-vraisemblance des cepstres différentiels s'écrit donc

$$\begin{aligned} L_{[1,N]}^\pm(\theta, \sigma^\pm) &\triangleq -\log \left(f_{[1,N]}^\pm(\theta, \sigma^\pm) \right) \\ &= \frac{N}{2} \ln(2\pi) + N \ln(\sigma) + \sum_{n=1}^N \frac{1}{2} \left(\frac{\check{y}^\pm(n) \mp \sum_{k=1}^{K^\pm} (a_k^\pm)^{\pm n}}{\sigma^\pm} \right)^2. \end{aligned} \quad (7.47)$$

Maximiser la vraisemblance (7.46) est équivalent à minimiser la log-vraisemblance (7.47). Cepen-

dant, il apparaît que le système d'équations

$$\begin{cases} \frac{\partial L_{[1,N]}^{\pm}}{\partial \sigma}(\theta, \sigma^{\pm}) = 0, \\ \frac{\partial L_{[1,N]}^{\pm}}{\partial a_k^{\pm}}(\theta, \sigma^{\pm}) = 0 \quad \forall k \in [1, K^{\pm}], \end{cases}$$

est non linéaire et fait apparaître des polynômes de degré N en les pôles a_k^{\pm} . Par conséquent, nous opterons pour une minimisation numérique de (7.47) à l'aide de l'algorithme du simplexe [LRWW98] pour estimer les pôles et la variance du bruit.

Remarques générales sur le deuxième algorithme. Comme pour le premier algorithme, nous verrons dans la section 7.3.4 que cet algorithme estime bien les paramètres pour des signaux de synthèse non bruités. De plus, on vérifie qu'il est assez robuste au fenêtrage. La robustesse au bruit dans les signaux d'analyse est meilleure qu'avec le premier algorithme mais reste limitée. Ceci est détaillé dans la section 7.4 avec des pistes pour améliorer la robustesse.

Ingrédients à ajouter pour traiter des signaux vocaux mesurés. L'algorithme présenté ici effectue l'estimation sur une réponse impulsionnelle et non sur un signal vocal périodique. De plus, seuls des GCIs synchronisés sur les échantillons du signal numérique peuvent être détectés. Une première application de l'algorithme en l'état à des signaux de voix sera néanmoins étudiée. Des pistes pour aborder ces deux problèmes seront présentées en section 7.4.

7.3.4 Performances de l'algorithme d'estimation

7.3.4.1 Signaux de synthèse

L'algorithme décrit ci-dessus a été testé sur un signal de type (7.32) avec les paramètres suivants.

- 4 paires de pôles complexes conjugués causaux (filtre vocal) de fréquences [900, 1200, 3000, 4000] Hz et de facteurs de qualités [5, 15, 40, 15],
- une paire de pôles complexes conjugués anticausaux (formants du filtre glottique) de fréquence 100 Hz et de facteur de qualité 0.7,
- $\mathcal{G} = 2$,
- $d = 251$,
- fréquence d'échantillonnage de 10 kHz.

Nous observons ce signal sur les échantillons [1, 550].

La figure 7.9 présente le signal observé $y(n)$ avec la position du GCI ainsi que le cepstre différentiel estimé $\hat{y}(n)$ et ses parties causale $\tilde{y}^+(n)$ et anticausale $\tilde{y}^-(n)$ (étapes 1 et 3 de l'algorithme). L'étape 2 de l'algorithme est immédiate, $\hat{y}(0)$ est précisément égal à d . Les résultats des étapes 4 et 5 (maximisation de vraisemblance sur les signaux $\tilde{y}^+(n)$ et $\tilde{y}^-(n)$ de la figure 7.9) sont présentés sur la figure 7.10. Nous pouvons observer sur cette figure la parfaite estimation du GCI, des pôles causaux (erreur moyenne de $3,4 \cdot 10^{-6} \%$), des pôles anticausaux (erreur moyenne de $2,1 \cdot 10^{-4} \%$) et du gain. L'erreur quadratique moyenne de reconstruction est de seulement $3,1 \cdot 10^{-3} \%$.

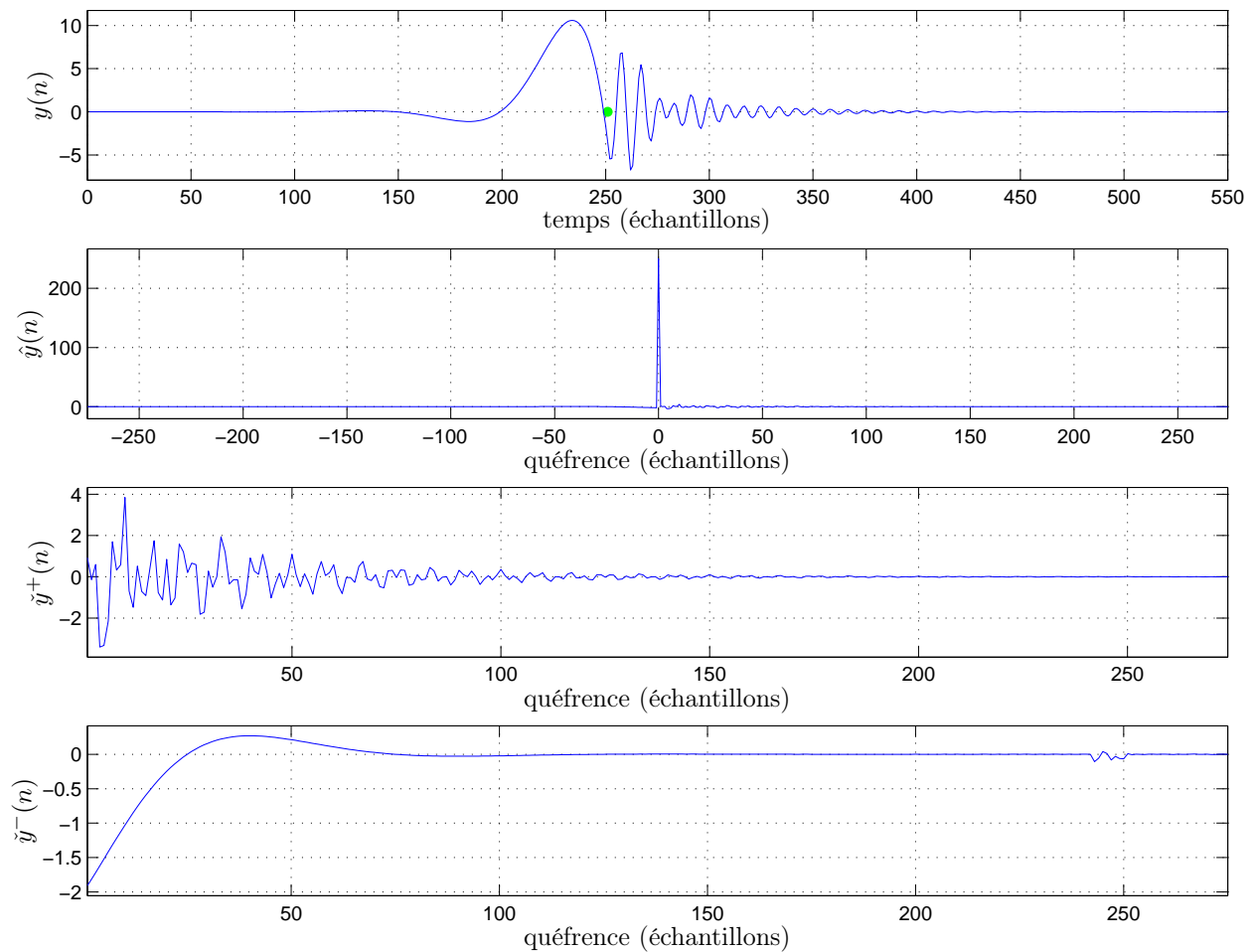


FIGURE 7.9 – Premier exemple : signal temporel et cepstre différentiel.

7.3.4.2 Robustesse de l'algorithme

L'algorithme proposé pour l'estimation des paramètres sur la réponse impulsionnelle d'un filtre est relativement robuste au fenêtrage du signal d'analyse. La figure 7.11 illustre les performances de l'algorithme lorsque le signal d'analyse est sévèrement tronqué. La troncature fait apparaître du bruit dans les cepstres différentiels causaux et anticausaux. Néanmoins, la maximisation de vraisemblance permet de retrouver la position des pôles avec une précision satisfaisante. En effet, l'erreur de position sur la paire de pôles anticausaux est de l'ordre de 1,3 % malgré une erreur de resynthèse (quadratique moyenne) de 15 %.

Un cas avec fenêtrage encore plus sévère est présenté en figure 7.12. Ce cas illustre l'intérêt de la maximisation de vraisemblance, notamment pour le cas du cepstre différentiel anticausal. Nous pouvons cependant constater que la paire de pôles causaux de fréquence la plus élevée n'est pas estimée. Une interprétation possible est que l'information portée par cette paire de pôle est noyée dans le bruit dans le signal $\hat{y}^+(n)$, l'algorithme n'arrive donc plus à la reconstruire. Par ailleurs, le bruit a contaminé l'échantillon de quéfrencé nulle dans le cepstre car la position du GCI estimé est erronée (décalage de 2 échantillons).

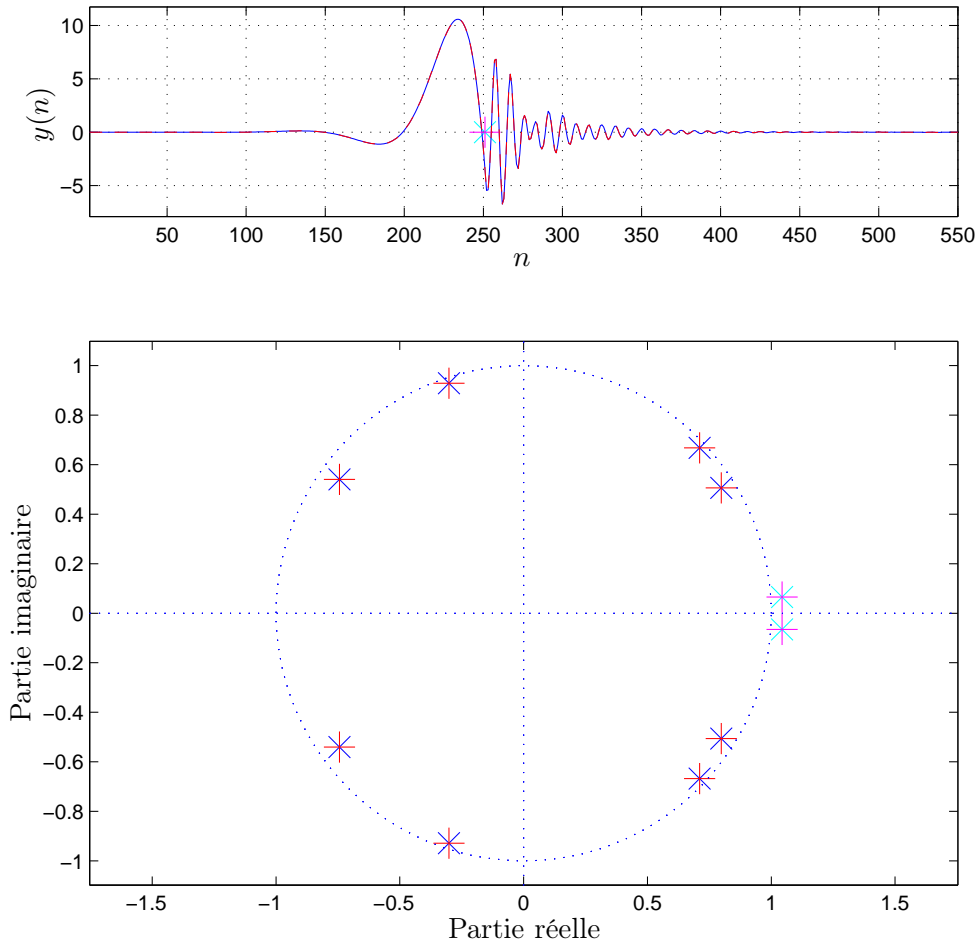


FIGURE 7.10 – Premier exemple : estimation des paramètres et reconstruction. En haut : signal $y(n)$ (-) et signal reconstruit (--) ainsi que le GCI (\times) et le GCI estimé ($+$), pôles du modèle de synthèse (\times et \times) and pôles estimés ($+$ et $+$).

La présence de bruit fait plus rapidement baisser les performances de l'algorithme. Deux exemples sont présentés en figures 7.13 et 7.14, respectivement pour des bruits de puissance 10^{-5} et 10^{-4} . Afin de rendre l'algorithme un peu plus robuste, la maximisation de vraisemblance a été réalisée sur les 100 premiers échantillons des signaux $\tilde{y}^{\pm}(n)$ plutôt que sur la totalité de l'horizon quéfrentiel disponible. Ceci permet de ne pas prendre en compte les hautes quéfrences où l'information est réduite et où le rapport signal sur bruit élevé perturbe l'algorithme. Nous pouvons remarquer que la présence de bruit a un impact très important sur l'estimation des cepstres différentiels. En effet, les signaux $\tilde{y}^{\pm}(n)$ sont très éloignés des signaux $\tilde{y}^{\pm}(n)$ théoriques (représentés en vert). Cependant, la procédure de maximisation de vraisemblance permet d'obtenir des résultats partiels prometteurs. Encore une fois, ce sont les pôles associés à des fréquences élevées et des facteurs de qualité faibles (donc moins porteurs d'information) qui sont le moins bien estimés.

Pour des bruits plus puissance plus élevée, les résultats sont rapidement dégradés.

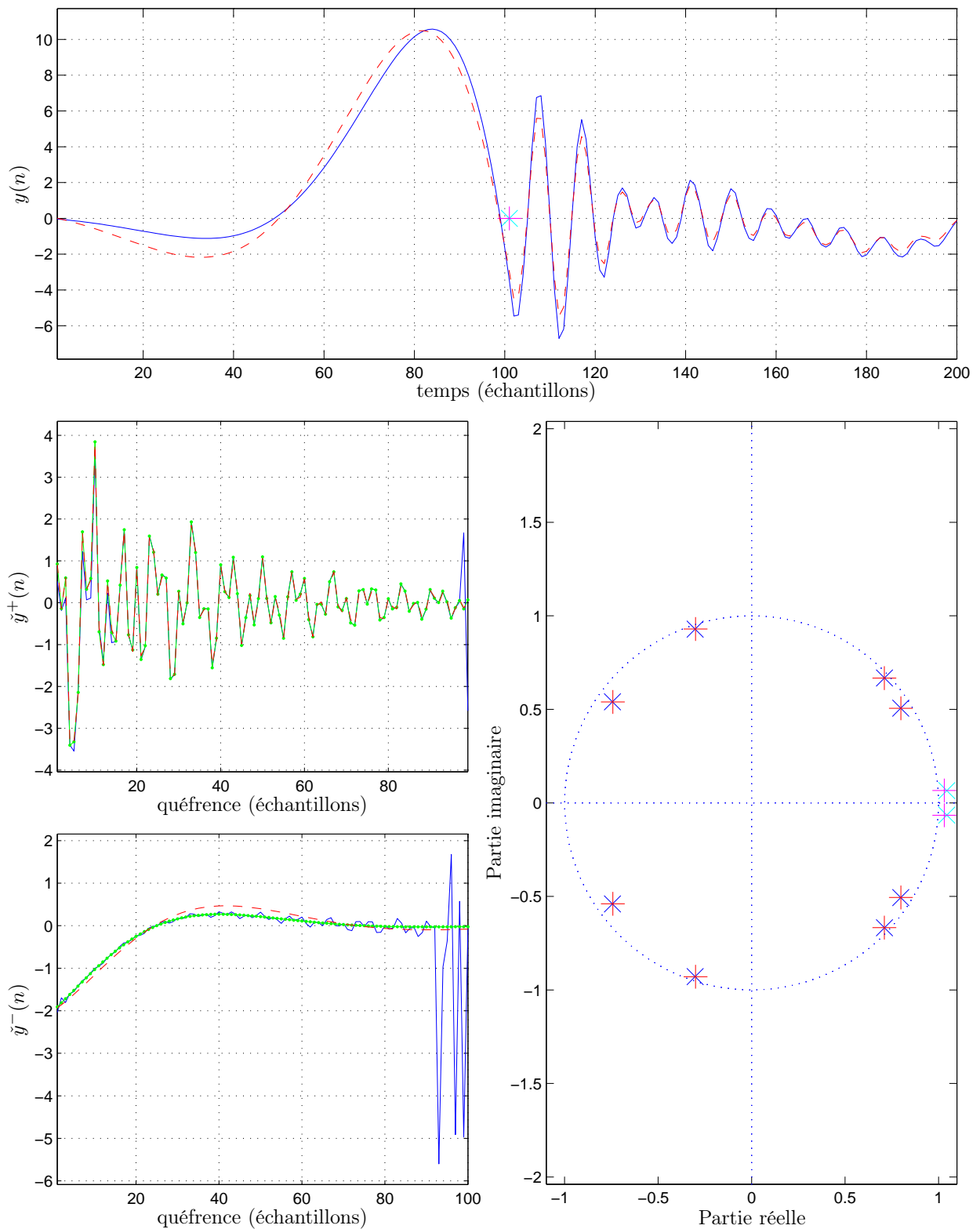


FIGURE 7.11 – Effet du fenêtrage 1. En haut : Signal $y(n)$ de synthèse (-) et signal reconstruit (--) ainsi que le GCI (x) et le GCI estimé (+). A gauche : signaux $\tilde{y}^\pm(n)$ théoriques (-●-), estimés (-) et reconstruits (--). A droite : pôles du modèle de synthèse (x et x) and pôles estimés (+ et +).

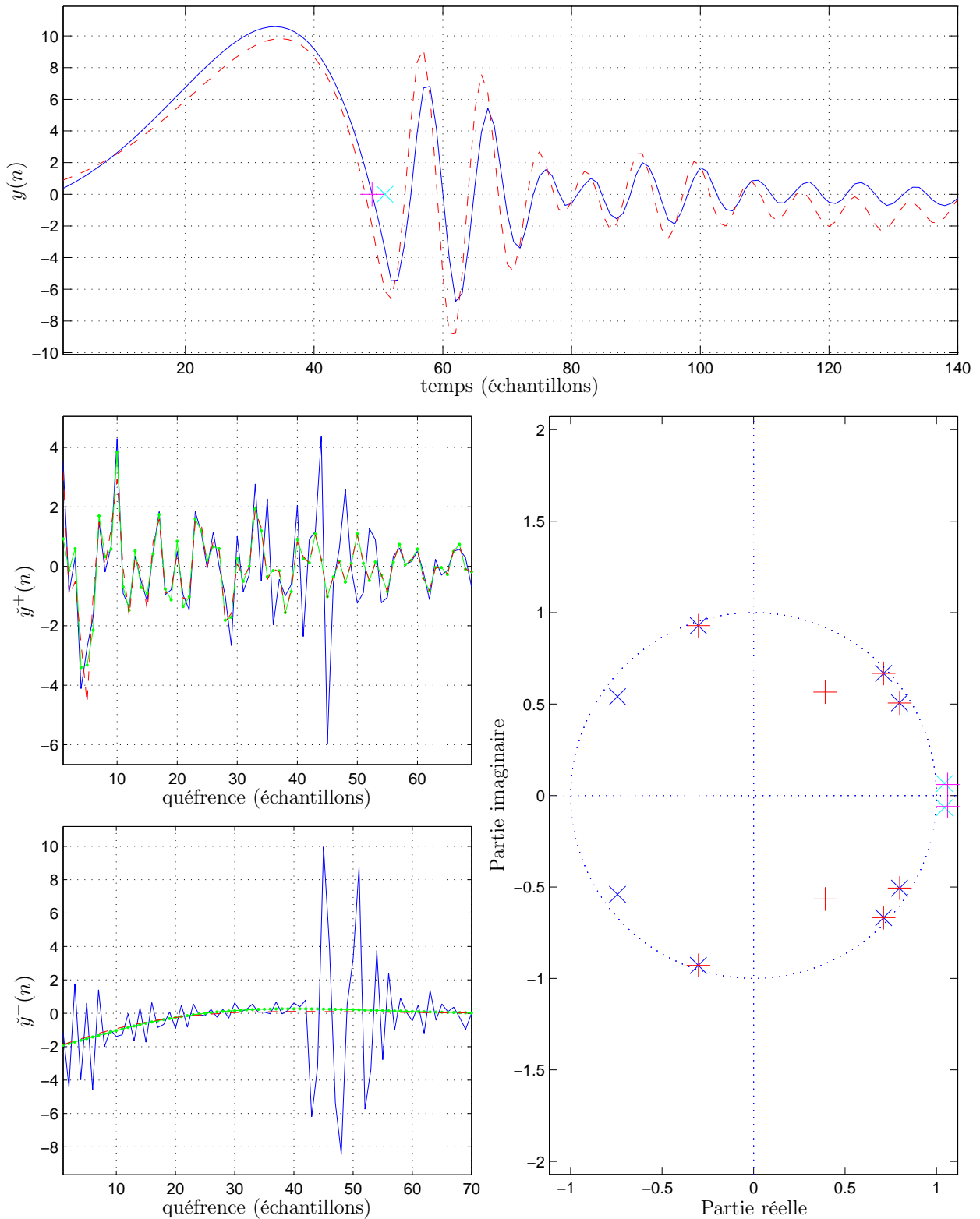


FIGURE 7.12 – Effet du fenêtrage 2. En haut : Signal $y(n)$ de synthèse (-) et signal reconstruit (- -) ainsi que le GCI (x) et le GCI estimé (+). A gauche : signaux $\tilde{y}^\pm(n)$ théoriques (-●-), estimés (-) et reconstruits (- -). A droite : pôles du modèle de synthèse (x et x) and pôles estimés (+ et +).

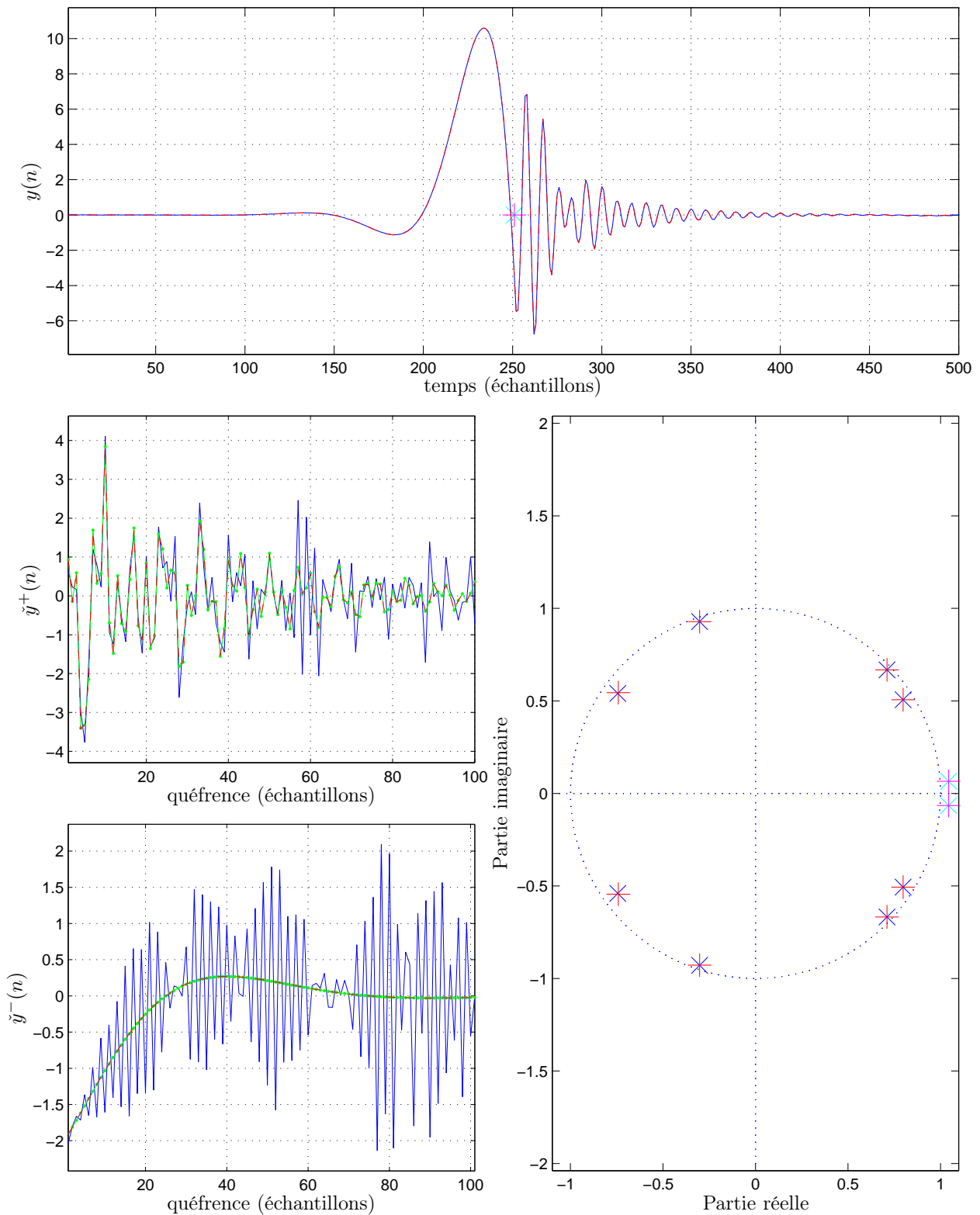


FIGURE 7.13 – Effet d'un bruit de puissance 10^{-5} . En haut : Signal $y(n)$ de synthèse (-) et signal reconstruit (- -) ainsi que le GCI (\times) et le GCI estimé ($+$). A gauche : signaux $\tilde{y}^\pm(n)$ théoriques (\bullet), estimés (-) et reconstruits (- -). A droite : pôles du modèle de synthèse (\times et \times) and pôles estimés ($+$ et $+$).

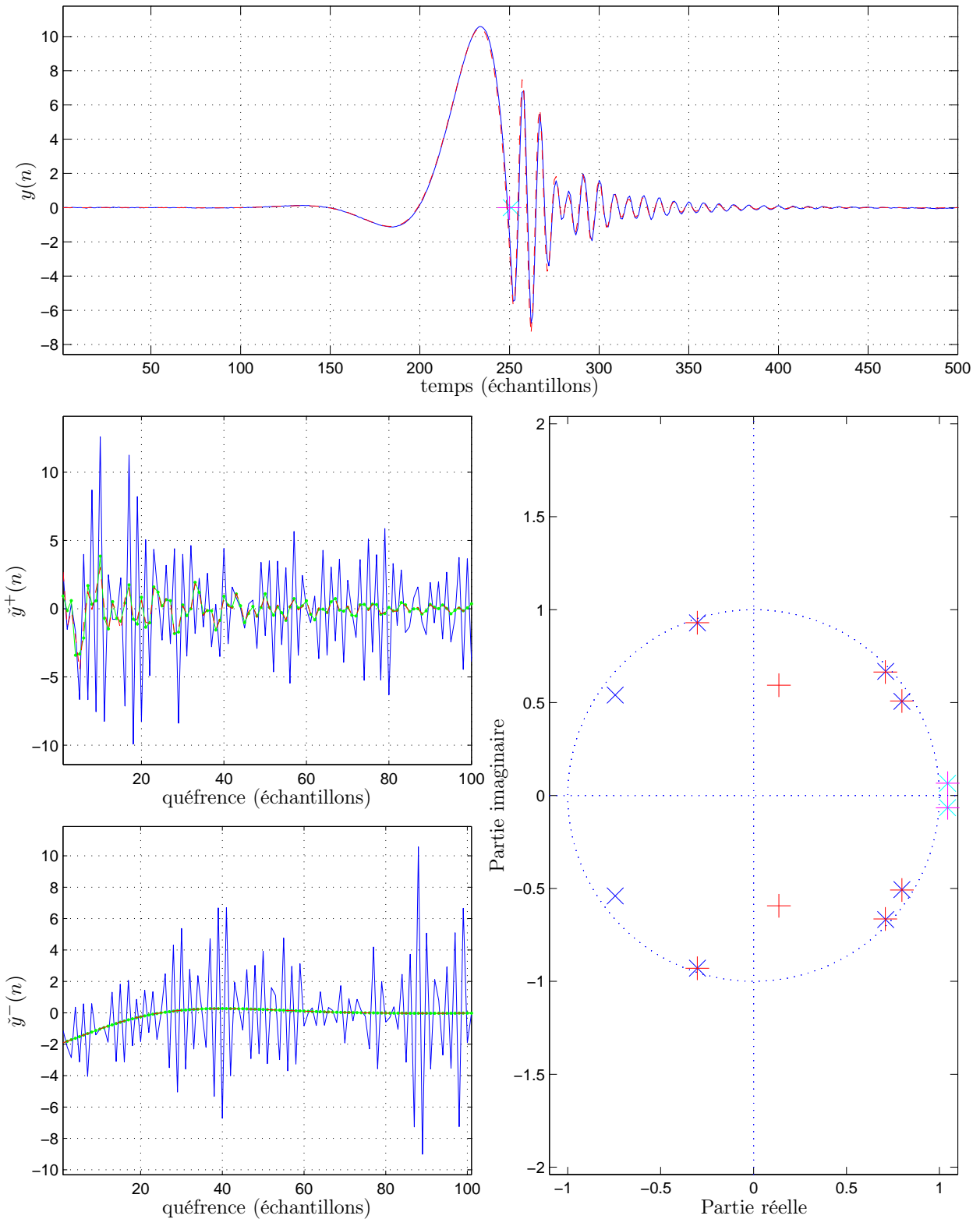


FIGURE 7.14 – Effet d’un bruit de puissance 10^{-4} . En haut : Signal $y(n)$ de synthèse (-) et signal reconstruit (- -) ainsi que le GCI (x) et le GCI estimé (+). A gauche : signaux $\tilde{y}^\pm(n)$ théoriques (-●-), estimés (-) et reconstruits (- -). A droite : pôles du modèle de synthèse (x et x) and pôles estimés (+ et +).

7.3.4.3 Signaux réels

J'ai testé l'algorithme dans son état d'avancement actuel sur des signaux vocaux. Afin d'illustrer les résultats obtenus, je présente ici la comparaison entre le signal de source glottique estimé par mon algorithme et celui estimé par la méthode proposée par Drugman dans [DBD11].

L'intégration de l'algorithme dans une chaîne de traitement par trames a été réalisé à l'aide de la toolbox Matlab® GLOAT proposée par Drugman et fondée sur les outils présentés dans [Dru11]. Le traitement se déroule selon les étapes suivantes,

- estimation du voisement et de la fréquence fondamentale [DA11],
- détection des GCI [DD09, DA11],
- découpage des trames selon la méthode proposée dans [DBD11] puis estimation paramétrique de la source glottique avec l'algorithme présenté ici. Afin d'améliorer la robustesse de l'estimation de la paire de pôles glottiques, une pénalisation est ajoutée lorsque la fréquence qui lui correspond est supérieure à 1 kHz.

La figure 7.15 présente un résultat obtenu pour un extrait de signal (voyelle anglaise /e/). La courbe magenta présente les résultats obtenus par la méthode de Drugman, dont la pertinence a été abondamment discutée dans [Dru11, DBD11]. La courbe bleue correspond au résultat obtenu avec l'algorithme proposé ici. Les résultats montrent que l'estimation de la source glottique proposée ici est tout à fait cohérente avec les méthodes existantes. De plus, peu de périodes présentent des paramètres estimés aberrants.

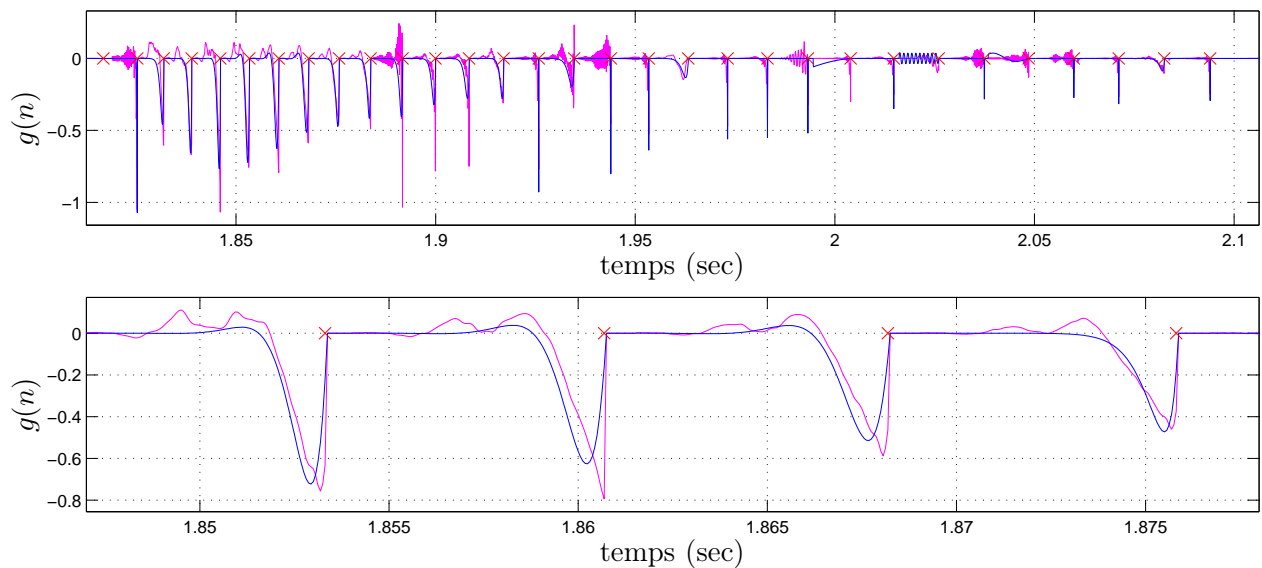


FIGURE 7.15 – (-) source glottique obtenue par la méthode [DBD11]. (-) source glottique obtenue par l'algorithme présenté ici. La figure inférieure présente un extrait de la figure supérieure.

Il est difficile d'estimer les performances de l'algorithme sur des signaux réels dans son état actuel, principalement en raison de l'absence de prise en compte de la périodisation dans l'algorithme d'estimation. De plus, la mesure de signal de source glottique n'est pas possible et seules des comparaisons avec les méthodes peuvent donner des informations sur la performance de l'algorithme. Un travail de recherche est nécessaire pour prendre en compte le recouvrement circulaire des réponses impulsionnelles qui intervient en cas d'excitation périodique ($e(n) = \text{III}_{T_0}(n)$ dans (7.31)). Cela permettra en outre une estimation plus robuste de la partie causale (filtre vocal). Des

premiers tests, non présentés ici, ont montré que l'algorithme est inutilisable en l'état pour cette estimation. La différence de performances dans l'estimation des parties causales et anticausales est vraisemblablement due aux temps de réponse des filtres vocaux (causaux), généralement plus longs que ceux des filtres glottiques (anticausaux). La perturbation apportée par la circularisation est donc plus sensible pour la partie causale que pour la partie anticausale.

7.4 Conclusion et perspectives

J'ai présenté dans ce chapitre deux méthodes d'estimation paramétrique pour le problème de la séparation source-filtre appliquée au signaux de voix. Ces deux méthodes sont fondées sur un modèle causal/anticausal tous-pôles.

Le premier algorithme repose sur un opérateur \mathcal{P} permettant de construire un signal infiniment périodique s à partir d'un échantillon de signal à analyser y après détection de la fréquence fondamentale. Les pôles de ce signal infiniment périodique sont ensuite transformés en zéros à l'aide de l'opérateur \mathcal{H} que j'ai introduit. Enfin, les pôles causaux et anticausaux, le gain et la position du GCI dans la trame d'analyse sont estimés à partir du signal $\mathcal{H} \circ \mathcal{P}(y)$. La pertinence et les performances de cet algorithme ont été montrées pour le cas des signaux synthétiques non bruités. Nous avons constaté que cet algorithme est encore très sensible à une mauvaise estimation de la fréquence fondamentale, mais de nombreux algorithmes robustes permettent d'estimer cette dernière avec précision (algorithme YIN [dCK02], analyse de l'auto-covariance, analyse cepstrale etc.). La sensibilité au bruit est elle aussi importante. En revanche, un mauvais dimensionnement du modèle d'analyse par rapport au modèle de synthèse ne conduit en aucun cas à des estimations aberrantes (estimation non perturbée si le nombre de pôles du modèle d'analyse est plus grand que le nombre de pôles du modèle de synthèse, estimation partielle dans le cas inverse). Les limitations se confirment lors de l'application de l'algorithme aux signaux réels, vraisemblablement en raison de la grande sensibilité de l'opérateur \mathcal{H} au bruit. Plusieurs solutions sont envisagées pour améliorer les performances de cet algorithme.

- Régularisation de l'estimation par contrainte de l'opérateur $\mathcal{H} \circ \mathcal{P}$: nombre de pôles, position des pôles causaux (basses fréquences) et anti-causaux.
- Réduction de la sensibilité au bruit en utilisant une déconvolution de type Wiener pour le calcul de l'opérateur \mathcal{H} .
- Modèle stochastique et optimisation au sens du maximum de vraisemblance.

Le choix ayant été fait de se diriger vers une méthode prenant en compte la robustesse dès le début de son élaboration et fondée sur la séparation causale-anticausale par représentation cepstrale, un deuxième algorithme a été proposé. Cet algorithme permet une prise en compte de la présence de bruit dans le signal par modèle stochastique. Nous avons tout d'abord étudié précisément la définition du cepstre complexe et ses propriétés. Ces considérations nous ont amené à examiner de près les propriétés d'holomorphie de la transformée de Laplace (ou en Z dans le cas discret) d'un signal et de son logarithme complexe sur un domaine contenant l'axe imaginaire pur (ou le cercle unité dans le cas discret). Comparé aux méthodes habituelles de calcul du cepstre complexe dans le traitement numérique de la parole, le cepstre différentiel apparaît alors comme une solution idéale pour le calcul du cepstre complexe car il permet à la fois

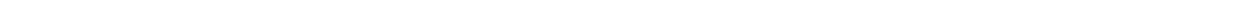
- un calcul rapide avec deux FFT et une IFFT à partir de la trame d'analyse,
- la préservation de l'holomorphie sur le cercle unité par absence de calcul direct du logarithme complexe, ce qui supprime la nécessité d'effectuer une recherche numérique des sauts artifi-

ciels (coupures) engendrés dans la partie imaginaire du logarithme complexe lors d'un calcul direct par détermination principale.

L'algorithme proposé permet une estimation des paramètres du modèle tous-pôles dans le cas d'un GCI isolé (réponse impulsionnelle au filtre glottique en série avec le filtre vocal). L'algorithme est robuste à la troncature de la réponse impulsionnelle. La présence d'un bruit faible est bien compensée par la recherche de pôles par maximisation de vraisemblance mais l'augmentation de la puissance de bruit altère assez rapidement les cepstres différentiels, ce qui ne permet plus à l'algorithme de faire l'estimation dans des conditions satisfaisantes. Les premiers tests sur signaux réels de l'algorithme dans son état actuel montrent des résultats déjà encourageants. Cependant, la prise en compte de la périodisation (circularisation de la réponse impulsionnelle) sera une étape nécessaire à l'obtention d'un algorithme entièrement fonctionnel. De plus, je pense que la périodisation ne pourra qu'avoir un effet bénéfique sur la résistance au bruit, car le RSB sera automatiquement plus élevé. Enfin, deux solutions sont envisagées pour diminuer la sensibilité au bruit :

- débruitage en amont de la séparation causal/anticausal,
- prise en compte d'un modèle de bruit plus fin (bruit gaussien centré sur le signal d'analyse et transformation pour modéliser son influence sur les cepstres différentiels).

Pour finir, notons que l'adaptation de la méthode à un filtre vocal pôle/zéro devrait être immédiate puisque les pôles et les zéros sont traités de façon quasi-identique dans le cepstre complexe (voir tableau 7.2 et remarque 6).





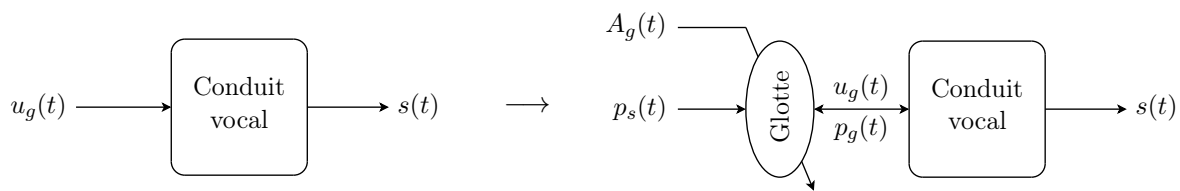
Troisième partie

Modèles à géométrie forcée et
analyse-synthèse



SOMMAIRE DE LA PARTIE

La partie **III** est consacrée à l'analyse-synthèse dans le cadre des modèles à géométrie forcée. Ainsi nous passons du choix **C1**, acoustique forcée (cf section 2.2.3), de la partie précédente au choix **C2**, système aéro-acoustique à géométrie pilotée, comme illustré ci-dessous.



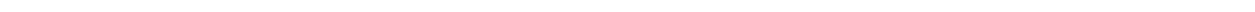
Un modèle à géométrie forcée est constitué de trois ingrédients,

- I1** un modèle de glotte,
- I2** un modèle de conduit vocal et de rayonnement,
- I3** un modèle de couplage entre les deux.

Je propose dans le chapitre 9 d'étudier l'apport du couplage à l'aide d'un modèle à géométrie forcée inspiré du modèle source-filtre étudié dans la partie précédente mais intégrant l'effet du couplage entre la glotte et le conduit vocal sur l'acoustique glottique. Pour rester proche de l'interface entre modèles physiques et modèles de signaux, le forçage géométrique est assuré par un modèle de glotte simple, sous forme de diaphragme piloté uniquement par l'aire glottique $A_g(t)$ et dont l'effet est pris en compte par l'équation de Bernoulli (**I1**). Le couplage (**I3**) est assuré par continuité de l'état acoustique à l'interface entre la glotte et le conduit vocal. Afin de rester proche du modèle source-filtre, le modèle de conduit vocal (**I2**) choisi est une concaténation de tubes droits dans laquelle se propagent des ondes planes sans pertes, modèle hérité des travaux de Lagrange et Bernoulli [Lag61, Ber62]. Ce modèle est rappelé et reformalisé en annexe G.

Pour améliorer la modélisation de la propagation des ondes dans le conduit vocal (**I2**), je propose ensuite de nous intéresser au modèle de Webster-Lokshin [Hé02] dans les tubes à géométrie variable. Les méthodes actuelles permettant de construire des structures de guides d'ondes à partir de ce modèles [Mig09] génèrent des instabilités dans le cas des tubes concaves. Une solution à ce problème est proposée.

Les modèles de fonctions d'aire glottique $A_g(t)$, qui pilotent le modèle de glotte 1D du chapitre 9, sont étudiés dans le chapitre 8. Deux approches sont envisagées : le modèle LF présenté dans le chapitre 6 et des modèles dynamiques de diversité [De 01] inspirés de modèles physiques auto-oscillants des plis vocaux.



Chapitre 8

Analyse-Synthèse de fonctions d'aire glottique

Sommaire

8.1	Construction d'un dictionnaire de test par estimation de fonctions d'aire glottique à partir de vidéo-endoscopies ultra-rapides	146
8.2	Modèle LF	147
8.2.1	Algorithme d'estimation des paramètres LF	150
8.2.1.1	Description générale de l'algorithme	150
8.2.1.2	Initialisation	150
8.2.1.3	Optimisation des paramètres LF	152
8.2.2	Performances de l'algorithme d'estimation des paramètres LF	154
8.2.2.1	Signaux de synthèse	154
8.2.2.2	Signaux réels	154
8.3	Modèles de diversité en systèmes dynamiques non linéaires	157
8.3.1	Contexte, motivations et premières applications	157
8.3.2	Modèle à une masse	158
8.3.2.1	Modèle Théorique	158
8.3.2.2	Réduction du problème	159
8.3.2.3	Système de Lur'e (à temps continu)	159
8.3.3	Modèle à deux masses	160
8.3.3.1	Modèle mécanique	160
8.3.3.2	Forces de pression	161
8.3.3.3	Adaptation à la structure des systèmes de Lur'e	162
	Approximations	162
	Cas où $\psi_1 = \psi_2$	163
	Application au modèle à deux masses	164
	Équation différentielle unique	164
	Système de Lur'e (à temps continu)	165
	Discrétisation	165
8.3.3.4	Utilisation pour l'analyse/synthèse	166
8.4	Conclusion	166

Dans ce chapitre, nous étudions la question de la modélisation de la fonction d'aire glottique $A_g(t)$ qui pilotera le modèle de glotte présenté au chapitre 9. Une revue historique des modèles de fonction d'aire glottique est proposée en annexe E. Je présente ici les deux approches que j'ai développées pour analyser et synthétiser des fonctions d'aire glottique à partir des données disponibles dans le corpus USC_2008_02.

Modèle LF. Une observation des fonctions d'aire glottique extraites des vidéos du corpus fait apparaître une forme générale de pulse avec des asymétries et des quotient ouverts variés. À partir de ces observations, je propose de modéliser les fonctions d'aire glottique à l'aide du modèle LF (annexe C), que nous ne considérons pas ici comme un modèle de débit glottique mais comme un modèle de pulse générique. La section 8.2 présente un algorithme d'estimation des paramètres LF à partir des fonctions d'aire glottique mesurées. Elle valide la proposition de modélisation des fonctions d'aire glottique par ce modèle. Cette représentation est efficace, mais elle ne contient pas une diversité suffisante pour représenter les signaux que j'ai pu observer dans la base de données (rebonds multiples notamment). Ceci m'a amené à proposer une seconde approche.

Modèle de diversité. Une approche possible pour modéliser la grande diversité des fonctions d'aire glottique observées est d'utiliser des modèles chaotiques. Cette approche a été explorée lors du stage de Mathias Le Borgne, encadré par moi-même et Thomas Hélie. Nous proposons, dans la section 8.3 d'utiliser les systèmes dits "de Lur'e", capables de produire une grande diversité de régimes pour générer des signaux d'aire glottique. Une méthode, originellement mise au point par O. De Feo et le professeur M. Hasler a déjà produit des résultats remarquables pour classifier diverses familles de signaux bio-métriques complexes (ECG, EGG, parole etc.) à l'aide de ces systèmes [De 01]. Ici, la difficulté est que l'on ne vise pas la catégorisation des régimes mais bien l'analyse et la synthèse des fonctions d'aire glottique avec leur variabilité. Les résultats pauvres obtenus par application directe de la méthode nous ont mené à proposer un modèle dynamique inspiré par la physique, ce qui a permis d'améliorer les résultats obtenus.

8.1 Construction d'un dictionnaire de test par estimation de fonctions d'aire glottique à partir de vidéo-endoscopies ultra-rapides

Afin d'établir un modèle de fonction d'aire glottique, commençons par observer une banque de signaux. Pour cela, nous avons à disposition le corpus USC_2008_02 déjà évoqué qui comporte, entre autres, des enregistrements endoscopiques en vidéo ultra-rapide de l'activité glottique. Les fonctions d'aire glottique $A_g(t)$ sont extraites des vidéos à l'aide de l'algorithme proposé par Deliyski dans [Del06], comme indiqué dans le chapitre 3.

L'algorithme de Deliyski, fondé sur l'analyse des moments de la luminance dans l'enregistrement vidéo, fait apparaître dans les signaux d'aire glottique estimés une composante basse fréquence, incluant un signal continu non nul, liée à la luminance moyenne de la vidéo, aux mouvements verticaux du larynx ainsi qu'aux mouvements de la caméra et de la source lumineuse. J'ai mis en place une détection de cette composante basse fréquence par filtrage moyenneur de largeur 8 périodes sur le signal $A_g(t)$ (la valeur moyenne de la fréquence fondamentale est estimée avec l'algorithme YIN [dCK02]). La figure 8.1 présente un exemple de signal $A_g(t)$ obtenu grâce à l'algorithme de Deliyski et sa version filtrée sans composante basse fréquence. Le désavantage de cet algorithme est que nous n'avons aucun moyen de retrouver la valeur absolue de l'aire glottique (facteur d'échelle) ainsi que

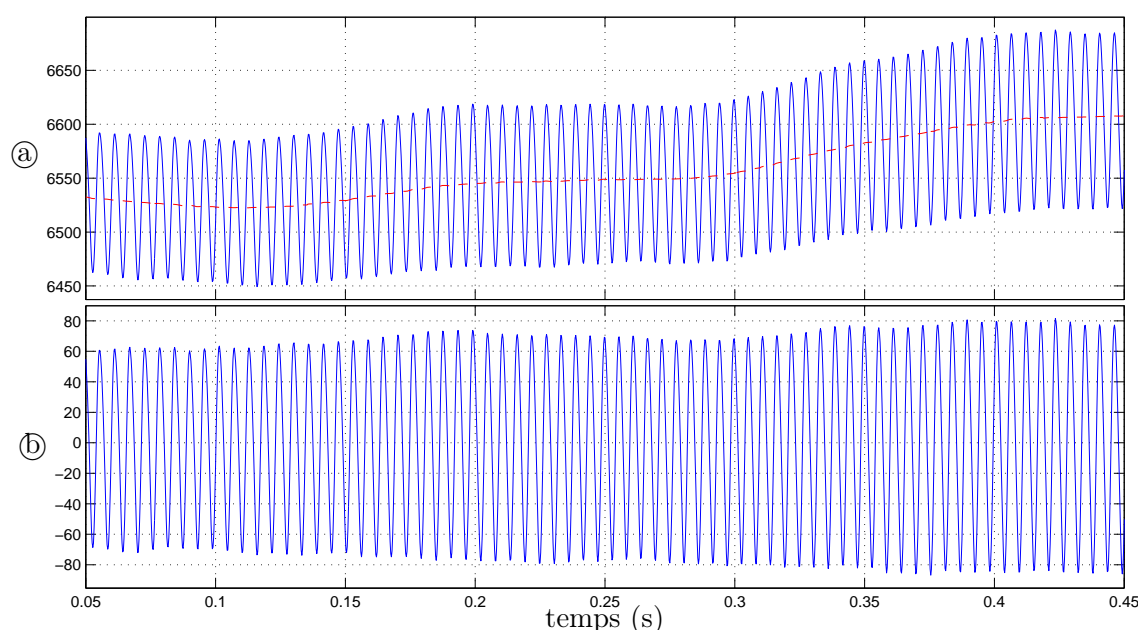


FIGURE 8.1 – Suppression de la composante BF des signaux d’aire glottique. (a) signal $A_g(t)$ estimé avec l’algorithme de Deliyski en bleu et sa moyenne glissante (composante BF) en rouge. (b) signal filtré sans composante basse fréquence.

sa valeur minimale (cas des glottes non complètement fermées). Dans le cas d’une utilisation pour la synthèse, le signal $A_g(t)$ obtenu par l’algorithme de Deliyski une fois filtré (figure 8.1 - (b)) sera donc redressé par ajout de la valeur minimale de $A_g(t)$ sur chaque trame. Dans le cas de la modélisation paramétrique par modèle LF qui nous intéresse dans la section suivante, nous travaillons sur $A'_g(t)$, ce qui évacue le problème de la valeur moyenne de $A_g(t)$.

Les figures 8.2 et 8.3 présentent respectivement des exemples de fonctions d’aires glottiques extraites du corpus et leur dérivées. Ces enregistrements ont été sélectionnés pour la variété de qualité vocale qu’ils présentent. En revanche, comme il l’a déjà été évoqué dans le chapitre 3, la fréquence d’échantillonnage de 4 kHz de la caméra ultra-rapide (et donc des signaux d’aire glottique) impose d’étudier des signaux de fréquence fondamentale relativement basse (inférieure à 500 Hz) afin de garder un minimum de contenu fréquentiel.

8.2 Modèle LF

Malgré la variabilité manifeste des signaux $A_g(t)$ observés, tous partagent la même forme générale d’impulsion présentant des asymétries et des quotients ouverts assez variés. Dans cette section, je propose un algorithme permettant de représenter les signaux $A_g(t)$ à l’aide du modèle de Liljencrants-Fant, ou LF (cf section 6.1.1 et annexes B et C). Le modèle LF, introduit pour représenter la source glottique, est ici considéré comme un modèle général d’impulsion paramétrique et nous verrons qu’il est adapté à la représentation paramétrique des signaux d’aire glottique. Le modèle LF a été choisi parmi les modèles d’impulsion existants car il est l’un de ceux présentant le plus grand nombre de paramètres (de degrés de liberté), et permet donc de représenter une grande variété de signaux temporels. De plus, la littérature abondante à son sujet en fait l’un des modèles les mieux documentés [FLL85, FKL94, Fan95, DDH06].

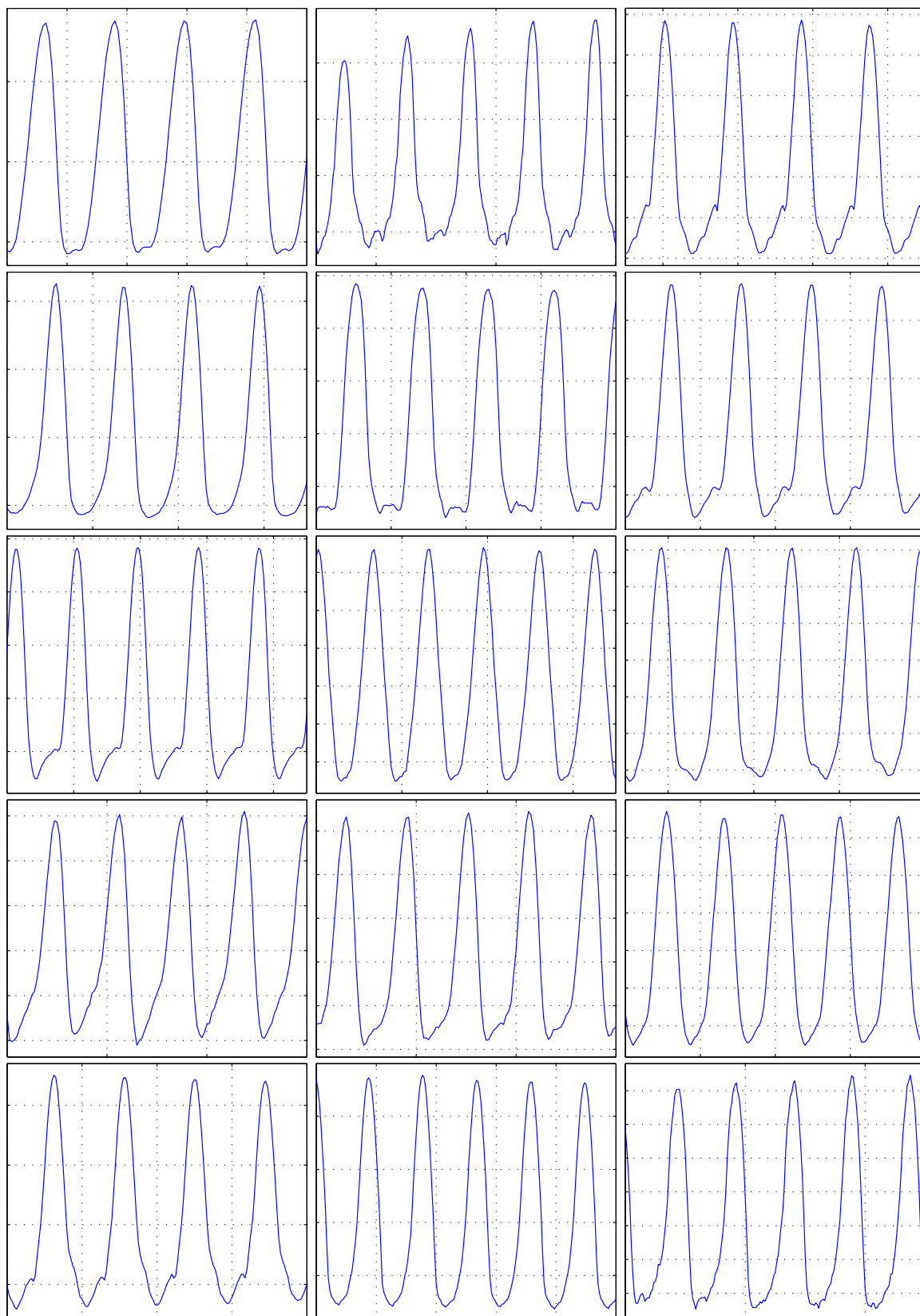
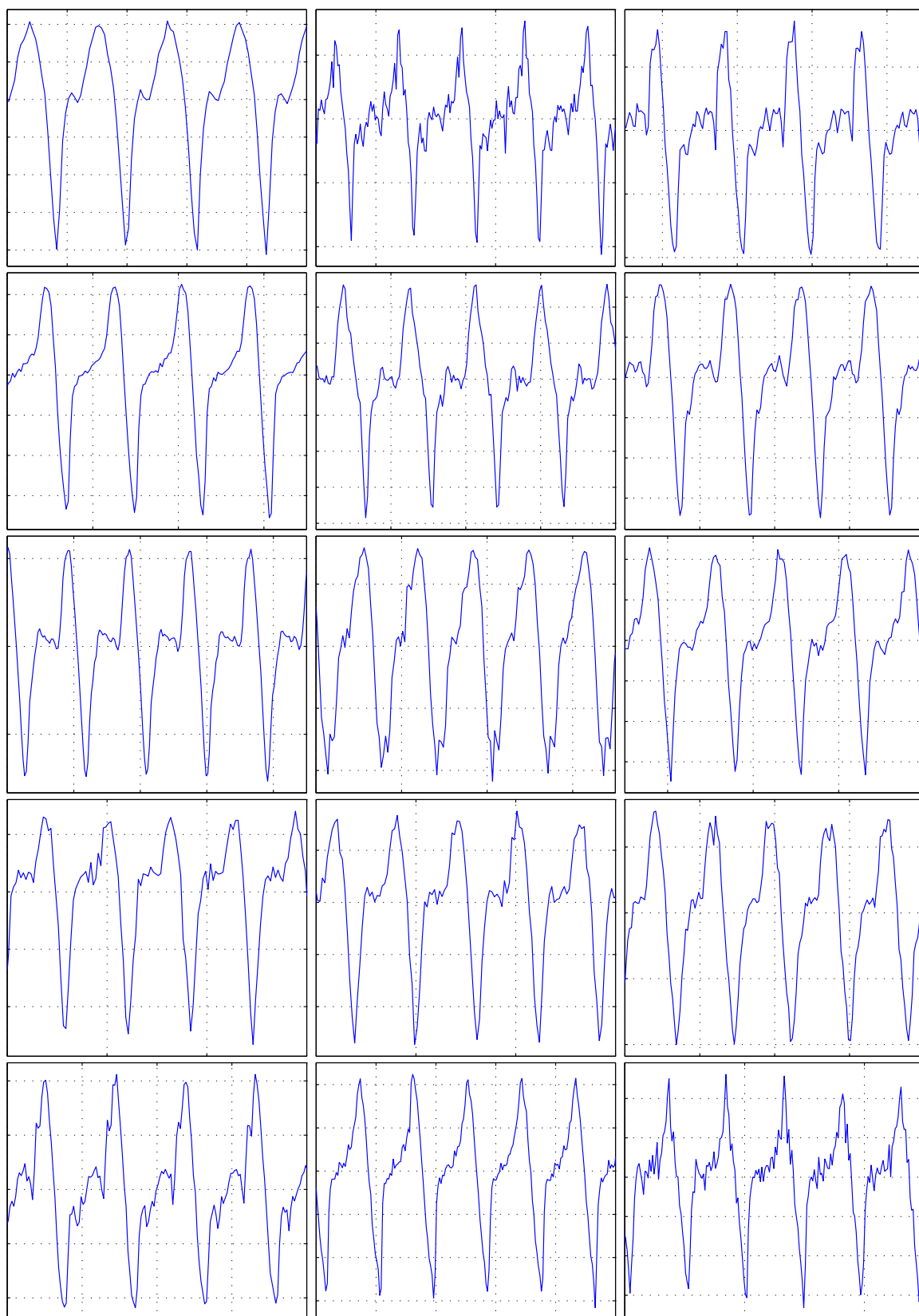


FIGURE 8.2 – Fonctions d'aire glottique $A_g(t)$ extraites du corpus USC_2008_02.

FIGURE 8.3 – Fonctions d'aire glottique dérivées $A'_g(t)$ extraites du corpus USC_2008_02.

8.2.1 Algorithme d'estimation des paramètres LF

L'algorithme présenté ici permet de représenter une dérivée de fonction d'aire glottique $A'_g(t)$ à l'aide du modèle LF. Le but de l'algorithme est donc de détecter les différentes périodes contenues dans le signal de mesure et de trouver les paramètres du modèle LF permettant d'approcher au mieux chacune des périodes. Pour rappel, la dérivée d'une période du modèle LF s'écrit

$$g'(t) = \begin{cases} g'_1(t) = -\mathcal{G} e^{\alpha(t-t_e)} \frac{\sin(\pi \frac{t}{t_p})}{\sin(\pi \frac{t_e}{t_p})} & 0 \leq t \leq t_e \\ g'_2(t) = -\frac{\mathcal{G}}{\varepsilon d_a} \left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)} \right) & t_e < t \leq t_c \\ g'_3(t) = 0 & t_c \leq t < T_0 \end{cases},$$

et les paramètres d'une période sont

$$\theta_{LF} = [\mathcal{G}, t_e, t_p, d_a, t_c, T_0]^T.$$

8.2.1.1 Description générale de l'algorithme

Supposons que le signal à analyser $A'_g(t)$ possède N périodes. L'algorithme d'estimation des paramètres se déroule alors en N étapes :

- première estimation rapide pour initialiser les paramètres,
- $1 \leq n \leq N - 1$: optimisation des paramètres LF des périodes n et $n + 1$: $\{\mathcal{G}(i), t_e(i), t_p(i), d_a(i), t_c(i), T_0(i)\}_{i \in \{n, n+1\}}$.

Les notations suivantes seront utilisées dans la suite de cette section :

- $A_g(t)$: signal échantillonné à analyser, contenant T échantillons et N périodes numérotées de 1 à N , l'horizon temporel est noté $\mathbb{T} = [1, T]$,
- \mathbb{T}_n : intervalle temporel comprenant les périodes n et $n + 1$,
- $\tilde{A}_g(t)$: représentation LF du signal $A_g(t)$, contenant les mêmes N périodes,
- $\mathcal{G}(n), t_e(n), t_p(n), d_a(n), t_c(n), T_0(n)$: paramètres LF de la période n ,
- ε : erreur quadratique de représentation paramétrique du signal $A_g(t)$ sur \mathbb{T} ,
- ε_n : erreur quadratique de représentation paramétrique du signal $A_g(t)$ sur \mathbb{T}_n .

8.2.1.2 Initialisation

La première étape est réalisée à l'aide de la méthode de détection d'événement décrite dans [KAZ00] (avec une fenêtre d'analyse gaussienne de 11 ms). La figure 8.4-@ présente un exemple d'événements détectés pour une dérivée de fonction d'aire glottique mesurée. Une fois les événements détectés, l'algorithme recherche, pour chaque période et dans cet ordre

- point $\tau_e(n)$: position du minimum entre deux événements adjacents,
- point $\tau_i(n)$: position du maximum entre les points $\tau_e(n - 1)$ et $\tau_e(n)$,
- point $\tau_p(n)$: moyenne des positions des points de valeur nulle entre les points $\tau_i(n)$ et $\tau_e(n)$,
- point $\tau_c(n)$: premier point de valeur nulle après le point $\tau_e(n)$,
- point $\tau_T(n)$ (début de la période n) : premier point de valeur nulle avant le point $\tau_i(n)$.

Les positions de points détectés permettent d'initialiser les paramètres LF pour chaque période,

$$\begin{aligned}
 t_e(n) &= \tau_e(n) - \tau_T(n), \\
 t_p(n) &= \tau_p(n) - \tau_T(n), \\
 d_a(n) &= \tau_a(n) - \tau_e(n), \\
 t_c(n) &= \tau_c(n) - \tau_T(n), \\
 T_0(n) &= \tau_T(n+1) - \tau_T(n), \\
 \mathcal{G}(n) &= \text{ordonnée du point } \tau_e(n),
 \end{aligned}
 \tag{8.1}$$

où les points $\tau_a(n)$ sont initialisés arbitrairement aux valeurs $\tau_e(n) + (\tau_c(n) - \tau_e(n))/5$. La figure 8.4-ⓑ illustre l'initialisation des paramètres LF correspondant aux événements détectés dans la figure 8.4-ⓐ.

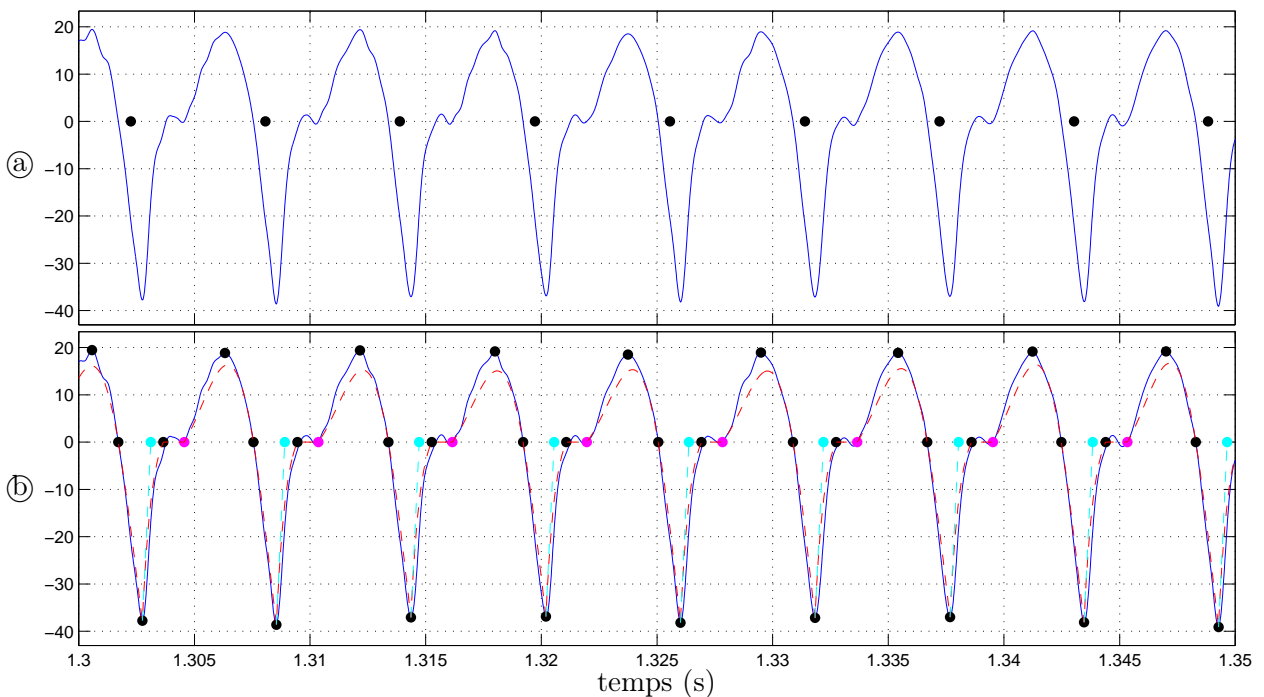


FIGURE 8.4 – Initialisation des paramètres LF. ⓐ événements détectés avec l'algorithme de Kawahara. ⓑ minima, maxima et points de valeur nulle détectés à partir des événements et initialisation des paramètres LF pour chaque période.

Il est apparu lors des premiers tests que, si la méthode de Kawahara est tout à fait adaptée au cas des signaux réels, elle est beaucoup moins performante pour les signaux de synthèse non bruités, sans que j'ai pu en comprendre les raisons. Une méthode *ad-hoc* a alors été mise au point pour ce cas. La détection d'événements est remplacée par une détection des minima locaux, ce qui permet d'accéder directement aux points $\tau_e(n)$. La suite de l'initialisation est identique.

8.2.1.3 Optimisation des paramètres LF

Après l'étape d'initialisation, les paramètres LF sont optimisés pour chaque couple de périodes adjacentes, comme présenté sur la figure 8.5. Dans la suite, nous noterons de façon identique les points $\tau_X(n)$ de la figure 8.5 et leur position temporelle. Cette optimisation des paramètres par paires de périodes permet d'optimiser la totalité des paramètres (y compris les durées des périodes $T_0(n)$) tout en gardant un nombre de paramètres relativement faible pour chaque optimisation. Chaque optimisation est réalisée par minimisation numérique (algorithme du simplexe) de la distance quadratique entre la cible et le modèle sur l'horizon temporel des deux périodes de travail

$$\varepsilon_n = \sqrt{\frac{\sum_{t \in \mathbb{T}_n} (A_g(t) - \widetilde{A}_g(t))^2}{\sum_{t \in \mathbb{T}_n} A_g(t)^2}}. \quad (8.2)$$

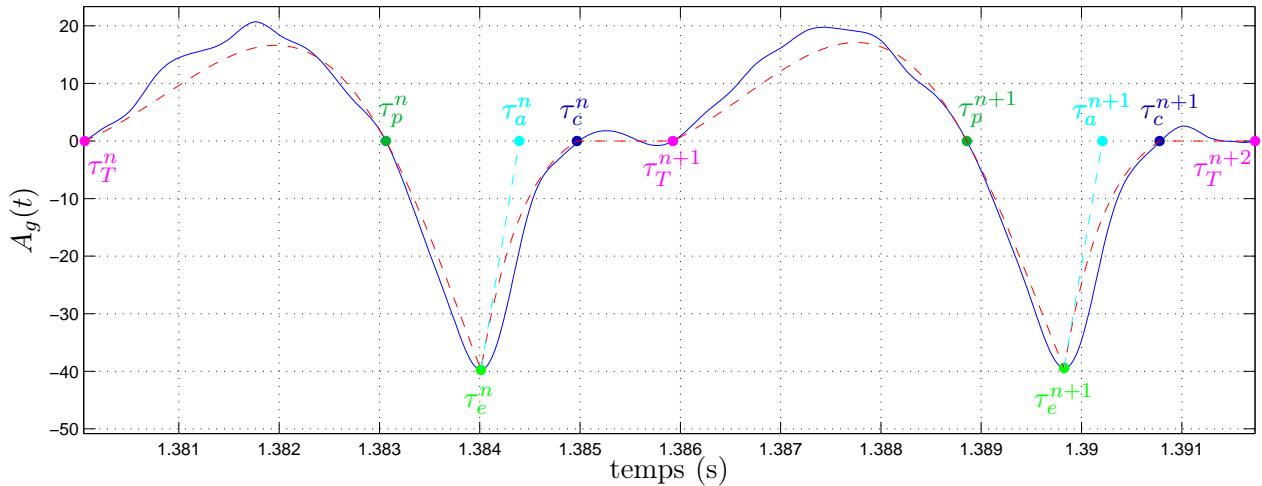


FIGURE 8.5 – Optimisation des paramètres LF sur deux périodes consécutives (les points $\tau_X(n)$ sont notés ici τ_X^n par souci de clarté). La position des points $\tau_T(n)$ et $\tau_T(n+2)$ sont fixes, les positions des autres points sont optimisées.

Les paramètres estimés par l'algorithme d'optimisation à l'étape n sont

- $\mathcal{G}(n)$ et $\mathcal{G}(n+1)$,
- $\tau_T(n+1)$,
- $\tau_p(n)$ et $\tau_p(n+1)$,
- $\tau_e(n)$ et $\tau_e(n+1)$,
- $\tau_a(n)$ et $\tau_a(n+1)$ et
- $\tau_c(n)$ et $\tau_c(n+1)$,

desquels sont déduit les paramètres de forme, le gain et la période du modèle LF pour les périodes n et $n+1$ par (8.1). Le respect des contraintes

$$\tau_T(n) < \tau_p(n) < \tau_e(n) < \tau_a(n) < \tau_c(n) < \tau_T(n+1) \quad \forall n \in [1, N] \quad (8.3)$$

$$\mathcal{G}(n) \geq 0 \quad \forall n \in [1, N] \quad (8.4)$$

est assuré par un mapping des paramètres. Plus précisément, les paramètres optimisés numériquement dans le cas des périodes n et $n+1$ ne sont pas les instants $\tau_p(n)$, $\tau_e(n)$, $\tau_a(n)$, $\tau_c(n)$, $\tau_T(n+1)$, $\tau_p(n+1)$, $\tau_e(n+1)$, $\tau_a(n+1)$, $\tau_c(n+1)$ et les valeurs de $\mathcal{G}(n)$ et $\mathcal{G}(n+1)$ mais les paramètres

$$\begin{aligned}
\tilde{\tau}_p(i) &= \operatorname{argth} \left(\frac{2\tau_p(i) - (\tau_e(i) + \tau_T(i))}{\tau_e(i) - \tau_T(i)} \right) \quad \forall i \in \{n, n+1\}, \\
\tilde{\tau}_e(i) &= \operatorname{argth} \left(\frac{2\tau_e(i) - (\tau_a(i) + \tau_p(i))}{\tau_a(i) - \tau_p(i)} \right) \quad \forall i \in \{n, n+1\}, \\
\tilde{\tau}_a(i) &= \operatorname{argth} \left(\frac{2\tau_a(i) - (\tau_c(i) + \tau_e(i))}{\tau_c(i) - \tau_e(i)} \right) \quad \forall i \in \{n, n+1\}, \\
\tilde{\tau}_c(i) &= \operatorname{argth} \left(\frac{2\tau_c(i) - (\tau_T(i+1) + \tau_a(i))}{\tau_T(i+1) - \tau_a(i)} \right) \quad \forall i \in \{n, n+1\}, \\
\tilde{\tau}_T(n+1) &= \operatorname{argth} \left(\frac{2\tau_T(n+1) - (\tau_p(n+1) + \tau_c(n))}{\tau_p(n+1) - \tau_c(n)} \right), \\
\tilde{\mathcal{G}}(i) &= \mathcal{G}(i)^2 \quad \forall i \in \{n, n+1\}.
\end{aligned} \tag{8.5}$$

Ainsi, l'optimisation numérique est effectuée sans contrainte, chacun des paramètres est optimisé sur \mathbb{R} et les points trouvés respectent par définition les contraintes (8.3-8.4). Pour calculer la valeur du critère d'optimisation (8.2), les paramètres LF sont retrouvés à l'aide de (8.1) après avoir inversé les équations (8.5) :

$$\begin{aligned}
\tau_p(i) &= \tau_T(i) \delta_p^+(i) + \tau_e(i) \delta_p^-(i) \quad \forall i \in \{n, n+1\}, \\
\tau_e(i) &= \tau_p(i) \delta_e^+(i) + \tau_a(i) \delta_e^-(i) \quad \forall i \in \{n, n+1\}, \\
\tau_a(i) &= \tau_e(i) \delta_a^+(i) + \tau_c(i) \delta_a^-(i) \quad \forall i \in \{n, n+1\}, \\
\tau_c(i) &= \tau_a(i) \delta_c^+(i) + \tau_T(i+1) \delta_c^-(i) \quad \forall i \in \{n, n+1\}, \\
\tau_T(n) &= \tau_c(n-1) \delta_T^+(n) + \tau_p(n) \delta_T^-(n) \quad \forall i \in \{n, n+1\}, \\
\mathcal{G}(i) &= \sqrt{\tilde{\mathcal{G}}(i)} \quad \forall i \in \{n, n+1\},
\end{aligned} \tag{8.6}$$

avec

$$\delta_X^\pm(i) = \frac{1 \pm \tanh(\tilde{\tau}_X(i))}{2}.$$

Remarque 7. Les équations (8.6) correspondent à un système tridiagonal pour lesquels il existe des algorithmes de résolution rapide.

Une fois les paramètres $\{\mathcal{G}(n), \tau_T(n), \tau_p(n), \tau_e(n), \tau_a(n), \tau_c(n)\}_{n \in [1, N]}$ optimisés, les paramètres LF permettant de représenter au mieux chaque période du signal d'entrée sont retrouvés grâce à (8.1).

8.2.2 Performances de l'algorithme d'estimation des paramètres LF

Sont ici présentés les résultats obtenus avec cet algorithme sur des signaux de synthèse dans un premier temps, puis sur des signaux d'aire glottique extraits des vidéos du corpus USC_2008_02 à l'aide de la méthode présentée dans la section 8.1.

8.2.2.1 Signaux de synthèse

Nous commençons par tester l'algorithme d'estimation des paramètres LF sur une série de 10 signaux de synthèse de 200 périodes générées par le modèle LF avec des paramètres tirés aléatoirement. Deux versions de l'algorithme ont été testées. La première estime l'intégralité des paramètres LF pour chaque période. La deuxième supprime la phase nulle du modèle, c'est-à-dire fixe $t_c = T_0$ pour chaque période. Notons que la deuxième version supprime un degré de liberté du modèle.

Le tableau 8.1 présente les résultats obtenus pour les deux versions de l'algorithme.

- T_{cal} est le temps de calcul moyen normalisé par le nombre de périodes,
- ε est l'erreur quadratique moyenne d'estimation sur le signal temporel (eq. (8.2) sur \mathbb{T}),
- ε_X est l'erreur relative moyenne pour chacun des paramètres, $\varepsilon_X = \text{moy}(|X(n) - \tilde{X}(n)|/X(n))$, où $X(n)$ est un des paramètres de synthèse et $\tilde{X}(n)$ le paramètre estimé lui correspondant.

	T_{cal}	ε	ε_g	ε_{T_0}	ε_{t_p}	ε_{t_e}	ε_{d_a}	ε_{t_c}
LF avec t_c	8.37 sec	0.1569	0.1838	0.1342	0.2793	0.2606	3.567	0.1979
LF sans t_c	6.83 sec	0.1645	0.2233	0.1475	0.1749	0.1558	2.063	X

TABLE 8.1 – Performances de l'algorithme de représentation par modèle LF pour les deux versions de l'algorithme (avec et sans le paramètre t_c) sur signaux de synthèse.

Si l'erreur quadratique d'estimation est relativement faible, il apparaît que les paramètres ne sont pas toujours très bien estimés. On remarque de plus que, si l'erreur d'estimation est légèrement plus faible pour la version avec le paramètre t_c , les paramètres sont globalement moins bien estimés. Une interprétation possible de ce phénomène est que le retrait d'un paramètre du modèle (d'un degré de liberté) permet de régulariser le problème d'estimation. Il semble apparaître donc ici que le modèle LF sans le paramètre t_c soit un bon compromis entre précision de la représentation des signaux et contrainte du modèle par le nombre de degrés de liberté. De plus, la suppression de ce paramètre entraîne un gain significatif en termes de temps de calcul.

8.2.2.2 Signaux réels

Afin de tester la pertinence du modèle LF pour représenter des signaux d'aire glottique $A_g(t)$, l'algorithme d'estimation a été testé sur des signaux $A'_g(t)$ du corpus. Nous n'avons pas ici de connaissance a priori des paramètres de la vibration glottique, aussi l'unique indicateur de performance que nous avons à disposition est l'erreur quadratique moyenne ε entre les signaux originaux et les signaux reconstruits.

Un ensemble d'extraits tirés du corpus présentant une variété de qualité vocale intéressante et totalisant environ 2000 périodes ont été analysés par les deux versions de l'algorithme. Le tableau 8.2 présente les valeurs moyennes de l'erreur de reconstruction ε et du temps de calcul normalisé par le nombre de périodes T_{cal} . Il apparaît de nouveau ici que la prise en compte du paramètre t_c ne permet qu'une très faible amélioration de la représentation tandis que sa non prise en compte

permet un gain significatif en terme de temps de calcul. La figure 8.6 présente des exemples de représentation LF avec et sans le paramètre t_c de signaux d'aire glottique mesurés. Comme on peut le voir, notamment dans les cas ④, ⑤ et ⑥, les résultats sont globalement assez similaires dans les deux cas. En revanche, on peut constater, notamment dans les cas ⑦ et ⑧, que le paramètre t_c permet de prendre en compte une phase complètement fermée non nulle. Cependant, j'ai pu constater aussi que sa prise en compte menait parfois à des résultats aberrants. Par exemple, on peut voir dans la dernière période de l'exemple ⑥ que la phase de retour a été quasiment supprimée dans le cas sans t_c . Nous savons que la phase de retour du paramètre LF a un impact fort sur le signal perçu dans le cas de la synthèse par modèle source-filtre [DDH06]. La non prise en compte de t_c permet donc d'une certaine façon de régulariser le signal $A_g(t)$.

	T_{cal}	ε
LF avec t_c	9.40 sec	0.2070
LF sans t_c	7.37 sec	0.2097

TABLE 8.2 – Performances de l'algorithme de représentation par modèle LF pour les deux versions de l'algorithme (avec et sans le paramètre t_c) sur signaux réels.

En conclusion, le modèle LF est un outil intéressant pour la représentation paramétrique des fonctions d'aire glottique. L'algorithme d'estimation des paramètres proposé ici amène à une estimation efficace des paramètres à partir d'un signal mesuré. La suppression du paramètre t_c permet un gain significatif en temps de calcul et en fiabilité (régularisation du problème d'estimation) pour une perte de performances faible en terme d'erreur d'estimation.

Cependant, il apparaît sur la figure 8.6 que le modèle LF, s'il permet une représentation paramétrique efficace des signaux d'aire glottique, ne permet que d'approcher la grande variabilité que présentent les signaux d'aire glottique. Les phénomènes fins, tels que les ouvertures et fermetures irrégulières des plis vocaux (les “rebonds” qui apparaissent sur les courbe de dérivée de fonction d'aire glottique), ne peuvent être pris en compte par une modèle paramétrique comme le modèle LF. Une possibilité pour générer plus de variabilité est d'utiliser des systèmes chaotiques. La section suivante présente comment ces systèmes peuvent être utilisés pour générer des signaux d'aire glottique.

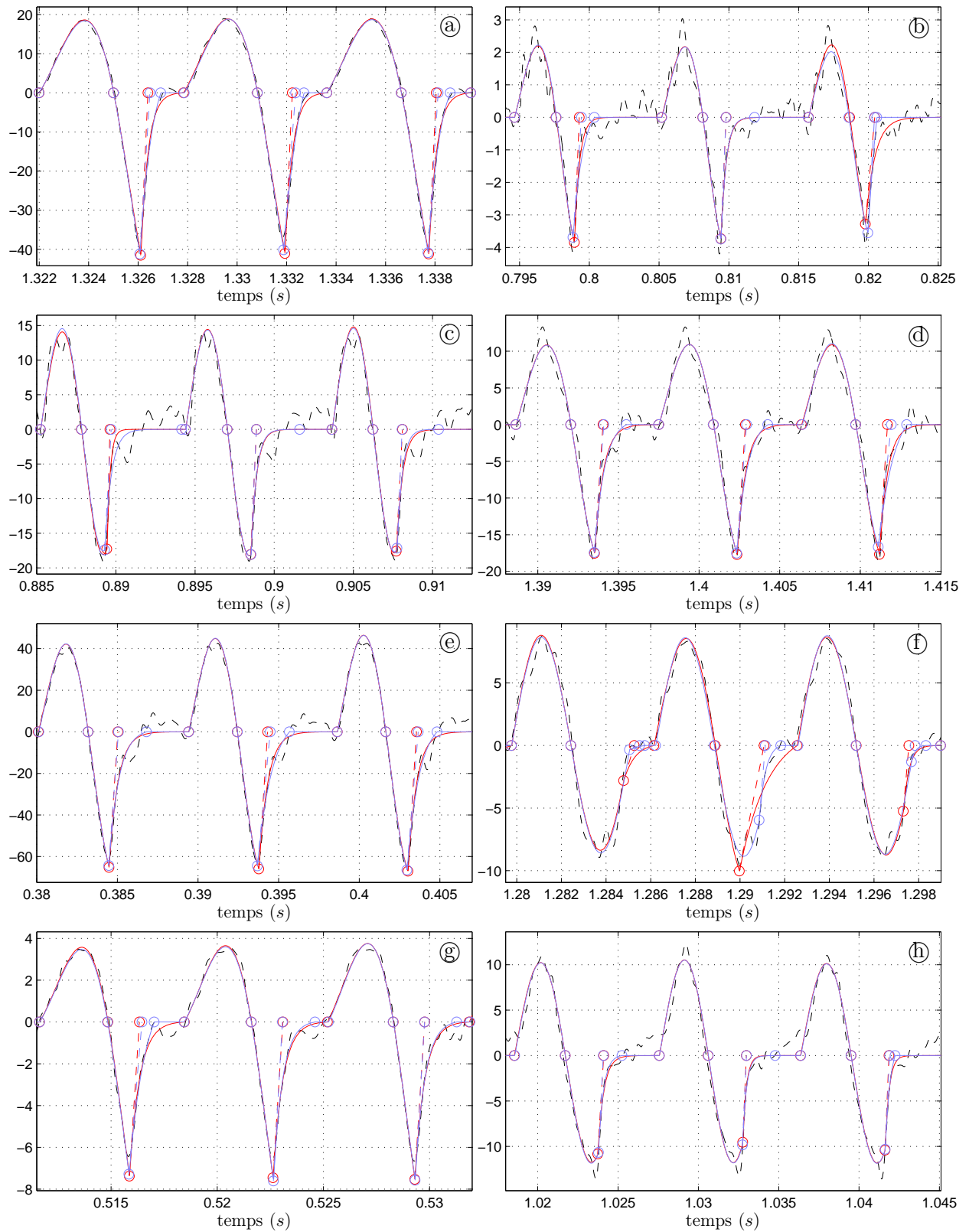


FIGURE 8.6 – Représentation LF de signaux d'aire glottique (avec et sans le paramètre t_c).
 (--) $\widetilde{A}_g(t)$, (-o-) $A_g(t)$ sans t_c , (-o-), $A_g(t)$ avec t_c .

8.3 Modèles de diversité en systèmes dynamiques non linéaires

8.3.1 Contexte, motivations et premières applications

Ce travail a fait l'objet du stage de M2R ATIAM de Mathias Le Borgne, encadré par moi-même et co-encadré par Thomas Hélie. Il a été impulsé suite à une discussion avec M. Hasler (EPFL) que nous remercions. Une partie du texte et des illustrations de cette section est issue de son rapport de stage [LB12].

Afin d'être capable de générer des signaux plus diversifiés, nous proposons de modéliser les signaux d'aire glottique à l'aide des systèmes dits "de Lur'e", capables de produire une grande diversité de régimes. Ces travaux suivent la démarche proposée par De Feo dans [De 01] où une méthode est proposée pour la classification de diverses familles de signaux bio-métriques complexes (ECG, EGG, parole etc.). Ici, le but visé n'est pas de catégoriser des régimes mais d'analyser et de synthétiser des fonctions d'aire glottique et leur variabilité.

Le modèle de Lur'e se présente sous la forme d'un système bouclé composé d'un filtre linéaire et d'une fonction non linéaire scalaire ψ [Kha02], comme présenté en figure 8.7 - (a).

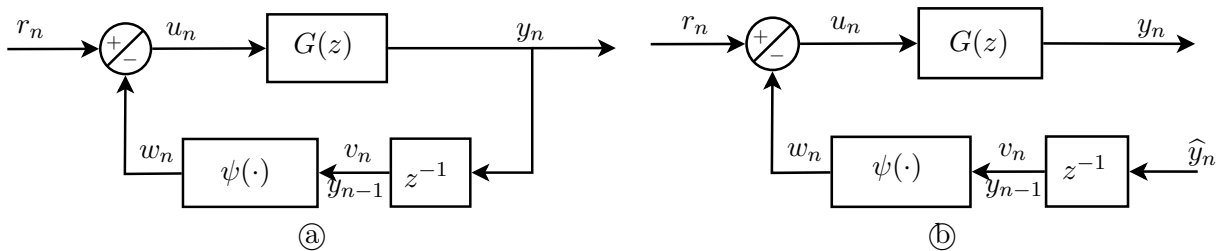


FIGURE 8.7 – (a) Structure générale d'un système de Lur'e à temps discret. (b) Système de Lur'e en boucle ouverte.

Une première tentative d'analyse des fonctions d'aire glottique fondée sur l'approche proposée par De Feo [De 01] est présentée en annexe F.3. Cette approche propose d'estimer les paramètres du système de Lur'e tel que le système en boucle ouverte (cf figure 8.7 - (b)) (i) ne soit pas l'identité et (ii) reproduise l'identité pour le signal observé. L'algorithme que nous proposons repose sur

1. un filtre $G(z)$ ARMA et une fonction non linéaire ψ paramétrique (vecteur de paramètres C),
2. une identification linéaire du filtre pour un C donné,
3. une mesure de la qualité de l'identification linéaire pénalisée dans le cas de la dégénérescence de la fonction non linéaire en fonction linéaire et du filtre en retard pur,
4. une optimisation des paramètres C par la méthode du recuit simulé.

Les résultats obtenus sont peu satisfaisants, vraisemblablement en raison de la trop grande généralité du modèle proposé. Afin d'améliorer les résultats de l'analyse/synthèse, nous nous proposons de régulariser le problème en informant le modèle de Lur'e. Pour ceci, nous allons construire un système de Lur'e à partir de modèles physiques simples des plis vocaux. Nous étudierons dans un premier temps un modèle à une masse afin de nous familiariser avec la mise en équations du système sous la forme d'un système de Lur'e. Nous verrons que ce système ne peut convenir et nous adapterons la méthode dans le cas d'un modèle à deux masses.

8.3.2 Modèle à une masse

Nous étudions dans un premier temps un modèle des plis vocaux à une masse (système masse-ressort-amortisseur, cf. figure 8.8). L'objectif est ici de transformer les équations régissant le système en un système de Lur'e à temps continu (cf annexe F.1).

8.3.2.1 Modèle Théorique

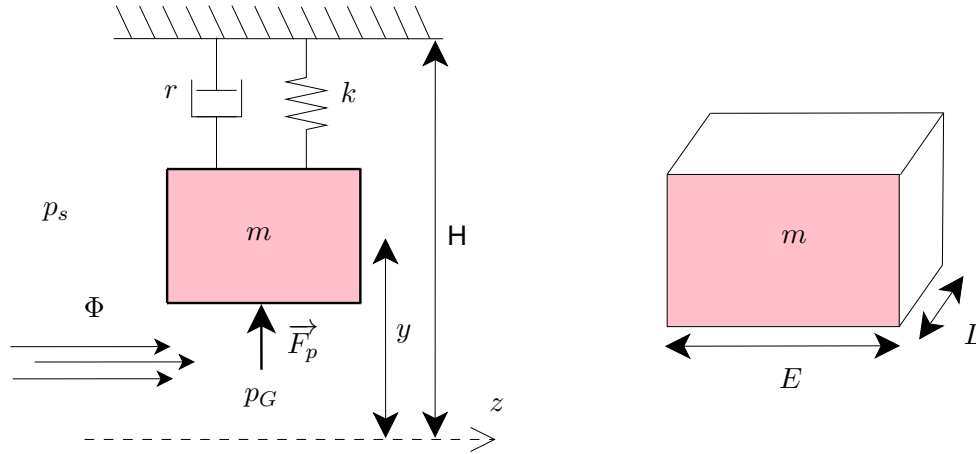


FIGURE 8.8 – Modèle à une masse

Nous considérons un modèle mécanique à une masse parallélépipédique présenté en figure 8.8. L'équation mécanique du système s'écrit

$$m\ddot{y} + r\dot{y} + k(y - y_0) = F_p = ELp_G, \quad (8.7)$$

où m est la masse, k le raideur du ressort, r le coefficient d'amortissement, p_G la pression glottique et (E, L) les dimensions de la face inférieure de la masse.

L'écoulement glottique est considéré comme parfait, incompressible, irrotationnel et permanent (cf [Cis08, p.38]). D'après la loi de Bernoulli, le flux d'air ϕ , constant sur toute la longueur de la glotte à un instant donné, suit en tout point M l'équation

$$p_M = \frac{1}{2}\rho\Phi^2 \frac{1}{A_M^2}, \quad (8.8)$$

où $A_M = Ly_M$ est l'aire de la glotte au point M suivant le plan normal à la direction z . Nous pouvons donc relier la pression glottique p_G à la pression sub-glottique p_s par

$$p_s - p_G = \frac{1}{2}\rho\Phi^2 \left[\frac{1}{(Ly)^2} - \frac{1}{(LH)^2} \right]. \quad (8.9)$$

Afin d'obtenir des équations compatibles avec un système de Lur'e, nous considérons que l'impédance d'entrée du conduit vocal est constante et vaut Z ,

$$Z\Phi = p_G. \quad (8.10)$$

En introduisant les paramètres réduits $\lambda = \frac{r}{m}$, $\omega_0 = \sqrt{\frac{k}{m}}$, $\mu = \frac{mH}{EL}$, $P_B = \frac{(ZLH)^2}{\rho}$, les variables réduites $x_H = \frac{y}{H}$ et $x_0 = \frac{y_0}{H}$ et en injectant (8.10) dans (8.9) pour éliminer ϕ , on obtient les équations adimensionnées

$$\frac{p_G}{\mu} = \ddot{x}_H + \lambda \dot{x}_H + \omega_0^2(x_H - x_0), \quad (8.11)$$

$$p_s - p_G = \frac{1}{2} \frac{p_G^2}{P_B} \left[\frac{1}{x_H^2} - 1 \right]. \quad (8.12)$$

8.3.2.2 Réduction du problème

D'après (8.12), on obtient

$$p_G = -p_s f \left(\frac{P_B}{p_s} \chi(x_H) \right), \quad (8.13)$$

où

$$\begin{cases} f(X) &= X \left(1 - \sqrt{1 + \frac{2}{X}} \right) \\ \chi(x) &= \frac{x^2}{1-x^2} \end{cases}.$$

Nous posons $F_K(x) = f \left(\frac{\chi(x)}{K} \right) = \frac{1}{K} \left(\frac{x^2}{1-x^2} \right) \left[1 - \sqrt{1 + 2K \left(\frac{1-x^2}{x^2} \right)} \right]$. La figure 8.9 présente le tracé de F_K pour deux valeurs de K . Une étude de cette fonction conduit à

- F_K est continue en 0, -1 et 1,
- pour $K \leq \frac{1}{2}$: $\mathcal{D}_{F_K} = \mathbb{R}$ (cf figure 8.9- (a)),
- Pour $K > \frac{1}{2}$: $\mathcal{D}_{F_K} = [-x_M, x_M]$ avec $x_M = \sqrt{\frac{K}{K-\frac{1}{2}}} > 1$ (cf figure 8.9- (b)).

En posant $x = x_H - x_0$, nous pouvons écrire

$$\ddot{x} + \lambda \dot{x} + \omega_0^2 x = -\frac{p_s}{\mu} f \left(\frac{P_B}{p_s} \chi(x + x_0) \right). \quad (8.14)$$

Nous définissons alors

$$\psi_{p_s}(x) = -p_s f \left(\frac{P_B}{p_s} \chi(x + x_0) \right). \quad (8.15)$$

L'équation du système de Lur'e associé à notre modèle est donc

$$\ddot{x} + \lambda \dot{x} + \omega_0^2 x = \frac{1}{\mu} \psi_{p_s}(x). \quad (8.16)$$

8.3.2.3 Système de Lur'e (à temps continu)

Nous avons défini un système de Lur'e à temps continu (figure 8.10) dont la fonction non linéaire est paramétrée par p_s :

$$\begin{cases} G(s) &= \frac{X(s)}{U(s)} = \frac{1/\mu}{s^2 + \lambda s + \omega_0^2}, & \psi_{p_s}(x) &= -p_s f \left(\frac{P_B}{p_s} \chi(x + x_0) \right). \\ u(t) &= \psi_{p_s}(x(t)) \end{cases}$$

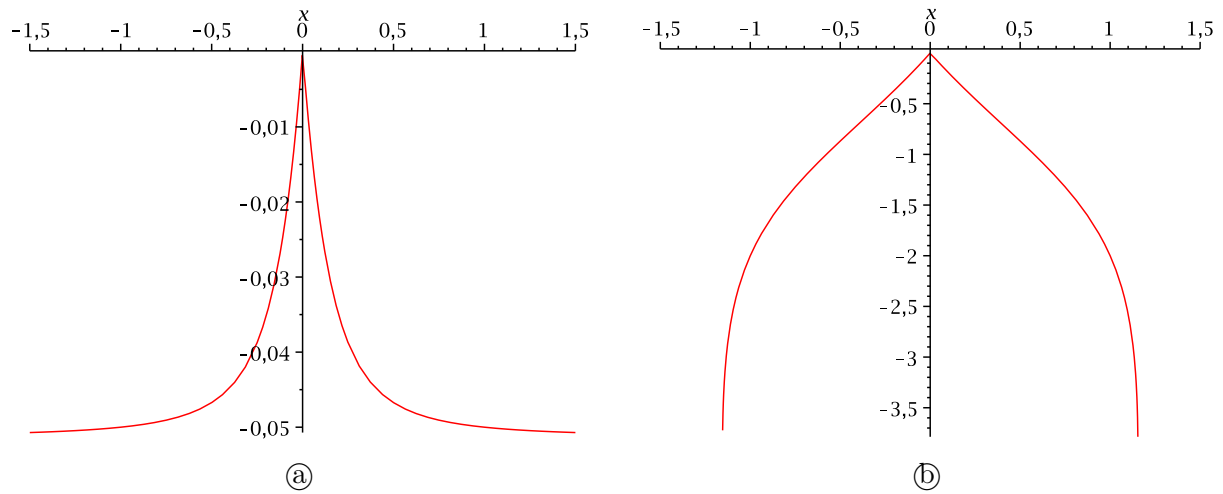


FIGURE 8.9 – Fonction F_K . (a) $K = 0.1$. (b) $K = 4$.

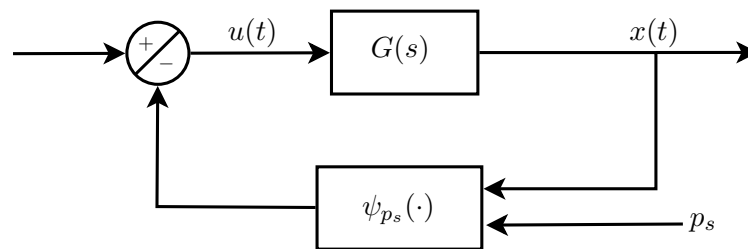


FIGURE 8.10 – Système de Lur'e à temps continu pour le modèle à une masse.

Ce modèle ne peut pas être chaotique, car on ne peut avoir de chaos pour une équation différentielle non linéaire d'ordre 2 à une dimension (le théorème de Poincaré-Bendixson affirme en effet qu'il ne peut y avoir dans ce cas là que des trajectoires convergentes, divergentes ou périodiques, cf [Sch96, p.112]). Il est donc nécessaire d'augmenter la complexité du modèle. Pour ceci, nous nous proposons de faire la même travail de mise en équation pour un modèle à deux masses.

8.3.3 Modèle à deux masses

Nous considérons maintenant le modèle à deux masses présenté en figure 8.11, classiquement utilisé pour modéliser les plis vocaux [IF72, Cis08, Bai09].

8.3.3.1 Modèle mécanique

Soient les paramètres réduits

$$\begin{cases} \lambda_1 = \frac{r_1}{m_1} \\ \lambda_2 = \frac{r_2}{m_2} \end{cases}, \quad \begin{cases} \omega_1 = \sqrt{\frac{k_1}{m_1}} \\ \omega_2 = \sqrt{\frac{k_2}{m_2}} \end{cases}, \quad \begin{cases} \omega_{12} = \sqrt{\frac{k_{12}}{m_1}} \\ \omega_{21} = \sqrt{\frac{k_{12}}{m_2}} \end{cases}.$$

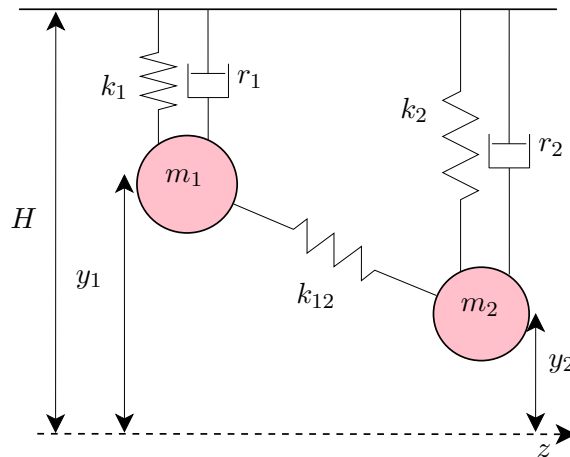


FIGURE 8.11 – Modèle à deux masses.

Nous choisissons des longueurs à vide égales pour simplifier les calculs, $y_{10} = y_{20} = y_0 = H - l_0$. Nous définissons alors les variables réduites

$$x_0 = \frac{y_0}{H}, \quad \begin{cases} x_{H1} = \frac{y_1}{H} \\ x_{H2} = \frac{y_2}{H} \end{cases}, \quad \begin{cases} x_1 = x_{10} - x_0 \\ x_2 = x_{20} - x_0 \end{cases},$$

ce qui amène aux équations mécaniques

$$\ddot{x}_1 + \lambda_1 \dot{x}_1 + \omega_1^2 x_1 + \omega_{12}^2 (x_2 - x_1) = \frac{F_{p1}}{H m_1}, \quad (8.17)$$

$$\ddot{x}_2 + \lambda_2 \dot{x}_2 + \omega_2^2 x_2 + \omega_{21}^2 (x_1 - x_2) = \frac{F_{p2}}{H m_2}. \quad (8.18)$$

Afin de retrouver une structure de Lur'e, nous devons simplifier le système pour obtenir une unique fonction non linéaire.

8.3.3.2 Forces de pression

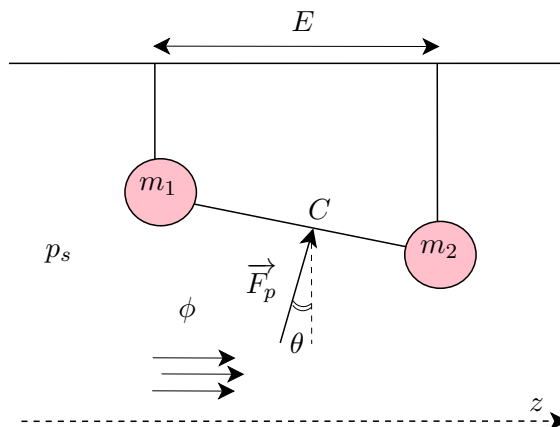


FIGURE 8.12 – Modèle trapézoïdal.

La force de pression est ici considérée comme constante sur toute la surface $S = EL$ délimitée par les masses m_1 et m_2 (voir figure 8.12, L est la largeur des plis vocaux). Elle dépend uniquement de p_C , la pression au point C , $F_{p_1} = F_{p_2} = F_{p_C}$. Nous obtenons par projection de \vec{F}_p sur l'axe vertical

$$F_{p_C} = EL \cos(\theta) p_C = \frac{EL}{\sqrt{1 + \left(\frac{H(x_{H_2} - x_{H_1})}{E}\right)^2}} p_C. \quad (8.19)$$

Nous pouvons alors adapter les équations (8.9) et (8.10) pour obtenir

$$Z\phi = p_C, \quad (8.20)$$

$$p_s - p_C = \frac{1}{2}\rho \left(\frac{\phi}{L}\right)^2 \left[\frac{1}{(Hx_C)^2} - \frac{1}{H^2} \right], \quad (8.21)$$

où $x_C = \frac{x_{H_2} + x_{H_1}}{2} = \frac{x_2 + x_1}{2} + x_0$. Nous retrouvons alors comme en (8.13)

$$p_C = -P_B \frac{x_C^2}{1 - x_C^2} \left[1 - \sqrt{1 + \frac{2p_s}{P_B} \frac{(1 - x_C^2)}{x_C^2}} \right] \quad (8.22)$$

avec $P_B = \frac{(ZLH)^2}{\rho}$. Finalement, la force de pression s'écrit

$$F_{p_C} = \frac{ELp_s}{\sqrt{1 + \left(\frac{H(x_2 - x_1)}{E}\right)^2}} f\left(\frac{P_B}{p_s} \chi\left(\frac{x_1 + x_2}{2} + x_0\right)\right) \quad (8.23)$$

avec

$$\begin{cases} f(X) &= X \left[1 - \sqrt{1 + \frac{2}{X}} \right], \\ \chi(x) &= \frac{x^2}{1 - x^2}. \end{cases} \quad (8.24)$$

Nous avons bien obtenu une unique fonction non linéaire ψ mais elle dépend des deux variables x_1 et x_2 , sans possibilité de réduction à une seule variable $x' = f(x_1, x_2)$ telle que $\psi = \psi(x')$. Nous devons donc faire des approximations supplémentaires qui nous permettront de retrouver la structure de Lur'e. Nous présenterons tout d'abord une série d'approximations des équations (8.17-8.18) permettant d'éliminer une variable puis une fonction non linéaire. Nous verrons ensuite comment appliquer ces approximations au cas présenté en 8.3.3.2.

8.3.3.3 Adaptation à la structure des systèmes de Lur'e

Approximations

1. **Changement de coordonnées** : Soit α tel que $0 < \alpha < 1$ et

$$\begin{cases} x_m &= (1 - \alpha)x_1 + \alpha x_2 \\ \epsilon &= x_2 - x_1 \end{cases}, \text{ soit } \begin{cases} x_1 &= x_m - \alpha\epsilon \\ x_2 &= x_m + (1 - \alpha)\epsilon \end{cases}. \quad (8.25)$$

Pour obtenir un système de Lur'e, nous souhaitons prendre $y = x_m$ et négliger ϵ .

2. **Changement de fonctions :** Pour deux fonctions non linéaires $\psi_1(x_1, x_2)$ et $\psi_2(x_1, x_2)$, soit β tel que $0 < \beta < 1$ et

$$\begin{cases} \Psi(x_m, \epsilon) &= (1 - \beta)\psi_1(x_1, x_2) + \beta\psi_2(x_1, x_2) \\ \delta\Psi(x_m, \epsilon) &= \psi_2(x_1, x_2) - \psi_1(x_1, x_2) \end{cases}, \quad (8.26)$$

ce qui est équivalent à

$$\begin{cases} \psi_1(x_m - \alpha\epsilon, x_m + (1 - \alpha)\epsilon) &= \Psi(x_m, \epsilon) - \beta\delta\Psi(x_m, \epsilon) \\ \psi_2(x_m - \alpha\epsilon, x_m + (1 - \alpha)\epsilon) &= \Psi(x_m, \epsilon) + (1 - \beta)\delta\Psi(x_m, \epsilon) \end{cases}. \quad (8.27)$$

Nous souhaitons prendre $\psi = F_m$ (soit $u = \psi(y) = F_m(x_m)$), et négliger δF_m .

3. **Développement limité :** Nous effectuons pour ceci un développement limité des ψ_i en supposant que $\epsilon \ll x_m$. À l'ordre 1,

$$\begin{cases} \Psi(x_m, \epsilon) = \underbrace{\Psi(x_m, 0)}_{\psi(y)} + \underbrace{[-\alpha(1 - \beta)\psi'_1(x_m, x_m) + (1 - \alpha)\beta\psi'_2(x_m, x_m)]}_{d\Psi(x_m) \simeq cste} \epsilon + o(\epsilon), \\ \delta\Psi(x_m, \epsilon) = \underbrace{\delta\Psi(x_m, 0)}_{\simeq cste} + \underbrace{[\alpha\psi'_1(x_m, x_m) + (1 - \alpha)\psi'_2(x_m, x_m)]}_{d(\delta\Psi)(x_m) \simeq cste} \epsilon + o(\epsilon). \end{cases} \quad (8.28)$$

4. **choix de α et β :** Nous souhaitons éliminer un maximum de termes dépendant de x_m dans le développement limité. À l'ordre 1, cela revient à rechercher α et β tels que les termes $d\Psi$ et $d(\delta\Psi)$ soient constants (éventuellement nuls). Nous pouvons considérer que cette contrainte de constance doit être appliquée en priorité à $d(\delta\Psi)$.
5. **Développement limité en x_m autour de l'équilibre :** Dans le cas où il serait impossible de rendre constants $d\Psi$ et $d(\delta\Psi)$ par simple choix de α et β , ou si $\delta\Psi(x_m, 0)$ n'est pas nulle, nous pouvons considérer que x_m est proche de la position d'équilibre x_q du point de vue de ces termes, ce qui revient à écrire

$$f(x_m) \simeq f(x_q) = K_q = cste \quad (8.29)$$

pour $f(x_m) = \delta\Psi(x_m, 0)$, $d\Psi(x_m)$ ou $d(\delta\Psi)(x_m)$ et donc

$$\psi_i = \underbrace{\Psi(x_m, 0)}_{\psi(y)} + K_q \epsilon. \quad (8.30)$$

Nous obtenons ainsi une unique fonction non linéaire ne dépendant que d'une seule variable, l'autre partie étant linéaire en ϵ .

Cas où $\psi_1 = \psi_2$

Nous obtenons dans ce cas

$$\begin{cases} \Psi(x_m, \epsilon) &= \underbrace{\Psi(x_m, 0)}_{\psi_1(x_m) = \psi_2(x_m)} + (\beta - \alpha)\psi'_1(x_m)\epsilon + o(\epsilon) \\ \delta\Psi(x_m, \epsilon) &= 0 + \psi'_1(x_m)\epsilon + o(\epsilon) \end{cases}. \quad (8.31)$$

Il apparaît que $\delta\Psi(x_m, 0) = 0, \forall x_m$, et que pour $\alpha = \beta$, $\Psi(x_m, \epsilon) = \Psi(x_m, 0) + o(\epsilon), \forall x_m$.

Application au modèle à deux masses

Dans le cas du modèle présenté en 8.3.3.2,

$$F_{pC} = F_{p1} = F_{p2} = \frac{ELp_s}{\sqrt{1 + \left(\frac{H(x_2-x_1)}{E}\right)^2}} f\left(\frac{P_B}{p_s} \chi\left(\frac{x_1+x_2}{2} + x_0\right)\right). \quad (8.32)$$

Cette expression suggère pour les approximations proposées $x_m = \frac{x_1+x_2}{2}$, soit $\alpha = \beta = \frac{1}{2}$. De plus, les fonctions non linéaires sont égales, nous sommes donc dans le cas où $\psi_1 = \psi_2$. Toujours dans l'objectif d'obtenir ψ homogène à une pression, nous choisissons $\psi_1 = \psi_2 = \frac{F_{pC}}{EL}$, ce qui donne

$$\Psi(x_m, \epsilon) = \frac{p_s}{\sqrt{1 + \left(\frac{H\epsilon}{E}\right)^2}} f\left(\frac{P_B}{p_s} \chi(x_m + x_0)\right).$$

Le développement limité de Ψ par rapport à ϵ nous permet d'obtenir

$$\Psi(x_m, \epsilon) = p_s f\left(\frac{P_B}{p_s} \chi(x_m + x_0)\right) \left[1 - \frac{1}{2} \left(\frac{H\epsilon}{E}\right)^2\right]$$

d'où, en négligeant ϵ^2 , nous pouvons définir la fonction non linéaire du système de Lur'e

$$\psi(x_m) = -P_B \frac{x_m^2}{1-x_m^2} \left[1 - \sqrt{1 + \frac{2p_s(1-x_m^2)}{P_B x_m^2}}\right].$$

Équation différentielle unique

Le système (8.17-8.18) devient donc

$$\ddot{x}_1 + \lambda_1 \dot{x}_1 + \omega_1^2 x_1 + w_{12}^2 (x_2 - x_1) = \frac{p_s}{\mu_1} \psi_{p_s}(x_m), \quad (8.33)$$

$$\ddot{x}_2 + \lambda_2 \dot{x}_2 + \omega_2^2 x_2 + w_{21}^2 (x_1 - x_2) = \frac{p_s}{\mu_2} \psi_{p_s}(x_m), \quad (8.34)$$

où $\begin{cases} \mu_1 = \frac{m_1 H}{E_1 L} \\ \mu_2 = \frac{m_2 H}{E_2 L} \end{cases}$ et $\begin{cases} K_1 = \frac{p_s}{\mu_1} \\ K_2 = \frac{p_s}{\mu_2} \end{cases}$. En reprenant les notations de 8.3.3.3, et en posant

$$\begin{cases} y = x_m \\ u = \psi_{p_s}(x_m) \end{cases},$$

nous obtenons à partir de (8.34-8.33)

$$[\lambda_\epsilon \dot{y} + \omega_\epsilon^2 y] + [\ddot{\epsilon} + \lambda_m \dot{\epsilon} + \omega_m^2 \epsilon] = K_\epsilon u, \quad (8.35)$$

$$[\ddot{y} + \lambda_M \dot{y} + \omega_M^2 y] + [\ddot{\epsilon} + \lambda_\epsilon \dot{\epsilon} + \omega_\epsilon^2 \epsilon] = K_M u, \quad (8.36)$$

avec

$$\begin{cases} \lambda_M = \frac{\lambda_1}{\alpha} + \frac{\lambda_2}{1-\alpha} \\ \omega_M^2 = \frac{1}{\alpha} \left(\omega_1^2 - \frac{\omega_{12}^2}{\alpha}\right) + \frac{1}{1-\alpha} \left(\omega_2^2 + \frac{\omega_{21}^2}{1-\alpha}\right) \\ K_M = \frac{K_1}{\alpha} + \frac{K_2}{1-\alpha} \end{cases},$$

$$\begin{cases} \lambda_m = \alpha \lambda_1 + (1 - \alpha) \lambda_2 \\ \omega_m^2 = \alpha \omega_1^2 + (1 - \alpha) \omega_2^2 + \omega_{21}^2 - \omega_{12}^2 \\ \lambda_\epsilon = \lambda_2 - \lambda_1 \\ \omega_\epsilon^2 = \omega_2^2 - \omega_1^2 \\ K_\epsilon = K_2 - K_1 \end{cases}$$

Nous pouvons finalement éliminer la variable ϵ en prenant une combinaison linéaire des équations (8.35) et (8.36) et de leurs dérivées (on montre qu'il est nécessaire d'utiliser les dérivées aux ordres 1 et 2 de ces équations). Nous obtenons alors une équation de la forme

$$\sum_{k=0}^4 \alpha_k y^{(k)}(t) = \sum_{l=0}^2 \beta_l u^{(l)}(t).$$

Système de Lur'e (à temps continu)

Le système de Lur'e correspondant (figure 8.13) s'écrit donc

$$\begin{cases} G(s) = \frac{Y(s)}{U(s)} = \frac{\sum_{l=0}^2 \beta_l s^l}{\sum_{k=0}^4 \alpha_k s^k}, \\ u(t) = \psi_{p_s}(y(t)) \end{cases}, \quad (8.37)$$

ou encore

$$\begin{cases} \dot{X}(t) = \tilde{A}X(t) + \tilde{b}u(t) \\ u(t) = \psi_{p_s}(y(t)) \\ y(t) = \tilde{c}^T X(t) \end{cases},$$

avec $\tilde{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{\alpha_3}{\alpha_4} & -\frac{\alpha_2}{\alpha_4} & -\frac{\alpha_1}{\alpha_4} & -\frac{\alpha_0}{\alpha_4} \end{bmatrix}$, $\tilde{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{\beta_4}{\alpha_4} \end{bmatrix}$, $\tilde{c} = \begin{bmatrix} \frac{\beta_0}{\alpha_4} \\ \frac{\beta_1}{\alpha_4} \\ \frac{\beta_2}{\alpha_4} \\ \frac{\beta_3}{\alpha_4} \end{bmatrix}$ et $X = \begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \\ y^{(3)} \end{bmatrix}$.

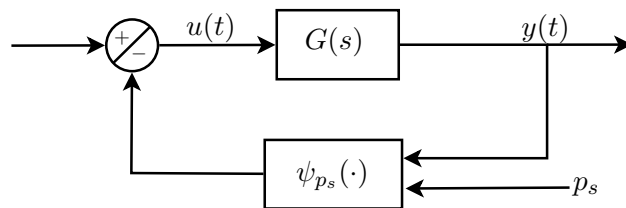


FIGURE 8.13 – Système de Lur'e à temps continu pour le modèle à deux masses.

Discrétisation

La discrétisation de $G(s)$ dans (8.37) est effectuée à l'aide de l'approximation des trapèzes sur la transformée de Laplace (voir [PR06] et l'annexe F.1.2). La fonction scalaire ψ_{p_s} est inchangée. Nous effectuons le changement de variable $s \mapsto \frac{2(z-1)}{T_S(z+1)}$ dans $G(s)$ pour obtenir

$$G(z) = \frac{\sum_{l=0}^4 b_l z^{-l}}{1 + \sum_{k=1}^4 a_k z^{-k}} \quad (\text{les coefficients } a_k \text{ et } b_l \text{ sont calculables analytiquement}).$$

8.3.3.4 Utilisation pour l'analyse/synthèse

L'utilisation d'un modèle physique fournit plusieurs simplifications par rapport au modèle utilisé dans l'algorithme d'identification lors de notre première tentative (section F.3.2). Notamment, nous disposons maintenant

- d'une structure forte pour la fonction non linéaire ψ qui dépend ici uniquement de trois paramètres,
- d'un ordre défini et faible pour les coefficients du filtre ($K = L = 4$),
- d'une bonne initialisation pour les paramètres. En effet, on peut trouver dans la littérature des valeurs pour les grandeurs physiques mises en jeu dans le modèle développé ici ($m_1, m_2, r_1, r_2, k_1, k_2, k_{12}, p_s, H, x_0$).

Nous pouvons donc maintenant envisager plusieurs stratégies d'optimisation pour les paramètres du modèle physique, en reprenant l'idée d'une minimisation de fonction d'objectif. Nous pouvons de plus choisir de réaliser l'optimisation sur tous les paramètres, ou en laisser certains fixes. La fonction d'objectif peut être simplement l'erreur quadratique entre signal synthétisé et signal cible, ou les mesures de qualité d'identification linéaire définies en F.3.2.2. Les contraintes ne sont par contre plus nécessaires puisque la régression de ψ en une fonction linéaire n'est pas possible.

Des premiers résultats sont présentés en figure 8.14. S'ils sont meilleurs qu'avec le modèle général proposé en section F.3.1, ils ne sont toujours pas satisfaisants. La périodicité est bien retrouvée mais les caractéristiques de la vibration sont en revanche mal estimées. L'explication la plus plausible au manque de précision de la re-synthèse est que la fonction non linéaire n'est finalement pas suffisamment adaptée pour approcher assez finement le signal cible.

Deux pistes d'amélioration sont : (i) introduire une excitation $r \neq 0$ additive stochastique en amont du filtre, maximiser la vraisemblance des données et inférer des raffinements de modèle en analysant ce signal, (ii) raffiner le modèle sur des considérations physiques en abandonnant la structure de Lur'e pour se rapprocher des modèles complets mis sous forme de systèmes passifs avec source.

8.4 Conclusion

Le modèle LF offre une représentation efficace à peu de paramètres. Un algorithme a été proposé pour estimer les paramètres du modèle LF permettant de représenter chaque période d'une fonction d'aire glottique mesurée sur un patient en phonation. Cet algorithme permet d'obtenir des résultats satisfaisants.

Un modèle chaotique de type système de Lur'e fondé sur des simplifications du modèle à deux masses des plis vocaux a été proposé dans le but d'obtenir un modèle de fonctions d'aire glottique plus diversifié. L'algorithme proposé par De Feo dans [De 01] a été adapté à notre cas, mais ne donne pas des résultats suffisants pour l'analyse-resynthèse de signaux d'aire glottique. Si cette étude ne permet pas aujourd'hui une application directe pour le problème de l'analyse-synthèse de la voix, elle pose les premiers éléments d'une recherche qui s'avère riche en perspectives pour modéliser la variabilité de la source glottique.

Dans la suite des travaux présentés ici, il sera donc fait appel au modèle LF lorsqu'un modèle paramétrique de fonction d'aire glottique sera nécessaire.

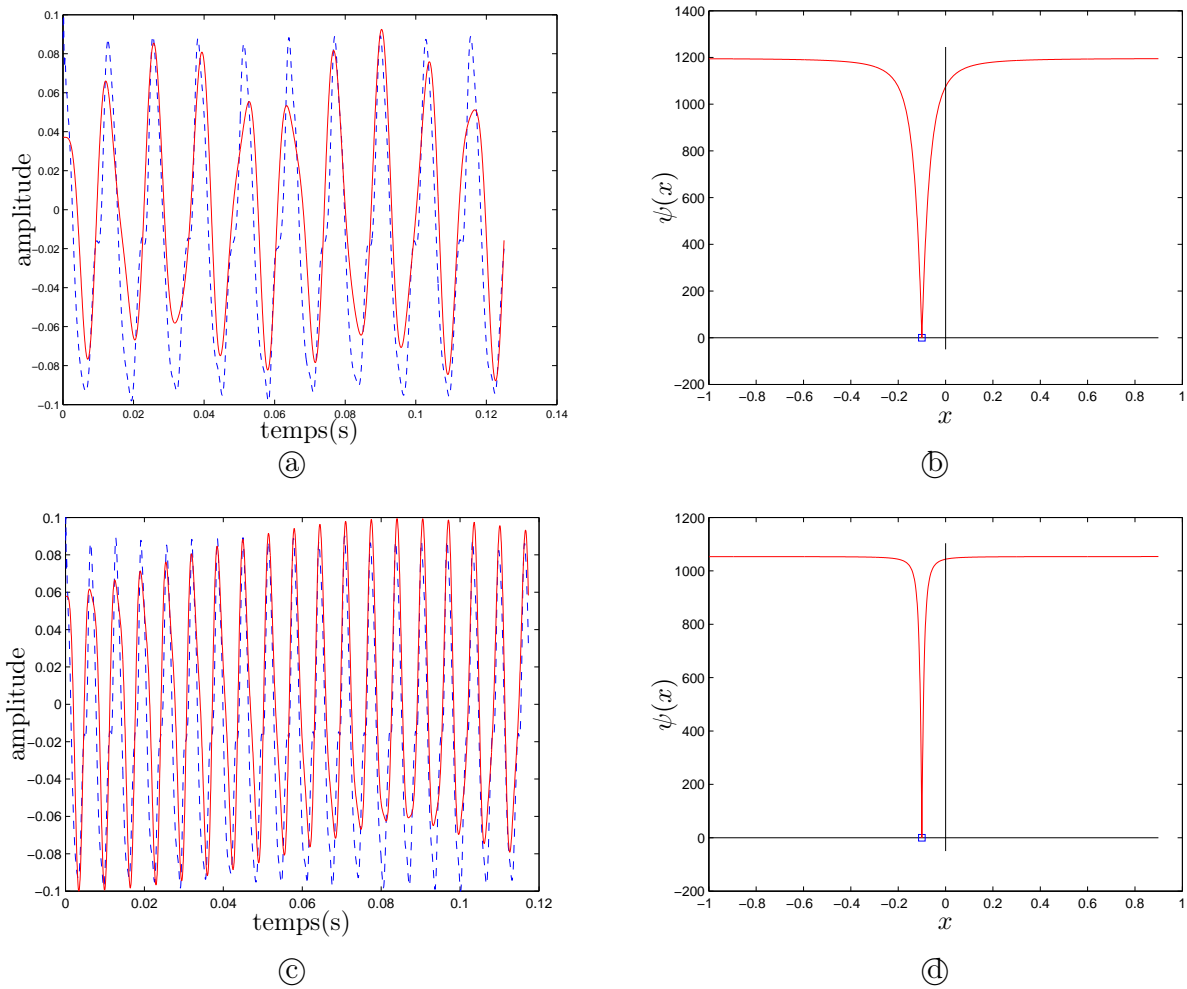


FIGURE 8.14 – Analyse-resynthèse de signaux d'aire glottique. (a),(c) signaux de synthèse (le signal synthétisé est en rouge, le signal cible est en bleu). (b),(d) fonctions non linéaires.



Chapitre 9

Modèles à géométrie forcée

Sommaire

9.1	Modèle couplé inspiré du modèle source-filtre	170
9.1.1	I1 Modèle de glotte	170
9.1.2	I2- I3 Modèle de conduit vocal et couplage	172
9.1.3	Modèle complet	173
9.1.4	Débouclage et discrétisation	173
9.1.5	Modèle de synthèse implémentable	174
9.2	Apport du couplage entre la source glottique et le conduit vocal	175
9.2.1	Influence de la pression sous-glottique	175
9.2.2	Influence du couplage glotte-conduit	178
9.3	Modélisation de conduit vocal à profil régulier en guides d'ondes acoustiques 1D	179
9.3.1	Formalisme guide d'ondes	179
9.3.2	Modèle de Webster-Lokshin dans les tubes à section variable	179
9.3.3	Profils à rayon positif	182
9.3.4	Diagonalisation de la matrice de transfert	184
9.3.5	Observations sur les nouveaux opérateurs de propagation et de dispersion	184
9.4	Conclusion et perspectives	188

Ce chapitre est consacré à l'étude des modèles dont la géométrie de la glotte et du conduit vocal est forcée (choix **C2**). Nous nous proposons ici, dans un premier temps, de développer un modèle simplifié, inspiré du modèle source-filtre mais intégrant le couplage entre la glotte et le conduit vocal (section 9.1), afin de caractériser l'apport de ce couplage (section 9.2). Les signaux d'aire glottique présentés dans le chapitre 8 pilotent la géométrie du modèle de glotte. Dans un deuxième temps, nous étudierons comment améliorer le modèle de conduit vocal avec le modèle de Webtser-Lokshin [Hé02] (section 9.3).

9.1 Modèle couplé inspiré du modèle source-filtre

Plusieurs modèles à géométrie forcée ont déjà été proposés, parmi lesquels nous pouvons citer les modèles de Rothenberg [Rot81b], d'Ananthapadmanabha et Fant [AF82], de Maeda [Mae82a] et d'Allen et Strong [AS85]. Ces modèles sont tous fondés sur le modèle de glotte de Van den Berg [vdBZD57] et un modèle électrique équivalent du système est construit pour la simulation. Le pilotage de l'aire d'ouverture glottique est fait par des fonctions *ad-hoc*.

Le modèle de production de la voix proposé ici, présenté dans la section 9.1, est une version simplifiée de ces modèles interactifs, afin de rester proche du modèle source-filtre, dans l'idée de la démarche présentée en section 2.2.3, et d'étudier l'intérêt de l'apport du couplage. Ce modèle est constitué de trois ingrédients que nous pouvons étudier séparément.

I1 Modèle de glotte, vibration glottique et écoulement à travers la glotte.

I2 Modèle de conduit vocal, acoustique du conduit vocal et rayonnement.

I3 Couplage glotte-conduit, qui lie entre l'écoulement à travers la glotte à l'acoustique dans le conduit.

Afin de rester proche du modèle source-filtre de la partie précédente, le choix est fait ici d'exploiter un modèle de conduit en concaténation de tubes droits, présenté dans la section 9.1.2. Le lien avec le modèle tous-pôles de filtre vocal est précisé. Le modèle de glotte proposé est un modèle à géométrie forcée à une dimension piloté par l'aire glottique. Je présente dans la section 9.1.1 un modèle simple de l'écoulement à travers la glotte, fondé sur l'équation de Bernoulli. L'interaction entre la glotte et l'acoustique du conduit vocal est assurée par la continuité de la pression et du débit acoustiques à l'interface entre les deux (section 9.1.2). Le modèle complet est représenté schématiquement sur la figure 9.1. Les variables de commande de ce modèle sont

- la pression sous-glottique $p_s(t)$,
- l'aire glottique $A_g(t)$,
- la géométrie du conduit vocal.

Les variables internes $u_g(t)$ et $p_g(t)$, le débit et la pression glottique sont calculées lors de la simulation du modèle et la variable de sortie est

- $s(t)$ le signal de synthèse.

9.1.1 I1 Modèle de glotte

Nous considérons un modèle géométrique simplifié des plis vocaux, qui se résume à un diaphragme dans un tube (axisymétrique), entièrement décrit par son aire d'ouverture, comme présenté sur la figure 9.2, où D_g représente le diamètre de l'ouverture glottique, p_s la pression sous-glottique, p_g la pression glottique, et p_c la pression à l'entrée du conduit vocal.

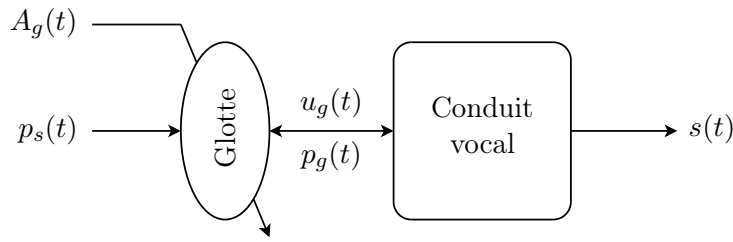


FIGURE 9.1 – Schéma général du modèle à géométrie forcée.

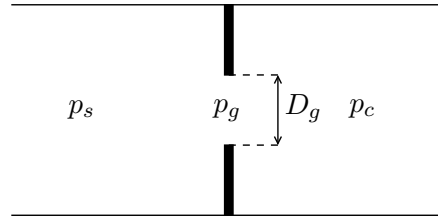


FIGURE 9.2 – Modèle de glotte à une dimension.

Les équations qui permettent de décrire un écoulement fluide sont fondées sur des principes de conservation. L'équation de Navier-Stokes traduit la conservation du moment d'une particule de fluide. Cette équation étant très complexe à résoudre analytiquement, nous faisons les approximations suivantes dans le cadre de notre étude.

- (H1) L'effet de la gravité peut être négligé sur l'écoulement dans la glotte.
- (H2) L'écoulement glottique est incompressible.
- (H3) L'écoulement glottique principal est isentropique.
- (H4) L'écoulement glottique est quasi-stationnaire.
- (H5) Les poumons sont une source de pression idéale (débit nul en sortie de la trachée).

Ces approximations correspondent à l'approche classique des modèles physiques de la production vocale [IF72, PHvH⁺94, LHVH98] et ont été validées par une étude des grandeurs caractéristiques de l'écoulement et des expérimentations in vitro [Vil02, Cis08].

Sous ces hypothèses, l'écoulement dans la glotte peut être décrit au moyen de l'équation de Bernoulli

$$p_g(t) + \frac{1}{2}\rho v_g(t)^2 = cste,$$

où $v_g(t)$ représente le flux glottique et ρ la densité de l'air. L'espace sous-glottique étant considéré comme une source idéale de pression (débit nul), nous pouvons donc écrire

$$p_s(t) = p_g(t) + \frac{1}{2}\rho v_g(t)^2. \quad (9.1)$$

De plus, le débit glottique $u_g(t)$ s'exprime

$$u_g(t) = A_g(t) \cdot v_g(t). \quad (9.2)$$

En suivant [Rod95], et en notant $\xi(t) = \text{sgn}(p_s(t) - p_g(t))$, nous déduisons de (9.1-9.2) le modèle de flux dans la glotte

$$u_g(t) = A_g(t)\xi(t)\sqrt{\frac{2}{\rho}|p_s(t) - p_g(t)|}. \quad (9.3)$$

9.1.2 I2-I3 Modèle de conduit vocal et couplage

Afin de rester proche du modèle source-filtre et du modèle tous-pôles de conduit vocal, nous choisissons pour ce modèle de production vocale à géométrie forcée le modèle de conduit en tronçons de tubes droits présenté en annexe G. Pour pouvoir établir un modèle complet et calculer l'impédance d'entrée (ramenée à la glotte) et la fonction de transfert acoustique du conduit vocal, il nous faut exprimer les conditions limites aux lèvres.

Le modèle de rayonnement choisi est un modèle idéalisé de rayonnement dans un demi-espace infini (pression nulle). Il s'écrit en reprenant les notations posées en section G.2.2,

$$\begin{cases} u_0^-(t + \tau) = u_0^+(t - \tau) \\ s(t) \propto \frac{\partial u}{\partial t}(L, t) = \frac{d}{dt} (u_0^+(t - \tau) + u_0^-(t + \tau)) = 2 \frac{du_0^+}{dt}(t - \tau) \end{cases} \quad (9.4)$$

Dans la suite, nous noterons $H_r(s) = k s$.

En combinant (G.17) et (9.4), nous pouvons calculer la fonction de transfert acoustique $H_t(s)$ et l'impédance d'entrée $Z_{in}(s)$ du conduit vocal, telles que $U(L, s) = H_t(s)U_g(s)$ et $P_g(s) = Z_{in}(s)U_g(s)$,

$$\begin{cases} H_t(s) = \frac{2c_{M-1} e^{-2M\tau s}}{H_{M-1}^+(s) + H_{M-1}^-(s)} \\ Z_{in}(s) = \frac{\rho c}{S_{M-1}} \frac{H_{M-1}^+(s) - H_{M-1}^-(s)}{H_{M-1}^+(s) + H_{M-1}^-(s)} \end{cases}, \quad (9.5)$$

avec $H_{M-1}^+(s) = H_{M-1}^{++}(s) + H_{M-1}^{-+}(s)$ et $H_{M-1}^-(s) = H_{M-1}^{+-}(s) + H_{M-1}^{--}(s)$ (cf annexe G.3).

Le couplage à la glotte repose sur la continuité de l'état acoustique (pression et débit) à l'entrée du conduit vocal

$$\begin{cases} p_G(t) = p(0, t) \\ u_G(t) = u(0, t) \end{cases}. \quad (9.6)$$

Remarque 8. On montre aisément par récurrence à partir de (G.14) que la fonction de transfert acoustique $H_t(s)$ s'écrit

$$H_t(s) = \frac{c_M e^{-2M\tau s}}{\sum_{i=0}^M a_i e^{-4i\tau s}}. \quad (9.7)$$

En échantillonnant ce système avec une période d'échantillonnage de $T_s = 4\tau$, (9.7) s'écrit dans le domaine en Z

$$H_t(z) = \frac{c_M z^{-\frac{M}{2}}}{\sum_{i=0}^M a_i z^{-i}}.$$

Ce résultat, déjà présenté dans [MG76], présente donc une motivation acoustique du modèle tous-pôles du conduit vocal étudié dans la partie II.

9.1.3 Modèle complet

Le modèle de synthèse est obtenu en combinant (9.3), (9.4), (9.5) et (9.6). La figure 9.3 présente le modèle complet sous forme de schéma fonctionnel. Les variables d'entrées du modèle sont la pression sous-glottique $p_s(t)$, l'aire glottique $A_g(t)$ et la géométrie permettant de calculer $H_t(s)$ et $Z_{in}(s)$. La variable de sortie est le son synthétisé $s(t)$.

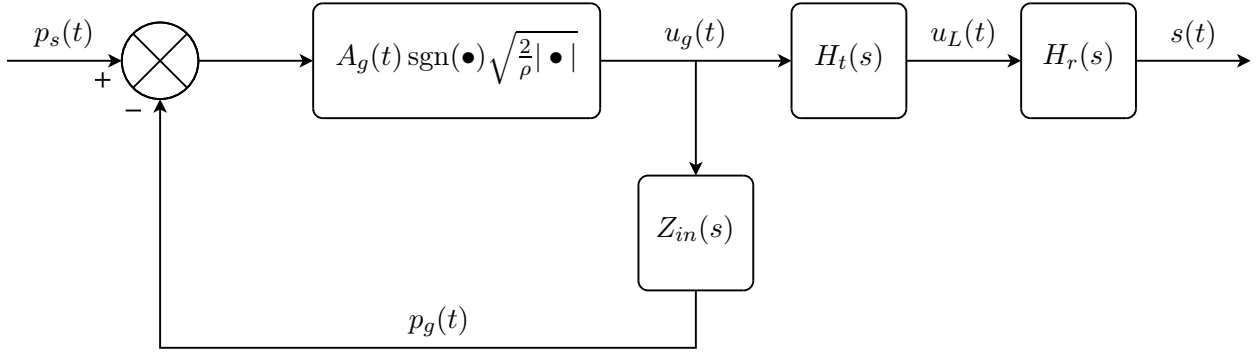


FIGURE 9.3 – Schéma fonctionnel du modèle couplé en temps continu.

9.1.4 Débouclage et discrétisation

Le calcul de $Z_{in}(s)$ fait apparaître un terme de réflexion instantanée égale à $\frac{\rho c}{S_{M-1}}$ (quelque soit la géométrie du conduit) ainsi qu'une somme de retards purs multiples de 4τ à différentes amplitudes. Autrement dit, nous pouvons mettre $Z_{in}(s)$ sous la forme

$$\begin{aligned} Z_{in}(s) &= \frac{\rho c}{S_{M-1}} + \tilde{Z}_{in}(s) e^{-4\tau s}, \text{ où} \\ \tilde{Z}_{in}(s) &= -\frac{\rho c}{S_{M-1}} \frac{2e^{4\tau s} H_{M-1}^-(s)}{H_{M-1}^+(s) + H_{M-1}^-(s)} \end{aligned} \quad (9.8)$$

ne contient que des réponses instantanées et retardées (mais pas de réponses avancées, le terme $e^{4\tau s}$ est compensé lors du calcul de $\tilde{Z}_{in}(s)$).

Le modèle de la figure 9.3 ne peut être implémenté sous forme numérique à cause de la réponse instantanée de $Z_{in}(s)$. Afin d'obtenir un système implémentable, il est nécessaire de trouver un système équivalent ne présentant pas de réponse instantanée dans la boucle de rétro-action. Pour résoudre la boucle instantanée, nous définissons

$$\tilde{P}_g(s) = \tilde{Z}_{in}(s) e^{-4\tau s} U_g(s)$$

tel que

$$P_g(s) = Z_{in}(s) U_g(s) = \frac{\rho c}{S_{M-1}} U_g(s) + \tilde{P}_g(s).$$

On montre alors que

$$\begin{cases} \xi(t) \geq 0 \Leftrightarrow p_s(t) - p_g(t) \geq 0 \Leftrightarrow p_s(t) - \tilde{p}_g(t) \geq \frac{\rho c}{S_{M-1}} u_g(t) \geq 0 \\ \xi(t) \leq 0 \Leftrightarrow p_s(t) - p_g(t) \leq 0 \Leftrightarrow p_s(t) - \tilde{p}_g(t) \leq \frac{\rho c}{S_{M-1}} u_g(t) \leq 0 \end{cases},$$

d'où

$$\xi(t) = \operatorname{sgn}(p_s(t) - p_g(t)) = \operatorname{sgn}(u_g(t)) = \operatorname{sgn}(p_s(t) - \tilde{p}_g(t)). \quad (9.9)$$

Nous obtenons alors

$$\begin{aligned} u_g(t) &= A_g(t)\xi(t)\sqrt{\frac{2}{\rho}|p_s(t) - p_g(t)|} \\ &= A_g(t)\xi(t)\sqrt{\frac{2}{\rho}\left|p_s(t) - \frac{\rho c}{S_{M-1}}u_g(s) + \tilde{p}_g(s)\right|} \end{aligned}$$

dont nous déduisons l'équation en $u_g(t)$

$$u_g(t)^2 + 2c\frac{A_g(t)^2}{S_{M-1}}\xi(t)u_g(t) - \frac{2}{\rho}A_g(t)^2\xi(t)(p_s(t) - \tilde{p}_g(t)) = 0 \quad (9.10)$$

D'après (9.9), on montre aisément que la seule et unique solution de (9.10) est

$$u_g(t) = A_g(t)F(p_s(t) - \tilde{p}_g(t)) \quad \text{avec} \quad (9.11)$$

$$F(x) = \operatorname{sgn}(x) \left(\sqrt{c^2\frac{A_g(t)^2}{S_{M-1}^2} + \frac{2}{\rho}|x|} - c\frac{A_g(t)}{S_{M-1}} \right). \quad (9.12)$$

La discrétisation est immédiate en prenant un période d'échantillonnage égale à 4τ . Le réglage de la fréquence d'échantillonnage du système discret peut alors être effectué en adaptant l'échantillonnage du profil du conduit.

9.1.5 Modèle de synthèse implémentable

Le modèle de la figure 9.3 devient, après débouclage et discrétisation, celui de la figure 9.4, avec

$$\begin{aligned} H_t(z) &= \frac{2c_{M-1}z^{-\frac{M}{2}}}{H_{M-1}^+(z) + H_{M-1}^-(z)}, \\ \tilde{Z}_{in}(z) &= -\frac{\rho c}{S_{M-1}} \frac{2zH_{M-1}^-(z)}{H_{M-1}^+(z) + H_{M-1}^-(z)}, \\ H_r(z) &= k(1 - \eta z^{-1}), \quad \eta \approx 1^-, \\ \begin{bmatrix} H_m^+(z) \\ H_m^-(z) \end{bmatrix} &= \prod_{i=m}^1 \begin{pmatrix} 1 & -\mu_i \\ -\mu_i z^{-1} & z^{-1} \end{pmatrix} \begin{bmatrix} 1 \\ z^{-1} \end{bmatrix}, \\ c_m &= \prod_{i=1}^m (1 - \mu_i), \\ \mu_i &= \frac{S_i - S_{i-1}}{S_i + S_{i-1}}. \end{aligned}$$

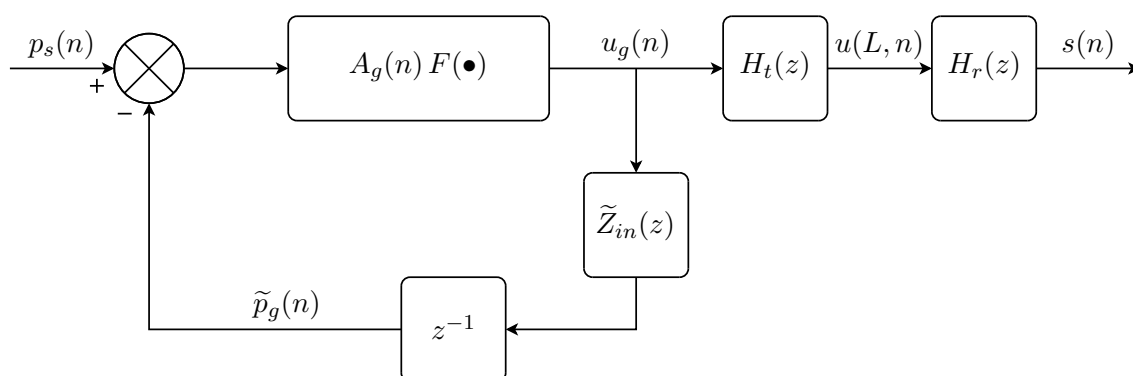


FIGURE 9.4 – Schéma fonctionnel du modèle couplé en temps discret.

9.2 Apport du couplage entre la source glottique et le conduit vocal

9.2.1 Influence de la pression sous-glottique

Une des entrées du modèle proposé est la pression sous-glottique. Afin de synthétiser des sons de voix, il est nécessaire de connaître un ordre de grandeur des valeurs de pression sous-glottique mises en jeu lors de la phonation. Comme il l'a été mentionné dans le chapitre 1, des méthodes permettent de mesurer et d'estimer la pression sous-glottique sur des sujets pendant la phonation [KL75, SH81, PH98]. D'autres méthodes permettent d'estimer la pression sous-glottique à partir de mesures moins invasives (voir [HHP88, BSA06] par exemple).

Cette littérature nous enseigne que la pression sous-glottique est de l'ordre de 5 – 10 cm H₂O pour les voix parlées en condition de phonation habituelle (confortable) [Hir81]. Elle peut descendre jusqu'à 1 cm H₂O en voix murmurée et s'élever peut s'élever jusqu'à des valeurs de 30 ou 40 cm H₂O, notamment pour les voix chantées et criées.

Afin d'étudier l'influence de la pression sous-glottique, cinq sons sont synthétisés avec une fonction d'aire glottique et une géométrie de conduit vocal similaires. La fonction d'aire glottique provient de la base de données USC_2008_02 et correspond à une voyelle \ə\. Le conduit vocal utilisé correspond aussi à une voyelle \a\ et provient de [ST98]. Seule la pression sous-glottique (constante) est modifiée et varie dans la plage indiquée ci-dessus : 1 cm H₂O, 5 cm H₂O, 10 cm H₂O, 20 cm H₂O, 40 cm H₂O. Les spectres normalisés des sons synthétisés sont présentés en figure 9.5. Nous pouvons constater dans cette figure un net renforcement des deux harmoniques autour de 600 Hz. Cette valeur correspond précisément au premier formant du conduit vocal utilisé, confirmant le couplage fort qui existe entre la glotte et le premier formant dans le cas de la voyelle \a\ [AS85].

Cette simulation montre que la prise en compte d'un modèle de glotte simple (diaphragme et équation de Bernoulli) qui amène à la relation non linéaire (9.3) entre la pression et le débit glottique p_g et u_g permet de rétablir les modifications spectrales du son rayonné lorsque la pression sous-glottique augmente ou diminue. Le cas de la voyelle \a\ met particulièrement en avant ce phénomène, grâce à une interaction forte avec le premier formant. Un deuxième exemple est présenté en figure 9.6, dans le cas de voyelle \u\. Une évolution est présente mais moins interprétable.

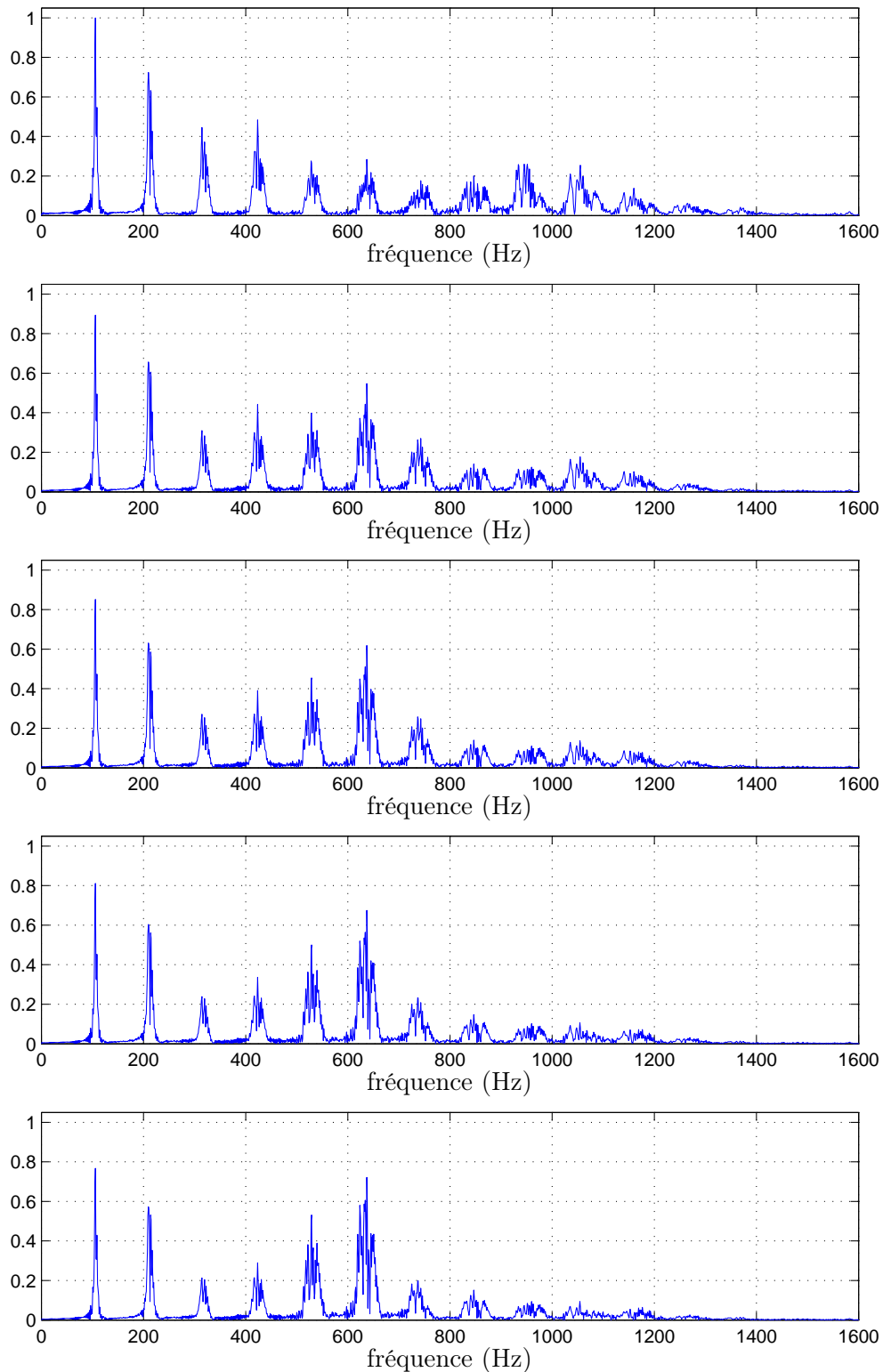


FIGURE 9.5 – Influence de la pression sous-glottique sur le spectre du son rayonné, voyelle $\backslash a \backslash$. De haut en bas : $p_s = 1 \text{ cm H}_2\text{O}$, $5 \text{ cm H}_2\text{O}$, $10 \text{ cm H}_2\text{O}$, $20 \text{ cm H}_2\text{O}$, $40 \text{ cm H}_2\text{O}$. La fonction d'aire glottique et le conduit vocal sont les mêmes. Les spectres sont normalisés en puissance. On constate un renforcement des fréquences autour de 600 Hz.

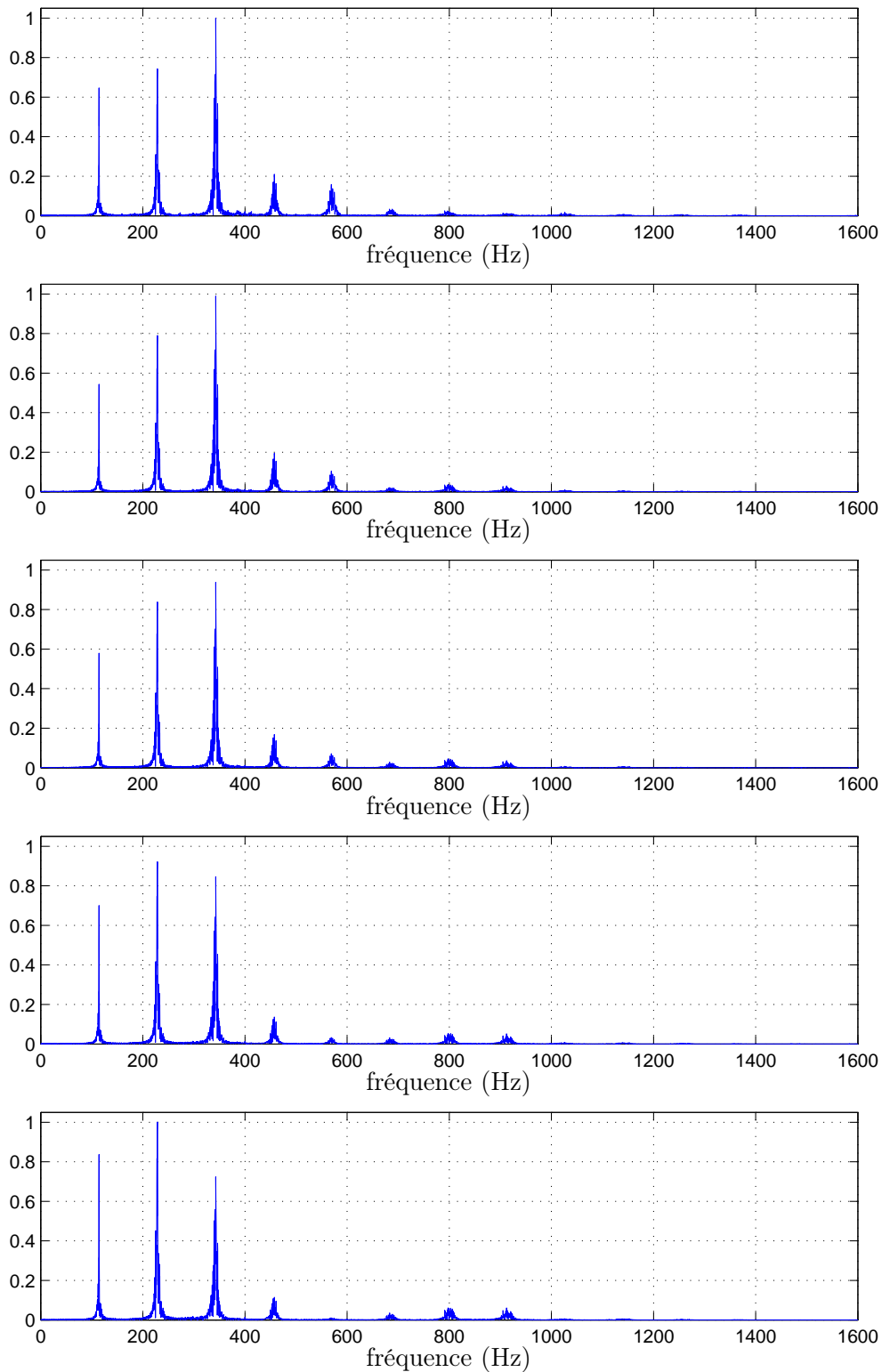


FIGURE 9.6 – Influence de la pression sous-glottique sur le spectre du son rayonné, voyelle $\backslash u \backslash$. De haut en bas : $p_s = 1 \text{ cm H}_2\text{O}$, $5 \text{ cm H}_2\text{O}$, $10 \text{ cm H}_2\text{O}$, $20 \text{ cm H}_2\text{O}$, $40 \text{ cm H}_2\text{O}$. La fonction d'aire glottique et le conduit vocal sont les mêmes. Les spectres sont normalisés en puissance.

9.2.2 Influence du couplage glotte-conduit

Un autre simulation permet de mettre en évidence l'influence du couplage entre la glotte et le conduit vocal sur l'acoustique glottique. Trois sons sont synthétisés avec la même fonction d'aire glottique (extraite du corpus) et la même pression sous-glottique (10 cm H₂O). Les trois sons diffèrent par les conduits vocaux utilisés, correspondant aux voyelles \u\, \o\, et \i\ (extraits de [ST98]). Il apparaît sur ces figures que la charge du conduit vocal peut éloigner significativement le débit glottique de la fonction d'aire glottique. Ce résultat, déjà présenté dans les nombreux travaux sur les modèles couplés à géométrie forcée [Rot81b, AF82, Mae82a, AS85] et auto-oscillants [IF72, Vil02, Rut07], est retrouvé ici avec un modèle simplifié. L'effet peut être particulièrement marqué dans nos simulations, comme dans le cas du \i\ présenté ici. Ceci peut trouver une explication dans l'interaction exacerbée constatée avec les modèles de glotte monodimensionnels (modèle auto-oscillant à une masse notamment [FL68, Mae82a]).

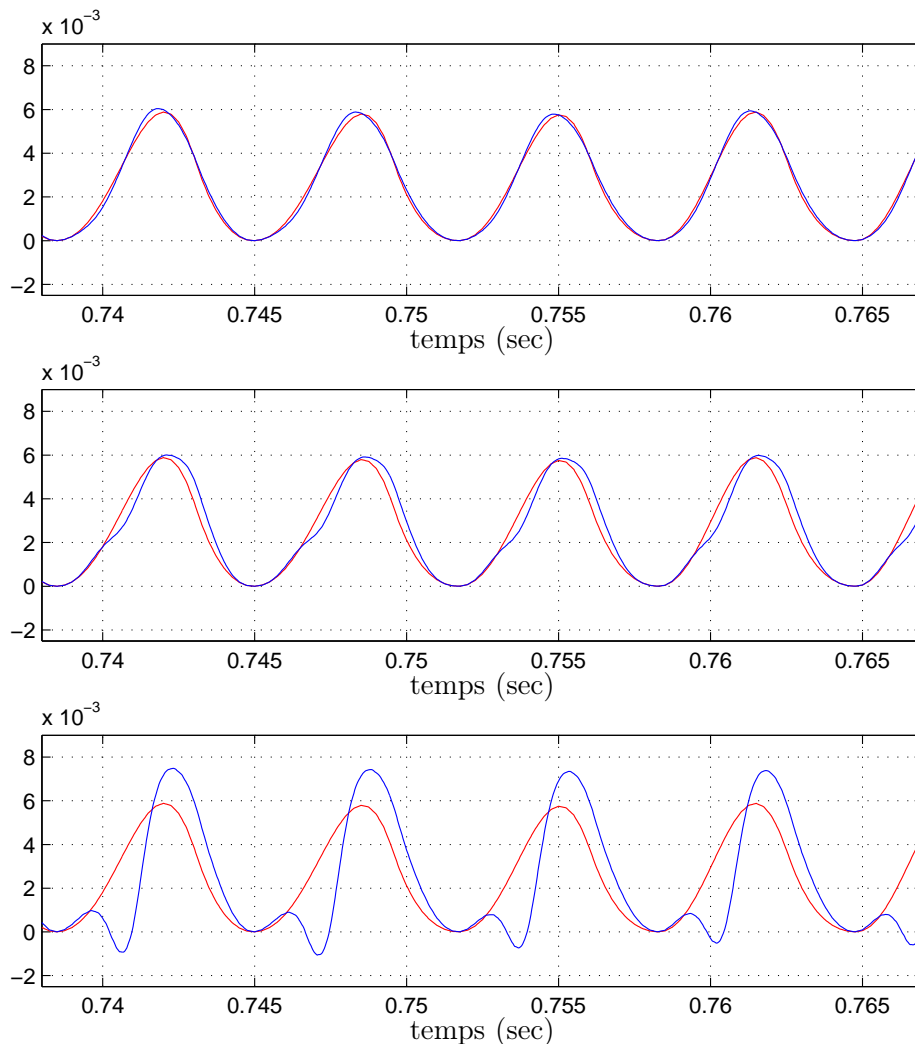


FIGURE 9.7 – Influence du couplage, comparaison de l'aire glottique (-) et du débit glottique (-). Les fonctions d'aire glottique et les pressions sous-glottiques sont les mêmes, les conduits vocaux correspondent, de haut en bas, aux voyelles \u\, \o\, et \i\.

9.3 Modélisation de conduit vocal à profil régulier en guides d'ondes acoustiques 1D

Je propose ici de repartir de mes résultats de stage M2R ATIAM [Hé10] pour améliorer le modèle de conduit vocal. Ces résultats s'appuient sur le modèle de Webster-Lokshin [Hé02] et l'intérêt de ce modèle comparé aux guides d'ondes en tubes droits ou tronçons de cônes est brièvement présenté. Malgré les avantages de ces raffinements, le cas (typique du conduit vocal) des profils concaves a longtemps posé problème pour la simulation en guides d'ondes numériques (cf [Mig09]). Je propose une résolution en ondes découplées progressives fondée sur des considérations physiques.

9.3.1 Formalisme guide d'ondes

L'annexe G présente le modèle général de propagation des ondes acoustiques dans les tubes acoustiques, mis en équation sous forme de l'équation des pavillons, ou équation de Webster [Web19]. Une solution pour la simulation faible coût de la propagation dans les tubes à section variable (non cylindrique) est celle que nous avons exploitée dans le modèle qui vient d'être présenté. Elle consiste à considérer le tube comme une concaténation de tubes droits. Ceci peut mener à la construction d'une impédance d'entrée et d'une fonction de transfert sous la forme de filtre numériques comme nous l'avons vu, ou bien à la concaténation de quadripôles simples, appelées structure de Kelly-Lochbaum [KL62] (figure 9.8), et de retards purs, dans le cadre d'une simulation de type guides d'ondes numériques [Fet86, Smi87, KB98]. Ce formalisme est particulièrement adapté à la simulation numérique. Välimäki propose en 1995 de prendre en compte une concaténation de tubes coniques plutôt que cylindriques, en supposant la sphéricité des ondes se propageant dans le tube afin de rétablir la continuité du profil du tube [Vä95]. Ce formalisme conduit aussi à des structures de Kelly-Lochbaum à la jonction entre deux tronçons de cônes (voir figure 9.8).

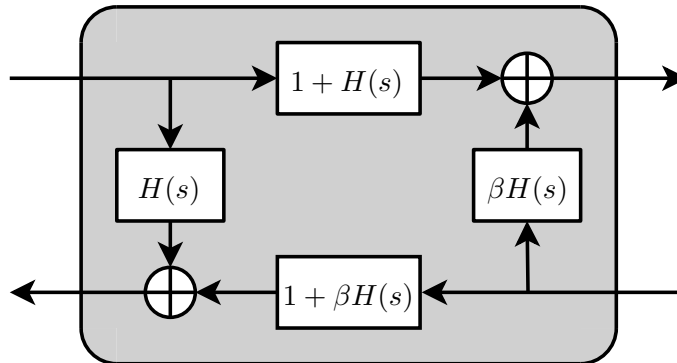


FIGURE 9.8 – Structure de Kelly-Lochbaum à la jonction entre deux tubes acoustiques. Cas des tubes droits : $H(s)$ est une constante (coefficients de réflexion k) et $\beta = -1$. Cas des tubes coniques : $H(s)$ est un filtre passe-bas et $\beta = 1$. D'après [Mig09].

9.3.2 Modèle de Webster-Lokshin dans les tubes à section variable

La propagation dans un tuyau est soumise à des phénomènes de pertes non pris en compte dans l'équation de Webster. Les trois mécanismes d'atténuation les plus importants dans le cas de la voix sont les effets de viscosité et de conduction thermique aux parois ainsi que celui des

parois vibrantes. La formule générale des pertes visco-thermiques pour un cylindre à symétrie de révolution est proposée par Kirchhoff dans [Kir68]. Plusieurs modèles simplifiés ont été proposés, notamment la théorie de Zwicker et Kosten [ZK49] et l'admittance équivalente de Cremer [Cre48]. La prise en compte des pertes visco-thermiques permet de raffiner le modèle et de rétablir les effets d'amortissement qui permettent un gain significatif de réalisme dans la synthèse sonore.

Un modèle mono-dimensionnel permettant de prendre en compte les pertes visco-thermiques par admittance équivalente de Cremer et de réduire les hypothèses sur les ondes qui se propagent dans le tube est le modèle de Webster-Lokshin [Hé02] qui s'exprime sous la forme d'une équation différentielle à dépendance en l'abscisse curviligne ℓ liant le profil du tube acoustique $R(\ell)$ et la distribution de pression $P(\ell, t)$:

$$\left(\partial_\ell^2 + 2 \frac{R'(\ell)}{R(\ell)} \partial_\ell - \frac{1}{c_0^2} \partial_t^2 - \frac{2\varepsilon(\ell)}{c_0^{\frac{3}{2}}} \partial_t^{\frac{3}{2}} \right) P(\ell, t) = 0, \quad (9.13)$$

où $\varepsilon(\ell) = \kappa_0 \frac{\sqrt{1-R'(\ell)^2}}{R(\ell)}$ quantifie les effets visco-thermiques ($\kappa_0 \approx 3 \times 10^{-4} m^{1/2}$ dans l'air) et c_0 est la vitesse du son.

Remarque 9. *L'impédance de Cremer peut être remplacée dans ce modèle par une autre impédance dissipative plus pertinente pour le conduit vocal (parois vibrantes, tissus mous) dont la passivité permet de conserver un problème bien posé, comme présenté dans [Hé03, section V].*

Ce modèle peut être résolu analytiquement dans le cas où $\Upsilon(\ell) = \frac{R''(\ell)}{R(\ell)}$ et $\varepsilon(\ell)$ sont constants. La matrice de transfert liant l'état acoustique (pression, vitesse) à l'entrée et à la sortie d'un tronçon de tube à Υ constant est donnée dans [Hé02, p. 63], elle est réécrite pour l'état acoustique (pression, débit) dans [Hé10, p. 21]. Mignot propose dans [Mig09] une méthode pour adapter le formalisme des guides d'ondes numériques et de la structure de Kelly-Lochbaum pour ce modèle dans le cas où $\Upsilon > 0$ (tubes convexes). La méthode est fondée sur la factorisation de l'équation (9.13) (de façon identique à la démarche présentée dans l'annexe G.2.1 pour les tubes droits sans pertes) et aboutit à la définition d'ondes découplées dont l'opérateur de propagation s'écrit, pour un tronçon de longueur L ,

$$\begin{aligned} T(s) &= e^{-L\Gamma(s)}, \text{ avec} \\ \Gamma^2(s) &= \frac{1}{c_0^2} s^2 + \frac{2\varepsilon}{c_0^{\frac{3}{2}}} + \Upsilon. \end{aligned} \quad (9.14)$$

Cette méthode conduit finalement à un outil modulaire permettant de construire un conduit par concaténation de portions de tubes droits, coniques ou à Υ constant, en prenant en compte d'éventuelles ruptures de pente et de rayon à l'interface entre les tronçons. Une stratégie de numérisation¹ des systèmes obtenus est proposée et conduit à des systèmes simulables en temps réels [Mig09, Hé09].

Contrairement à la concaténation de tubes droits ou coniques pour lesquels l'approximation géométrique d'un profil donné peut être faite aisément en discrétisant le profil, la représentation d'un tube quelconque par une concaténation de tronçons de tubes à Υ constant est non triviale. Lors de travaux précédents [Hé09], j'ai proposé un outil pour l'estimation automatique des paramètres géométriques de la concaténation de tronçons de tube à Υ constant (positif ou négatif) permettant de représenter au mieux un tube cible. Ainsi il est possible de construire un modèle numérique

1. La numérisation est non triviale car l'opérateur $\Gamma(s)$ fait intervenir des fonctions comportant des coupures qu'il est nécessaire d'approximer par des ensembles finis de pôles ou zéros pour la simulation numérique.

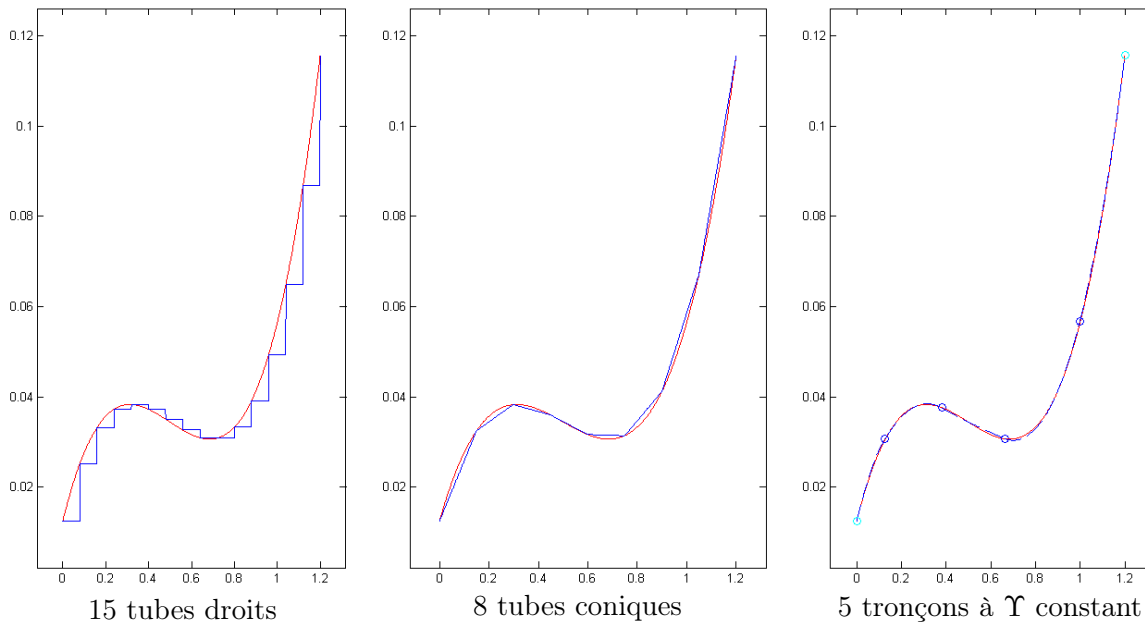


FIGURE 9.9 – Représentation d'un profil arbitraire en concaténation de tronçons.

de la propagation acoustique dans un tube quelconque par le modèle de Webster-Lokshin. De plus, l'outil de représentation géométrique proposé permet d'imposer la régularité \mathcal{C}^2 du profil du conduit virtuel, ce qui permet de régulariser la réponse (une régularité \mathcal{C}^k du profil implique une régularité \mathcal{C}^{k-1} de la réponse impulsionnelle du conduit [HHMM13]). La figure 9.9 illustre l'intérêt de cette représentation géométrique et de l'outil d'estimation des paramètres géométriques, les tronçons à Υ constant permettent une représentation fidèle de la géométrie avec un nombre réduit de tronçons.

L'intérêt du modèle de Webster-Lokshin a été démontré dans le cas d'un pavillon de trombone dont le profil et l'impédance d'entrée ont pu être mesurés. Le profil est représenté par une concaténation de cinq tronçons de tubes à Υ constant dont les paramètres géométriques ont été calculés avec mon algorithme. La figure 9.10 présente les impédances mesurée et simulée pour ce profil (extrait de [HHMM13]).

Les structures de simulation numérique proposées dans [Mig09] sont malheureusement instables dans le cas des tubes à $\Upsilon < 0$ (tubes concaves) car elles impliquent des tubes qui ne sont pas physiquement réalistes (ils sont périodiquement négatifs) lors de la décomposition en quadripôles. Ces tubes non physiques ont pour conséquence de faire apparaître des singularités dans la demi-plan droit de Laplace pour l'opérateur de propagation $\Gamma(s)$, menant à des instabilités. Si cela n'est pas très pénalisant pour la famille des cuivres (les pavillons des cuivres sont rarement concaves), cela rend par contre ces outils inutilisables pour la simulation des conduits vocaux qui ont généralement des portions concaves. Je propose ici une solution permettant d'obtenir des structures de guides d'ondes à temps continu pour le modèle de Webster-Lokshin garantissant le réalisme physique des tubes considérés dans la décomposition en ondes découplées quelque soit la valeur de Υ . La démarche adoptée se déroule en deux étapes,

1. poser des conditions pour garantir que les tubes considérés sont physiquement réalistes (de longueur finie et à rayon positif) en opérant un changement de variable dans la paramétrisation de la géométrie garantissant la positivité du rayon $R'(\ell)$, puis
2. diagonaliser la matrice de transfert acoustique.

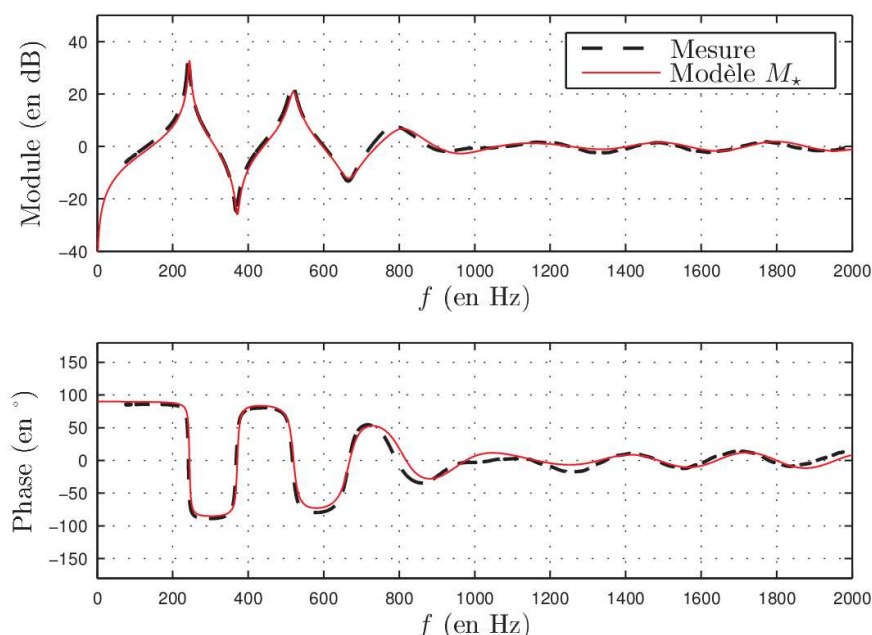


FIGURE 9.10 – Impédance d'entrée d'un pavillon de trombone mesurée et simulée. Le simulation repose sur une concaténation de cinq tronçons à Υ constant, le modèle de propagation de Webster-Lokshin et un modèle de rayonnement en portion de sphère pulsante. Extrait de [HHMM13].

9.3.3 Profils à rayon positif

Considérons maintenant un tronçon de tube de longueur L sur l'intervalle $[-\frac{L}{2}, \frac{L}{2}]$. Comme nous l'avons dit, le modèle (9.13) peut être résolu analytiquement dans le cas où les paramètres $\Upsilon(\ell)$ et $\varepsilon(\ell)$ sont constants. Pour le paramètre ε , on fait le choix $\varepsilon(\ell) \approx \frac{1}{L} \int_{-L/2}^{L/2} \varepsilon(\ell) d\ell$. La constance du paramètre $\Upsilon(\ell)$ induit l'équation

$$R''(\ell) - \Upsilon R(\ell) = 0. \quad (9.15)$$

Les solutions de cette équation s'écrivent

$$R(\ell) = R_l S^l(\ell, \Upsilon) + R_r S^r(\ell, \Upsilon), \quad (9.16)$$

avec

$$\begin{aligned} S^l &: (\ell, \Upsilon) \mapsto -\frac{\sinh\left(\sqrt{L^2\Upsilon}\left(\frac{\ell}{L} - \frac{1}{2}\right)\right)}{\sinh\left(\sqrt{L^2\Upsilon}\right)} \text{ et} \\ S^r &: (\ell, \Upsilon) \mapsto \frac{\sinh\left(\sqrt{L^2\Upsilon}\left(\frac{\ell}{L} + \frac{1}{2}\right)\right)}{\sinh\left(\sqrt{L^2\Upsilon}\right)}, \end{aligned} \quad (9.17)$$

et où R_l est le rayon à l'extrémité gauche du tube ($\ell = -L/2$) et R_r le rayon à l'extrémité droite ($\ell = L/2$). On monte alors [Hé10, p. 18 et suivantes] qu'un profil de la forme (9.16) est positif sur l'intervalle $[-\frac{L}{2}, \frac{L}{2}]$ si et seulement si

$$L^2\Upsilon > -\pi^2 \text{ et } \begin{cases} R_l \geq 0 \\ R_r \geq 0 \end{cases}. \quad (9.18)$$

Nous définissons alors le coefficient d'asymétrie θ et le rayon moyen ρ

$$\begin{cases} \theta &= \ln\left(\frac{R_r}{R_l}\right) \\ \rho &= \sqrt{R_r R_l} \end{cases}, \quad (9.19)$$

de sorte que $R_l = \rho e^{-\frac{\theta}{2}}$ et $R_r = \rho e^{\frac{\theta}{2}}$. Cette paramétrisation du profil du tube $R(\ell)$ permet d'établir la catégorisation présentée dans le tableau 9.1 et la figure 9.11.

$L^2\Upsilon$	$-\pi^2$	0	$+\infty$
$\theta^2 < \Upsilon$	×	×	Géométrie de type cosh
$\theta^2 = \Upsilon$	×	Tube droit	Géométrie de type exponentiel
$\theta^2 > \Upsilon$	Tube concave	Tube conique	Géométrie de type sinh

TABLE 9.1 – Catégorisation géométrique des profils en fonction de θ et Υ .

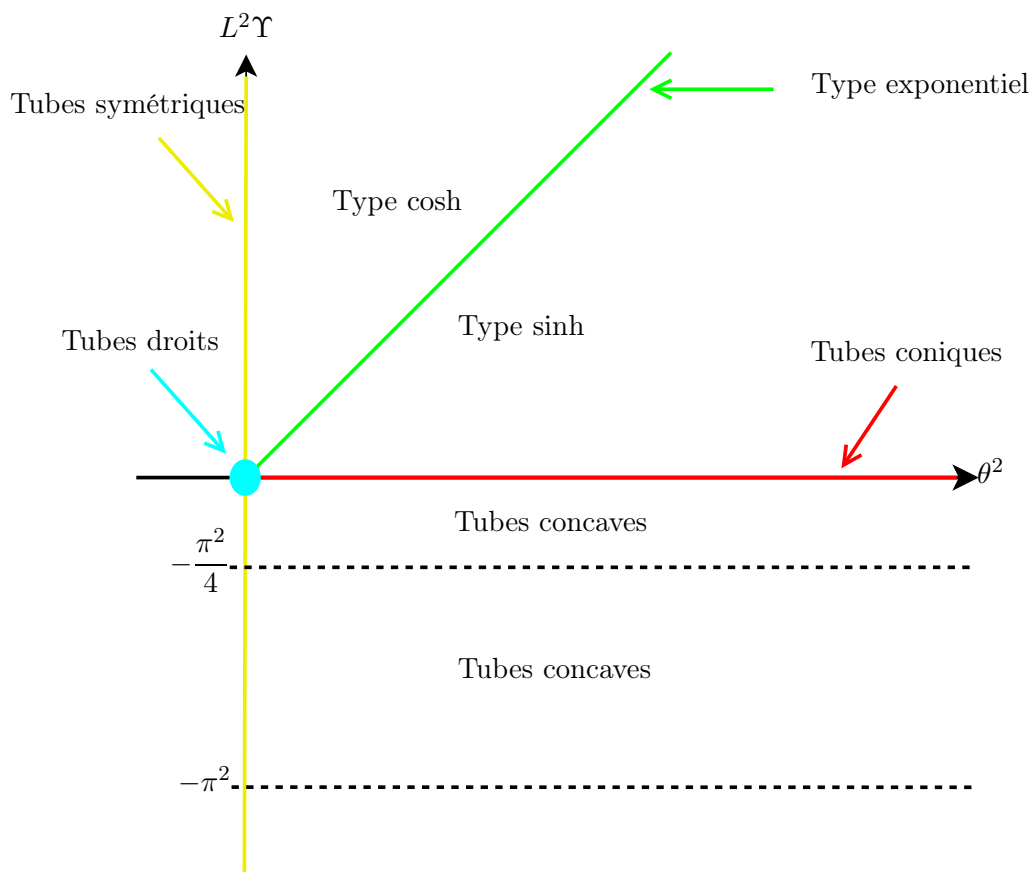


FIGURE 9.11 – Catégorisation géométrique des profils en fonction de θ et Υ .

9.3.4 Diagonalisation de la matrice de transfert

On montre que la matrice de transfert acoustique d'un tube à Υ constant possède un déterminant unitaire et son polynôme caractéristique s'écrit [Hé10, HHDM12] (la dépendance en la variable de Laplace s n'est pas représentée ici pour alléger l'écriture) :

$$P(\lambda) = \lambda^2 - 2 \cosh(L\tilde{\Gamma})\lambda + 1, \quad (9.20)$$

avec

$$\cosh(L\tilde{\Gamma}) = \cosh(L\Gamma) + \frac{\sqrt{L^2\Upsilon} [\cosh(\theta) - \cosh(\sqrt{L^2\Upsilon})]}{\sinh(\sqrt{L^2\Upsilon})} \frac{\sinh(L\Gamma)}{L\Gamma}. \quad (9.21)$$

Les valeurs propres de la matrice de passage, opérateurs de propagation de nos nouvelles ondes découplées, s'écrivent finalement

$$\tilde{T}^+(s) = e^{-L\tilde{\Gamma}(s)} \text{ et } \tilde{T}^-(s) = e^{+L\tilde{\Gamma}(s)}. \quad (9.22)$$

Le signe $+$ dans l'exponentielle de l'opérateur $\tilde{T}^-(s)$ s'explique par le fait qu'il caractérise le trajet de l'onde retour dans le sens opposé à son sens de propagation. Une écriture quadripolaire prenant en entrée l'onde aller à gauche et l'onde retour à droite, et en sortie l'onde aller à droite et l'onde retour à gauche, fait bien apparaître une matrice diagonale dont les deux éléments sont l'opérateur $\tilde{T}^+(s)$, que nous noterons dans la suite simplement $\tilde{T}(s)$, comme présenté dans la figure 9.12.

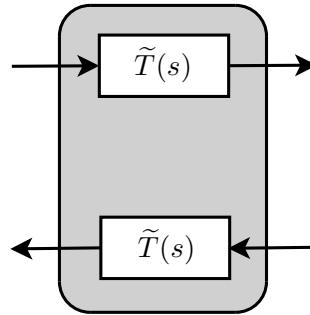


FIGURE 9.12 – Structure de propagation des ondes découplées stables.

9.3.5 Observations sur les nouveaux opérateurs de propagation et de dispersion

Nous pouvons remarquer deux propriétés intéressantes de cet opérateur de propagation.

- L'opérateur de propagation permet de retrouver les opérateurs précédemment proposés dans la littérature pour les cas des tubes droits et coniques ($\Upsilon=0$). Notamment le cas des tubes droits sans pertes ($\varepsilon=0$, $\Upsilon=0$ et $\theta=0$) permet de retrouver un retard pur ($\tilde{T}(s)=e^{-Ls/c_0}$).
- Dans les cas des tubes droits et des profils exponentiels, $\tilde{\Gamma}(s)=\Gamma(s)$, nous retrouvons l'opérateur obtenu par factorisation de l'équation (9.13).

Nous appelons opérateur de dispersion $D(s)$ l'opérateur de propagation privé du terme de retard pur,

$$\begin{cases} D(s) &= e^s T(s) &= e^{-L\Gamma(s)+s} \\ \tilde{D}(s) &= e^s \tilde{T}(s) &= e^{-L\tilde{\Gamma}(s)+s} \end{cases} \quad (9.23)$$

Les figures 9.13 et 9.14 présentent des tracés des modules et phase de $D(s)$ et $\tilde{D}(s)$ dans le plan de Laplace pour différentes valeurs de Υ (tubes convexes, coniques et concaves) et ε (sans et avec pertes). La valeur de θ est ici fixée à $\ln(2)$. Des figures complémentaires sont présentées dans [Hé10].

L'étude de ces figures nous amène aux considérations suivantes.

1. L'absence de déroulement de phase dans le demi-plan droit de Laplace pour les tracés de la phase de $\tilde{D}(s)$ nous prouve qu'il s'agit bien d'un opérateur de dispersion ne contenant pas de terme de retard.
2. Les opérateurs de dispersion $\tilde{D}(s)$ ont toujours un module inférieur à 1 et ne présente jamais de coupure dans le demi-plan droit de Laplace sur leur tracé de phase, contrairement aux opérateurs $D(s)$. Ces nouveaux opérateurs de dispersion nous assure donc la stabilité dans le transport des ondes découplées nouvellement définies.
3. On remarque une structure de modes non transportés (module plus petit que 1) pour les tronçons à $\Upsilon \neq 0$. Ces modes, dûs probablement à la finitude du tronçon nous assurent que l'information géométrique (finitude du tronçon et positivité du rayon) a bien été transmise dans la matrice de transfert acoustique.

La figure 9.15 présente l'évolution des diagrammes de Bode (module à gauche, phase à droite, mêmes palettes de couleur que dans les figures 9.13 et 9.14) de $\tilde{T}(s)$ en fonction de la valeur de Υ . Sur une image, chaque ligne correspond à un diagramme de Bode (avec la pulsation ω en abscisse) et la position en ordonnée indique la valeur du paramètre Υ .

Étudions d'abord le premier cas : $\theta^2=0$, géométrie parfaitement symétrique. Le cas $\Upsilon=\theta^2=0$ correspond au tube droit donc le propagateur est un retard pur, le module vaut 1 sur toute la droite de Fourier. Au dessous, (courbure négative), le module est plus petit que 1 et finit par être complètement nul quand on atteint le cas limite $\Upsilon=-\pi^2$. Cela s'interprète de la façon suivante. Quand $\Upsilon=0$, nous sommes dans le cas d'un tube droit, tout se propage. Quand $\Upsilon=-\pi^2$, le tube concave devient fermé et rien ne se propage, tout est mode, les propagateurs sont nuls. Tout est mode parce que, précisément, $\Upsilon=-\pi^2$ est le seul cas (limite) où on a des vraies conditions frontière : il correspond à un tronçon fermé. On retrouve cette dernière propriété dans le cas d'un tube dissymétrique ($\theta=4$). Pour un tube dissymétrique, les basses fréquences ne se propagent pas, même dans le cas des tubes exponentiels (cas $\Upsilon=\theta^2$). Pour un tube dissymétrique qui n'est pas exponentiel, des modes sont retenus, pour $\Upsilon>0$ et $\Upsilon<0$. Nous n'avons pas encore aujourd'hui totalement interprété ces modes. La seule chose dont nous sommes sûrs, c'est que (i) ces modes ne sont pas des modes dûs à des conditions frontières donc il ne s'agit pas de décomposition modale et (ii) ils sont dûs à la non exponentialité du profil.

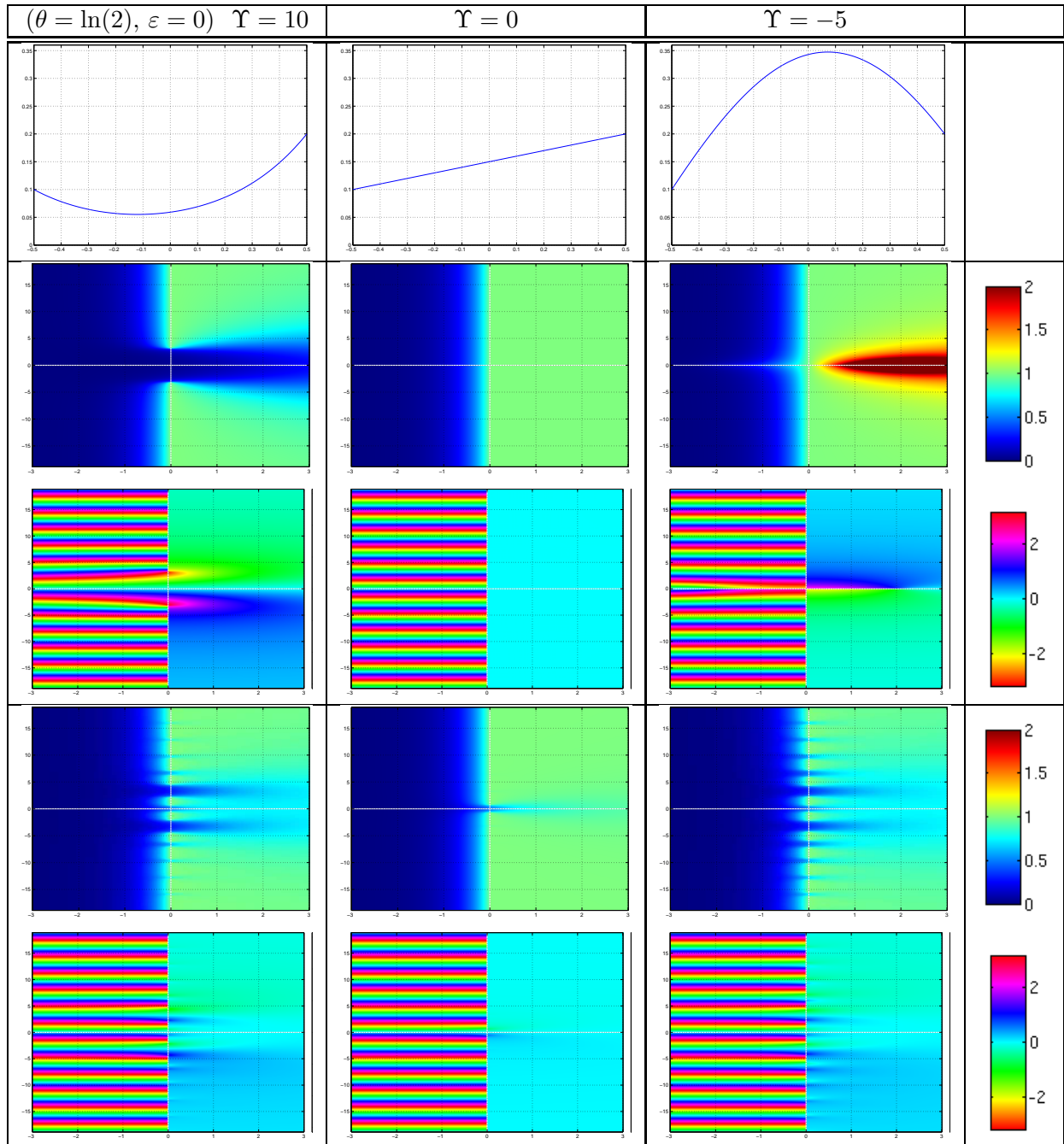


FIGURE 9.13 – Opérateurs de dispersion $D(s)$ et $D_n(s)$ dans le plan de Laplace. Cas $\theta = \ln(2)$ et $\varepsilon = 0$. De haut en bas : Profil du tube $R(\ell)$, module de $D(s)$, phase de $D(s)$, module de $D_n(s)$, phase de $D_n(s)$. Pour les phases, la palette de couleur est circulaire entre $-\pi$ et π .

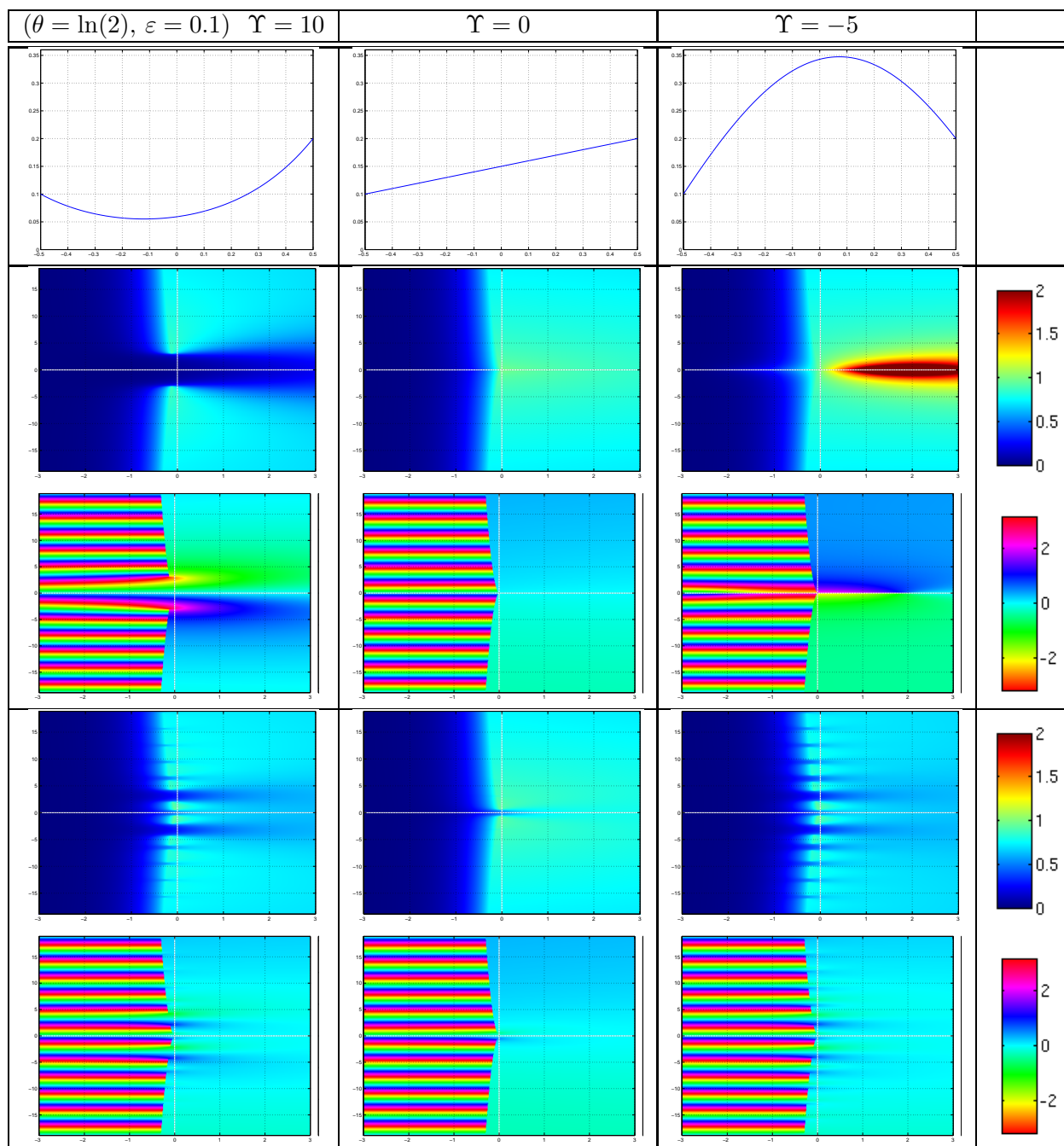


FIGURE 9.14 – Opérateurs de dispersion $D(s)$ et $D_n(s)$ dans le plan de Laplace. Cas $\theta = \ln(2)$ et $\varepsilon = 0.1$. De haut en bas : Profil du tube $R(\ell)$, module de $D(s)$, phase de $D(s)$, module de $D_n(s)$, phase de $D_n(s)$. Pour les phases, la palette de couleur est circulaire entre $-\pi$ et π .

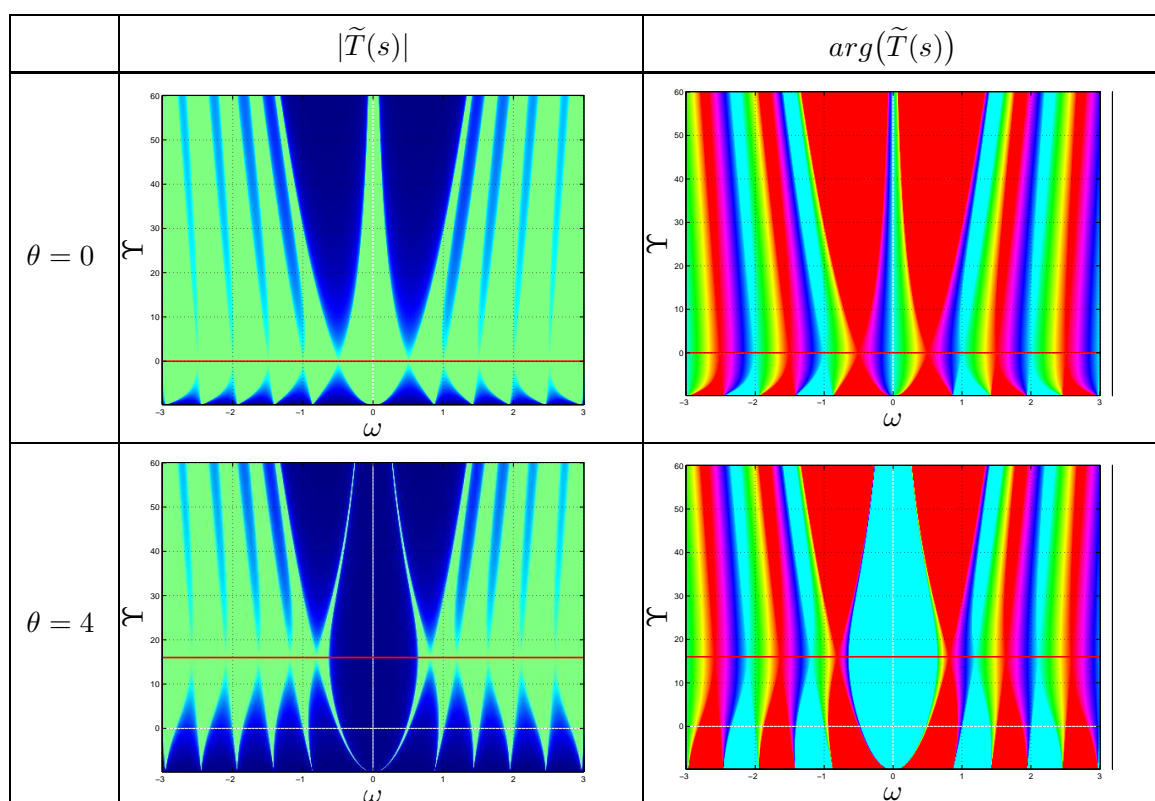


FIGURE 9.15 – Ces figures présentent l'évolution du diagramme de Bode (module à gauche, phase à droite, mêmes palettes de couleur que dans les figures 9.13 et 9.14) de $\tilde{T}(s)$ en fonction du paramètre Υ . Dans tous les cas, $\varepsilon = 0$. Dans la première ligne, la ligne rouge représente la valeur $\Upsilon = \theta^2 = 0$, dans la deuxième ligne, la ligne rouge représente la valeur $\Upsilon = \theta^2 = 16$.

9.4 Conclusion et perspectives

J'ai tout d'abord présenté dans ce chapitre un modèle inspiré du modèle source-filtre mais intégrant le couplage entre la source glottique et le conduit vocal. Un modèle de glotte monodimensionnel fondé sur l'équation de Bernoulli permet d'exprimer le débit glottique en fonction de l'aire glottique, de la pression sous-glottique et de la pression glottique. Le conduit vocal est modélisé par une concaténation de tubes droits dans lesquels les ondes acoustiques se propagent en ondes planes sans pertes. Le rayonnement est modélisé par un filtre dérivateur. Enfin, le couplage est assuré par la continuité de l'état acoustique à l'interface entre la glotte et le conduit vocal. Cette modélisation fait apparaître une fonction de transfert acoustique du conduit vocal sous forme de filtre tous-pôles, donnant ainsi une motivation acoustique du modèle exploité dans la partie II.

Les premières simulations effectuées avec ce modèle permettent de mettre en évidence l'intérêt des modèles couplés par rapport au modèle source-filtre. Tout d'abord, il apparaît dans les simulations que le débit glottique peut s'éloigner quantitativement de la fonction d'aire glottique. Ce résultat, déjà présenté dans les nombreux travaux sur les modèles couplés à géométrie forcée [Rot81b, AF82, Mae82a, AS85] et auto-oscillants [IF72, Vil02, Rut07], est retrouvé ici avec un modèle simplifié.

Ensuite, les simulations font apparaître le rôle important que prend la pression sous-glottique dans le cas couplé. Dans le cas du modèle source-filtre, le gain du modèle de synthèse n'a qu'une

influence sur le niveau sonore du son rayonné. Ici, les modifications de la pression sous-glottique génèrent des modifications dans le spectre du son synthétisé. Ainsi, comme il l'apparaît spécialement dans le cas de la voyelle \a\, une augmentation de la pression sous-glottique a des conséquences sur la qualité vocale.

Ce modèle reste toute fois limité, notamment dans le réalisme des sons produits. Une piste pour améliorer le modèle est d'utiliser un modèle de conduit plus réaliste. Dans cette idée, je me suis proposé d'étudier le modèle de Webster-Lokshin pour les tubes à section variable. Des structures en guides d'ondes adaptées à la simulation numérique existent mais ne permettent pas des simulations stables dans le cas des tubes concaves, très fréquent pour les conduits vocaux. Une solution pour remédier à ce problème est proposée. Tout d'abord, une paramétrisation du rayon permet d'assurer la prise en compte de sa finitude et de sa positivité. Ensuite, la factorisation de la matrice de passage plutôt que de l'équation du modèle, permet de définir des nouvelles ondes découplées dont les opérateurs de propagation sont toujours stables, quelle que soit la géométrie du conduit. De plus, les nouveaux opérateurs de propagation proposés sont compatibles avec plusieurs opérateurs de propagation proposés dans la littérature pour des cas plus simples (tubes droits, coniques, exponentiels).

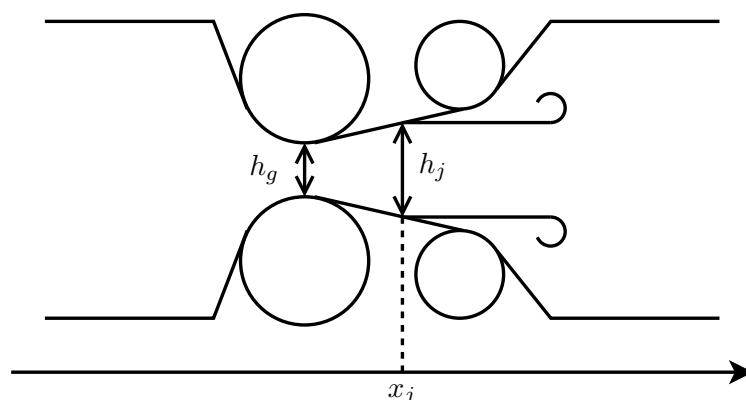
Afin d'obtenir des structures guides d'ondes numériques simulables à partir de ces définitions d'ondes découplées, il reste plusieurs travaux à accomplir.

- La diagonalisation de la matrice de transfert fait apparaître une famille paramétrique d'ondes découplées à deux degrés de liberté (l'opérateur $\tilde{T}(s)$ est valable pour toutes ces ondes) [Hé10, p. 24]. Il est donc nécessaire de faire un choix pour ces deux degrés de liberté afin d'obtenir une décomposition quadripolaire unique.
- Une étude rigoureuse de la stabilité et de la passivité du système complet obtenu à partir de cette décomposition devra être menée.
- Enfin, la discrétisation des opérateurs mis en jeu dans la nouvelle structure sera nécessaire à l'obtention d'un outil de simulation numérique.

Pour finir, une autre piste d'amélioration du modèle serait un travail sur le modèle de glotte. Nous avons mentionné dans la première partie de ce document, notamment dans le chapitre 4, que le mouvement dans l'épaisseur des plis vocaux peut être important dans certains cas, notamment dans le mécanisme 1 pour lequel la masse du muscle *vocalis* est entraînée en vibration. Pour modéliser ce mouvement, il est nécessaire de prendre en compte des modèles bidimensionnels de l'ouverture glottique. Ainsi Ishizaka et Flanagan ont proposé un modèle à 2 masses [IF72] après des premières simulations à l'aide d'un modèle à une masse [FL68], récupérant ainsi la possibilité de modéliser des mouvements plus complexes et notamment des différences de phase entre l'oscillation de la masse inférieure et celle de la masse supérieure (mouvement dans l'épaisseur, effet de buse).

Une modification du modèle initial d'Ishizaka et Flanagan (figure 2.6) a été proposée par Pelorson et al. afin de considérer une géométrie plus réaliste [PHvH⁺94]. Dans ce modèle 2D, les masses ne sont plus rectangulaires mais circulaires et un point de décollement du jet mobile est pris en compte, comme présenté sur la figure 9.16. Liljencrants [Lil93] propose une méthode empirique pour déterminer la position x_j du point de décollement de jet, consistant à supposer que le ratio $\frac{h_j}{h_g}$ est constant. Lous et al. proposent dans [LHVH98] de prendre 1,2 pour cette valeur et montre que ce critère est suffisant pour obtenir des auto-oscillations entretenues du modèle à deux masses ainsi modifié.

Un degré de raffinement supplémentaire dans la prise en compte de décollement de jet a été proposé par Lagrée et al. [LG01, LBD⁺05, LVP07]. Le modèle proposé est fondé sur la prise en

FIGURE 9.16 – Modèles à deux masses modifié [PHvH⁺94].

compte d'une couche limite à la paroi du conduit laryngé et donne une expression de la distribution de pression et de débit tout au long du conduit.

Parallèlement à ces recherches sur le décollement du jet dans la glotte, une simplification du modèle à deux masses est proposée par Avanzini et al. dans [AAK01]. Le modèle à deux masses modifié est ici simplifié en considérant que les deux masses ont un mouvement identique mais décalé par un délai t_0 . Cette approche nous intéresse particulièrement car elle permet une simulation des débits glottiques plus fins qu'avec les modèles à une masse tout en réduisant considérablement le nombre de paramètres du modèle.

Au vu de ces résultats, une approche permettant d'améliorer le modèle présenté dans la section 9.1 tout en maintenant un faible nombre de paramètres et un faible coût de calcul consisterait à

- exploiter la géométrie du modèle à deux masses proposée dans [PHvH⁺94],
- piloter le mouvement des deux masses par la même fonction d'aire glottique (cf chapitre 8) avec un décalage entre les deux masses, tel que proposé par [AAK01], et
- prendre en compte un des modèles de décollement de jet proposé ci-dessus.

Un tel modèle permettrait de récupérer du réalisme dans l'activité laryngée (mouvement dans l'épaisseur des plis vocaux) et par conséquent dans l'onde de débit glottique [IF72, Vil02, Rut07]. Le pilotage ne s'en trouverait pas beaucoup complexifié puisqu'un seul paramètre sera ajouté au pilotage glottique : le délai t_0 entre le mouvement des deux masses. Du fait de la faible dimensionnalité du modèle, l'inversion du modèle devrait aussi être envisageable.





Conclusion générale

L'objectif du travail présenté dans cette thèse était d'étudier et de faire avancer la connaissance autour de la production de la voix selon trois axes :

- A1** l'exploration de l'activité laryngée,
- A2** la modélisation de la phonation,
- A3** l'analyse et la synthèse sonore.

Des résultats selon ces trois axes de recherche ont été présentés tout au long de ce document.

Partie I : Outils pour l'exploration et la mesure de signaux biométriques

L'échographie ultra-rapide [SCTH99, TBSF02, DGTF06] est une technique récente ayant déjà des applications médicales performantes. Elle s'est malheureusement révélée inutilisable en l'état pour l'imagerie des plis vocaux en phonation.

Les deux techniques utilisées aujourd'hui pour l'exploration laryngée sont l'électroglottographie et la vidéo-endoscopie ultra-rapide. J'ai développé un outil de visualisation synchronisée pour l'exploration d'une base de données de signaux électroglottographiques et vidéo-endoscopiques. Cet outil a été exploité dans le cadre d'une étude comparative de la vibration des plis vocaux en phonation et des lèvres d'un joueur de trombone en situation de jeu. Cette étude nous a permis de montrer la pertinence de l'algorithme DECOM [HDDC04] pour l'estimation du quotient ouvert sur des signaux DELG. Ces estimations montrent des valeurs de quotient ouvert globalement plus grandes pour les lèvres que pour les plis vocaux, confirmant les résultats de la littérature. De plus, il apparaît dans nos mesures une catégorisation de la vibration labiale selon la fréquence d'oscillation, similaire à celle de la vibration laryngée. Une notion de mécanisme de vibration semble apparaître pour les lèvres et une étude approfondie du fonctionnement anatomique et mécanique des lèvres en vibration permettra de la valider. Enfin, les observations menées avec ces deux méthodes d'exploration mettent en évidence l'impossibilité d'explorer le mouvement des plis vocaux dans la dimension perpendiculaire au plan glottique avec les outils actuels, bien que l'EKG en donne en aperçu.

Pour pallier cette limitation, un nouveau dispositif, modulaire et évolutif pour une utilisation expérimentale, a été proposé dans le chapitre 5. Ce dispositif s'inspire des EKG multi-canaux existants [Rot92, KF09] et de la technique de l'EIT [Hol10] pour aller vers une reconstruction spatiale tridimensionnelle de l'activité glottique. L'utilisation de la technique de l'EIT pour l'imagerie de l'activité glottique est inédite mais les résultats obtenus par simulation dans la littérature [SNK⁺11] et par mes premières expériences (section 5.2) montrent qu'elle est envisageable. Ce dispositif inclut des canaux d'excitation multiples par multiplexage fréquentiel, des électrodes à géométrie variable grâce à un routeur analogique et la protection des canaux analogiques d'excitation et de mesure est assurée par un blindage des câbles et dans les cartes électroniques multi-couches.

Partie II : Modèle source-filtre et analyse-synthèse

Dans le cadre du paradigme source-filtre en modèle tous-pôles paramétrique, deux algorithmes d'estimation et de séparation source-filtre ont été proposés. Nous avons montré que l'estimation des pôles par calcul des racines de $TFD^{-1}[1/TFD[s]]$, où s est un signal parfaitement périodique construit à partir d'un extrait du signal d'analyse (premier algorithme, section 7.2), conduisait à des résultats encourageants : cette méthode est capable d'estimer correctement les paramètres et de discriminer la contribution de la source glottique de celle du filtre vocal sur des signaux de synthèse non bruités. Mais le calcul de $1/TFD[s]$ pose un problème classique d'amplification du bruit et des perturbations. Des méthodes régularisantes pour ce problème mal posé devraient pouvoir améliorer la méthode pour les signaux de synthèse bruités et réels.

Un deuxième algorithme a été proposé prenant en compte la robustesse au bruit de mesure dès sa conception en maximisant la vraisemblance des cepstres différentiels causaux et anticausaux (section 7.3). Ayant compris l'importance de la prise en compte de l'holomorphie de la transformée du signal et de son logarithme (qui est la transformée du cepstre), je propose de calculer le cepstre complexe par le biais du cepstre différentiel. Ce dernier permet d'assurer l'holomorphie du logarithme de la transformée en Z du signal analysé sur sa couronne de convergence en calculant sa dérivée, qui n'est pas soumise au problème de déterminations multiples (branches multiples du logarithme complexe). Il permet en outre de profiter des algorithmes de transformation de Fourier rapide tout en évitant le problème de repliement fréquentiel lié au calcul du cepstre par logarithme de la transformée de Fourier, et en améliorant la robustesse comparé au calcul du cepstre par ZFT. L'algorithme proposé permet d'estimer correctement tous les paramètres du modèle causal-anticausal sur une réponse impulsionnelle. Il présente une bonne robustesse au fenêtrage de la réponse et une robustesse limitée mais meilleure que celle de l'algorithme précédent à la présence de bruit dans le signal d'analyse. Une première application de l'algorithme en l'état à l'estimation de source glottique présente des résultats encourageants.

Partie III : Modèles à géométrie forcée et analyse-synthèse

Deux approches ont été proposées dans le chapitre 8 pour l'analyse-synthèse des signaux d'aire glottique. J'ai tout d'abord montré que le modèle LF permet une représentation paramétrique efficace des fonctions d'aire glottique. Un algorithme d'estimation des paramètres a été développé et ses performances ont été démontrées sur une base de données de fonctions d'aire glottique estimées à partir d'enregistrements vidéo-endoscopiques ultra-rapides.

Ensuite, J'ai proposé une adaptation du modèle à deux masses au système dynamique de Lur'e. Une adaptation de la méthode de classification proposée dans [De 01] au problème de l'estimation paramétrique des fonctions d'aire glottique nous a permis de montrer la pertinence de ce modèle mais la méthode d'estimation nécessite un effort de recherche complémentaire pour aboutir à une méthode d'analyse entièrement fonctionnelle.

Un modèle inspiré du modèle source-filtre mais intégrant l'interaction entre la source glottique et le filtre vocal a été proposé. Ce modèle, constitué d'un modèle de glotte monodimensionnel et d'un conduit vocal en concaténation de tubes droits, permet de retrouver des phénomènes observés dans les modèles plus complexes [Mae82a, IF72]. Tout d'abord, la pression sous-glottique devient un véritable paramètre de contrôle de la qualité vocale, contrairement au gain du modèle source-filtre (section 9.2.1). Ensuite nous retrouvons des effets de rebonds sur le débit glottique (section 9.2.2), présentés dans la littérature comme la conséquence de l'interaction avec le conduit vocal [IF72]. Cet

effet peut être sur-dimensionné dans nos simulations, ce qui trouve une explication dans l'interaction exacerbée constatée avec les modèles de glotte monodimensionnels (modèle à une masse notamment [FL68, IF72]).

Afin d'améliorer le réalisme de la synthèse, le modèle de Webster-Lokshin [Hé02] permettant la prise en compte des pertes visco-thermiques et d'un profil régulier a été envisagé comme modèle de conduit vocal. Ce modèle ne conduit à des simulations numériques stables que pour le cas des tubes convexes [Mig09]. J'ai montré que la diagonalisation de la matrice de transfert acoustique dans un tube à Υ constant (section 9.3) permet de définir des ondes découplées progressives pour lesquelles les opérateurs de propagation sont toujours stables, quelle que soit la géométrie du conduit. J'ai proposé, dans des travaux passés, un outil de représentation d'un profil de conduit par concaténation régulière de tronçons de tubes à Υ constant. Grâce à ces résultats, l'étude de la discrétisation des nouveaux opérateurs proposés permettra, à terme, l'utilisation du modèle de Webster-Lokshin pour des conduits à section variable, quelle que soit leur géométrie.

Perspectives

Les questions ouvertes et les travaux pouvant poursuivre ces recherches sont nombreux.

L'étude de l'anatomie de l'appareil phonatoire d'après la littérature [GL18, LPV88, L05, Fin08] et mes discussions avec des spécialistes de l'anatomie montre que le contrôle de la vibration laryngée implique de nombreux muscles et articulations et n'est pas encore aujourd'hui compris dans son intégralité. Le rôle des cartilages aryténoïdes est crucial. Ce sont eux qui permettent de mettre les plis vocaux en position de phonation à travers deux mouvements (translation et rotation des cartilages aryténoïdes) présentés dans la figure 1.4-Ⓐ et Ⓑ. Le rôle de ces cartilages et du muscle vocal dans le contrôle des mécanismes laryngés est aujourd'hui discuté et la littérature n'est pas unanime à ce sujet [Hir81, Tit94, Hen01, Fin08]. Une hypothèse correspondant aux illustrations pré-citées et à celles présentées dans [Mic88, p. 22] que je propose est que la rotation des cartilages aryténoïdes permet, en mécanisme 2, d'imposer sur les plis vocaux une condition aux limites de type force (les plis vocaux sont très tendus) tandis que l'accolement des aryténoïdes permet, en mécanisme 1, d'imposer une condition aux limites de type position. Cette hypothèse nécessite une étude plus approfondie. Une meilleure compréhension du contrôle de la vibration laryngée pourra mener à des améliorations significatives dans les modèles de production de la voix et dans la chirurgie du larynx.

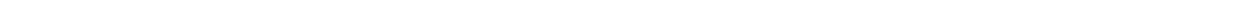
En lien avec la notion de mécanisme laryngé, il m'apparaît aujourd'hui crucial de pouvoir observer le mouvement des plis vocaux dans les trois dimensions de l'espace. L'attente est d'ailleurs grande à ce sujet dans le monde médical. J'ai bon espoir que les travaux initiés dans le cadre du projet VoiceTronics aboutiront à terme à un appareil permettant la reconstruction tridimensionnelle de l'activité laryngée à partir de mesures électriques effectuées à la surface du cou du patient. En ce qui concerne le problème de reconstruction de la géométrie, un moyen de régulariser le problème inverse de Calderón qui me paraît pertinent est la prise en compte d'un modèle paramétrique de la carte volumique de conductivité et l'utilisation des transformations conformes. Les récents travaux de Haddar et al [HK05, HK06, HK10, BHBH10] ont montré les potentialités d'une telle démarche pour l'EIT et l'adaptation de ces méthodes à notre dispositif pourra faire l'objet de travaux futurs.

Dans le domaine de l'analyse-synthèse, un résultat important de mes travaux est l'étude complète du cepstre complexe et la conclusion quant à l'utilisation du cepstre différentiel pour son estimation. Le deuxième algorithme que j'ai proposé nécessite néanmoins des recherches complémentaires pour

aboutir à une méthode complète d'analyse-synthèse du signal vocal. Notamment, l'extension des résultats au cas des fonctions périodiques devra être réalisée. Une difficulté pour ce travail sera l'analyse de l'holomorphie du logarithme de la transformée en Z du signal dans les espaces discrets correspondant aux espaces circulaires des fonctions périodiques dans le domaine temporel. Par ailleurs, une piste qui me paraît intéressante pour améliorer la résistance au bruit est d'améliorer le modèle stochastique de bruit sur les cepstres différentiels. Une possibilité pour cela est de définir un bruit gaussien sur le signal d'analyse et d'étudier analytiquement sa répercussion sur le cepstre différentiel.

Enfin, le champ des possibles est immense dans le domaine de la modélisation couplée de la phonation. Nous avons vu en fin de chapitre 9 des perspectives qui pourraient permettre d'améliorer les modèles existants. Notamment, l'étude de stabilité et de passivité des ondes découplées proposées dans la section 9.3 permettra de valider définitivement l'approche proposées. Une fois les degrés de liberté de la décomposition résolus, la discrétisation pourra être réalisée selon la démarche de [Mig09] en approchant les coupures par un ensemble de pôles discrets. Par ailleurs, le dispositif électronique de mesure d'imagerie des pis vocaux en phonation proposé dans ce document pourra fournir des signaux de pilotage de modèles de glotte à géométrie forcée.





Bibliographie

- [AAK01] F. Avanzini, P. Alku, and M. Karjalainen. One-delayed-mass model for efficient synthesis of glottal flow. In Paul Dalsgaard, Borge Lindberg, Henrik Benner, and Zheng-Hua Tan, editors, *EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, Aalborg, Denmark, September 3-7, 2001*, pages 51–54. ISCA, 2001.
- [ABT00] F. Alipour, D. A. Berry, and I. R. Titze. A finite-element model of vocal-fold vibration. *The Journal of the Acoustical Society of America*, 108(6) :3003–3012, December 2000.
- [AF82] T. V. Ananthapadmanabha and G. Fant. Calculation of true glottal flow and its components. *Speech Communication*, 1(3-4) :167–184, 1982.
- [aix] <http://www.supersonicimagine.com/aixplorer-r/presentation>.
- [AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.
- [AS85] D. R. Allen and W. J. Strong. A model for the synthesis of natural sounding vowels. *The Journal of the Acoustical Society of America*, 78(1) :59–69, 1985.
- [AS95] S. Adachi and M.-A. Sato. Time-domain simulation of sound production in the brass instrument. *The Journal of the Acoustical Society of America*, 97(6) :3850—3861, 1995.
- [AS96] S. Adachi and M.-A. Sato. Trumpet sound simulation using a two-dimensional lip vibration model. *The Journal of the Acoustical Society of America*, 99(2) :1200–1209, 1996.
- [Bai09] L. Bailly. *Interaction entre cordes vocales et bandes ventriculaires en phonation : exploration in-vivo, modélisation physique, validation in-vitro*. Thèse de doctorat, Université du Maine, June 2009. Département Parole et Cognition Département Parole et Cognition.
- [BB84] D. C. Barber and B. H. Brown. Applied potential tomography. *Journal of Physics E : Scientific Instruments*, 17(9) :723–733, 1984.
- [BBP⁺78] A. L. Benabid, L. Balme, J.C. Persat, M. Belleville, J.P. Chirossel, M. Buyle-Bodin, J. de Rougemont, and C. Poupot. Electrical impedance brain scanner : principles and preliminary results of simulation. *Tower International Technomedical Journal of Life Sciences*, 8 :59–68, 1978.

- [BCG10] S. Bromage, M. Campbell, and J. Gilbert. Open areas of vibrating lips in trombone playing. *Acta Acustica united with Acustica*, 96 :603–613, 2010.
- [BCT⁺03] J. Bercoff, S. Chaffai, M. Tanter, L. Sandrin, S. Catheline, M. Fink, J.L. Gennisson, and M. Meunier. In vivo breast tumor detection using transient elastography. *Ultrasound in Medicine & Biology*, 29(10) :1387–1396, October 2003.
- [BD03] B. Bozkurt and T. Dutoit. Mixed-phase speech modeling and formant estimation, using differential phase spectrums. In *Voice Quality : Functions, Analysis and Synthesis VOQUAL'03*, pages 27–30, 2003.
- [Ber62] D. Bernoulli. Physical, mechanical and analytical researches on sound and on the tones of differently constructed organ pipes. Mémoires de l'Academie Scientifique de Paris, 1762.
- [BHBH10] F. Ben Hassen, Y. Boukari, and H. Haddar. Inverse impedance boundary problem via the conformal mapping method : the case of small impedances. *Revue ARIMA*, 13 :47–62, 2010.
- [BHT63] B. Bogert, M. Healy, and J. Tukey. The quefrency alanalysis of time series for echoes : Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe-crackin. In M. Rosenbatt, editor, *Proceedings of the Symposium on Time Series Analysis*, volume 15, pages 209–243, 1963.
- [BHW⁺09] A. Borsic, R. Halter, Y. Wan, A. Hartov, and K. D. Paulsen. Sensitivity study and optimization of a 3D electric impedance tomography prostate probe. *Physiological Measurement*, 30(6) :S1–18, June 2009.
- [BIN00] R. S. Blue, D. Isaacson, and J. C. Newell. Real-time three-dimensional electrical impedance imaging. *Physiological Measurement*, 21(1) :15–26, February 2000.
- [BLB⁺03] A. P. Bagshaw, A. D. Liston, R. H. Bayford, A. Tizzard, A. P. Gibson, A. T. Tidswell, M. K. Sparkes, H. Dehghani, C. D. Binnie, and D. S. Holder. Electrical impedance tomography of human brain function using reconstruction algorithms based on the finite element method. *NeuroImage*, 20(2) :752–764, 2003.
- [BLM83] T. Baer, A. Löfquist, and N. S. McGarr. Laryngeal vibrations : A comparison between high-speed filming and glottographic techniques. *The Journal of the Acoustical Society of America*, 73(4) :1304–1308, April 1983.
- [BO00] R. J. Baken and R. F. Orlikoff. *Clinical Measurement of Speech and Voice*. Speech Science. Singular Thomson Group, San Diego, California, USA, 2nd edition edition, 2000.
- [Bor02] L. Borcea. Electrical impedance tomography. *Inverse Problems*, 18(6) :R99–R136, December 2002.
- [Boz05] B. Bozkurt. *Zeros of the z-transform (ZZT) representation and chirp group delay processing for the analysis of source and filter characteristics of speech signals*. Phd thesis, Faculté Polytechnique de Mons, 2005.
- [BSA06] E. Björkner, J. Sundberg, and P. Alku. Subglottal pressure and normalized amplitude quotient variation in classically trained baritone singers. *Logopedics, Phoniatrics, Vocology*, 31(4) :157–65, January 2006.

-
- [BSM01] B. Brandstätter, H. Scharfetter, and C. Magele. Multi frequency electrical impedance tomography. *COMPEL : The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 20(3) :828–847, 2001.
- [Cal80] A. P. Calderón. On an inverse boundary value problem. In W.H. Meyer and M.A. Raupp, editors, *Seminar on Numerical Analysis and its Applications to Continuum Physic*, pages 65–73, Rio de Janeiro, Brasil, 1980. Sociedade Brasileira de Matematica.
- [Cam04] M. Campbell. Brass instruments as we know them today. *Acta Acustica united with Acustica*, 90 :600–610, 2004.
- [CC95] K. E. Cummings and M. A. Clements. Glottal models for digital speech processing : A historical survey and new results. *Digital Signal Processing*, 5(1) :21–42, 1995.
- [CCS83] M. Castellengo, R. Caussé, and B. Sluchin. Étude acoustique de l'émission multiphonique aux cuivres. In *Proceeding of the 11th International Congress on Acoustics*, pages 355–357, Paris, France, 1983.
- [CGC00] J.S. Cullen, J. Gilbert, and D.M. Campbell. Brass instruments : Linear stability analysis and experiments with an artificial mouth. *Acta Acustica united with Acustica*, 86(4) :704–724, 2000.
- [Cha01] A. Chaigne. *Ondes acoustiques*. Les éditions de l'École Polytechnique, 2001.
- [CHM⁺90] D. G. Childers, D. M. Hicks, G. P. Moore, L. Eskenazi, and A. L. Lalwani. Electroglottography and vocal fold physiology. *Journal of Speech and Hearing Research*, 33(2) :245–254, 1990.
- [CIN99] M. Cheney, D. Isaacson, and J. Newell. Electrical impedance tomography. *SIAM Review*, 41(1) :85–101, 1999.
- [Cis08] J. Cisonni. *Modélisation et inversion d'un système complexe de production de signaux acoustiques*. PhD thesis, Institut Polytechnique de Grenoble, 2008.
- [CK85] D. G. Childers and A. K. Krishnamurthy. A critical review of electroglottography. *Critical Reviews in Biomedical Engineering*, 12(2) :131–161, 1985.
- [CK08] A. Chaigne and J. Kergomard. *Acoustique des instruments de musique*. BELIN, 2008. ISBN 978-2-7011-3970-8.
- [CL84] D. G. Childers and J. N. Larar. Electroglottography for laryngeal function assessment and speech analysis. *Biomedical Engineering, IEEE Transactions on*, BME-31(12) :807–817, 1984.
- [CM88] F. Charpentier and E. Moulines. Text-to-speech algorithms based on fft synthesis. In *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, volume 1, pages 667–670, 1988.
- [CNL⁺83] D. G. Childers, J. M. Naik, J. N. Larar, A. K. Krishnamurthy, and G. P. Moore. Electroglottography, speech and ultra-high speed cinematography. *Vocal fold physiology and biophysics of voice*, pages 202–220, 1983.
-

- [CO98] M. Campedel Oudot. *Étude du modèle “sinusoïdes et bruit” pour le traitement des signaux de parole. estimation robuste de l’enveloppe spectrale*. PhD thesis, École Nationale Supérieure des Télécommunications, 1998.
- [Cre48] L. Cremer. On the acoustic boundary layer outside a rigid wall. *Archiv für Elektronik und Übertragungstechnik*, 2 :136–139, 1948.
- [CS86] F. Charpentier and M. Stella. Diphone synthesis using an overlap-add technique for speech waveforms concatenation. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, volume 11, pages 2015–2018, 1986.
- [CS96] D. C. Copley and W. J. Strong. A stroboscopic study of lip vibrations in a trombone. *The Journal of the Acoustical Society of America*, 99(2) :1219–1226, 1996.
- [CSK77] D. G. Childers, D. P. Skinner, and R. C. Kemerait. The Cepstrum : A guide to Processing. *Proceedings of the IEEE*, 65(10) :1428–1443, 1977.
- [CSM84] D. G. Childers, A. M. Smith, and G. P. Moore. Relationships between electroglottograph, speech, and vocal cord contact. *Folia Phoniatrica et Logopaedica*, 36 :105–118, 1984.
- [CT65] J. Cooley and J. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90) :297–301, 1965.
- [CT89] Calliope and J.-P. Tubach. *La Parole et son traitement automatique*. Masson (Collection technique et scientifique des télécommunications), 1989.
- [Cve12] L. Cveticanin. Review on mathematical and mechanical models of the vocal cord. *Journal Of Applied Mathematics*, 2012 :18, 2012.
- [DA11] T. Drugman and A. Alwan. Joint robust voicing detection and pitch estimation based on residual harmonics. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association, INTERSPEECH 2011*, pages 1973–1976. ISCA, 2011.
- [DBD09] T. Drugman, B. Bozkurt, and T. Dutoit. Complex cepstrum-based decomposition of speech for glottal source estimation. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, pages 116–119, 2009.
- [DBD11] T. Drugman, B. Bozkurt, and T. Dutoit. Causal-anticausal decomposition of speech using complex cepstrum for glottal source estimation. *Speech Communication*, 53(6) :855–866, 2011.
- [DBR08a] G. Degottex, E. Bianco, and X. Rodet. Estimation of glottal area with high-speed videoendoscopy. In *Speech Production Workshop : Instrumentation-based approach, ParisIII/ILPGA, Paris, France, July 2008*.
- [DBR08b] G. Degottex, E. Bianco, and X. Rodet. Usual to particular phonatory situations studied with high-speed videoendoscopy. In *The 6th International Conference on Voice Physiology and Biomechanics, ICVPB*, pages 19–26, Tempere, Finland, August 2008.

-
- [dCK02] A. de Cheveigne and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4) :1917–1930, 2002.
- [DD09] T. Drugman and T. Dutoit. Glottal Closure and Opening Instant Detection from Speech Signals. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, pages 2891–2894, 2009.
- [DDH03] B. Doval, C. D’Alessandro, and N. Henrich. The voice source as a causal/anticausal linear filter. In *Voice Quality : Functions, Analysis and Synthesis VOQUAL’03*, pages 16–20, 2003.
- [DDH06] B. Doval, C. D’Alessandro, and N. Henrich. The Spectrum of Glottal Flow Models. *Acta Acustica united with Acustica*, 92(6) :1026–1046, 2006.
- [De 01] O. De Feo. *Modeling diversity by strange attractors with application to temporal pattern recognition*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2001.
- [Deg10] G. Degottex. *Glottal source and vocal-tract separation*. PhD thesis, 2010.
- [Del06] D. Deliyski. High-speed videoendoscopy : recent progress and clinical prospects. In *Proceedings of the 7th International Conference on Advances in Quantitative Laryngology Voice and Speech Research (AQL ’06)*, volume 7, pages 1–12, University of Groningen, Groningen, the Netherlands, 2006.
- [DGTF06] T. Deffieux, J.-L. Gennisson, M. Tanter, and M. Fink. Ultrafast ultrasonic imaging of in vivo muscle contraction. In *Ultrasonics Symposium, 2006. IEEE*, pages 1001–1004, 2006.
- [DH03] O. De Feo and M. Hasler. Temporal signal classification using chaos. Number september, 2003.
- [DL93] T. Dutoit and H. Leich. Mbr-psola : Text-to-speech synthesis based on an {MBE} re-synthesis of the segments database. *Speech Communication*, 13(3–4) :435–440, 1993.
- [dNCH10] B. d’Andréa Novel, J.-M. Coron, and T. Hélie. Asymptotic state observers for a simplified brass instrument model. *Acta Acustica united with Acustica*, 96(4) :733–742, 2010.
- [dORPGA03] M. de Oliveira Rosa, J. C. Pereira, M. Grellet, and A. Alwan. A contribution to simulating a three-dimensional larynx model using the finite element method. *The Journal of the Acoustical Society of America*, 114(5) :2893–2905, 2003.
- [DP03] D. Deliyski and P. Petrushev. Methods for objective assessment of high-speed videoendoscopy. In *Proceedings of the 6th International Conference : Advances in Quantitative Laryngology, Voice and Speech Research, AQL-2003*, pages 1–16, Hamburg, Germany, April 2003.
- [DPB⁺08] D. Deliyski, P. Petrushev, H. S. Bonilha, T. T. Gerlach, B. Martin-Harris, and R. E. Hillman. Clinical implementation of laryngeal high-speed videoendoscopy : challenges and evolution. *Folia Phoniatrica et Logopaedica*, 60(1) :33–44, January 2008.
-

- [Dru11] T. Drugman. *Advances in Glottal Analysis and its Applications*. Phd thesis, University of Mons, 2011.
- [DSH⁺05] H. Dehghani, N. Soni, R. Halter, A. Hartov, and K. D. Paulsen. Excitation patterns in three-dimensional electrical impedance tomography. *Physiological Measurement*, 26(2) :S185–97, April 2005.
- [DSH⁺10] B. Denby, T. Schultz, K. Honda, T. Hueber, J.M. Gilbert, and J.S. Brumberg. Silent speech interfaces. *Speech Communication*, 52(4) :270–287, 2010.
- [EDCG12] P. Eveno, J.-P. Dalmont, R. Caussé, and J. Gilbert. Wave propagation and radiation in a horn : Comparisons between models and measurements. *Acta Acustica united with Acustica*, 98(1) :158–165, 2012.
- [EDS⁺10] M. Echternach, S. Dippold, J. Sundberg, S. Arndt, M. F. Zander, and B. Richter. High-speed imaging and electroglottography measurements of the open quotient in untrained male voices’ register transitions. *Journal of Voice*, 24(6) :644–50, November 2010.
- [Eis67] E. Eisner. Complete solutions of the “webster” horn equation. *The Journal of the Acoustical Society of America*, 41(4B) :1126–1146, 1967.
- [End] www.richard-wolf.com/en/human-medicine/ent-ork/products/phoniatrics.html.
- [Fab57] P. Fabre. Un procédé électrique percutané d’inscription de l’accolement glottique au cours de la phonation : Glottographie de haute fréquence. *Bulletin de l’Académie Nationale de Médecine*, pages 66–69, 1957.
- [Fan60] G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, 1960.
- [Fan79] G. Fant. Glottal source and excitation analysis. *STL-QPSR*, 20(1) :085–107, 1979.
- [Fan95] G. Fant. The LF-model revisited. Transformations and frequency domain analysis. *STL-QPSR*, 36(2-3) :119–156, 1995.
- [Far40] D. W. Farnsworth. High-speed motion pictures of the vocal cords. *Bell Laboratories Record*, 18 :203–208, 1940.
- [FEG⁺08] M.J. Fagan, S.R. Ell, J.M. Gilbert, E. Sarrazin, and P.M. Chapman. Development of a (silent) speech recognition system for patients following laryngectomy. *Medical Engineering & Physics*, 30(4) :419–425, 2008.
- [Fet86] A. Fettweis. Wave digital filters : Theory and practice. *Proceedings of the IEEE*, 74(2) :270–327, 1986.
- [Fin08] C. Finck. *Anatomie et Physiologie de la Phonation et de la déglutition*. 2008.
- [FKLB94] G. Fant, A. Kruckenberg, J. Liljencrants, and M. Bavegard. Voice source parameters in continuous speech. Transformation of LF-parameters. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP 1994)*, (September) :1451–1454, 1994.
- [FL68] J. L. Flanagan and L. Landgraf. Self-oscillating source for vocal-tract synthesizers. *Audio and Electroacoustics, IEEE Transactions on*, 16(1) :57–64, 1968.

-
- [FL87] G. Fant and Q.-G. Lin. Glottal source-vocal tract acoustic interaction. *The Journal of the Acoustical Society of America*, 81(S1) :S68–S68, 1987.
- [Fla68] J. L. Flanagan. Source-system interaction in the vocal tract. *Annals of the New York Academy of Sciences*, 155(1) :9–17, 1968.
- [Fla72] J. L. Flanagan. *Speech analysis; synthesis and perception*, by James L. Flanagan. Springer-Verlag Berlin, New York, 2nd ed. edition, 1972.
- [Fle93] N. H. Fletcher. Autonomous vibration of simple pressure-controlled valve in gas flows. *The Journal of the Acoustical Society of America*, 93(4) :2172–2180, 1993.
- [FLL85] G. Fant, J. Liljencrants, and Q. Lin. A four-parameter model of glottal flow. *STL-QPSR*, 26(4) :1–13, 1985.
- [FMT99] K. Funaki, Y. Miyanaga, and K. Tochinal. Recursive {ARMAX} speech analysis based on a glottal source model with phase compensation. *Signal Processing*, 74(3) :279–295, 1999.
- [Fra06] T. Frauenrath. Tomographie der Glottis durch Messung der elektrischen Transferimpedanz. Master’s thesis, RWTH Aachen University, 2006.
- [Fre00] I. Frerichs. Electrical impedance tomography (EIT) in applications related to lung and ventilation : a review of experimental and clinical activities. *Physiological Measurement*, 21(2) :R1–21, May 2000.
- [FS11] V. Fréour and G. P. Scavone. Development of an electrolabiograph embedded in a trombone mouthpiece for the study of lip oscillation mechanisms in brass instrument performance. *Journal of the Canadian Acoustical Association*, 39(3) :130–131, 2011.
- [FS12] V. Fréour and G. P. Scavone. Investigation of the effect of upstream airways impedance on regeneration of lip oscillations in trombone performance. In *Proceedings of the Acoustics 2012 conference*, pages 2225–2230, Nantes, France, 2012.
- [Gar54] M. Garcia. Observations on the Human Voice. In *Proceedings of the Royal Society of London*, volume 7, pages 399–410, January 1854.
- [Gar94] M. Garcia. *Hints on singing*. London : E. Ascherberg, New York : E. Schubert, 1894.
- [GC86] B. Guerin and Y. Cheng. Dynamically controlled excitation source for a time-varying formant synthesizer. In *ICASSP ’86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 2003–2006. Institute of Electrical and Electronics Engineers, 1986.
- [GDOM09] M. Golla, D. Deliyski, R. Orlikoff, and H. Moukalled. Objective comparison of the electroglottogram to synchronous high-speed images of vocal-fold contact during vibration. In Manfredi C., editor, *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, volume 6, pages 141–144, 2009.
- [GIR07] Y. Granot, A. Ivorra, and B. Rubinsky. Frequency-division multiplexing for electrical impedance tomography in biomedical applications. *International Journal of Biomedical Imaging*, 2007 :1–9, January 2007.
-

- [GL18] H. Gray and W. H. Lewis. *Anatomy Of The Human Body*. Philadelphia : Lea & Febiger, 1918.
- [GS91] S. K. Gupta and J. Schroeter. Low update rate articulatory analysis/synthesis of speech. In *ICASSP 91 : IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 481–484. IEEE, 1991.
- [Gun03] H. E. Gunter. A mechanical model of vocal-fold collision with high spatial and temporal resolution. *The Journal of the Acoustical Society of America*, 113(2) :994–1000, 2003.
- [HBD12] T. Hueber, G. Bailly, and B. Denby. Continuous Articulatory-to-Acoustic Mapping using Phone-based Trajectory HMM for a Silent Speech Interface. In *InterSpeech 2012 proceedings*, page Tue.P3c.01, Portland, États-Unis, September 2012.
- [HBE⁺08] N. Henrich, P. Bezar, R. Expert, M. Garnier, C. Guerin, C. Pillot, S. Quattrocchi, B. Roubeau, and B. Terk. Towards a common terminology to describe voice quality in Western lyrical singing : Contribution of a multidisciplinary research group. *Journal of Interdisciplinary Music Studies*, 2 :71–93, spring/fall 2008.
- [HDDC04] N. Henrich, C. D’alessandro, B. Doval, and M. Castellengo. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America*, 115(3) :1321–32, March 2004.
- [Hel54] H. L. F. Helmholtz. *On the Sensations of Tone*. Dover, New York, 1954.
- [Hen01] N. Henrich. *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception*. These, Université Pierre et Marie Curie - Paris VI, November 2001.
- [Hen06] N. Henrich. Mirroring the voice from Garcia to the present day : some insights into singing voice registers. *Logopedics, Phoniatrics, Vocology*, 31(1) :3–14, January 2006.
- [Hez] <http://anasynth.ircam.fr/home/people/thomashezard>.
- [HF38] W. Herriott and D. W. Farnsworth. High speed motion pictures of the vocal cords. *The Journal of the Acoustical Society of America*, 9(3) :274–274, 1938.
- [HGB90] D. G. Hanson, B. R. Gerratt, and G. S. Berke. Frequency, intensity, and target matching effects on photoglottographic measures of open quotient and speed quotient. *J Speech Hear Res*, 33(1) :45–50, 1990.
- [HGG12] A. E. Hartinger, R. Guardo, and H. Gagnon. Eit system and reconstruction algorithm adapted for skin cancer imaging. In *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, pages 798–803, 2012.
- [HHCD12] T. Hézard, T. Hélie, R. Caussé, and B. Doval. Analysis-synthesis of vocal sounds based on a voice production model driven by the glottal area. In *Proceedings of Acoustics 2012*, pages 1–5, April 2012.
- [HHD⁺12] T. Hézard, T. Hélie, B. Doval, N. Henrich, and M. Kob. Non-invasive vocal-folds monitoring using electrical imaging methods. In *100 years of electrical imaging*, pages 145–148, Paris, 2012.

-
- [HHD13] T. Hézar, T. Hélie, and B. Doval. A source-filter separation algorithm for voiced sounds based on an exact anticausal/causal pole decomposition for the class of periodic signals. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pages 54–58, 2013.
- [HHDM12] T. Hélie, T. Hézar, L. Delebecque, and R. Mignot. Considerations on travelling waves in the horn equation and energetic aspects. In *Proceedings of the Acoustics 2012 conference*, pages 1–5, Nantes, France, Avril 2012.
- [HHMM13] T. Hélie, T. Hézar, R. Mignot, and D. Matignon. One-dimensional acoustic models of horns and comparison with measurements. *Acta Acustica united with Acustica*, 99 :960–974, 2013.
- [HHP88] E. B. Holmberg, R. E. Hillman, and J. S. Perkell. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America*, 84(2) :511–529, 1988.
- [Hir81] M. Hirano. *Clinical examination of Voice*, volume 5 of *Disorders of human communication*. Springer-Verlag, New York, 1981.
- [HK05] H. Haddar and R. Kress. Conformal mappings and inverse boundary value problems. *Inverse Problems*, 21(3) :935–953, 2005.
- [HK06] H. Haddar and R. Kress. Conformal mapping and an inverse impedance boundary value problem. *Journal of Inverse ill-posed Problems*, 14(8) :785–804, 2006. RR-HH RR-HH.
- [HK10] H. Haddar and R. Kress. Conformal mapping and impedance tomography. *Inverse Problems*, 26(7) :1–18, 2010.
- [HLA90] D. M. Howard, G. A. Lindsey, and B. Allen. Toward the quantification of vocal efficiency. *Journal of Voice*, 4(3) :205–212, 1990.
- [HLG⁺94] T. A. Hughes, P. Liu, H. Griffiths, B. W. Lawrie, and C. M. Wiles. An analysis of studies comparing electrical impedance tomography with x-ray videofluoroscopy in the assessment of swallowing. *Physiological measurement*, 15 Suppl 2 :A199–209, May 1994.
- [HLG⁺96] T. A. Hughes, P. Liu, H. Griffiths, B. W. Lawrie, and C. M. Wiles. Simultaneous electrical impedance tomography and videofluoroscopy in the assessment of swallowing. *Physiological measurement*, 17(2) :109–19, May 1996.
- [Hol10] D. S. Holder. *Electrical Impedance Tomography : Methods, History and Applications*. Series in Medical Physics and Biomedical Engineering. Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK, 2010.
- [How95] D. M. Howard. Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice*, 9(2) :163–172, 1995.
- [HP00] C. M. Hassan and N. A. Peppas. *BIOPOLYMERS/PVA HYDROGELS/ANIONIC POLYMERISATION NANOCOMPOSITES*, volume 153 of *Advances in Polymer Science*, chapter Structure and applications of poly(vinyl alcohol) hydrogels produced by conventional crosslinking or by freezing/thawing methods, pages 37–65. Abe, A., 2000.
-

- [HVLR99] T. Hélie, C. Vergez, J. Lévine, and X. Rodet. Inversion of a physical model of a trumpet. In *IEEE CDC - Conference on Decision and Control*, volume 38.3, pages 2593–2598, Phoenix Arizona, USA, Décembre 1999.
- [Hvv05] J. Horáček, P. Šidlof, and J. G. Švec. Numerical simulation of self-oscillations of human vocal folds with hertz model of impact forces. *Journal of Fluids and Structures*, 20(6) :853–869, 2005.
- [HW78] R. P. Henderson and J. G. Webster. An impedance camera for spatially specific measurements of the thorax. *Biomedical Engineering, IEEE Transactions on*, BME-25(3) :250–254, 1978.
- [Hé02] T. Hélie. *Modélisation physique d'instruments de musique en systèmes dynamiques et inversion*. PhD thesis, Paris 11, 2002. Th. : physique ; 7106.
- [Hé03] T. Hélie. Mono-dimensional models of the acoustic propagation in axisymmetric waveguides. *The Journal of the Acoustical Society of America*, 114 :2633–2647, 2003.
- [Hé09] T. Hézard. Construction de famille d'instruments à vent virtuels. Master's thesis, École Nationale Supérieure de l'Électronique et des ses Applications, 2009.
- [Hé10] T. Hézard. Ondes découplées et ondes progressives dans les tubes acoustiques à section variable pour la représentation en gudies d'ondes. Master's thesis, M2R ATIAM (UPMC/IRCAM/LAM/Telecom ParisTech), 2010.
- [IF72] K. Ishizaka and J. L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal chords. *Bell System Technical Journal*, 51(51) :1233–1368, July-August 1972.
- [Kar10] S.-Z. Karakozoglou. Glottal source analysis : a combinatory study using high-speed videoendoscopy and electroglottography. Master's thesis, University of Paris-Sud XI, Computer Science Department - University of Crete, Computer Science Department - LIMSI-CNRS, Audio & Acoustics Group, June 2010.
- [Kaw97] H. Kawahara. Speech representation and transformation using adaptive interpolation of weighted spectrum : vocoder revisited. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 2, pages 1303–1306, 1997.
- [KAZ00] H. Kawahara, Yoshinori A., and Parham Z. Accurate vocal event detection method based on a fixed-point analysis of mapping from time to weighted average group delay. In *Proceedings of the 6th International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000*, volume 4, pages 664–667, Beijing, China, 2000.
- [KB98] M. Kahrs and K. Brandenburg. *Applications of Digital Signal Processing to Audio and Acoustics*. The Springer International Series in Engineering and Computer Science. Springer, 1998.
- [KBMK13] S. Kazemirad, H. Bakhshaei, L. Mongeau, and K. Kost. Non-invasive in vivo measurement of the mechanical properties of human vocal fold tissue. In *Acoustical Society of America, editor, Proceedings of Meetings on Acoustics*, volume 19, pages 1–8, 2013.

- [KF09] M. Kob and T. Frauenrath. A system for parallel measurement of glottis opening and larynx position. *Biomedical Signal Processing and Control*, 4(3) :221–228, July 2009.
- [Kha02] H. K. Khalil. *Nonlinear Systems*. Prentice Hall PTR, 3rd edition edition, 2002.
- [Kir68] G. Kirchhoff. Ueber die Einflußder Wärmeleitung in einem Gase auf die Schallbewegung. *Annalen der Physik Leipzig*, 134 :177–193, 1868. (English version : R. B. Lindsay, ed., *Physical-Acoustics*, Dowden, Hutchinson and Ross, Stroudsburg, 1974).
- [KK90] D. H. Klatt and L. C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2) :820–857, 1990.
- [KL62] J. L. Kelly and C. C. Lochbaum. Speech Synthesis. In *Proceedings of the 4th International Congress on Acoustics*, pages 1–4, 1962.
- [KL75] P. Kitzing and A. Löfqvist. Subglottal and oral air pressures during phonation-preliminary investigation using a miniature transducer system. *Medical & biological engineering*, 13(5) :644–648, September 1975.
- [KM79] I. Konvalinka and M. Matausek. Simultaneous estimation of poles and zeros in speech analysis and itif-iterative inverse filtering algorithm. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(5) :485–492, 1979.
- [KNR02] M. Kob and C. Neuschaefer-Rube. A method for measurement of the vocal tract impedance at the mouth. *Medical Engineering & Physics*, 24(7–8) :467–471, 2002.
- [Kri83] A. K. Krishnamurthy. *Study of vocal fold vibration and the glottal sound source using synchronized speech, electroglottography and ultra-high speed laryngeal films*. PhD thesis, University of Florida, 1983.
- [L05] G. Léothaud. Théorie de la phonation. Cours de DEUG 2ème année - DMU3D1B, 2004-2005.
- [Lag61] J. L. Lagrange. Nouvelles recherches sur la nature et la propagation du son. Misc. Taurinensia (Mélanges Phil. Math., Soc. Roy. Turin), 1760–1761.
- [Lam10] S. Lamesch. *Mécanismes laryngés et voyelles en voix chantée*. Doctor of philosophy, Université Pierre et Marie Curie (Paris VI), 2010.
- [LB77] F. L. E. Lecluse and M. Brocaar. Quantitative measurements in the electroglottogram. In *17th International Congress of Logopedics and Phoniatrics*, 1977.
- [LB12] M. Le Borgne. Système dynamique non linéaire de production de signaux de source glottique. M2R ATIAM, UPMC, Sept. 2012.
- [LBD⁺05] P.-Y. Lagrée, E. Berger, M. Deverge, C. Vilain, and A. Hirschberg. Characterization of the pressure drop in a 2d symmetrical pipe : Some asymptotical, numerical, and experimental comparisons. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 85(2) :141–146, 2005.

- [Lec77] F. L. E. Lecluse. *Elektroglottografie - Een experimenteel onderzoek betreffende de elektrische impedantie van het mannelijkstrottehoofd*. Phd, Universiteit te Rotterdam, 1977.
- [LG01] P.-Y. Lagrée and K. Goorman. The flow in the glottis : self oscillation of a 2D elastic stenosis. In *Archives of Physiology and Biochemistry V109*, volume 363, page 80, 2001.
- [LHVH98] N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, and A. Hirschberg. A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design. *Acta Acustica united with Acustica*, 84(6) :1135–1150, 1998.
- [Lil93] J. Liljencrants. Personal communication, 1993.
- [Lju99] L. Ljung. *System identification*. Wiley Online Library, 1999.
- [LMZ⁺08] H. Luo, R. Mittal, X. Zheng, Bielałowicz S. A., R. J. Walsh, and J. K. Hahn. An immersed-boundary method for flow–structure interaction in biological systems with application to phonation. *Journal of Computational Physics*, 227(22) :9303–9332, 2008.
- [LPV88] F. Legent, L. Perlemuter, and C. Vandembrouck. *Cahiers d’anatomie ORL*, volume 2. Masson, 1968-1988.
- [LRWW98] J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1) :112–147, 1998.
- [LSGR94] J. Lacau St Guily and B. Roubeau. Voies nerveuses et physiologie de la phonation. *Encyclopédie Médico-Chirurgicale, Oto-rhino-laryngologie*, 1994.
- [LVP07] P.-Y. Lagrée, A. Van Hirtum, and X. Pelorson. Asymmetrical effects in a 2D stenosis. *European Journal of Mechanics - B/Fluids*, 26(1) :83–92, January 2007.
- [Mae82a] S. Maeda. A digital simulation method of the vocal-tract system. *Speech Communication*, 1(3-4) :199–229, 1982.
- [Mae82b] S. Maeda. The role of the sinus cavities in the production of nasal vowels. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, volume 7, pages 911–914, 1982.
- [Mak75] J. Makhoul. Linear prediction : A tutorial review. *Proceedings of the IEEE*, 63(4) :561–580, 1975.
- [Mal00] S. Mallat. *Une exploration des signaux en ondelettes*. Édition de l’École Polytechnique, Palaiseau, 2000.
- [Mar42] D. W. Martin. Lip vibrations in a cornet mouthpiece. *The Journal of the Acoustical Society of America*, 13 :305–308, 1942.
- [Mau12] A. Mauffrey. Réalisation d’un système électronique pour l’exploration non invasive de l’activité glottique pendant la phonation. Rapport de stage de 2ème année de l’École Nationale Supérieure de l’Électronique et de ses Applications, 2012.

- [Mau13] A. Mauffrey. Mesure in vivo non invasive de l'activité glottique pendant la phonation. Master's thesis, École Nationale Supérieure de l'Électronique et de ses Applications, 2013.
- [MBA⁺87] Y. F. Mangnall, A. J. Baxter, R. Avill, N. C. Bird, B. H. Brown, D. C. Barber, A. D. Seagar, A. G. Johnson, and N. W. Read. Applied potential tomography : a new noninvasive technique for assessing gastric function. *Clinical Physics and Physiological Measurement*, 8(4A) :119–129, 1987.
- [MC90] E. Moulines and F. Charpentier. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9(5–6) :453–467, 1990.
- [MG76] J. D.. Markel and A. H. Gray. *Linear prediction of speech*. 1976.
- [Mic88] U. Michels. *Guide illustré de la musique*. Les indispensables de la Musique. Fayard, Novembre 1988.
- [Mig09] R. Mignot. *Réalisation en guides d'ondes numériques stables d'un modèle acoustique réaliste pour la simulation en temps-réel d'instruments à vent*. Doctor of philosophy, Telecom-ParisTech, 2009.
- [Mil59] R. L. Miller. Nature of the Vocal Cord Wave. *Journal of The Acoustical Society of America*, 31(6) :667–677, 1959.
- [MQ86] R. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(4) :744–754, 1986.
- [MvS02] D. G. Miller, J. G. Švec, and H. K. Schutte. Measurement of characteristic leap interval between chest and falsetto registers. *Journal of Voice*, 16(1) :8–19, 2002.
- [NKGB07] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes. Estimation of glottal closure instants in voiced speech using the dypsa algorithm. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(1) :34–43, 2007.
- [NKvS05] X. Niu, A. Kain, and J. P. H. van Santen. Estimation of the acoustic properties of the nasal tract during the production of nasalized vowels. In *Proceedings of the 9th European Conference on Speech Communication and Technology (INTERSPEECH 2005)*, pages 1045–1048. ISCA, 2005.
- [Nol64] A. M. Noll. Short-Time Spectrum and “Cepstrum” Techniques for Vocal-Pitch Detection. *The Journal of the Acoustical Society of America*, 36(2) :537–543, 1964.
- [Nol67] A. M. Noll. Cepstrum pitch determination. *The Journal of the Acoustical Society of America*, 41(2) :293–309, 1967.
- [Opp65] A. V. Oppenheim. Superposition in a class of nonlinear systems. Technical Report 432, Massachusetts Institute of Technology, Research Laboratory of Electronics, March 1965.
- [Orl91] R. F. Orlikoff. Assessment of the dynamics of vocal fold contact from the electroglottogram : Data from normal male subjects. *Journal of Speech and Hearing Research*, 34(5) :1066–1072, 1991.

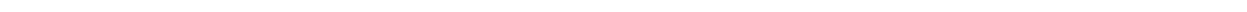
- [OS75] A. V. Oppenheim and R. W. Schaffer. *Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1975.
- [OSB99] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck. *Discrete-time signal processing*. Prentice Hall, 1999.
- [OSS68] A. V. Oppenheim, R. W. Schaffer, and T. G. Stockham. Nonlinear filtering of multiplied and convolved signals. *Proceedings of the IEEE*, 56(8) :1264–1291, 1968.
- [Per01] W. Pereira. *Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency*. Master of engineering, McGill University, 2001.
- [PF81] A. Polydoros and A. T. Fam. The differential cepstrum : definition and properties. In *Proceedings of IEEE International Symposium on Circuits and Systems*, volume 1, pages 77–80, 1981.
- [PH98] R. L. Plant and A. D. Hillel. Direct measurement of subglottic pressure and laryngeal resistance in normal subjects and in spasmodic dysphonia. *Journal Of Voice*, 12(3) :300–314, January 1998.
- [PHvH⁺94] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, and Y. Auregan. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model. *The Journal of the Acoustical Society of America*, 96(6) :3416–3431, 1994.
- [PR06] C. Piccardi and S. Rinaldi. A method for parameter estimation in lur’e systems. *International Journal of Bifurcation and Chaos*, pages 1461–1470, 2006.
- [Qua02] T. F. Quatieri. *Discrete-Time Speech Signal Processing : Principles and Practice*. Prentice Hall, 2002.
- [Rad05] T. Radomski. Manuel Garcia (1805-1906) : A Bicentenary Reflection. *Australian Voice*, 11 :24–41, 2005.
- [RGV92] P. M. Record, R. Gadd, and F. Vinther. Multifrequency electrical impedance tomography. *Clinical Physics and Physiological Measurement*, 13(A) :67–72, 1992.
- [RHC09] B. Roubeau, N. Henrich, and M. Castellengo. Laryngeal vibratory mechanisms : The notion of vocal register revisited. *Journal of Voice*, 23(4) :425–438, 2009.
- [RHI⁺86] M. Rokkaku, K. Hashimoto, S. Imaizumi, Nimi S., and S. Kirtani. Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, 20 :47–54, 1986.
- [Ric03] O. Richards. *Investigation of the Lip Reed Using Computational Modelling and Experimental Studies with an Artificial Mouth*. PhD thesis, University of Edinburgh, Edinburgh, United Kingdom, 2003.
- [RM88] M. Rothenberg and J. J. Mahshie. Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, 31(3) :338–351, 1988.
- [Rod95] X. Rodet. One and two mass model oscillations for voice and instruments. In *ICMC : International Computer Music Conference*, Banff, Canada, Septembre 1995.

- [Ros71] A. E. Rosenberg. Effect of Glottal Pulse Shape on the Quality of Natural Vowels. *The Journal of the Acoustical Society of America*, 49(2B) :583–590, February 1971.
- [Rot81a] M. Rothenberg. Acoustic interaction between the glottal source and the vocal tract. In N. Stevens K and University of Tokyo Press M. Hirano, Eds., editors, *Vocal Fold Physiology*, pages 305–328, 1981.
- [Rot81b] M. Rothenberg. An interactive model for the voice source. *STL-QPSR*, 22(4) :001–017, 1981.
- [Rot92] M. Rothenberg. A multichannel electroglottograph. *Journal of Voice*, 6(1) :36–43, January 1992.
- [Rou93] B. Roubeau. *Mécanismes vibratoires laryngés et contrôle neuro-musculaire de la fréquence fondamentale*. PhD thesis, Université Paris XI, 1993.
- [RPB85] X. Rodet, Y. Potard, and J.-B. Barrière. Chant : de la synthèse de la voix chantée à la synthèse en général. Rapports de recherche IRCAM 35, IRCAM, Paris, France, 1985.
- [RR87] G. R. Reddy and V. V. Rao. On the computation of complex cepstrum through differential cepstrum. *Signal Processing*, 13(1) :79–83, 1987.
- [RS11] L. Rabiner and R. Schafer. *Theory and applications of digital speech processing*. Pearson Education, 2011.
- [Rut07] N. Ruty. *Modèles d'interactions fluide/parois dans le conduit vocal. Applications aux voix et aux pathologies*. PhD thesis, Insitut National Polytechnique de Grenoble, 2007.
- [Sar07] D. Sarason. *Complex Function Theory*. American Mathematical Society, second edition edition, 2007.
- [SCB⁺09] S. Stevenson, M. Campbell, S. Bromage, J. Chick, and J. Gilbert. Motion of the lips of brass players during extremely loud playing. *The Journal of the Acoustical Society of America*, 125(4) :EL152–EL157, 2009.
- [Sch69] R. W. Schafer. Echo removal by discrete generalized linear filtering. Technical Report 466, Massachusetts Institute of Technology, Research Laboratory of Electronics, 1969.
- [Sch96] E. R. Scheinerman. *Invitation to dynamical systems*. Prentice Hall Upper Saddle River, NJ, 1996.
- [SCTH99] L. Sandrin, S. Catheline, M. Tanter, and X. Hennequin. Time-resolved pulsed elastography with ultrafast ultrasonic imaging. *Ultrasonic Imaging*, 21(4) :259–272, December 1999.
- [SD77] K. Steiglitz and B. Dickinson. Computation of the complex cepstrum by factorization of the z-transform. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '77.*, volume 2, pages 723–726, 1977.
- [SH81] J. R. Smitheran and T. J. Hixon. A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech and Hearing Disorders*, 46(2) :138–146, 1981.

- [SHL⁺11] Rudolph Sock, Fabrice Hirsch, Yves Laprie, Pascal Perrier, Béatrice Vaxelaire, Gilbert Brock, Fayssal Bouarourou, Camille Fauth, Véronique Ferbach-Hecker, Liang Ma, Julie Busset, and Jean Sturm. An X-ray database, tools and procedures for the study of speech production. In L. Ménard, S. R. Baum, V. L. Gracco, and D. J. Ostry, editors, *Proceedings of the 9th International Seminar on Speech Production (ISSP2011)*, pages 41–48, Montréal, Canada, June 2011. Département Parole et Cognition de GIPSA-lab Département Parole et Cognition de GIPSA-lab.
- [SL08] K. Schnell and A. Lacroix. Time-varying linear prediction for speech analysis and synthesis. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3941–3944, 2008.
- [SL09] K. Schnell and A. Lacroix. Iterative inverse filtering by lattice filters for time-varying analysis and synthesis of speech. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 4017–4020, 2009.
- [Smi87] J. O. III Smith. Music applications of digital waveguides. Technical report, Stanford University, Center for Computer Research in Music and Acoustics, 1987.
- [SNK⁺11] A. Seppänen, A. Nissinen, V. Kolehmainen, S. Siltanen, and A.-M. Laukkanen. Electrical impedance tomography imaging of larynx. In *Proceedings of the 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA2011)*, pages 27–29, August 2011.
- [sol] <http://www.solidexpress.com>.
- [ST98] Brad H. Story and Ingo R. Titze. Parameterization of vocal tract area functions by empirical orthogonal modes. *Journal of Phonetics*, 26(3) :223–260, July 1998.
- [Ste77] K. Steiglitz. On the simultaneous estimation of poles and zeros in speech analysis. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 25(3) :229–234, 1977.
- [Sto05] B. H. Story. A parametric model of the vocal tract area function for vowel and consonant simulation. *The Journal of the Acoustical Society of America*, 117(5) :3231–3254, 2005.
- [Sty95] Y. Stylianou. *Harmonic plus Noise Models for Speech, combined with Statistical Méthodes, for Speech and Speaker Modification*. PhD thesis, École Nationale Supérieure des Télécommunications, 1995.
- [SW10] T. Schultz and M. Wand. Modeling coarticulation in emg-based continuous speech recognition. *Speech Communication*, 52(4) :341–353, 2010.
- [SWKG01] J. A. Stammen, S. Williams, D. N. Ku, and R. E. Guldberg. Mechanical properties of a novel PVA hydrogel in shear and unconfined compression. *Biomaterials*, 22(8) :799–806, 2001.
- [TBLT10] V.-A. Tran, G. Bailly, H. Loevenbruck, and T. Toda. Improvement to a nam-captured whisper-to-speech system. *Speech Communication*, 52(4) :314–326, 2010.
- [TBSF02] M. Tanter, J. Bercoff, L. Sandrin, and M. Fink. Ultrafast compound imaging for 2-d motion vector estimation : application to transient elastography. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 49(10) :1363–1374, 2002.

-
- [Tit73] I. R. Titze. The human vocal cords : A mathematical model. i. *Phonetica*, 28(3-4) :129–170, 1973.
- [Tit74] I. R. Titze. The human vocal cords : A mathematical model. ii. *Phonetica*, 29(1-2) :1–21, 1974.
- [Tit82] I. R. Titze. Synthesis of sung vowels using a time domain approach. In New York) V.L. Lawrence (The Voice Foundation, editor, *Transcripts of the Eleventh Symposium : Care of the Professional Voice*, pages 90–98, 1982.
- [Tit84] I. R. Titze. Parameterization of the glottal area, glottal flow, and vocal fold contact area. *The Journal of the Acoustical Society of America*, 75(2) :570–580, February 1984.
- [Tit89] I. R. Titze. A four-parameter model of the glottis and vocal fold contact area. *Speech Communication*, 8(3) :191–201, September 1989.
- [Tit94] I. R. Titze. *Principles of Voice Production*. Prentice Hall Inc., Englewood Cliffs, New York, 1994.
- [TS97] I. R. Titze and B. H. Story. Acoustic interactions of the voice source with the lower vocal tract. *The Journal of the Acoustical Society of America*, 101(4) :2234–2243, April 1997.
- [Uhl09] G. Uhlmann. Electrical impedance tomography and calderón’s problem. *Inverse Problems*, 25(12) :1–39, 2009.
- [Val66] G. Valiron. *Théorie des fonctions*. Number vol. 1 in Cours d’analyse mathématique. Masson, 3ème édition edition, 1966.
- [vdB58] J. van den Berg. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, 1(3) :227–244, 1958.
- [vdBZD57] J. van den Berg, J. T. Zantema, and R. Jr. Doornenbal. On the Air Resistance and the Bernoulli Effect of the Human Larynx. *The Journal of the Acoustical Society of America*, 29(5) :626–631, 1957.
- [Vel98] R. Veldhuis. A computationally efficient alternative for the liljencrants–fant model and its perceptual evaluation. *The Journal of the Acoustical Society of America*, 103(1) :566–571, 1998.
- [Ven67] W. Vennard. *Singing : The Mechanism and the Technic*. Music Instruction Bks. Carl Fischer, L.L.C, 1967.
- [VET92] H. Valbret, Moulines ; E., and J. P. Tubach. Voice transformation using {PSOLA} technique. *Speech Communication*, 11(2–3) :175–187, 1992.
- [Vil02] C. E. Vilain. *Contribution à la synthèse de parole par modèle physique : applications à l’étude des voix pathologiques*. PhD thesis, Institut national polytechnique (Grenoble), 2002.
- [VPF+04] C.E. Vilain, X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, and J. Willems. Experimental validation of a quasi-steady theory for the flow through the glottis. *Journal of Sound and Vibration*, 276(3–5) :475–490, 2004.
-

- [VRR06] F. Villavicencio, A. Röbel, and X. Rodet. Improving lpc spectral envelope extraction of voiced speech by true-envelope estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006 (ICASSP'06)*, page 1, Toulouse, Mai 2006.
- [Š00] J. G. Švec. *On Vibration Properties of Human Vocal Folds : Voice Registers, Bifurcations, Resonance Characteristics, Development and Application of Videokymography*. PhD thesis, University of Groningen, Groningen, the Netherlands, 2000.
- [VVSK99] P. J. Vauhkonen, M. Vauhkonen, T. Savolainen, and J. P. Kaipio. Three-dimensional electrical impedance tomography based on the complete electrode model. *IEEE Transactions on Bio-medical Engineering*, 46(9) :1150–1160, September 1999.
- [Vä95] V. Välimäki. *Discrete-time modeling of acoustic tubes using fractional delay filters*. Doctor of philosophy, Helsinki University of Technology, Faculty of Electrical Engineering, 1995.
- [Web19] A. G. Webstern. Acoustical impedance, and the theory of horns and of the phonograph. *Proceedings of the National Academy of Sciences of the United States of America*, 5(7) :275–282, 1919.
- [WP94] É. Walter and L. Pronzato. *Identification de modèles paramétriques à partir de données expérimentales*. Masson, 1994.
- [WTMD01] W. Wang, M. Tang, M. McCormick, and X. Dong. Preliminary results from an EIT breast imaging simulation system. *Physiological Measurement*, 22(1) :39–48, February 2001.
- [YM03] S. Yoshikawa and Y. Muto. Lip-wave generation in horn players and the estimation of lip-tissue elasticity. *Acta Acustica united with Acustica*, 89 :145–162, 2003.
- [ZG03] Y. Zou and Z. Guo. A review of electrical impedance techniques for breast cancer detection. *Medical Engineering & Physics*, 25(2) :79–90, 2003.
- [ZK49] C. Zwicker and C. W. Kosten. *Sound absorbing materials*. Elsevier New York, 1949.





Annexe A

Projet VoiceTronics

Le projet VoiceTronics a été initié au début de ma thèse de doctorat, en septembre 2010, par Thomas Hélie, Nathalie Henrich et moi-même. Malte Kob, Thierry Legou et Aude Lagier ont rejoint le projet en septembre 2011. Le projet VoiceTronics est constitué de trois sous-projets

SP1 un dispositif électronique de mesure de l'activité laryngée,

SP2 une maquette de larynx robotisée à échelle réelle et à géométrie simplifiée,

SP3 une maquette de larynx robotisée à l'échelle 5 :1 présentant les cartilages du larynx et leurs articulations,

et s'est déroulé dans deux cadres différents.

1. Projets mécatroniques avec des étudiants de deuxième année à l'école Mines ParisTech. Le projet VoiceTronics a été suivi par trois générations d'étudiants, en 2010-2011, 2011-2012 et 2012-2013.

2. Stages d'Antonin Mauffrey. Antonin Mauffrey, étudiant brillant de l'École Nationale Supérieure de l'Électronique et de ses Applications, a effectué son stage de fin de deuxième année en Juillet et Août 2011 et son Projet de Fin d'Études d'Ingénieur entre avril et septembre 2013 à l'IRCAM, encadré par Thomas Hélie, Emmanuel Fléty et moi-même, sur le dispositif électronique de mesure de l'activité laryngée.

Le sous-projet **SP1** est présenté en détail dans le chapitre 5. Après avoir présenté les projets mécatroniques de l'école Mines ParisTech et notre rôle dans le projet VoiceTronics dans la section [A.1](#), je présente dans la section [A.2](#) la maquette à échelle réelle et dans la section [A.3](#) la maquette à l'échelle 5 :1, toutes deux entièrement réalisées par les étudiants qui ont travaillé sur le projet.

A.1 Projets mécatroniques à l'école Mines ParisTech

Des projets mécatroniques sont organisés à l'école Mines ParisTech pour les étudiants de seconde année. Ces projets constituent une première expérience de mise en situation industrielle. Un "client" extérieur à l'école propose un projet qui est assigné à un groupe de dix à quinze étudiants pendant une année, à raison d'un après-midi par semaine consacré au projet. Les étudiants ont la charge de

- définir un cahier des charges précis avec le client,
- s'organiser en groupes de travail pour
- réaliser l'objectif industriel défini par le cahier des charges et
- établir une stratégie de communication autour du projet.

Les étudiants sont encadrés par un professeur d'électronique (Denis Brousse dans notre cas) et un professeur de mécanique (Frédéric Bruyère puis Jérémie Rupil dans notre cas) qui les conseillent dans l'organisation et les développements techniques. Ils collaborent avec trois groupes de trois à cinq étudiants, en BTS électronique, mécanique et communication. Ces étudiants proviennent de lycées éloignés géographiquement (en région parisienne) afin de mettre les étudiants dans la situation d'une entreprise travaillant avec des prestataires externes.

Nous avons donc tenu le rôle de client avec Thomas Hélie et Nathalie Henrich (puis avec Malte Kob, Thierry Legou et Aude Lagier). J'ai pu personnellement suivre de près les trois générations d'étudiants qui ont travaillé sur le projet afin de suivre les décisions prises et ré-orienter les choix techniques pour que les prototypes soient au plus proche de ce que nous attendions. Parmi les trois sous-projets, le dispositif de mesure électronique de mesure de l'activité glottique (**SP1**) a peu avancé. Pour cette raison, nous avons préféré proposer un stage à un étudiant d'école d'ingénieur spécialisé en électronique. Nous avons ainsi recruté Antonin Mauffrey, grâce à qui le dispositif a bien avancé (voir le chapitre 5). Les sous-projets **SP2** et **SP3** ont bien mieux avancé, bien que de façon inégale selon le niveau d'investissement et les compétences des étudiants affectés à ces sous-projets chaque année.

A.2 SP2 : Maquette de larynx robotisée à échelle réelle et à géométrie simplifiée

L'objectif de cette maquette de larynx robotisée à échelle réelle et à géométrie simplifiée était le test, la calibration et la validation *in vitro* du dispositif électronique de mesure de l'activité glottique avant les premiers essais sur cobayes humains. Le cahier des charges résumé de ce sous-projet est le suivant.

- La maquette doit posséder des propriétés électriques et mécaniques analogues à celles des tissus humains de la région laryngée, mais avec une conductivité homogène.
- La géométrie de la maquette doit être simplifiée (forme extérieure cylindrique) mais des plis vocaux doivent être présents et mobiles.
- Les plis vocaux sont pilotés en vibration avec soit un asservissement en position, soit un contrôle en temps-réel de la position des plis vocaux.

La solution adoptée par les étudiants a été d'utiliser un hydrogel conducteur¹ [HP00, SWKG01]. La maquette présente une forme cylindrique à l'extérieur. Le profil interne a été dessiné par simplification de planches anatomiques, sous la supervision d'Aude Lagier. Des barres d'entraînement sont

1. Nous remercions Messieurs Laurent Corté et David Moreau du Centre des Matériaux, Mines-ParisTech, CNRS UMR 7633, pour avoir validé la faisabilité, fourni la matériau et permis le moulage.

intégrées à la maquette lors du moulage pour permettre l'entraînement des plis vocaux en vibration. Ces barres sont fines, rigides et non métalliques pour perturber le moins possible la conductivité du tissu artificiel. Une mesure de l'ouverture glottique à l'aide d'une source lumineuse et d'une diode photo-sensible a été proposée mais non réalisée. L'hydrogel nécessitant une hydratation permanente, un caisson étanche à hygrométrie contrôlée a été étudié. La réalisation de ce caisson a été entamée par les étudiants mais n'a pas été menée au terme.

La figure A.1 présente la maquette réalisée.

A.3 SP3 : Maquette de larynx robotisée à l'échelle 5 :1

Sous l'impulsion de l'équipe pédagogique en charge des projets mécatroniques, exigeant un sous-projet mettant en œuvre des liaisons mécaniques plus complexes, une maquette à l'échelle 5 :1 présentant les cartilages du larynx avec une géométrie réaliste et leurs articulations simplifiées a été ajoutée au projet. Cette maquette présente a posteriori un intérêt double.

1. Elle nous a servi de support de discussion lors de nos échanges avec des spécialistes d'anatomie.
2. Elle constituera à terme un outil de démonstration très intéressant pour la pédagogie médicale et intéresse de fait particulièrement Aude Lagier, chirurgienne ORL et enseignante en anatomie.

Nous souhaitons une maquette dont la géométrie n'est pas entièrement idéalisée, comme c'est habituellement le cas dans les maquettes que l'on trouve dans les salles de cours d'anatomie. Pour la conception des cartilages, nous avons envisagé deux approches.

La première consistait à exploiter un scanner de la région laryngée. J'ai retracé sur chaque image, avec l'aide d'Aude Lagier, le contour des différents cartilages pour ensuite reconstruire un objet 3D en compilant les différentes coupes. La faible résolution des images et la difficulté de suivre le contour des cartilages sur les images en raison de leur calcification inégale a abouti à des cartilages 3D difficilement exploitables car très irréguliers. Cette solution a donc été abandonnée.

La deuxième approche a consisté à scanner en 3D des moulages en plâtre de cartilages prélevés sur un larynx humain. Le prélèvement et le moulage ont été effectués par Aude Lagier. Le scan 3D a été confié à la société SolideXpress [sol]. Ces scans 3D ont été ensuite lissés par les étudiants et imprimés sur une imprimante 3D.

Les articulations du larynx ont été présentées au chapitre 1. Pour notre maquette, nous avons simplifier ces articulations en nous limitant à

1. une liaison pivot pour modéliser l'articulation crico-thyroïdienne, et
2. une liaison pivot pour modéliser l'articulation thyro-aryténoïdienne.

Cette deuxième approximation n'est pas entièrement satisfaisante mais les moyens de production des étudiants ne nous permettaient pas d'aller plus loin. Les deux articulations sont commandées par des servo-moteurs (les deux cartilages aryténoïdes sont mis en rotation de façon symétrique à l'aide d'un engrenage, un seul servo-moteur pilote les deux rotations).

La figure A.2 présente des photos de la maquette réalisée.

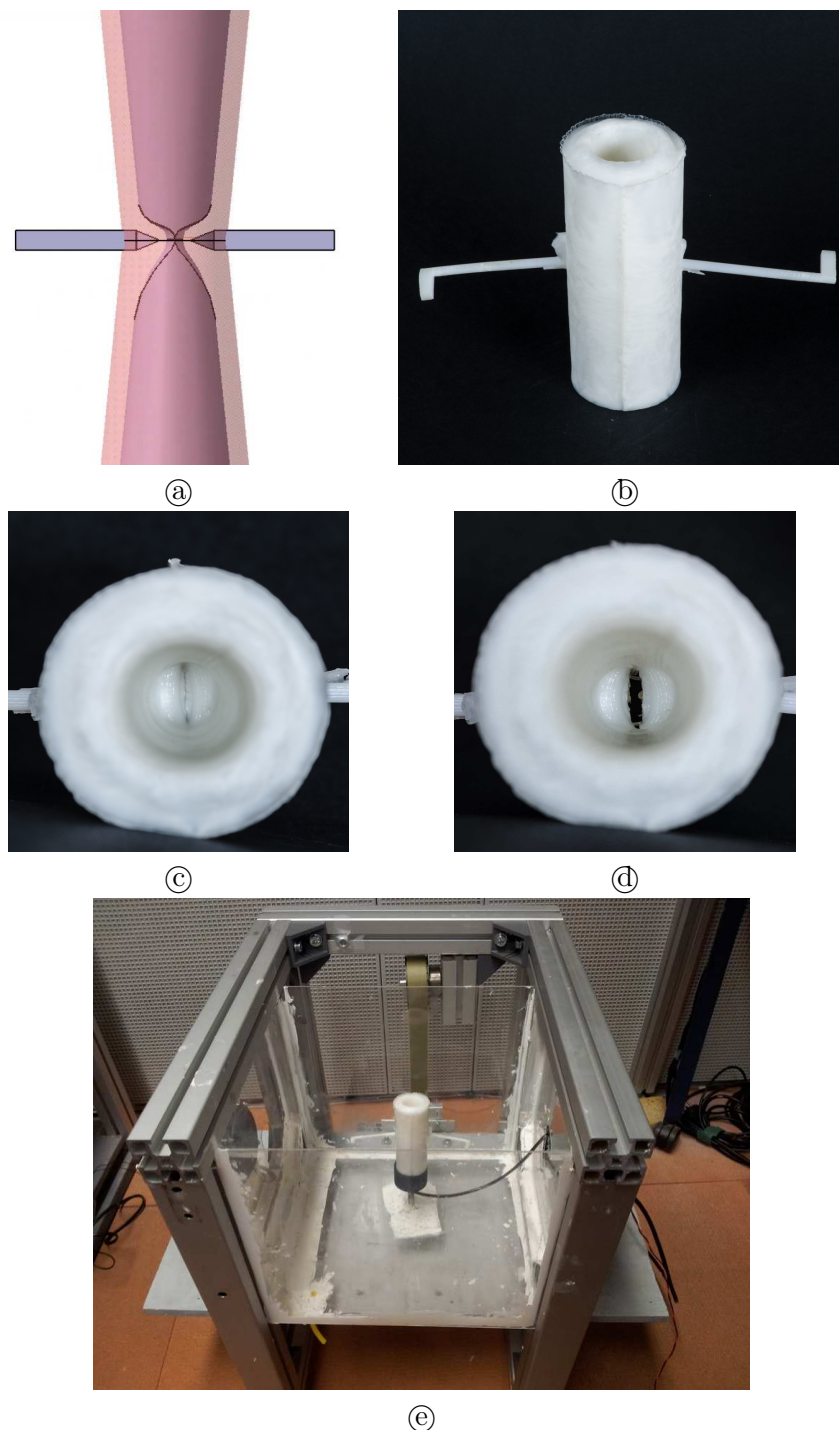


FIGURE A.1 – Maquette de larynx à échelle réelle et à géométrie simplifiée. (a) vue en transparence du profil interne du conduit et des barres d'entraînement des plis vocaux. (b) vue de ma maquette avec les barres d'entraînement des plis vocaux. (c) - (d) vue interne en glotte fermée et glotte ouverte. La longueur de l'ouverture glottique est de 1 cm et l'amplitude de vibration est d'environ 3 mm. (e) caisson accueillant la maquette, des haut-parleurs assurent la mise en vibration.

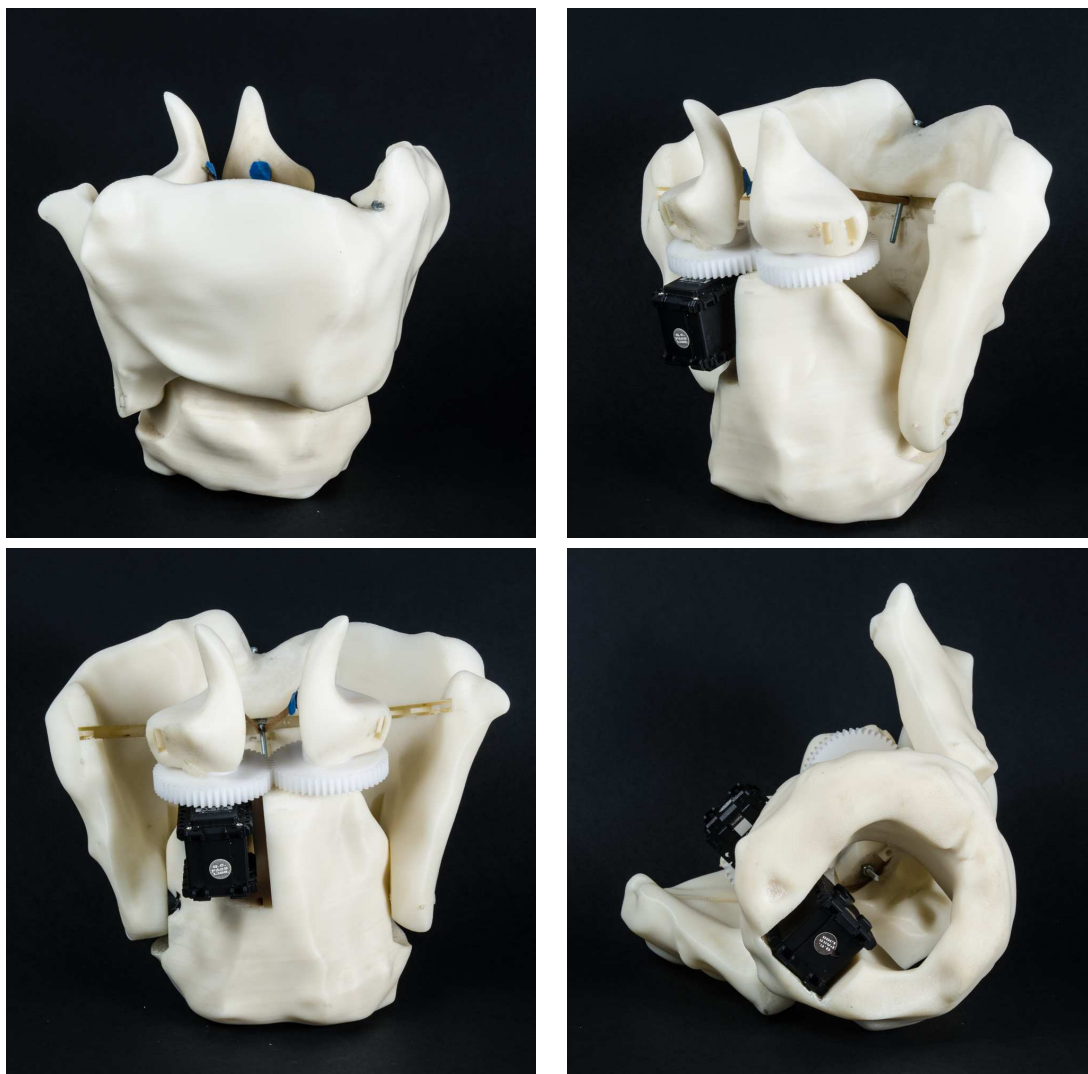
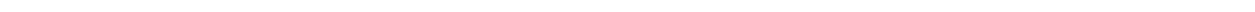


FIGURE A.2 – Maquette de larynx robotisée à l'échelle 5 :1. Les deux servo-moteurs (bloc noirs) pilotent les articulations crico-thyroïdienne et thyro-aryténoïdiennes. Les engrenages (cercles blancs) permettent de piloter de façon symétrique les deux cartilages aryténoïdes. D'autres photos de cette maquette sont présentées tout au long du chapitre 1.

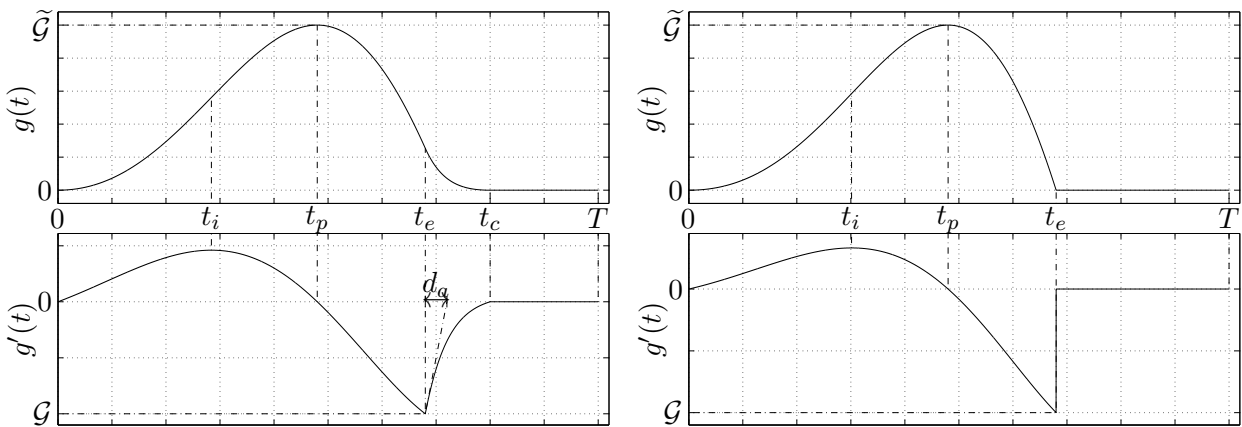


Annexe B

Modèles de pulse glottique par fonctions paramétriques définies par morceaux

B.1 Forme générique

Les modèles présentés ici sont définis par morceaux selon la forme générique de la figure ci-après. Nous pouvons distinguer une phase montante sur $[0, t_p]$ (avec un maximum de dérivée en t_i), une phase descendante $[t_p, t_e]$, une phase de retour $[t_e, t_c]$ paramétrée par d_a , et une phase nulle $[t_c, T]$, où T est la période glottique. Pour chacun des modèles présentés, nous précisons quels sont les paramètres et si les phases de retour et nulles sont prises en compte. Notons que les gains \mathcal{G} et $\tilde{\mathcal{G}}$ sont liés pour chacun des modèles. Lorsqu'un des deux gains est choisi, le deuxième est aisément calculé en écrivant les égalités $\tilde{\mathcal{G}} = g(t_p)$ et $\mathcal{G} = g'(t_e)$.



Forme générique avec phase de retour

Forme générique sans phase de retour

(les figures sont issues ici du modèle LF)

B.2 Expressions analytiques des modèles

B.2.1 Rosenberg-B (1971)

Rosenberg présente dans [Ros71] plusieurs modèles de pulse glottique parmi lesquels le modèle B (ici dénommé Rosenberg-B) ressort comme étant particulièrement pertinent. Ce modèle est composé de deux parties polynomiales :

$$g(t) = \begin{cases} \tilde{\mathcal{G}} \left[\left(\frac{t}{t_p} \right)^2 \left(3 - 2 \frac{t}{t_p} \right) \right] & 0 \leq t \leq t_p \\ \tilde{\mathcal{G}} \left[1 - \left(\frac{t - t_p}{t_e - t_p} \right)^2 \right] & t_p < t \leq t_e \\ 0 & t_e < t < T \end{cases} . \quad (\text{B.1})$$

Ce modèle ne possède que deux paramètres de formes t_e et t_p . Il possède pas de phase de retour.

B.2.2 KLGLOTT88 (1990)

Le modèle KLGLOTT88 est une version modifiée du modèle Rosenberg-B, proposée par Klatt et Klatt dans [KK90]. La phase descendante du modèle précédent est supprimée et remplacée par le prolongement de la phase ascendante, supprimant un degré de liberté au modèle en fixant

$$t_e = \frac{3}{2} t_p . \quad (\text{B.2})$$

$$g(t) = \begin{cases} \tilde{\mathcal{G}} \left[\left(\frac{t}{t_p} \right)^2 \left(3 - 2 \frac{t}{t_p} \right) \right] & 0 \leq t \leq t_e = \frac{3}{2} t_p \\ 0 & t_e < t < T \end{cases} . \quad (\text{B.3})$$

Ce pulse est ensuite filtré par un filtre passe-bas résonnant à 3 kHz afin de régulariser sa forme, c'est-à-dire récupérer une phase de retour. Les deux paramètres de forme de ce filtre sont donc t_p (ou t_e , qui est lié à t_p par B.2) et la pente du filtre passe-bas.

B.2.3 R++ (1998)

Veldhius propose dans [Vel98] un modèle inspiré du modèle Rosenberg-B et du modèle de Fant que nous verrons plus loin. La partie ascendante est la même que celle de KLGLOTT88, la partie descendante est remplacée par une exponentielle décroissante et la partie nulle est supprimée. Ce modèle s'exprime donc sous sa forme dérivée par

$$g'(t) = \begin{cases} \tilde{\mathcal{G}} \left[\frac{6t}{t_p^2} \left(1 - \frac{t}{t_p} \right) \right] & 0 \leq t \leq t_e \\ g'(t_e) \frac{e^{-\frac{t-t_e}{d_a}} - e^{-\frac{T-t_e}{d_a}}}{1 - e^{-\frac{T-t_e}{d_a}}} & t_e < t < T \end{cases} . \quad (\text{B.4})$$

Ce modèle possède trois paramètres de forme $\{t_p, t_e, d_a\}$. Il possède une phase de retour mais pas de phase nulle.

B.2.4 Fant (1979)

Le premier modèle de Fant, présenté dans [Fan79], est le premier modèle de source glottique faisant apparaître des fonctions circulaires. Il est composé d'une phase ascendante, d'une phase descendante et d'une phase nulle.

$$g(t) = \begin{cases} \tilde{\mathcal{G}} \cdot \frac{1}{2} \left[1 - \cos\left(\pi \frac{t}{t_p}\right) \right] & 0 \leq t \leq t_p \\ \tilde{\mathcal{G}} K \cos\left(\pi \frac{t-t_p}{t_p}\right) - K + 1 & t_p < t \leq t_e = t_p \left(1 + \frac{\arccos \frac{K+1}{K}}{\pi}\right) \\ 0 & t_e < t < T \end{cases} . \quad (\text{B.5})$$

Il possède deux paramètres de forme t_p et K . Ce dernier représente la pente de la phase descendante à l'instant t_e . Il est aussi possible de paramétrer le modèle avec les paramètres t_e et t_p puisque t_e , t_p et K sont liés par l'équation $t_e = t_p \left(1 + (\arccos \frac{K+1}{K})/\pi\right)$.

B.2.5 LF (Liljencrants-Fant) (1985)

Le modèle introduit par Fant, Liljencrants et Lin dans [FLL85], communément appelé LF est un modèle très couramment utilisé pour les applications d'analyse-synthèse du signal de parole. Les paramètres t_e , t_p , d_a et t_c sont réglables indépendamment. Il s'écrit sous sa forme dérivée

$$g'(t) = \begin{cases} g'_1(t) = -\mathcal{G} e^{\alpha(t-t_e)} \frac{\sin(\pi \frac{t}{t_p})}{\sin(\pi \frac{t_e}{t_p})} & 0 \leq t \leq t_e \\ g'_2(t) = -\frac{\mathcal{G}}{\varepsilon d_a} \left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)} \right) & t_e < t \leq t_c \\ g'_3(t) = 0 & t_c \leq t < T \end{cases} . \quad (\text{B.6})$$

Les paramètres α et ε sont déterminés par les deux équations implicites

$$\lim_{t \rightarrow t_e} g'_2(t) = g'_1(t_e) \quad \text{et} \quad (\text{B.7})$$

$$\int_0^T g'(t) dt = \int_0^{t_e} g'_1(t) dt + \int_{t_e}^{t_c} g'_2(t) dt + \int_{t_c}^T g'_3(t) dt = 0, \quad (\text{B.8})$$

qui expriment la continuité de la dérivée en t_e (B.7) et la nullité de la moyenne de la source glottique (B.8). Ces deux équations peuvent s'écrire plus explicitement

$$\varepsilon d_a = 1 - e^{-\varepsilon(t_c-t_e)}, \quad (\text{B.9})$$

$$\frac{\frac{\pi}{t_p} \left(\cos\left(\pi \frac{t_e}{t_p}\right) - e^{-\alpha t_e} \right) - \alpha \sin\left(\pi \frac{t_e}{t_p}\right)}{\sin\left(\pi \frac{t_e}{t_p}\right) \left(\alpha^2 + \frac{\pi^2}{t_p^2} \right)} = \frac{1 - e^{-\varepsilon(1-t_e)} (1 + \varepsilon(1-t_e))}{\varepsilon^2 d_a}. \quad (\text{B.10})$$

Une étude du modèle LF incluant une ré-écriture en version adimensionnée ainsi qu'un calcul analytique de son spectre est proposée en annexe C. Différents paramétrages de ce modèle sont présentés dans [FKLB94]. Notons que ce modèle, en raison de son grand nombre de degrés de libertés, permet d'obtenir des formes de pulse glottique très variées. Comme il l'est montré dans l'annexe C, il est possible de supprimer la phase de retour, les formules ci-dessus dégénèrent de façon régulière lorsque d_a tend vers 0. De même, il est possible de supprimer la phase nulle en fixant $t_c = T$.

B.2.6 LF modifié (LF-Rd) (1995)

Le modèle LF a été modifié par Fant en 1995 dans [Fan95]. Le modèle LF modifié, souvent dénommé LF-Rd dans la littérature, est en fait une réduction du modèle à un seul paramètre de forme Rd . Des formules pour retrouver les paramètres de forme $\{t_e, t_p, d_a, t_c\}$ du modèle LF à partir du seul paramètre Rd sont proposées dans [Fan95].

Annexe C

Étude analytique du modèle LF

C.1 Écriture temporelle du modèle de dérivée du débit

Une période fondamentale du modèle LF s'écrit :

$$U'(t) = E(t) = \begin{cases} E_1(t) = -E_e e^{\alpha(t-t_e)} \frac{\sin(\pi \frac{t}{t_p})}{\sin(\pi \frac{t_e}{t_p})} & 0 \leq t \leq t_e \\ E_2(t) = -\frac{E_e}{\varepsilon d_a} \left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)} \right) & t_e < t \leq t_c \\ E_3(t) = 0 & t_c \leq t \leq T \end{cases} \quad (\text{C.1})$$

Les paramètres du modèles sont :

- E_e excitation maximale,
- T période fondamentale,
- t_e instant d'excitation maximale,
- t_c instant de fermeture complète,
- t_p instant de débit maximal,
- d_a constante de temps de la fermeture glottique,
- α ,
- ε .

Ce modèle est contraint par plusieurs équations implicites :

$$\text{Continuité en } t_e : \quad \lim_{t \rightarrow t_e} E_2(t) = E_1(t_e). \quad (\text{C.2})$$

$$\text{Moyenne nulle : } \int_0^T E(t) dt = \int_0^{t_e} E_1(t) dt + \int_{t_e}^{t_c} E_2(t) dt + \int_{t_c}^T E_3(t) dt = 0. \quad (\text{C.3})$$

Nous verrons que la contrainte (C.2) n'est valable que dans le cas $d_a > 0$.

C.2 Écriture temporelle adimensionnée du modèle de dérivée du débit

L'adimensionnement des paramètres temporels suivant permet de simplifier l'écriture du modèle :

$$\tilde{t}_e = \frac{t_e}{t_c}, \quad \tilde{t}_p = \frac{t_p}{t_c}, \quad \tilde{d}_a = \frac{d_a}{t_c}, \quad \tilde{T} = \frac{T}{t_c}, \quad \tilde{\alpha} = \alpha t_c, \quad \tilde{\varepsilon} = \varepsilon t_c. \quad (\text{C.4})$$

On écrit alors

$$U'(\tilde{t}) = E(\tilde{t}) = \begin{cases} E_1(\tilde{t}) = -E_e e^{\tilde{\alpha}(\tilde{t}-\tilde{t}_e)} \frac{\sin\left(\pi \frac{\tilde{t}}{\tilde{t}_p}\right)}{\sin\left(\pi \frac{\tilde{t}_e}{\tilde{t}_p}\right)} & 0 \leq \tilde{t} \leq \tilde{t}_e \\ E_2(\tilde{t}) = -\frac{E_e}{\tilde{\varepsilon}\tilde{d}_a} \left(e^{-\tilde{\varepsilon}(\tilde{t}-\tilde{t}_e)} - e^{-\tilde{\varepsilon}(1-\tilde{t}_e)} \right) & \tilde{t}_e < \tilde{t} \leq 1 \\ E_3(\tilde{t}) = 0 & 1 \leq \tilde{t} \leq \tilde{T} \end{cases}, \quad (\text{C.5})$$

où $\tilde{t} = \frac{t}{t_c}$. La suite du document sera écrit selon la variable temporelle adimensionnée et les $\tilde{\cdot}$ seront omis dans un soucis de clarté.

C.3 Équations implicites et bilan des degrés de liberté

C.3.1 Cas $d_a > 0$: “fermeture douce”

L'équation (C.2) est équivalente à

$$\varepsilon d_a = 1 - e^{-\varepsilon(1-t_e)}. \quad (\text{C.6})$$

Cette équation en ε n'a pas de solution analytique générale mais il est possible de trouver une solution numériquement pour d_a , et t_e donnés.

L'équation (C.3) est équivalente à

$$\frac{\frac{\pi}{t_p} \left(\cos\left(\pi \frac{t_e}{t_p}\right) - e^{-\alpha t_e} \right) - \alpha \sin\left(\pi \frac{t_e}{t_p}\right)}{\sin\left(\pi \frac{t_e}{t_p}\right) \left(\alpha^2 + \frac{\pi^2}{t_p^2} \right)} = \frac{1 - e^{-\varepsilon(1-t_e)} (1 + \varepsilon(1-t_e))}{\varepsilon^2 d_a}. \quad (\text{C.7})$$

De même, cette équation en α n'a pas de solution analytique générale mais il est possible de trouver une solution numériquement, pour d_a , t_e , t_p et ε donnés.

Le modèle possède donc

$$8 \text{ paramètres} - 2 \text{ équations implicites} = 6 \text{ degrés de liberté}. \quad (\text{C.8})$$

Le paramétrage usuel est : E_e l'excitation maximale, T la période fondamentale (qui s'exprime dans le modèle en tant que variable d'adimensionnement), t_e l'instant d'excitation maximale, t_c l'instant de fermeture complète, t_p l'instant de débit maximal et d_a la constante de temps de la fermeture glottique. α et ε sont calculés à partir des équations implicites (C.6-C.7).

C.3.2 Cas $d_a = 0$ “fermeture abrupte”

Le cas $d_a = 0$ est un cas limite. En effet, on montre que

$$\lim_{d_a \rightarrow 0} E_2(t) = 0, \quad \forall t \in]t_e, 1]. \quad (\text{C.9})$$

L'équation (C.2) n'a alors plus de sens et elle n'intervient pas dans ce cas. De plus, on remarque que le modèle (en variable temporelle dimensionnée) ne dépend plus de la valeur de t_c . On peut donc fixer (en variable temporelle adimensionnée) $t_c = 1$ dans ce cas et exprimer $E(t)$ uniquement à l'aide des fonctions E_1 et E_2 (qui est la fonction nulle dans ce cas limite). Enfin, on remarque que le paramètre ε n'intervient plus dans le modèle. On montre par ailleurs à l'aide de (C.6) que $\lim_{d_a \rightarrow 0} \varepsilon = +\infty$ et $\lim_{d_a \rightarrow 0} \varepsilon d_a = 1$. On peut alors dire que (C.6) est toujours “vérifiée par continuité” dans le cas $d_a = 0$.

L'équation (C.3) est équivalente à

$$\frac{\pi}{t_p} \left(\cos \left(\pi \frac{t_e}{t_p} \right) - e^{-\alpha t_e} \right) - \alpha \sin \left(\pi \frac{t_e}{t_p} \right) = 0. \quad (\text{C.10})$$

Cette équation en α n'a pas de solution analytique générale mais il est possible de trouver une solution numériquement, pour t_e et t_p donnés.

Le modèle possède donc

$$\begin{aligned} & 8 \text{ paramètres} - 1 \text{ équation implicite} - 3 \text{ paramètres ne s'exprimant plus} \\ & = 4 \text{ degrés de liberté.} \end{aligned} \quad (\text{C.11})$$

Le paramétrage usuel est le suivant : E_e l'excitation maximale, T la période fondamentale qui s'exprime dans le modèle en tant que paramètre d'adimensionnement, t_e l'instant d'excitation maximale et t_p l'instant de débit maximal. α n'est pas directement interprétable physiquement et est calculé à partir des autres paramètres grâce à l'équation implicite (C.10).

C.4 Écriture temporelle adimensionnée du modèle de débit

$$\begin{aligned} U(t) &= \int_0^t E(\tau) d\tau \\ &= \begin{cases} t_c T U_1(t) & 0 \leq t \leq t_e \\ t_c T U_2(t) & t_e \leq t \leq 1 \\ t_c T U_3(t) & 1 \leq t \leq \frac{1}{t_c} \end{cases}, \text{ avec} \end{aligned} \quad (\text{C.12})$$

$$\begin{cases} U_1(t) = \frac{E_e e^{\alpha(t-t_e)}}{\sin(\omega_G t_e)(\alpha^2 + \omega_G^2)} (\omega_G \cos(\omega_G t) - \alpha \sin(\omega_G t)) + k_1 \\ U_2(t) = \delta \left[\frac{E_e}{\varepsilon^2 d_a} \left(e^{-\varepsilon(t-t_e)} + \varepsilon e^{-\varepsilon(1-t_e)} t \right) + k_2 \right] \\ U_3(t) = k_3 \end{cases}. \quad (\text{C.13})$$

Les constantes d'intégration k_1 , k_2 et k_3 sont calculées pour vérifier les contraintes

$$U_1(0) = U_2(1) = U_3(1) = 0, \quad (\text{C.14})$$

et on obtient

$$\begin{cases} k_1 = -\frac{E_e e^{\alpha t_e} \omega_G}{\sin(\omega_G t_e) (\alpha^2 + \omega_G^2)}, \\ k_2 = -\frac{E_e}{\varepsilon^2 d_a} e^{-\varepsilon(1-t_e)} (1 + \varepsilon), \\ k_3 = 0. \end{cases} \quad (\text{C.15})$$

On a alors

$$\begin{cases} U_1(t) = \frac{E_e e^{-\alpha t}}{\sin(\omega_G t_e) (\alpha^2 + \omega_G^2)} \left[e^{\alpha t} (\omega_G \cos(\omega_G t) - \alpha \sin(\omega_G t)) - \omega_G \right] \\ U_2(t) = \delta \frac{E_e e^{\varepsilon t}}{\varepsilon^2 d_a} \left[e^{-\varepsilon t} - e^{-\varepsilon} (\varepsilon(1-t) + 1) \right] \\ U_3(t) = 0 \end{cases}. \quad (\text{C.16})$$

De plus, on montre aisément que (C.7) assure la continuité du débit en t_e .

C.5 Transformées de Fourier des modèles temporel et dérivé

Les fréquences sont adimensionnées, $\tilde{f} = f/F_c$ avec $F_c = 1/t_c T$.

$$E(t) = [\mathbb{1}_{[0,t_e]}(t)E_1(t) + \mathbb{1}_{[t_e,t_c]}(t)E_2(t)] * \frac{1}{t_c} \text{III}_{\frac{1}{t_c}}(t). \quad (\text{C.17})$$

$$\begin{aligned} \mathbf{F}_1(f) &= TF[\mathbb{1}_{[0,t_e]}E_1](f) \\ &= \frac{-t_c T E_e}{\sin(\omega_G t_e) ((\alpha - 2i\pi f)^2 + \omega_G^2)} \left[e^{-2i\pi f t_e} \left((\alpha - 2i\pi f) \sin(\omega_G t_e) \right. \right. \\ &\quad \left. \left. - \omega_G \cos(\omega_G t_e) \right) + e^{-\alpha t_e} \omega_G \right], \end{aligned} \quad (\text{C.18})$$

$$\begin{aligned} \mathbf{F}_2(f) &= TF[\mathbb{1}_{[t_e,t_c]}E_2](f) \\ &= -\delta \frac{t_c T E_e e^{\varepsilon t_e}}{\varepsilon d_a} \left[\frac{\varepsilon e^{-(\varepsilon+2i\pi f)}}{2i\pi f (\varepsilon + 2i\pi f)} + e^{-2i\pi f t_e} \left(\frac{e^{-\varepsilon t_e}}{\varepsilon + 2i\pi f} - \frac{e^{-\varepsilon}}{2i\pi f} \right) \right]. \end{aligned} \quad (\text{C.19})$$

$$TF[E](f) = (\mathbf{F}_1 + \mathbf{F}_2(f))(f) \times \text{III}_1(f). \quad (\text{C.20})$$

$$TF[U](f) = \frac{TF[E](f)}{2i\pi f F_c} = \frac{(\mathbf{F}_1 + \mathbf{F}_2(f))(f) \times \frac{1}{t_c} \text{III}_{\frac{1}{t_c}}(f)}{2i\pi f F_c}. \quad (\text{C.21})$$

Annexe D

Dictionnaire de cepstres en temps discret

Cette annexe présente le calcul du cepstre de plusieurs signaux utiles en analyse/synthèse de la voix ainsi que quelques propriétés remarquables du cepstre tel qu'il est défini dans la section 7.3.1.2. Les domaines de convergence –DDC– sont précisés pour la transformée en Z des signaux et pour le logarithme de la transformée en Z s'il est différent.

D.1 Impulsion de Dirac

$$\begin{aligned}x(n) &= \delta(n) \\X(z) &= 1, \quad \text{DDC} = \mathbb{C} \\ \log[X](z) &= 0 \\ \hat{x}(n) &= 0\end{aligned}$$

D.2 Pôle

D.2.1 Pôle causal

$$\begin{aligned}x(n) &= a^n H(n), \quad |a| < 1 \\X(z) &= \frac{1}{1 - \frac{a}{z}}, \quad \text{DDC} = \{z \in \mathbb{C} / |z| > |a|\} \\ \log[X](z) &= -\log\left(1 - \frac{a}{z}\right) \\ &= \sum_{n=1}^{\infty} \left(\frac{a^n}{n}\right) z^{-n} \\ \hat{x}(n) &= \frac{a^n}{n} H(n-1)\end{aligned}$$

D.2.2 Pôle anticausal

$$\begin{aligned}
 x(n) &= a^n H(-n), & |a| > 1 \\
 X(z) &= \frac{1}{1 - \frac{z}{a}}, & \text{DDC} = \{z \in \mathbb{C} / |z| < |a|\} \\
 \log[X](z) &= -\log\left(1 - \frac{z}{a}\right) \\
 &= \sum_{n=1}^{\infty} \left(\frac{a^{-n}}{n}\right) z^n \\
 \hat{x}(n) &= -\frac{a^n}{n} H(-n-1)
 \end{aligned}$$

D.3 Zéro

D.3.1 Zéro causal

$$\begin{aligned}
 x(n) &= \delta(n) - b\delta(n-1), & |b| < 1 \\
 X(z) &= 1 - \frac{b}{z}, & \text{DDC} = \mathbb{C}^* \\
 \log[X](z) &= \log\left(1 - \frac{b}{z}\right) \\
 &= -\sum_{n=1}^{\infty} \left(\frac{b^n}{n}\right) z^{-n}, & \text{DDC} = \{z \in \mathbb{C} / |z| > |b|\} \\
 \hat{x}(n) &= -\frac{b^n}{n} H(n-1)
 \end{aligned}$$

D.3.2 Zéro anticausal

$$\begin{aligned}
 x(n) &= \delta(n) - \frac{1}{b}\delta(n+1), & |b| > 1 \\
 X(z) &= 1 - \frac{z}{b}, & \text{DDC} = \mathbb{C} \\
 \log[X](z) &= \log\left(1 - \frac{z}{b}\right) \\
 &= -\sum_{n=1}^{\infty} \left(\frac{b^{-n}}{n}\right) z^n, & \text{DDC} = \{z \in \mathbb{C} / |z| < |b|\} \\
 \hat{x}(n) &= \frac{b^n}{n} H(-n-1)
 \end{aligned}$$

D.4 Propriétés remarquables

D.4.1 Convolution de deux signaux temporels

$$\begin{aligned}x(n) &= s_1 * s_2(n) \\X(z) &= S_1(z) \times S_2(z) \\ \log[X](z) &= \log[S_1](z) + \log[S_2](z) \\ \hat{x}(n) &= \hat{s}_1(n) + \hat{s}_2(n)\end{aligned}$$

D.4.2 Multiplication par une constante

$$\begin{aligned}x(n) &= \alpha s(n), & \alpha \in \mathbb{C}^* \\X(z) &= \alpha S(z) \\ \log[X](z) &= \ln(\alpha) + \log[S](z) \\ \hat{x}(n) &= \ln(\alpha)\delta(n) + \hat{s}(n)\end{aligned}$$

D.4.3 Homothétie temporelle

$$\begin{aligned}x(n) &= s(\alpha n), & \alpha \in \mathbb{R}^* \\X(z) &= \frac{1}{\alpha} S\left(\frac{z}{\alpha}\right) \\ \log[X](z) &= -\ln(\alpha) + \log[S]\left(\frac{z}{\alpha}\right) \\ \hat{x}(n) &= -\ln(\alpha)\delta(n) + \alpha\hat{s}(\alpha n)\end{aligned}$$



Annexe E

Modèles de fonction d'aire glottique

Un des premiers modèles de fonction d'aire glottique est celui proposé par Ananthapadmanabha et Fant dans [AF82] dans le cadre de leur modèle à géométrie forcée. Deux fonctions A_g^1 et A_g^2 sont proposées pour représenter une période de fonction d'aire glottique.

$$\begin{aligned} A_g^1(t) &= A_{max} \sin\left(\frac{\pi t}{T_0}\right) & 0 \leq t < T_0, \\ A_g^2(t) &= \frac{A_{max}}{2} \left(1 - \cos\left(\frac{\pi t}{t_p}\right)\right) & 0 \leq t < t_p, \\ &= A_{max} \cos\left(\frac{\pi(t - t_p)}{2(T_0 - t_p)}\right) & t_p \leq t < T_0. \end{aligned}$$

Les paramètres de ces modèles sont (1) le gain A_{max} et (2) la période T_0 pour A_g^1 et A_g^2 , et (3) le paramètre de forme t_p pour A_g^2 uniquement. L'impédance glottique est ensuite exprimée à partir de l'aire glottique [vdBZD57] afin de pouvoir calculer le débit glottique en fonction de l'aire glottique, de la pression sub-glottique et des charges acoustiques de la trachée et du conduit vocal.

Un modèle inspiré de la physiologie des plis vocaux est proposé par Titze dans [Tit82]. Ce modèle peut s'exprimer de la façon suivante

$$\begin{aligned} A_g(\theta) &= A_{max} \left[\left(\frac{\theta}{\theta_m}\right)^{-\theta_m \cot \theta_m} \left(\frac{\sin(\theta)}{\sin(\theta_m)}\right) \right]^\beta & 0 \leq \theta < \Pi, \\ &= 0 & \Pi \leq \theta < \Pi\gamma^{-1}, \end{aligned}$$

avec $\theta = \Pi t / \gamma \Pi$ et $\theta_m = \Pi \delta / (1 + \delta)$. Les paramètres de ce modèle sont (1) le gain A , (2) la période T_0 , (3) le quotient ouvert γ , (4) δ qui pilote l'asymétrie du pulse et (5) β qui pilote la pente de l'ouverture et de la fermeture. Ce modèle de signaux d'aire glottique a été utilisé dans plusieurs systèmes de synthèse vocal [Tit82, AS85], relativement similaires à celui d'Ananthapadmanabha et Fant.

En 1991, Gupta et Schroeter proposent un modèle de fonction d'aire glottique à vocation générique [GS91]. Dans ce modèle, la dérivée de la fonction d'aire glottique, présentée en figure E.1,

s'exprime par morceaux

$$A'_g(t) = \begin{cases} b_1 \cdot t^2, & 0 \leq t < T_1, \\ a_2 \sin\left(\frac{\pi(t - T_1)}{2T_2}\right) + b_2 \cos\left(\frac{\pi(t - T_1)}{2T_2}\right) + c_2, & T_1 \leq t < T_1 + T_2, \\ D + a_3(t - T_1 - T_2) + b_3(t - T_1 - T_2)^2, & T_1 + T_2 \leq t < T_1 + T_2 + T_3, \\ 0, & T_1 + T_2 + T_3 \leq t < T. \end{cases}$$

Ce modèle possède un grand nombre de degrés de liberté ($T_1, T_2, T_3, T_4, b_1, a_2, b_2, c_2, a_3, b_3$ et D), avantage pour la diversité des formes d'ondes synthétisables mais pénalisant dans le cadre d'une méthode d'analyse. Il permet de générer des discontinuités aux points A-B et C-D correspondant aux pics dans la dérivée de la fonction d'aire glottique, que l'on peut rapprocher des instants d'ouverture et de fermeture glottique. De plus, la nullité de la moyenne de $A_g(t)$ sur une période n'est pas imposée, de sorte que ce modèle puisse être utilisé pour la synthèse des sons voisés et non voisés, ainsi que des transitoires.

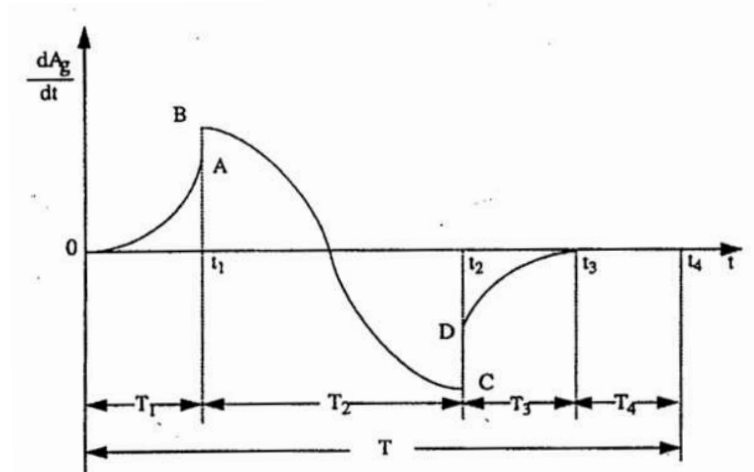


FIGURE E.1 – Fonction d'aire glottique proposée par Gupta et Schroeter. Extrait de [GS91].

De nombreux autres modèles existent, à une, deux ou trois dimensions. Nous pouvons notamment citer les travaux de Titze sur la modélisation de l'ouverture glottique en deux et trois dimensions [Tit84, Tit89]. Par ailleurs, Rothenberg propose dès 1981 [Rot81b] un modèle semblable au modèle d'Ananthapadmanabha et Fant mais dans lequel la variation temporelle de l'impédance glottique est directement modélisée, sans passer par la fonction d'aire glottique. Enfin, le lecteur pourra se référer à l'article de Cummings [CC95] qui présente un bestiaire des modèles articulatoires de production de la voix et passe en revue un certain nombre de modèles de fonction d'aire glottique.

Annexe F

Modele de Lur'e

F.1 Modèle de Lur'e en temps continu

F.1.1 Formulation

Le modèle de Lur'e en temps continu est très similaire à celui à temps discret (figure F.1), l'équation récursive du filtre dans le cas discret étant équivalente ici à une équation différentielle linéaire de transformée de Laplace $G(s)$. On peut donc écrire

— Equation différentielle (transformée de Laplace $G(s)$) :

$$\sum_{k=0}^{K-1} \alpha_k y^{(k)}(t) = \sum_{l=0}^{L-1} \beta_l u^{(l)}(t)$$

— Fonction non linéaire ψ :

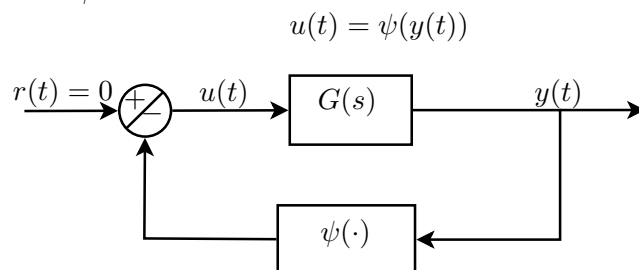


FIGURE F.1 – Système général de lur'e à temps continu.

F.1.2 Méthode de discrétisation à l'aide de l'approximation des trapèzes

Une méthode générale de discrétisation des systèmes de Lur'e est présentée dans [PR06]. Nous la présentons ici succinctement pour montrer ses limites dans notre cas.

Soit le système

$$\begin{cases} \dot{x}(t) &= Ax(t) + bu(t) \\ y(t) &= c^\top x(t) \\ u(t) &= \psi(y(t)) \end{cases} .$$

Nous calculons la transformée de Laplace de (F.1)

$$\begin{aligned} X(s)(sI - A) &= bU(s), \\ Y(s) &= \underbrace{c^\top (sI - A)^{-1} b}_{R(s)} U(s). \end{aligned} \quad (\text{F.1})$$

Nous cherchons ensuite à exprimer la fonction de transfert $Q(z)$ du système, $Q(z) = \frac{Y(z)}{U(z)}$. Nous utilisons pour cela l'approximation des trapèzes pour une période d'échantillonnage T_S .

$$\begin{aligned} x(t + T_S) - x(t) &= \int_t^{t+T_S} Ax(t) + bu(t) dt \\ &\simeq \frac{T_S}{2} [(Ax(t + T_S) + bu(t + T_S)) + (Ax(t) + bu(t))]. \end{aligned}$$

En notation discrète, pour $x_n = x(nT_S)$ (de même pour u et y),

$$x_{n+1} - x_n \simeq \frac{T_S}{2} [(Ax_{n+1} + bu_{n+1} + (Ax_n + bu_n))]. \quad (\text{F.2})$$

D'où par transformée en z

$$\begin{aligned} X(z) \left[(z - 1)I - \frac{T_S}{2}(z + 1)A \right] &= \frac{T_S}{2} bU(z). \\ \Rightarrow Y(z) &= \underbrace{c^\top \left(\frac{2(z - 1)}{T_S(z + 1)} I - A \right)^{-1} b}_{Q(z)} U(z). \end{aligned} \quad (\text{F.3})$$

Nous avons donc finalement une équivalence entre transformée en z et transformée de Laplace par la transformation

$$s \mapsto \frac{2(z - 1)}{T_S(z + 1)} \quad (\text{F.4})$$

A partir de (F.3), on déduit une équation de récurrence du type

$$\alpha_0 y_n + \alpha_1 y_{n-1} + \dots + \alpha_{K-1} y_{n-K} = \beta_0 u_n + \beta_1 u_{n-1} + \dots + \beta_{K-1} u_{n-K}, \quad (\text{F.5})$$

où $u_n = \psi(y_n)$, sans retard à la différence du modèle à temps discret exposé en F.3.1. Le retard sera alors introduit artificiellement dans le système, $u_n = \psi(y_{n-1})$ (au lieu de $u_n = \psi(y_n)$) dans l'équation (F.5). Dans ce cas les $\{a_k\}$ et $\{b_l\}$ calculés par discrétisation ne sont pas exacts pour le système avec retard, mais il peuvent constituer une initialisation sensée pour une identification paramétrique. Cela nous permet de plus d'avoir une idée des ordres du filtre du système de Lur'e à temps discret.

F.2 Bases de fonctions pour la non-linéarité ψ

Plusieurs classes de fonctions sont proposées.

F.2.1 Polynômes de Tchebychev

On définit les polynômes de Tchebychev (de type 1)

$$\begin{cases} T_0(x) & = & 1 \\ T_1(x) & = & x \\ T_{n+1}(x) & = & 2xT_n(x) - T_{n-1}(x) \end{cases} .$$

F.2.2 Smoothed Piecewise linear Functions

Les "Smoothed Piecewise linear Functions" ont une définition similaire à celle des fonctions continues affines par morceaux.

Fonctions continues affines par morceaux : Nous considérons N segments, donc $N-1$ points de rupture x_k , N paramètres de pente σ_k , et un paramètre de décalage vertical y_0 , ce qui donne la définition

$$y = \sum_{k=1}^{N-1} \sigma_k |x - x_k| + \sigma_N |p(x - x_N)| + y_0, \text{ avec } p(x) = \begin{cases} x & \text{pour } x > 0 \\ 0 & \text{sinon} \end{cases} .$$

Ces fonctions sont faciles à implémenter (ψ est de plus une combinaison linéaire des ψ_j), mais il est difficile de savoir comment faire varier les coefficients $\{c_j\}_{j=1\dots J}$ pour que ψ évolue de façon satisfaisante, ce qui est gênant dans l'algorithme d'optimisation numérique.

Smoothed Piecewise linear Functions : Nous remplaçons l'opérateur valeur absolue par la fonction $s_{\tilde{a}}(x) = \frac{2x}{\pi} \arctan(\tilde{a}x)$ où \tilde{a} est appelé paramètre d'attraction. On ajoute donc $N-1$ paramètres d'attraction, et la définition de y devient

$$y = \sum_{k=1}^{N-1} \sigma_k s_{\tilde{a}_k}(x - x_k) + \sigma_N s_{\tilde{a}_{N-1}}(p(x - x_N)) + y_0 .$$

Remarquons que nous n'avons alors plus véritablement une base de fonctions dont les paramètres C seraient les coefficients, comme dans le cas des polynômes. Par contre, nous disposons d'un homomorphisme entre les ordonnées des points de rupture et les coefficients σ_k , ce qui permet d'utiliser les points de rupture, les attractions ainsi que σ_1 et σ_N comme paramètres. Dans ce cas, le rapport entre paramètres et la forme de la fonction est intuitif. C'est donc cette paramétrisation que nous avons majoritairement utilisé.

F.3 Première tentative d'analyse de fonctions d'aire glottique

F.3.1 Modèle de Lur'e à temps discret

F.3.1.1 Modèle de synthèse paramétrique

Le modèle de Lur'e (figure F.2) est un système bouclé composé d'un filtre linéaire et d'une fonction non linéaire scalaire ψ [Kha02]. Les notations suivantes seront utilisées dans cette section.

- $n \in [0, N - 1]$: temps discret,
- y_n : sortie du système,
- r_n : entrée du système, nous choisissons ici $r_n = 0$,
- u_n : signal intermédiaire scalaire (entrée du filtre),
- $G(z)$: filtre ARMA, $G(z) = \frac{\sum_{l=0}^{L-1} b_l z^{-l}}{1 + \sum_{k=1}^K a_k z^{-k}}$,
- ψ : fonction non linéaire scalaire. Nous choisissons une fonction paramétrique, paramétrées par le vecteur de paramètres $C = [\{c_j\}_{j \in [1, J]}]^T$.

La mise en équation du système de Lur'e à temps discret présenté en figure F.2 donne

$$\begin{cases} y_n = -\sum_{k=1}^K a_k y_{n-k} + \sum_{l=0}^{L-1} b_l u_{n-l}, \\ w_n = \psi(v_n) = \sum_{j=1}^J c_j \psi_j(y_n), \\ v_n = y_{n-1}, \\ u_n = -w_n + r_n, \end{cases} \quad (\text{F.6})$$

avec les conditions initiales

$$\{y_k\}_{k \in [-K, -1]}. \quad (\text{F.7})$$

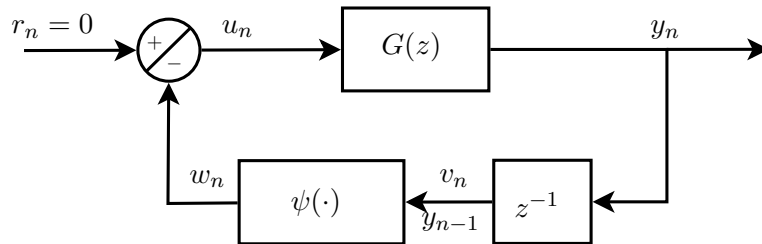


FIGURE F.2 – Structure générale d'un système de Lur'e

Les paramètres du système sont donc

- conditions initiales $\{y_k\}_{k \in [-K, -1]}$,
- coefficients $\{a_k\}_{k \in [1, K]}$ et $\{b_l\}_{l \in [1, L]}$ (nous noterons par la suite A le vecteur contenant les paramètres a_k suivis des paramètres $-b_k$), ordres K et L ,
- paramètres $\{c_j\}_{j \in [1, J]}$ et J .

Nous choisissons dans un premier temps pour la base $\{\psi_j\}$ les fonctions linéaires définies par morceaux adoucies (smoothed piecewise linear functions, cf annexe F.2).

Le système d'équations (F.6) peut se mettre sous la forme de représentations d'états

$$\begin{cases} X_{n+1} = \tilde{A}X_n + \tilde{b}u_n \\ y_n = \tilde{c}^\top X_n \\ u_n = \psi(y_{n-1}) \end{cases}, \quad (\text{F.8})$$

avec

$$\tilde{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_K & -a_{K-1} & -a_{K-2} & \cdots & -a_1 \end{bmatrix}, \quad X_n = \begin{bmatrix} y_{n-K+1} \\ \vdots \\ y_n \end{bmatrix},$$

$$\tilde{b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad \text{et} \quad \tilde{c}^\top = [0 \quad \cdots \quad 0 \quad b_{L-1} \quad \cdots \quad b_0].$$

F.3.1.2 Quelques exemples de synthèse

Nous présentons ici à titre d'exemple quelques signaux produits avec un système de Lur'e.

En jouant librement avec les coefficients du filtre et la fonction non linéaire, il est facile de produire des signaux chaotiques. La figure F.3 montre un exemple de synthèse avec des coefficients ajustés de manière aléatoire. Ce signal ne présente pas de pseudo-périodicité et semble très irrégulier.

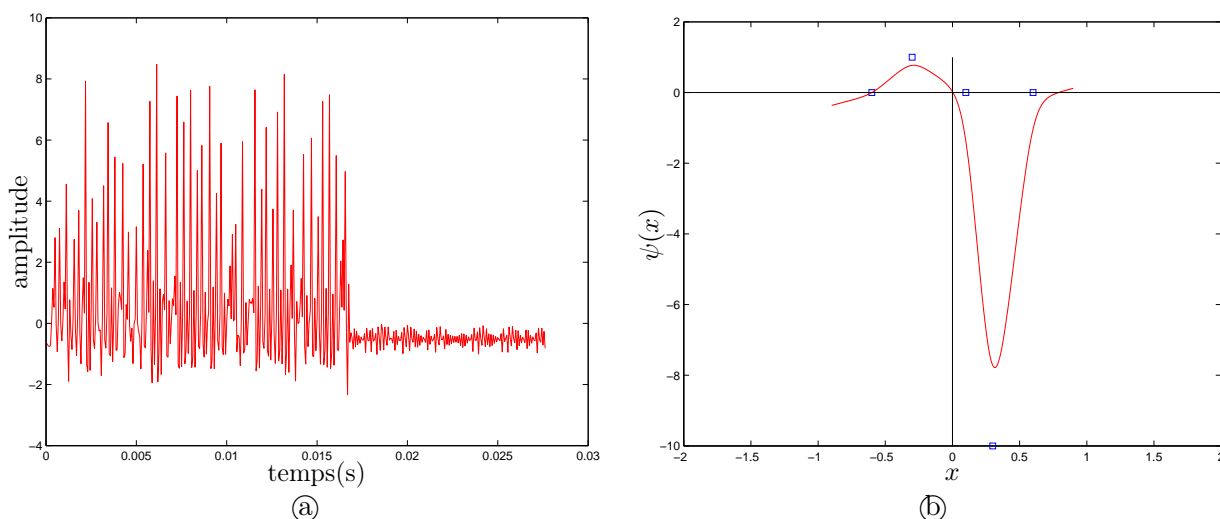
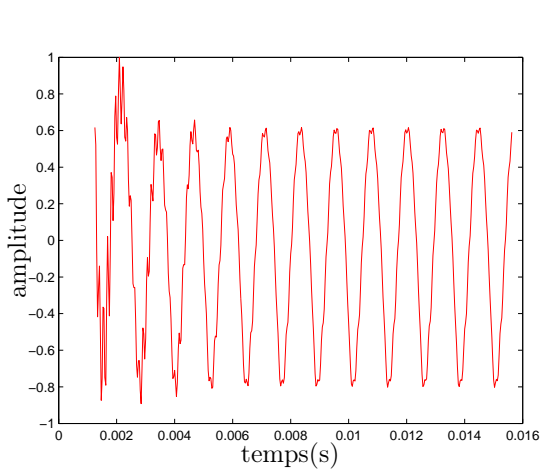
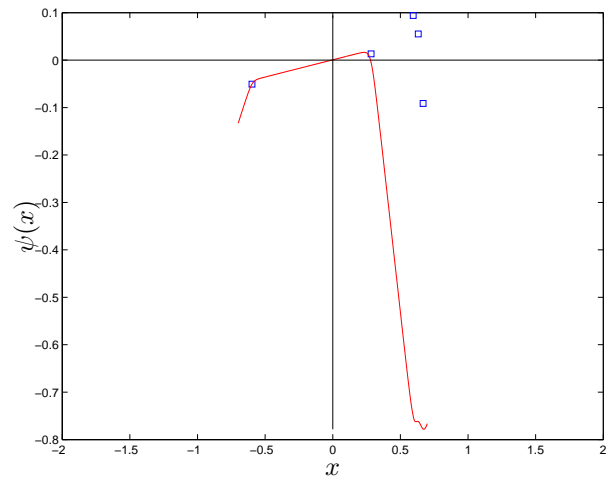


FIGURE F.3 – Synthèse par système de Lur'e, paramètres aléatoires. (a) Signal de synthèse. (b) - Fonction non linéaire.

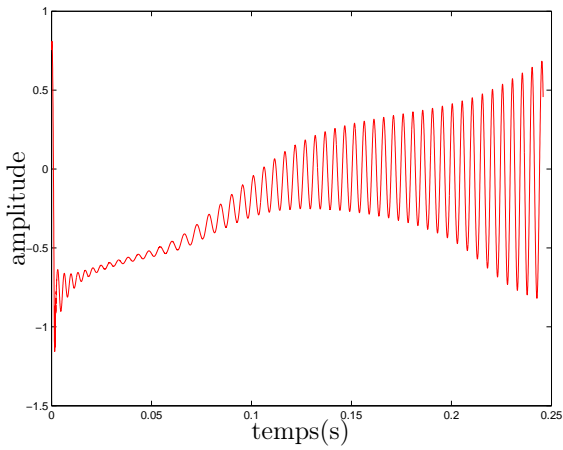
En recherchant de manière automatique des paramètres permettant d'obtenir un signal plus régulier, nous pouvons observer plusieurs comportements intéressants : des signaux pseudo-périodiques, des oscillations convergentes et des oscillations divergentes (figure F.4).



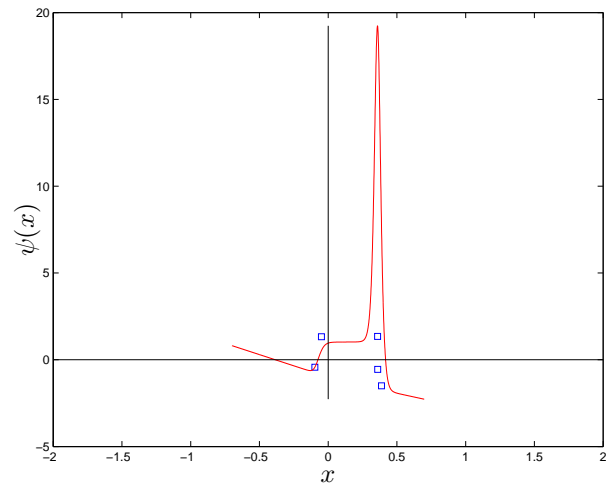
(a)



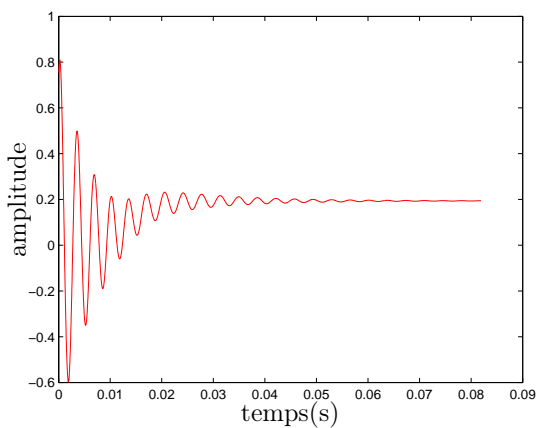
(b)



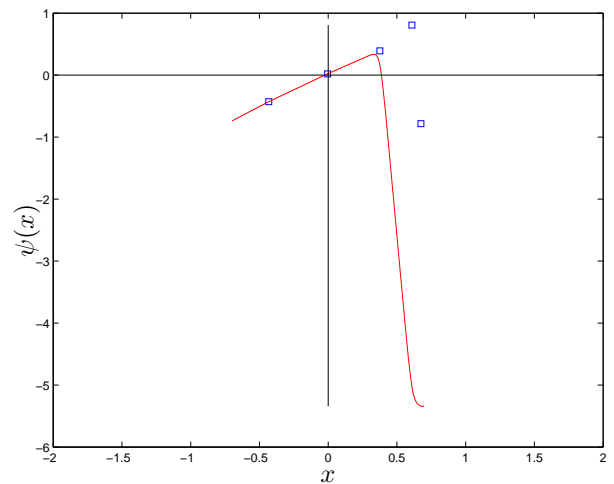
(c)



(d)



(e)



(f)

FIGURE F.4 – Synthèse pseudo-périodique par système de Lur'e. (a),(c),(e) Signaux de synthèse. (b),(d),(f) Fonctions non linéaires.

En étudiant la suite linéaire récurrente $y_n = 2\cos(\Theta)y_{n-1} - y_{n-2}$, on peut calculer l'expression générale $y_n = A\cos(n\Theta) + B\sin(n\Theta)$, où A et B dépendent de y_0 et y_1 . Il s'agit d'une suite périodique. Si nous définissons un système de Lur'e en prenant $\psi(x) = x$ et $u_n = -\psi(y_{n-1})$, nous obtenons $y_n = 2\cos(\Theta)\psi(y_{n-1}) - y_{n-2} = -y_{n-2} - 2\cos(\Theta)u_n$, où ψ est l'identité. En modifiant légèrement ψ pour l'écartier de l'identité (figure F.5- (a)) nous pouvons observer des comportements chaotiques (figure F.5- (b)).

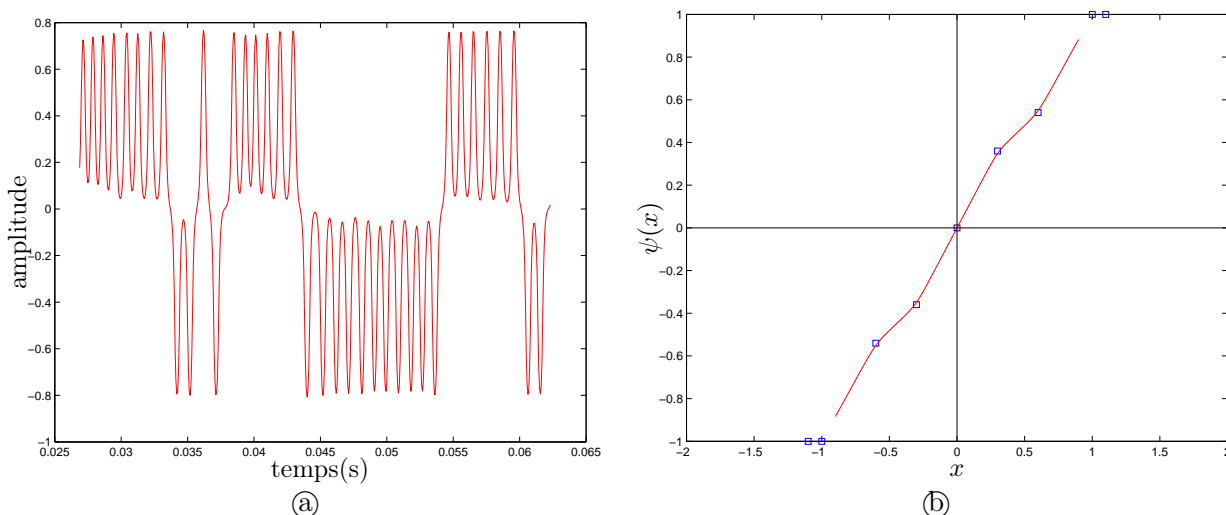


FIGURE F.5 – Synthèse par système de Lur'e chaotique. (a) Synthèse. (b) Fonction non linéaire.

F.3.2 Un algorithme d'identification des paramètres du système de Lur'e

L'idée développée dans [De 01] consiste à estimer les paramètres de G et de ψ de sorte que le système en boucle ouverte (figure F.6) reproduise l'identité pour le signal \hat{y} étudié mais sans que G et ψ ne dégénère vers l'identité. On cherche ainsi le combinaison d'une déformation linéaire à mémoire et d'une déformation non-linéaire sans mémoire, non triviale, qui se compense uniquement pour le signal d'intérêt. On espère ensuite que le système bouclé va régénérer des signaux qualitativement proches de la classe d'intérêt et isoler un régime type au sein de la diversité des signaux d'aire glottique observés dans la base de données.

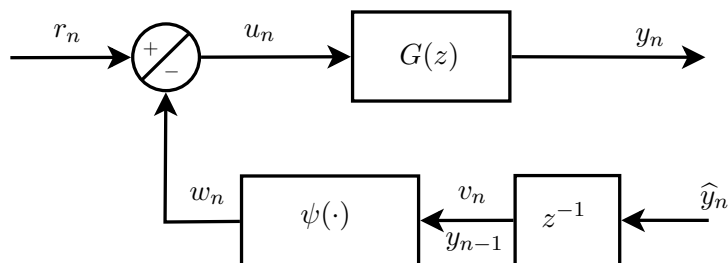


FIGURE F.6 – Système de Lur'e en boucle ouverte.

[De 01, p.128] propose pour ce problème un algorithme associant une identification linéaire pour les coefficients du filtre étant donné C et une optimisation pour les paramètres de la fonction non linéaire. Il s'agit de construire une identification linéaire des paramètres du filtre pour un C donné et d'optimiser une fonction d'objectif (fonction de C) prenant en compte

une mesure de la qualité de l'identification linéaire ($f_I(C)$) qui est d'autant plus faible que les paramètres C sont proches de ceux du système de Lur'e recherché, et

des contraintes sur les paramètres ($f_{P_j}(C)$) qui permettent d'éviter la dégénérescence de la fonction non-linéaire en une fonction linéaire et celle du filtre en retard pur.

F.3.2.1 Identification linéaire

Les coefficients $\{a_k\}_{k \in [1, K]}$ et $\{b_l\}_{l \in [1, L]}$ du filtre sont estimés pour un C donné par une technique d'identification linéaire classique (ARX, ARMAX, Box-Jenkins, etc. [Lju99]) avec un critère de moindres carrés. Ici, le signal d'entrée est u_n , et le signal de sortie y_n , avec $y_n = -\sum_{k=1}^K a_k y_{n-k} + \sum_{l=0}^{L-1} b_l u_{n-l}$. L'identification linéaire s'écrit donc

$$\underbrace{\begin{pmatrix} y_K \\ y_{K+1} \\ \vdots \\ y_N \end{pmatrix}}_{Y_N} = \underbrace{\begin{pmatrix} y_{K-1} & y_{K-2} & \cdots & y_0 & u_K & u_{K-1} & \cdots & u_{K-L+1} \\ y_K & y_{K-1} & \cdots & y_1 & u_{K+1} & u_K & \cdots & u_{K-L+2} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-K} & u_N & u_{N-1} & \cdots & u_{N-L+1} \end{pmatrix}}_{T_N} \underbrace{\begin{pmatrix} -a_1 \\ \vdots \\ -a_K \\ b_1 \\ \vdots \\ b_L \end{pmatrix}}_{-A}. \quad (\text{F.9})$$

$$\Rightarrow A = -\underbrace{(T_N^\top T_N)^{-1} T_N^\top}_{T_N^\dagger} Y_N. \quad (\text{F.10})$$

F.3.2.2 Mesure de la qualité de l'identification linéaire

Par prédiction

[De 01] propose la mesure suivante.

$$\begin{cases} f_I(C) &= \frac{1}{r_e} \sum_{i=1}^{r_e} e'_{k_e}(i)^2 && \text{(erreur quadratique moyenne)} \\ e_{k_e}(i) &= y_{i+k_e|i} - \hat{y}_{i+k_e} && \text{(erreur comprenant une erreur statique systématique)} \\ e'_{k_e}(i) &= e_{k_e}(i) - \mathbb{E}[e_{k_e}(n)] && \text{(erreur centrée)} \end{cases}$$

où r_e et k_e sont des nombres d'échantillons correspondant à des multiples de la pseudo-période T_p , soit $r_e T_s = n T_p$ et $k_e T_s = m T_p$ (où T_s est la fréquence d'échantillonnage, et n et m sont deux entiers choisis arbitrairement). $y_{i+k_e|i}$ est l'estimation de la valeur du signal de sortie du système de Lur'e calculé par prédiction linéaire en connaissant u_n jusqu'à l'instant $i + k_e$ et \hat{y}_n jusqu'à l'instant i . Le choix de $e'_{k_e}(i) = e_{k_e}(i) - \mathbb{E}[e_{k_e}(n)]$ plutôt que $e_{k_e}(i)$ dans le calcul de l'erreur quadratique moyenne est, d'après [De 01], motivé par le fait que pour des conditions initiales nulles dans la prédiction, l'erreur statique systématique ($\mathbb{E}[e_{k_e}(n)]$) sera nulle.

Avec la matrice de corrélation

Une autre méthode¹ pour obtenir une mesure de la qualité de l'identification linéaire (cf [De 01, p.134]) est de calculer l'intervalle (ou tolérance) des paramètres obtenus par identification linéaire. En appelant θ le vecteur de paramètres, et $\bar{\theta}$ le vecteur de paramètres moyens, la matrice Θ de covariance des paramètres est définie par

$$\Theta = \mathbb{E} \left[(\theta - \bar{\theta})(\theta - \bar{\theta})^\top \right].$$

La tolérance vaut alors

$$R = \Theta / (\bar{\theta} \bar{\theta}^\top).$$

La qualité de l'identification linéaire est finalement donnée par l'inverse de norme infinie de R . D'après [WP94], Θ peut être approchée par la matrice $(T_N^\top T_N)^{-1}$ définie dans (F.10). Cette méthode a donc l'avantage d'être plus rapide que la précédente.

F.3.2.3 Contraintes sur les paramètres

Les f_{P_j} expriment les contraintes sur les paramètres et seront du type

$$f_{P_j}(C) = \frac{1}{g^2(C)}, \quad (\text{F.11})$$

$$f_{P_j}(C) = \frac{1}{|a_i|}, \quad (\text{F.12})$$

$$f_{P_j}(C) = \frac{1}{|\mu_i|}, \quad (\text{F.13})$$

où

- (F.11) pénalise les régressions de ψ en fonction linéaire, $g(C) = 0$ traduisant la linéarité de ψ ,
- (F.12) pénalise les dégénérescences d'ordre du filtre,
- (F.13) pénalise les régressions du filtre en retard pur (les μ_i sont les pôles du filtre).

F.3.2.4 Optimisation des paramètres

La fonction d'objectif est minimisée par rapport à C à l'aide d'une méthode d'optimisation numérique. Nous avons choisi la méthode du recuit simulé pour des raisons pratiques. A chaque tirage de C , la valeur de la fonction d'objectif est évaluée comme suit.

- Calcul des valeurs de $u_n = \psi(\hat{y}_{n-1})$,
- identification des coefficients $\{a_k\}_{k \in [1, K]}$ et $\{b_l\}_{l \in [1, L]}$ du filtre (étant donné u et \hat{y}),
- mesure de la qualité de l'identification linéaire $f_I(C)$,
- calcul des contraintes f_{P_j} ,
- calcul de la valeur de la fonction d'objectif $f_{obj}(C) = f_I(C) + \sum_j f_{P_j}(C)$.

1. Les deux méthodes de F.3.2.2 et F.3.2.2 ont été testées et chacune peut être intéressante suivant les cas.

F.3.2.5 Premiers résultats

Des exemples de signaux re-synthétisés sont présentés en figures F.7 et F.8. Les résultats obtenus sont assez éloignés des signaux cibles. Plusieurs explications sont possibles.

Algorithme mal adapté. La fonction d'objectif telle qu'elle a été définie ne dépend pas directement de l'erreur de resynthèse. Nous pouvons nous poser la question de la pertinence de cette définition. Cependant, la pertinence de cette méthode est montrée dans [De 01] pour la resynthèse de signaux proche des nôtres.

Paramètres de l'optimisation numérique. Les paramètres du recuit simulé (contrôlant l'initialisation et l'évolution de la température, la détermination de paramètres voisins, etc.) sont difficiles à déterminer, mais on peut cependant raisonnablement penser que la souplesse de recherche de l'algorithme peut compenser un choix imparfait dans les paramètres. On écarte donc aussi cette explication.

Manque de précision sur les paramètres estimés. Il apparaît, en examinant l'évolution de l'algorithme de recuit simulé, que les signaux synthétisés sont très sensibles aux paramètres du modèle. Il est donc possible que notre méthode de recherche de paramètres, couplée à un paramétrage de la fonction d'objectif impliquant un grand nombre de paramètres, ne soit pas assez précise pour trouver des paramètres permettant une synthèse satisfaisante.

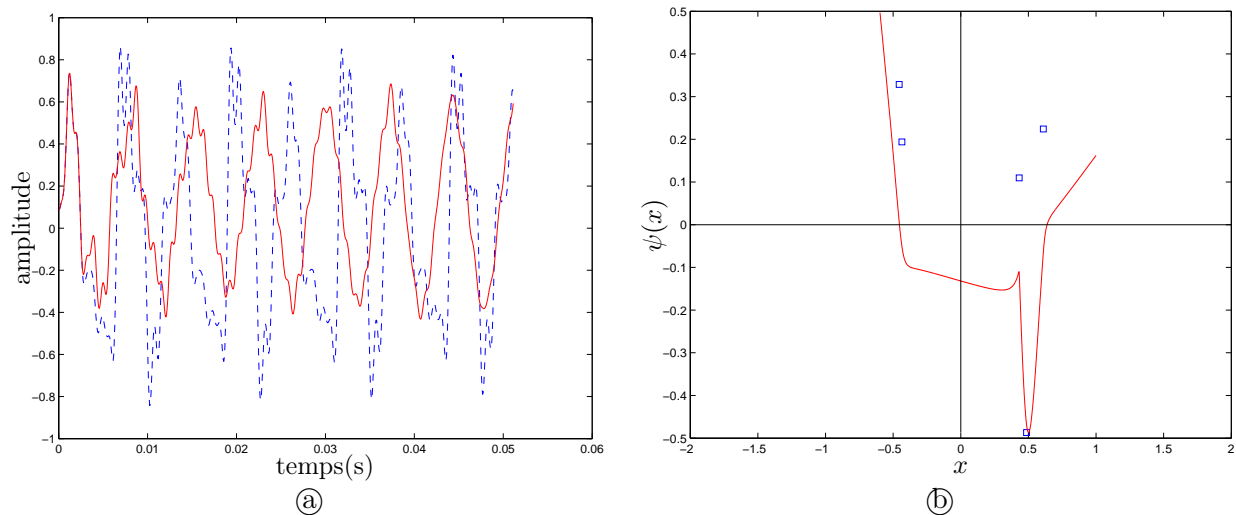


FIGURE F.7 – Analyse-resynthèse, premiers exemples 1. (a) signal de synthèse (le signal synthétisé est en rouge, le signal cible est en pointillés bleu). (b) Fonction non linéaire.

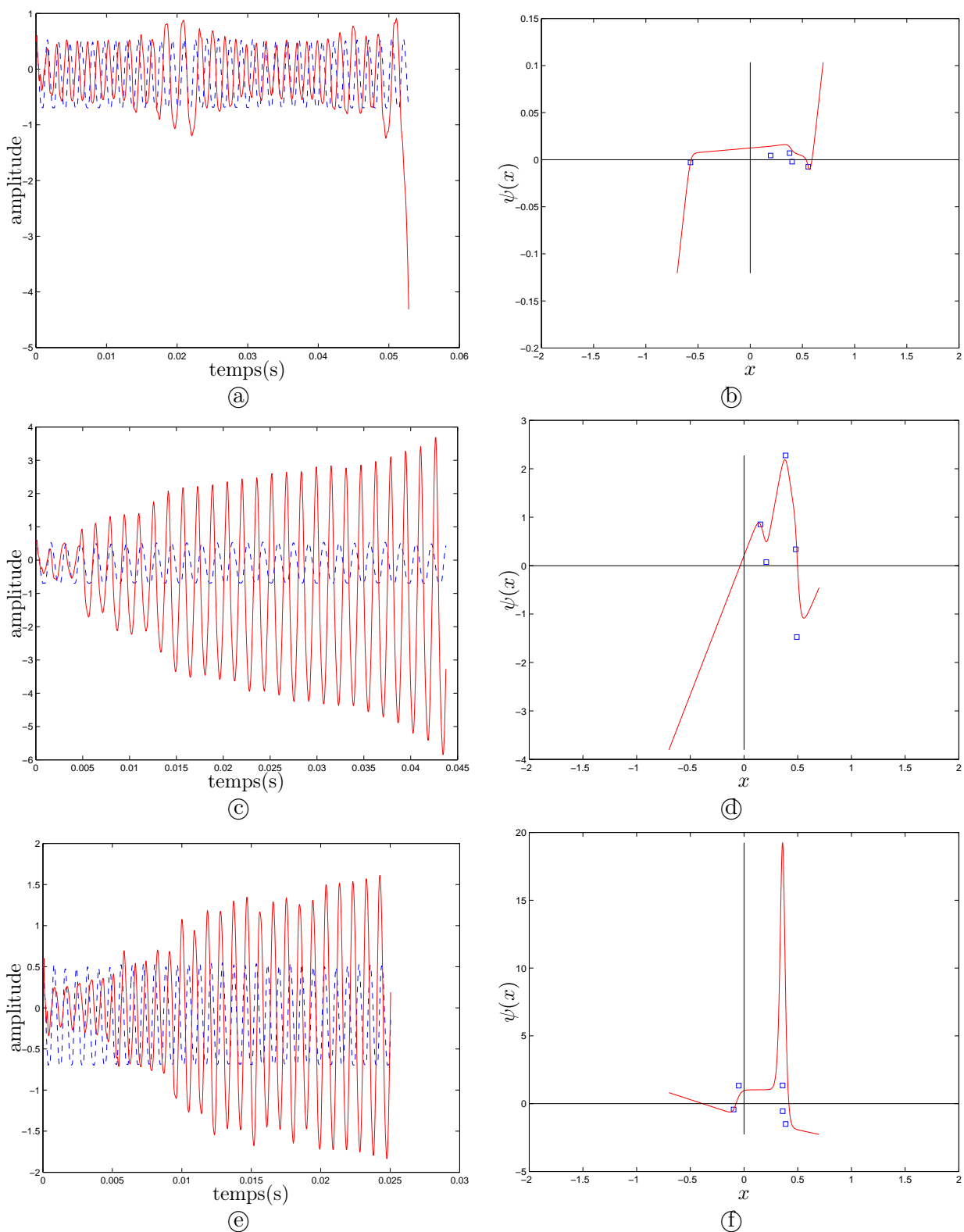


FIGURE F.8 – Analyse-resynthèse, premiers exemples 2. (a),(c),(e) signaux de synthèse (le signal synthétisé est en rouge, le signal cible est en pointillés bleu). (b),(d),(f) fonctions non linéaires correspondant respectivement à (a),(c) et (e).



Annexe G

Propagation dans les tubes acoustiques

G.1 Modèle général et matrices de transfert acoustique

Dans un tube axisymétrique de forme quelconque représenté par son rayon $R(z)$ (où z est la variable longitudinale, voir la figure G.1), nous notons $p(z, r, t)$, $u(z, r, t)$ et $v(z, r, t)$ respectivement la pression, le débit et la vitesse dans le direction des z positifs dans le tube à la position z et à l'instant t .

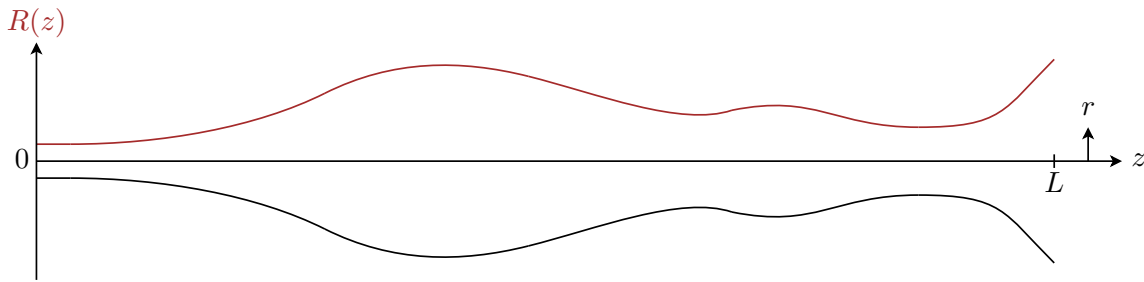


FIGURE G.1 – Tube axisymétrique, cas général.

Considérons un fluide idéal sans source et soumis à aucune force. La propagation linéaire des ondes acoustiques est gouvernée par les équations de conservation de la masse et du mouvement [CK08]

$$\frac{\partial p}{\partial t}(z, r, t) + \rho c^2 \operatorname{div} \vec{v}(z, r, t) = 0 \quad \text{et} \quad (\text{G.1})$$

$$\frac{\partial \vec{v}}{\partial t}(z, r, t) + \frac{1}{\rho} \operatorname{grad} p(z, r, t) = 0, \quad (\text{G.2})$$

où ρ est la densité de l'air et c la vitesse du son dans l'air. On déduit de (G.1 - G.2) l'équation des ondes (équation de d'Alembert) pour la pression (une équation analogue est obtenue pour le débit)

$$\square p(z, r, t) = \Delta p(z, r, t) - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}(z, r, t) = 0. \quad (\text{G.3})$$

Un modèle de propagation unidimensionnel (dépendant d'une unique variable spatiale x) a été

proposé par Lagrange [Lag61] et Bernoulli [Ber62] pour un profil régulier $R(x)$ et a donné lieu à l'équation de Webster [Web19, Eis67] (ou équation des pavillons [CK08]), simplification de (G.3),

$$\left(\frac{\partial^2}{\partial x^2} + 2 \frac{R'(x)}{R(x)} \frac{\partial}{\partial x} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) p(x, t) = 0. \quad (\text{G.4})$$

De plus, nous savons [Cha01, p.58] que dans ce cadre, la propagation est caractérisée par une matrice de transfert acoustique correspondant à un quadripôle réciproque symétrique,

$$\begin{bmatrix} P(z_1) \\ U(z_1) \end{bmatrix} = \begin{bmatrix} \cosh(\Gamma) & Z_c \sinh(\Gamma) \\ Z_c^{-1} \sinh(\Gamma) & \cosh(\Gamma) \end{bmatrix} \begin{bmatrix} P(z_n) \\ U(z_2) \end{bmatrix}. \quad (\text{G.5})$$

Remarque 10. La loi de composition permet d'écrire, pour n tronçons identiques en cascade

$$\begin{bmatrix} P(z_0) \\ U(z_0) \end{bmatrix} = \begin{bmatrix} \cosh(n\Gamma) & Z_c \sinh(n\Gamma) \\ Z_c^{-1} \sinh(n\Gamma) & \cosh(n\Gamma) \end{bmatrix} \begin{bmatrix} P(z_n) \\ U(z_n) \end{bmatrix}.$$

Nous verrons dans la suite de ce chapitre qu'à chacun des cas correspondant à notre cadre d'étude correspond une impédance itérative Z_c et une fonction Γ particulières. Nous écrirons ces cas sous forme de matrice de transfert pour les variables d'Euler p et u , et nous étudierons le changement de variable en ondes progressives.

G.2 Cas des tubes droits

G.2.1 Matrice de transfert acoustique dans un tube droit sans pertes

Dans un tube droit, c'est à dire un tube pour lequel $R(z) = R$ est constant, les ondes se propagent en ondes planes [Lag61, Ber62] et G.4 devient

$$\frac{\partial^2 p}{\partial z^2}(z, t) - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}(z, t) = 0 \quad \text{ou} \quad \frac{\partial^2 u}{\partial z^2}(z, t) - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}(z, t) = 0 \quad (\text{G.6})$$

qui se réécrit par factorisation

$$\left(\frac{\partial}{\partial z} + \frac{1}{c} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial z} - \frac{1}{c} \frac{\partial}{\partial t} \right) u(z, t) \quad \text{ou} \quad \left(\frac{\partial}{\partial z} + \frac{1}{c} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial z} - \frac{1}{c} \frac{\partial}{\partial t} \right) p(z, t).$$

On en déduit alors la solution générale en ondes progressives

$$\begin{cases} u(z, t) &= u^+ \left(t - \frac{z}{c} \right) - u^- \left(t + \frac{z}{c} \right) \\ p(z, t) &= p^+ \left(t - \frac{z}{c} \right) + p^- \left(t + \frac{z}{c} \right) \end{cases} \quad (\text{G.7})$$

D'après (G.1) et (G.7), on déduit

$$\begin{cases} p^+ \left(t - \frac{z}{c} \right) &= \frac{\rho c}{S} u^+ \left(t - \frac{z}{c} \right) \\ p^- \left(t + \frac{z}{c} \right) &= -\frac{\rho c}{S} u^- \left(t + \frac{z}{c} \right) \end{cases}. \quad (\text{G.8})$$

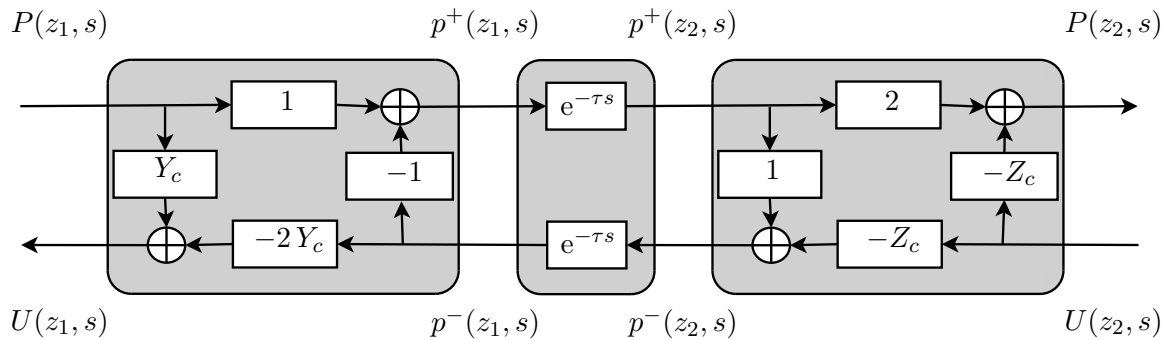
où S représente la section du tube perpendiculaire à l'axe Oz ($S = \pi R^2$).

D'après (G.7-G.8), la propagation dans un tube droit entre un point d'abscisse z_1 et un point d'abscisse z_2 s'écrit, dans le domaine de Laplace,

$$\begin{bmatrix} P(z_2, s) \\ U(z, s) \end{bmatrix} = \begin{bmatrix} \cosh(-\tau s) & Z_c \sinh(-\tau s) \\ Z_c^{-1} \sinh(-\tau s) & \cosh(-\tau s) \end{bmatrix} \begin{bmatrix} P(z_1, s) \\ U(z_1, s) \end{bmatrix}, \quad (\text{G.9})$$

avec $\tau = \frac{z_2 - z_1}{c}$ et $Z_c = \frac{\rho c}{S}$. Nous retrouvons bien la forme (G.5), avec $\Gamma(s) = e^{-\tau s}$.

La représentation quadripolaire des équations de propagation permet un formalisme plus adapté à la simulation numérique des guides d'ondes. Dans ce formalisme, (G.9) peut s'écrire pour les ondes progressives p^\pm définies en (G.7) [Mig09]



G.2.2 Concaténation de tubes droits

Pour représenter un tube quelconque en concaténation, nous discrétisons son profil $R(z)$ comme représenté en figure G.2, en notant R_m et $S_m = \pi R_m^2$ respectivement le rayon et la section de chaque section cylindrique.

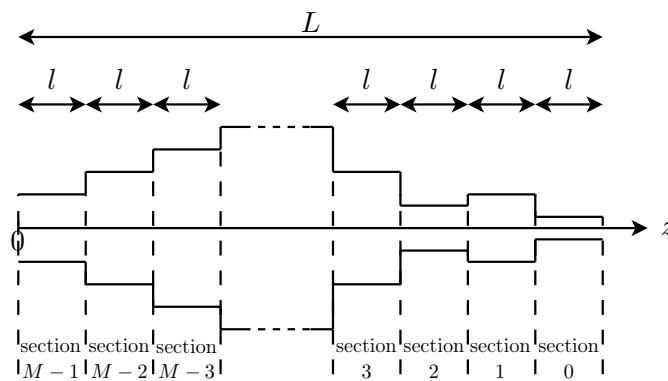


FIGURE G.2 – Modèle de conduit en tronçons de tubes droits

Dans chaque section, la propagation des ondes est régie par les matrices de transfert proposées précédemment. Nous définissons pour chaque section les ondes progressives de débit $u_m(z, t) = u_m^+(t - z/c) u_m^-(t - z/c)$ pour $z \in [-l/2, l/2]$. A l'interface entre la section m et la section $m - 1$,

la continuité de pression et de débit amène aux égalités

$$\begin{cases} u_m^+(t - \tau) = \frac{u_{m-1}^+(t + \tau) - \mu_m u_{m-1}^-(t - \tau)}{1 - \mu_m} \\ u_m^-(t + \tau) = \frac{u_{m-1}^-(t - \tau) - \mu_m u_{m-1}^+(t + \tau)}{1 - \mu_m} \end{cases}, \quad (\text{G.10})$$

où

$$\mu_m = \frac{S_m - S_{m-1}}{S_m + S_{m+1}} = \frac{Z_{m-1} - Z_m}{Z_{m-1} + Z_m} \text{ et } \tau = \frac{l}{2c}. \quad (\text{G.11})$$

En définissant

$$t_m = 2(m+1)\tau, \quad c_m = \prod_{i=1}^m (1 - \mu_i) \text{ et } \begin{cases} y_m^+(t) = c_m u_m^+(t + \tau - t_m) \\ y_m^-(t) = c_m u_m^-(t - \tau - t_m) \end{cases}, \quad (\text{G.12})$$

le système (G.10) se réécrit, en temporel et en transformée de Laplace,

$$\begin{cases} y_m^+(t) = y_{m-1}^+(t) - \mu_m y_{m-1}^-(t) \\ y_m^-(t + 4\tau) = y_{m-1}^-(t) - \mu_m y_{m-1}^+(t) \end{cases} \text{ et} \quad (\text{G.13})$$

$$\begin{cases} Y_m^+(s) = Y_{m-1}^+(s) - \mu_m Y_{m-1}^-(s) \\ Y_m^-(s) = e^{-4\tau s} (Y_{m-1}^-(s) - \mu_m Y_{m-1}^+(s)) \end{cases}. \quad (\text{G.14})$$

En définissant $H_m^{++}(s)$, $H_m^{-+}(s)$, $H_m^{+-}(s)$ et $H_m^{--}(s)$ tels que

$$\begin{cases} Y_m^+(s) = H_m^{++}(s)Y_0^+(s) + H_m^{-+}(s)Y_0^-(s)e^{4\tau s} \\ Y_m^-(s) = H_m^{+-}(s)Y_0^+(s) + H_m^{--}(s)Y_0^-(s)e^{4\tau s} \end{cases}, \quad (\text{G.15})$$

le système (G.14) devient

$$\begin{cases} H_m^{++}(s) = H_{m-1}^{++}(s) - \mu_m H_{m-1}^{+-}(s) \\ H_m^{-+}(s) = H_{m-1}^{-+}(s) - \mu_m H_{m-1}^{--}(s) \\ H_m^{+-}(s) = e^{-4\tau s} (H_{m-1}^{+-}(s) - \mu_m H_{m-1}^{++}(s)) \\ H_m^{--}(s) = e^{-4\tau s} (H_{m-1}^{--}(s) - \mu_m H_{m-1}^{-+}(s)) \end{cases}, \quad (\text{G.16})$$

avec $H_0^{++}(s) = 1$, $H_0^{--}(s) = e^{-4\tau s}$ et $H_0^{-+}(s) = H_0^{+-}(s) = 0$.

Le système (G.16) nous permet finalement d'écrire les relations entrées/sorties du tuyau complet pour les ondes progressives de débit

$$\begin{cases} U_0^+(s) e^{-\tau s} = c_{M-1} e^{-2M\tau s} \frac{H_{M-1}^{--}(s)U_{M-1}^+(s)e^{\tau s} - H_{M-1}^{-+}(s)U_{M-1}^-(s)e^{-\tau s}}{H_{M-1}^{++}(s)H_{M-1}^{--}(s) - H_{M-1}^{-+}(s)H_{M-1}^{+-}(s)} \\ U_0^-(s) e^{\tau s} = c_{M-1} e^{-2M\tau s} \frac{H_{M-1}^{+-}(s)U_{M-1}^-(s)e^{-\tau s} - H_{M-1}^{++}(s)U_{M-1}^+(s)e^{\tau s}}{H_{M-1}^{++}(s)H_{M-1}^{--}(s) - H_{M-1}^{-+}(s)H_{M-1}^{+-}(s)} \end{cases}, \quad (\text{G.17})$$

ou bien

$$\left\{ \begin{array}{l} U_0^+(s) e^{-\tau s} = -\frac{H_{M-1}^{-+}(s)}{H_{M-1}^{++}(s)} U_0^-(s) e^{\tau s} + \frac{c_{M-1} e^{-2M\tau s}}{H_{M-1}^{++}(s)} U_{M-1}^+(s) e^{\tau s} \\ U_{M-1}^-(s) e^{-\tau s} = \frac{H_{M-1}^{+-}(s)}{H_{M-1}^{++}(s)} U_{M-1}^+(s) e^{\tau s} + \\ \frac{H_{M-1}^{++}(s) H_{M-1}^{--}(s) - H_{M-1}^{+-}(s) H_{M-1}^{-+}(s)}{c_{M-1} e^{-2M\tau s} H_{M-1}^{++}(s)} U_0^-(s) e^{\tau s} \end{array} \right. , \quad (G.18)$$

ou encore

$$\left\{ \begin{array}{l} U_{M-1}^+(s) e^{\tau s} = \frac{H_{M-1}^{++}(s)}{c_{M-1} e^{-2M\tau s}} U_0^+(s) e^{-\tau s} + \frac{H_{M-1}^{-+}(s)}{c_{M-1} e^{-2M\tau s}} U_0^-(s) e^{\tau s} \\ U_{M-1}^-(s) e^{-\tau s} = \frac{H_{M-1}^{+-}(s)}{c_{M-1} e^{-2M\tau s}} U_0^-(s) e^{\tau s} + \frac{H_{M-1}^{--}(s)}{c_{M-1} e^{-2M\tau s}} U_0^+(s) e^{-\tau s} \end{array} \right. . \quad (G.19)$$

G.3 Observations sur les fonctions H_t et Z_{in}

On remarque que $H_t(s)$ et $Z_{in}(s)$ peuvent s'écrire sous la forme

$$\left\{ \begin{array}{l} H_t(s) = \frac{2c_{M-1} e^{-2M\tau s}}{H_{M-1}^+(s) + H_{M-1}^-(s)} \\ Z_{in}(s) = \frac{\rho c}{S_{M-1}} \frac{H_{M-1}^+(s) - H_{M-1}^-(s)}{H_{M-1}^+(s) + H_{M-1}^-(s)} \end{array} \right. \quad (G.20)$$

avec

$$\left\{ \begin{array}{l} H_m^+(s) = H_{m-1}^+(s) - \mu_m H_{m-1}^-(s) \\ H_m^-(s) = e^{-4\tau s} (H_{m-1}^-(s) - \mu_m H_{m-1}^+(s)) \end{array} \right. , \quad \left\{ \begin{array}{l} H_0^+(s) = 1 \\ H_0^-(s) = e^{-4\tau s} \end{array} \right. . \quad (G.21)$$

De plus, on montre que

$$H_m^-(s) = e^{-4(m+1)\tau s} H_m^+(-s). \quad (G.22)$$

Après discrétisation du système par $z = e^{4\tau s}$, on note

$$\left\{ \begin{array}{l} H_m^+(z) = \sum_{k=1}^{m+1} a_k^m z^{-k} \\ H_m(z) = H_m^+(z) + H_m^-(z) = \sum_{k=1}^{m+1} h_k^m z^{-k} \end{array} \right. ,$$

on montre alors les propriétés suivantes valables $\forall m \geq 0$:

$$H_m^-(z) = \sum_{k=1}^{m+1} a_{m+1-k}^m z^{-k}, \quad (G.23)$$

$$\begin{cases} h_0^m &= 1 \\ h_k^m &= h_{m+1-k}^m = a_k^m + a_{m+1-k}^m \quad 1 \leq k \leq m \\ h_{m+1}^m &= 1 \end{cases} \quad , \quad (\text{G.24})$$

$$\begin{cases} a_0^m &= 1 \\ a_k^m &= a_k^{m-1} - \mu_m a_{m-k}^{m-1} \quad 1 \leq k \leq m-1 \\ a_m^m &= -\mu_m \end{cases} \quad . \quad (\text{G.25})$$



Notations

\mathbb{Z}	Ensemble des nombres entiers naturels
\mathbb{N}	Ensemble des nombres entiers relatifs
\mathbb{R}	Ensemble des nombres réels
\mathbb{C}	Ensemble des nombres complexes
\mathbb{S}	Cercle unité dans \mathbb{C}
<hr/>	
i	Unité imaginaire
\bar{z}	Conjugué de z
$\mathbb{1}_I(x)$	Fonction indicatrice de l'ensemble I
$\delta(n - d) = \delta_d(n)$	Signal de Kroenecker centré en d
<hr/>	
$S(f)$	Transformée de Fourier à temps continu du signal $s(t)$
$S(\omega)$	Transformée de Fourier à temps discret du signal $s(n)$
$S(k)$	Selon de contexte : Développement en série de Fourier du signal périodique $s(t)$ Transformée de Fourier discrète du signal périodique $s(n)$
$S(z)$	Transformée en Z du signal $s(n)$
$S(s)$	Transformée de Laplace du signal $s(n)$
$\hat{s}(n)$	Cepstre complexe du signal $s(n)$
$\check{s}(n)$	Cepstre différentiel du signal $s(n)$
<hr/>	
$*$	Opérateur de convolution
$\ln(x)$	Logarithme népérien défini sur \mathbb{R}^{+*}
$\log(x)$	Détermination principale du logarithme complexe défini sur \mathbb{C}^*
$\text{Log}(x)$	Logarithme complexe multivalué défini sur \mathbb{C}^*
$\text{Log}[F](x)$	Logarithme complexe multivalué de la fonction complexe F
$\log[F](x)$	Logarithme complexe de la fonction complexe F
$H(n)$	Fonction de Heaviside $H(n) = \begin{cases} 0 & \text{si } n \geq 0 \\ 1 & \text{si } n < 0 \end{cases}$
$\mathbb{E}[X]$	Espérance mathématique de la variable aléatoire X

Acronymes

CALM Causal-Anticausal Linear Model

CAN Convertisseur analogique-numérique

CNA Convertisseur numérique-analogique

DDS Direct Digital Synthesizer

EGG Électroglottographie

EIT Electrical Impedance Tomography

ELG Électrolabiographie

FPGA Field-Programmable Gate Array

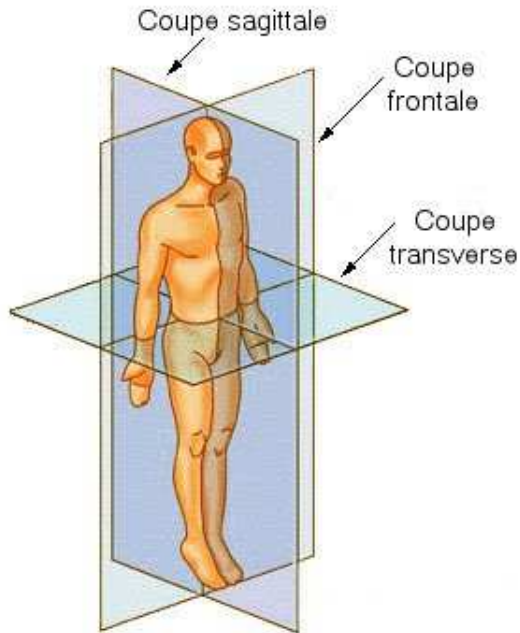
GCI Glottal CLosure Instant

LF Liljencrants-Fant

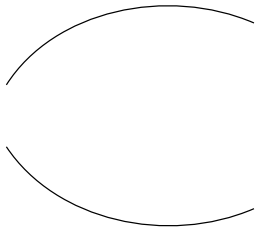
MADI Multichannel Audio Digital Interface

Expressions usuelles

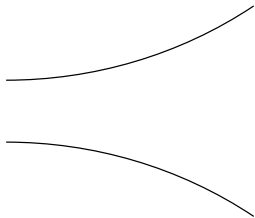
Système de référence en anatomie (Wikimedia Commons)



Tube à profil concave



Tube à profil convexe





Liste des tableaux

4.1	Performances of the LF parameters estimation algorithm	60
4.2	Numerical results	61
5.1	Test qualitatif de validation du principe de l'EIT sur une maquette de larynx simplifiée : résultats numériques	74
6.1	Modèles paramétriques d'impulsion glottique définis par morceaux	97
6.2	Pôles causaux/anticausaux stricts/non stricts	99
7.1	Vocabulaire de l'analyse cepstrale	116
7.2	Dictionnaire de cepstres utiles.	120
8.1	Performances de l'algorithme de représentation par modèle LF sur signaux de synthèse	154
8.2	Performances de l'algorithme de représentation par modèle LF sur signaux réels . . .	155
9.1	Catégorisation géométrique des profils en fonction de θ et Υ	183

Table des figures

1.1	Anatomie de l'appareil phonatoire humain	14
1.2	Anatomie du larynx	15
1.3	Anatomie des plis vocaux	16
1.4	Adduction et abduction des plis vocaux	17
1.5	Articulation crico-thyroïdienne	18
1.6	Mécanismes laryngés 1 et 2	20
1.7	Laryngoscope de Manuel Garcia	21
2.1	Exemple de signal vocal (voix parlée)	26
2.2	Classification des modèles de production de la voix : modèles de signaux et modèles physiques	27
2.3	Modèle source-filtre général	28
2.4	Modèle source-filtre simplifié	29
2.5	Modèle source-filtre, CALM	29
2.6	Modèle à deux masses d'Ishizaka et Flanagan	30
2.7	Modèle de Maeda	31
3.1	Images échographiques des plis vocaux	41
3.2	Image obtenue par échographie ultra-rapide (plan transverse)	41
3.3	Vidéo-endoscopie ultra-rapide des plis vocaux	42
3.4	Description d'un appareil EGG et exemple de signaux EGG et DEGG sur une période de vibration glottique	43
3.5	Résistance du larynx mesurée en fonction de la fréquence	43
4.1	Schematic description of an EGG device and typical glottal cycle on EGG and EGG derivative signals	50
4.2	Transparent mouthpiece equipped with copper foil shielding tape electrodes	51

4.3	Data acquisition set-up for brass instruments	52
4.4	Synchronous display of multimodal measurements of lips and glottis	53
4.5	Extracts of EI signals of the lips and the vocal folds	56
4.6	Electrical impedance derivative DEI signals of the lips and the vocal folds	57
4.7	Open area (OA) signals of the lips and the vocal-folds	59
4.8	One period of LF model	60
4.9	LF estimation examples	62
4.10	Geometrical parameters D and d on a lip frame	63
4.11	Opening Area OA versus geometrical parameters D and d	64
4.12	Electrical impedance EI versus opening area OA	65
5.1	Problème de Calderón	71
5.2	Principe de la tomographie d'impédance électrique	72
5.3	Simulation de l'EIT appliquée à la mesure de l'activité glottique, extrait de [SNK ⁺ 11]	73
5.4	Test qualitatif de validation du principe de l'EIT sur une maquette de larynx simplifiée	74
5.5	Électrodes à géométrie variable	76
5.6	Système complet	77
5.7	Modules d'excitation et de mesure	77
5.8	Architecture détaillée du dispositif	79
5.9	DDS et transformateur d'isolement	80
5.10	EKG à un canal réalisé à l'aide d'un module d'excitation	81
5.11	Module d'excitation réalisé	81
5.12	Carte de commutation 1 vers 32	82
5.13	Câble coaxial et triaxial	83
5.14	Blindage des pistes du routeur analogique par la méthode du <i>Stitching</i>	84
5.15	Carte fond de panier et huit cartes 1 vers 32	84
5.16	Accès au lien MADi physique : carte réalisée	85
5.17	Traitement séquentiel des ordres de configuration	87
5.18	Programme MAX/MSP pour le pilotage du routeur	87
6.1	Forme générique d'un pulse glottique paramétrique.	96
7.1	General scheme of the all-pole causal/anticausal decomposition algorithm	108

7.2	Illustration of the algorithm	111
7.3	Results of the algorithm	112
7.4	Results of the algorithm for real speech with high order model	113
7.5	Results of the algorithm for real speech with low order model	114
7.6	Partie imaginaire du logarithme de la transformée d'un signal à deux pôles	117
7.7	Surface de Riemann associée au logarithme et choix d'une branche	119
7.8	Partie imagine du logarithme de la transformée d'un signal tous-pôles	122
7.9	Premier exemple : signal temporel et cepstre différentiel	130
7.10	Premier exemple : estimation des paramètres et reconstruction	131
7.11	Effet du fenêtrage 1	132
7.12	Effet du fenêtrage 2	133
7.13	Effet d'un bruit de puissance 10^{-5}	134
7.14	Effet d'un bruit de puissance 10^{-4}	135
7.15	Tests sur signaux réels	136
8.1	Suppression de la composante BF des signaux d'aire glottique	147
8.2	Fonctions d'aire glottique $A_g(t)$ extraites du corpus USC_2008_02.	148
8.3	Fonctions d'aire glottique dérivées $A'_g(t)$ extraites du corpus USC_2008_02.	149
8.4	Initialisation de l'estimation des paramètres LF	151
8.5	Optimisation des paramètres LF sur deux périodes consécutives	152
8.6	Représentation LF de signaux d'aire glottique (avec et sans le paramètre t_c)	156
8.7	Structure générale d'un système de Lur'e à temps discret	157
8.8	Modèle à une masse	158
8.9	Fonction F_K	160
8.10	Système de Lur'e à temps continu pour le modèle à une masse	160
8.11	Modèle à deux masses	161
8.12	Modèle trapézoïdal	161
8.13	Système de Lur'e à temps continu pour le modèle à deux masses	165
8.14	Analyse-resynthèse de signaux d'aire glottique	167
9.1	Schéma général du modèle à géométrie forcée	171
9.2	Modèle de glotte à une dimension	171
9.3	Schéma fonctionnel du modèle couplé en temps continu	173

9.4	Schéma fonctionnel du modèle couplé en temps discret	175
9.5	Influence de la pression sous-glottique sur le spectre du son rayonné, voyelle \a\ . . .	176
9.6	Influence de la pression sous-glottique sur le spectre du son rayonné, voyelle \u\ . . .	177
9.7	Influence du couplage, comparaison de l'aire glottique et du débit glottique	178
9.8	Structure de Kelly-Lochbaum à la jonction entre deux tubes acoustiques	179
9.9	Représentation d'un profil arbitraire en concaténation de tronçons	181
9.10	Impédance d'entrée d'un pavillon de trombone mesurée et simulée	182
9.11	Catégorisation géométrique des profils en fonction de θ et Υ	183
9.12	Structure de propagation des ondes découplées stables	184
9.13	Opérateurs de dispersion $D(s)$ et $D_n(s)$ dans le plan de Laplace. Cas $\theta = \ln(2)$ et $\varepsilon = 0$	186
9.14	Opérateurs de dispersion $D(s)$ et $D_n(s)$ dans le plan de Laplace. Cas $\theta = \ln(2)$ et $\varepsilon = 0.1$	187
9.15	Évolution des diagrammes de Bode des valeurs propres en fonction de Υ	188
9.16	Modèles à deux masses modifié [PHvH ⁺ 94]	190
A.1	Maquette de larynx à échelle réelle et à géométrie simplifiée	222
A.2	Maquette de larynx robotisée à l'échelle 5 :1	223
E.1	Fonction d'aire glottique proposée par Gupta et Schroeter	238
F.1	Système général de Lur'e à temps continu	239
F.2	Structure générale d'un système de Lur'e	242
F.3	Synthèse par système de Lur'e, paramètres aléatoires	243
F.4	Synthèse pseudo-périodique par système de Lur'e	244
F.5	Synthèse par système de Lur'e chaotique	245
F.6	Système de Lur'e en boucle ouverte	245
F.7	Analyse-resynthèse, premiers exemples 1	248
F.8	Analyse-resynthèse, premiers exemples 1	249
G.1	Tube axisymétrique, cas général	251
G.2	Modèle de conduit en tronçons de tubes droits	253





Table des matières

Préface	i
Résumé	iii
Abstract	v
Remerciements	vii
Notes liminaires	ix
Sommaire	xi
Introduction générale	1
Partie préliminaire : Présentation de l’objet d’étude et position du problème	9
1 Production de la voix humaine	13
1.1 Anatomie de l’appareil phonatoire	14
1.1.1 Présentation générale, situation du larynx et des plis vocaux	14
1.1.2 Auto-oscillation des plis vocaux	16
1.1.3 Articulations des cartilages et actions musculaires	17
1.1.4 Contrôle des résonateurs	18
1.2 Voix parlée, voix chantée et mécanismes laryngés	18
1.2.1 Voix parlée et voix chantée	18
1.2.2 Notion de qualité vocale	19
1.2.3 Mécanismes laryngés	19
1.3 Méthodes exploratoires	21
1.3.1 Visualisation des plis vocaux en vibration	21
1.3.2 Méthodes indirectes d’observation de l’activité glottique	22

1.3.3	Corpus de données multi-modales	22
1.3.4	Estimation de la géométrie et des propriétés acoustiques du conduit vocal	22
1.4	Conclusion	23
2	Différents modèles pour l'analyse et la synthèse sonores de la voix	25
2.1	Signal vocal et quasi-stationnarité	26
2.2	Modèles physiques et modèles de signaux	26
2.2.1	Modèle source-filtre	28
2.2.2	Modèles physiques simples	29
2.2.3	Différents choix pour la modélisation de la phonation	30
2.3	Cuivres et voix, des liens privilégiés	31
2.4	Position du problème et organisation du document	31
I	Première partie : Outils pour l'exploration et la mesure de signaux biométriques	35
3	Méthodes pour l'exploration laryngée	39
3.1	Ordres de grandeur temporel et spatial du mouvement glottique	40
3.2	Échographie	40
3.3	Vidéo-endoscopie ultra-rapide	41
3.4	Électroglottographie	42
4	Mesures multimodales sur les lèvres et la glotte, comparaison entre deux systèmes à valve humaine	45
4.1	Introduction	47
4.1.1	Context	47
4.1.2	Problem statement and document organisation	48
4.2	Experimental setup	49
4.2.1	Human voice production system	50
4.2.2	Brass instruments	51
4.3	Visualisation tool	52
4.3.1	Opening area extraction	52
4.3.2	Synchronous display of multimodal measurements	53
4.3.3	First observations from videos visualisation	54

4.4	Features extraction, analysis and comparisons	54
4.4.1	Qualitative comparison between vocal-fold & lip EI signals (Obs. 1 & 2) . . .	54
4.4.2	Qualitative comparison between vocal-fold & lip DEI signals (Obs.1 &3) . . .	55
4.4.3	Lip and vocal-fold OA and DOA signals and parametric representation (Obs. 4 & Prop. 1)	58
4.4.4	Geometrical parameters of lip and glottis opening areas (Obs. 2)	61
4.4.5	Opening geometries and electrical impedance signals (Obs. 2)	63
4.5	Conclusion	66
5	Imagerie électrique multicanal du larynx	69
5.1	Tomographie d'impédance électrique	71
5.2	Test préliminaire de validation du principe de l'EIT sur une maquette de larynx simplifiée	73
5.3	Conception d'un dispositif mixte EGG/EIT multicanal à géométrie variable	75
5.3.1	Mesures mixtes EGG/EIT	75
5.3.2	Multiplexage fréquentiel et multiplexage temporel	75
5.3.3	Électrodes à géométries d'activation configurables	75
5.3.4	Vue d'ensemble et mise en équation	75
5.3.5	Choix et dimensionnement	78
5.4	Développements et réalisations techniques	78
5.4.1	Vue générale du dispositif	78
5.4.1.1	Identification des contraintes et des problèmes techniques	78
5.4.1.2	Architecture du dispositif et état d'avancement	79
5.4.2	Éléments matériels	80
5.4.2.1	Génération de signaux isolés galvaniquement	80
5.4.2.2	Routeur analogique	81
5.4.2.3	Accès protégé aux électrodes	83
5.4.2.4	Traitement synchronisé des ordres de configuration et des mesures	84
5.4.3	Premier élément logiciel et premiers tests	85
5.5	Conclusion et perspectives	86

II	Deuxième partie : Modèle source-filtre et analyse-synthèse	91
6	Le modèle source-filtre	95
6.1	Source glottique	96
6.1.1	Modèles paramétriques de source glottique	96
6.1.1.1	Modèles paramétriques définis par morceaux	97
6.1.1.2	CALM (Causal-Anticausal Linear Model)	97
6.1.2	Choix du modèle d'étude	98
6.1.2.1	Pôles causaux et pôles anticausaux	99
6.1.2.2	Définition du modèle de source glottique	100
6.2	Filtre vocal et rayonnement	100
6.2.1	Conduit vocal et filtre tous-pôles	101
6.2.2	Filtre pôles-zéros	101
6.2.3	Filtre de rayonnement	101
6.2.4	Définition du modèle d'étude	101
6.3	Modèle source-filtre paramétrique complet étudié	102
7	Analyse du modèle source-filtre	103
7.1	État de l'art des méthodes d'estimation et de séparation	104
7.2	Premier algorithme : Séparation source-filtre pour la classe des signaux périodiques	105
7.2.1	General description of the algorithm	107
7.2.2	Space Π of infinitely periodic signals	108
7.2.3	Turning poles into zeros	109
7.2.4	Extracting the desired poles	109
7.2.5	Estimating the parameters d and \mathcal{G}	109
7.2.6	General remarks on the algorithm	110
7.2.7	Tests on synthetic signals	110
7.2.8	Discussion on robustness	111
7.2.9	Conclusion and perspectives	112
7.3	Deuxième algorithme : Analyse cepstrale et séparation source-filtre	115
7.3.1	Considérations générales sur le cepstre	115
7.3.1.1	Cepstre réel, cepstre complexe et déconvolution causale-anticausale	115
7.3.1.2	(re)Définition du cepstre complexe	117

Rappels sur le logarithme complexe	117
Cepstre complexe	120
Méthodes de calcul	122
Cepstre complexe et cepstre différentiel	124
7.3.2 Cepstre d'un signal tous-pôles causal/anti-causal	125
Cepstre complexe	125
Cepstres différentiels causaux et anti-causaux	126
7.3.3 Algorithme d'estimation des paramètres du modèle causal-anticausal	127
Présentation de l'algorithme	127
Maximisation de la vraisemblance des cepstres différentiels	128
Remarques générales sur le deuxième algorithme.	129
Ingrédients à ajouter pour traiter des signaux vocaux mesurés.	129
7.3.4 Performances de l'algorithme d'estimation	129
7.3.4.1 Signaux de synthèse	129
7.3.4.2 Robustesse de l'algorithme	130
7.3.4.3 Signaux réels	136
7.4 Conclusion et perspectives	137

III Troisième partie : Modèles à géométrie forcée et analyse-synthèse 141

8 Analyse-Synthèse de fonctions d'aire glottique 145

8.1 Construction d'un dictionnaire de test par estimation de fonctions d'aire glottique à partir de vidéo-endoscopies ultra-rapides	146
8.2 Modèle LF	147
8.2.1 Algorithme d'estimation des paramètres LF	150
8.2.1.1 Description générale de l'algorithme	150
8.2.1.2 Initialisation	150
8.2.1.3 Optimisation des paramètres LF	152
8.2.2 Performances de l'algorithme d'estimation des paramètres LF	154
8.2.2.1 Signaux de synthèse	154
8.2.2.2 Signaux réels	154
8.3 Modèles de diversité en systèmes dynamiques non linéaires	157

8.3.1	Contexte, motivations et premières applications	157
8.3.2	Modèle à une masse	158
8.3.2.1	Modèle Théorique	158
8.3.2.2	Réduction du problème	159
8.3.2.3	Système de Lur'e (à temps continu)	159
8.3.3	Modèle à deux masses	160
8.3.3.1	Modèle mécanique	160
8.3.3.2	Forces de pression	161
8.3.3.3	Adaptation à la structure des systèmes de Lur'e	162
	Approximations	162
	Cas où $\psi_1 = \psi_2$	163
	Application au modèle à deux masses	164
	Équation différentielle unique	164
	Système de Lur'e (à temps continu)	165
	Discrétisation	165
8.3.3.4	Utilisation pour l'analyse/synthèse	166
8.4	Conclusion	166
9	Modèles à géométrie forcée	169
9.1	Modèle couplé inspiré du modèle source-filtre	170
9.1.1	I1 Modèle de glotte	170
9.1.2	I2-I3 Modèle de conduit vocal et couplage	172
9.1.3	Modèle complet	173
9.1.4	Débouclage et discrétisation	173
9.1.5	Modèle de synthèse implémentable	174
9.2	Apport du couplage entre la source glottique et le conduit vocal	175
9.2.1	Influence de la pression sous-glottique	175
9.2.2	Influence du couplage glotte-conduit	178
9.3	Modélisation de conduit vocal à profil régulier en guides d'ondes acoustiques 1D	179
9.3.1	Formalisme guide d'ondes	179
9.3.2	Modèle de Webster-Lokshin dans les tubes à section variable	179
9.3.3	Profils à rayon positif	182

9.3.4	Diagonalisation de la matrice de transfert	184
9.3.5	Observations sur les nouveaux opérateurs de propagation et de dispersion	184
9.4	Conclusion et perspectives	188
 Conclusion générale		193
 Bibliographie		199
 Annexes		219
 A Projet VoiceTronics		219
A.1	Projets mécatroniques à l'école Mines ParisTech	220
A.2	SP2 : Maquette de larynx robotisée à échelle réelle et à géométrie simplifiée	220
A.3	SP3 : Maquette de larynx robotisée à l'échelle 5 :1	221
 B Modèles de pulse glottique par fonctions paramétriques définies par morceaux		225
B.1	Forme générique	225
B.2	Expressions analytiques des modèles	226
B.2.1	Rosenberg-B (1971)	226
B.2.2	KLGLOTT88 (1990)	226
B.2.3	R++ (1998)	226
B.2.4	Fant (1979)	227
B.2.5	LF (Liljencrants-Fant) (1985)	227
B.2.6	LF modifié (LF-Rd) (1995)	228
 C Étude analytique du modèle LF		229
C.1	Écriture temporelle du modèle de dérivée du débit	229
C.2	Écriture temporelle adimensionnée du modèle de dérivée du débit	230
C.3	Équations implicites et bilan des degrés de liberté	230
C.3.1	Cas $d_a > 0$: "fermeture douce"	230
C.3.2	Cas $d_a = 0$ "fermeture abrupte"	231
C.4	Écriture temporelle adimensionnée du modèle de débit	231
C.5	Transformées de Fourier des modèles temporel et dérivé	232

D	Dictionnaire de cepstres en temps discret	233
D.1	Impulsion de Dirac	233
D.2	Pôle	233
D.2.1	Pôle causal	233
D.2.2	Pôle anticausal	234
D.3	Zéro	234
D.3.1	Zéro causal	234
D.3.2	Zéro anticausal	234
D.4	Propriétés remarquables	235
D.4.1	Convolution de deux signaux temporels	235
D.4.2	Multiplication par une constante	235
D.4.3	Homothétie temporelle	235
E	Modèles de fonction d'aire glottique	237
F	Modele de Lur'e	239
F.1	Modèle de Lur'e en temps continu	239
F.1.1	Formulation	239
F.1.2	Méthode de discrétisation à l'aide de l'approximation des trapèzes	239
F.2	Bases de fonctions pour la non-linéarité ψ	241
F.2.1	Polynômes de Tchebychev	241
F.2.2	Smoothed Piecewise linear Functions	241
F.3	Première tentative d'analyse de fonctions d'aire glottique	242
F.3.1	Modèle de Lur'e à temps discret	242
F.3.1.1	Modèle de synthèse paramétrique	242
F.3.1.2	Quelques exemples de synthèse	243
F.3.2	Un algorithme d'identification des paramètres du système de Lur'e	245
F.3.2.1	Identification linéaire	246
F.3.2.2	Mesure de la qualité de l'identification linéaire	246
	Par prédiction	246
	Avec la matrice de corrélation	247
F.3.2.3	Contraintes sur les paramètres	247
F.3.2.4	Optimisation des paramètres	247

F.3.2.5 Premiers résultats	248
G Propagation dans les tubes acoustiques	251
G.1 Modèle général et matrices de transfert acoustique	251
G.2 Cas des tubes droits	252
G.2.1 Matrice de transfert acoustique dans un tube droit sans pertes	252
G.2.2 Concaténation de tubes droits	253
G.3 Observations sur les fonctions H_t et Z_{in}	255
 Notations, acronymes et expressions usuelles	 259
 Liste des tableaux	 263
 Table des figures	 265
 Table des matières	 271

Résumé :

Ce travail porte sur l'étude de la production de la voix. L'objectif principal est de proposer des solutions pour son analyse-synthèse qui permettent une large représentation de la diversité des signaux biométriques (observés en pratique) dans le cadre du paradigme source-filtre ou de modèles physiques simplifiés. L'étude est menée selon trois axes, l'exploration de l'activité laryngée, la modélisation de la phonation et l'analyse-synthèse sonore. Après une description de l'appareil phonatoire et de modèles d'analyse-synthèse actuels, la première partie s'intéresse à l'exploration de signaux biométriques. À partir du recensement des dispositifs actuels, nous proposons un outil de synchronisation qui révèle la complexité du mouvement 3D des plis vocaux. Une étude comparative avec un système acoustique à valve humaine similaire (lèvres d'un tromboniste) est menée et le développement d'un nouvel outil combinant deux techniques récentes d'imagerie électrique est proposé.

Dans la deuxième partie, deux algorithmes d'analyse sont proposés pour un modèle source-filtre fondé sur une représentation tous-pôles causale-anticausale du signal vocal. Un calcul du cepstre complexe est proposé, garantissant l'holomorphie sur une couronne de convergence dans le domaine en Z . Cet outil est alors exploité pour la séparation de la source et du filtre.

Enfin, des représentations paramétriques des signaux d'aire glottique sont exploitées pour piloter des modèles à géométrie forcée. Un modèle de synthèse, inspiré du modèle source-filtre mais intégrant l'effet du couplage entre la glotte et le conduit vocal, et des avancées sur la modélisation de conduit vocal à profil régulier sont présentées.

Mots clés : production de la voix, analyse-synthèse, séparation source-filtre, modèles physiques, mesures in vivo, impédancemétrie.

Abstract :

This work focuses on the study of voice production. The main objective is to propose solutions for its analysis-synthesis which enable a wide representation of the diversity of biometric signals (observed in practice) within the framework of source-filter paradigm or simplified physical models. The study is conducted according to three axes : exploration of laryngeal activity, modelisation of the phonation and sound analysis-synthesis.

After a description of the vocal apparatus and current analysis-synthesis models, the first part focuses on the exploration of biometric signals. From an inventory of current devices, we propose a synchronisation tool which reveals the complexity of the 3D motion of vocal folds. A comparative study with a similar human-valve-based sound production system (lips of a trombonist) is conducted and the development of a new tool combining two recent electrical imaging techniques is proposed.

In the second part, two new analysis algorithms are proposed for a source-filter model based on a causal-anticausal all-pole representation of speech signals. A computation method of the complex cepstrum is proposed, which ensures the holomorphy on a convergence domain in the Z plane. This tool is then used for the separation of the source and the filter.

Finally, parametric representations of glottal area waveforms are used to drive models controlled by the geometry. A synthesis model, based on the source-filter model but including the effect of the coupling between the glottis and the vocal tract, and advances in the modeling of the vocal tract with regular profile are presented.

Keywords : voice production, analysis-synthesis, source-filter separation, physical models, in vivo measurements, impedancemetry.

Travaux réalisés à

Institut de Recherche et Coordination Acoustique/Musique
IRCAM - CNRS UMR 9912 STMS - UPMC
1, place Igor Stravinsky
75004 PARIS, FRANCE