

Master 2 MIS ATIAM

Rapport de Stage

Analyse de la source glottique par ZZT

[UPMC paris VI - IRCAM - LIMSI/CNRS]
encadré par Christophe D'Alessandro

Nicolas Sturmel

22 juin 2006

Résumé

Ce stage a pour sujet la détection des paramètres de la source glottique à l'aide de l'analyse par Zéros de la Transformée en Z (ZZT). En utilisant une description spectrale de la source glottique, des travaux précédent on montré que la ZZT permet d'estimer les paramètres de source. Cette étude est un approfondissement des travaux initiaux.

Elle s'est articulée autour de l'implémentation de l'algorithme de décomposition par ZZT, et de l'étude de sa robustesse. Nous avons notamment mesuré avec quelle finesse la position des points d'analyse, synchrones au fondamental (GCI, glottal closing instant), devaient être déterminés pour arriver à une décomposition optimale. Nous avons aussi étudié différentes sources de problèmes supplémentaires lors de la décomposition : erreurs sur le calcul des zéros, rapport son/bruit, influence du fenêtrage du signal.

Nous avons ensuite montré qu'il était possible, via une mesure conjointe du quotient ouvert par électroglottographie (EGG) et du formant glottique par ZZT, d'arriver à estimer un nouveau paramètre de la source glottique : le coefficient d'asymétrie α_m . Nous avons alors créé une base de données de signaux vocaux pour éprouver l'implémentation de la décomposition et la mesure de ce paramètre. Cette base s'articule autour de signaux émis en mécanisme laryngé I et II par 3 locuteurs, sur 3 voyelles choisies pour leurs caractéristiques spectrales typiques : A, I et OU.

Il apparaît que l'analyse donne surtout de bons résultats pour les signaux à grande période fondamentale et/ou à faible quotient ouvert. On mesure alors des valeurs de α_m typiquement comprises entre 0.6 et 0.8. Nous montrons donc, que sous certaines hypothèses vis à vis des signaux analysés, nous pourrions faire une mesure dont la précision serait inférieure au différentiel perceptif. Par contre il semble que la méthode de décomposition implémentée n'est pas assez robustes pour des signaux émis en mécanisme II lorsque les périodes sont courtes et les quotients ouverts forts.

Table des matières

I	Introduction	6
1	Présentation du Sujet	7
2	L'appareil vocal	8
2.1	La production de la voix	8
2.2	Modélisation source/filtre	9
3	Modélisation de la source glottique	11
3.1	Les modèles de la source glottique	11
3.1.1	Modèle LF	11
3.1.2	Notation et paramètres	12
3.2	Le spectre du modèle générique [7]	12
4	Mesure du quotient ouvert : Etat de l'art	14
4.1	Différence entre les deux premiers harmoniques	14
4.2	Mesure par signaux EGG	14
4.2.1	Utilisation de l'EGG	14
4.2.2	Principe de l'EGG	15
4.2.3	Mesure du quotient ouvert	15
4.2.4	Critique	15
4.3	Seuil différentiel	16
5	ZZT et déconvolution	17
5.1	Principe	17
5.1.1	Rappels, transformée en Z	17
5.1.2	Causalité, Stabilité	17
5.1.3	Définition de la ZZT	18
5.2	Théorie des polynômes et interprétation dans les domaines fréquentiel et temporel	18
5.2.1	Passage d'une forme à une autre du polynôme	18
5.2.2	Organisation des Zéros pour un signal périodique	19
5.3	Exemples	19
5.3.1	ZZT de la réponse impulsionnelle d'un AR1	19
5.3.2	ZZT d'une sinusoïde	20
5.4	Utilisation de la ZZT pour l'analyse de la voix	20
5.4.1	Modèle causal/anti causal [6]	20
5.4.2	ZZT de signaux synthétiques	21
5.4.3	Application à un signal réel [4]	22
II	Estimation des propriétés basse fréquence de la source glottique	25
6	Déconvolution de la source par ZZT : implémentation	26
6.1	Implémentation de la ZZT	26
6.1.1	Composante continue	27

6.2	Plage de Recherche	28
6.3	Evaluation de la précision de l'analyse	28
6.3.1	Déplacement de F_g par ZZT	28
6.3.2	Influence des erreurs sur les zéros	28
6.3.3	Conclusion	29
6.4	Mesure de α_m	29
7	Etude de la robustesse	31
7.1	Vis à vis d'une mauvaise détermination du GCI	31
7.1.1	Sur des signaux de synthèse	31
7.1.2	Sur des signaux réels	33
7.2	Vis à vis d'une mauvaise décomposition	36
7.2.1	Origine	36
7.2.2	Masquage de F_g	36
7.2.3	Limites de la mesure	37
7.3	Autres points importants	37
7.3.1	Rapport signal bruit	37
7.3.2	Fréquence fondamentale	37
7.3.3	Quotient ouvert	38
7.4	Résumé	38
III	Analyse d'une base de données d'échantillons	39
8	Base de données	40
8.1	Enregistrement de la base	40
8.1.1	Protocole expérimental	40
8.2	Classification des échantillons	41
9	Evaluation et résultats	42
9.1	Protocole d'évaluation	42
9.2	Critères d'évaluation	43
9.3	Interprétation des résultats	44
9.3.1	Influence du phonème et de la période du signal	44
9.3.2	Influence du quotient ouvert	46
9.3.3	Variation de α_m et interprétation physique	46
9.3.4	Qualité de la décomposition, commentaire des annexes	46
9.3.5	Mesure directe de O_q par ZZT	46
IV	Finalemment...	48
10	Conclusion	49
10.1	Bilan	49
10.2	Critique	49
10.3	Travail à venir	50

Abréviations

ν	: fréquence réduite
ZZT	: Zéros of Z Transform (zéros de la transformée en Z)
UV	: UnVoiced (signal non voisé)
V	: Voiced (signal voisé)
GCI	: Glottal Closing Instant (instant de fermeture glottique)
O_q	: Open Quotient (quotient ouvert)
Fg	: Formant Glottique
F_0	: Fréquence fondamentale
T_0	: période fondamentale
egg, EGG	: électroglottograph{e e ique}
DEGG	: dérivée du signal electroglottographique
ODG	: onde de débit glottique
JND	: just noticeable difference (seuil différentiel)

Préambule

Que ce soit dans les GSM, sur internet, dans les interfaces homme-machine, dans la synthèse musicale, ou tout simplement dans le cabinet d'un orthophoniste, les domaines d'application de l'analyse-synthèse des signaux vocaux sont vastes. Mais encore aujourd'hui un grand nombre de problèmes ne peuvent être résolus. Nos fonctions cognitives font naturellement la différence entre plusieurs locuteurs, et il n'est pas étonnant de se rendre compte, juste en écoutant la voix d'un interlocuteur, de déceler chez lui une tension, une peine, ou de la fatigue.

Si les techniques actuelles permettent de résoudre avec brio des problèmes liés à la dictée et à la synthèse de voix monotonale, il semblerait que les implémentations du modèle source filtre utilisée industriellement (et ses techniques d'analyse idoines) arrivent à ses propres limites.

L'analyse par ZZT¹ est un domaine de recherche innovant. Rendue possible grâce aux progrès informatiques, cette technique d'analyse a pour ambition d'ouvrir un nouveau champ de recherche qui permettrait de résoudre ces problèmes en utilisant un modèle plus complet de l'appareil de production vocal.

Ce stage a pour sujet la détermination d'un jeu complet de paramètres menant à l'analyse des paramètres de la source glottique. Il a été démontré rapidement dans [2] qu'il était possible, par décomposition ZZT, de déterminer ces paramètres.

Après avoir expliqué et modélisé le fonctionnement du mécanisme vocal et dressé un état de l'art des techniques de détermination du quotient ouvert, nous présenterons le principe de la ZZT avant de l'appliquer à un échantillon réel.

Ce travail préliminaire posera les jalons d'un travail de réflexion autour de la possibilité d'estimer un autre coefficient du modèle de la source glottique. Nous établirons un protocole expérimental de mesure, et appliquerons la dite méthode à une base de données d'échantillons spécialement créée pour l'occasion.

¹ *Zeros of Z Transform*, ou zéros de la transformée en Z

Première partie

Introduction

Chapitre 1

Présentation du Sujet

La détermination des paramètres de la source glottique est un sujet relativement difficile, car il porte sur la mesure de paramètres convolués de la source de la production vocale. Si, à l'aide d'un enregistrement EGG, il est possible d'atteindre deux des principaux paramètres d'analyse, il n'existe actuellement pas de méthode fiable pour déterminer les paramètres restants.

Ce stage se concentre donc sur l'analyse des paramètres de la source glottique, en particulier le coefficient d'asymétrie α_m , par la détermination du maximum du spectre de la source glottique (formant glottique F_g). Une nouvelle méthode, l'analyse par ZZT, est étudiée en profondeur et évaluée pour la détermination de ce nouveau paramètre par le biais d'une séparation source-filtre.

Si en théorie cette méthode est être simplement mise en oeuvre, son implémentation aux signaux réels fait aussi partie de l'étude de ce stage, et une partie sera consacrée au protocole expérimental, à son implémentation et à sa robustesse.

Une base de données de voix a été spécialement enregistrée pour l'occasion, et a servi à une application *grande échelle* de la méthode développée. La première partie introduit les connaissances acquises et nécessaires au stage, considérées indispensables à la compréhension du cheminement intellectuel menant aux résultats présentés dans la suite du document.

Chapitre 2

L'appareil vocal

2.1 La production de la voix

Le larynx est composé de 11 cartilages, de membranes, d'os, de muscles et de ligaments reliant les différentes pièces (cf 2.1). Le jeu présent entre ces éléments est primordial pour le mécanisme de phonation, car il va permettre la modification de la géométrie des cordes vocales.

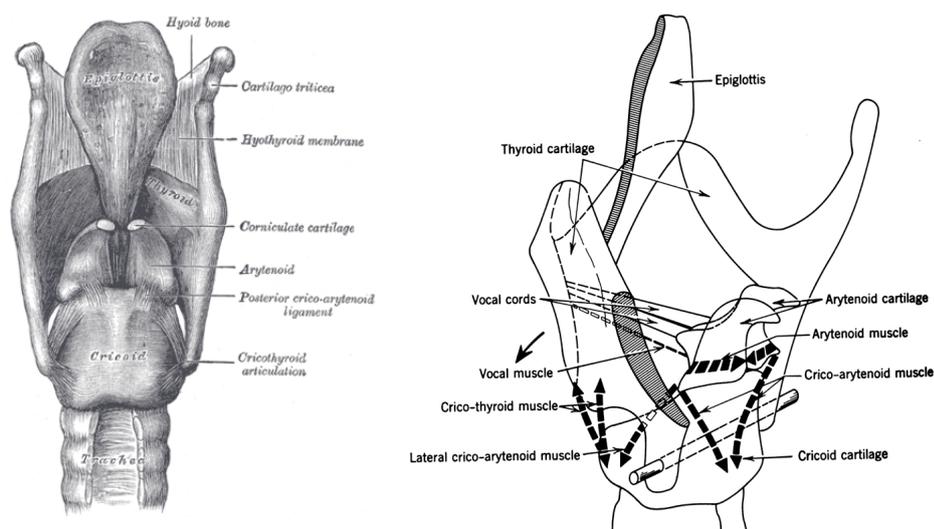


FIG. 2.1 – Fonctionnement du larynx (source web : education.yahoo.fr et RANKE, LULLIES, 1953)

Les cordes vocales (cf figure 2.2) sont situées dans le vestibule laryngé, visibles à gauche de la figure 2.1, elles se constituent en un repli de muqueuse de petite taille (environ 2cm) et permettent l'émission de sons : la phonation.

On appelle glotte l'espace situé entre les cordes vocales, et on dira d'une composante du son qu'elle est glottique si elle est directement produite par les cordes vocales.

La source : L'air expulsé par les poumons va se comprimer sous les cordes vocales jusqu'à provoquer leur ouverture; on parle alors de pression due au débit sous-glottique. Le seuil

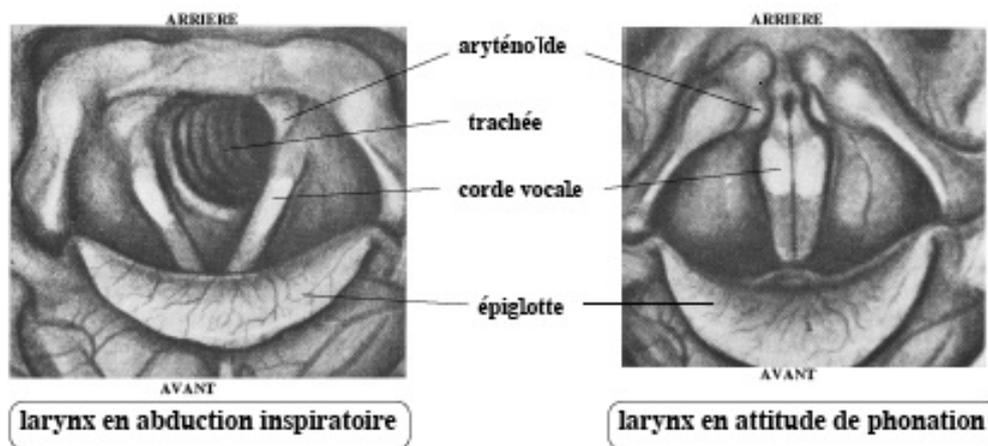


FIG. 2.2 – Larynx en position de respiration et de phonation (ref : thèse de Nathalie Henrich, figure 1.2)

de cette pression à partir duquel l'air va passer dépend bien évidemment de la géométrie instantanée du larynx. (surface et raideur des cordes vocales, entre autre)

Atteindre les périodes d'ouverture et de fermeture de la glotte nous permet donc d'atteindre certaines caractéristiques des sons émis.

Conduit vocal : De l'autre côté des cordes vocales vont se former des ondes planes stationnaires se propageant dans le conduit vocal (composé du pharynx, des cavités buccales et nasales). Ce dernier, pouvant être modélisé comme une succession de tubes de longueur et section différentes, va générer des résonances qui se traduisent par des maxima locaux sur la fonction de transfert du conduit, maxima appelés *formants*.

Une séquence composée d'une ouverture puis d'une fermeture de cette glotte correspond à une période du signal glottique ; cette période détermine directement la fréquence fondamentale des sons issus de cette phonation (phonation voisée [V]) Les parties bruitées de la phonation (omniprésentes en phonation non voisée [UV]) sont elles générées par les turbulences de l'écoulement d'air, notamment au niveau de la glotte.

L'articulation, élément essentiel de la parole, est réalisée en aval de la glotte, par toutes les cavités sur le trajet de l'air : on parle alors de résonateurs qui, indépendamment de la fréquence fondamentale d'un signal voisé, vont lui donner son caractère propre à chaque locuteur et à chaque phonème prononcé.

Pour l'ODG, on parle de débit, alors que nous intéressons à la pression instantanée de l'émission vocale (mesurable). Comme nous mesurons la pression de la voix, nous mesurons de même la dérivée de l'ODG.

2.2 Modélisation source/filtre

En tenant compte des propriétés énoncées ci-avant, il est donc possible de modéliser *du point de vue du traicteur de signal* la production vocale comme le succession de trois actions :

- La production d'un signal périodique par la glotte,
- Le filtrage de ce signal par les cavités orales et nasales,
- Une dérivation appliquée suite au rayonnement par les lèvres

Le modèle source filtre considère que la voix est le résultat d'un filtrage linéaire à variations lentes dans le temps [9] d'un signal source : la source glottique. On considérera, pour simplifier le problème et en vertu des propriétés de linéarité du filtre, que la dérivation du signal due au rayonnement par les lèvres s'applique directement à la source glottique. Ainsi, les sons voisés sont modélisés par un filtrage de la dérivée de la source glottique. [13]

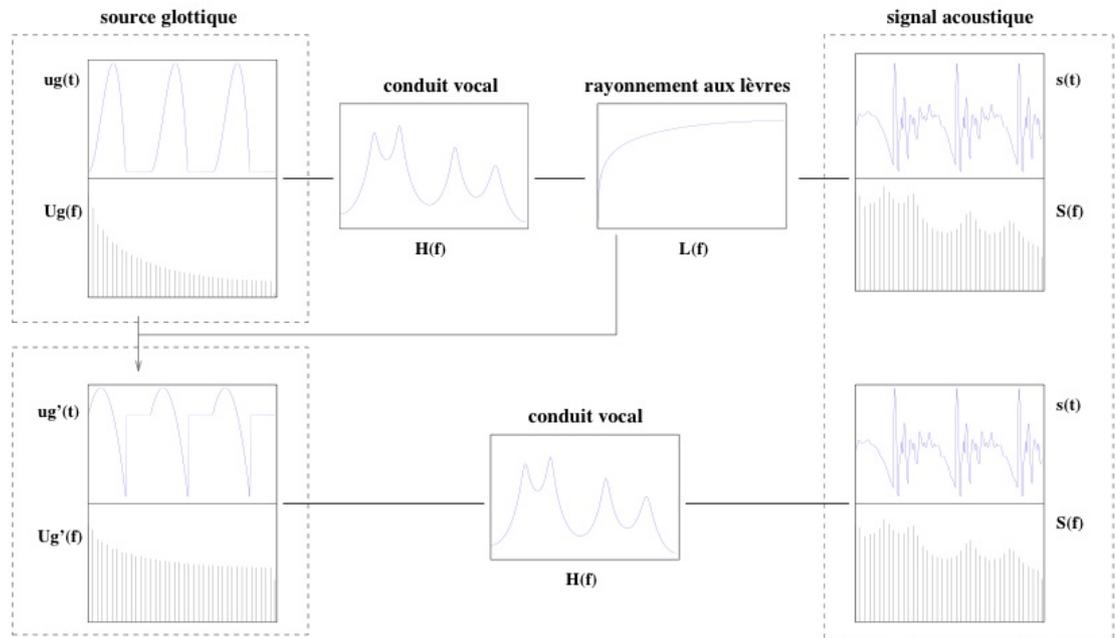


FIG. 2.3 – Illustration du modèle source-filtre (ref : thèse de Nathalie Henrich, figure 2.1)

Chapitre 3

Modélisation de la source glottique

3.1 Les modèles de la source glottique

Le but du modèle choisi est de représenter la forme d'onde glottique avec un nombre restreint de paramètres ayant chacun une réalité physiologique ou physique. On se place donc dans le domaine du traitement du signal et on attend du modèle qu'il reproduise avant tout, le plus fidèlement possible, la forme d'onde du débit glottique ou sa dérivée. Avant de définir le jeu de paramètres sur lequel nous nous baserons par la suite, nous allons voir plus en détail le modèle LF.

3.1.1 Modèle LF

C'est un modèle de signal temporel, introduit dans [14] par Liljencrants et Fant, de la forme d'onde dérivée de la source glottique. L'avantage de son calcul immédiat dans le domaine temporel est compensé par la difficulté de détermination de certains de ses paramètres cachés :

Les paramètres α et ϵ du modèle explicités ci dessous sont principalement déduits de la continuité de la dérivée de l'ODG et de la nullité du débit à la fin de chaque période (l'ODG nulle à la fin d'un cycle).

présentation : Le modèle LF considère l'onde dérivée de source glottique comme la concaténation temporelle de deux phases : la phase d'ouverture et la phase de fermeture glottique. La phase d'ouverture est modélisée par le produit d'une sinusoïde et d'une exponentielle croissante. La phase de fermeture est quant à elle réduite à une exponentielle décroissante.

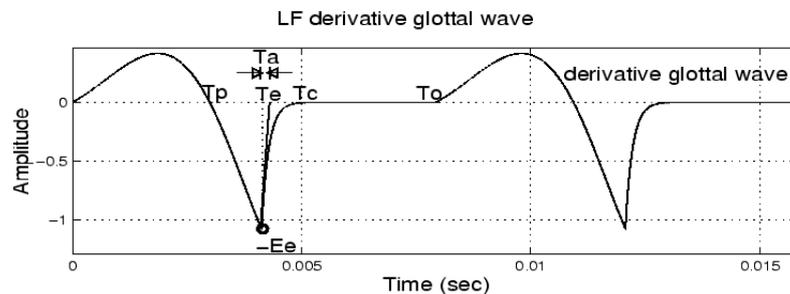


FIG. 3.1 – Illustration du modèle LF de la dérivée de l'ODG (source ccrma.stanford.edu)

explicitation :

$$\begin{aligned}
g(t) &= E_e e^{at} \sin \omega_g t && \text{pour } 0 \leq t \leq T_e \\
g(t) &= -\frac{E_e}{\epsilon T_a} [e^{-\epsilon(t-T_e)} e^{-\epsilon(T_c-T_e)}] && \text{pour } T_e \leq t \leq T_c \leq T_0 \\
g(t) &= 0 && \text{pour } T_e \leq t \leq T_0 \\
\end{aligned}$$

pour trouver ϵ : continuité à $t = T_e$

pour trouver a : $\int_0^{T_0} g(t, a) dt = 0$

T_a est le temps de la phase de fermeture, T_0 est la période fondamentale du signal, T_e le temps de la phase ouvert, et E_e l'amplitude de voisement comme montrés sur la figure 3.1.

3.1.2 Notation et paramètres

Il est ici question de définir un modèle qui convient le mieux aux observations et mesures qui sont le sujet du stage. Pour cela nous utilisons le jeu de paramètres définis dans [7] qui correspondent plus finement à la représentation temporelle de la dérivée de l'ODG.

- O_q : Le quotient ouvert
- T_0 : La période fondamentale
- A_v : L'amplitude de voisement
- α_m : le coefficient d'asymétrie, analogue au quotient de vitesse.
- T_a : la durée effective de la phase de retour, permet d'atteindre la pente spectrale.

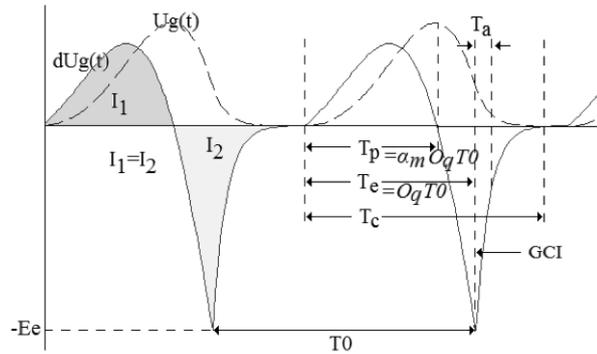


FIG. 3.2 – Modèle du flux glottique. U_g : débit acoustique glottique ; dU_g : dérivée du débit acoustique (source [5])

3.2 Le spectre du modèle générique [7]

On remarque sur la figure 3.3 que le spectre possède un maximum caractéristique, on appelle ce maximum le formant glottique, qui ne fait pas partie des formants résonantiels. On montre que, comme sur le schéma, la fréquence du formant glottique est en directe relation avec T_0 , α_m et O_q . Il est donc possible d'exprimer une fonction de ces 3 paramètres.

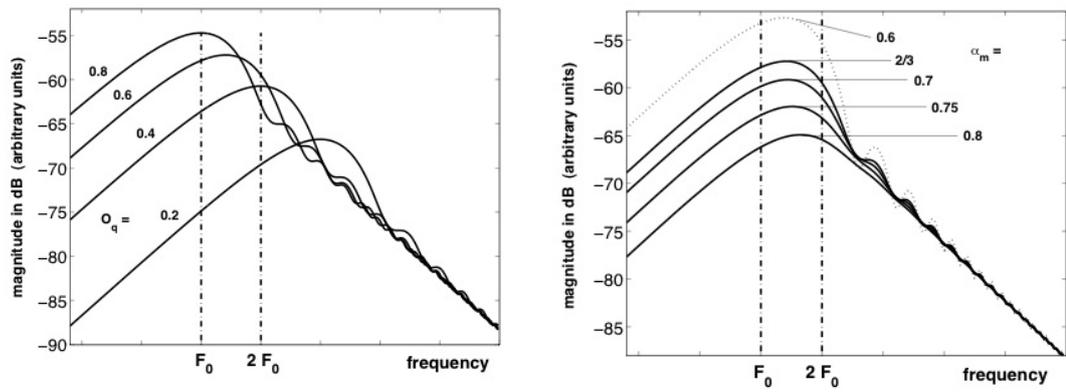


FIG. 3.3 – Spectre du modèle LF pour différentes valeurs de O_q à gauche, et différentes valeurs de α_m à droite. (source [7])

Chapitre 4

Mesure du quotient ouvert : Etat de l'art

Le quotient ouvert est donc défini comme le rapport entre le temps d'ouverture et le temps total d'une période de l'onde glottique. Tout le problème de la détermination de O_q réside dans la détermination de l'onde glottique elle-même, ou la plupart du temps, des instants d'ouverture et de fermeture glottique.

La méthode la plus évidente est de réaliser un filtrage inverse du signal vocal en évaluant la fonction de transfert du canal par LPC [15]. Cette méthode, bien que simple, ne donne pas de résultats fiables, c'est pourquoi nous intéresserons ici à deux méthodes : une méthode de détermination spectrale de O_q et une méthode non invasive de mesure directe sur le locuteur par signaux EGG.

4.1 Différence entre les deux premiers harmoniques

Il existe une technique de mesure du quotient ouvert par le rapport d'amplitude des deux premiers harmoniques. Soit $H_1 - H_2$ la différence d'amplitude (en dB) des deux premiers harmoniques, alors il est montré dans [10] qu'il existe une bonne corrélation entre le $H_1 - H_2$ et O_q en utilisant le modèle LF dans un cas de fermeture abrupte ($T_a = 0$). Cette corrélation quasiment linéaire suit l'équation :

$$H_1 - H_2 = -6 + 0.27e^{5.5O_q}$$

Cependant un calcul plus poussé ([7]) montre que le rapport des amplitudes de ces deux harmoniques est fonction à la fois de O_q et de α_m . Cette méthode ne permet donc pas d'avoir directement le quotient ouvert, mais une valeur proche de celui-ci, modifiée par la valeur courante du coefficient d'asymétrie.

4.2 Mesure par signaux EGG

4.2.1 Utilisation de l'EGG

La détermination de O_q par signaux EGG¹ [11] est un des sujets du travail de thèse de Nathalie Henrich [13]. Plusieurs méthodes y sont présentées et comparées, mais elles

¹Electroglottographie

résident toutes dans l'exploitation d'enregistrements egg, qui nécessitent une préparation et une instrumentation du locuteur.

4.2.2 Principe de l'EGG

Mise au point par Fabre en 1956 [8], l'électroglottographie est une méthode non invasive de l'investigation du mouvement vibratoire des cordes vocales.

Deux électrodes placées sur le sujet, au niveau du cartilage thyroïdien, sont parcourues par un courant haute fréquence. Le circuit de mise en forme des signaux opère une démodulation afin de garder uniquement les variations en tension de la porteuse haute fréquence qui représentent les variations d'impédance entre les deux électrodes.

Quand la glotte est fermée, la résistance aux courants est moindre, ce qui permet donc (via quelques précautions) de retrouver une image du mouvement glottique à travers la mesure de l'impédance entre les deux électrodes.

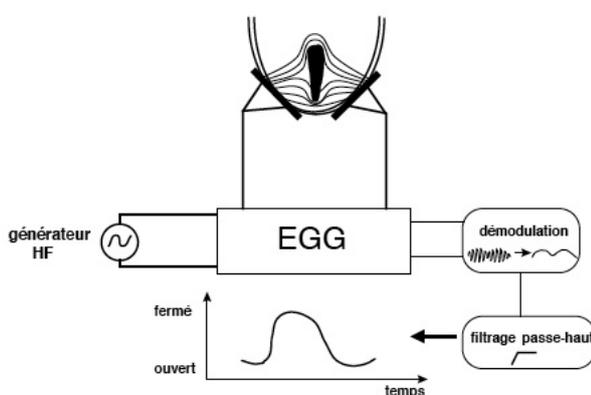


FIG. 4.1 – Principe de l'EGG (source, thèse de N. Henrich)

4.2.3 Mesure du quotient ouvert

Plusieurs méthodes de mesure ont alors été mises au point et comparées entre elles pour la détermination du quotient ouvert. Certaines sur le signal EGG lui même, d'autres sur sa dérivée (DEGG) où les instants d'ouverture et de fermeture glottique sont mis en évidence.

On peut citer [13] les méthodes à seuil 30% ou 50% qui donnent le quotient ouvert par le rapport du temps où le signal EGG est supérieur au seuil sur le temps d'une période du signal ; ou la méthode DECOM qui se base sur l'analyse des signaux DEGG et la détermination précise des instants de fermeture et d'ouverture glottique.

4.2.4 Critique

Cette méthode se révèle très fiable, et beaucoup plus légère en matière de calcul que la méthode par différence des deux premiers harmoniques. En outre il est possible d'accéder directement aux instants critiques du mouvement des cordes vocales en totale indépendance au canal supraglottique.

Cependant, si cette méthode présente un intérêt scientifique évident, elle nécessite d'instrumenter un patient avant d'analyser sa voix : elle est donc cantonnée aux laboratoires de recherche et aux cabinets des orthophonistes.

4.3 Seuil différentiel

Une étude complète du seuil différentiel de perception d'une variation de O_q a été réalisée dans [12] ; nous y apprenons que le JND sur le quotient ouvert est relatif à la valeur ciblée et de l'ordre de 14% pour un sujet non entraîné et 10% pour un sujet entraîné. Par conséquent, seules de grandes variations de O_q seront perçues, et dans un but de déterminer les paramètres de la source glottique, si il n'est donc pas nécessaire d'obtenir une grande précision pour des mesures perceptives de O_q .

Chapitre 5

ZZT et déconvolution

Ce chapitre présente le deuxième point bibliographique important du stage : l'analyse par ZZT introduite dans la thèse de Baris Bozkurt [5]. Après avoir rappelé le principe de la transformée en Z et énoncé quelques propriétés immédiates de l'étude de ces polynômes, des exemples vont être abordés.

5.1 Principe

5.1.1 Rappels, transformée en Z

On définit la transformée en Z, $X(z)$ d'un signal $x(n)$ de N échantillons par :

$$X(z) = \sum_{i=0}^{N-1} x(i)z^{-i}$$

comme le pendant dans le domaine du signal discret de la transformée de Laplace.

En calculant $X(z)$ pour $z \in \mathbb{C}_1$ le cercle unité, on retrouve la transformée de Fourier discrète du signal x :

$$TFD_{x(n)}(\nu) = \sum_{i=0}^{N-1} x(i)e^{j2\pi\nu i}$$

5.1.2 Causalité, Stabilité

Phase minimum : Un système est à phase minimum si ses zéros et ses pôles se trouvent à l'intérieur de cercle unité. Un tel système sera toujours causalement stable. (anticausalement instable)

Phase maximum : Un système est à phase maximum si ses zéros et ses pôles se placent à l'extérieur du cercle unité, il sera alors anti-causalement stable, en causalement instable.

Phases multiples : Un système à phases multiples est un système dont les pôles ne sont pas exclusivement dans l'une des deux zones délimitées par le cercle unité. Un tel système peut être vu comme la convolution temporelle (produit spectral) d'un système à minimum phase et d'un autre à maximum phase.

5.1.3 Définition de la ZZT

On appelle alors ZZT la représentation des zéros du polynôme $X(z)$, transformée en Z du signal $x(n)$. (de l'anglais *Zeros of the Z Transform*)

5.2 Théorie des polynômes et interprétation dans les domaines fréquentiel et temporel

Lorsqu'on étudie la ZZT, on peut voir trois domaines distincts :

- Un domaine purement formel où intervient principalement l'algèbre des polynômes,
- Le domaine temporel discret, identifié par les coefficients du polynôme,
- Le domaine fréquentiel discret, calculé à partir des zéros de ce même polynôme.

Il est important de noter que du point de vue formel on accède bien plus facilement au domaine fréquentiel discret lorsqu'on utilise la factorisation du polynôme. On peut donc voir la forme développée du polynôme comme sa forme temporelle, et sa forme factorisée comme sa forme fréquentielle. Comme le polynôme en Z est toujours exprimé dans le domaine Z (fréquentiel discret), cette distinction se base sur l'observation des composantes qui définissent le polynôme :

- les coefficients sont les échantillons,
- les zéros sont les points caractéristiques de la fonction de transfert.

5.2.1 Passage d'une forme à une autre du polynôme

En pratique il est loin d'être évident de passer d'une forme à une autre du polynôme. Le nombre de coefficients (et a fortiori de zéros) étant très élevé dans le cas d'analyse de signaux vocaux (plusieurs centaines d'échantillons par période), il n'existe pas de méthode formelle rapide pour factoriser ou développer un polynôme.

Factorisation : Dans cette étude nous utilisons la fonction `roots()` de MATLAB © qui utilise, pour trouver les zéros d'un polynôme, la diagonalisation de la matrice associée [16]. A savoir pour le polynôme :

$$c_4x^4 + c_3x^3 + c_2x^2 + c_1x^1 + c_0$$

Ses zéros sont les valeurs propres issues de la diagonalisation de la matrice :

$$\begin{array}{cccc} -\frac{c_4}{c_0} & -\frac{c_3}{c_0} & -\frac{c_2}{c_0} & -\frac{c_1}{c_0} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}$$

Développement : Le passage de la forme factorisée à la forme développée du polynôme pose d'énormes problèmes de précision au niveau du calcul. En effet, cette reconstruction se fait par le passage de produit entre les zéros, produits pondérés par le coefficient du binôme $\binom{n}{p}$ qui atteint des valeurs astronomiques pour un nombre de coefficient élevé. Dans la pratique, matlab est incapable de reconstruire un polynôme à partir de ses zéros si le nombre de ses zéros dépasse la centaine. Il est donc nécessaire de contourner le problème.

Ce problème est abordé dans [5], et il est proposé de construire le spectre à partir des zéros, puis d'atteindre le signal temporel (forme développée du polynôme) par transformée de Fourier inverse.

5.2.2 Organisation des Zéros pour un signal périodique

Soit un signal composé de KN échantillons tels que les échantillons $(k-1)N$ à $k(N-1)$ soient identiques aux échantillons 0 à $N-1$ (pour $k \in [0; K]$), il possède donc une période de N échantillons. Considérant la transformée en Z , $P(Z)$ de ce signal ; on sait que les coefficients $j+kN$ quelque soit $k \in [0; K]$ et pour $j \in [0; N-1]$ sont identiques, et on peut donc factoriser le polynôme P en le mettant sous la forme suivante :

$$P(Z) = \left(\sum_{i=0}^{N-1} x(i)Z^{-i} \right) \left(\sum_{k=0}^{K-1} Z^{-kN} \right)$$

Propriétés

On retrouve deux propriétés :

- Les zéros du polynôme P sont de deux types : une partie va définir une période du signal (membre de gauche, N zéros) et une partie va définir sa périodicité (racines n -ième de l'unité, ou **polynôme cyclotomique** [17]) membre de droite, $(K-1)N$ zéros). Ceci explique de manière simple et précise la présence d'un abondant nombre de zéros répartis sur le cercle unité lors d'un calcul de ZZT pour des signaux périodiques.
- On retrouve enfin la propriété qu'un signal périodique est la convolution d'une de ses périodes (membre de gauche) avec un peigne de Diracs (membre de droite, où on reconnaît la transformée en Z d'un peigne de Diracs)

5.3 Exemples

5.3.1 ZZT de la réponse impulsionnelle d'un AR1

Soit un filtre autorégressif d'ordre 1 tel que sa transformée en Z est :

$$H(Z) = \frac{1}{1 - aZ^{-1}}$$

Sa réponse impulsionnelle est :

$$h(n) = a^n$$

Calculer la ZZT de $h(n)$ pris sur N échantillons (figure 5.1), revient à trouver les zéros du polynôme $P(Z)$ tel que :

$$P(Z) = \sum_{i=0}^{N-1} Z^{N-1-i} a^i = Z^{N-1} \sum_{i=0}^{N-1} \left(\frac{a}{Z}\right)^i$$

Si Y_i sont les racines du polynôme $\sum_{i=0}^{N-1} Y^{-i}$ alors les zéros Z_i du polynôme $P(Z)$ sont : $Z_i = aY_i$ (décalage des zéros sur l'axe des modules pour $a \in \mathbf{R}$)

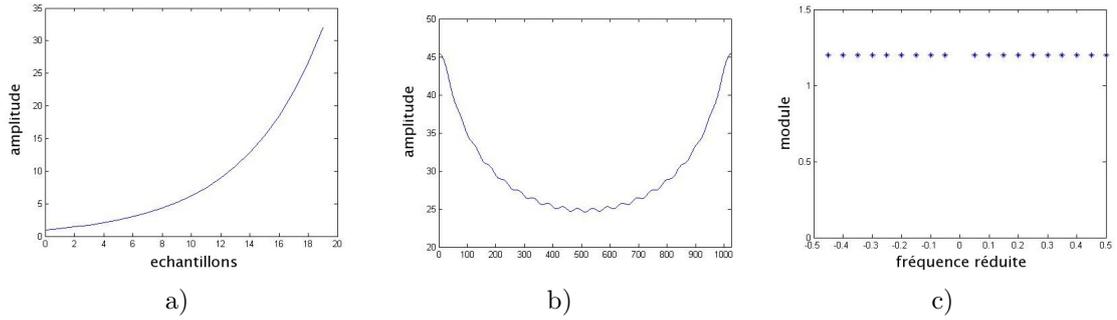


FIG. 5.1 – ZFT d'un AR1 pour $a = 1.2$ a) réponse impulsionnelle ; b) spectre sur 1024 échantillons ; c) représentation ZFT polaire

5.3.2 ZFT d'une sinusoïde

dans le cas de la sinusoïde montrée figure 5.2, on s'attend à avoir des zéros pour toutes les fréquences réduites sauf celles propres au signal. C'est en effet ce qu'on observe : il "manque" des zéros autour de la fréquence réduite $\nu = 0.1$, traduisant la présence de résonances visibles sur le spectre.

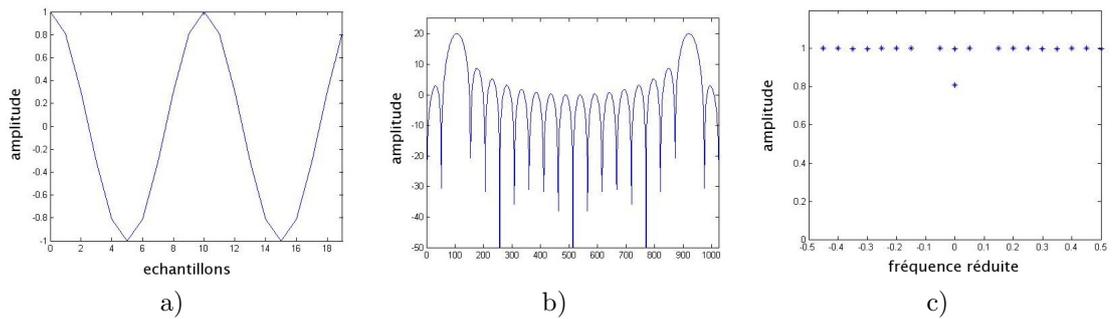


FIG. 5.2 – ZFT d'une portion de sinusoïde de fréquence réduite $\nu = 0.1$. a) réponse impulsionnelle ; b) spectre sur 1024 échantillons ; c) représentation ZFT polaire

5.4 Utilisation de la ZFT pour l'analyse de la voix

Dans cette section, il est question de l'utilité de la représentation par ZFT appliquée au signaux vocaux. Après la présentation des propriétés du modèle choisi, nous voyons la représentation de signaux synthétiques avant de présenter la mise en oeuvre de l'analyse pour des signaux réels.

5.4.1 Modèle causal/anti causal [6]

Le modèle LF vu précédemment est composé principalement de deux phases : la première phase correspond à la phase d'ouverture de la glotte, alors que la deuxième phase correspond à la phase de fermeture.

La première phase est modélisée par le produit d'une sinusoïde avec une exponentielle croissante. Ce signal est à maximum phase en raison de la présence de cette exponentielle. Les zéros dus à ce signal seront donc en dehors du cercle unité dans notre plan complexe.

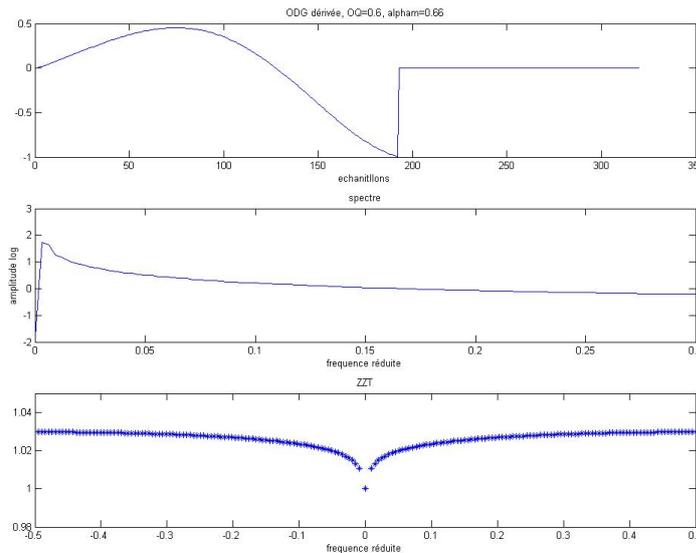
La deuxième partie, elle, se modélise par une exponentielle décroissante et est donc causalement stable : cette partie du signal est à minimum phase et ses zéros seront placés à l'intérieur du cercle unité.

C'est donc le caractère *phase mélangée* des signaux qui rend la représentation par ZZT intéressante. En effet, en prenant comme critère la localisation des zéros par rapport au cercle unité, il est possible de faire une déconvolution du signal vocal et d'en extraire *en un seul processus* le signal d'excitation glottique **et** la réponse impulsionnelle du conduit vocal [3].

5.4.2 ZZT de signaux synthétiques

Pour l'expérimentation et la validation des théories il est nécessaire de passer par des signaux synthétiques. Pour cela on génère une ou plusieurs périodes de signal à partir du modèle LF d'onde glottique.

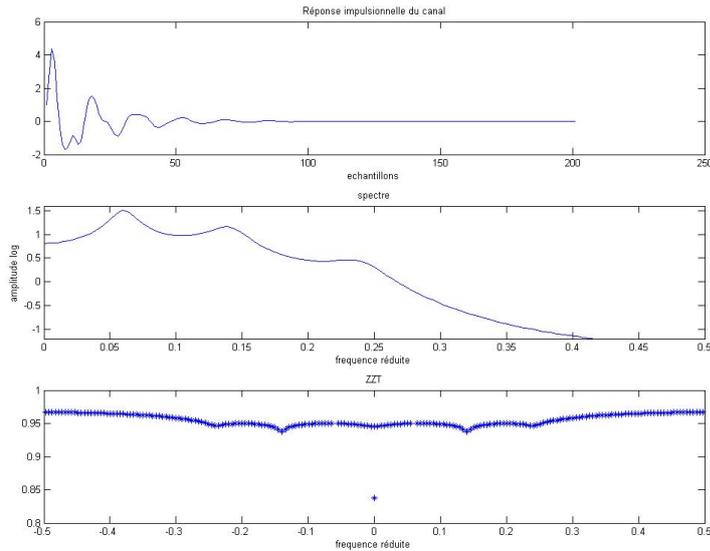
Dans un premier temps on génère une période de dérivée d'ODG, comme sur la figure 5.3, puis la réponse impulsionnelle d'un filtre multipôle (cf figure 5.4), et enfin la convolution des deux, sur la figure 5.5.



On y retrouve un spectre caractéristique possédant un pic pour $f = Fg$, pic se traduisant à nouveau par un "trou" dans la ZZT pour les fréquences réduites correspondantes.

FIG. 5.3 – Onde glottique : forme temporelle, spectrale, puis ZZT

Des trois représentations proposées, seule la ZZT permet la visualisation (et la déconvolution par critère sur l'emplacement des zéros) des deux composantes du signal.



On y retrouve un spectre multirésonant, et les "trou" dans la ZFT pour les fréquences réduites correspondantes.

FIG. 5.4 – Réponse impulsionnelle : forme temporelle, spectrale, puis ZFT

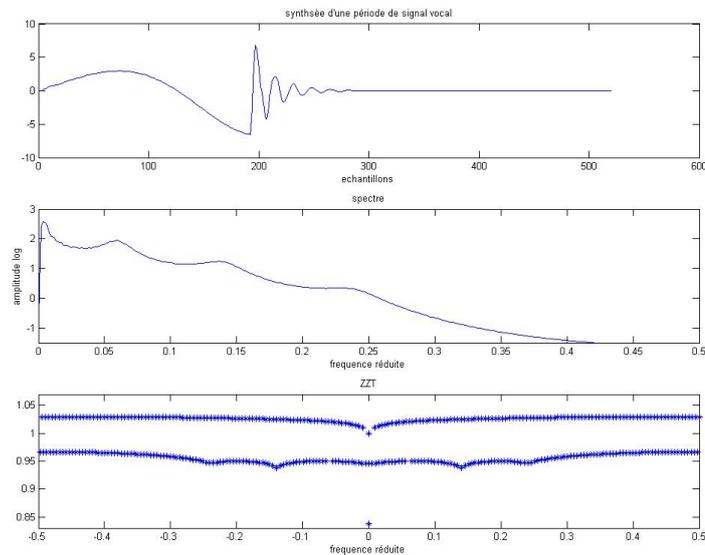


FIG. 5.5 – Période simulée d'un signal vocal : forme temporelle, spectrale, puis ZFT

5.4.3 Application à un signal réel [4]

Si, pour des signaux propres et synthétiques, la ZFT se pose comme un outil puissant d'analyse synthèse, un certain nombre de facteurs vont rendre très difficile l'analyse de signaux réels. La présence de bruit, la difficulté de la détermination du GCI et le manque de précision du calcul des zéros vont demander de prendre un certain nombre de précautions pour mener une analyse à bien.

Fenêtrage

Le fenêtrage est un des points essentiels décrit dans [5] [1]. Il est inutile ici de reprendre toute l'étude faite alors, mais l'utilisation d'une fenêtre, outre le fait d'exacerber d'un côté la phase anticausale, et de l'autre la phase causale, va permettre, si son choix est judicieux, de visualiser au mieux les zéros dans le plan complexe (en les éloignant du cercle unité).

En effet, les zéros des signaux réels sont naturellement très proches du cercle unité. Pour éviter que les erreurs de calcul de Matlab ne les placent du mauvais "coté" du cercle, un fenêtrage tel que celui de Blakman, va permettre de multiplier chaque phase par sa forme "exponentielle" correspondante (croissante ou décroissante pour respectivement la phase anticausale ou causale). Nous avons vu précédemment que multiplier un signal par une forme exponentielle déplace ses zéros.

Détection des instants de fermeture glottique

Afin que le fenêtrage du signal fonctionne au mieux, il faut placer le centre de la fenêtre à l'instant où s'arrête la phase d'excitation du conduit vocal et où commence sa réponse. Cet instant est l'instant de fermeture glottique (GCI) [5].

Il existe une littérature très fournie sur les différentes techniques de détermination de cet instant, littérature qui pourrait faire elle même l'objet de toute une étude. Cet aspect ne sera donc pas étudié dans le cas présent ; mais il est essentiel de rappeler que la détection du GCI est un point crucial pour l'organisation de la ZZT.

Exemple

Si la représentation par ZZT du signal, figure 5.6 met bien en évidence les deux phases du signal, on remarque que le spectre de la source n'est pas exactement celui attendu (une seule résonance). En effet, l'effet du canal, contrairement au signal de synthèse présenté ci avant, est aussi présent (mais moindre) sur la partie anticausale du signal : on voit en particulier que les pics de résonance du canal correspondent à des creux de la TF de la source.

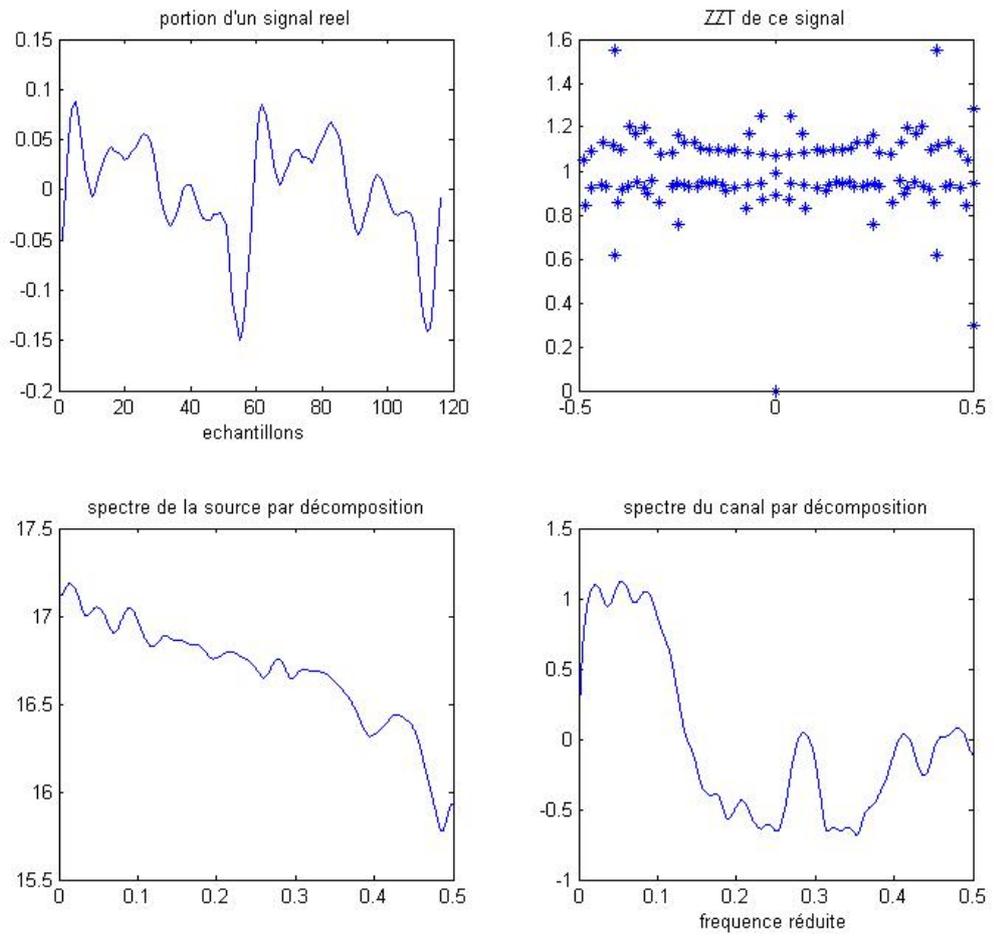


FIG. 5.6 – ZFT d'un signal vocal : période à analyser (GCI au centre de la fenêtre), ZFT du signal, spectre de la source, spectre du canal

Deuxième partie

Estimation des propriétés basse fréquence de la source glottique

Chapitre 6

Déconvolution de la source par ZZT : implémentation

L'analyse par ZZT nous permet d'accéder au spectre de la source glottique. Nous avons vu que, selon son modèle générique, ce spectre possède un maximum appelé *formant glottique* dont la fréquence correspondante est principalement fonction de 3 des principaux paramètres du modèle : O_q , T_0 et α_m .

Mesurer F_g , en concordance avec une mesure de O_q et T_0 , permettrait donc d'atteindre un paramètre de la source non encore mesuré jusqu'alors : le coefficient d'asymétrie α_m .

6.1 Implémentation de la ZZT

L'algorithme décrit dans [4] et [5] nécessite un haut degré de compréhension des mécanismes sous jacents à la ZZT. Lors de sa mise en oeuvre, on se rend compte que :

- **Le fenêtrage corrompt la représentation des signaux**, et rend indispensable son centrage sur le GCI. La période fenêtrée avant le GCI verra ses zéros déplacés vers l'extérieur du cercle unité, alors que la partie post GCI verra ses zéros déplacés vers l'intérieur de ce cercle. Ceci est problématique surtout au niveau de forte résonance du canal vocal qui vont être présentes dans la partie anticausale du plan.
- **Le positionnement de la composante continue est ambigu**, et un certain nombre de zéros sur l'axe des réels (fréquence nulle) ne sont pas placés du bon côté. Le problème peut être résolu en choisissant judicieusement un zéro à inclure dans la partie *anticausale*, mais le choix de ce zéro reste ambigu et il n'existe aucun moyen de le déterminer avec certitude.
Le formant glottique étant généralement positionné très bas dans le spectre, le problème posé est évident, car un mauvais choix du zéro de la composante continue va fortement influencer sur la position du formant glottique.
- **Une instabilité de la décomposition en hautes fréquences** trouve probablement sa source dans les erreurs de calcul générées par la fonction matlab. Il en résulte un mauvais placement des zéros, notamment en haute fréquence, par rapport au cercle unité.

6.1.1 Composante continue

Le principal problème de la mesure du formant glottique par ZZT est sa position en fréquence. Le *trou* qui matérialise le formant se trouve souvent entre le zéro de la fréquence nulle et le zéro suivant.

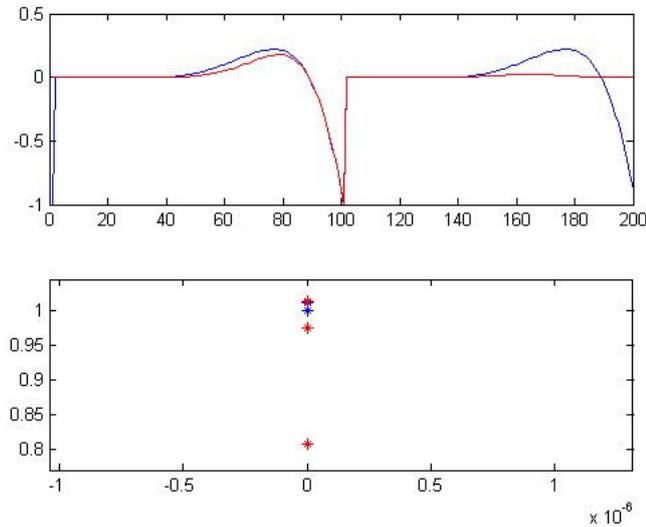


FIG. 6.1 – Déplacement des zéros de la composante continue sur un signal synthétique. En rouge, signal fenêtré, en bleu signal non fenêtré ; en bas, position des zéros de la fréquence nulle.

Principal problème : nous l'avons vu, le zéro à fréquence nulle matérialise l'absence de composante continue. De manière intrinsèque, la composante continue est difficilement caractérisable comme causale ou anticausale ; de plus, le fenêtrage rend encore plus difficile le bon placement de ce zéro. Le fenêtrage va rendre non nulle la moyenne de l'échantillon de signal analysé et donc déplacer les zéros d'argument 0 loin du cercle unité (en dedans ou en dehors) comme le montre la figure 6.1.

Or, nous l'avons dit précédemment, le placement de ce zéro joue de manière prépondérante sur la détection du formant glottique. En l'absence de ce zéro, le formant du spectre de la source ainsi reconstituée se trouve autour de 0Hz, alors que si le zéro est mal placé, sa valeur peut varier jusqu'à plus de 25% de la valeur attendue de F_g . Erreur totalement inacceptable.

A l'instar du travail de Baris Bozkurt, l'expérience nous montre que :

Si un zéro se trouve tel que son module est compris entre 1 et 1.05 ($1 < \|z_i\| \leq 1.05$) alors le spectre reconstruit est conforme à ce qu'on pourrait attendre

Dans le cas contraire, un zéro se trouve nécessairement proche de 1 de l'autre côté du cercle unité. Ce zéro, généralement compris entre 0.98 et 1, est alors enlevé de la partie causale et inséré dans la partie anticausale.

Il semble cohérent d'établir un tel équilibre, car nous savons que la source recherchée est la dérivée de l'onde glottique, donc ne possède pas, ou très peu, de composante continue ; alors que le conduit vocal ne doit logiquement pas présenter de coupure des basses fréquences.

6.2 Plage de Recherche

Nous savons, d'après le modèle de la source glottique retenu, que le formant glottique se trouve dans les basses fréquences. En règle générale, ce formant se trouve entre $\frac{F_0}{2}$ et $4F_0$; c'est dans cette plage là que la recherche va se concentrer.

Comme nous disposons des zéros de la fonction de transfert, nous pouvons obtenir le spectre complexe pour n'importe quel jeu de fréquence : le calcul du spectre est alors vu comme une interpolation de l'information donnée par les zéros. L'algorithme calculera donc le spectre sur une plage bien précise, fonction de F_0 , et sur un grand nombre de points.

6.3 Evaluation de la précision de l'analyse

6.3.1 Déplacement de F_g par ZZT

En générant un signal de synthèse, on s'aperçoit que la décomposition par ZZT déplace le formant glottique. Ce déplacement se trouve être fonction de O_q , T_0 et α_m . Nous l'avons vu précédemment, ce déplacement du formant glottique trouve probablement sa source dans la position du zéro de fréquence nulle déplacé par fenêtrage et maladroitement replacé lors de la séparation des zéros.

Ce déplacement semble somme toute être systématique, et peut donc être compensable afin d'augmenter conséquemment la précision de la mesure.

6.3.2 Influence des erreurs sur les zéros

Il est très difficile de quantifier l'erreur sur les zéros. Ils sont déterminés de telle manière à ce que l'erreur sur le polynôme tout entier soit minimale. En faisant la différence entre un signal recomposé après une ZZT et son original, on obtient un bruit de l'ordre de 10^{-16} (soit plus de $300dB$, Matlab travaillant en 64bits), mais lors de la décomposition, l'erreur sur chaque composante séparée, par rapport à son originale peut dépasser les $120dB$. La décomposition est donc réalisée telle que les erreurs sur chaque zéro se compensent de manière globale (cf figure 6.2).

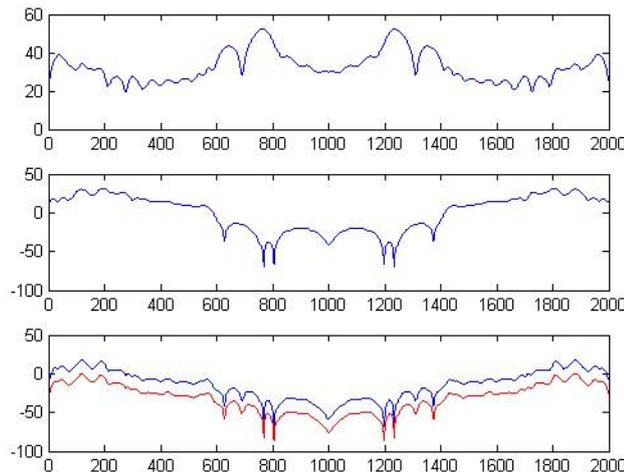


FIG. 6.2 – Illustration de l'imprécision du calcul des racines

La source de ce problème est que suite à l'imprécision du calcul un zéro peut aisément passer d'un côté à l'autre du cercle unité. Si ce passage n'a que très peu d'effet sur le spectre global, il va bien sur influencer de manière notable le spectre de chaque composante.

Nous voyons sur la figure 6.2 (analyse d'une trame de l'échantillon N9 de la base constituée), que les bosses du spectre supérieur sont placées au même endroits que les creux du spectre inférieur. Cet exemple illustre parfaitement le cas où des zéros fortement résonants changent subitement de coté du cercle unité (de 0.999 à 1.002 par exemple, soit une erreur a priori minime)

Cette erreur, principalement présente dans les hautes fréquences, est limitée par les précautions prises pour rendre la mesure robuste (recherche du maximum sur une plage réduite, déplacement de certains zéros etc...). Il en découle que ce problème a un effet minime sur la plage spectrale qui nous intéresse. Par ailleurs, il n'existe aucune méthode pour quantifier cette erreur, à ce jour.

6.3.3 Conclusion

La précision de la mesure sur F_g varie donc en fonction de nombreux facteurs. En sachant d'ores et déjà qu'on peut atteindre une erreur *systematique* supérieure à 30% sur F_g il faut rappeler que nous cherchons à mesurer au final le coefficient d'asymétrie qui est fonction de O_q et $\frac{F_0}{F_g}$.

La précision est F_g est un facteur critique d'une bonne mesure de α_m et ne saurait souffrir d'une erreur de mesure supérieure à un JND sur le paramètre en question. Il est donc important de mettre tous les moyens en oeuvre pour chercher à limiter et corriger cet erreur afin d'avoir une mesure fiable.

6.4 Mesure de α_m

Une fois mesuré F_g , nous pouvons tenter de déterminer un nouveau paramètre de la source glottique. En s'aidant de signaux EGG pour déterminer le quotient ouvert de la période de signal en question, nous pouvons, grâce à une table précalculée, estimer la valeur de α_m associée au couple $(\frac{1}{T_0 F_g}, O_q)$ mesuré.

Nous avons montré, lors des points précédents, que la mesure de F_g par ZZT était tres éloignée de la réalité mais de manière systématique, c'est pourquoi nous ne nous basons pas sur une expression théorique de la relation $\alpha_m = R(O_q, F_g, T_0)$ mais que nous une table (exemple sur la figure 6.3) afin de corriger l'erreur systématique sur F_g et retrouver la valeur originale de la fréquence du formant..

Cette méthode est à la fois la plus simple et la plus fiable : elle permet une rapidité de calcul et une réalité vis à vis de la mesure physique. Avec une telle méthode on peut s'attendre à une mesure de F_g précise à moins de 10% près.

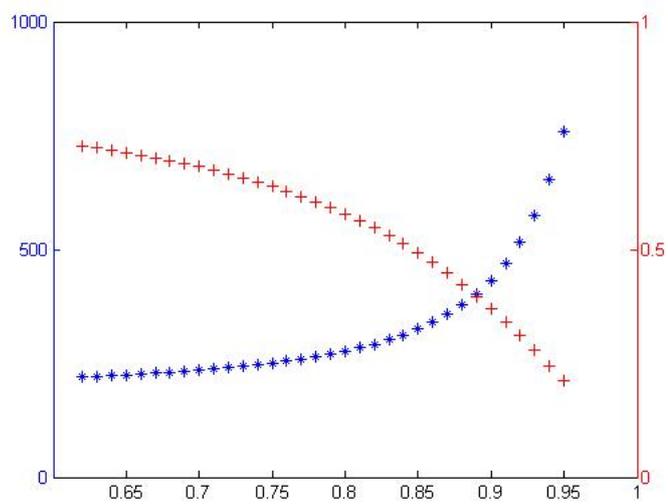


FIG. 6.3 – Valeurs prises par F_g pour ($f_0 = 160Hz; O_q = 0.6$) en fonction de α_m en bleu, rapport $\frac{f_0}{F_g}$ en rouge.

Chapitre 7

Etude de la robustesse

Cette méthode d'analyse étant destinée à une utilisation sur une base de données, il est indispensable d'en étudier la robustesse dans un premier lieu, afin de comprendre l'origine des erreurs ou impossibilités de mesure.

L'expérimentation permet de distinguer les deux facteurs les plus importants, constituant les fondations du protocole expérimental. La cohérence des mesures va de concert avec le respect d'une marge d'erreur que nous définirons sur ces deux facteurs.

Nous verrons ensuite les autres points moins critiques pour la cohérence des mesures, qui nécessitent tout de même d'être critiqués ou évalués, avant de s'attaquer à l'application de la méthode sur une grande base de données.

7.1 Vis à vis d'une mauvaise détermination du GCI

Il est clairement établi dans [5], ainsi que dans tout ce que nous avons dit au paravant que la détermination de la position du GCI est critique pour le calcul de la ZZT. Cependant, il n'existe aucun moyen, à part peut-être l'EKG, de connaître avec précision l'emplacement de ce GCI. Et même en utilisant les enregistrements EKG pour pointer les GCI sur un signal acoustique associé, il existe une certaine imprécision de la détection du GCI (retard entre la prise de son et la prise de signal EKG, non systématique car fonction des mouvements du locuteur).

De quelle ampleur est l'erreur sur la détermination de F_g lorsque le GCI est mal placé ? A partir de quel moment la mesure devient-elle inexploitable, et quelle erreur reste acceptable ? C'est ce que nous allons montrer, tout d'abord sur des signaux synthèse où F_g est précisément connu, puis sur des signaux réels.

7.1.1 Sur des signaux de synthèse

L'expérimentation montre que l'amplitude de résonance des formants ainsi que leur placement a un effet radical sur la plage acceptable d'erreur du GCI ainsi que la précision de la mesure de F_g .

Cas typique Sur cet exemple, figure 7.1, on prend deux formants vocaliques à 100 et 400 Hz, leurs pôles correspondant ont respectivement un module de 0.95 et 0.97. Le graphique du milieu de cette figure nous montre qu'au GCI effectif, l'erreur est bel et bien de 0% par rapport à F_g mesuré par ZZT, et de l'ordre de 20% par rapport au F_g réel. Non seulement ceci démontre que l'analyse par ZZT déplace de manière systématique le formant glottique,

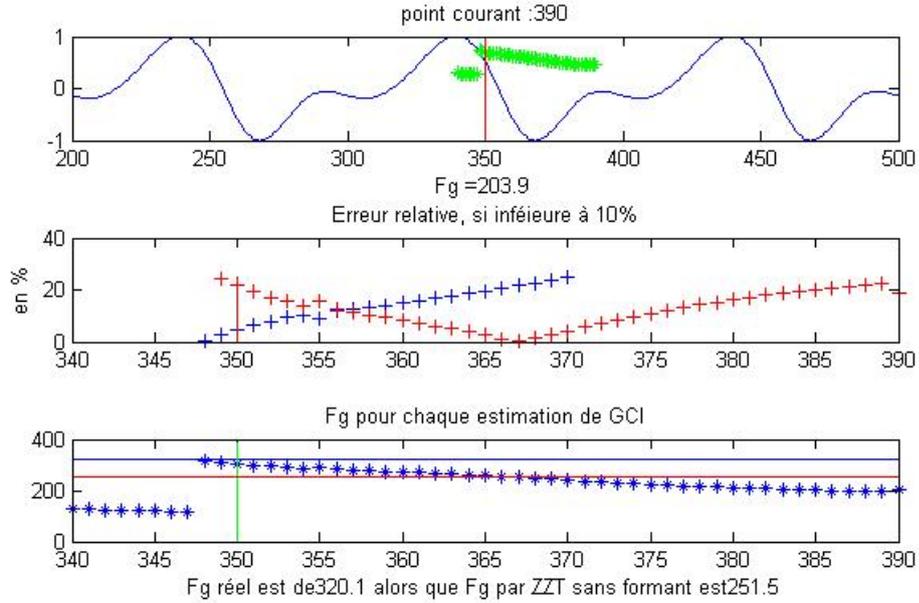


FIG. 7.1 – Influence de la détection de GCI pour un filtre vocalique faiblement résonant. **en haut** : Le signal synthétique et en vert, la variation relative de \tilde{F}_g mesurée au point en rapport. **au milieu** : L'erreur relative entre la mesure et, en bleu la valeur réelle de F_g , en rouge la valeur \tilde{F}_g mesurée au GCI sur la dérivée de l'ODG source par ZZT **en bas** : Les valeurs mesurées. en trait bleu F_g , en trait rouge \tilde{F}_g .

mais cela démontre qu'une erreur de positionnement sur le GCI entraîne nécessairement une erreur supplémentaire de positionnement du formant glottique. L'erreur sur \tilde{F}_g reste inférieure à 10% dans une plage de 10 points autour du GCI effectif du signal filtré (aux alentours du point 367).

En conséquence, on remarque que la précision sur le GCI est un facteur critique vis à vis de la bonne mesure de \tilde{F}_g . Mais une variation sur la position de GCI inférieure à 10% de la période ne donnera pas des résultats incohérents.

Accrochage sur un formant Dans le deuxième cas, figure 7.2, le filtre vocalique est synthétisé pour être très résonant (largeur de bande $< 60\text{Hz}$) et en particulier à 600Hz , fréquence proche de F_g (environ 500Hz). On remarque alors que sur toute la plage de variation de GCI, la détection *accroche* sur le premier formant à 600Hz , ne donnant aucun résultat valable. (graphique du milieu vide) Bien plus que la position du GCI, c'est la résonance des formants résidus de la décomposition qui induisent ici la décomposition en erreur.

Pour augmenter la précision de la mesure, nous utiliserons les signaux EGG pour déterminer avec précision les GCI du signal, et nous compenserons le retard acoustique.

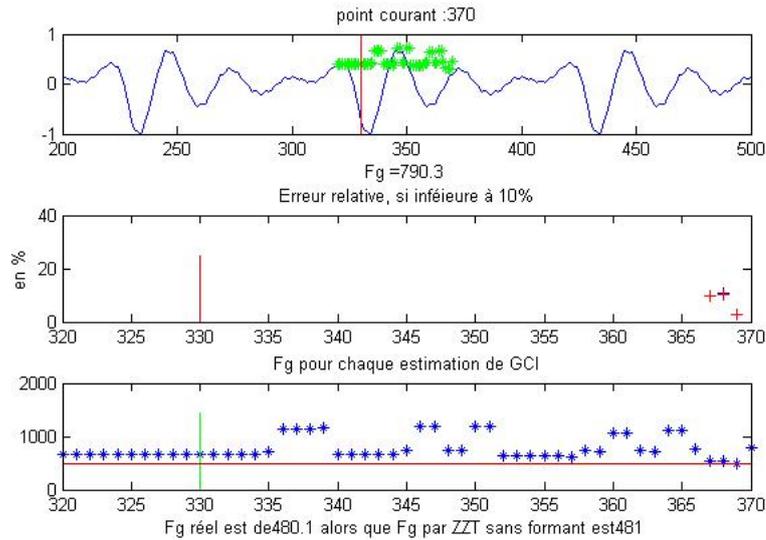


FIG. 7.2 – Détection de F_g accrochant sur un formant, l'absence de point sur le graphique du milieu nous indique qu'aucun point ne présente une erreur relative inférieure à 25%

7.1.2 Sur des signaux réels

Sur des signaux réels, il n'existe aucun moyen à ce jour de prédire la valeur du formant glottique que nous cherchons à évaluer. Nous ne verrons donc que deux cas différents, un où la plage de variation acceptable du GCI est large (cf 7.3), dans les 10% de la période, et un où la mesure ne semble cohérente que sur un nombre très restreint de points (cf 7.4). Un dernier cas (figure 7.5) illustre la possibilité de ne pas arriver du tout à estimer F_g même avec une mesure au GCI.

Une mesure sera considérée cohérente si F_g est dans une plage acceptable vis à vis de O_q et si les spectres visualisés de concert sont cohérents avec la mesure.

Dans un premier cas, figure 7.3, nous voyons que la décomposition est relativement stable dans une bonne plage autour du GCI. L'erreur calculée par rapport à F_g est mesurée au GCI effectif, qui est supposé être le F_g réel, moyennant erreur par ZTT.

Dans le cas de la figure 7.4, les mesures obtenues sont nettement moins stables, mais une latence est toujours acceptable autour du GCI. La mesure s'avère néanmoins nettement moins précise.

Une fois encore, nous voyons que le positionnement du GCI est critique pour arriver à estimer correctement le formant glottique. Mais à nouveau, une certaine imprécision est acceptable pour la position du GCI, dans la mesure où elle ne dépasse pas 10% de la période du signal.

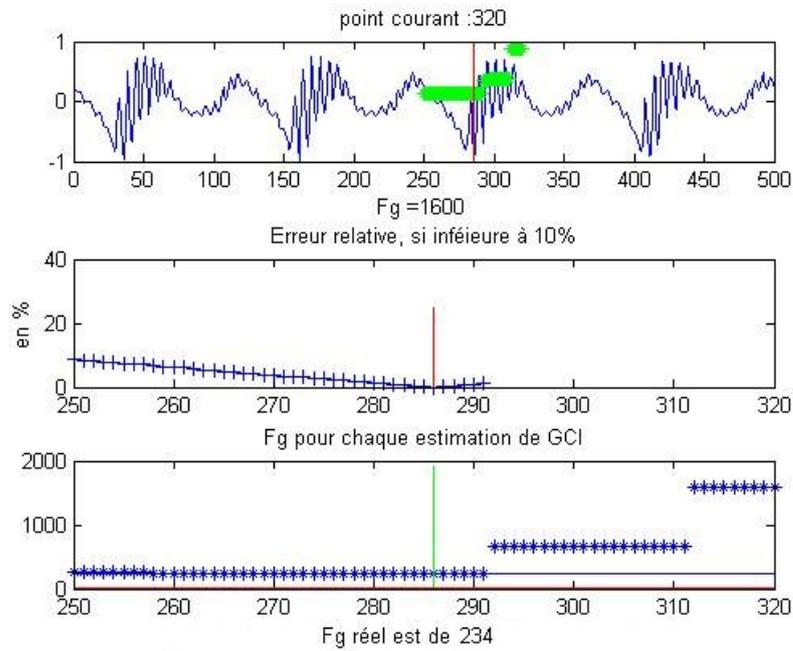


FIG. 7.3 – Evolution de la mesure de F_g pour un signal réel, cas où la marge d'erreur sur GCI est grande. **en haut** signal, **au milieu** erreur par rapport à la mesure au GCI, **en bas** valeur des F_g estimés

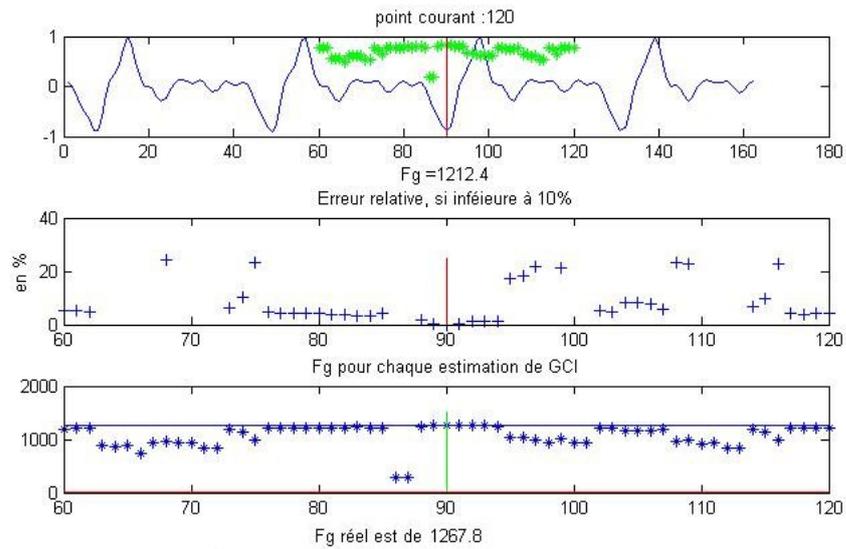


FIG. 7.4 – Evolution de la mesure de F_g pour de la voix chantée en mécanisme II

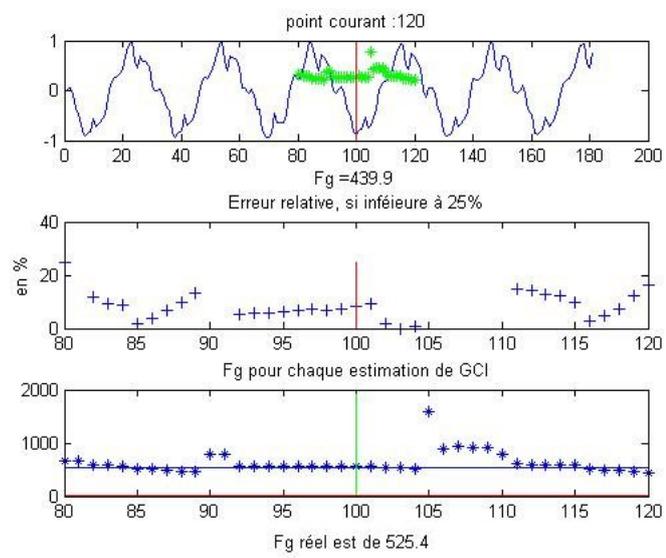


FIG. 7.5 – Cas où la mesure présente une erreur importante très vite lorsqu'on s'éloigne du GCI.

7.2 Vis à vis d'une mauvaise décomposition

On désigne par mauvaise décomposition, le cas où l'algorithme n'arrive pas à dissocier complètement le spectre de la source et celui du canal. Dans ce cas, la mesure de F_g peut être faussée comme nous venons de le voir.

7.2.1 Origine

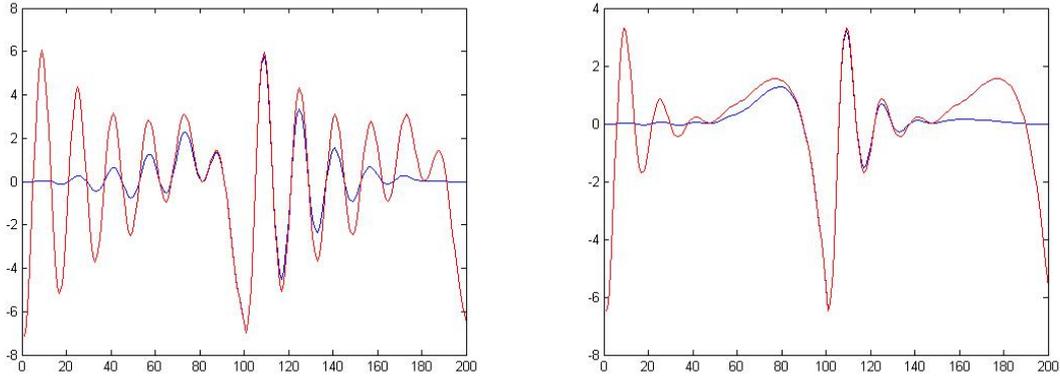


FIG. 7.6 – Influence du fenêtrage sur un signal synthétique, formant fortement résonant à gauche, faiblement résonant à droite. En rouge le signal original, en bleu le signal fenêtré. On voit que dans certains cas le fenêtrage donne une forme exponentielle croissante à la R.I. du canal vocalique.

Un des phénomènes le plus problématique du fenêtrage est son action sur les formes temporelles des signaux. Comme nous le voyons sur la figure 7.6, une forte résonance du conduit vocalique peut être rendue exponentielle croissante par le fenêtrage, et on trouvera donc un pic associé à ce formant vocalique dans la partie source de la décomposition.

La figure 7.7, donnant la décomposition spectrale d'un signal réel nous montre dans quelle mesure les formants vocaliques peuvent être présents dans le spectre de la source

Dans certains cas, ce phénomène ne pose pas de problèmes pour l'analyse de signaux vocaux, en particulier pour T_0 élevé et des réponses impulsionnelles du filtre vocal plutôt courtes. De cette manière, la phase d'ouverture ne comportera que peu de résidus de la réponse du canal vocalique et sa multiplication par la partie croissante de la fenêtre va faciliter la séparation des zéros.

Dans le cas d'un canal très résonant (R.I. très longue) et/ou de T_0 court ; les parties source et filtre ne sont pas dissociables temporellement, et on observe une décomposition incomplète.

7.2.2 Masquage de F_g

Cette décomposition incomplète peut avoir une forte conséquence sur la mesure de F_g quand α_m tend vers 1. En effet, dans ce cas la largeur de bande du formant augmente, et il voit son amplitude largement diminuée. Une décomposition incomplète peut alors mener au masquage du formant glottique par un formant vocalique proche, et la mesure de F_g se verra radicalement faussée.

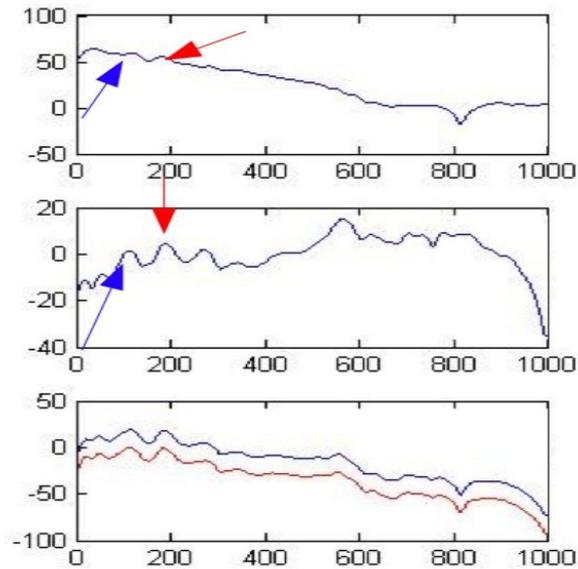


FIG. 7.7 – Décomposition incomplète, spectres. **en haut** spectre estimé de la source ; **au milieu** spectre estimé du canal vocalique ; **en bas** spectre original en bleu, spectre recomposé en rouge. *Trame extraite de l'échantillon N9*

7.2.3 Limites de la mesure

L'expérience nous permet de déduire que la limite de mesurabilité est fonction de la résonance du premier ou du deuxième formant vocalique. Cela dit, il n'y a guère de chance d'arriver à mesurer de valeurs de α_m supérieures à 0.9 sans se poser des questions sur la fiabilité de la mesure.

7.3 Autres points importants

7.3.1 Rapport signal bruit

On s'aperçoit par le biais de la figure 7.8, que la présence de bruit n'est pas un facteur réellement limitant dans l'analyse des signaux. En effet la méthode s'avère très robuste de ce point de vue, et on arrive encore à analyser facilement des signaux au RSB supérieur à 20 dB.

Comme on cherche un maximum local sur le spectre, un bruit blanc ne faussera pas les mesures, du moment qu'il n'a aucun effet sur la précision du calcul, notamment des racines du polynôme associé.

7.3.2 Fréquence fondamentale

De la fréquence fondamentale du signal va dépendre le nombre d'échantillons sur lequel va se porter l'analyse. Un grand nombre d'échantillons va être synonyme d'une grande précision fréquentielle tant du point de vue de la position spectrale du formant glottique que de la précision de la décomposition ZZT. On peut logiquement prévoir que les résultats de l'analyse seront bien meilleurs pour des signaux à grand T_0 (sujets masculins).

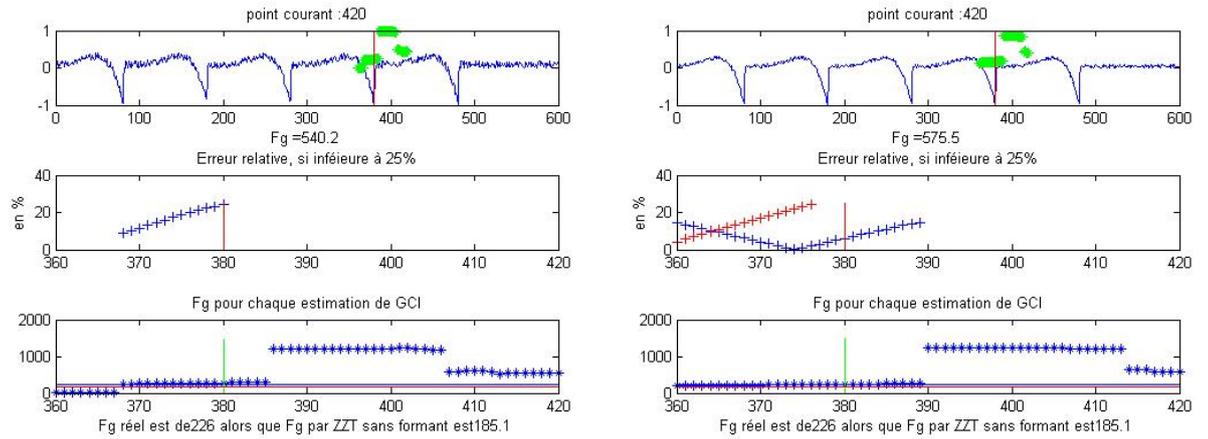


FIG. 7.8 – Influence du Rapport signal sur bruit. A droite un RSB de 18 dB, à gauche 12 dB. On voit que la méthode est encore efficace pour 18 dB, mais ne l'est plus à 12 dB

7.3.3 Quotient ouvert

L'expérience montre que le quotient ouvert n'a que peu d'effet sur la précision de la mesure. On peut donc s'attendre à une mesure homogène sur toute la plage de variation de O_q . Cependant, une forte valeur de O_q va souvent de paire avec des sons émis en deuxième mécanisme vocalique, donc faible période fondamentale. Dans ce cas, une augmentation de O_q pourra se trouver fortement corrélée avec une baisse de fiabilité de la décomposition.

7.4 Résumé

Si la méthode de décomposition par ZZZT s'applique convenablement à la théorie, sa mise en oeuvre se base sur un certain nombre de méthodes qui ne s'appliquent pas bien sur des signaux réels. La détection de GCI pourrait être le sujet d'une étude à part entière tant le problème est vaste vis à vis de son application à la ZZZT ; les imprécisions de calculs sur les racines et le fenêtrage sont des points faibles qui remettent sérieusement en cause la mise en oeuvre d'une analyse systématique de la voix.

Troisième partie

Analyse d'une base de données d'échantillons

Chapitre 8

Base de données

Les mesures effectuées étant les premières dans le genre, il a été nécessaire de constituer une base de données de signaux couplés Voix-Egg. Ce chapitre s'articule autour du protocole expérimental mis en place pour l'enregistrement et la classification des échantillons.

Nous verrons enfin la repartition de ce corpus.

8.1 Enregistrement de la base

L'enregistrement s'est articulé autour de la variation de deux paramètres de la source glottique : α_m et O_q .

Variation de α_m : Ce paramètre est fortement lié à la largeur de bande du formant glottique. Or, nous savons que perceptivement, un signal à bande large sera perçu avec une intensité acoustique subjective plus forte. Ainsi, nous avons recherché à mesurer des variations de α_m à travers une variation de l'intensité et de l'effort vocal.

Variation de O_q : On déplace le quotient ouvert en *serrant* la voix. Diminuer O_q c'est chercher à augmenter la pression sur le larynx (abduction), ce qui revient à déplacer le formant glottique dans les hautes fréquences.

8.1.1 Protocole expérimental

- Tous les enregistrements se font à f_0 **quasi contsant**
- On ne cherchera pas à distinguer la voix chantée de la voix parlée. Les efforts à produire pour fournir les échantillons ont amené les sujets à souvent changer entre ces deux types.
- 3 voyelles ont été retenues pour les tests : I et OU ; pour la position basse en fréquence de leur premier formant vocalique, et A pour sa position haute dans le triangle vocalique, et le fait qu'elle n'ai justement pas de premier formant vocalique bas.
- Le registre vocal sera aussi pris en compte, aussi a-t-on demandé à chaque sujet de préciser le registre dans lequel il émettait l'échantillon en question

Les échantillons obtenus sont des paires d'échantillons monophoniques, enregistrés de manière synchrone. Le premier canal est l'enregistrement du signal EGG à la sortie directe de l'electroglottographe ; le deuxième canal est l'enregistrement du signal acoustique associé.

8.2 Classification des échantillons

Nombre total d'échantillons : 71

Locuteur	Nombre d'échantillons total	En mécanisme I	En mécanisme II
C	22	15	7
M	19	10	9
N	30	30	0
Total	71	55	16

O_q moyen	Nbre	f_0 moyen en Hz	Nbre	Phonème	Nbre
$0.3 \leq . < 0.5$	4	< 140	12	A	36
$0.5 \leq . < 0.6$	28	$140 \leq . < 260$	43	I	16
$0.6 \leq . < 0.7$	24	$260 \leq . < 400$	12	OU	19
$0.7 \leq . < 0.9$	15	$400 \leq .$	4		

FIG. 8.1 – Population de la base de données

Mécanisme : Chaque locuteur a précisé, après l'émission d'un phonème, dans quel mécanisme il était. Une grande partie des échantillons sont en mécanisme 1 ; les quelques échantillons en mécanisme II ne sont étudiés que pour définir les limites de fonctionnement de la méthode.

O_q moyen : La valeur O_q moyen est définie par le calcul de la moyenne de toutes les valeurs de O_q de l'échantillon en question.

f_0 moyen : Les expériences ont été faites en cherchant à avoir f_0 constant, la valeur retenue est la moyenne de la mesure sur toutes les trames.

Chapitre 9

Evaluation et résultats

Après avoir implémenté les algorithmes et analysé leur comportement, défini un corpus d'étude qui couvre assez largement le domaine de mesure, nous pouvons donc nous attaquer à l'étude de la robustesse de la mesure de α_m sur une base de données.

9.1 Protocole d'évaluation

Pour ce faire on utilise le principe suivant :

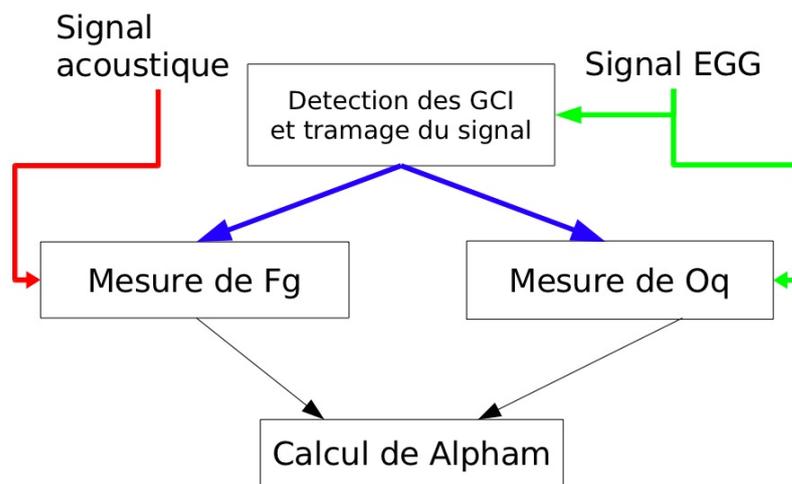


FIG. 9.1 – Organigramme du script d'analyse

Dans un premier temps, on utilise le signal EGG pour effectuer une détection des GCI. Avec les instants ainsi obtenus, on arrive non seulement à la détermination de f_0 , mais aussi à une décomposition *pitch synchrone* du signal, décomposition qui sera utilisée dans les processus suivants.

Dans un deuxième temps, on réalise la mesure de O_q sur la forme temporelle de la DEGG. Cette mesure n'est certes pas d'une grande précision, mais largement assez au regard de la précision de mesure des autres paramètres, ainsi que du JND sur O_q .

Dans un troisième temps, c'est la fréquence du formant glottique que nous cherchons à déterminer, évidemment par ZZT, sur des trames de longueur $2f_0$ centrées sur les GCI. On

peut choisir, pour accélérer l'analyse, de ne mesurer F_g que pour une trame sur 10 ou 20 ; dans la mesure où les signaux sont à variations lente, comme dans la base qui nous intéresse.

Finalement, nous cherchons quelle valeur de α_m représente le meilleur compromis pour atteindre la fréquence F_g du formant glottique pour le quotient ouvert mesuré. Une succession de spectres de source glottique est donc modélisée, et nous choisissons la valeur de α_m correspondant au spectre donnant le maximum le plus proche de la valeur de F_g mesurée. (calcul de la table de triplets $(\alpha_m, \frac{F_0}{F_g}, O_q)$).

9.2 Critères d'évaluation

On considérera une mesure réussie si elle correspond à certains critères :

- Les valeurs de F_g trouvées permettent d'obtenir des valeurs de α_m cohérentes. Le modèle considère $0.6 < \alpha_m < 0.99$, nous nous basons sur cette considération. Compte tenu des difficultés de mesure d'un $\alpha_m > 0.9$ mises en exergue dans le chapitre précédent, nous pouvons nous attendre à ce qu'une mesure supérieure à 0.9 soit un résultat faussé.
- Que les mesures F_g ne soient pas le résultat d'un accrochage sur un formant vocalique. Pour ce faire on peut visualiser les représentations temps-fréquence de la décomposition.

La partie concernant la détection de GCI est prépondérante pour l'analyse, mais ne fait aucunement partie du problème posé pour le stage. Il va de soi qu'un fichier ne sera comptabilisé dans le corpus des fichiers à "analyse défailante" qu'après avoir vérifié que cette défaillance n'est pas le résultat d'une mauvaise détection du GCI.

Fort heureusement, l'utilisation des signaux EGG pour la détermination du GCI est très efficace et ne pose problème que dans très peu de cas.

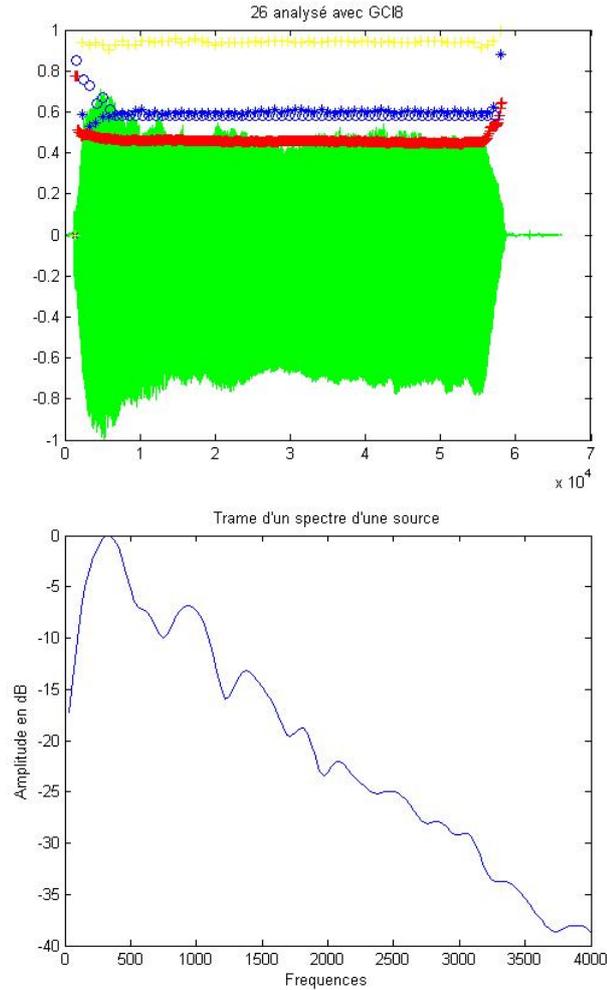


FIG. 9.2 – Une analyse valide (échantillon N26). **En haut** : résultats de l’analyse, mesures de OQ en rouge, de F_g en étoiles bleues et de α_m en ronds bleu. **En bas** : un spectre de la source, obtenu par décomposition ZZT.

9.3 Interprétation des résultats

Dans un premier temps, nous présentons les résultats de la même manière que nous avons classifié les échantillons de la base de données. Nous donnons 4 qualités possibles à une analyse :

- Analyse Bonne (1) : plus de 90% des valeurs mesurées sont cohérentes.
- Analyse Moyenne (2) : entre 30 et 90% des valeurs mesurées sont cohérentes.
- Analyse Mauvaise (3) : entre 5 et 30% des valeurs sont cohérentes.
- Analyse défailante (0) : moins de 5% de valeurs cohérentes, l’analyse est un échec.

9.3.1 Influence du phonème et de la période du signal

Nous remarquons que l’analyse est particulièrement efficace pour les sons à grande période fondamentale et notamment le phonème A. Dans ces cas, on obtient alors une décomposition presque parfaite, et des mesures cohérentes de α_m et F_g . Alors que les grandes périodes

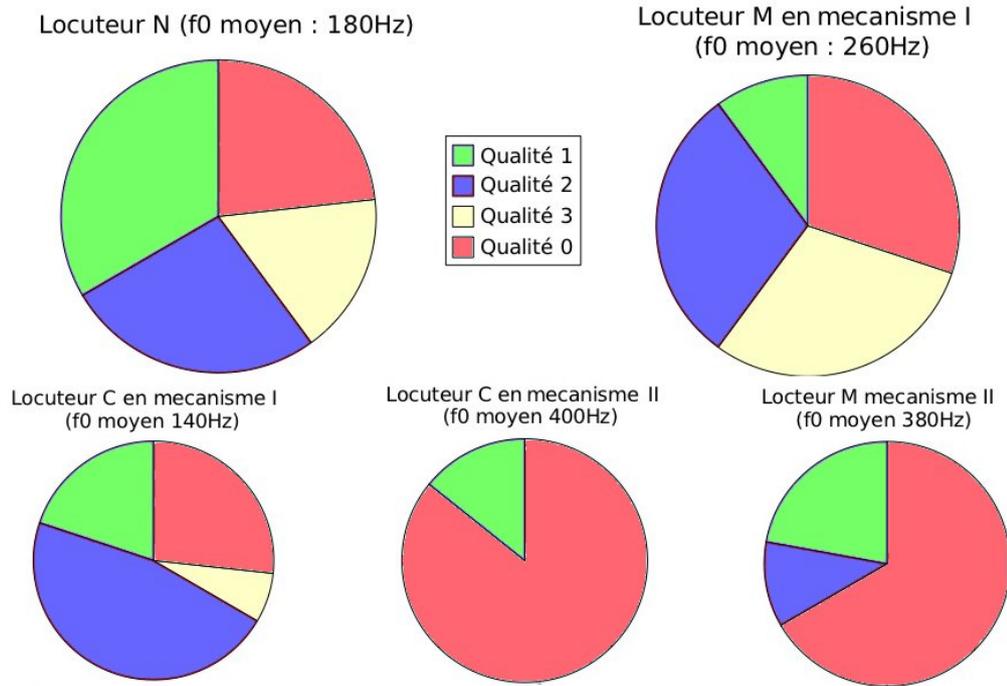


FIG. 9.3 – Résultats d’analyse par locuteur et mécanisme

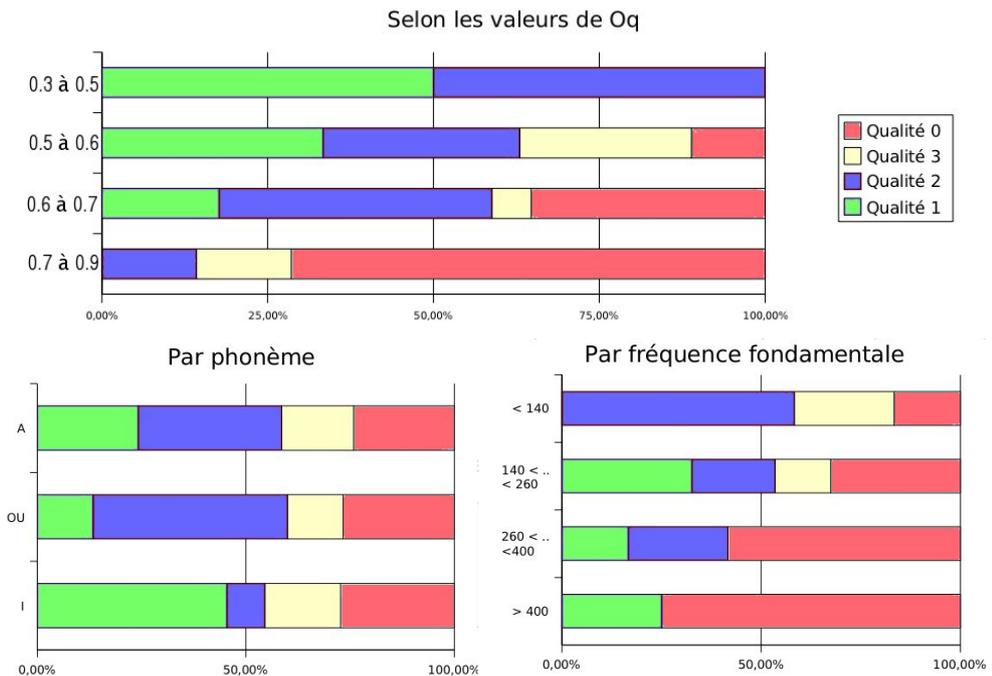


FIG. 9.4 – Résultats d’analyse par valeur de quotient ouvert, phonème et fréquence fondamentale

vont fiabiliser la séparation, il est à noter que le phonème A, de part sa complexité, est parfaitement adapté à une déconvolution.

Dans les autres cas, les résultats sont nettement plus aléatoires et vont principalement

dépendre de la qualité du signal et sa propension intrinsèque à être déconvolué. En effet, pour des phonèmes fortement résonants, et dans le cas des signaux analysés ici, la forme de signal est quasi sinusoïdal ce qui traduit une résonance non seulement très marquée, mais aussi confondant un formant vocalique et le formant glottique. Il est alors difficile de juger objectivement de la qualité de la décomposition quand les spectres obtenus pour la source et le filtre sont quasiment identiques.

En mécanisme II, les périodes fondamentales associées sont généralement très courtes, et rendent l'analyse difficile. Tous les signaux étudiés ici en mécanisme II présente un caractère très résonant et se déconvoluent donc difficilement.

9.3.2 Influence du quotient ouvert

Le quotient ouvert est fortement corrélé avec le mécanisme vocal d'émission, nous l'avons déjà vu précédemment. C'est pourquoi on retrouve que de fortes valeurs de quotient ouvert donnent de mauvais résultats d'analyse ; non pas spécifiquement en raison de la valeur elle même du paramètre, mais surtout car on retrouve plus souvent cette valeur du paramètre pour des sons à faible période fondamentale et prouvés moins faciles à analyser.

Par ailleurs, les faibles valeurs de O_q donnent de bons résultats car elles sont produites en stressant la glotte et résultent, non seulement en une diminution de O_q mais aussi dans une accentuation du formant glottique par rapport aux formants vocaliques. De plus, la production de tels signaux ne peut se faire généralement que pour de grandes périodes fondamentales.

9.3.3 Variation de α_m et interprétation physique

Sur certains échantillons nous avons essayé de faire varier l'effort de production vocale. En effet, α_m est un paramètre agissant sur la largeur de bande du formant glottique, une valeur élevée de ce paramètre va traduire une densité spectrale plus importante, assimilée à un effort de production vocale accru. On s'attend donc, lorsque le locuteur parle avec un effort, à une augmentation de la valeur du paramètre.

En ronds bleus sur la figure 9.5 nous voyons les valeurs prises par α_m et en vert le signal acoustique enregistré. Nous voyons aisément que lorsque le signal est plus fort, les valeurs du paramètre augmentent. Cette constatation permet de nous conforter dans l'idée que les mesures effectuées ont une réalité physique.

9.3.4 Qualité de la décomposition, commentaire des annexes

En annexes sont présentées quelques une des analyses considérées comme les plus caractéristiques et intéressantes. On y verra 4 graphiques, le premier étant lié à la mesure directe (O_q en rouge et $\frac{f_0}{F_g}$ en bleu, les deuxième et troisième étant les spectrogrammes issus de la décomposition et le quatrième donnant les valeurs de α_m estimées ; tout cela en fonction du temps. On y remarquera que arriver à estimer correctement α_m n'est pas nécessairement synonyme d'une bonne décomposition. En effet, notre mesure se concentre sur la partie basse fréquence du spectre ($f < 1600Hz$), et nous ne prenons aucune précaution, dans cette étude, pour obtenir des spectres cohérents dans les basses fréquences. Cependant, avec un petit effort on retrouve les formants caractéristiques sur les deux spectres.

9.3.5 Mesure directe de O_q par ZTT

Nous avons vu que la fréquence du formant glottique est fonction à la fois de O_q , T_0 et F_g . Mais à la vue des résultats de cette étude, on peut se demander si il ne serait pas possible de

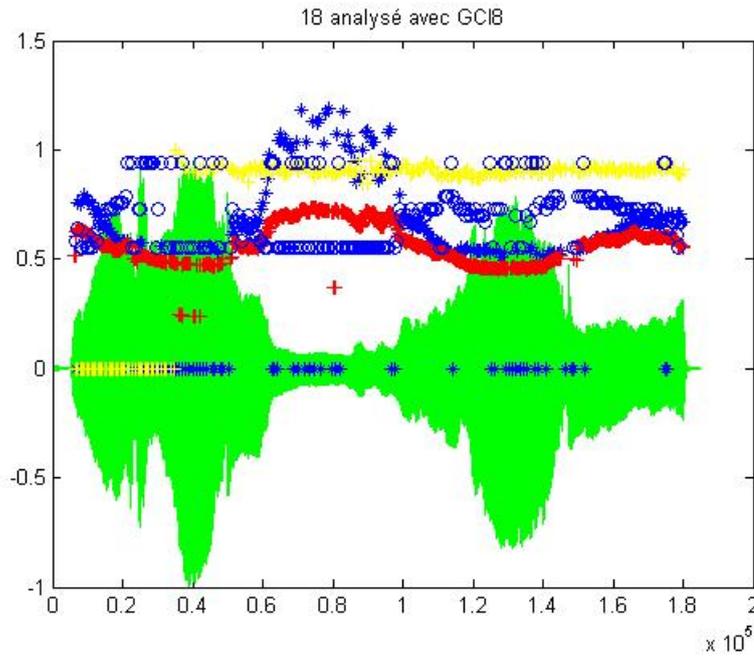


FIG. 9.5 – Variation de α_m avec l’effort vocal, sur l’échantillon C18. **En vert** le signal original. **En rouge** la mesure du quotient ouvert. **En bleu** les mesures effectuées, F_g représenté par les croix bleues, α_m par les ronds bleus. **En jaune** la variation relative de F_0 .

mesurer directement le coefficient ouvert par ZZT. En effet, pour des valeurs de α_m comprises entre 0.7 et 0.8, la position du formant glottique varie très peu (10 à 20%). En considérant une émission vocale normale (donc non extrême, supposant α_m variant dans la plage décrite précédemment) et une précision sur le quotient ouvert de l’équivalent d’un JND, il doit être possible d’arriver à estimer directement une valeur perceptive de O_q par ZZT.

Quatrième partie

Finalemment...

Chapitre 10

Conclusion

10.1 Bilan

Au cours de cette étude nous avons donc tenté de mesurer un nouveau paramètre de la source glottique par le biais d'une analyse par ZZT. Nous avons tout d'abord prouvé, par les travaux passés, qu'il était possible, en mesurant l'emplacement du formant glottique, d'exprimer une fonction des paramètres du modèle de la dérivée de l'ODG et, par le biais d'une mesure du quotient ouvert, d'arriver au coefficient d'asymétrie α_m .

Nous avons ensuite montré que la mesure était possible sur des signaux synthétiques, seule et unique base dont nous disposons pour tester la mesure.

Après une étude de la robustesse de cette mesure vis à vis des facteurs déterminants de la ZZT, nous avons alors constitué une base de données pour tester un nombre varié de situations de production vocale. Finalement nous avons analysé ces résultats afin de critiquer la dite méthode.

10.2 Critique

Si les résultats obtenus peuvent sembler faibles, il ne faut pas oublier que la mesure des paramètres de la source glottique est très problématique. Aucune équipe de recherche à ce jour ne peut se targuer de réussir avec précision une détermination des paramètres de la source glottique par déconvolution, et en ce sens la méthode présentée et étudiée ici réalise un travail conséquent.

Il n'existe aucune base de données marquée pour évaluer la précision d'une méthode de mesure du coefficient d'asymétrie, tout comme il n'a encore jamais été tenté de mesurer ce coefficient. Compte tenu de la qualité perceptive de ce coefficient, une précision de l'ordre de 10% semble être nécessaire pour effectuer une analyse-resynthèse, la méthode présentée ici prétend coïncider avec cette nécessité.

Par conséquent, même si le manque de repères fiables pour la mesure réalisée peut rendre le lecteur sceptique, le protocole expérimental mis en place met en valeur la cohérence des mesures avec le modèle et la réalisé physique. Ce travail permet donc d'ouvrir la voie vers une nouvelle approche de la mesure des paramètres de la source glottique, en combinant déconvolution (par ZZT) et mesures physiques de l'activité glottique par EGG.

Néanmoins, la ZZT reste une méthode aux fondations fragiles. Les présupposés à une bonne décomposition ne sont pas toujours atteints avec la précision nécessaire : ainsi il n'existe pas encore, à ce jour, de méthode fiable de détection du GCI, comme il n'existe pas à ce jour de méthode précise de calcul de racines de polynômes à grand nombre de coefficients.

10.3 Travail à venir

Les mesures réalisées ne sont que la preuve succincte que la ZZT peut être exploitée dans des buts bien précis d'analyse-synthèse. Il est évident que la méthode est loin d'être parfaite, et nécessite encore une grande quantité de travail. En outre, il serait judicieux de réaliser une analyse perceptive pour vérifier que les paramètres estimés sont corrects ; le temps manquant cruellement pour effectuer ce travail lors de ce stage.

Par ailleurs, la ZZT se pose comme une méthode d'analyse innovante, qui devrait aussi trouver sa place dans le domaine de la synthèse ou de la mesure, comme nous le montrons dans cette étude. Même si technologiquement, les moyens d'exploitation de la ZZT sont encore limités, les capacités de calculs ne cessent d'augmenter. Au jour d'aujourd'hui, on peut réaliser une décomposition dans un temps équivalent à 10 ou 20 fois le temps réel en programmation Matlab.

Apport personnel, Remerciements

Ce stage m'aura permis d'aborder le domaine à la fois vaste, complexe et passionnant qu'est l'analyse des signaux vocaux. En 3 mois et demi, il a été nécessaire d'assimiler un grand nombre de connaissances diverses, des bases mathématiques de la ZZT au fonctionnement mécanique sommaire de la glotte. J'ai aussi appris que dans ce domaine où les mesures de références sont rares sinon inexistantes, l'établissement d'un protocole expérimental solide est une étape indispensable dans la démarche de recherche. Apprendre à relativiser un résultat, chercher sa cohérence ou son incohérence et adapter une méthode développée sur de la théorie, à des signaux réels.

Merci à Christophe d'Alessandro pour son encadrement pendant ce stage ; sa sagesse et sa maîtrise du sujet *vocal* ont su me guider dans ce monde où chaque pas doit être mesuré et réfléchi pour ne pas être mal interprété. Merci à Boris Doval pour ses encouragements et son intérêt envers mon travail.

Merci à Michelle Castellango pour avoir donné de son temps pour la mise en place de l'expérience permettant l'enregistrement de la base de données qui a supporté ce stage ; et merci à Maeva Garnier et Christophe d'Alessandro de s'être prêtés au jeu pour le-dit enregistrement.

Table des figures

2.1	Fonctionnement du larynx (source web : education.yahoo.fr et RANKE, LULLIES, 1953)	8
2.2	Larynx en position de respiration et de phonation (<i>ref : thèse de Nathalie Henrich, figure 1.2</i>)	9
2.3	Illustration du modèle source-filtre (<i>ref : thèse de Nathalie Henrich, figure 2.1</i>)	10
3.1	Illustration du modèle LF de la dérivée de l'ODG (source cerma.stanford.edu)	11
3.2	Modèle du flux glottique. U_g : débit acoustique glottique; dU_g : dérivée du débit acoustique (source [5])	12
3.3	Spectre du modèle LF pour différentes valeurs de O_q à gauche, et différentes valeurs de α_m à droite. (source [7])	13
4.1	Principe de l'EGG (source, thèse de N. Henrich)	15
5.1	ZZT d'un AR1 pour $a = 1.2$ a) réponse impulsionnelle; b) spectre sur 1024 échantillons; c) représentation ZZT polaire	20
5.2	ZZT d'une portion de sinusoïde de fréquence réduite $\nu = 0.1$. a) réponse impulsionnelle; b) spectre sur 1024 échantillons; c) représentation ZZT polaire	20
5.3	Onde glottique : forme temporelle, spectrale, puis ZZT	21
5.4	Reponse impulsionnelle : forme temporelle, spectrale, puis ZZT	22
5.5	Période simulée d'un signal vocal : forme temporelle, spectrale, puis ZZT . .	22
5.6	ZZT d'un signal vocal : période à analyser (GCI au centre de la fenêtre), ZZT du signal, spectre de la source, spectre du canal	24
6.1	Déplacement des zéros de la composante continue sur un signal synthétique. En rouge, signal fenêtré, en bleu signal non fenêtré; en bas, position des zéros de la fréquence nulle.	27
6.2	Illustration de l'imprécision du calcul des racines	28
6.3	Valeurs prises par F_g pour ($f_0 = 160Hz$; $O_q = 0.6$) en fonction de α_m en bleu, rapport $\frac{f_0}{F_g}$ en rouge.	30
7.1	Influence de la détection de GCI pour un filtre vocalique faiblement résonant.	32
7.2	Détection de F_g accrochant sur un formant, l'absence de point sur le graphique du milieu nous indique qu'aucun point ne présente une erreur relative inférieure à 25%	33
7.3	Evolution de la mesure de F_g pour un signal réel, cas où la marge d'erreur sur GCI est grande. en haut signal, au milieu erreur par rapport à la mesure au GCI, en bas valeur des F_g estimés	34
7.4	Evolution de la mesure de F_g pour de la voix chantée en mécanisme II . . .	34

7.5	Cas où la mesure présente une erreur importante très vite lorsqu'on s'éloigne du GCI.	35
7.6	Influence du fenêtrage sur un signal synthétique, formant fortement résonant à gauche, faiblement résonant à droite. En rouge le signal original, en bleu le signal fenêtré. On voit que dans certains cas le fenêtrage donne une forme exponentielle croissante à la R.I. du canal vocalique.	36
7.7	Décomposition incomplète, spectres. en haut spectre estimé de la source ; au milieu spectre estimé du canal vocalique ; en bas spectre original en bleu, spectre recomposé en rouge. <i>Trame extraite de l'échantillon N9</i>	37
7.8	Influence du Rapport signal sur bruit. A droite un RSB de 18 dB, à gauche 12 dB. On voit que la méthode est encore efficace pour 18 dB, mais ne l'est plus à 12 dB	38
8.1	Population de la base de données	41
9.1	Organigramme du script d'analyse	42
9.2	Une analyse valide (échantillon N26). En haut : résultats de l'analyse, mesures de OQ en rouge, de Fg en étoiles bleues et de α_m en ronds bleu. En bas : un spectre de la source, obtenu par décomposition ZZT.	44
9.3	Résultats d'analyse par locuteur et mécanisme	45
9.4	Résultats d'analyse par valeur de quotient ouvert, phonème et fréquence fondamentale	45
9.5	Variation de α_m avec l'effort vocal, sur l'échantillon C18. En vert le signal original. En rouge la mesure du quotient ouvert. En bleu les mesures effectuées, F_g représenté par les croix bleues, α_m par les ronds bleus. En jaune la variation relative de F_0	47
10.1	Analyse de l'échantillon N9, fréquence fondamentale : 191Hz, phonème [A] . .	56
10.2	Analyse de l'échantillon N20, fréquence fondamentale : 187Hz, phonème [OU]	57
10.3	Analyse de l'échantillon N33, fréquence fondamentale : 125Hz, phonème [OU]	57
10.4	Analyse de l'échantillon C20, fréquence fondamentale : 130Hz, phonème [OU]	58
10.5	Analyse de l'échantillon C20, fréquence fondamentale : 128Hz, phonème [A] .	58
10.6	Analyse de l'échantillon M1, fréquence fondamentale : 243Hz, , phonème [A] .	59
10.7	Analyse de l'échantillon M17, fréquence fondamentale moyenne : 191Hz . . .	59

Bibliographie

- [1] B. Bozkurt, B. Doval, C. d'Alessandro, and T. Dutoit. Appropriate windowing for group delay analysis and roots of z-transform of speech signals. *eusipco*, 2004.
- [2] B. Bozkurt, B. Doval, C. d'Alessandro, and T. Dutoit. A method for glottal formant frequency estimation. *icslp*, 2004.
- [3] B. Bozkurt, B. Doval, C. d'Alessandro, and T. Dutoit. Zeros of z-transform representation with application to source-filter séparation in speech. *IEEE*, 12 :344–347, 2004.
- [4] B. Bozkurt, B. Doval, C. d'Alessandro, and T. Dutoit. Zeros of z-transform (zzt) decomposition of speech for source-tract separation. *icslp*, 2004.
- [5] Baris 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.
- [6] B. Doval, C. d'Alessandro, and N. Henrich. The voice source as a causal/anticausal linear filter. *VOQUAL'03, Geneva*, 08 2003.
- [7] B. Doval, C. d'Alessandro, and N. Henrich. The spectrum of glottal flow models. *to be published in acta acustica*, 2006.
- [8] P. Fabre. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation. *Bull. Nat. Med.*, 1957.
- [9] G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, Netherlands, 1960.
- [10] G. Fant. The lf-model revisited. *STL-QPSR*, 2-3 :119–156, 1995.
- [11] N. Henrich, C. d'Alessandro, B. Doval, and M. Castellengo. Glottal open quotient in singing : Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *JASA*, 117(3) :1417–1430, 03 2005.
- [12] N. Henrich, Gunila Sundin, Daniel Ambroise, C. d'Alessandro, B. Doval, and M. Castellengo. Just noticeable differences of open quotient and asymmetry coefficient in singing voice. *JOV03*, 17 :481–494, 2003.
- [13] Nathalie Henrich. *Etude de la source glottique en voix parlée et chantée*. PhD thesis, Université Paris VI, 2001.
- [14] J. Liljencrants., Q. Lin, and G. Fant. A four-parameter model of glottal flow. *STL-QPSR*, 4 :1–13, 1985.
- [15] J. Makhoul. Linear prediction : A tutorial review. *Proceedings of the IEEE*, 63(4) :561–580, 04 1975.
- [16] the MathWorks Inc. roots (matlab function). <http://www.mathworks.com/.../ref/roots.html>, 05 2006.
- [17] Eric W. Weisstein. Cyclotomic polynomial. *From MathWorld—A Wolfram Web Resource.*, 05 2006.

Annexes

Résultats d'analyse de la base de données

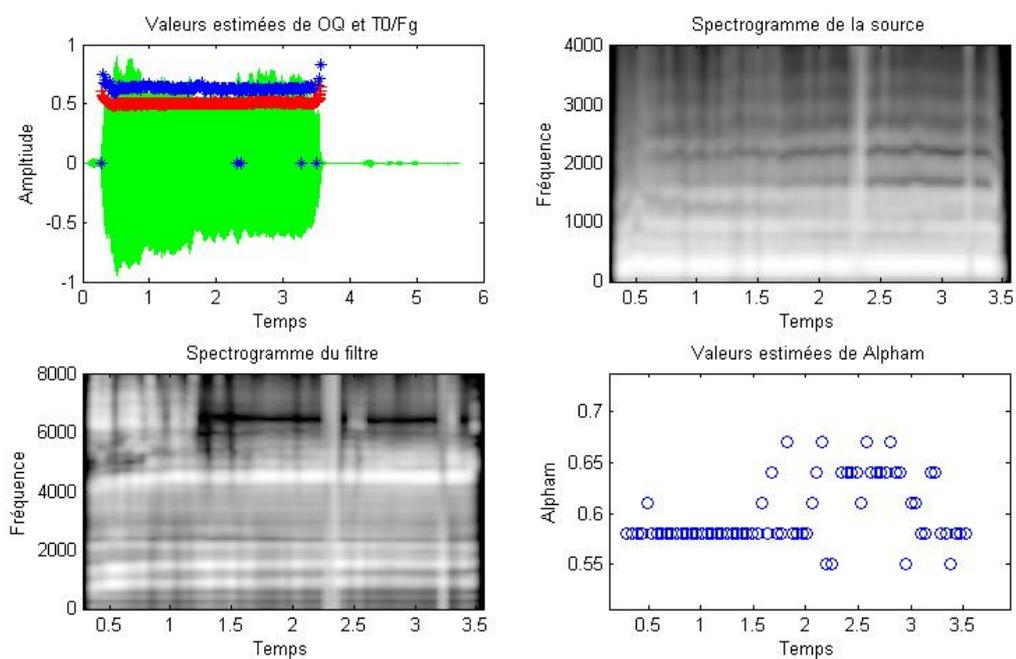


FIG. 10.1 – Analyse de l'échantillon N9, fréquence fondamentale : 191Hz, phonème [A]

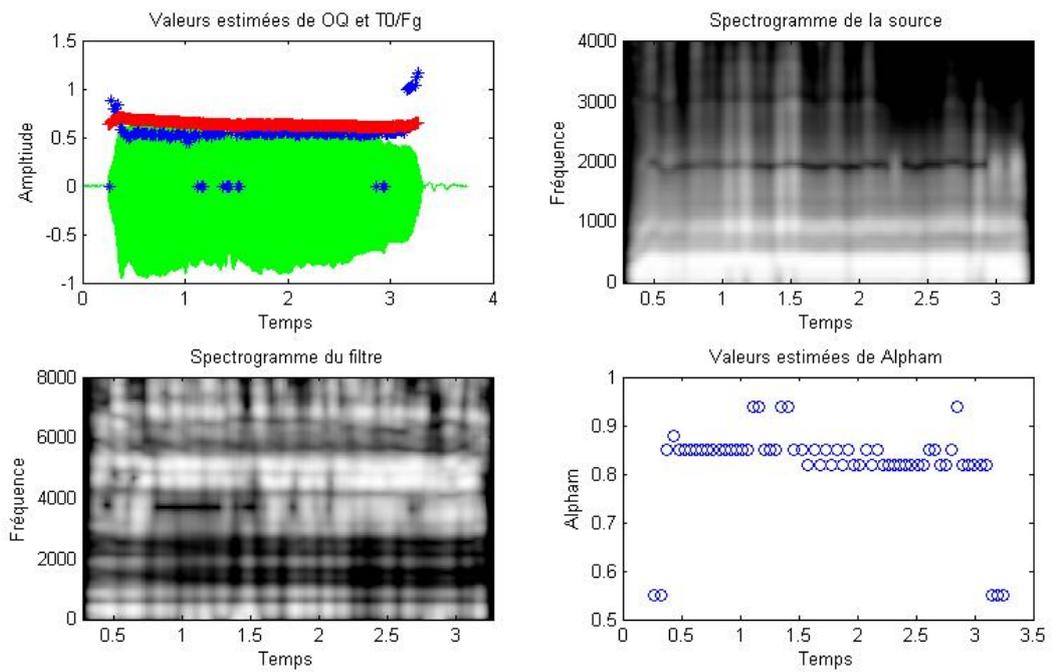


FIG. 10.2 – Analyse de l'échantillon N20, fréquence fondamentale : 187Hz, phonème [OU]

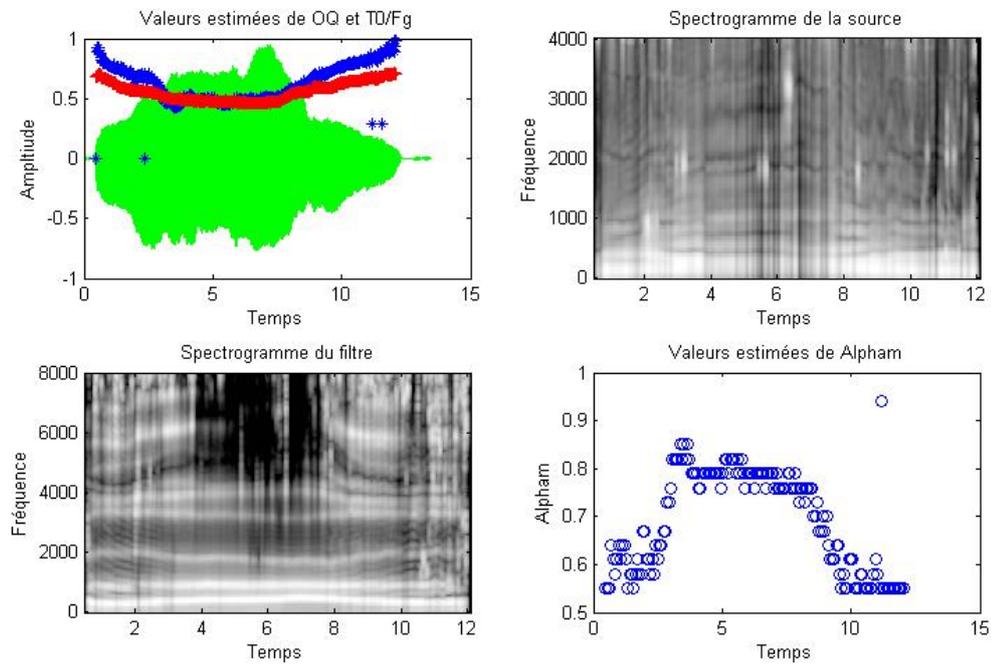


FIG. 10.3 – Analyse de l'échantillon N33, fréquence fondamentale : 125Hz, phonème [OU]

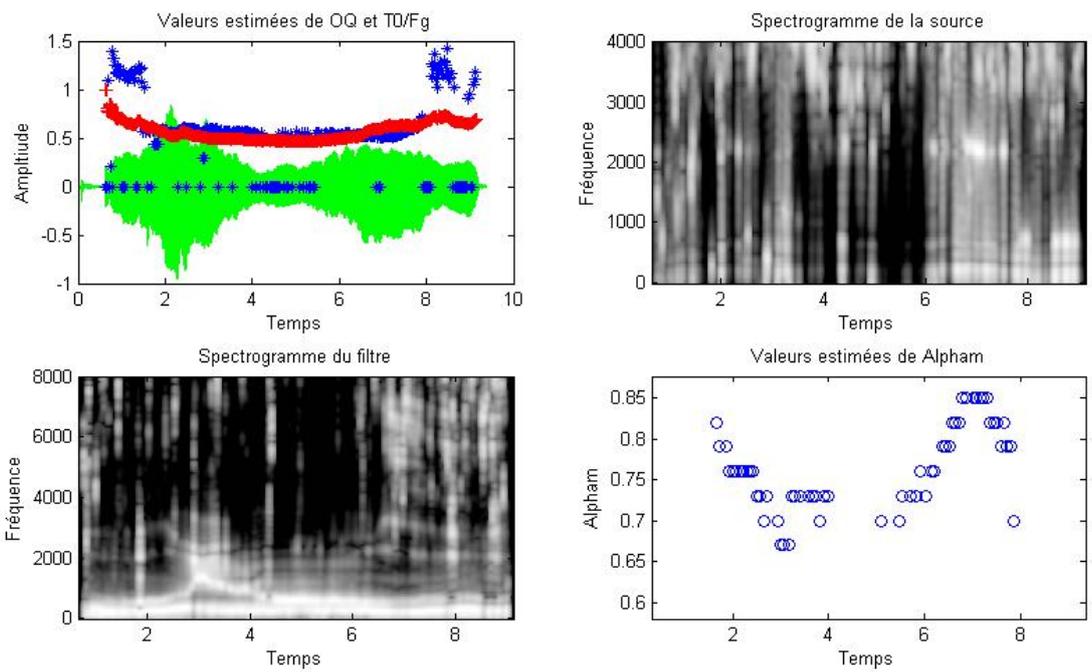


FIG. 10.4 – Analyse de l'échantillon C20, fréquence fondamentale : 130Hz, phonème [OU]

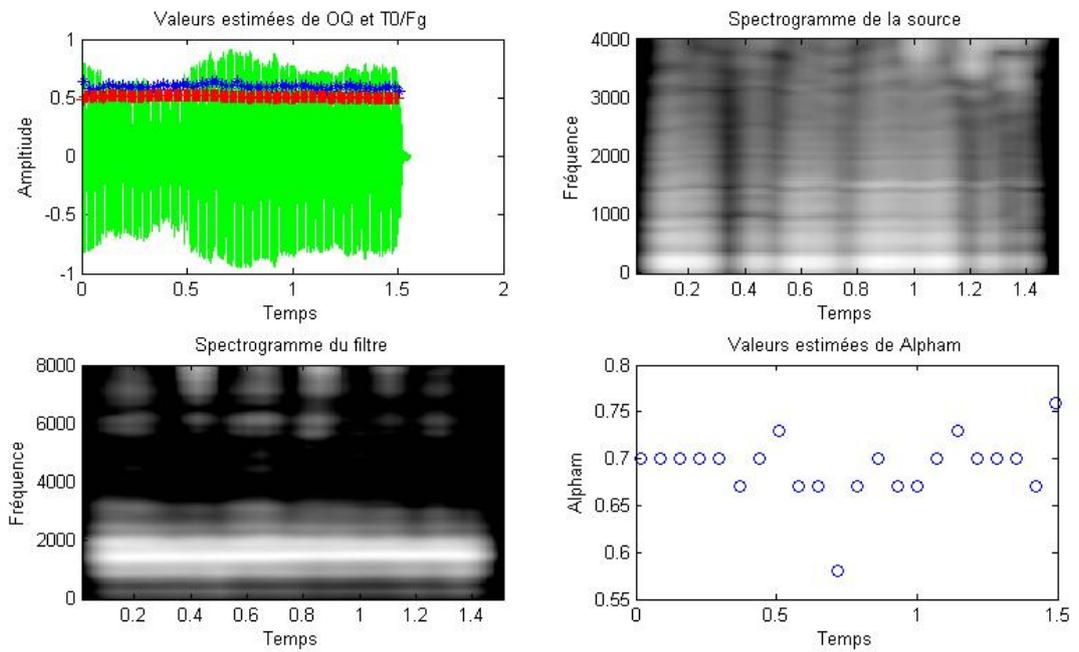


FIG. 10.5 – Analyse de l'échantillon C20, fréquence fondamentale : 128Hz, phonème [A]

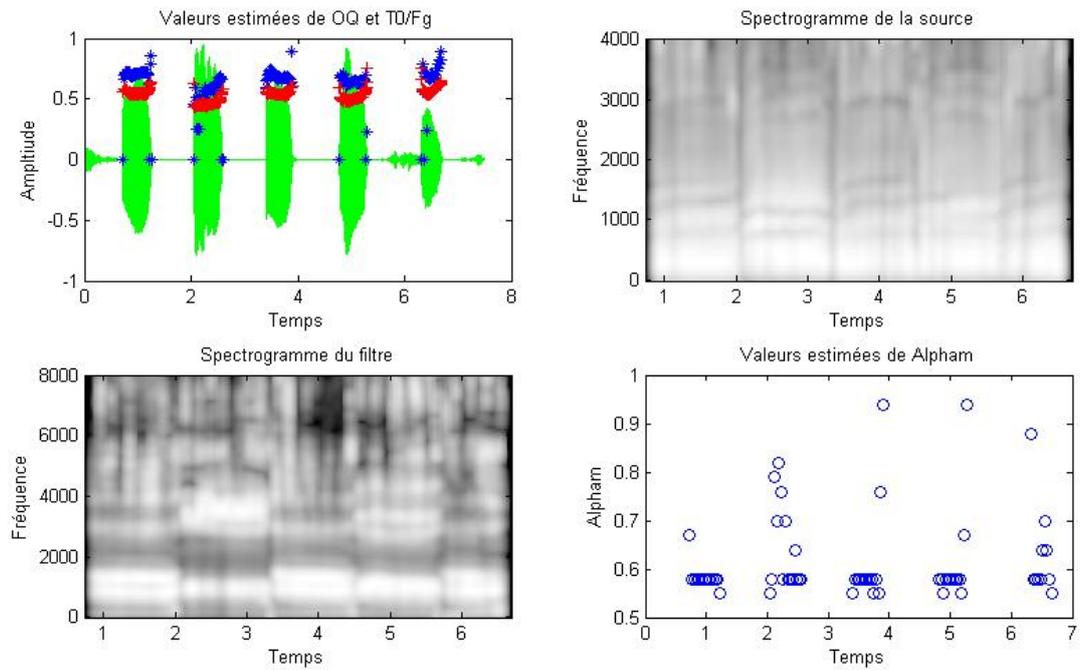


FIG. 10.6 – Analyse de l'échantillon M1, fréquence fondamentale : 243Hz, , phonème [A]

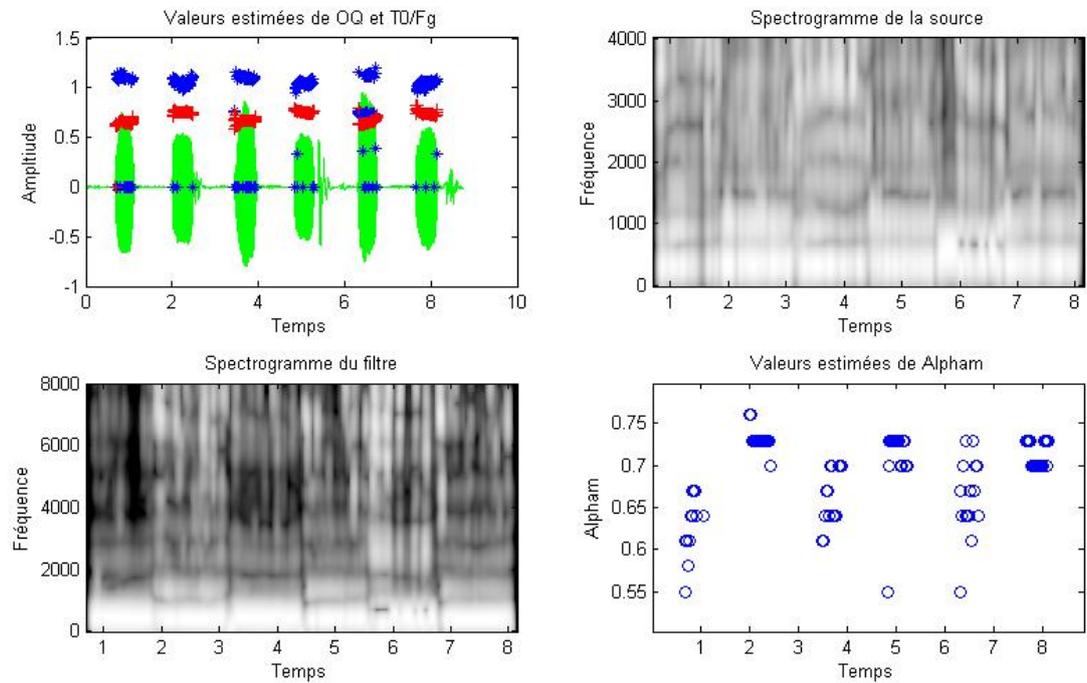


FIG. 10.7 – Analyse de l'échantillon M17, fréquence fondamentale moyenne : 191Hz